ANALYSIS OF SPEECH FEATURES AS POTENTIAL INDICATORS FOR DEPRESSION AND HIGH RISK SUICIDE AND POSSIBLE PREDICTORS FOR THE HAMILTON DEPRESSION RATING (HAMD) AND BECK DEPRESSION INVENTORY SCALE (BDI-II)

By

Nik Nur Wahidah Nik Hashim

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfilments of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Electrical Engineering

May, 2014

Nashville, Tennessee

Approved:

Dr. Mitchell D. Wilkes

Dr. Ronald M. Salomon

Dr. Alfred B. Bonds

Dr. Daniel J. France

Dr. Asli Ozdas Weitkamp

Dr. Richard A. Peters II

ACKNOWLEDGEMENT

Foremost, I would like to express my sincere gratitude to my supervisor, Dr. Mitch Wilkes. This work would not have been possible without his guidance, support and encouragement. Under his guidance, I successfully overcame many obstacles and learned a lot. I am extremely lucky to have him as a supervisor who cared so much about my work, and was always prompt in responding to my questions and inquiries. His genuinely good character always shines, inspires and keeps me motivated to finish this thesis.

My sincere thanks also go to Dr. Salomon who took time out of his busy schedule to provide valuable feedback on the thesis and shared his immense knowledge relating to psychiatry. His quick response to my infinite inquiry was also greatly appreciated. I am also very grateful to The Vanderbilt University Department of Psychiatry for providing valuable database for us to use in this study.

Special thanks to my thesis committee Dr. A.B. Bonds, Dr. Dan France, Dr. Asli Ozdas and Dr. Alan Peters for their support, insightful comments and helpful suggestions.

Finally, I wish to thank my parents, Nik Hashim Nik Pa and Rohani Mohd Musa for their continuous prayer and unconditional love that had thus became my driving force to complete this work. To my dear and loving husband, Wan Ahmad Hasan Wan Ahmad Sanadi, who had always believe in me, calms me with words of motivation and provides encouragement during tough times. Not to forget, a heartfelt love to my baby, Wan Abdurrahman for showing me his wonderful smiles despite still trying to understand the world around him. I can only ask The Best of Rewarder to reward them immensely.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

When talking about violence, majority of people often relate it with homicide, war or abuse. Acts of fatal and non-fatal suicide attempts and suicidal ideation are also considered to be a self-directed violence. In the United States, the rate of suicide has been steadily increasing every year since the year 2000 [77]. The current available summary by the National Center for Health Statistics reported an increment of 1.7 percent from 2008 to 2009. The list of cause of death revealed that suicide ranks in 10^{th} with 36,909 numbers of suicides and ranks in 3^{rd} for the age group of 15-24 with 4,371 numbers of suicides reported in the year 2009 [78]. To view the importance of this issue, on average one suicide occurs every 14.2 minutes in the United States. On average, one young person dies in suicide every two hours. Non-completed suicide attempts numbered 922,725 during this interval, translating to an average of one attempt every 34 seconds. Male exhibit a greater risk of death from suicide as a gender wise analysis reported a ratio of 3.7 male to 1 female by suicide [79]. A surge in the military suicide rate with 154 deaths in the first 155 days of 2012, an increase of 18 percent compared to the statistic reported at the first half of the previous year. Deaths by suicide among military personnel during the period outnumbered those U.S. soldiers killed in action by an estimated of a two to one ratio [80]. Suicide does not only cost emotional consequences for family and friends, but there are also substantial economic costs of approximately $34.6 billion associated with medical and work loss [77].

Prediction and identification of suicidal risk as opposed to major depressive illness is a complex task that requires clinicians to use specific interviewing approaches, sometimes relying on their intuition, to deploy the skills they develop through proper knowledge, training and experience. Despite decades of research, accurate prediction of suicide and imminent suicide attempts still remains elusive. There are no reliable objective methods to assist with clinical assessment that have been empirically tested [81, 82]. Identifying suicidal predisposition at an early stage is essential in order to identify acute and appropriate treatment for each unique patient. Inaccurate assessment may mislead clinicians to believe that patients who are actually at

imminent risk of committing suicide are experiencing a less severe psychological disorder and thus putting the lives of patients at risk. An alternative assessment tool may assist non-specialists with no proper training in psychiatry in providing objective metrics that might signal a need for extensive interviewing and provide precautions to the patients. On the other hand, clinicians with advanced psychiatric training may also benefit by the information from a second source that can give quantitative results and thus yield a better identification of a near-term suicidal predisposition.

Speech is a rich source of information which people use to express ideas and communicate. Aside from the physical speech information, speech also contains implicitly hidden information that reflects psychological states [11], [31], [34]-[47], [51]-[54], [64]-[65]. Previous studies have suggested that depression is associated with distinctive speech patterns. Among the characteristics are decreases in intonation, phonation stress, loudness, inflection and intensity, increase in duration of speech, sluggishness in articulation, narrow pitch range, monotonous, and lack in vitality [32], [41]. These characteristics correlate with the disturbances occurring in the respiratory, laryngeal, resonance and articulatory system which is then embedded in the acoustic signals. The past few years have seen growing interest in research which uses speech to identify psychological disorders, Parkinson disease, stress, emotions and affective states. It is considered worthwhile for researchers to investigate actively with the problem of describing these conditions rather than depending solely on solutions that are currently ready-made from other disciplines. Speech researches should be able to make distinctive contributions into the mainstream research revolving those fields.

This research focuses on the psychological disorders, particularly on two distinct groups of suicidal predisposition and major depression. The investigations of speech within the area of depressive disorders are more widely studied compared to suicidality. Marilyn and Stephen Silverman [62] have proposed and explored the effect of suicidality in speech. The knowledge on how acoustic parameters of speech are modulated when a patient begins to experience an imminent risk of suicidal as opposed to only experiencing a major depression have been accumulated for over 20 years and still it continues. The future of this research hopes to develop a diagnostic tool that could be used by trained clinicians to provide quantitative measures during a psychological assessment and for assisting the primary care physicians to determine whether to send the patient to a psychiatrist.

The main purpose of this research is to identify possible voice parameters that can be used as an indicator for near term suicidal risk and depression. Previous studies relating to investigation of the vocal cues for depression and imminent suicidal risk detection often revolves around spectrum-based measures of the voice signal [51]-[54]. As an alternative of looking at a precise frequency distribution which can be influence by the nature of the microphone or the room, this research also aims to discover feature that is a robust approach toward changes in recording devices and environments. The success of identifying such feature will allow a more practical and robust application in real world. This research also attempts to demonstrate the effectiveness of using acoustic measurements as a possible means to predict ratings from well-known medical diagnostic tools known as the Hamilton Depression Scale (HAMD) and Beck Depression Inventory (BDI-II).

The work presented in this thesis contributes to the findings of features relating to pauses and phonation (timing based measures) in speech that were extracted using a new approach. When attempting to distinguish between the groups of high risk suicidal and depressed patients, these features were identified to be robust across two different male populations and yielded an effective classification performance within each population of male and female patients using only at most two combined features. Secondly, the thesis also demonstrated that the regression analysis of the HAMD and BDI-II score by means of speech measures has successfully predicted the well-known clinical diagnostic ratings with minimal error. To our best knowledge, this is the only work that has extensively studied the ability of speech measure in predicting the clinical scores.

The research is divided into three major studies and a small-scale study. Following this introductory part is the Chapter II which presents a general overview of the joint mechanism of speech production and perception and also a background overview of several studies relating to the area of speech and psychopathology. Chapter III explores the ability of the timing based measures in distinguishing between the groups of high risk and depressed patients. Chapter IV demonstrates the regression analysis of HAMD and BDI-II ratings with the speech acoustic measures. Chapter V presents another classification analysis which is a partial replication and extension work that was done by France [IEEE T-BME 47(7) (2000)] concerning the investigation of root mean squared Amplitude Modulation feature in identifying the imminent suicidal patients and the depressed patients. Chapter VI presents a small-scale study on the

ability of the Power Spectral Density (PSD) to demonstrate the significance of separation between the groups of high risk and depressed patients and to examine the significant improvements in patients' mental condition after a few days of receiving treatments. Finally, Chapter VI summarizes the work performed in this research and presents suggestion for future work.

CHAPTER II

BACKGROUND AND SIGNIFICANCE

1.0 Introduction

It is beyond the scope of this thesis to provide an in-depth discussion on the acoustic analysis of speech. However, this chapter is designed to provide a useful introduction to the wide range of important concepts throughout this work. This chapter also provides the essential details to equip the readers with basic knowledge of understanding the transformation of the dynamic process of speech into a quantitative form for detailed analysis of speech within the scope of this study. The chapter begins with a general overview of the joint mechanism of speech production and perception. Then, a brief explanation of speech production mechanism and the model of speech production will be introduced in order to develop an understanding of the analysis of speech for information extraction. The latter part of the chapter provides a background overview of several studies in the area relating to the association between acoustic properties of speech with psychopathology and Schizophrenia. Focus on discussion is given to acoustic features that are correlated with depression and suicidal speech and acoustic features that are pertinent to studies of speech in suicidal and depression are reviewed.

2.0 Speech Communication Mechanism

Speech is a complex process of transmitting information from the speaker's brain to the listener's brain. This process is generally referred to as speech chain [1]. The speech chain divides the entire process of direct information transfer into five stages as shown in figure 2.1.

The speaker will produce a linguistically meaningful speech by arranging his/her thoughts and organizing in linguistic form by combining words and phrases according to the grammatical rules in the language. This linguistic level takes place in the speaker's brain. The information is then conveyed in the physiological level through the nervous system and articulatory muscle movement. The brain provides control by sending impulses to the muscles that are involved in speech production. The information is being *modulated* onto the *carrier* in the form of acoustic wave that is produced during speaking as it travels from the mouth (and in

some cases through the nose) to the ear. Gathering information at the acoustic level is the most accessible element for practical applications because the acoustic wave can be captured by using a microphone and converted into a digital form. At the listener's side, the process of physiological and linguistic is reversely repeated. The transferred information is intended to be heard and understood by the listener. At the physiological level, the acoustic wave impinges on the eardrum and activates the physical auditory system. The content of the speech is decoded by the listener's active cognitive interpretation of the signal during linguistic level.

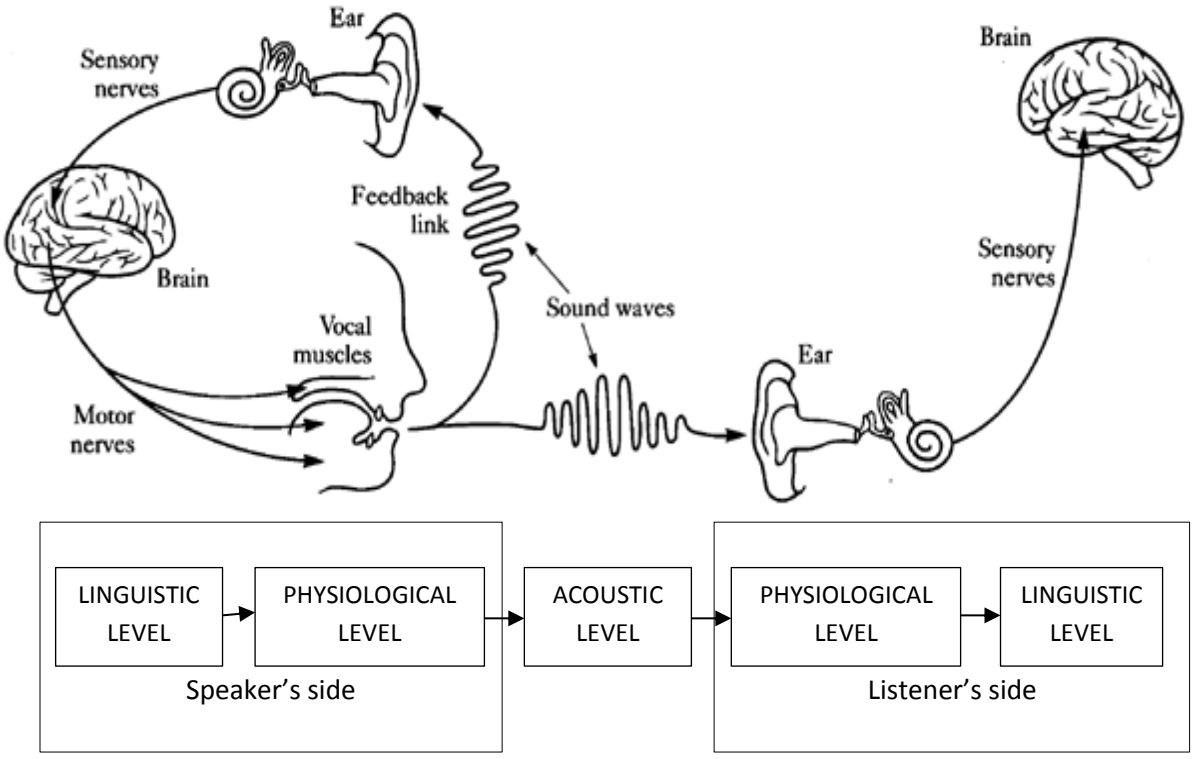


***Figure 2.1**: The speech chain [1]*

In parallel to the speaker-listener information transfer, the speaker also acts as a listener. The speaker controls and modifies the produced speech sound in order to make appropriate sound according to their intentions. This notion of shared knowledge between the speaker and the listener creates a feedback loop from the speaker's acoustic signal of speech to the speaker's brain. Thus, speech and hearing together function like a closed-loop system.

3.0 Linguistic Basis of Speech

Linguistic utterance that comprises of words, phrases and sentences are generally considered to be discrete unlike speech in acoustic signals, is continuous. Basic language unit in speech that has linguistic distinction of meaning and a unique set of articulatory gestures is called a phoneme. A phoneme can be divided into groups of vowels and consonants. Combination of phonemes makes up a syllable and sequence of phonemes and syllable create a larger language unit called a word. Mixture of words based on grammatical structure of the language creates a sentence. Besides phoneme identities, other carriers of linguistic information in speech include the prosodic features of speech such as timing, stress and intonations. Although these features do not alter the meaning of the word, however, they provide additional useful information about what is being said. Differences in prosody may affect the grammatical functions such as turning a question into a statement. Prosody also portrays attitudes of the speaker. Together, they form a linguistic basis of speech. A detailed discussion on phonemics and phonetics can be found in [2].

4.0 Speech Production

This section explores the broad outline of the physiological method of describing the anatomy of human speech production and recognizing relevant anatomical structures with regards to speech production. Emphasis is given on the important role of the anatomical structures in the process of formulating the speech production model and identifying its parameters.

4.1 Anatomical and Physiological Characteristic of Speech Production

Figure 2.2 shows the left cross-sectional view of the upper portion of a human anatomical structure involved in the production of speech. Three major components of the speech mechanism are respiratory system, laryngeal (vibration) system and articulatory system.

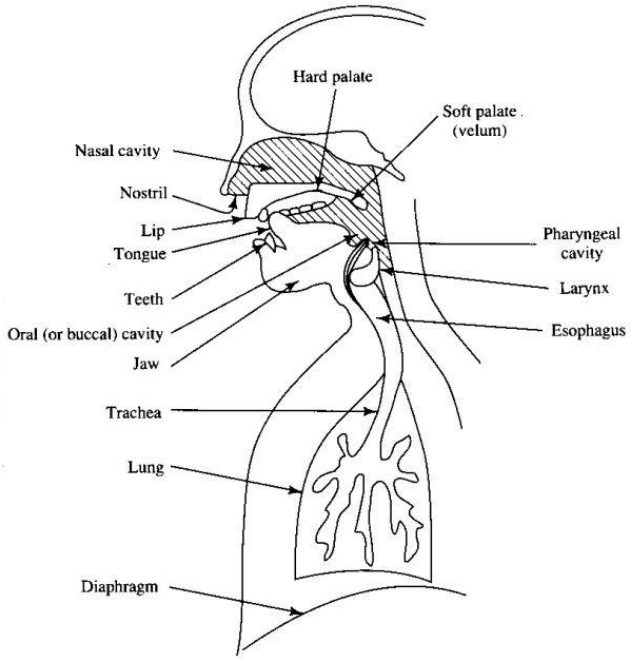


Figure 2.2: *Cross-sectional view of an anatomy structure for human vocal production [3]*

Figure 2.3 shows a simplified block diagram of the speech production process. The lungs provide airflow. Muscle force pushes air through trachea, bronchi, glottis (located between the vocal chords and larynx) and finally, into the three main cavities consisting of the vocal tract, the pharynx, and the oral and nasal cavities.

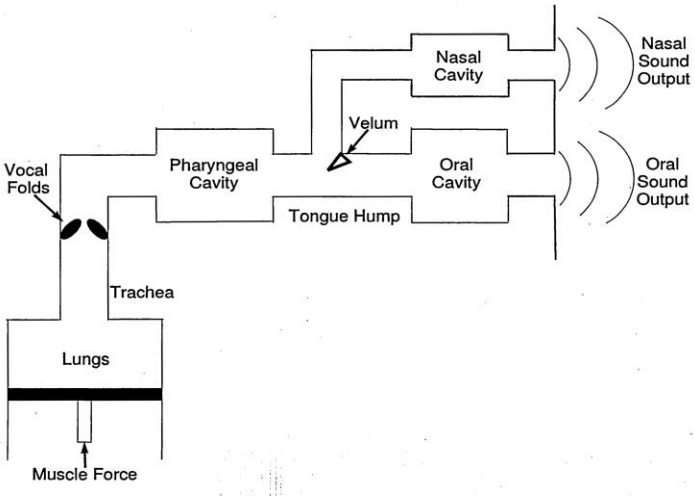


Figure 2.3: *Simplified speech production model [4]*

4.1.1 Respiratory system

Respiratory system comprises of lungs, bronchi, trachea and other associated muscles. During breathing, air is sucked into the lungs (inspiration) and squeezed out of the lungs (expiration). During speaking, inspiration of air moves rapider and the process of expiration is lengthened. The system acts as the main energy source by supplying air and responsible for the amplitude of the sound as the displacement of vocal chords changes with respect to air flow energy. The amount of air that is being inspired and the amount of pressure given during expiration determines the characteristic of the speech such as total duration, loudness, stress pattern, and pauses components [6].

4.1.2 Laryngeal System

The laryngeal system consists of a tabular structure muscles and cartilages found on top of the trachea and is called the larynx. The larynx functions as a source generator. A smaller muscular valve that is part of the larynx is formed by the vocal chords (vocal folds) and the space in-between the vocal chords is called the glottis. The vibrations of vocal chords convert air pressures and flows from the respiratory system into sound waves.

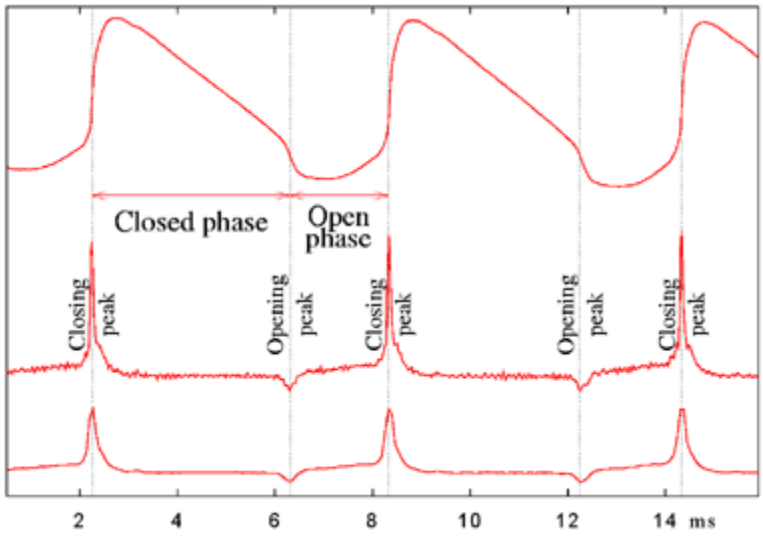


Figure 2.4: *EGG signal corresponding to opening and closing of the glottis (top), DEGG Derivative of the signal (middle), smoothed DEGG (bottom) [7]*

Three major types of sound source produced by the air stream from the lungs are voiced sounds, unvoiced sounds and glottal stop. Voiced sounds include some consonants such as [‘m’, ‘n’, ‘w’] and all vowels. Unvoiced sounds include consonants that are excited primarily by air turbulence such as [‘s’, ’h’] and is known as fricatives. Glottal stop sounds are plosive sounds that are produced by the blocking of glottis, tongue, lips or nasal.

Voiced sounds are formed through the vibration of the vocal chords. Its input can be modeled as a quasi-periodic excitation at the glottal passage caused by the opening and closing of the glottis. Figure 2.4 shows an electrogastrogram (EGG) recording of electrical signals that travel through the glottis. The derivative of EGG (DEGG) produced an alternating positive and negative peak, where positive peak corresponds to the immediate closing of the glottis and the negative peak corresponds to the opening of the glottis which is represented by the steep fall in EGG signal [7]. Through a process called adduction, Bernoulli force causes the vocal cords to be brought together and creates a closed air space under the glottis thus, provisionally blocking the air flow from the lungs. This process leads to an increase in sub-glottal pressure when air pressure from the lungs continues to build up below the vocal cords. Figure 2.5 represents the schematic representation of the vocal tract model when the vocal chords are closed. Once this pressure becomes greater than the resistance of the vocal chords, the vocal cords re-open and release a single waft of air. Due to elasticity, laryngeal muscle tension, and Bernoulli effect, the vocal chords rapidly close to its original position. This process is sustained by a continuous supply of pressurized air in a quasi-periodic manner. The cycle continues until thousands wafts of air are released and filtered through the vocal tract thus, producing sounds [8]. The fundamental period (F_0) or pitch period (T_0) corresponds to the time between consecutive vocal chords cessations (frequency pulses).

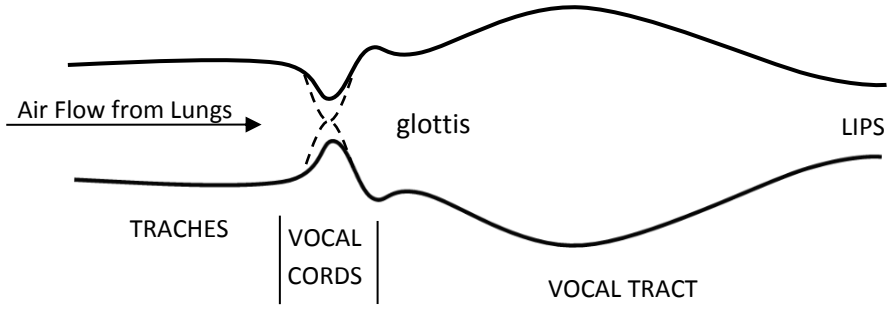


***Figure 2.5**: Vocal tract organ pipe model [5]*

Unvoiced sound has more of a noise-like quality. A narrow passage opening that exists between partially adducted vocal chords causes constrictions to the steady air flow that passes through the passage, thus, causing turbulence or noise-like behavior in the air flow. The developed sound typically exhibit smaller amplitude and faster oscillation compared to voiced sound.

When vocal chords are adducted more firmly, air pressure builds beneath the folds. This thus creates a short explosive burst of air as the vocal chords are quickly released. The gentle pop sound is perceived as the glottal stop. Turbulent noise usually follows the release of the constriction.

4.1.3 Articulatory System

The primary resonant structure is known as the vocal tract. Vocal tract starts at the larynx and extends through the pharynx (main resonating chamber for the voice), oral cavity and nasal cavity [3]. The vocal chords vibrate and create the opening and closing of the glottal passage. The glottis is the opening of the vocal tract where pulses beginning to be filtered. Opening and closing of the vocal chords periodically controls air in a quasi-periodic manner. The vocal tract can be viewed as a filter that applies spectral shaping to the pulses produced and this shaping varies over time. The generated pulses are considered to be a quasi-stationary process where static parameters of speech remain reasonably constant over short time intervals, typically 10ms to 30ms. The vocal tract has natural frequencies and is frequently described in terms of its resonances frequency (formants, F_i). Formants represent the spectral peaks of acoustic energy around a particular frequency in the speech wave depending on the shape of the vocal tract. F_1 is often observed as the strongest formant because as the frequency increase, the power decreases due to the low-pass nature of glottal excitation [9].

The movable articulator structures include velum (or soft-palate), jaw, lips, tongue and teeth. These articulators shape the vocal tract and alter the speech into a comprehensible utterance called speech. With air flowing through the vocal tract, the articulation of the velum is used to produce constriction. Lowering the velum causes air from the vocal tract region up to the lips to be restricted thus, allowing more openings towards the nose passage. For voiced speech, the velum is articulated upward, thus causing the nasal passage to be blocked temporarily in order for sound to be produced through the lips. The larynx functions as the airflow regulator

into the vocal tract, which causes the formant frequencies to increase or decrease by altering the tract length via raising or lowering the larynx [9]. Besides the larynx and velum, the tongue and lips are the two other major organs in an articulatory system which produces various sounds.

4.2 Models of Speech Production

Factors affecting the spectral structure of a vowel are (1) vocal excitation, (2) vocal tract transfer function and (3) transmission characteristics (i.e. lips radiation and room acoustics). A conceptual representation of speech production is derived in order to extract important information from speech. According to source-filter theory, speech signals can be viewed as a glottal source excitation followed by a linear time-varying filter that shapes the resonance characteristic of the vocal tract. So, the radiated speech signal is a product of the source energy (source) and the resonator (filter). As represented in figure 2.6, the assumption of speech production as a linear process is an oversimplified model of a more complex model and is further described in [5]. Even so, such an approximate model permits an examination of the effects of glottal excitation and vocal tract independently. Therefore, modification of the properties in the vocal tract will not affect the properties in the source excitation and vice versa.

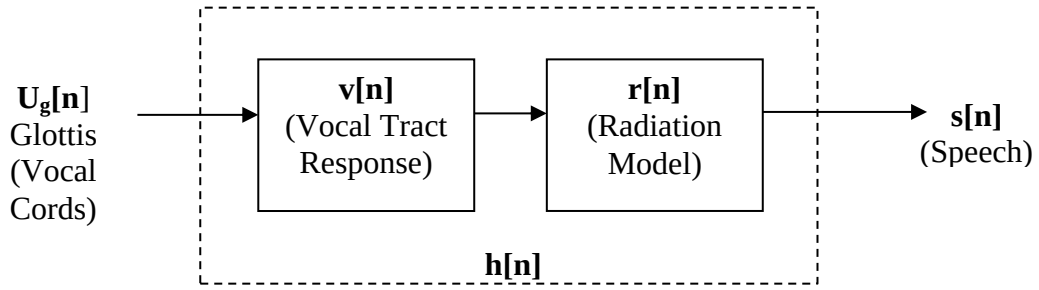


***Figure 2.6**: Linear filter of a voice production model*

In most speech analysis, the main focus would be on the voiced part of the speech. Voiced speech can be defined to be the convolution of the input waveform with its impulse response in time-domain. Referring to figure 2.7, voiced speech is modeled as periodic pulses of air-flow with a desired fundamental frequency that is shaped by the glottis represented by $U_g[n]$. The glottal shaped pulse passes through a pulse shape modifier with a tube-like passage way

called the vocal tract that is represented by $v[n]$. Finally, the produced sound is emitted to the surrounding air through radiation (lips), $r[n]$. For mathematical purpose, $v[n]$ and $r[n]$ can be grouped together and represented as $h[n]$. Also, radiation (lips) only plays a role in shaping the quasi-periodic train of glottal pulses.

$$S(f) = Ug(f)V(f)R(f) \tag{2.1}$$

The voiced speech in terms of sound pressure spectrum can also be viewed as multiplying the input spectrum by its frequency response which can be represented in a function of frequency as shown in equation 2.1. The source-filter model allows the modeling of speech production as a linearly separable filter. In this acoustic system, the vocal tract is assumed to be approximately linear by disregarding the effect of vibrating walls or external radiation [9], thus allowing it to be characterized as a frequency response or impulse response.

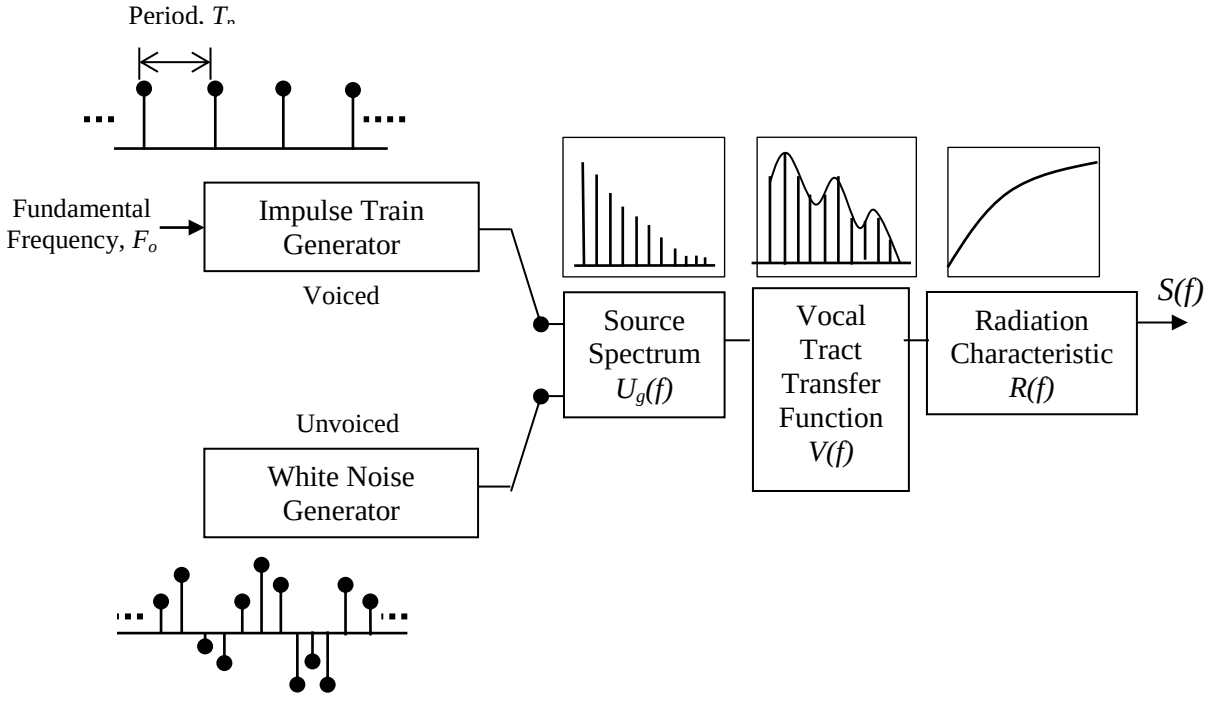


Figure 2.7: *Source-filter model of speech production*

4.2.1 *Excitation Models*

When modeling the source of voice production, there is a difference between the acoustic model and the source-filter model. In the acoustic model, the glottal flow is dependent on the vocal tract shape due to the acoustic load above the glottis that is defined by the output of the vocal tract. Meanwhile, assuming that the source-filter model is independent of the vocal tract shaping variations, the glottal source is defined as a non-interactive signal description of the voice source [10]. Implementation of this model inputs a random white noise for unvoiced sound and a discrete-time periodic impulse train with a certain fundamental period between each pulse that acts as the source excitation signal for voiced sound.

Voiced speech is considered to be non-stationary over a large interval of time but the characteristics and information in the voiced speech can be measured to be relatively constant over a short period of time. Similarly, the glottal pulse can be represented by Equation 2.2 [11] where $g[n]$ represents the discrete-time impulse train pulses and T_o is the fundamental period which can be represented in time and frequency domain as shown in figure 2.8.

$$g[n] = \sum_k \delta[n - k\,T_o] \tag{2.2}$$

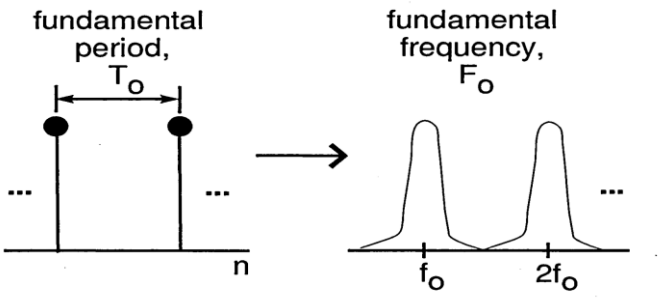


Figure 2.8: *Time and frequency domain representation of glottal pulses [11]*

4.2.2 Vocal Tract Models

The resonance frequencies varies according to the size of the vocal tract but are not affected appreciably by the shape of the vocal tract (i.e. straight or curve). Longer vocal tract corresponds to lower resonance frequencies and smaller separation in frequency. Therefore, straight pipes of different cross-sectional area as shown in figure 2.9 will serve the purpose for

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

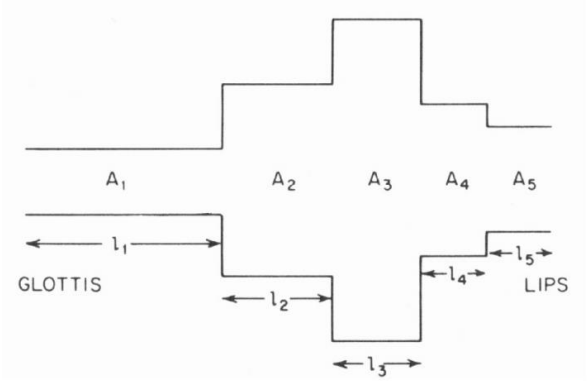


Figure 2.9*: Concatenation of lossless tubes for N = 5. [5]*

The output speech is related to the relationship between pressure and volume velocity which are determined by the cross-sectional area of the tube and the speed of air. At the joint of two tubes, continuity must be obtained in order to keep constant pressure on both sides as the waves traveling from one tube to the other. The excitation propagates through the series of tubes with some partially reflected and some waves partially propagated across the two joint tubes. Besides the joint of two tubes, boundary conditions at the lips and glottis must also be taken into account [5]. A linear prediction (LP) analysis involves the prediction of signal parameters based on the previous values and is a technique that is used to model the vocal tract as an all-pole filter called an inverse filter as shown in Equation 2.3 where $\{a_k, 1 \leq k \leq N\}$ are the predictor coefficient and the N order of the filter (number of poles).

$$V(z) = \frac{1}{A(z)} \quad and \quad A(z) = 1 - \sum_{k=1}^{N} a_k z^{-k} \tag{2.3}$$

When modeling the vocal tract by an all-pole filter, the nasal and unvoiced sounds are not taken into account. According to [12], inclusion of nasal and unvoiced sound into the current all-pole model can be achieved by including more poles rather than including zeros. All poles will remain inside the unit circle considering the areas of the concatenated tubes to be positive.

4.2.3 Lips Radiation

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

5.0 The Acoustic Signal

In speech, the acoustic wave is a process that takes place starting from the mouth to the ear. Figure 2.10 illustrates propagation of sound in the air. For speech sound generation, the source of energy is considered to be the exhalation of air from the lungs and vibrator to be the vocal chords. When the vocal chords are completely adducted, pressure builds up from beneath. Thus, rapid opening and closing of the vocal chords causes a series of compression and refraction of wave or fluctuation in air pressure in the surroundings. The ear picks up the pressure variation and transforms the pressure into vibration for the brain to interpret it as speech.

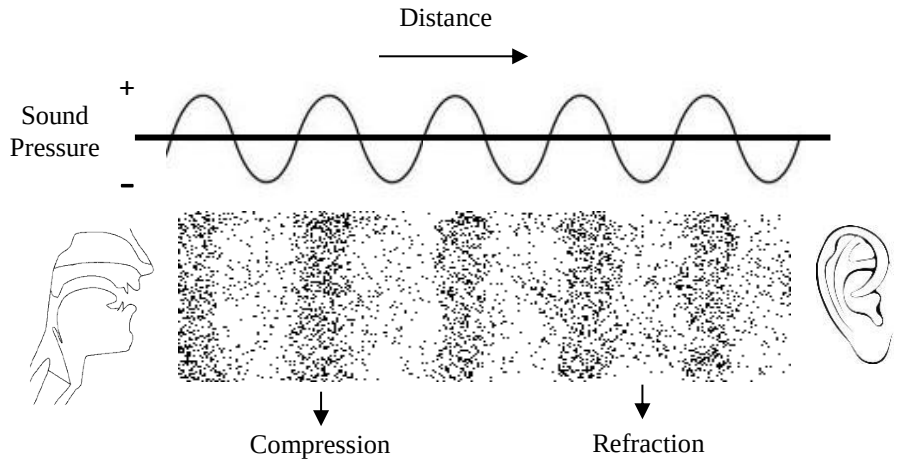


***Figure 2.10**: Illustration of sound propagation in air*

A microphone is used to capture the sound waves and converts it into audio signal. Changes in air pressure causes a material in the microphone called diaphragm to vibrate and produces a variation in an electrical voltage which is proportional to the air pressure. Three main features that are used to describe the audio signal are time, frequency and amplitude. Figure 2.11 displays the time waveform that tracks changes in air pressure (amplitude) over time and the spectrogram that graphs energy content in a signal as a function of frequency and time.

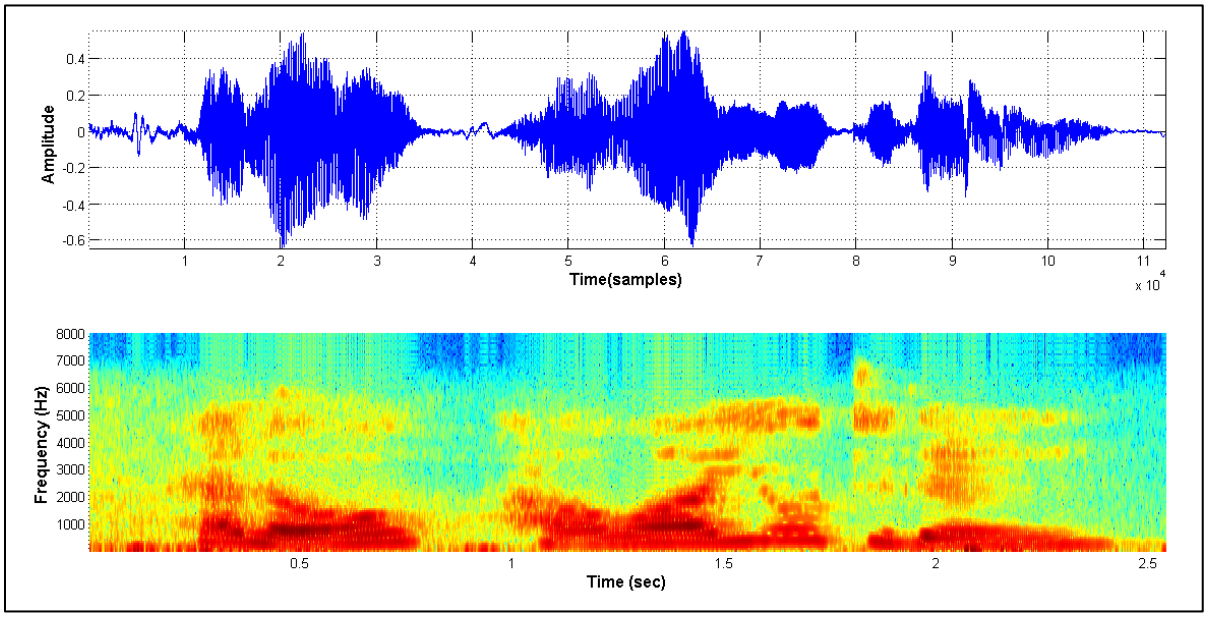


***Figure 2.11**: Example of time waveform and spectrogram plot*

6.0 Speech Perception

Understanding the structure and information contains in speech signal has been generally explored in the context of speech production. However, the auditory system has also been recognized to have an association with speech signals and to be used as an explanatory framework in the studies of speech. Human don't perceive frequency, level, spectral shape, modulation depth or frequency of modulation but instead perceive pitch, loudness, sharpness, fluctuation strength or roughness. The information received by auditory system can be described most effectively in the three dimensions of loudness, critical band rate and time. The ear gathers sounds from the surroundings and converts it a form that can be interpreted by the brain. Unconsciously, the auditory system has the capability to identify and decode variations of spectrum, pitch and amplitude that is constantly occurring in speech. The first part in this section explores the anatomy and physiology of the auditory system and the second part deals with the auditory psychophysics relating to the quantitative modeling of auditory perception.

6.1 Anatomy and physiology of the Auditory System

Referring to figure 2.12, peripheral auditory system can be divided into three sections; the outer ear, middle ear and inner ear [1].

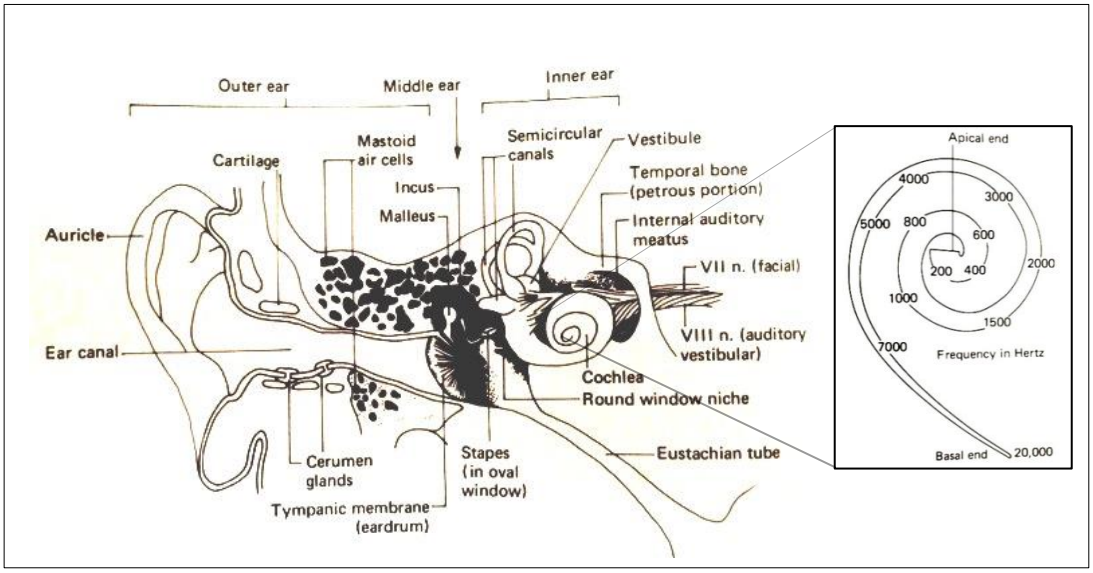


***Figure 2.12**: The auditory system and close-up image of the maximum amplitude distribution in the cochlea [13].*

6.1.1 Outer Ear

The outer ear consists of an air-filled tubed that begins at the opening of the ear canal and ends at the tympanic membrane (eardrum). Sound waves from the surrounding are channeled down through the ear canal to the eardrum. The tube-like passageway also functions as an acoustic resonator that enhances sound energy (vibration).

6.1.2 Middle Ear

One of the three smallest bones is connected to the drum membrane. The three bones called malleus, incus and stapes form a chain and conducts energy from the middle ear to the inner ear. Incus and Stapes is connected by a joint in order to permit motion between these two bones. Motion is important due to the function of stapes that vibrates and moves in a certain direction depending on the intensity of the received sound. The middle ear performs two major functions which are delivering sound energy from eardrum to the inner ear effectively with minimum loss and protecting the inner ear by reducing any excessive amount of energy that enters.

6.1.3 Inner Ear

The main component in the inner ear is the cochlea which consists of snail-shaped tube that begins at the basal end and reaches the apical end (apex) as it coils inward. The cochlea is divided into two large fluid-filled cavities separated by a stiff element called basilar membrane. The membrane is narrow near the basal end and more elastic and wider at the apex. As the membrane traverse into the apical end, portions of the membrane respond to different range of frequencies beginning with 20000 Hz near the basal end and approximately 20 Hz at the apex as illustrated in the close up image in figure 2.12. The motion of the basilar membrane is in the form of a traveling wave that will respond to the frequencies range with maximum amplitude. So, a low-frequency tone will produce higher traveling wave amplitude near the apex and high-frequency tone will produce smaller amplitude near the basal. Hair cells located at the base of the membrane converts the motion of the basilar membrane into an electrical signal which then generates waveform that resembles the original acoustic pressure wave at the eardrum.

6.2 Auditory Psychophysics

Fletcher [14] formalized the ability of the auditory system to identify and separate frequencies using a concept called *critical band*. Each place on the membrane reacts to a certain range of frequencies. The nature of traveling wave that peaks at different positions as it navigate along the basilar membrane allows the membrane to be modeled as an array of band-pass filters with overlapping frequencies and correspond to a filter with different center frequency.

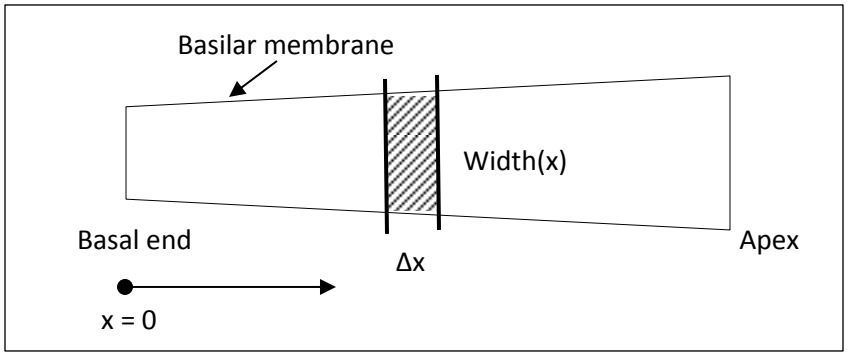


Figure 2.13*: Illustration of an unrolled basilar membrane [4]*

Shown in figure 2.13, when analyzing the membrane, we can take a small section (Δx) of the membrane and model it as a digital filter. The width of the membrane increases as the distance increase. The shorter width of the membrane corresponds to the higher frequency. Thus, the membrane can be represented by many cascaded second order digital filters as shown in figure 2.14. Each resonating filter depends on the frequency of the membrane. The deflection on the membrane that is caused by the input frequency will be sent to the brain through hair cells as an electrical signal.

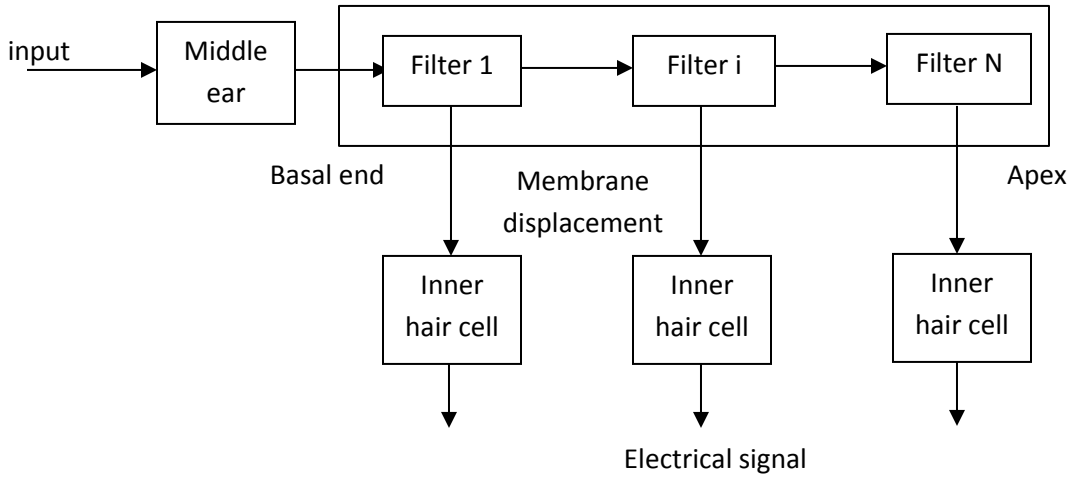


***Figure 2.14**: Digital filter model of the basilar membrane [4]*

In real time domain, the digital filter model of the basilar membrane can also be modeled as parallel bandpass filter bank in time domain in order to reduce the delay time caused by cascading filters as shown in figure 2.15. Each filter has different bandwidths with small bandwidth at higher frequency and larger bandwidth at lower frequency. They are called the 'critical band filters'. A set of 24 bandpass filters is identified to be sufficient in modeling the basilar membrane.

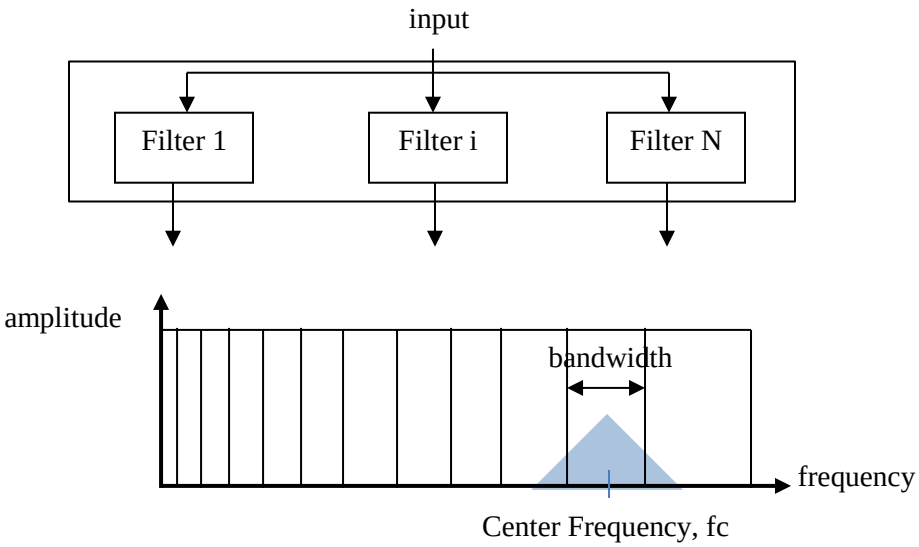


***Figure 2.15**: Parallel filter bank model [5]*

Signal component within a given critical band can be masked by other components within the same critical band (intra-band). Also the signal in one band can suppress the other filters as well-sounds on one critical band can mask sounds in different critical bands (inter-band). These concepts are called the auditory masking.

These filters may be spaced on a perceptual frequency scale known as either bark scale or mel scale. These scales are a mapping of the linear frequency spectrum based on human auditory perception since human ear is more sensitive to changes in lower frequency portion of the spectrum. The relationship between the frequency in Hz and the critical band rate with the unit of Bark or Mel can be approximated by the following equations [66]:

$$F_{Bark} = 13.0tan^{-1}(0.76F_{Hz}) \quad for \quad F_{Hz} < 1.5kHz \qquad (2.4)$$

$$F_{Bark} = 8l7 + 14.2log_{10}(F_{Hz}) \quad for \quad F_{Hz} > 1.5kHz \qquad (2.5)$$

or

$$F_{Mel} = 2595log_{10}\left[1 + \frac{F_{Hz}}{700}\right] \qquad (2.6)$$

7.0 Suicidal Indicators in Psychiatric Disorders

Diagnosis revealed that people with psychiatric disorders are at the highest risk of committing suicide. Particularly, diagnosed disorders that are most related to suicidal are depression and schizophrenia. Untreated depression is one of the most common cause for suicidal. According to [17], more than 90 percent of people who committed suicide are believed to have clinical depression or diagnosable mental disorders. A study in [18] also observed certain level of depression present in all 26 patients who died by suicide where 18 of them were diagnosed as severely depressed, six as moderately depressed and two as mildly depressed. In some cases, schizophrenia also precipitate suicide attempts. Between 18 to 55 percent of patients experiencing schizophrenia attempted suicide and approximately 10 percent of those with schizophrenia died by taking their lives [19].

Depression is a mental type disorders that disturbs an individual's mood. People with chronic depression usually indicates symptom such as severe fatigue, loss of interest, insomnia,

suicidal thoughts and social withdrawal accompanying with presence of certain physical complaints. These symptoms persist for at least two weeks and on average last for nine months. Bipolar disorder (manic depression) is one type of depression where a person's feeling alternates between feeling depressed in one occasion and suddenly feeling extremely joyful or experiencing dysfunctional behavior. People with bipolar usually exhibit poor judgment and reckless behavior because they tend to be hyperactive, irritable and overly confident. Another form of depression is called dysthymia which is a less common form of depression. A person with dysthymia experience intermittent feeling of similar to those with major depression but only for a brief time [15].

Schizophrenia is a medical illness that affects the function of the brain by causing disturbances which include delusions, confused thinking and hallucinations. Some major symptoms relating to schizophrenia include disturbances in thought process that causes the individual to be unable to think clearly, thus can make it very difficult for individual to develop proper communication. Also, they will have difficulties in distinguishing between reality and fantasy by having beliefs that are not true such as believing that people are trying to harm them. Another symptom that frequently occurs is unpredictable emotion changes. In some occasions, the individual can be extremely happy, sad or have no feeling at all. In relating to speech, cognitive symptoms such as poor concentration, difficulty understanding, difficulty expressing thoughts and slow thinking due to this illness may alter the speech and voice during communication [16].

Reports in [20] has shown that there is a relationship between major depression and suicidal in schizophrenia. Individual with schizophrenia that attempted suicide has been reported that depression has been considered to be the most important risk factor for suicidal behavior besides feeling hopelessness and despair about their current situation. There are many other factors relating to psychiatric disorders that also contribute to suicidal behavior such as mood disorders, organic brain syndrome, and personality disorders. However, major depression is the most common factors of suicide and thus its relation to suicide will be study in depth in this research.

8.0 Speech and Voice Characteristics in Psychiatric Patients

8.1 Subjective Analysis: Assessment by Trained Listeners/Clinicians

Voice is a highly expressive instrument that not only conveys thought but also reveals the affective and psychological state of a person. Information on psychological state is often encoded in 'how' and 'what' we say. However, in this study, the question of 'how' is much more important. For many years, trained clinicians have detected the differences in speech and voice pattern between normal people and patients that are diagnosed with psychiatric disorder. As early as 1938, Newman and Mather [21] analyzed the characteristic of voice and spoken language of 40 patients with affective disorders and associating it with certain psychiatric syndromes. Voice of patients with depressive syndromes was described as dead, listless quality. The pitch was described as a 'narrow tonal range', mainly 'step-wise rather than gliding' and broadest pitch changes occurring at the end of sentences. Hesitation pauses disturbed their speech flow giving the impression of being slow and halting. Moses [22] identifies various vocal characteristics that differentiate the voice of neurotic and psychotic individuals. He described voice patterns in individuals who were depressed as monotonous, uniform speech, having 'regular repetition of gliding intervals' and lower pitch ranges. His findings supported [21].

Case studies by Ostwald [23] on acoustic changes during psychotherapy reported that a change in the loudness of voice in psychiatric patients revealed to be the most reliable indicator of clinical change. Another acoustic measure that was shown to be reliable is a change in time-duration where it was reported that after treatment speech in younger patients tend to speed up and speech in older patients tend to slow down. Hargreaves and Starkweather [24] identified the existence of two subgroups of speech patterns within the 10 chronically depressed patients. One subgroup displayed lifeless quality of voice with decrease in loudness and diminished inflection changes, results which agree with previous findings. However, another subgroup of manic-depressive displayed different voice and articulation profile where they exhibit louder voice, wider pitch range, vigorous articulation, clear, lively and vital voice. As these patients became less manic, their pitch narrows and emphasis patterns are reduced. Another study by Hargreaves, Starkweather and Blacker [25] observed 32 depressed patients interviewed daily over the period of five weeks and observed low in overall loudness and lack of high overtones giving their a dull, lifeless quality with diminished inflection. Whitman and Flicker [26] observed higher pitch,

lower loudness, reduction in vocal intensity and an increased in articulatory rate correlated with greater severity of depression.

Hollien and Darby [27] observed slower reading time and lower phonation ratio in depressed patients compared to control patients. On the other hand, another study performed by Darby [28] later on suggested an unexpected result. The speech of depressed patients was reported to exhibit rapid rate, short rushed of speech and was more predominant than slow rate. Based on the results from the Hamilton Rating Score (HAMD) conducted on 13 depressed patients, he also reported reduce of stress pattern in speaking, lack of pitch variation and loudness, and increase in harshness.

These studies revealed speech characteristics such as slow, delayed, low voice, monotonous speech, lower pitch, decrease in loudness and decrease in intensity are prominent features that can be observed in patients with severe depression. However, evidences about linguistic changes provided by the clinicians are subjective in nature and hard to verify. Nowadays, there are a growing number of studies that focuses on finding objective nature in speech and voice for the purpose of psychiatric identification especially in depression. However, assessment by clinicians is still used as the bases for these studies.

8.2 Objective Analysis: Acoustic and Temporal Measurements

8.2.1 Previous Analysis of Features in Depressed Speech

Throughout these years, some of the most generally used speech and vocal features that were investigated in association to the state of depression include pitch-related features, energy-related features, formant features, timing features and glottal features.

Fundamental frequency (F0) describes the vibratory rate of the vocal folds during speech and is perceives by human listeners as an individual's vocal pitch. Pitch is considered to be one of the closest measures in the human perception of depression. Higher-pitched voice corresponds to higher F0 which usually heard in female and children while lower F0 value is perceived as a deeper voice which is usually heard in male. Listeners perceives an individual's speech pattern as monotonic (or small pitch variation) and highly intonated (or large pitch variation. In the frequency domain analysis, F0 was identified to be the most common feature studied in terms of its relationship to depressed speech [30, 31, 32, 34, 37, 42, 43, 46, 47, 48, 50].

Among F0 related features extracted are mean F0, standard deviation of the F0 and variance in F0 (inflection) [47, 32, 34, 37, 42]. Coefficient of variation of F0 is calculated by dividing the standard deviation of F0 the mean of F0 and is used to derive a standardized measure of pitch variation that is comparable between male and female [43, 46]. A study in [48] performed a small scale test for F0 feature extraction using the method of autocorrelation, cepstrum and average magnitude difference function from Roger Jang's audio toolbox. Results yielded comparable F0 values and thus, autocorrelation was chosen as the method of F0 extraction. Another investigation in [50] extracted F0-6bB-bandwidth, F0-amplitude and F0-contour. F0-contour is described as a specific form of F0 distribution curve that reflects variation around the mean F0. In [5], F0 rate of change measurement was calculated for a particular time interval and normalized by dividing by F0. The rate of change was measure in percentage rather than in Hz in order to eliminate speaker's pitch level.

Investigation on F0 speech pattern found out that depressed patients either showed a decrease in F0 after receiving treatment [31, 32, 38, 43, 47, 50] or display higher F0 range in comparison to control patients [39, 42, 51] . According to [31], speech F0 observed in depressed patients indicated that the patient's speech became more relaxed after receiving treatment. During depressed state, tension in the vocal cords increased due to higher muscle tone and thus, causing the vocal cord to vibrate faster. A study by [50] compared F0-related features extracted in the first recording session when patients were first admitted into the hospital, the sixth recording session after 14 days admitted and the final recording session during hospital discharge. F0-amplitude and F0-6dB-bandwidth started to show statistically significant difference in the sixth recording session and remained so until the point of hospital discharge, whereas the F0 only reached significance during the final recording. The pattern of improvement and recovery from depression was signified by a decrease in F0, F0-amplitude, and F0-contour. Higher F0-6bB-bandwidth however was observed in the final recording speech indicating a negative correlation with the psychological state.

It is also reported in [46] that depressed patients that had more than 50% drop in the depression severity showed significant increase in pitch variability about F0 (coefficient of variance in F0). On the other hand, patients with less than 50% drop in depression severity exhibit small decrease in pitch variability. Coefficient of variance in F0 was identified to correlate with clinical change in response to treatment. A study in [43] looked at correlation

between pitch variation and Hamilton Depression Rating Score (HAMD). The result demonstrated negative correlation between pitch variation and HAMD scores but not significant due to sample size. Negative correlation indicates that an increase in HAMD scores (which indicates greater symptom severity) corresponds to a decrease in pitch variation (monotonous).

The effect of observations labeling and classification result was studied in [34, 42]. He divided the passage into two grouping of 13 observations of 5 sentences and 5 observations of 13 sentences respectively. In [42], the entire set of observations was labeled as a patient group (depressed or control) based on the majority of the observation's label. Feature statistics related to F0 demonstrated high classification accuracy between depressed and control speech in male and female patients. However, in [34], Moore performed the analysis on two different observational groupings to eliminate the effect of dependency using observation subsets. In this study, it was reported that pitch did not provide the best classification between control and depressed subjects. His argument was based on the assumption that reading methods affects the pitch variations because patients natural reading ability were inclined to properly match their voice inflection with the content of the text rather than portraying their mood.

Another most common feature that can be observed through human perception is the temporal aspect of speech. Besides the content of speech, the human ear also listens to the prosodic characteristics such as sounding of vocal inflection, rhythm and stress placement in speech. Time-based measure of speech production relating to pauses, timing structure, phonation, and syllable prolongations forms an important part of these prosodic natures of speech. Many evidence in the literature has studied the relationship of psychological state and temporal features [30, 32, 37, 40, 41, 43, 44, 45, 46, 47, 50].

The choice of speech sample seems to be an important factor when performing timing-based analysis. Free speech or automatic speech will have an effect on the timing-based feature analysis especially in pause-related measurements. In automatic speech, variation that exists in speech is most probably due to the temporal aspect because reading a passage eliminates the involvement of complex cognitive planning processes. However, free speech involves both cognitive planning and other speech measurements including the temporal aspect of speech [41]. Temporal features extracted from automatic speech task were reported to better correlate with depression severity [46]. Automatic speech might have an effect to the improvement that is observed in speaking rate and fluency of speech. Patients might have gotten used to the recording

procedure after performing it a few times and their improvement might be due to the 'practice' effect from each recording sessions [41]. On the other hand, [44] stated that the improvement demonstrated by depressive patients was not because of this 'practice' effect since control patients did not exhibit any change in pause time over the period of study.

Example of automatic speech tasks are counting from one to ten [46, 47, 50, 44], pronouncing vowels for a short period of time or repeating syllables [46] and reading a standardized passage that is commonly used in the assessment of communication disorders such as *The Grandfather Passage* or *The Rainbow Passage* [46, 50]. Example of free speech task is a standardized interview on conversational topics between a patient and an interviewer [47, 46, 45, 41].

Common timing-based features extracted are grouped into phonation length that is described by periods of voicing of uninterrupted by pauses, pause-related measures such as total period of pauses and patient's response time to interviewer's questions or comments and speaking rate which is calculated by dividing the number of syllables spoken by the length of the sampled measured in seconds.

When comparing between two groups of control and depressed, pause duration and pauses between the interviewer's questions and patient's answers in the control group were found to be lower than the depressed group. However, phonation length was observed to be higher in the control group compared to the depressed group [39, 44, 47].

For investigation relating to the correlation of depression severity with the timing-based features, results in [46, 45] suggested that total recording length that is dominated by longer and more variable pauses was observed to increase with greater severity symptom in depression. It is based on the ratio of pause time measure that was calculated as the percent of total pause time relative to the total time of the speech session. This also implies that patients exhibit lower phonation length and slower speaking rates. Other findings also observed similar outcome which they reported that depressed patients showed improvement by demonstrating a decrease in pause-related measures and/or an increase in speaking rate after patients receive treatments [30, 32, 38, 47, 50, 44, 41]. However, inconsistent results were reported for vocalization-related measures. A study in [47] reported increment in phonation length with regards to treatment but studies in [38, 44] reported no change in phonation length throughout the whole period. Studies by [43, 46, 45] supported previous finding by showing significantly high negative correlations between speaking

rate and HAMD scores indicating greater symptom severity when speaking rates decrease. Investigations by [40, 43, 45] demonstrated a moderately correlated relationship between pause-related measures and HAMD total score but other studies achieved strongest correlation between pause-related measures with HAMD total score [44, 46] and Retardation Rating Scale (ERD) [40]. This inconsistency was possibly due to the difference in speech sample used or the accuracy of the clinical ratings. Increase in pause duration and higher speaking rate that are observed in depressed speech imply that patients speak with greater hesitancy and takes longer time to express themselves.

Most subjective analysis reported in the review regarding the F0 speech pattern and timing-based measurements seem to agree with the objective analysis performed by trained clinicians/listeners based on perceptual observation.

Besides F0 and temporal features, another aspect of prosodic measures that has been investigated is the distribution of spectral energy. The spectral content of speech can be considered as a measure of loudness in phonation [37]. The use of spectral energy has led to a number of studies on the correlation with clinical depression and distinguishing between the depressed group and the control group [25, 30, 31, 37, 50, 51, 55, 56, 57]. During patient's recovery from depression and after treatment, studies revealed an increase in the overall energy content of speech. The energy content in speech contains more power in the frequency range below 500 Hz [51]. Analysis of the distribution of spectral energy in proportions of frequency band revealed that the energy content in the speech of depressed patients below 500Hz range (lower frequency range) exhibited an increment when measured from admission to the time of discharge. On the contrary, proportion of frequency band in the range of 500Hz to 1000Hz (middle frequency rage) exhibited a decrease in energy content [23, 25, 31, 56, 57]. A contradictory result was reported in [24, 50] where they observed an increase in both lower and middle frequency range of 200Hz to 1000Hz when patients recover from severe depression.

Relationship between spectral energy with depression and measure of changes in energy content in patient's speech during admission and discharged has been reported to be significantly different [24, 30, 50, 51]. But investigation in [30] identified that change in the spectral energy of speech with respect to the severity of depression exhibited positive as well as negative correlation. Some patient's displayed speech transformation from low-voiced, flat and

monotonous towards a standard speech observed in healthy patients, whereas in others, their speech displayed transformation from loud and abrupt speech towards standard speech.

A number of literatures have looked at the relationship between formant features and depression. The study of formant allows researchers to observe patient's vocal tract behavior during speech production and articulation. The investigation concerning formant feature and depression severity correlation was initially conducted by [31] based on the usefulness of formant feature in speaker identification. As patient's depression severity reduces, this study reported a significant frequency increment for the first formant and the overall formant frequencies resembles more closely to the neutral formant frequencies of 500, 1500 and 2500 Hz. A study by [36, 51] also replicated this result, reporting a greater standard deviation and an increase in the first formant frequency for depressed patients in comparison to the healthy patients. The neutral formant frequencies occurred when the vocal tract is in the resting position. This observation suggests that depressed patient speaks with less articulatory effort.

On the contrary, another study performed by [29] reported a decrease in the second formant transition when comparing the speech of major depressed groups to control groups. Control patients were assumed to experience a much relaxed tongue movement and less sluggish articulation compared to depressed patients.

8.2.2 Previous Analysis of Features in Suicidal Speech

The idea of recognizing distinctive patterns and tone of voice in patients with high risk suicide was introduced by two clinical psychologists, Drs. Stephen and Marilyn Silverman. Both had experience in treating patients with near term suicidal risk. They began research in the 1980s by collecting and analyzing suicidal tape recordings obtained through therapy sessions in an uncontrolled environment, and notes and interviews made shortly before suicide attempts. They describe the similarity of vocal speech between depressed and suicidal patient but notice changes occur considerably in the tonal quality and acoustical characteristics when the patient enters the suicidal state. Criteria in speech that were identified as a precursor to suicidal are the substantial non instantaneous amplitude decay upon conclusion of the utterance, low or minimal amplitude modulation (thinner and less rich vocal content), decrease variation of fundamental frequency and low frequency amplitude modulation [58-62].

A small scale study performed by Campbell [63] investigated the statistical properties of the fundamental frequency distribution on two male patients and one female patient that were observed over a period of time ranging from nine hours to nine years. Throughout recording sessions, there was a certain range of time which patients became suicidal. Therefore, recordings of high risk suicidal and control were extracted using their own speech segments extracted during times when they were either in the state of suicidal or non-suicidal. The statistical properties and variations in fundamental frequency distribution served as the discriminating features in this study based on the result of 22.7% misclassification error. The study reported that the recording environment must be controlled in order for the patient to be accurately classified.

France [51] investigated multiple acoustical properties of speech including fundamental frequency (F0), Amplitude Modulation (AM), formant, and power distribution (PSD) on 21 male depressed patients and 22 male high risk suicidal patients. The following six statistics which include range, variance, mean, skewness, kurtosis and coefficient of variance were extracted for F0, AM and formant analysis. As for power distribution, spectral distribution from four 500Hz equal bands within 0-2000Hz was obtained for analysis. AM range, AM coefficient of variance and band three PSD were selected as the best discriminator yielding 86% overall correct classification using a hold-one-out method of quadratic discriminant analysis. The classifier was significantly more effective in classifying major depressed speech (86%) than suicidal speech (77%).

Ozdas [52] divided her feature analysis into source domain analysis and filter domain analysis. Vocal tract characteristic using a non-model based approach for near term suicidal risk assessment was the focus of this study. The effects of source (excitation) and filter (vocal tract) on suicidal state were the two domains examined. The source domain method analyzed the effectiveness of small cycle-to-cycle variations of fundamental frequency known as voice 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 low order Mel-Frequency Cepstral Coefficient (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.

Yingthawornsuk [53] extracted features based on the Power Spectral Density (PSD) and Gaussian mixture model (GMM) based spectral modeling of the vocal tract which contains information on spectral pattern (intensity, responding frequency and bandwidth). 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.

Keskinpala [54] did a follow-up to Ozdas' and Yingthawornsuks' studies where she proposed an optimization analysis of multiple MFCC coefficients and different numbers, ranges and edges of spectral energy bands using a new set of data. Continuing the work of Yingthawornsuk, she tested out seven types of changes with energy bands in order to find an optimized energy bands for better classification rates on all pairwise groups of high-risk suicidal, depressed and remitted. The optimization techniques were:

- Increase the number of energy bands within 0 Hz to 2000 Hz (two to 10 equal bands)
- Increase the energy band range to 0 Hz to 3000 Hz
- Increase the number of energy bands and range
- Exponential band edges
- Exponential band edges and increasing the energy band range
- Non-uniform band edges
- Non-uniform band edges and increasing the energy band range

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. This study introduced the method of cepstral mean normalization for compensating spectral variability due to different recording environments. The importance of environmental compensation was tested by performing classification with and without compensation and results demonstrated that using no

compensation provided better results. Text-dependent speech samples (automatic speech) were shown to provide better discrimination analysis in distinguishing suicidal patients compared to the interview speech sample (free speech). Cross-validation and testing with all data training were two methods of resampling that were used to obtain the classification qualitative measurements. Cross validation classifier based-method was demonstrated to perform well as an assessment approach in identifying high risk suicidal patients.

Wan Ahmad Hasan [55] and Nik Nur Wahidah [56] performed analysis of power spectral density (PSD) on female and male patients, respectively. Different method of PSD extraction was used in these studies compared to the Welch method that was previously used by Yingthawornsuk and France. Their studies reported a significant increase in classification of high risk suicidal group and depressed group after removing outlier patients, thus, suggesting that there are a small group of people with unusual speech characteristics. These people might indicate that suicidality does not change the characteristic of the voice or perhaps their voice mechanism is damaged. Also, since the database was obtained from patients in a wide range of 18-65 years old, some of the speech recordings might belong to older people. Elderly people may experience changes in voice and thus contribute to the existence of the sub-population.

Although they have contributed far more than what was written here, this section only covers parts of their work that are pertinent to this research.

References for Chapter I and II

[1] P. B. Denes, E. N. Pinson, "The Speech Chain", 2nd edition, Worth Publishers, New York, 1993.

[2] J. R. Deller, J. H. L. Hansen, Proakis J.G., "Discrete-Time Processing of Speech Signals", Macmillan Publishing Company, New York, NY, 1993.

[3] E. J. Yannakoudakis, P. J. Hutton, "Speech Synthesis and Recognition Systems", pp. 16-26, 1987.

[4] J.L. Flanagan, "Speech Analysis, Synthesis, and Perception", 2nd ed., Springer-Verlag, New York, 1983.

[5] L. R. Rabiner, R. W. Schafer, "Digital Processing of Speech Signals", Prentice-Hall Signal Processing Series, pp. 38-105, 1978.

[6] T. J. Hixon, L. D. Shriberg, J. H. Saxman, "Introduction to Communication Disorders" Englewood Cliffs Prentice-Hall, 1980.

[7] Tools for Electroglottographic Analysis: Software, Documentation and Databases, website, http://voiceresearch.free.fr/egg

[8] R. T. Ingo, "The Myoelastic Aerodynamic Theory of Phonation ", PhD, pp. 26-45, 2006.

[9] L. Deng, D. O'Shaughnessy, "Speech Processing: A Dynamix and Optimization-Oriented Approach", *Signal Processing and Communication Series*, pp. 203-261, 2003.

[10] G. Digottex, "Glottal source and vocal tract separation: Estimation of Glottal Parameters, Voice Transformation and Synthesis using a Glottal Model, PhD, University Pierre and Marie Curie, December, 2010.

[11] E. Moore II, "Evaluating Objective Feature Statistics of Speech as Indicator of Vocal Affect and Depression", Ph.D Thesis, Georgia Institute of Technology, November, 2003.

[12] B. S. Atal, S. L. Hanauer, "Speech Analysis and Synthesis by Linear Prediction of the Speech wave," *J. Acoust. Soc. Am.*, vol. 50, no. 2 (Part 2), pp. 637-665, 1971.

[13] H. Davis, S. R. Silverman, "Hearing and Deafness", 3rd Edition, New York Holt, Rinehart and Winston, 1970.

[14] H. Fletcher, "Speech and Hearing in Communication", D. Van Nostrand Co. Inc., Princeton, N.J., 1953.

[15] I. H. Gotlib, C. L. Hammen, Handbook of Depression,The Guilford Press, 2nd Edition, 2010

[16] R. Miller, S. E. Mason, "Diagnosis: Schizophrenia : a Comprehensive Resource for Patients, Families and Helping Professionals", Columbia University Press, 2002

[17] Y. Conwell, D. Brent, "Suicide and aging I: patterns of psychiatric diagnosis", *International Psychogeriatrics*, 7(2): 149-64, 1995

[18] H. Hendin, J. T. Maltsberger, A. P. Haas, K. Szanto, H. Rabinowicz, "Desperation and Other Affective States in Suicidal Patients, Suicide and Life-Threatening Behavior", *The American Association of Suicidology*, vol. 34, no. 4, 2004.

[19] S. G. Siris, "Suicide and schizophrenia", *Journal of Psychopharmacology*, vol. 15, pp. 127–135, 2001.

[20] J. M. Harkavy-Friedman, E. A. Nelson, D. F. Venarde, J. J. Mann, "Suicidal Behavior in Schizophrenia and Schizoaffective Disorder: Examining the Role of Depression, Suicide and Life-Threatening Behavior", *The American Association of Suicidology*, vol. 34, no. 1, 2004.

[21] S. S. Newman, V. G. Mather, "Analysis of Spoken Language of Patients with Affective Disorders", *Am. J. Psychiat*, vol. 94, pp. 912-942, 1938.

[22] P. J. Moses, "The Voice of Neurosis", Grune and Stratton, New York 1954.

[23] P. F. Ostwald, "Soundmaking: The Acoustic Communication of Emotion", Thomas, Springfield, 1963.

[24] W. A. Hargreaves, J. A. Starkweather, "Voice Quality Changes in Depression", Language and Speech, vol. 7, pp. 84-88, 1964.

[25] W. A. Hargreaves, J. A. Starkweather, K. H. Blacker, "Voice Quality in Depression", *J. Abnorm. Psychol*, vol. 70, pp. 218-220, 1965.

[26] E. Whitman, D. J. Flicker, "A Potential New Measurement of Emotional State: A Preliminary Report", Newark Beth-Israel Hospital, vol. 17, pp. 167-172, 1966.

[27] H. Hollien, J. K. Darby, "Acoustic comparisons of psychotic and non-psychotic voices", In: Current issues in the phonetic sciences, vol. 9, pp. 79-103, 1979.

[28] J. K. Darby, "Speech and Voice Parameters of Depression: A Pilot Study", *J. Commun. Disord.*, vol. 7, pp. 75-85, 1984.

[29] A. J. Flint, S. E. Black, I. Campbell-Taylor, G. F. Gailey, and C. Levinton, "Abnormal Speech Articulation, Psychomotor Retardation, and Subcortical Dysfunction in Major Depression", *J. Psychiat. Res*, vol. 27, no. 3, pp. 309-319, 1993.

[30] H. H. Stassen, G. Bomben, and E. Gunther, "Speech Characteristics in Depression", *Psychopathology*, vol. 24, pp. 88-105, 1991.

[31] F. Tolkmitt, H. Helfrich, R. Standke, and K. R. Scherer, "Vocal Indicators of Psychiatric Treatment Effects in Depressives and Schizophrenics", *Journal of Communication Disorders*, vol. 15, pp. 209-222, 1982.

[32] J. K. Darby, and H. Hollien, "Vocal and Speech Patterns of Depressive Patients", *Folia. Phoniat*., vol. 29, pp. 279-291, 1977.

[33] B. S. Helfer, T. F. Quatieri, J. R. Williamson, D. D. Mehta, R. Horwitz, B. Yu, "Classification of Depression State Based on Articulatory Precision*", Proceedings of Interspeech: 14th Annual Conference of the International Speech Communication Association*, Lyon, France, 2013.

[34] E. Moore, M. Clements, J. Peifer, L. Weisser, "Comparing Objective Feature Statistics of Speech for Classifying Clinical Depression", *Proceedings of the 26h annual international conference of the IEEE*, 2004.

[35] E. Moore, M. A. Clements, J. W. Peifer, L. Weisser, "Critical Analysis of the Impact of Glottal Features in the Classification of Clinical Depression in Speech", *IEEE Transactions on biomedical engineering*, vol. 55, no. 1, 2008.

[36] N. Cummins, J. Epps, M. Breakspear, R. Goecke, "An Investigation of Depressed Speech Detection: Features and Normalization", *INTERSPEECH 2011, 12th Annual Conference of the International Speech Communication Association*, Florence, Italy, 2011.

[37] A. Nilsonne, J. Sundberg, S. Tenstrom, A. Askenfelt, "Analyzing Voice Fundamental Mobility and Some Other Aspects of Temporal Dynamics in Reading", *KTH Computer science and communication*, vol. 26, no. 4, pp. 59-84, 1985.

[38] A. Nilsonne, "Acoustic Analysis of Speech Variables during Depression and After Improvement", *Acta psychiatry. Scand*., vol. 76, pp. 235-245, 1987.

[39] A. Nilsonne, "Speech Characteristics as Indicators of Depressive Illness", *Acta. Psychiatr. Scand*, vol. 77, no. 3, pp. 253-63, 1988.

[40] P. Hardy, R. Jouvent, and D. Widlocher, "Speech Pause Time and the Retardation Rating Scale for Depression (ERD)." *Journal of Affective Disorders*, vol. 6, pp. 123-127, 1984.

[41] H. Ellgring, and K. R. Scherer, "Vocal Indication of Mood Change in Depression", *Journal of Nonverbal Behavior*, vol. 20, no. 2, pp. 83-110, 1996.

[42] E. Moore, M. Clements, J. Peifer, L. Weisser, "Analysis of Prosodic Variation in Speech for Clinical Depression", *Proceedings of the 25th annual international conference of the IEEE*, 2003.

[43] M. Cannizaro, B. Harel, N. Reilly, P. Chappell, and P. J. Snyder, "Voice Acoustical Measurement of the Severity of Major Depression. *Brain and Cognition*, vol. 56, no. 1, pp. 30-35, 2004.

[44] E. Szabadi, C. M. Bradshaw, and J. A. O. Besson, "Elongation of Pause-Time in Speech: A Simple, Objective Measure of Motor Retardation in Depression", *The British Journal of Psychiatry*, 129, 592-597, 1976.

[45] A. C. Trevino, T. F. Quatieri, and N. Malyska, "Phonologically-Based Biomarkers for Major Depressive Disorders", *Journal on Advances in Signal Processing*, vol. 1, no. 42, 2011.

[46] J. C. Mundt, P. J. Snyder, M. S. Cannizzaro, K. Chappie, and D. S. Geralts, "Voice Acoustic Measures of Depression Severity and Treatment Response Collected Via Interactive Voice Response (IVR) Technology", *Journal of Neurolinguistic*, vol. 20, pp. 50-64, 2007

[47] M. Alpert, E. R. Pouget, and R. R. Silva, "Reflections of Depression in Acoustic Measures of the Patient's Speech. *Journal of Affective Disorders*, vol. 66, pp. 59-69, 2011.

[48] Alex Low Lu-Shih, N. C. Maddage, M. Lech, L. B. Sheebar, N. B. Allen, "Detection of Clinical Depression in Adolescents' Speech During Family Interaction", *IEEE Transactions on biomedical engineering*, vol. 58, no. 3, 2011.

[49] Kuan Ee Brian Ooi, Lu-Shih Alex Low, M. Lech, N. Allen, "Early Prediction of Major Depression in Adolescents using Glottal Wave Characteristics and Teager Energy Parameters," *ICASSP*, 2012.

[50] S. Kunny, H. H. Stassen, "Speaking Behavior and Voice Sound Characteristics in Depressive Patients during Recovery", *J. Psychiat. Res*., vol. 27, no. 3, pp. 289-307, 1993

[51] D. J. France, "Acoustical Properties of Speech as Indicators of Depression and Suicidal Risk", Ph.D, Thesis, Vanderbilt University, August, 1997.

[52] A. Ozdas, "Analysis of Paralinguistic Properties of Speech for Near-term Suicidal Risk Assessment", Ph.D, Thesis, Vanderbilt University, May, 2001.

[53] T. Yingthawornsuk, "Acoustic Analysis of Vocal Output Characteristics for Suicidal Risk Assessment", Ph.D, Thesis, Vanderbilt University, December, 2007.

[54] H. K. Keskinpala, "Analysis of Spectral Properties of Speech for Detecting Suicide Risk and Impact of Gender Specific Differences", PhD Thesis, Vanderbilt University, 2011.

[55] R. Roessler, and J. Lester, "Voice Predict Effect during Psychotherapy", *The Journal of Nervous and Mental Disease*, vol. 163, no. 3, 1976.

[56] K. S. Scherer, "Nonlinguistic Vocal Indicators of Emotion and Psychopathology", *Emotions in personality and psychopathology*, pp. 595-529, 1979.

[57] J. D. Laver, "Individual Features in Voice Quality", Ph.D thesis, University of Edinburgh, 1976.

[58] S. E. Silverman, "Method for Detecting Suicidal Predisposition" U.S patent 4 675 904, June 23, 1987.

[59] S. E. Silverman, "Method for Detecting Suicidal Predisposition" U.S patent 5 148 483, Sept 15, 1992.

[60] S. E. Silverman, "Method for Detecting Suicidal Predisposition" U.S patent 5 976 081, Nov 2, 1999.

[61] S. E. Silverman, "Method for Detecting Suicidal Predisposition" U.S patent 5 591 238, June 8, 2003.

[62] S. E. Silverman, M. K. Silverman, "Method and Apparatus for Evaluating Near-Term Suicidal Risk using Vocal Parameters" U.S patent 7 062 443, June 13, 2006.

[63] L. Campbell, "Statistical Characteristics of Fundamental Frequency Distributions in the Speech of Suicidal Patients", Master's Thesis, Vanderbilt University, 1995.

[64] W. A. S Wan Ahmad Hasan, "Acoustical Analysis of Speech Based on Power Spectral Density Features in Detecting Suicidal Risk among Female Patients", Master's Thesis, Vanderbilt University, 2011.

[65] N. H. Nik Nur Wahidah, "Analysis of Power Spectrum Density of Male Speech as Indicators for High Risk and Depressed Decision", Master's Thesis, Vanderbilt University, 2011.

[66] S. Greenberg, W. A. Ainsworth, R. R. Fay, "Speech Processing in the Auditory System", Springer, 2004.

[67] R. M. Salomon, H. K. Keskinpala, M. H. Sanchez, T. Yingthawornsuk, N. H. Nik Nur Wahidah, W. A. S. Wan Ahmad Hasan, N. Taneja, D. Vergyri, B. H. Knoth, P. E. Garcia, D. M. Wilkes, R. Shiavi, "Analysis of Voice Speech Indicators in Suicidal Patients", Manuscript submitted for publication, 2012.

[68] 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 in press)

[69] V. Franc, V. Hlavac, "Linear and Quadratic Classification Toolbox for Matlab", Czech Pattern Recognition Workshop, 2000.

[70] H. M. Kalayeh, and D. A. Landgrebe, "Predicting the Required Number of Training Samples", *Pattern analysis and machine intelligence*, *IEEE transaction*, pp. 664-667, 1983.

[71] D. H. Klatt, L. C. Klatt, "Analysis, Synthesis and Perception of Voice Quality Variations Among Female and Male Talkers*", Journal Acoustical Society of America*, 1989.

[72] Y. Conwell, "Management of Suicidal Behavior in the Elderly", *Psychiatric Clinics of North America*, vol. 2, no. 3, 1997.

[73] Interactive Screening Program [http://www.afsp.org/]

[74] M. Hamilton, "A Rating Scale for Depression", *Journal Neurol. Neurosurg. Psychiat*., vol. 23, no. 56, 1960.

[75] J. C. Mundt, A. P. Vogel, D. E. Feltner, W. R. Lenderking, "Vocal Acoustic Biomarkers of Depression Severity and Treatment Response", *Journal of Biological Psychiatry*, vol. 72, 580-587, 2012.

[76] S. Theodoridis, K. Koutroumbas, "*Pattern Recognition, Fourth Edition*" Ac-ic Press, 2008

[77] Centers for Disease Control and Prevention [http://www.cdc.gov]

[78] D. K. Kenneth, X. Jiaquan, L. M. Sherry, M. M. Arialdi, K. Hsiang-Ching, "Deaths: Final Data for 2009, National Vital Statistic Reports", vol. 60, no. 3, 2011.

[79] J. L. McIntosh, "U.S.A Suicide: 2009 Official Final Data", American Association of Suicidology, 2010, http://www.suicidology.org. Accessed 26 July 2012

[80] R. Burns, "An impact: Suicides are Surging among US Troops" (The Associated Press, 2012 in press)

[81] K. A. Busch, J. Fawcett, D. G. Jacob, "Clinical Correlates of Inpatient Suicide", *J. Clin. Psychiatry*, vol. 64, no. 1, pp. 14-19, 2003.

[82] J. C. Fowler, "Suicide Risk Assessment in Clinical Practice: Pragmatic Guidelines for Imperfect Assessment", *American Psychological Association*, vol. 49, no. 1, pp. 81-90, 2012.

[83] C. M. Perlman, E. Neufeld, L. Martin, M. Goy, J. P. Hirdes, "Suicide Risk Assessment Inventory: A Resource Guide for Canadian Health care Organizations", Toronto, ON: Ontario Hospital Association and Canadian Patient Safety Institute, 2011.

[84] L. R. Wingate, T. E. Joiner, R. L. Walker, M. D. Rudd, D. A. Jobes, "Empirically Informed Approaches to Topics in Suicide Risk Assessment", Behavioral Sciences & the Law (Suicide and the Law, 2004), pp. 651–665

[85] J. J. Man, "The Neurobiology of Suicide", Nature Medicine, vol. 4, pp. 25-30, 1998.

[86] M. E. Ayadi, M. S. Kamel, F. Karray, "Survey on Speech Emotion Recognition: Features, Classification Schemes and Databases", Pattern Recognition, vol. 44, no. 3, pp. 572-587, 2011.

[87] S. L. Murphy, J. Q. Xu, and K. D. Kochanek, "Final data for 2010", National Vital Statistics Reports," MD: National Center for Health Statistics, Hyattsville, vol. 61, no. 4, 2013.

[88] E. Kraepelin, "Manic depressive insanity and paranoia", pp. 38, Livingstone, Edinburgh, 1921.

[89] P. F. Ostwald, "The Sounds of Emotional Disturbance", *Arch Gen Psychiatry*, vol. 5, no. 6, 1961.

[90] A. W. Siegman and S. Boyle, "Voices of Fear and Anxiety and Sadness and Depression: The Effects of Speech Rate and Loudness on Fear and Anxiety and Sadness and Depression", *Journal of Abnormal Psychology*, vol 102, no. 3, pp. 430-437, 1993.

[91] R. V. Shannon, Fan-Gang Zeng and J. Wygonski, "Title of the Speech Recognition using Only Temporal Cues", from the book of The auditory processing of speech: From sounds to words, edited by MEH Schouten, 1992.

[92] J. Kubanek, P. Brunner, A. Gunduz, D. Poeppel and G. Schalk, "The Tracking of Speech Envelope in the Human Cortex", PLoS ONE 8(1): e53398, 2013.

[93] H. Riquimaroux, "The Extent to Which Changes in the Amplitude Envelope Can Carry Information for Perception of Vocal Sound without the Fundamental Frequency or Formant Peaks", *Dynamics of speech production and perception*, 2006.

[94] J. C. R. Licklider, I Pollack, "Effects of Differentiation, Integration and Infinite Peak Clipping on the Intelligibility of Speech", *J Acoust. Soc. Am*, vol. 20, pp. 42-51, 1948.

CHAPTER III

ANALYSIS OF FEATURES BASED ON THE TIMING PATTERN OF SPEECH AS POTENTIAL INDICATORS OF HIGH RISK SUICIDE AND DEPRESSION

Abstract

Patients who are diagnosed with depression without appropriate clinical recognition of their hidden suicidal tendencies are at elevated risk of making a suicide attempts. An important clinical problem remains the differentiation between non-suicidal and more lethal episodes of depression. In an effort to find a reliable method that could assist clinicians in risk assessment, information in the speech signal has been found to contain characteristic changes associated with high risk suicidal states. This paper addresses the question of whether the information contain in the speech timing based measures are able to discriminate the high risk suicidal speech from the depressed speech. Data sets were collected from readings of a standard "rainbow passage" essay. Use of the leave-one-out procedure as a means to measure a classifier performance for all-data classification revealed single speech timing based measure to be a significant discriminator with 74% and 72% correct classification for male and female speech from Database A, respectively. Certain combinations of the timing based measures increase the accuracy up to 70%-90% for male and 79% for female patients. For male patients, using the trained features and classifiers on Database A and testing on Database B achieved up to100% detection of high risk speech in Database B. This finding suggests that the timing based measures are robust across databases despite the less than ideal recordings conditions and different equipment used during the recordings sessions.

1.0 Introduction

Suicide continues to be a major concern to public health worldwide. In the United States, the current available summary by the National Center for Health Statistics reported an age-adjusted increase of 2.4 percent from 2008 to 2009 followed by an increase in 2010 by another 2.5% [1]. The 2009 list of cause of death revealed that suicide ranks in 10^{th} with 36,909 suicides for all age groups, ranks 3^{rd} for the age group of 15-24 with 4,371 suicides [2]. A surge in the

military suicide rate with an increase of 18 percent in the year 2012 compared to the statistic reported at the first half of the previous year [3].

Despite decades of research, accurate prediction of suicide and imminent suicide attempts still remains elusive. A detection of biometric characteristics known to be associated with imminent suicidal risk at an early stage may prompt the clinician to propose and promote a more intensive treatment plan. Currently, commonly used suicide risk assessment tools comprise a series of questionnaires and checklists with rating scales that can be evaluated with fair reliability by trained clinicians [4]. It requires clinicians to use specific interviewing approaches, sometimes relying on their intuition, and to deploy specialized skills they develop through formal education, clinical training and clinical experience. Clinical interview tools remain the standard of care, but are often highly sensitive and commonly reveal "high" risk in Emergency Department evaluations. The interview quality, in evoking truthful admissions of suicidal thought content, determines reliability. This traditional method is time- and energy-intensive, and still frequently misses true positives. There might be more than 75 known risk factors that could lead to greater potential for suicidal behavior [5]. These risk factors include psychological milieu (life events, environmental factors and medical illnesses), presence of psychiatric disorders, biological factors, health records, family history and the history of any previous suicide attempts.

True risk and true sensitivity of the tools are difficult to assess since positive clinical findings often lead to intensive treatment, and natural outcome trials are unacceptable. No objective tool is available to assess (or to assist with) the false negative finding where a person at risk denies suicidal thoughts [6, 7]. Negative screenings can bring the clinician to believe that persons who are actually at imminent risk of committing suicide are experiencing a less severe (often depressive) disorder. Even for specialists, and especially for non-specialists, objective metric might signal a critical need for more extensive interviewing and other precautions. Misdiagnosis in such situations may result in an untoward, unfortunate outcome.

Studies have shown that speech contains implicitly hidden information that reflects psychological states and brain function, including affective states or the presence of diseases such as Parkinson's [8]-[14], [47]. Among the distinctive speech patterns that have been associated with depression are decreases in intonation, stress, loudness, inflection, intensity and speech rate, sluggishness in articulation, monotonous, and lack in vitality [15]-[17]. These characteristic correlates with the changes occurring in the speech production mechanism by

affecting the respiratory, laryngeal, resonance and articulatory subsystem that in turn are encoded in the acoustical signal. Investigations on vocal characteristics in terms of their relationship to depression include speech prosody (e.g., pitch, energy, and speaking rate) [8, 16], [17]-[19], spectral features (e.g., power spectral density, formants and their associated bandwidth) [8, 10, 11], glottal features [9, 20] and MFCCs [9, 11].

2.0 Previous Work

Investigations on the acoustic features for identifying depression and imminent suicidal risk detection often revolve around spectrum based measures of speech. France [8] examined speech features that are characterized by the long-term Fundamental Frequency statistics (mean, variance, skewness and kurtosis), Amplitude Modulation, Formants (including bandwidth and ratios) and Power Distribution. Ozdas [9] employed the low order Mel-Frequency Cepstral Coefficient (MFCC), small cycle-to-cycle variations of fundamental frequency known as voice jitter, and glottal flow spectral slope as discriminating features among the near-term suicidal, depressed and remitted groups. Yingthawornsuk [10] extracted features based on the Power Spectral Density (PSD) and Gaussian mixture model (GMM) based spectral modeling of the vocal tract which contains information on spectral pattern (intensity, responding frequency and bandwidth). Keskinpala [11] proposed an optimization study of multiple MFCC coefficients and performed an extensive study on different numbers, ranges and edges of the spectral energy bands.

Besides using clinical based assessment and speech based measurements, diagnosis of the psychological disorders particularly in depression has also been examined by means of visual based expression. Results from the analyses performed in [36]-[40] suggest that depressed patients exhibit particular facial expressions, behavior patterns and physical movements. Another study in [35] has explored the multimodal fusion for detecting depression by combining visual and verbal cues based on the hypothesis that the information from individual cues complements each other thus the multiple feature fusion will improve the performance of the system. However, this paper will further examine the ability of the speech based measurements in distinguishing depressed speech from near-term suicidal speech.

Several studies have observed the correlation between different characteristics of prosody and speech rate with major depression. According to Monrad-Krohn [21], the definition

of prosody consists of the normal variation of pitch, stress and rhythm which includes silent intervals of pauses. Alpert [18] on the other hand separated speech productivity and pausing under the term fluency and defined prosody as emphasis and inflection. Speech rate comprises a combination of phonation length (voiced), frequency of short pauses and the duration of pauses. The study reported herein focuses on the use of certain features related to the rhythm, fluency of speech and speech rate in an attempt to capture information related to voiced and silent pauses and quantify these features as an indicator of depressed and high risk suicidal speech.

Patients experiencing major depressive disorder often experience psychomotor retardation and cognitive disturbances [16], [22]-[24]. One of the effects observed is abnormalities in verbal productivity such as slower speech rate and increase pause time in between responses. Psychomotor retardation occurs due to the condition in which the brain has difficulty in communicating with the rest of the body, thus increasing the response time and radically reducing muscle activity. The disturbances of the interactions between numbers of neuromuscular systems thus affect the motor execution and production of speech. On the other hand, cognitive function relates to impairment in attention, information processing, working memory and decision-making processes. Cognitive impairment might also affect the number and duration of speech pauses as opposed to articulation due to hesitancy and reduction in attentiveness.

A number of studies [18], [26]-[28], [31] have examined the effect of depressive symptoms' severity on speech pauses and phonation. Methods of recordings include non-spontaneous speech where a patient counts from 1 to 10 [26], readings of standardized text [27, 31] such as the *grandfather passage* and collection of pauses that occurred in between a series of questions and answers during an interview with the clinician [18, 28]. The outcomes were consistent from one study to another where they reported that during the period of improvement, the patients exhibited a decrease in pause time and displayed no significant changes in phonation time. Although the use of pauses within counting and text reading revealed a positive relationship with depression, the effect of shorter pause time after improvement might also be connected to the practice effects from repeated events of measurements. Thus according to [16], it should be considered with some caution. On the other hand, a constant length of pause time that was observed in the control and healthy patients throughout a period of improvement might suggest that pauses in speech are independent of the practice effect [26].

Investigation on the correlation between the pauses in speech and the patient's clinical rating evaluations such as the 17-item Hamilton Depression Rating Scale (HAMD) and the Retardation Rating Scale for Depression (ERD) has recently attracted the interest of researchers in this field. The method of Pearson product-moment [30] and Spearman correlation [32] coefficients were adapted in these studies. According to a study in [29], speech pause time was shown to be significantly correlated to the ERD scores as opposed to the HAMD score. On the contrary, [30] reported moderate correlation and [16], [31]-[33] reported significant correlation between speech features that are related to pauses and the HAMD score. Among these highly correlated features are the total recording duration, total pause time, variability of pauses, vocalization to pause ratio, speaking rate and minimal fundamental frequency. However, [16] only demonstrated the correlation for their female subjects but not on the male subjects which is most likely due to small number of samples. A study performed by [32] investigated acoustic features within the phonemes of speech signals and their relationship with individual symptom sub-topic ratings of each 17 HAMD score. The paper claimed that changes in speech patterns correlate with different HAMD symptom sub-topic ratings.

This paper investigated whether the information contained in the speech timing based measures are able to discriminate the high risk suicidal (HR) speech from the depressed (DP) speech. In the effort to address this question, we initially identified that the summation of pauses and the summation of vocalizations in the previous studies were manually collected. The study in [34, 35] extracted the switching pauses (silence between turns) by manually transcribing the recordings and then forced-aligned the speech in order to obtain start time, stop time and utterance. In this study, we introduced a new approach to represent the pauses and vocalization using the Markov model and also constructing a histogram using the voiced, unvoiced and silent sections in a speech signal. We refer to these features as Transition Parameters and Interval pdf, respectively.

This distinction would be highly useful for use in real-world applications, especially when using a method that is unobtrusive to patients and practicable for the use of researchers and clinicians. Our focus is to address concerns from the clinical side of the problem and to find viable acoustic features in speech that have good reliability and can make the clinically critical separations between DP and HR.

3.0 Database

3.1 Database Collection

All recording sessions were conducted at the Vanderbilt Emergency Department or Psychiatric Hospital with patient's documented informed consent. Patients who volunteered were made aware of the aim of the study with assurance of maximum identity protection procedures. Patients under the influence of alcohol, toxicity or experiencing respiratory problems such as shortness of breath were excluded. All recordings were made in a standard, empty psychiatric interview room without the benefit of soundproof or acoustically ideal environment, mimicking the real-world clinical environment. For the purpose of this research, only the automatic speech gathered from the high risk suicidal and depressed patients were used. Group assignment was made according to assessment made by experienced clinicians using the Hamilton Depression Rating Scale (HAMD), Beck Depression Inventory (BDI-II), MINI International Neuropsychiatric Interview and Pierce Suicide Intent Scale (SIS) [41]. Patients were asked to read from a standardized "rainbow passage" which contains every sound in the English language and is considered to be phonetically balanced with the ratios of assorted phonemes similar to the ones in normal speech [42]. In automatic speech, variations in phonemes and articulation can almost be eliminated because each patient was reading from the same passage. According to Ellgring [16], the use of automatic speech disregards the involvement of complex cognitive planning processes and variation of the pause time is emphasized.

Two types of databases were used for this study. All speech samples were digitized using a 32-bit analog to digital converter at 44.1 kHz sampling rate for both databases. Table 3.1 shows the information regarding the two databases. In the first database (Database A), we only used the reading speech for the purpose of performing this analysis. Recordings were collected once per patient and each recording was categorized as either high risk suicidal or depressed. Audio acquisitions were made using a high-quality Audix SCX-one cardioid microphone with a frequency response of 40Hz to 20kHz, Sony VAIO laptop with Pentium IV 2GHz CPU 512 Mb memory, Windows XP, a Digital Audio MBox for digital audio interface and recording software PROTools LE for the digital audio editor.

In the second database (Database B), some patients attended three recording sessions, some had two recordings and the rest only had one recording. The second and third recordings were collected a few days after receiving treatments. Group assignment was determined for each

recording using clinical information in written form by an experienced clinician while blind to the recording sounds. The rating scale content was supplanted, when necessary, by the interviewer's notes. The interviewer conducted all rating scales with each recorded interview. Strong inter-rater reliability was maintained for the first set of recordings by regular training sessions. The second set ratings were all performed by one interviewer (who was supervised closely by a trained clinician). Entry criteria restricted inclusion to patients who were labeled as high risk suicidal during their first recording session. The state of the patient during the next recording sessions were made blind to the acoustic engineering researchers and categorized as *others*. In this case, *other* may or may not indicate that the patient is no longer considered high risk. Audio acquisitions for Database B were made using a portable high-quality field recorder, a TASCAM DR-1, with a frequency response of 40Hz to 20kHz, Samsung Q40 laptop with Intel Core i5 2.4GHz 4G memory and Windows 7.

Table 3.1 Information on Database A and Database B

Database A	**Male**		**Female**	
Total number of patients	HR*	DP*	HR	DP
	7	12	10	18
Database B	**Male**		**Female**	
Total number of patients	8		13	
Total number of recordings	18		32	
• Number of patients with three recordings • Number of patients with two recordings • Number of patients with one recording	4 2 2		7 5 1	
Number of recordings labeled as HR	7		12	
Number of recordings labeled as *others*	11		20	

*HR - High Risk Suicidal; DP - Depressed

3.2 Data Pre-processing

In the preprocessing stage, recordings were edited using a free audio digital editor called Audacity 2.0.1 to remove any identifying information, to preserve patient privacy. Undesirable sounds such as the interviewer's voice, voices other than the patient, sneezing, coughing and door slams were removed from the de-identified recordings. The reading passages are unedited in order to preserve the lengths of the silent intervals. For this study, short and long pauses are both important information that needed to be preserved thus, they were kept unedited. Each

edited speech sample was detrended by subtracting the mean value to compensate for possible variability that exists during recording.

4.0 Methodology

4.1 Voiced, Unvoiced and Silence Extraction

Speech signals comprise a mixture of voiced, unvoiced, and silence intervals. Voiced, unvoiced and silence speech samples can be estimated by segmenting the sampled signals based on their energy values. Voiced speech samples exhibit the quasi-stationary behavior and are composed of low frequency characteristics. On the other hand, unvoiced speech samples exhibit noise-like behavior and contain higher frequencies. A voiced/unvoiced/silence decision was made for each frame based on the method in [9].

4.2 Feature Extraction

4.2.1 Transition Parameters

A sampled signal contains a combination of voiced, unvoiced and silence frames that were represented as three different states labeled 1, 2, and 3 respectively. The words spoken in all sampled signals are the same (using the "rainbow passage") but the variation occurs in the timing pattern of the speech. The idea is to capture the variations in the form of transition from one state to another. These states can interchange with each other or perhaps return to the same state according to a set of probabilities that pertains to the states. The probabilities were estimated with a method of an observable discrete-time Markov process [43] implemented using the statistics toolbox available in MATLAB. The state and sequence are initially known with the emission probabilities set to be the matrix identity. One of the output parameters given is the estimated transition matrix (T) which in this case is a three-by-three matrix where $t_{ij} = \Pr(X_{k+1} = j \mid X_k = i)\ for\ i = 1,2,3\ and\ j = 1,2,3$. Each row i is a conditional probability given that you are in state i at time k and column j is the possible next state at time k+1. For example, t_{13} denotes the conditional probability of going from a voiced frame to a silent frame (voiced-to-silence).

$$Transition\ Matrix, T = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix}$$

The nine features were concatenated into a row vector representing each patient as $\{t_{11}, t_{12}, t_{13}, t_{21}, t_{22}, t_{23}, t_{31}, t_{32}, t_{33}\}$.

4.2.2 Interval Length Probability Density Function

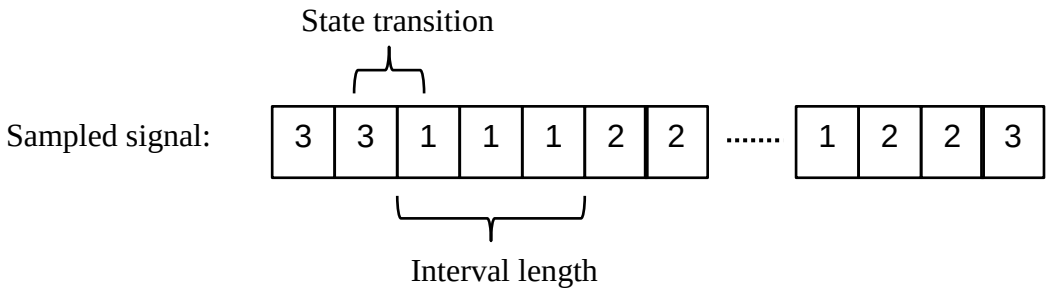


Figure 3.1: *Graphical representation of state transition and interval length pdf in a sampled signal*

The idea behind this feature is to observe any variations in the distribution of the number of frames per interval for each collection of voiced, unvoiced and silence intervals. Some questions that arise regarding the distribution of intervals are whether a person in high risk suicidal or depressed holds their vowels longer? Do they slur their speech and end up producing longer unvoiced segments or do they have longer silences? The shape of the pdf describes where the variability mostly occurred. The pdf is estimated by counting the number of occurrence for a consecutive number of 40ms frames per interval, that belongs to voiced, unvoiced or silence, that are mixed within a sampled signal. The implementation procedure to obtain the Interval Length pdf for a sampled signal is as follows:

1) For the Interval Length pdf of voiced intervals, find all the voiced intervals in the signal. Figure 3.1 shows a voiced interval (denoted by 1's) of the length three.
2) Count all the intervals of length one (40ms) and divide by the total number of voiced intervals for normalization.
3) Do the same for voiced interval of lengths two (80ms) through 24 (0.96 sec) and normalize.

4) Count all the intervals of length 25 (one sec or longer) and normalize. At this point, you have a vector of interval length percentages, i.e., a histogram.
5) Repeat step 1-4 for unvoiced (labeled '2') and silence (labeled '3') for a maximum of 0.24s (six frames per interval) and 2.0s (50 frames per interval) respectively.

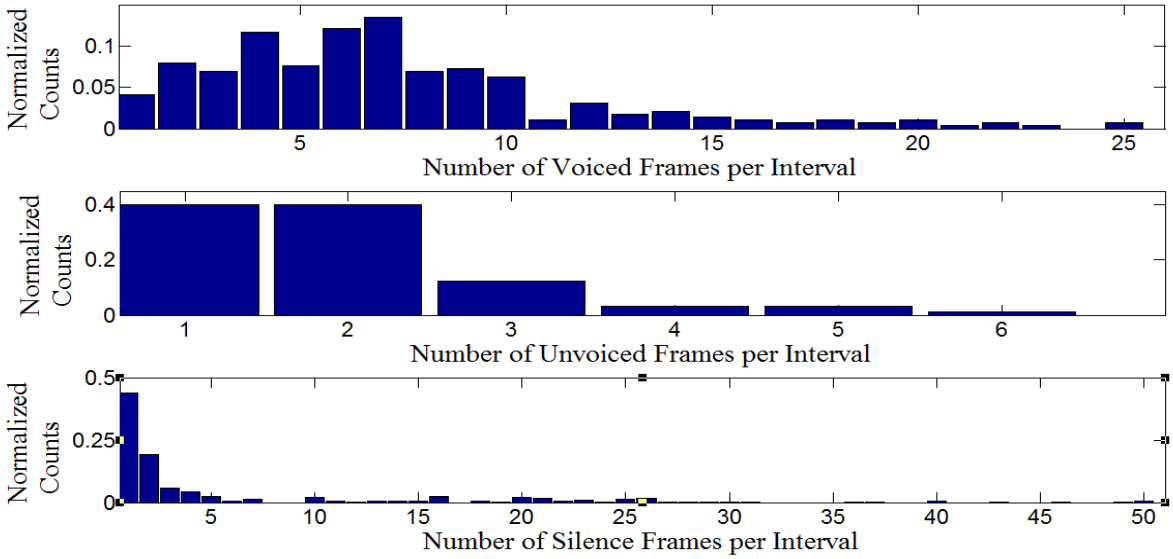


Figure 3.2: *Examples of the voiced, unvoiced and silence interval pdf distributions*

Examples of the resulting pdfs are shown in figure 3.2. Each bin is treated as a feature. For the silence interval distribution, every five consecutive interval ratios were combined in order to reduce the number of features from 50 to 10. Therefore, each bin in the silence interval distribution represents multiple numbers of interval lengths that occurs within an increment of 0.2s.

4.3 Quadratic and Linear Classifier

The discriminant analyses performed on the acquired features were done on the basis of pairwise analysis classification of high risk suicidal and depressed. The decision boundaries for the two-class classification were obtained using a quadratic classifier and a linear classifier. Because of having small data sets, a resampling method was necessary to be used when performing linear and quadratic classifications. The resampling methods that were adopted in

this research were Equal Test-Train, Jackknife (Leave-One-Out) and Cross-Validation. The discriminant functions were applied using the "classify" command provided in the MATLAB statistical toolbox, which implements the linear and quadratic classifiers as you find described in [46].

4.4 Methods of Resampling

4.4.1 Equal-Test-Train

Classifications were first performed using the quadratic and linear classifier with the resampling method of Equal Test-Train data where all data in the training set are also used for testing. In other words, this would be an optimistic estimate due to the fact that the testing data are duplicates of the training data. When testing on the same data that is train on, it does not really reflect the performance of the classifier on realistic problems. However, it does show whether the data can really be separated. To verify the accuracy of the classifier model, other resampling methods are applied.

4.4.2 *Jackknife (Leave-One-Out)*

The use of the jackknife is to show whether information obtained from or within a subpopulation can predict the behavior of the unknown individual. It gives a basic idea as to whether the classifier has good predictive characteristics. The overall set consists of *N* patients from all classes to be classified. The implementation of the jackknife method in this research is on the basis of leave-one-out patient. The procedure involves leaving out one patient's data from the data set and develops a training data set with the remaining *N*-1 patients. The excluded patient's data is then tested. This process is repeated by excluding the next patient from the overall set of data until all patients have been chosen as testing data.

4.4.3 Cross-validation

When performing classification, it is best to have as much training data as possible to prevent instability and high variance in parameter estimation. Cross-validation is an effective resampling method without replacement for the problem of small data sets. It is similar to the jackknife technique, but instead of predicting the behavior of a single individual, cross-validation is used to predict the behavior in another small subpopulation. When performing classification,

the data sets are partitioned into two sets of samples for testing and training. For this study, the partitioned samples were chosen to be 30% testing data and 70% training data. The testing data are chosen randomly from the original data sets. Similar to the jackknife method, the sample data were chosen according to patients. Using a random pick, 3 patients were chosen from class 1 (ω_1) and another 3 patients from class 2 (ω_2), and the resulting six patients were excluded from the training data.

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

4.5 Two Stages Classification Analysis

The classification analysis was divided into two stages where in stage 1, the classification analysis was performed within the Database A (Section 3.0), to determine how well the two pairwise groups of suicidal/depressed were able to be separated when using the features proposed. The classifications were performed on a single and multiple combinations of features using the three methods of resampling within each feature category (i.e., Transition Parameters, voiced, unvoiced and silence interval pdf). In the search for an optimal performance of classification, features that yielded the best classification result within each category were then combined

For stage 2, features that were recognized to perform well according to analysis results in stage 1 were then used as the features to identify recordings where the patient was initially in the condition of high risk suicidal in Database B. The rest of the recordings in Database B were labeled as 'other'. Classification of suicidal/others were implemented by treating all patients in Database A as the training data and one recording in Database B as the test data. This process is repeated until all recordings in Database B have been chosen as the test data. This method will determine how well the information from Database A translates to Database B, and to see if it is possible to classify patients from Database B with prior knowledge from a different subpopulation (Database A).

5.0 Results

5.1 Stage 1: Analysis on a Subpopulation of Database A

5.1.1 Statistical Analysis

Table 3.2 displays the estimated means and standard deviations of the nine Transition Parameters and the Interval pdf of voiced, unvoiced and silence in units of interval frames collected from recordings in Database A. Since each row in the transition matrix adds up to one, the probabilities in a row vector interrelate with each other, implying that if silence-to-silence is larger, silence-to-voiced and silence-to-unvoiced will also be affected. The mean and standard deviation of the Interval pdf are given in units of frames which can be translated into time by multiplying the number of interval frames by 40ms.

Table 3.2 Mean and standard deviation of the nine Transition Parameters and the Interval pdf of voiced, unvoiced and silence for recordings in Database A

Transition Parameter	**Mean and Standard Deviation**			
	Male		**Female**	
	HR	**DP**	**HR**	**DP**
$\mathbf{t_{11}}$	0.8450 ± 0.0285	0.8011 ± 0.0298	0.8167 ± 0.0292	0.7955 ± 0.0413
$\mathbf{t_{12}}$	0.0098 ± 0.0072	0.0161 ± 0.0122	0.0066 ± 0.0053	0.0039 ± 0.0028
$\mathbf{t_{13}}$	0.1451 ± 0.0326	0.1828 ± 0.0309	0.1767 ± 0.0296	0.2006 ± 0.0403
$\mathbf{t_{21}}$	0.1385 ± 0.0960	0.1979 ± 0.0600	0.2418 ± 0.0416	0.2226 ± 0.0569
$\mathbf{t_{22}}$	0.4511 ± 0.1975	0.5020 ± 0.0915	0.4360 ± 0.0451	0.4032 ± 0.0882
$\mathbf{t_{23}}$	0.2855 ± 0.1908	0.3001 ± 0.0501	0.3222 ± 0.0465	0.3742 ± 0.0946
$\mathbf{t_{31}}$	0.1524 ± 0.0320	0.2020 ± 0.0363	0.1790 ± 0.0313	0.2050 ± 0.0446
$\mathbf{t_{32}}$	0.0286 ± 0.0197	0.0553 ± 0.0141	0.0390 ± 0.0134	0.0359 ± 0.0210
$\mathbf{t_{33}}$	0.8190 ± 0.0428	0.7428 ± 0.0486	0.7820 ± 0.0345	0.7591 ± 0.0575
Interval pdf	**Mean and Standard Deviation (Interval of Frames)**			
Voiced	5.7143 ± 1.1127	4.5000 ± 0.6742	4.6364 ± 0.9244	4.2778 ± 1.1785
Unvoiced	1.8571 ± 0.3780	1.9167 ± 0.5149	1.4545 ± 0.5222	1.4444 ± 0.5113
Silence	2.1429 ± 0.3780	2.0000 ± 0	2.0909 ± 0.5394	2.2778 ± 0.5745

Both male and female speech revealed a higher voice-to-voice (t_{11}) and silence-to-silence (t_{33}) mean transition probability in high risk suicidal group compared to depressed group. The

high mean transition probability of silence-to-silence (t_{33}) indicates that the silence pauses are longer and the high mean transition probability of voiced-to-voiced demonstrated that patients were inclined to hold their vowels longer. These behaviors were also demonstrated by the larger mean value of voiced and silence intervals in high risk suicidal speech when compared to the depressed speech, with the exception of female silence.

For unvoiced-to-unvoiced (t_{22}), male depressed speech and female high risk speech exhibited a higher occurrence of unvoiced, which may indicate that male depressed and female high risk suicidal patients experienced more sluggishness in speech. The same trends were observed in the Interval pdf of unvoiced speech. Although the mean and standard deviation does show some correlation or a redundancy in the information between the Transition Parameters and Interval pdf, the shape of the overall Interval pdf does contain information that is distinct from the information conveyed through the Transition Parameters.

5.1.2 Classification of High Risk Suicidal and Depressed Speech in Male Reading

A. Transition parameters

Table 3.3 presents two finest performance obtained by classification using Silence-to-Voiced (t_{31}) and Voiced-to-Silence (t_{13}) for discriminating between high risk suicidal and depressed in male reading speech.

Table 3.3 Results for male reading speech classification using Transition Parameters

Transition Parameter	Feature: **Silence-to-Voiced (t_{31})**		
	All-Data %	High Risk %	Depressed %
Equal-Test-Train	74	71	75
Jackknife	74	71	75
Cross-Validation	73	73	72
	Feature: **Voiced-to-Silence (t_{13})**		
	All-Data %	High Risk %	Depressed %
Equal-Test-Train	74	71	75
Jackknife	74	71	75
Cross-Validation	71	69	73

The all-data percentage indicates vectors that are correctly classified over both groups. High risk and depressed percentages denote the percentage of vectors that are correctly classified within each group respectively.

Results show that the classifier works equally well in classifying both high risk and depressed using linear and quadratic classifiers. Also, equal performance was demonstrated by all methods of resampling. Approximately five of the seven (~70%) high risk suicidal patients were correctly classified as suicidal and about nine out of 12 (~75%) depressed patients were correctly classified as depressed using all methods of resampling.

B. Interval pdf

Classification analysis using Interval pdfs were divided into three parts; voiced, unvoiced and silence. Classification on unvoiced features did not yield good results. For voiced and silence intervals, classification was performed with every single feature (i.e., a histogram bin) from the collection of 25 voiced bins and 10 silence bins. The analysis continues for all possible combinations of two and selected combinations of three features. The number of occurrences that a histogram bin contributes to a classification performance within the range of 75% to 100% using the jackknife procedure is represented in figure 3.3(a) and figures 3.3(b) for voiced and silence intervals. This suggests which portions of the pdfs contain the most information.

Referring to the histogram in figure 3.3(a) and 3.3(b), the discriminative information in the male reading speech occurred when patients hold their vowels for a range of time intervals from 0.16s (eight consecutive frames) to 0.48s (12 consecutive frames) with a peak at an interval of 0.36s (nine consecutive frames). On the other hand, silence pauses that occurred within an approximately 40ms (one frame) to 1.2s (30 consecutive frames) time interval contained most of the information relating to the variability characteristics in the speech of high risk and depressed.

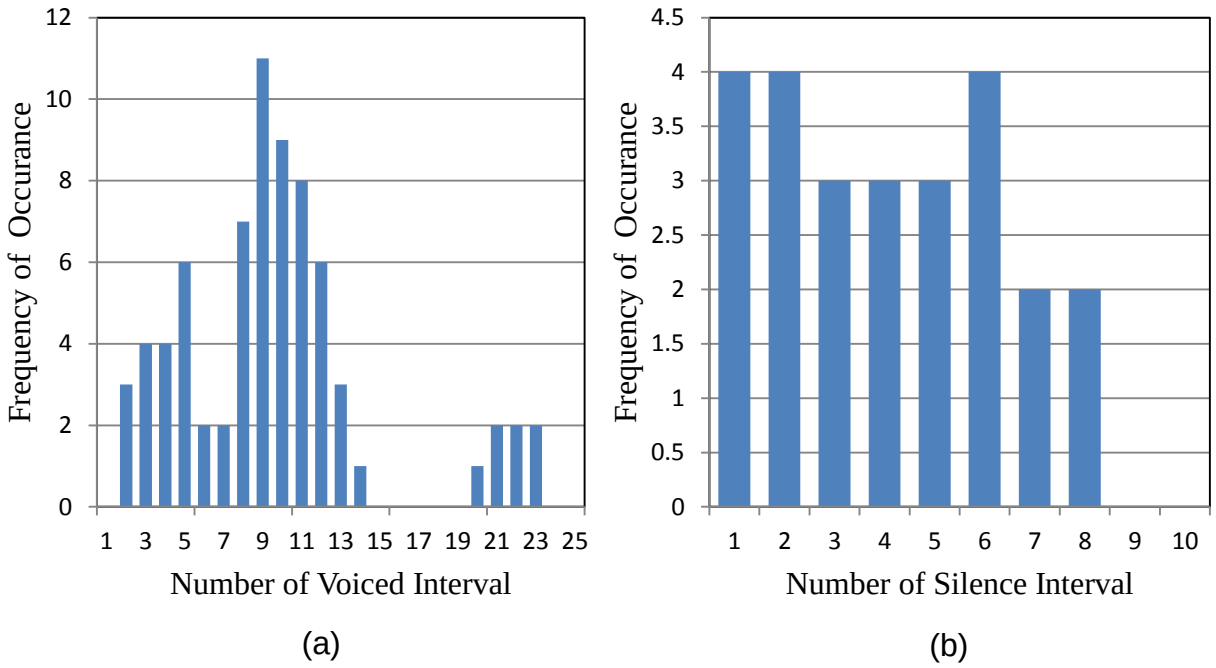


Figure 3.3: *Histogram of the individual (a) 25 voiced interval ratios and (b) 10 silence interval ratios that contributed 75% to 100% correct jackknife classification using a single and/or combination of features for male high risk and depressed speech.*

C. Combined Feature Set

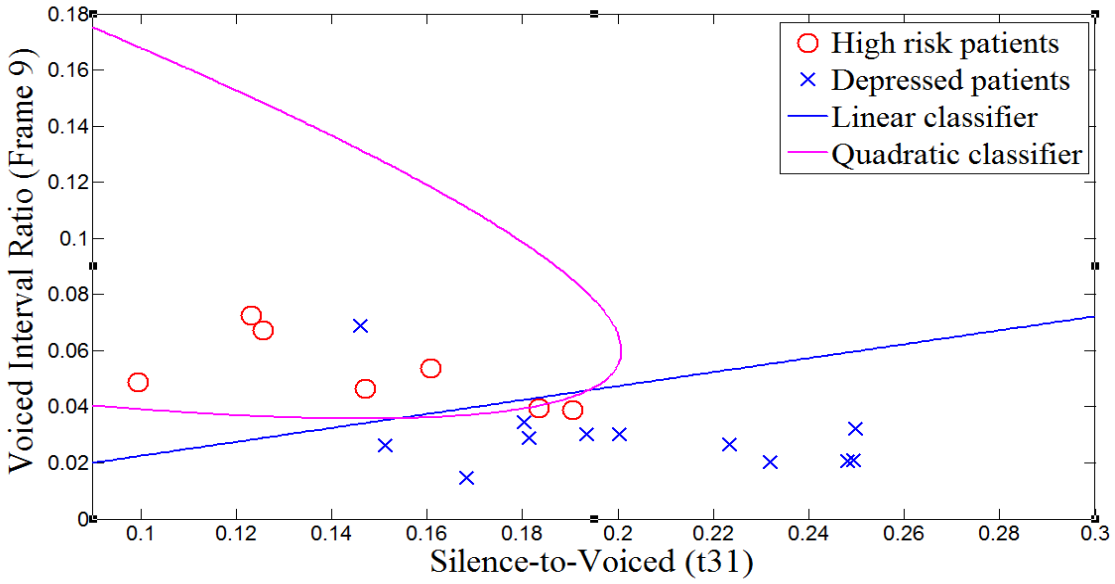
A combined feature set can increase, retain or decrease the performance of the classifier depending on the variability of the information. The variability that exists within each feature can either complement or nullify each other. Even if the variability is high (i.e., high classification performance from each set), both features can still be carrying similar information and maintain the performance as it is.

We performed a classification analysis using combinations of Silence-to-Voiced (t_{31}) with each single feature of eighth to 12th voiced bin and the first to sixth silence bin. The same procedure was then repeated for Voiced-to-Silence (t_{13}). We identified that classification on a combined feature set that comprised of Silence-to-Voiced (t_{31}) outperformed Voiced-to-Silence (t_{13}). The ninth voiced interval (Voiced9) and fourth silence interval (Silence4) produced the best classification when combined with Silence-to-Voiced (t_{31}) as shown in table 3.4. The overall results demonstrated that the classifier performed better on the depressed speech compared to the high risk suicidal speech

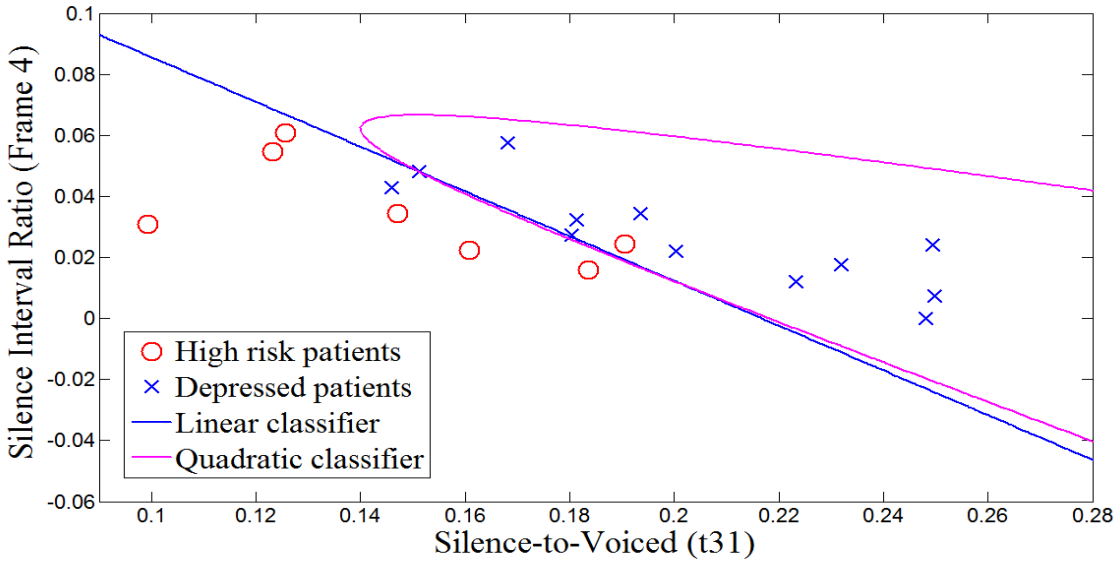
Table 3.4 Results of the combined feature sets classification for high risk and depressed male reading speech

	Feature: $\mathbf{t_{31}}$ **+ Voiced9**			
	All-Data %	High Risk %	Depressed %	Classifier
Equal-Test-Train	84	71	92	Linear/Quadratic
Jackknife	84	71	92	Linear
Cross-Validation	79	71	88	Linear
	Feature: $\mathbf{t_{31}}$ **+ Silence4**			
	All-Data %	High Risk %	Depressed %	Classifier
Equal-Test-Train	89	86	92	Linear/Quadratic
Jackknife	74	71	75	Linear
Cross-Validation	79	72	85	Linear

Figure 3.4(a) and 3.4(b) plot the distribution of high risk and depressed patients using the combined feature set. By observation, the distributions of high risk patients and depressed patients were distinct from each other and vectors that are misclassified were fairly close to the boundary except for one of the depressed patients as shown in figure 3.4(a).



(a)



(b)

Figure 3.4: *Plot of the high risk and depressed patient distribution for the combined feature set of Silence-to-Voiced (t_{31}) with (a) voiced interval ratios in frame 9 and with (b) silence interval ratios in frame 4 using linear and quadratic discriminant classifier.*

5.1.3 Classification of High Risk Suicidal and Depressed Speech in Female Reading

A. Transition Parameters

An analysis of a single and multiple combinations of features from Transition parameters were performed on the classification of high risk and depressed for female reading speech. Classification using linear discriminant function on a single feature of Voiced-to-Silence (t_{13}) from the Transition Parameters revealed the best classifier performance for female reading speech as shown in table 3.5. Overall, the classifier performed equally well in classifying both high risk and depressed patients.

Table 3.5 Optimal Result for high risk and depressed female reading speech classification using Silence-to-Voiced (t_{31})

Transition Parameter	Feature: **Voiced-to-Silence (t_{13})**		
	All Data %	High Risk %	Depressed %
Equal-Test-Train	72	73	72
Jackknife	72	73	72
Cross-Validation	71	72	70

B. Interval pdf

Classification on unvoiced and silence frame intervals did not reveal good results. However, shown in table 3.6, classification using quadratic discriminant function with a single feature of 16 voiced frame per interval (Voiced16) and a combination of 16 to 20 voiced frames per interval (Voiced16:20) produced the best classification performance in identifying between high risk suicidal and depressed female patients.

Table 3.6 Optimal result for high risk and depressed female reading speech classification using interval pdf

Interval pdf	Feature: **Voiced16:20**		
	All-Data %	High Risk %	Depressed %
Equal-Test-Train	93	82	100
Jackknife	79	82	78
Cross-Validation	75	65	84
	Feature: **Voiced16**		
	All-Data %	High Risk %	Depressed %
Equal-Test-Train	83	91	78
Jackknife	79	82	78
Cross-Validation	76	83	69

The classifier performed equally effective on Voiced16:20 and Voiced16 using the jackknife method. Cross-validation on the other hand effectively classified depressed speech using Voiced16:20 and high risk speech when using Voiced16. However, the higher percentage of correctly classified high risk suicidal patients is more preferable because identifying high risk is more critical than depressed. Therefore, if we are required to choose between the two features, Voiced16 is preferable because of the higher percentage of correctly classified high risk suicidal.

C. Combined Feature Set

Results of the classification analysis on female reading speech using the combination of Voiced-to-Silence (t_{13}) with the 16^{th} to the 20^{th} voiced bins and another combination with only the 16^{th} voiced bin are demonstrated in table 3.7. However, we observe that classification using the Interval pdf feature by itself performed remarkably better compared to the single Transition Parameter and the combined feature set.

Table 3.7 Results of the combined feature sets classification for high risk and depressed female reading speech

	Feature: **t_{13} + Voiced16:20**			
	All-Data %	High Risk %	Depressed %	Classifier
Equal-Test-Train	93	82	100	Quadratic
Jackknife	76	64	83	Quadratic
Cross-Validation	65	47	84	Quadratic
	Feature: **t_{13} + Voiced16**			
	All-Data %	High Risk %	Depressed %	Classifier
Equal-Test-Train	79	82	78	Quadratic
Jackknife	72	64	78	Quadratic
Cross-Validation	71	66	76	Quadratic

5.2 Stage 2: Analysis of Classification between Two Populations

5.2.1 Testing Classifier on Database B for Male Reading Speech

In the second stage, we test the ability of the classifier to identify high risk recordings in Database B using the trained features and classifier that we analyzed in stage 1 on Database A. As shown in table 3.8, the high percentages in group *other* cannot be verified because except for the high risk recordings, the rest of the recordings in Database B were made unknown to the researcher. Thus the recordings labeled *other* can be a mixture of high risk, depressed or remitted. Using the feature of Silence-to-Voiced (t_{31}) alone, seven out of eight recordings that were labeled as high risk were successfully identified. The results improve significantly to a perfect identification of high risk suicidal when Silence-to-Voiced (t_{31}) was combined with Voiced9 and Silence4. Remarkably, the classifier that was trained using only Voiced9 produced similar result as the combined features.

Table 3.8 Results of the tested classifier for the identification of high risk suicidal recordings in male patient database B

Feature Combination	All Data %	High Risk %	Other %	Classifier
t_{31}	83	86	82	Quadratic
Voiced9	94	100	91	Linear/Quadratic
t_{31} + Voiced9	94	100	91	Linear/Quadratic
t_{31} + Silence4	89	100	82	Quadratic

5.2.2 Testing Classifier on Database B for Female Reading Speech

Although classification results of the trained classifier in male reading speech were outstanding, features obtained from female reading speech did not translate well between the two populations. However, Transition Parameters and Interval pdf do work effectively within each database but with different features.

6.0 Discussion

This paper presented new methods of extracting features based on the timing patterns of speech using the Markov Transition Matrix and the Interval pdf of voiced, unvoiced and silence for the analysis of vocal characteristics in high risk suicidal and depressed detection. Previous research [16]-[19], [27]-[28], [30]-[32], [42] has reported the importance of information contained in the patterns of voiced speech and silent pauses as a possible indicator of depressive illness. The results of this investigation correlate with the previous findings where it was shown that features relating to voice and silence from Transition Parameters and Interval pdf provided prominent results in classification of high risk suicidal and depressed patients. Besides the preliminary process of separating voiced, unvoiced and silence segments, the process of obtaining the Transition Parameters and Interval pdf are not related to the spectrum based measures.

The Transition Parameters represents the decision or transition probabilities between speech frames. Results demonstrated that information on the distinguishing characteristics of high risk suicidal and depressed in both male and female are mostly embedded in the transition probabilities of silence and voiced speech. Silence-to-Voiced(t_{31}) and Voiced-to-Silence(t_{13}) was found to be the most significant features in distinguishing between high risks suicidal and depressed male patients and the latter to female patients. However, the strong consistency of this feature can be observed at least for the male patients and thus will be discussed in further detail. The probability that the current frame is silence and the next frame is voiced is affected by the length of the silent pauses because a row in the transition matrix sums to one. If silent pauses are longer, the probability of silence-to-silence will increase and will force other probabilities to decrease. The probability of the observed significant feature that revolves around the interaction between silent and voiced frames can also be affected by either longer voiced sections or longer silences. Therefore, four features of Voiced-to-Voiced (t_{11}), Voiced-to-Silence (t_{13}), Silence-to-

Voiced (t_{31}) and Silence-to-Silence (t_{33}) can be used to analyze the characteristic between high risk suicidal and depressed. Referring back to table 3.2, the means of the inter-transitions between silence and voice for depressed patients were higher than the high risk patients thus signifying a more frequent and active start and stop in depressed speech. The higher means in the intra-transitions within silence and voiced for high risk patients indicate a slower speech rate, holding out vowels longer and taking longer time for pauses.

The Interval pdf describes the overall shape of the distribution within voiced, unvoiced and silence where longer intervals are expected to have more variability. The overall pdf for voiced, unvoiced and silence exhibit similar characteristic of an asymmetrical right-skewed distribution. The distribution's peak is off centered with a tail stretching in the opposite direction away from it. Most variability occurs in the tail end shape of the pdf which are demonstrated by the results obtained from both male and female categories. For male interval distribution, the significant feature of the nine voiced frame per interval and the four silence frame per interval were located in the tail direction away from their mean. Similarly for female interval distribution, the significant feature of 16 to 20 voiced frames per interval was nearly close to the tail end shape of the pdf.

Only a small feature set was used to generate a strong performance, developed using two completely different databases. One database was used to train the paradigm, and it was tested on the second dataset. We were able to find a single Transition Parameter Silence-to-Voiced (t_{31}) and the ninth bin of voiced interval pdf (Voiced9) that produced 86% and 100% separation on high risk recordings, respectively. Also, using two combined parameters of Silence-to-Voiced (t_{31}) with the ninth bin of voiced interval pdf (Voiced9) and combination of Silence-to-Voiced (t_{31}) with the fourth bin of silence interval pdf (silence4), both revealed 100% separation on high risk recordings. The fact that only one or two parameters were able to produce the quality of discrimination and also perform across two datasets recorded in different environments (a variety of clinical interview rooms) using different devices strengthens the argument that these results are not coming from over-modeling or from spurious environmental factors. It is a strong indication that there is significant information within these parameters. Additionally, the parameters are easily calculated.

Automatic speech was used in this study due to the consistency in the spoken words. Each patient was saying the exact same words. Patients were given the standard "rainbow

passage" essay that contains all phonemes found in the English language. However, the difference in the transition probability might have existed because of the variability in the decision made between voiced, unvoiced and silence. For the same word, some that are marked as voiced in one patient might have been marked as something else in another patient. This can contribute to a low classification result when a trained classifier is tested on a different population as demonstrated by the female group. Aside from that, female speech has been reported to be more breathy than male. Breathy voice quality occurs because of the incomplete closure of vocal folds that allows air to flow through the glottis and thus presents an existence of noise in the higher frequency spectrum, domination of harmonic excitation by aspiration noise and alternations in vocal tract which are shown by the extra poles and zeros in the vowel spectrum [44]. Therefore, a word that should have ended with a full voicing might be influenced by the aspiration.

When discussing the issue of small sample size, it is important to keep the dimensionality of studied features low. High dimensional feature spaces often lack generalization and can lead to over-modeling the limited dataset. Results could become highly questionable when using huge dimensional spaces on a very small amount of data because it could easily be modeling things that are not the characteristics of general population but just the individuals of the small data sets. Thus, it might work well for the corresponding dataset but failed to generalize well to novel datasets. According to a rule of thumb for an adequate sample size, an appropriate number of samples per estimated feature are of the 5:1 ratio [45]. Nevertheless, this study only used one and/or two features to obtain those results within a number of sample set that varies from 18 to 32. This suggests that there is valuable information embedded in the small number of features relative to the data.

The two databases were recorded at different times, during collection intervals that used two different types of recording devices (Audix SCX-one cardioid and TASCAM DR-1). The fact that the trained classifier performed so well when tested on the second database of male participants, demonstrated that the features were not affected by different recording devices.

References

[1] S. L. Murphy, J. Q. Xu, K. D. Kochanek, "Final Data for 2010. National Vital Statistics Reports", Hyattsville, MD: National Center for Health Statistics, vol. 61, no. 4, 2013.

[2] M. Heron, "Deaths: Leading Causes for 2009: National Vital Statistics Reports", Hyattsville, MD: National Center for Health Statistics, vol. 61, no. 7, 2012.

[3] R. Burns, "An Impact: Suicides are Surging among US Troops", *The Associated Press*, 2012.

[4] C. M. Perlman, E. Neufeld, L. Martin, M. Goy, J. P. Hirdes, "Suicide Risk Assessment Inventory: A Resource Guide for Canadian Health care Organizations", Toronto, ON: Ontario Hospital Association and Canadian Patient Safety Institute, 2011.

[5] L. R. Wingate, T. E. Joiner, R. L. Walker, M. D. Rudd, D. A. Jobes, "Empirically Informed Approaches to Topics in Suicide Risk Assessment", Behavioral Sciences & the Law, pp. 651–665, 2004.

[6] K. A. Busch, J. Fawcett, D. G. Jacob, "Clinical Correlates of Inpatient Suicide", *J. Clin. Psychiatry*, vol. 64, no. 1, pp. 14-19, 2003.

[7] J. C. Fowler, "Suicide Risk Assessment in Clinical Practice: Pragmatic Guidelines for Imperfect Assessment", *American Psychological Association,* vol. 49, no. 1, pp. 81-90, 2012.

[8] D. J. France, R. G. Shiavi, S. E. Silverman, M. K. Silverman, D. M. Wilkes, "Acoustical Properties of Speech as Indicators of Depression and Suicidal Risk", *IEEE Transaction on Biomedical Engineering*, vol. 47, no. 7, 2000.

[9] A. Ozdas, R. G. Shiavi, S. E. Silverman, M. K. Silverman, D. M. Wilkes, "Investigation of Vocal Jitter and Glottal Flow Spectrum as Possible Cues for Depression and Near-Term Suicidal Risk", *IEEE Transaction on Biomedical Engineering*, vol. 51, no. 9, 2004.

[10] T. Yingthawornsuk, H. K. Keskinpala, D. M. Wilkes, R. G. Shiavi, R. M. Salomon, "Direct Acoustic Feature using Iterative EM Algorithm and Spectral Energy for Classifying Suicidal Speech", *INTERSPEECH*, pp. 766-769, 2007.

[11] H. K. Keskinpala, T. Yingthawornsuk, D. M. Wilkes, R. G. Shiavi, R. M. Salomon, "Screening for High Risk Suicidal States using Mel-Cepstral Coefficients and Energy in Frequency Bands", *In European Signal Processing Conf.*, pp. 2229-2233, 2007.

[12] K. R. Scherer, "Vocal affect expression: A Review and Model for the Future Research", *Psychological Bulletin*, vol. 99, no. 2, pp. 143-145, 1986.

[13] D. Ververidis, C. Kotropoulos, "Emotional Speech Recognition: Resources, Features and Methods", *Speech Communication*, vol. 48, no. 9, pp. 1162-1181, 2006.

[14] H. K. Rouzbahani, M. R. Daliri, "Diagnosis of Parkinson's Disease in Human using Voice Signal", *Basic and Clinical Neuroscience*, vol. 2, no. 3, pp. 12-20, 2011.

[15] A. Askenfelt, S. Nilsonne, "Voice Analysis in Depressed Patients: Rate of Change of Fundamental Frequency Related to Mental State", *STL-QPSR*, vol. 21, no. 2-3, pp. 71-84, 1980.

[16] H. Ellgring, K. R. Scherer, "Vocal Indication of Mood Change in Depression", *J. of Nonverbal Behavior*, vol. 20, no. 2, pp. 83-110, 1996.

[17] J. K. Darby, H. Hollien, "Vocal and Speech Patterns of Depressive Patients", *Int. J. of Phoniatrics, Speech Therapy and Communication Pathology*, vol. 29, no. 4, pp. 279-91, 1997.

[18] M. Alpert, E. R. Pouget, R. R. Silva, "Reflections of Depression in Acoustic Measures of the Patient's Speech", *J. of Affective Disorders*, vol. 66, no. 1, pp. 59-69, 2001.

[19] E. Szabadi, C. M. Bradshaw, J. A. O. Besson, "Elongation of Pause-Time in Speech: A simple, Objective Measure of Motor Retardation in Depression", *The British J. of Psych*., vol. 129, no. 7, pp. 592-597, 1976.

[20] E. Moore, M. A. Clements, J. W. Peifer, L. Weisser, "Critical Analysis on the Impact of Glottal Features in the Classification of Clinical Depression in Speech", *IEEE Trans. On Biomed. Eng.*, vol. 55, no. 1, pp. 96-107, 2008.

[21] G. H. Monrad-Krohn, "The Third Element of Speech: Prosody in the Neuro-Psychiatric Clinic", *The British J. of Psych.*, vol. 103, no. 431, 326-331, 1957.

[22] R. M. Lane, J. F. O'Hanlon, "Cognitive and psychomotor effects of antidepressants with emphasis on selective serotonin reuptake inhibitors and the depressed elderly patient", *German J. of Psych*, 1999.

[23] D. Schrijvers, W. Hulstijn, B. G. C Sabbe, "Psychomotor symptoms in depression: A diagnostic, pathophysiological and therapeutic tool", *J. of Affective Disorders*, vol. 109, no. 1-2, 2008.

[24] D. Marazziti, G. Consoli, M. Picchetti, M. Carlini, L. Faravelli, "Cognitive Impairment in Major Depression", *European J. of Pharmachology*, vol. 626, pp. 83-86, 2010.

[25] A. Ghozlan, D. Widlocher, "Decision Time and Movement Time in Depression: Differential Effects of Practice Before and After Clinical Improvement", *J. of Perceptual and Motor Skills*, vol. 68, no. 1, pp. 187-92, 1989.

[26] G. M. Hoffman, J. C. Gonze, J. Mendlewicz, "Speech Pause Time as a Method for the Evaluation of Psychomotor Retardation in Depressive Illness", *The British J. of Psych.*, vol. 146, pp. 535-538, 1985.

[27] A. Nilsonne, "Acoustic Analysis of Speech Variables During Depression and After Improvement", *Acta. Psychiatr. Scand*, vol. 76, no. 3, pp. 235-45, 1987.

[28] A. Nilsonne, :Speech Characteristics as Indicators of Depressive Illness", *Acta. Psychiatr. Scand*, vol. 77, no. 3, pp. 253-63, 1988.

[29] P. Hardy, R. Jouvent, D. Widlocher, "Speech Pause Time and the Retardation Rating Scale for Depression (ERD)", *J. of Affective Disorders*, vol.6, pp. 123-127, 1984.

[30] M. Cannizaro, B. Harel, N. Reilly, P. Chappell, P. J. Snyder, "Voice Acoustical Measurement of the Severity of Major Depression", *Brain and Cognition*, vol. 56, no. 1, pp.30-35, 2004.

[31] J. C. Mundt, P. J. Snyder, M. S. Cannizzaro, K. Chappie, D. S. Geralts, "Voice Acoustic Measures of Depression Severity and Treatment Response Collected Via Interactive Voice Response (IVR) Technology", *J. of Neurolinguistic*, vol. 20, pp. 50-64, 2007.

[32] A. C. Trevino, T. F. Quatieri, N. Malyska, "Phonologically-based Biomarkers for Major Depressive Disorders", *EURASIP J. on Advances in Signal Processing*, vol. 42, 2011.

[33] J. C. Mundt, A. P. Vogel, D. E. Feltner, W. R. Lenderking, "Vocal Acoustic Biomarkers of Depression Severity and Treatment Response", *J. of Biological Psychiatry*, vol. 72, no. 7, pp. 580-587, 2012.

[34] Ying Yang, F. Catherine, J. F. Cohn, "Detecting Depression Severity from Vocal Prosody", *IEEE Trans. On Affective Computing*, vol. 4, no. 2, 2013.

[35] J. F. Cohn, T. S. Kruez, I. Matthews, Ying Yang, Minh Hoai Nguyen, M. T. Padilla, Feng Zhou, F. De la Torre, "Detecting Depression from Facial Actions and Vocal Prosody", *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference*, pp. 1-7, 2009.

[36] J. M. Girard, J. F. Cohn, M. H. Mahoor, S. Mavadati, D. P. Rosenwald, "Social Risk and Depression: Evidence from Manual and Automatic Facial Expression Analysis" *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops*, pp. 1-8, 2013.

[37] J. Joshi, A. Dhall, R. Goecke, J. F. Cohn, "Relative Body Parts Movement for Automatic Depression Analysis", *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference* pp. 492-497, 2013.

[38] S. Alghowinem, R. Goecke, M. Wagner, G. Parker, M. Breakspear, "Head Pose and Movement Analysis as an Indicator of Depression" *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference*, pp. 283-288, 2013.

[39] J. Joshi, R. Goecke, G. Parker, M. Breakspear, "Can Body Expressions Contribute to Automatic Depression Analysis?", *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops*, pp. 1-7, 2013.

[40] S. Scherer, G. Stratou, M. Mahmoud, J. Boberg, "Automatic Behavior Descriptors for Psychological Disorder Analysis", *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops,* 2013.

[41] R. M. Salomon, H. K. Keskinpala, M. H. Sanchez, T. Yingthawornsuk, N. H. Nik Wahidah, W. S. Hasan, N. Taneja, D. Vergyri, B. H. Knoth, P. E. Garcia, D. M. Wilkes, R. Shiavi, "Analysis of Voice Speech Indicators in Suicidal Patients", Manuscript submitted for publication, 2012.

[42] 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.

[43] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", *Proceedings of the IEEE,* vol. 77, no. 2, pp. 257-286, 1989.

[44] D. H. Klatt, L. C. Klatt, "Analysis, Synthesis and Perception of Voice Quality Variations Among Female and Male Talkers", J. Acoust. Soc. of America, vol. 87, no. 2, 1989.

[45] H. M. Kalayeh, D. A. Landgrebe, "Predicting the required number of training samples, Pattern analysis and machine intelligence", *IEEE transaction on Pattern Analysis and Machine Learning,* vol. 5, no.6, pp. 664-667, 1983.

[46] V. Franc, V. Hlavac, Linear and Quadratic Classification Toolbox for Matlab, Czech Pattern Recognition Workshop, 2000.

[47] M. H. Sanchez, D. Vergyri, L. Ferrer, C. Richey, P. Garcia, B. Knoth, W. Jarrold, "Using Prosodic and Spectral Features in Detecting Depression in Elderly Males", *In conference of the International Speech Communication Association,* pp. 28-31, 2011.

CHAPTER IV

INVESTIGATION ON ACOUSTIC MEASURES OF SPEECH AS A POTENTIAL PREDICTOR FOR THE HAMILTON DEPRESSION RATING SCALE (HAMD) AND BECK DEPRESSION INVENTORY SCALE (BDI-II)

Abstract

Negative screenings can bring the clinician to believe that persons who are actually at imminent risk of committing suicide are experiencing a less severe (often depressive) disorder. Even for specialists, and especially for non-specialists, objective metric might signal a critical need for more extensive interviewing and other precautions. Identifying imminent suicidal risk at an early stage may prompt the clinician to propose and promote a more intensive treatment plan. In this study, we attempted to demonstrate the effectiveness of using patient's acoustic measurements as a possible means to predict ratings from well-known medical diagnostic tools known as the Hamilton Depression Scale (HAMD) and Beck Depression Inventory (BDI-II). The results are based on method of linear regression with the implementation of jackknife analysis and applying the forward (SFS) and backward (SBS) feature selection to increase the performance of the predictions. Evaluation of model performance error that measures how close predictions are to the actual scores was based on the measure of mean absolute error (MAE). The results revealed an average MAE between the predicted and the actual HAMD score was approximately two scores and the MAE for the BDI-II score was approximately one score for male and eight scores for female. This finding demonstrates the ability of using speech measures to predict the psychological condition of an individual through their clinical scores.

1.0 Introduction

Despite extensive research into reliable methods for psychiatric assessment, it still continues to be clinically challenging and scientifically deficient. Conventional methods are generally performed according to a series of questionnaires and rating scales that measure various aspects including thoughts, behaviors and symptoms that are evaluated by a trained clinician or self-reported measurements. The information gathering process is a non-static and

time-consuming process that requires the trained clinician to maintain regular interactions with the patient for obtaining accurate assessment. It is desirable to have a second assessment tool to aid the clinicians with the diagnosis of the patient's actual psychological state especially when it comes to distinguishing the state of a major depression and an imminent risk of committing suicide. Symptoms for the near-term suicidal patient are similar to depression thus leading to possible misdiagnosis of an imminent suicide.

The steadily increasing rate of suicide every year in the United State [1] has motivated researchers to investigate a possible method to reduce the statistics. The list of causes of death revealed that suicide ranks 10^{th} with 36,909 suicides, and ranks 3^{rd} for the age group of 15-24 with 4,371 suicides reported in the year 2009 [2]. To view the importance of this issue, on average one suicide occurs every 14.2 minutes in the United States. Unsuccessful suicide attempts numbered 922,725 during this period, translating to an average of one attempt every 34 seconds. Males exhibit a greater risk of death from suicide as a gender-wise analysis reported a ratio of 3.7 males to 1 female by suicide [3]. A surge in the military suicide rate occurred with 154 deaths in the first 155 days of 2012, an increase of 18 percent compared to the statistic reported for the first half of the previous year. Deaths by suicide among military personnel during this period outnumbered U.S. soldiers killed in action by an estimated two to one ratio [4]. Suicide not only has significant emotional consequences for family and friends, but there are also substantial economic costs of approximately $34.6 billion associated with medical bills and work loss [1].

Suicide are caused by a range of factors such as psychological milieu (life events, environmental factors and medical illnesses), presence of psychiatric disorders, biological factors, health records, family history and history of any previous suicide attempts [5]. Major depression is often the psychiatric diagnosis associated with suicide. Even though most people with depression do not end up committing suicide, it has been reported that more than 90 percent of the suicide victims experience depression and other psychiatric disorders [6]. Information from a second source may provide instantaneous quantitative results and thus identify imminent suicidal risk at an early stage and allowing the patient to receive proper hospitalization and treatment.

1.1 Voice acoustic as a measure of suicidality

A growing number of studies have demonstrated the identification of psychological disorders using information extracted from speech signals that are known as vocal features [9-18]. A large part of these works centered on distinguishing between groups within the major depressive disorders (MDD) or comparing among the control patients. Recognizing that suicide has a profound public health significance, Drs. Stephan and Marilyn Silverman began in the 1980s by collecting and analyzing suicidal tape recordings obtained through therapy sessions or notes and interviews made shortly before suicide attempts. They describe the similarity of the vocal speech between the depressed and the high risk suicidal patients, but observed that changes occur in the tonal quality and acoustical characteristic when a patient enters the suicidal state [21]. Later on, several studies have expand the scope of the investigation on analyzing the speech features for discriminating between the groups of high risk suicidal, depression and remission. Among the speech features that have been identified to exhibit distinguish characteristics are the Frequency (F_0), Amplitude Modulation, Power Spectral Density (PSD), Mel-Frequency Cepstral Coefficients (MFCC), formants, voice jitter and glottal flow spectral slope [9-15].

1.2 Suicide assessment by the Hamilton Depression Rating Scale (HAMD)

The most common interview scale and administered diagnostic tool to measure the severity of depression in an inpatient population is the Hamilton Depression Rating Scale (HAMD). The HAMD assessment has also been considered as the primary standard for determining suicidal risk. It contains 17-items questionnaires including one item on suicidal thoughts with rating scales that can be evaluated only by trained clinicians. Clinicians rely on their intuitions during evaluation and determining the ratings for the provided questionnaires. Generally accepted opinions by clinicians on interpretation of the total HAMD scores is that score between 0 to 7 shows no presence of depression, 8 to 13 indicates mild depression, 14 to 18 indicates moderate depression, 19 to 22 indicates severe depression and score over 22 indicates very severe depression [7]. For a single suicide item, patients scoring 2 or higher were found to be 4.9 times more likely to die by suicide [8]. Even though it has been found to be reliable, the application of clinician-administered instrument is time-consuming and requires extensive effort by clinicians to obtain repeated comprehensive evaluations. There is a risk of improper assessment due to an accidental failure to inquire about specific information relating to

suicide risk and the clinician's lack of well-defined conceptual clarity concerning suicidal and depression behavior. The measurement tool is considered inappropriate for widespread adoption such as in situations where a trained clinician is not available.

1.3 Suicide assessment by the Beck Depression Inventory (BDI-II)

Beck Depression Inventory is a reliable and valid 21-item self-rating screening tool for identifying depressive symptoms that includes one item on suicidal thoughts. Each item is assigned a score ranging from zero to three indicating the severity of the symptom. The suggested guidelines for interpreting the BDI-II total score is as follows: 0 to 13 represent minimal depression symptom, 14 to 19 indicate mild depression, 20 to 28 for moderate depression and 29 to 63 represent severe depression [35]. Patients scoring two or more on the BDI single suicide item were found to be 6.9 times more likely to die by suicide [8]. It is believed that some patients are more comfortable revealing their thoughts and feelings through the self-scale questionnaires instead of discussing such information with another individual [36]. The BDI-II test can be quickly and effortlessly administered and thus reduces the time it takes to continually observe the development of the patient's psychological state. However, there are limitations when performing a self-rating assessment. Whether in a group or an individual setting, patients that are completing the test may easily overestimate or understate the answers when trying to assess their own psychological state [37].

1.4 Significance of paper

Patients suffering from depression are more likely to seek help from the primary care physician than a psychiatrist for diagnosis and treatment relating to various somatic complaints. A study performed on suicidal behavior in the elderly revealed that there are approximately 60% of patients visited their primary care physicians within 30 days prior to committing suicide, 35% within seven days and 3% on the day before [20]. Therefore, having a second assessment tool may not only help mental health professionals, but also can be used practically by other physicians. A primary care physician that may not necessarily be a specialist in psychiatry can perform a test that would provide an additional diagnostic tool to determine whether to send the patient to a psychiatrist. Having a recording device in a doctor's office would be practical instead

of doing all recordings in a studio with a perfectly quiet environment. Another possible application is for an interactive screening program within schools and colleges. It could be used to effectively identify student that suffers from major depression and those who are at risk of imminent suicide. Early detection will allow those in charge to encourage the students to get help and seek treatment [22].

2.0 Previous Work

An early study conducted by Harvy [23] analyzed the relationship between the clinical subjective ratings and the acoustic measures in patients with major depression. The extracted acoustic measures were speech pause time (SPT), phonation time (PT) and total time (TT) and the clinical ratings were the Hamilton Depression Rating Scale (HAMD: 17 item version) and the Retardation Rating Scale for Depression (ERD). The correlation analysis was performed according to changes in the acoustic measures and changes in the rating scores that took place before the onset of treatment (D_o) and the final evaluation (D_f) made within 48 hours of discharge or change of medication. A paired t-test revealed a significant correlation between the SPT and changes in the ERD. However, changes in HAMD score did not correlate as well as the ERD score. The study also reported no significant changes for the PT and TT measures.

In a longitudinal study performed by Ellgring [24], correlation between voice parameters and the Voice Analogue Scale for Subjective Well-being (VAS) was computed separately on 5 male and 11 female depressed patients. Samples were taken from standardized interviews for depression, and analyses were chosen from samples obtained within 5 days of admission and after 50 days of treatment. The mean fundamental frequency (MF0), speech rate (SR) and mean pause duration (MPD) were found to be significantly correlated with the subjective well-being for female subjects. On the contrary, no significant correlations were identified in males, possibly due to a small number of patients and less improvement throughout therapy.

Another investigation performed by Stasen [25] was aimed at evaluating the relationship between the psychopathology scales and the changes in acoustic characteristics of speech in hospitalized depressive patients throughout the first 2 weeks of antidepressant treatment. The study demonstrated a similar development and close relationship between the two curves of the HAMD-17 depression score with a single acoustic measure of mean pause duration and

fundamental frequency amplitude. Both relationships exhibit significant correlations when measured using the Spearman rank correlation.

A study by Cannizzaro [26] attempted to replicate the results reported by Stasen [25] with an exception of using recordings that were made in less than ideal conditions and using relevant voice acoustical metrics from samples of spontaneous speech (interview). Three acoustic measures of speaking rate, percent pause time and pitch variation were extracted from 5 male and 2 female recordings. The Pearson product-moment correlation analysis revealed a significant negative correlation between speaking rate and the HAMD. Although HAMD scores demonstrated large negative correlation with pitch variation and moderate significance with percent pause time, however neither achieved statistically significant, possibly due to the small sample size. The results corroborate previous findings and demonstrated the ability of voice acoustical analysis to objectively track the severity of depression despite imperfect recording conditions.

Two separate studies were performed by Mundt [27,28] using different depression severity measures and different methods of assessment for 35 depressed and non-depressed patients. In the initial study, pitch variability across the second harmonic and vocal acoustics relating to pauses and vocalization in speech were significantly correlated with the Interactive Voice Response (IVR) HAMD scores when measured using Pearson's correlation. Also reported, total pause time during automatic speech where patients were asked to read from a standardized passage, counting from 1 to 20 and pronouncing vowels for 5s reveal stronger correlation, whereas vocalization-pause-ratio reveals better correlation during free speech. In the second study, logistic regression analyses were conducted on acoustics measures of speech in order to classify between 105 patients who responded to treatment and those who did not. Among the seven acoustic measures that were found to be significantly correlated with depression severity measures in the previous study, six of them were also found to be significant in this study. Results across both studies were consistent thus provide strong evidence for the value of vocal acoustic features as an indicator of depression severity.

The study performed by Trevino [29] builds up upon the initial study by Mundt [27]. Besides looking at global features in speech, this research also investigated acoustic features within the phonemes of speech signals and their relationship, using Spearman correlation with individual symptom sub-topic ratings of each 17 HAMD score instead of the total score. A large

set of average phone lengths, a linear combination of phone-specific length measures and the energy of a phone were shown to be significantly correlated with the HAMD Psychomotor Retardation sub-topic.

This paper attempts to address the question of how well do speech features, specifically the timing based measures predict the ratings from well-known medical diagnostic tools known as the Hamilton Depression Scale (HAMD) and Beck Depression Inventory (BDI-II). Previous research mainly investigates the correlation between the clinical ratings and speech measurements. Correlation merely describes the relationship between two variables whereas regression predicts the value of a dependent variable (i.e., clinical scores) using one or more measurements of independent variables (i.e., acoustic features). In [39], they used naïve listeners with no experience in making clinical judgment to predict the participant's and interviewer's HAMD ratings. This method reported moderate predictability of the HAMD ratings.

Another objective of this study is to analyze the regression between the speech features and the patient's clinical score using two speech databases that were collected using different recording devices and environment. Finally, the analysis will also compare the performance of the prediction when using an interview speech as opposed to a reading speech and whether the characteristic of the prediction varies in terms of gender-wise.

3.0 Database

3.1 Assessment Procedures

All recording sessions were conducted at the Vanderbilt Emergency Department or Psychiatric Hospital with patient's documented informed consent. Patients who volunteered were made aware of the aim of the study with assurance of maximum identity protection procedures. Patients under the influence of alcohol, toxicity or experiencing respiratory problems such as shortness of breath were excluded. All recordings were made in a standard, empty psychiatric interview room without the benefit of soundproof or acoustically ideal environment, mimicking the real-world clinical environment. Group assignment was made according to assessment made by experienced clinicians using the Beck Depression Inventory (BDI-II), MINI International Neuropsychiatric Interview, Pierce Suicide Intent Scale (SIS) and Hamilton Depression Rating Scale (HAMD) [38].

Two types of databases were used for this study. All speech samples were digitized using a 32-bit analog to digital converter at 44.1 kHz sampling rate for both databases. In the first database (Database A), recordings were collected once per patient and each recording was categorized as either high risk suicidal or depressed. Audio acquisitions were made using a high-quality Audix SCX-one cardioid microphone with a frequency response of 40Hz to 20kHz, Sony VAIO laptop with Pentium IV 2GHz CPU 512 Mb memory, Windows XP, a Digital Audio MBox for digital audio interface and recording software PROTools LE for the digital audio editor.

In the second database (Database B), a number of patients attended three recording sessions, some had two recordings and the rest only had one recording. The second and third recordings were collected a few days after receiving treatments. Group assignment was determined for each recording according to clinical information in written form by an experienced clinician while blind to the recording sounds. The rating scale content was supplanted, when necessary, by the interviewer's notes. The interviewer conducted all rating scales with each recorded interview. Strong inter-rater reliability was maintained for the first set of recordings by regular training sessions. The second set ratings were all performed by one interviewer who was supervised closely by the experienced clinician. Entry criteria restricted inclusion to patients who were labeled as high risk suicidal during their first recording session. The state of the patient during the next recording sessions were made blind to the acoustic engineering researchers and categorized as *others*. In this case, *other* may or may not indicate that the patient is no longer considered high risk. Audio acquisitions for the second database (B) were made using a portable high-quality field recorder, a TASCAM DR-1, with a frequency response of 40Hz to 20kHz, Samsung Q40 laptop with Intel Core i5 2.4GHz 4G memory and Windows 7.

Two types of speech samples that were collected from the male and the female patients are called the interview speech and the reading speech. For the interview speech, patients were engaged in an interview with a clinician answering a series of questions such as feeling of guilt, thoughts of suicide, interest level, presence of anxiety and somatic complaints. For the reading speech, patients were asked to read from a standardized "rainbow passage" which contains every sound in the English language and is considered to be phonetically balanced with the ratios of assorted phonemes similar to the ones in normal speech [30].

In this study, the regression analysis on Database A with their HAMD scores was conducted using the interview and reading speech recordings from the male and female patients. On the other hand, the reading speech for male and female patients who were diagnosed with high risk suicidal (HR), major depression (DP), suicidal ideation (ID) and remission (RM) from Database B were used for the regression analysis with their HAMD and BDI-II scores. Only the reading speech was used for the latter analysis due to an inadequate number of acoustic measures that could be extracted from the interview speech.

Table 4.1 displays the information on the number of patients and recordings in Databases A and B that were used in the regression analysis. In Database A, there are a number of nine and 35 male reading patients and 14 and 58 female reading patients with the HAMD scores and BDI-II scores, respectively. In Database B, HAMD scores were associated with eight male interview patients, seven male reading patients, 14 female interview patients and 13 female reading patients.

Table 4.1: The number of patients and recordings in Database A and Database B that were used in the regression analysis

Database A	**Male Reading**		**Female Reading**	
Total number of patients with HAMD scores	9		14	
• Total number of 20 second segments	21		33	
Total number of patients with BDI-II	35		58	
• Total number of 20 second segments	69		140	
Database B	**Male Interview**	**Male Reading**	**Female Interview**	**Female Reading**
Total number of patients with HAMD scores	8	8	14	13
Total number of recordings	19	18	34	32
Total number of 20 second segments	473	49	479	78
• Number of patients with three recordings	5	4	8	7
• Number of patients with two recordings	1	2	4	5
• Number of patients with one recording	2	2	2	1
Number of recordings labeled as high risk (HR)	8	7	13	12

Figures 4.1 to 4.4 display the HAMD scores and the BDI-II scores distribution for the male and female patients in Database A.

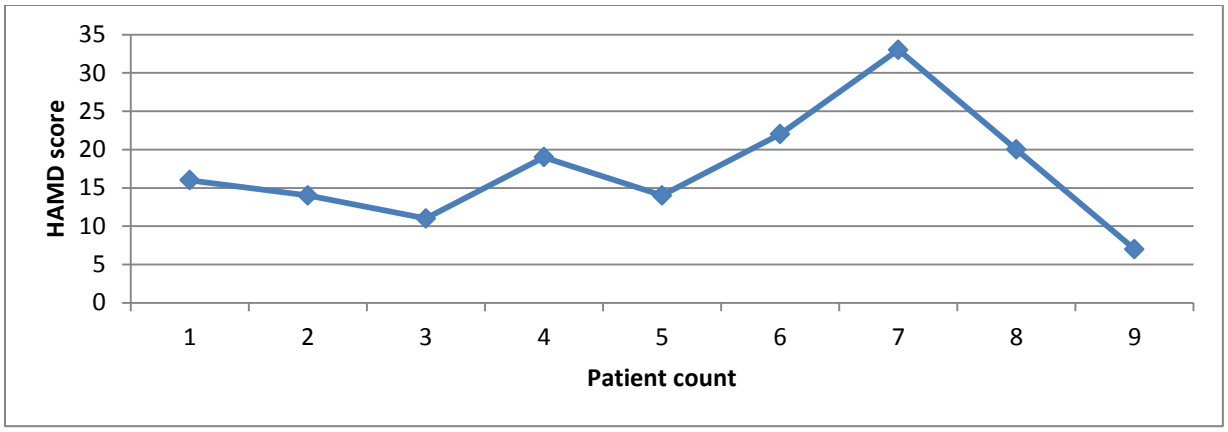


***Figure 4.1**: HAMD scores for male patients from Database A*

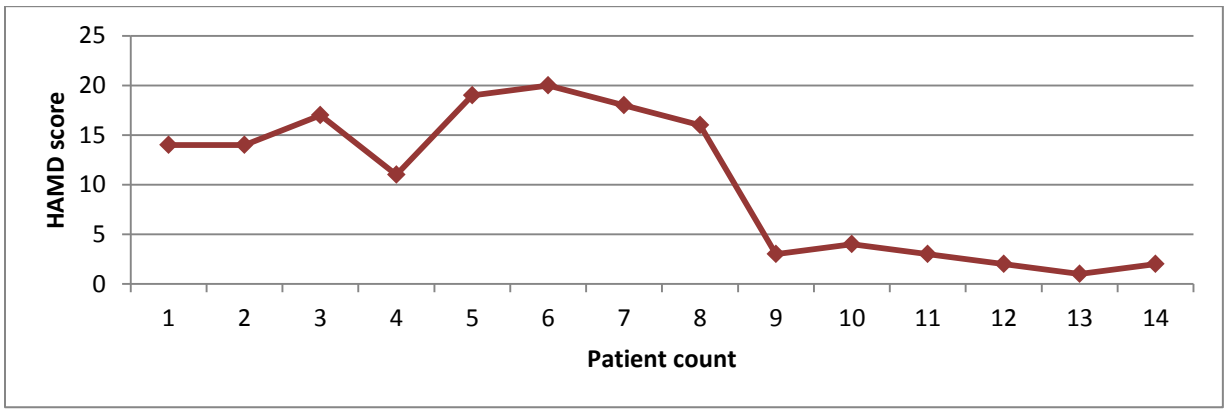


***Figure 4.2**: HAMD scores for female patients from Database A*

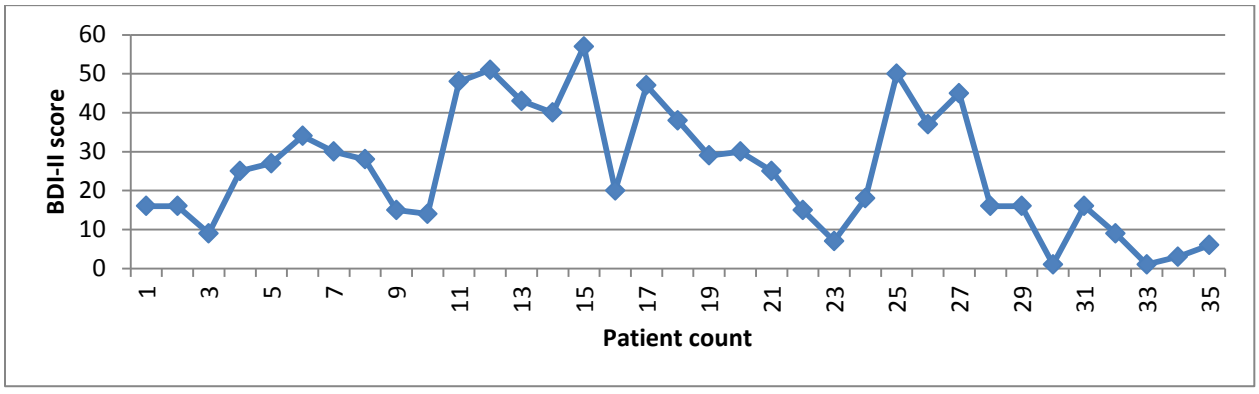


***Figure 4.3**: BDI-II scores for male patients from Database A*

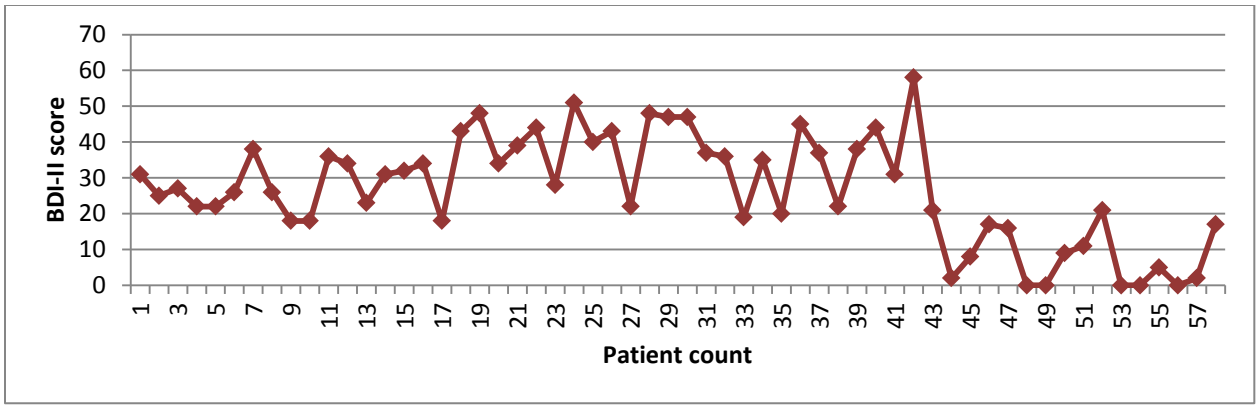


***Figure 4.4**: BDI-II scores for female patients from Database A*

Figures 4.5 and 4.6 display the HAMD scores for the male and female patients from Database B according to their recording sessions. Each patient will have a certain number of 20 second segments depending on the length of the recording, thus each 20 second segment will be represented by the same HAMD score.

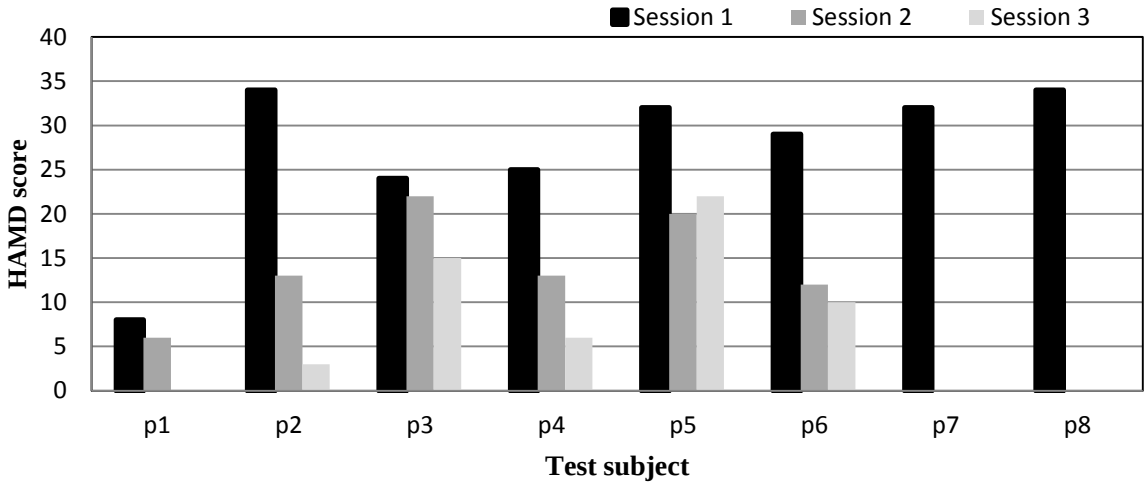


***Figure 4.5**: HAMD scores for male patients from Database B*

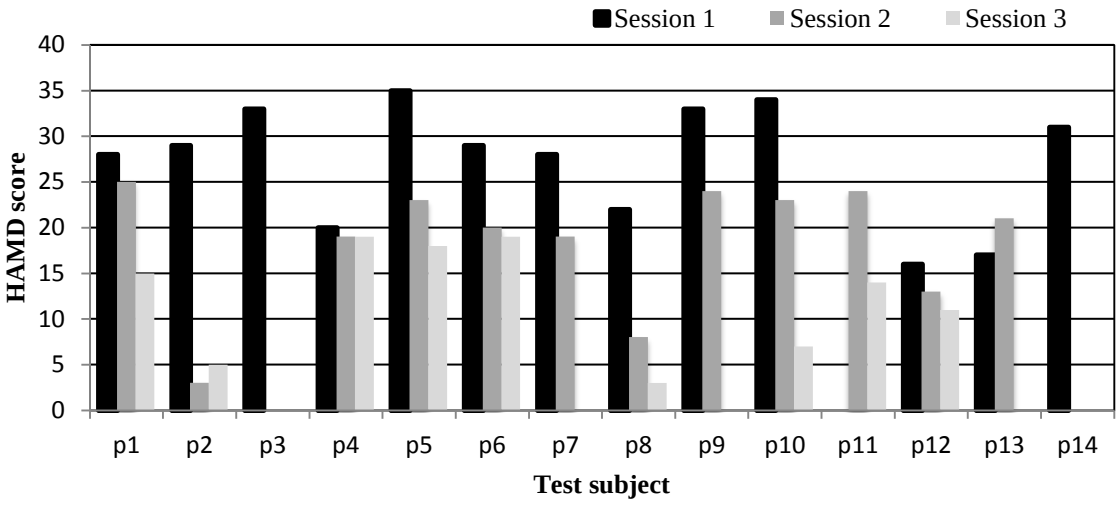


***Figure 4.6**: HAMD scores for female patients from Database B*

3.2 Acoustic Procedures

In the preprocessing stage, recordings were edited using a free audio digital editor called Audacity 2.0.1 to remove any identifying information, to preserve patient privacy. Undesirable sounds such as the interviewer's voice, voices other than the patient, sneezing, coughing and door slams were removed from the de-identified recordings. Features are then collected according to methods explained in [13] - [15].

For the Mel-Frequency Cepstral Coefficients, long pauses that are present for more than 0.5 second were also removed from speech samples. The sampled signals were then divided into window frames of 40ms. A voiced/unvoiced/silence decision was made for each frame based on the method in [10]. Unvoiced and silence terms were removed and the voiced terms were concatenated into one new speech signal for further analysis. The voiced speech sample was then divided into 20 second segments and 13-MFCCs were calculated for each 40ms frame using Slaney's Matlab Auditory Toolbox [31]. For one segment of 20 second voiced sample recording, the mean of 13-MFCC was obtained. Therefore, each patient will have a certain set of the mean of 13-MFCC depending on the quantity of 20 second segments it has

4.0 Methodology

4.1 Voice Acoustic Features

Voice acoustic features were extracted from the speech samples of interview and reading speech tasks. So far there is no common agreement on which feature contains the most distinguishable information for the identification of psychological state. Common approaches include a high number of features and then applying a feature selection algorithm for dimensionality reduction. Considering the small size of the dataset, it is beneficial to include relevant features that are expected to be good predictors according to results from previous studies of feature classification. Redundant, correlated or irrelevant features may negatively influence the performance of the predictor. The initial set consists of the following 67 acoustic features:

(1) Seven equal bands of Power Spectral Density (PSD) from 0 to 1750 Hz [13,14].
(2) 13 Mel-Frequency Cepstral Coefficients (mfcc) [31]
(3) Transition parameters [15]; voiced-to-voiced(t_{11}), voiced-to-silence(t_{13}), unvoiced-to-voiced(t_{21}), unvoiced-to-unvoiced(t_{22}), silence-to-voiced(t_{31}), silence-to-silence(t_{33})
(4) Interval pdfs [15]; 25 voiced bins (voi), 6 unvoiced bins (unv), 10 silence bins (sil)

Features in items (1) and (2) are related to spectrum based measures while features in items (3) and (4) are associated with time timing based measures.

4.2 Method of Resampling

To validate the performance of the predictive regression model, a method of resampling known as jackknife (or leave-one-out) was applied to the data set. This resampling method is appropriate to be used with a small to moderately sized data set and the observations are independent of each other. Also, as the number of unknown variables begins to approach the dimension of the output space, the issue of over-modeling arises. Thus, the prediction model may only be able to predict within the population due to a large number of features but fail to predict a new population. The jackknife method in a sense creates *new* observation each time by removing one patient and estimating the model coefficients using the remainder population (training data set). The model coefficients will then be used to predict the score of the left-out

patient (test sample). This process is repeated by excluding the next patient from the data set until all patients have been chosen as left-out.

4.3 Multiple Linear Regressions Model

The method of multiple linear regressions using least squares was applied to examine the relationship and to obtain the model coefficients of the features (independent variables) to the clinical HAMD score (dependent variable). The general regression model equation for $\boldsymbol{P}$ number of features is $\underline{\boldsymbol{h}} = [\boldsymbol{D}_{1,2,..,P}\ 1]\ \boldsymbol{x}\ \underline{\boldsymbol{a}}$, where $\underline{\boldsymbol{h}}$ is a column vector of the actual clinical scores associated with the training data set that matches the number of rows in the matrix $\boldsymbol{D}$. Matrix $\boldsymbol{D}$ consists of the input vectors where the number of rows represents observations and each column represents the independent variables. The dimension of columns for $\boldsymbol{D}$ is represented by: $\boldsymbol{D}$ = [PSD, Transition parameters, Interval pdfs, MFCC]. $\underline{\boldsymbol{a}}$ is a matrix that consists of a column vector of the model coefficients resulting from the multiple regression process and having the same size as matrix $\underline{\boldsymbol{h}}$. The error minimization was performed in two steps:

Step 1: The least square solutions minimize the sum of the squared error for each prediction which is the differences between each test sample's actual clinical score and the predicted score using the estimated model coefficients from the training data set.

Step 2: For $\boldsymbol{N}$ number of patients, mean sum of absolute error is calculated by summing the absolute values of the error and then dividing the total error by $\boldsymbol{N}$ which can be written as, $\boldsymbol{MAE} = \frac{1}{N}\sum_{i=1}^{N}|e_i|$ and $\boldsymbol{e_i} = \underline{\boldsymbol{h_i}} - [\boldsymbol{D}_i\ 1]\ \underline{\boldsymbol{a_{-i}}}$. Then, using methods of feature selection, combination of features that will give the minimum sum of absolute error were calculated.

4.4 Feature Selection

In order to optimize the jackknife method, the next step is to identify what combination of features that would provide the best predictability for the left-out patient. Essentially, feature selection will try to find a set of features that are generalizable from the rest of the populations. By having smaller number of features, besides improving the generalization capability, it will

also reduce complexity and run-time. Also, one particular problem with the linear regression is when the number of features exceeds the number of observations, the least square solution will not be unique. Too many features may lead to a bad prediction due to the method finding a way to fit itself not just to the underlying structure but also to the irrelevant information in the training data set as well. One way to solve this is by finding out what features that are relevant and eliminating features that contribute less to the prediction without involving a transformation. This approach was carried out analytically by applying the two most common sequential search algorithms which are called the Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS).

4.4.1 Sequential Forward Selection (SFS)

SFS starts with an empty set $\boldsymbol{Z}_0 = \{0\}$ and let $\mathbf{X} = [\boldsymbol{x}_1\ \boldsymbol{x}_2\ \ldots\ \boldsymbol{x}_f]$ be the feature matrix. Form a linear regression estimator using exactly one feature and continue to evaluate until each feature has been chosen. Select one feature that produces the minimum mean sum of absolute error and add it to the empty set. Next, add each of the remaining features $\mathbf{X} = [\boldsymbol{x}_1\ \boldsymbol{x}_{2\ \ldots}\ \boldsymbol{x}_{f\text{-}1}]$ one at a time to the new subset and evaluate the two features combination that yields the best performance. Repeat the process until all features are chosen, $\mathbf{X} = \{0\}$.

4.4.2 Sequential Backward Selection (SBS)

SBS is a process of sequential discarding bad features. SBS starts with all features $\boldsymbol{Z}_0 = \boldsymbol{X}$. Repeatedly, remove one feature at a time and form a linear regression estimator using the remaining features. Discard the one feature that when remove from the set, yields the minimum mean sum of absolute error. The process is repeated until there is only one feature left in the set.

4.5 Measuring the Fit of the Regression Model

To assess the quality and significance of the multiple regression models, a table Analysis of Variance (ANOVA) that contains information concerning variances that exist within a regression model is constructed. It is a general statistical technique that analyzes the dissimilarities or resemblances between two or more groups of data. The table consists of sum of squares, degrees of freedom and F statistics [33].

Sum of Squared

In this study, we are interested in the Sum of Squared Error (SSE), Sum of Squared Regression (SSR) and Sum of Squared Total (SST). For $\boldsymbol{M}$ number of 20 second segments and $\boldsymbol{P}$ number of features, the variation that exists in the dependent variable $\underline{\boldsymbol{h}}$ can be expressed and written using sum of squared as:

$$\boldsymbol{SST = SSR + SSE}$$

SST is the total variance in the collected sample that measures the distances from the actual HAMD scores, $\underline{\boldsymbol{h}}$ to its mean, $\overline{\underline{\boldsymbol{h}}}$ and having $\boldsymbol{M}$-1 degrees of freedom. SSR is the total variance in the model prediction sample that measures the distances of the predicted clinical scores $\underline{\widehat{\boldsymbol{h}}}$ from the mean $\overline{\underline{\boldsymbol{h}}}$ with $\boldsymbol{P}$ degrees of freedom. SSE is the total squared error that measures the model predicted scores from the actual scores with $\boldsymbol{M}$-$\boldsymbol{P}$-1 degrees of freedom. If all the prediction scores matches exactly the actual scores, SST = SSR.

$$\boldsymbol{SST} = \sum_{j=1}^{M} (\underline{\boldsymbol{h}}_j - \overline{\underline{\boldsymbol{h}}}_j)^2 , \qquad \boldsymbol{SSR} = \sum_{j=1}^{M} (\underline{\widehat{\boldsymbol{h}}}_j - \overline{\underline{\boldsymbol{h}}}_j)^2 , \qquad \boldsymbol{SSE} = \sum_{j=1}^{M} (\underline{\widehat{\boldsymbol{h}}}_j - \underline{\boldsymbol{h}}_j)^2$$

R^2 and Adjusted R^2

R-squared or R^2 is the measure of how well the total amount of variation in the dependent variable is explained by the independent variables in the estimated equation. The value of R^2 is bounded by $0 \leq R^2 \leq 1$ where $R^2 = 1$ represents a perfect prediction and $R^2 = 0$ indicating the independent variables are not suitable for prediction.

$$R^2 = 1 - \frac{SSE}{SST}$$

F-Statistics

The F-test is the overall or joint significance that indicates whether the relationship between the independent variables and the set of predictor's coefficients are statistically significant. It evaluates the following hypothesis test,

$$H_0\colon\ \alpha_1 = \alpha_2 = \ \ldots\ \ = \alpha_P = 0$$
$$H_1 = at\ least\ one\ or\ more\ \alpha\ \neq 0$$

The F-test is an upper-tailed test that is calculated as $F = \frac{SSR/_P}{SSE/_{(M-P-1)}}$ where F is distributed as a random variable. Under the null, $F \sim F(P, M-P-1)$ and the critical value is based on P numerator degrees of freedom, df_n and $(M-P-1)$ denominator degrees of freedom, df_d. A larger F-statistic signifies a stronger relationship between the actual clinical scores and the features used for the model's prediction. The F-statistic does not point out which particular features are significant only that at least one of the features is. Thus, rejection of H_0 or $F > F_\alpha(df_n, df_d)$ indicates at least one of the features in the set has a partial effect in the clinical score variation. If H_0 is accepted or $F \leq F_\alpha(df_n, df_d)$, this implies that all features in the set are insignificant in explaining the variations. It can also be shown using p-values where the null hypothesis is rejected for significance p-value that is larger than α = 0.05. For a p-value that is less than α = 0.05 but larger than α = 0.01, it is considered to be probably significantly different. And for p-value that is less than α = 0.01, the two populations are considered to be extremely significantly different.

5.0 Results

5.1 Regression Analysis on Speech Features and HAMD using Database B

5.1.1 Statistical Analysis

Table 4.2 displays the estimated number of features, Mean Sum of Absolute Error (MAE), Standard Deviation Sum of Absolute Error (SDAE) and Median Sum of Absolute Error (MdAE), maximum error (maxE) and percentage of absolute error above MAE (%>MAE) using methods of SFS and SBS for predicting the patient's clinical scores according to groups of male and female interview and reading speech in Database B. Plus or minus sign on maximum error signify direction of error. A negative error indicates that the clinical score prediction is less than the actual score and vice versa. The numbers α(β,γ) in row labeled %>MAE represents percentage of the absolute errors that are above the MAE (α), percentage of errors that are above +MAE (β), and percentage of errors that are below –MAE (γ).

Table 4.2 Statistical comparison on the application of the forward (SFS) and backward (SBS) feature selection procedure using the interview and reading speech from the male and female patients in Database B for predicting the HAMD scores

		Male Interview	Male Reading	Female Interview	Female Reading
SFS	# features	29	11	11	10
	MAE	1.5240	2.0657	3.5750	5.0121
	SDAE	2.2586	3.1741	5.1323	6.9470
	MdAE	-0.0064	0.1971	0.0376	0.0978
	MaxE	-12.0652	-9.8811	-18.8601	-15.4992
	% >MAE	37 (18,19)	39 (12, 27)	32 (11, 20)	49 (20, 29)
SBS	# features	19	17	33	25
	MAE	1.2127	1.0580	2.1106	1.4017
	SDAE	1.6756	1.2868	3.2200	1.8073
	MdAE	-0.0295	-0.3814	0.1653	-0.0986
	MaxE	-4.0328	-3.2377	-12.4477	-5.0437
	% > MAE	41 (22, 19)	39 (27, 12)	39 (18, 21)	45 (16, 28)

With the exception of the male reading group, the total number of 20 second segments that consist of 473 vectors for male interview, 479 vectors for female interview and 78 vectors for female reading exceed the number of features in the initial set (67 features). In terms of the mathematical justification, having more equations than unknowns (overdetermined system) generally will not produce a unique solution because it may be impossible to satisfy all equation simultaneously. Instead, the best approximate solution is based on the solution that minimizes the residuals. Having large dimension of data points and using it to obtain a linear combination of 67 features or less is a nice solution because this shows that the vector lies in the range space and there is information that strongly relates to the prediction. Therefore, considering the number of observations to be the total number of 20 second segments is reasonable. Also, the total number of features selected using methods of SFS and SBS are less than the total number of recordings except for the SFS method on the male interview speech.

It can be seen from table 4.2 that the method of SFS outperformed SBS in terms of the total number of features that yielded minimum MAE except for the male interview. On the other hand, SBS outperformed SFS with regards to its ability to select a group of features that produces smaller value of MAE, SDAE and maxE estimation. The MAE value directly expresses the measure of how close the prediction scores are to the actual HAMD scores without

considering the direction of error. If for example a patient has an actual HAMD score of 23, an error of minus four could misplace the patient in the lower level category of severe depression considering 23 is the threshold score for the high risk suicidal. By observation, an error of approximately three or less is considered to be decent. The MaxE produced by the suboptimal SBS features are smaller than the MaxE produced by the SFS feature combination. Besides the female interview, prediction errors by the SBS features are kept within a smaller range of error.

For the male interview speech, prediction of the HAMD scores using a feature combination obtained by the SBS method is more desirable than the SFS method because the total number of features by the SFS method exceed the total number of recordings. Also, the SBS method exhibit a smaller MAE value compared with the SFS.

Results for the male reading speech using both methods are acceptable. There is however a small trade-off between the number of features and the MAE. The MAE of 2.0657 by the SFS method is not considered to be substantial. The combination of the 17 feature obtained by the SBS method is still less than the number of 20 second segments and the number of recordings.

For the female interview, both methods produced a suboptimal number of features that is less than the number of observations. Considering SDAE, errors produced by methods of SFS are more dispersed and had a higher maximum compared to SBS. Also, since the percentages of absolute errors above MAE are almost similar, choosing the one with a lower MAE would be more desirable.

The number of features generated by the SFS procedure is marginally smaller compared to the SBS procedure but both numbers of features are still less than the number of observations. However, MAE of 5.0121 is considered too significant. Therefore, predicting the HAMD scores using 25 features and MAE of 1.4017 is preferable with the female reading speech.

Even though the method of SFS achieved the highest level of dimensionality reduction by selecting the least number of features, generalized features obtained by the method of SBS outweighs the preference due to its ability to gather features that yield the least MAE.

The complete plots of MAE with respect to the number of features obtained by the SFS and the SBS procedure for all four subject categories are also illustrated in figures 4.7 to 4.10.

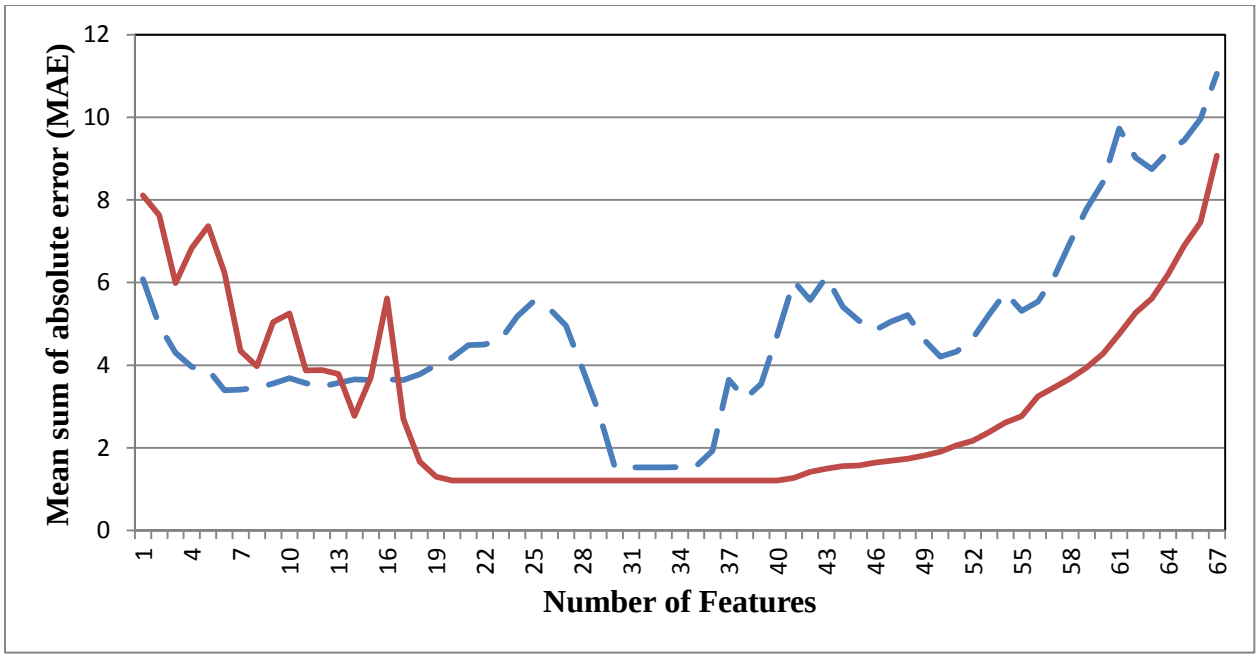


***Figure 4.7**: Characteristic plot of the SFS (blue line '--') and the SBS (red line) methods using the male interview speech from Database B to predict the HAMD scores*

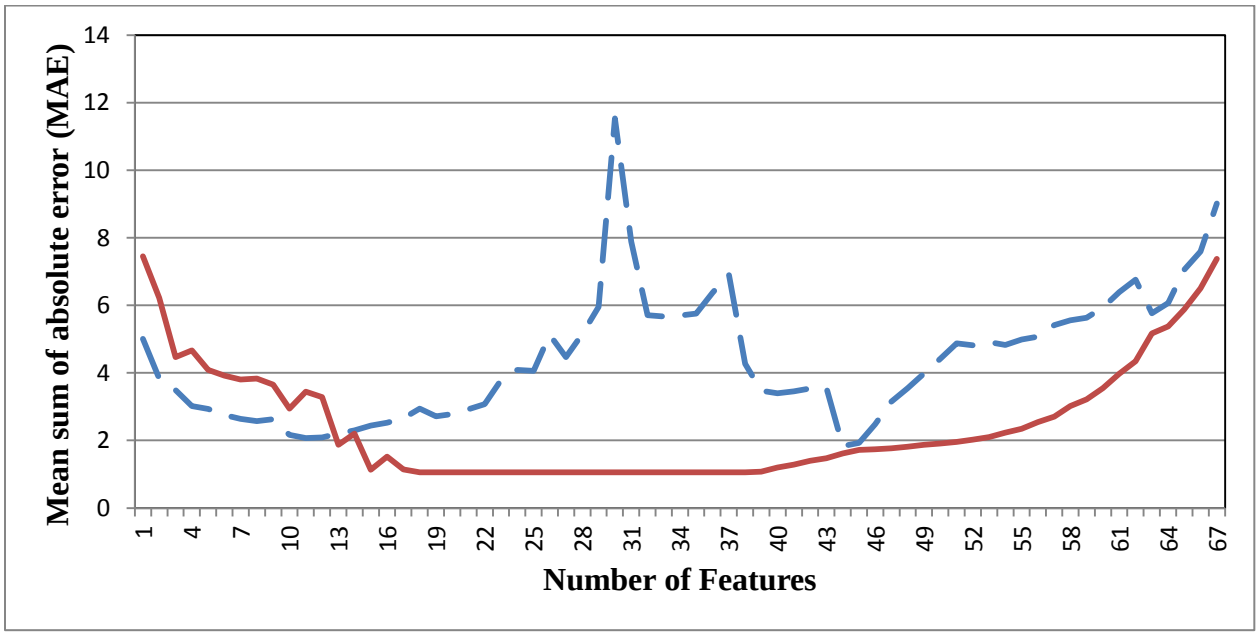


***Figure 4.8**: Characteristic plot of the SFS (blue line '--') and the SBS (red line) methods using the male reading speech from Database B to predict the HAMD scores*

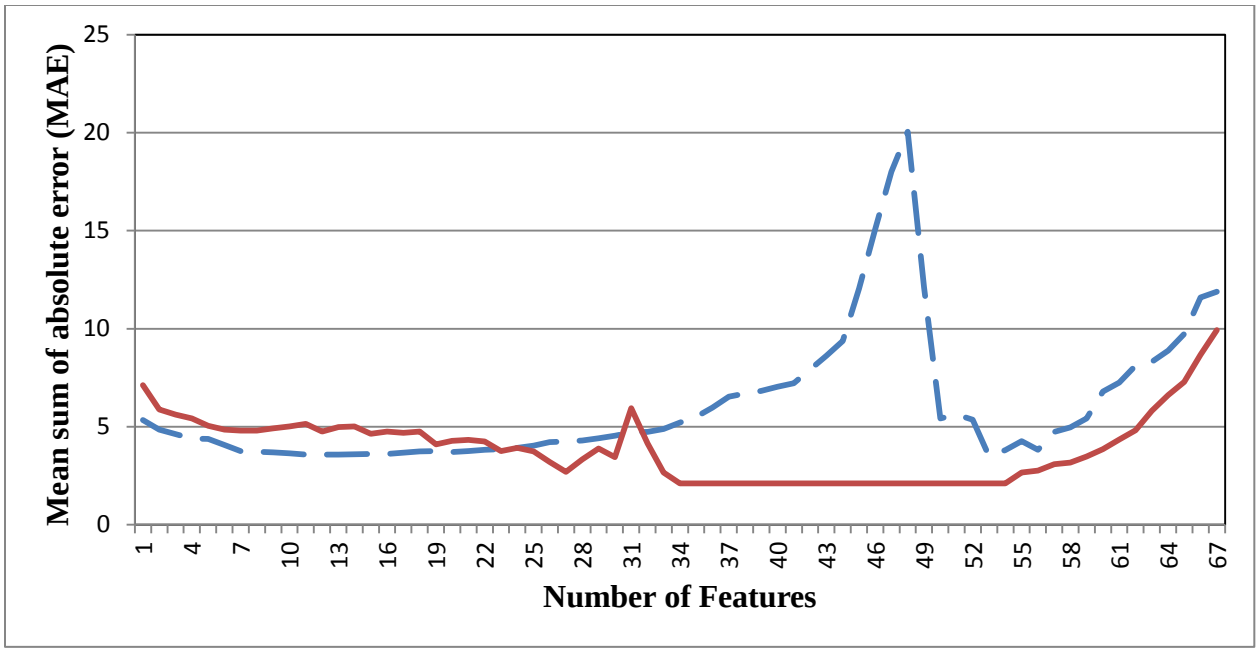


***Figure 4.9**: Characteristic plot of the SFS (blue line '--') and the SBS (red line) methods using the female interview speech from Database B to predict the HAMD scores*

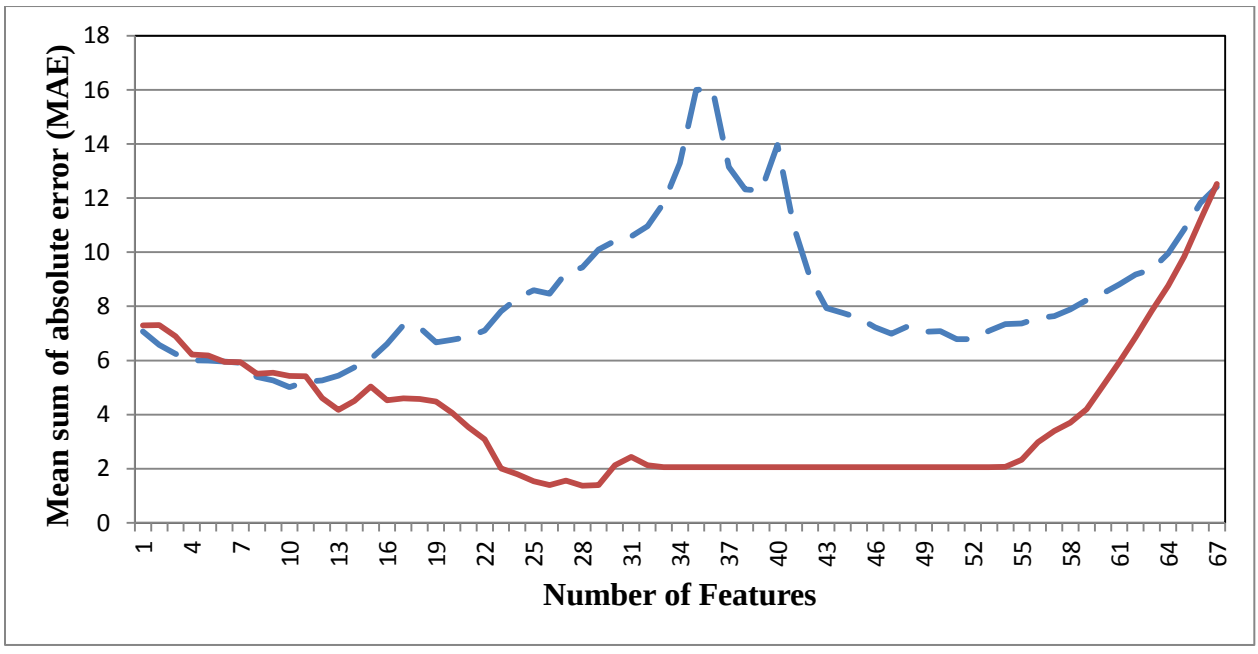


***Figure 4.10**: Characteristic plot of the SFS (blue line '--') and the SBS (red line) methods using the female reading speech from Database B to predict the HAMD scores*

Apart from figure 4.7, we observed that the graphs of SFS initially reduces the value of MAE at a slow rate as features were added one at a time until it reaches a local minimum. Each added features contains relevant information that increases the accuracy of the prediction. The graph then exhibit a steady increase until it reaches a global maximum which shows that the added features greatly degraded the prediction due to the features that contain high degree of irrelevant and redundant information. But after reaching a peak, the graphs again decline to another local minimum. Even though the previous irrelevant features were not removed, the prediction still improved probably because of the new added information that counters the irrelevant features.

All plots in figures 4.7 to 4.10 demonstrated a similar trend for the backward (SBS) feature selection. The graph of SBS is read from right to left because the initial set contains 67 features and the feature that was identified to produce the minimum MAE when removed from the set was then discarded one by one. The graph exhibits a gradual decrease of the MAE value as each irrelevant feature was discarded from the set. What is interesting is that, after removing a certain number of features, the MAE remained the same for the next 20 discarded features. These 20 features were identified to be the spectrum-based measure that consists of the 13-MFCC and the seven equal bands of PSD. This behavior was observed throughout the regression analysis when the SBS method was implemented. The minimum MAE was achieved after removing all spectrum-based measures for the male interview and reading speech and the female interview speech. However, for the female reading speech, the minimum MAE was obtained after removing a few more timing-based measures.

Figures 4.11 to 4.14 display comparison plots of the actual HAMD scores and the predicted HAMD scores obtained using the selected SFS feature combination for the male and female patients with their interview and reading speech.

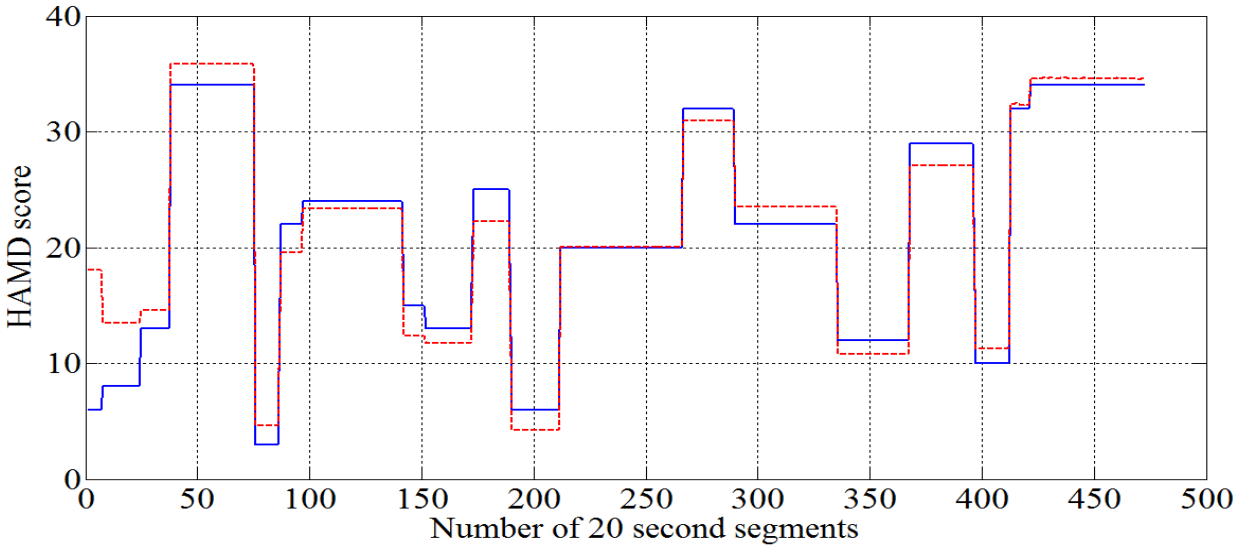


***Figure 4.11**: The actual (blue '—') and the predicted (red '--') HAMD scores for male interview patients in Database B using the SFS procedure*

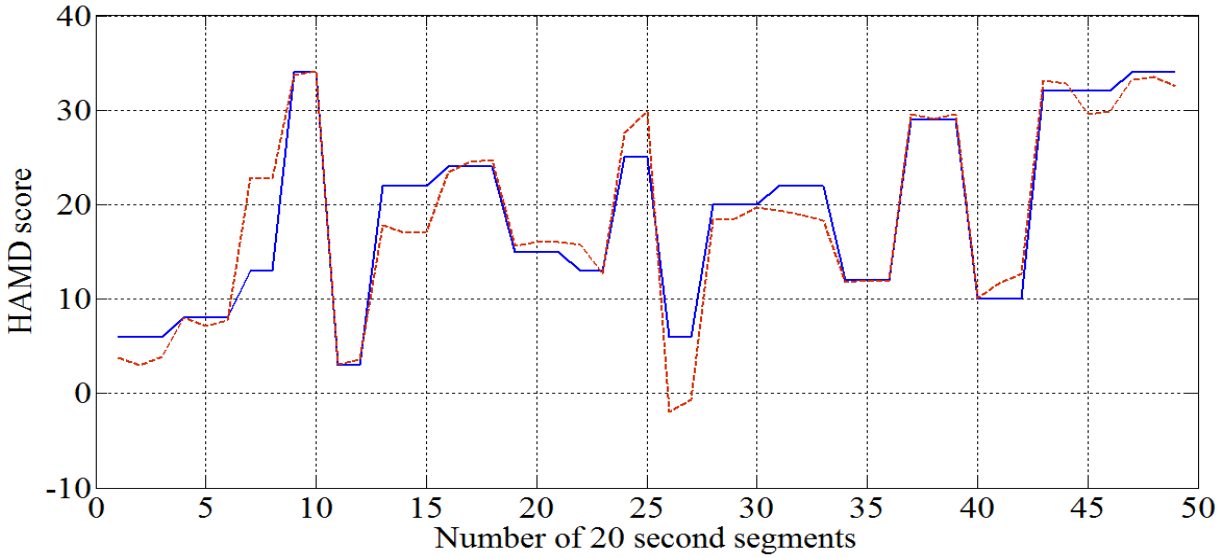


***Figure 4.12**: The actual (blue '—') and the predicted (red '--') HAMD scores for male reading patients in Database B using the SFS procedure*

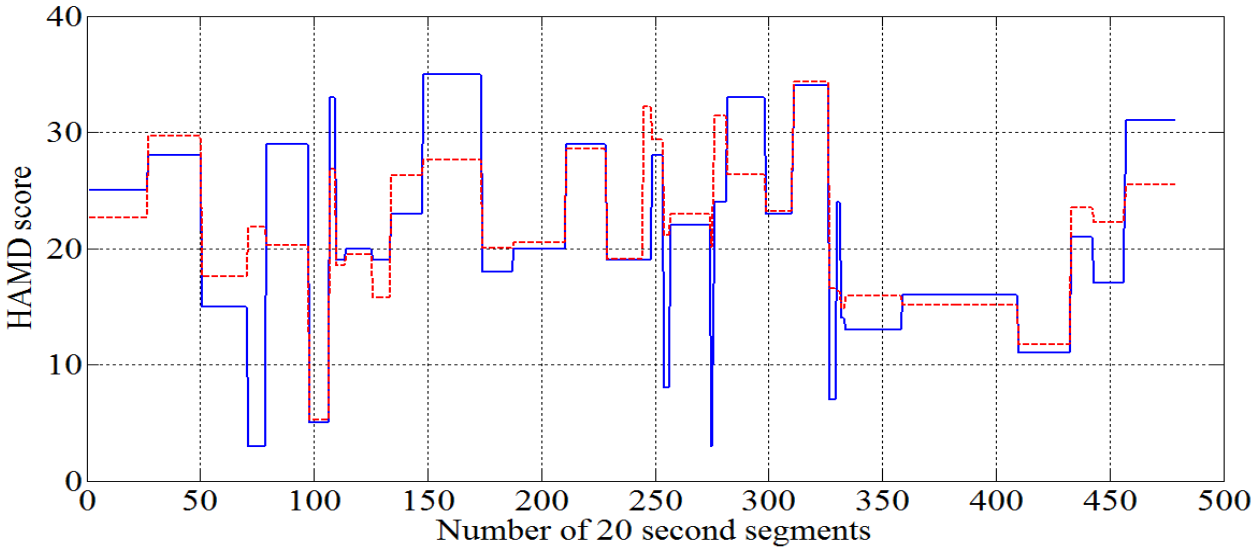


***Figure 4.13**: The actual (blue '—') and the predicted (red '--') HAMD scores for female interview patients in Database B using the SFS procedure*

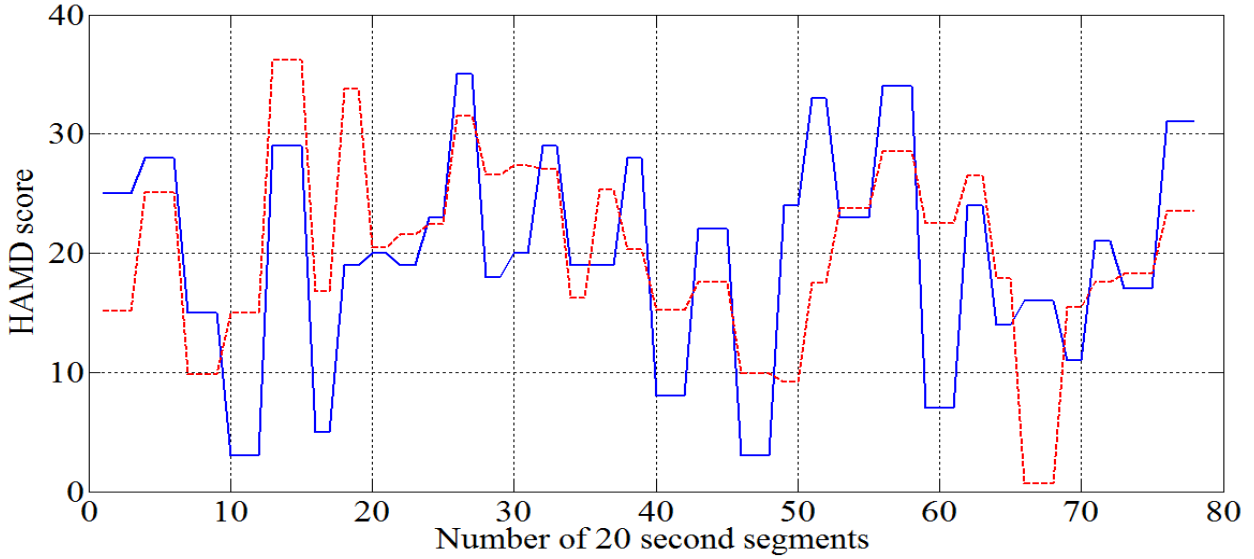


***Figure 4.14**: The actual (blue '—') and the predicted (red '--') HAMD scores for female reading patients in Database B using the SFS procedure*

Figures 4.15 to 4.18 display comparison plots of the actual HAMD scores and the predicted HAMD scores obtained using the selected SBS feature combination for the male and female patients with their interview and reading speech.

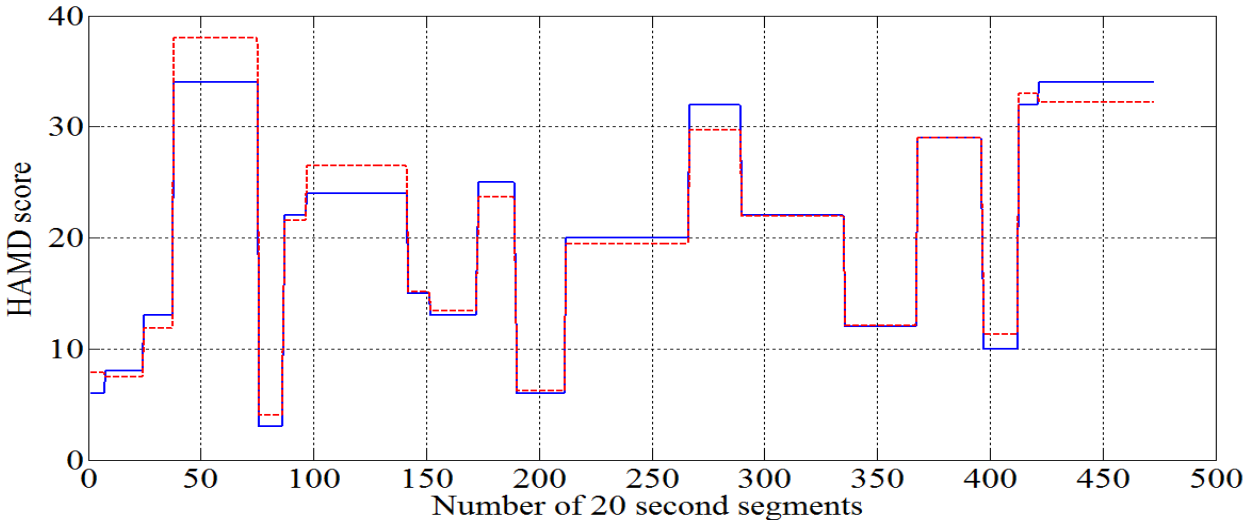


***Figure 4.15**: The actual (blue '—') and the predicted (red '--') HAMD scores for male interview patients in Database B using the SBS procedure*

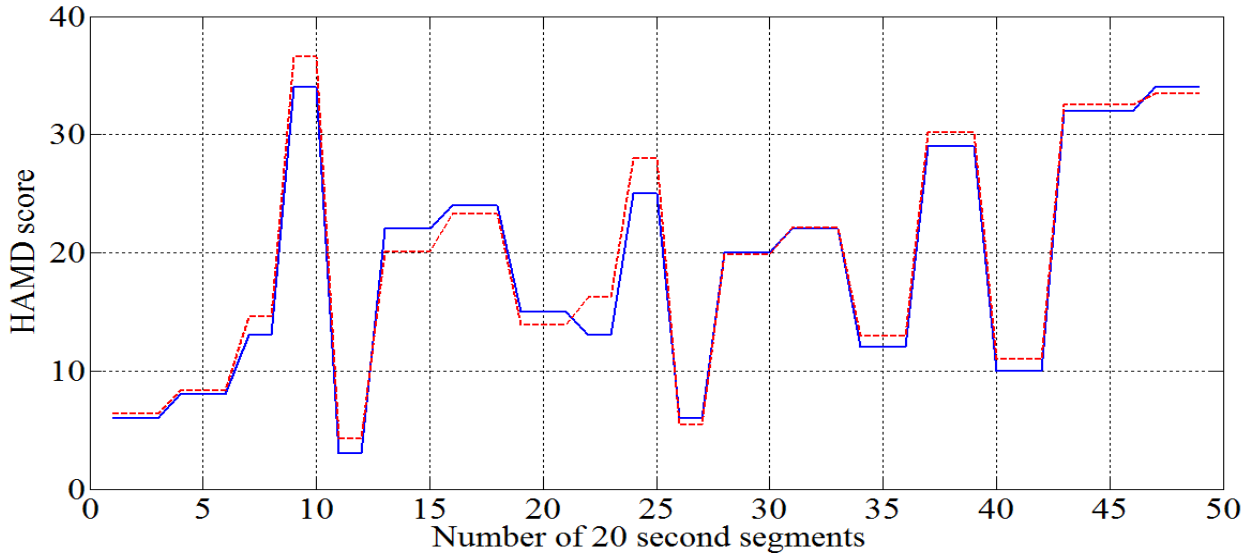


***Figure 4.16**: The actual (blue '—') and the predicted (red '--') HAMD scores for male reading patients in Database B using the SBS procedure*

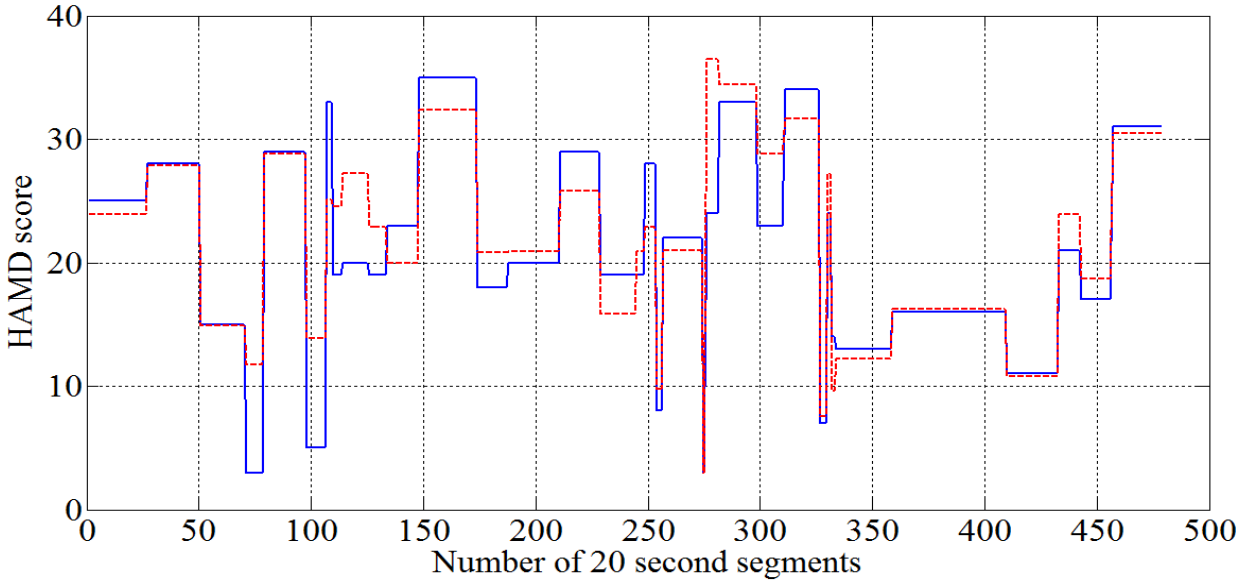


***Figure 4.17**: The actual (blue '—') and the predicted (red '--') HAMD scores for female interview patients in Database B using the SBS procedure*

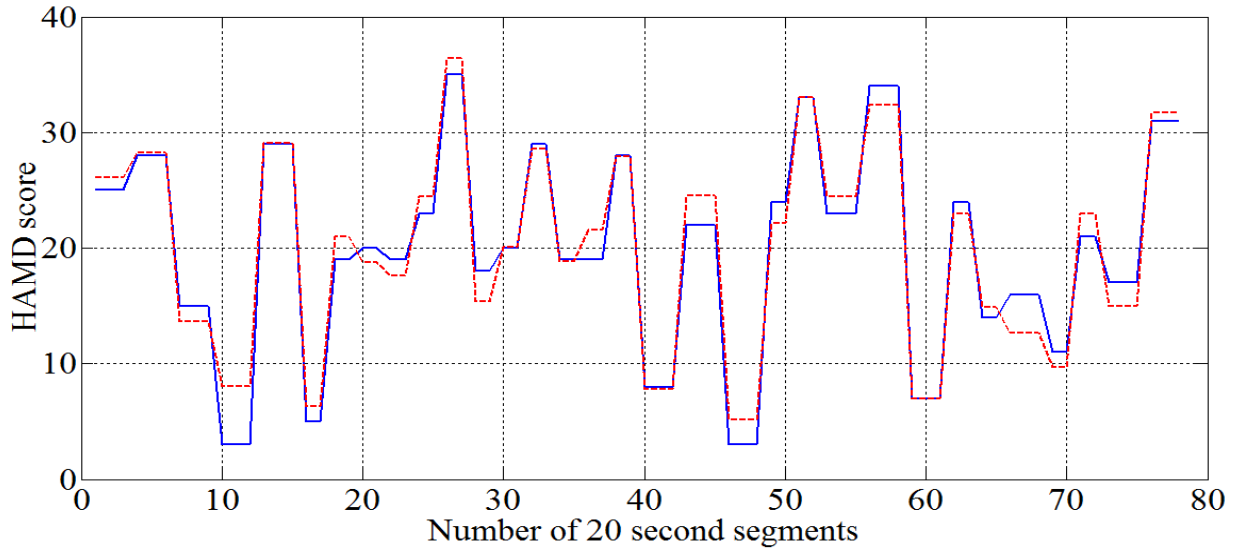


***Figure 4.18**: The actual (blue '—') and the predicted (red '--') HAMD scores for female reading patients in Database B using the SBS procedure*

By observation, there are three noteworthy cases of the prediction type error for HAMD score:

Type I:

Consider the threshold for the HAMD score of a potential high risk suicidal patient to be above 23. If the prediction score occurred on either one side of the actual score and both are above the threshold, the errors are considered insignificant and thus, the decision would still be interpreted as potential high risk suicidal and does not implicate any major misprediction.

Type II:

If the prediction score is above the threshold and higher than the actual score, the prediction error may still be acceptable due to lower risk of fatal outcome. If the actual score is above the threshold and higher than the predicted score, the patient's condition might be at risk of misinterpretation depending on the significance level of error differences. However, in both cases, the results may imply the necessity of performing a second assessment to determine the need for hospitalization.

Type III:

If the prediction score occurred on either one side of the actual score and both are below the threshold, results may still indicate unnecessary hospitalization. However, the significant error may suggest performing a second assessment for the benefit of reassurance.

Table 4.3 Percentage of patients with an error prediction of the HAMD score of less than one, two or three for the male and female interview and reading speech by methods of SFS and SBS

	SFS			SBS		
Percent Error (% Er)	< 1	< 2	< 3	< 1	< 2	< 3
Male Interview	34.04	87.10	94.93	51.16	77.59	91.97
Male Reading	49.64	61.22	79.59	55.10	87.76	95.92
Female Interview	25.26	45.09	58.25	48.02	61.38	75.16
Female Reading	8.97	15.38	26.92	34.62	75.64	92.31

Table 4.3 displays the error percentages of patients with difference between the actual and the predicted HAMD scores less than one, two and three. In this case, we consider an error larger than three to be significant.

For male interview regression using the SFS and the SBS feature combinations, 5.07% and 8.03% of the predictions produced errors of more than three, respectively. Referring to figure 4.11, these errors occurred at the first to the 24th segments thus demonstrating a type III error. Whereas, in figure 4.15, the 38th to the 75th segments demonstrate a type I error. In both cases, the predicted scores were higher than the actual scores.

For the male reading speech regression, prediction using the feature combination obtained by the SBS method did the best compared to the SFS method with a performance of 95.92% and 79.59% error of less than three, respectively. Referring to figure 4.12, 20.41% of errors that are larger than three occurred at the 7th, 8th, 13th to 15th, 32nd and 33rd segments thus signifying a type II error while at the 25th to 27th segments demonstrate a type III error. Meanwhile, referring to figure 4.16, 4.08% of the errors are larger than three occurring at the 22nd and 23rd segments indicate a type I error.

For the female interview speech regression, 41.75% and 24.84% of the segments have errors larger than three for prediction using the SFS and the SBS features, respectively. Most major errors caused by prediction using the SBS features are type I and III except for the segments at approximately within 110th to 150th, which display a type II error.

For the female reading speech regression, the SBS features outperformed the SFS feature combination. The SBS features yielded an error prediction of 7.69% larger than three which consists of the 10th to 12th and the 66th to 68th segments thus demonstrating a type III error.

5.1.2 Goodness of Fit in the Multiple Regression Model using Database B

Table 4.4 displays the analysis of variance between the actual HAMD scores data set and the prediction HAMD scores data set resulting from multiple regression models with the SFS and the SBS method for our four subject groups.

Table 4.4 Analysis of Variance on Multiple Regression Model for Male and Female (Interview and Reading) using SFS and SBS

		Male Interview	Male Reading	Female Interview	Female Reading
	SST	41551.00	4605.30	29964.00	6181.10
SFS	df_n	29	11	11	10
	df_d	443	37	467	67
	SSE	2429.80	489.03	10905.00	5287
	SSR	40853.00	5110.50	18949.00	4630.60
	R^2	0.9415	0.8938	0.6361	0.1448
	F-stats	256.8379	35.1509	73.7709	5.8682
	p-value	0.0001	0.0819	0.6361	0.5741
	$F_{\alpha=0.05}(df_n, df_d)$	1.49	2.06	1.81	1.98
	$F_{\alpha=0.001}(df_n, df_d)$	2.08	3.83	2.91	3.48
SBS	df_n	19	17	33	25
	df_d	453	31	445	52
	SSE	1349.00	90.05	5038.60	255.53
	SSR	42753.00	4666.80	25023.00	5911.50
	R^2	0.9675	0.9804	0.8318	0.9587
	F-stats	755.6127	94.5038	66.9693	48.1192
	p-value	0.0001	0.0001	0.0001	0.0022
	$F_{\alpha=0.05}(df_n, df_d)$	1.61	1.96	1.46	1.72
	$F_{\alpha=0.001}(df_n, df_d)$	2.38	3.59	2.01	2.76

The R^2 value indicates the amount of variation in the actual HAMD scores that is explained by the elected feature combination. According to table 4.4, for the regression models using the SFS method, 94.15% variability that exists in the actual HAMD scores are explained by the variation in the 29 features in the male interview speech, 89.38% are explained by variation in the 11 features in the male reading speech, 63.61% are explained by variation in the 11 features in the female interview speech and 14.48% are explained by variation in the 10 features in the female reading speech. For the regression models using the SBS method, 96.75% variability that exists in the actual HAMD scores are explained by the variation in the 19 features in the male interview speech, 98.04% are explained by variation in the 17 features in the male reading speech, 81.18% are explained by variation in the 33 features in the female interview speech and 95.87% are explained by variation in the 25 features in the female reading speech. The higher the percentage of variation explained by a feature set, the closer the model's predictor hyperplane is to an actual fitted hyperplane.

For the male reading, female reading and female interview speech regression using the features from SFS, p-values indicate that there were no significant differences between the model prediction's variables and the actual HAMD scores and thus failed to reject the null hypothesis at the 5% significance level. Regression analysis using the SBS features from our four groups of speech and using the SFS features from the male interview speech demonstrated p-values that are considered to be extremely statistically significant in the variation and so the null hypothesis is rejected at the 0.01% significance level. These results agree with the MAEs obtained in table 4.2.

5.2 Regression Analysis on speech features and HAMD scores in Database A

Table 4.5 Statistical comparison on the application of the forward (SFS) and backward (SBS) feature selection procedure using the reading speech from the male and female patients in Database A for predicting the HAMD scores

		Male Reading	Female Reading
SFS	# features	4	9
	MAE	1.4282	1.6893
	SDAE	-0.0064	-0.1341
	MdAE	1.8551	2.3055
	MaxE	-4.5200	7.1571
	% >MAE	38 (14,24)	36 (18,18)
SBS	# features	14	13
	MAE	1.3389	2.1558
	SDAE	0.1629	0.1750
	MdAE	2.9023	2.7834
	MaxE	-9.5037	-6.3280
	% > MAE	10 (0,10)	48 (21,27)

Table 4.5 displays the statistical comparison for the regression analysis between the reading speech features in Database A and the HAMD score. The analysis revealed that the implementation of SBS and SFS feature selection procedure on male and female speech effectively predicted the HAMD scores with minimal MAEs (MAE less than three is considered insignificant). The SFS outperformed the SBS method in determining the suboptimal number of feature that will predict the HAMD scores with a minimal MAE. The maximum error (MaxE) for all categories is considered significant due to the value being larger than three. For the SFS procedure, 38% of the male reading speech yielded an absolute error prediction larger than 1.4282 and 36% of the female reading speech produced an absolute error prediction larger than

1.6893. On the other hand, using the SBS method, 10% of the male reading speech produced an absolute error prediction larger than 1.3389 and 48% of the female reading speech yielded an absolute error prediction larger than 2.1558. Referring to table 4.1, there are nine male patients and 14 female patients. Apart from the application of SBS method in the male reading speech, the rest of the groups achieved a minimal MAE using a total number of features that is less than the total number of patients thus indicating that the combinations of features are generalizable.

The plots of MAE HAMD scores prediction with respect to the total number of features obtained by the SFS and the SBS procedure for the male and female reading patients are illustrated in figures 4.19 to 4.20.

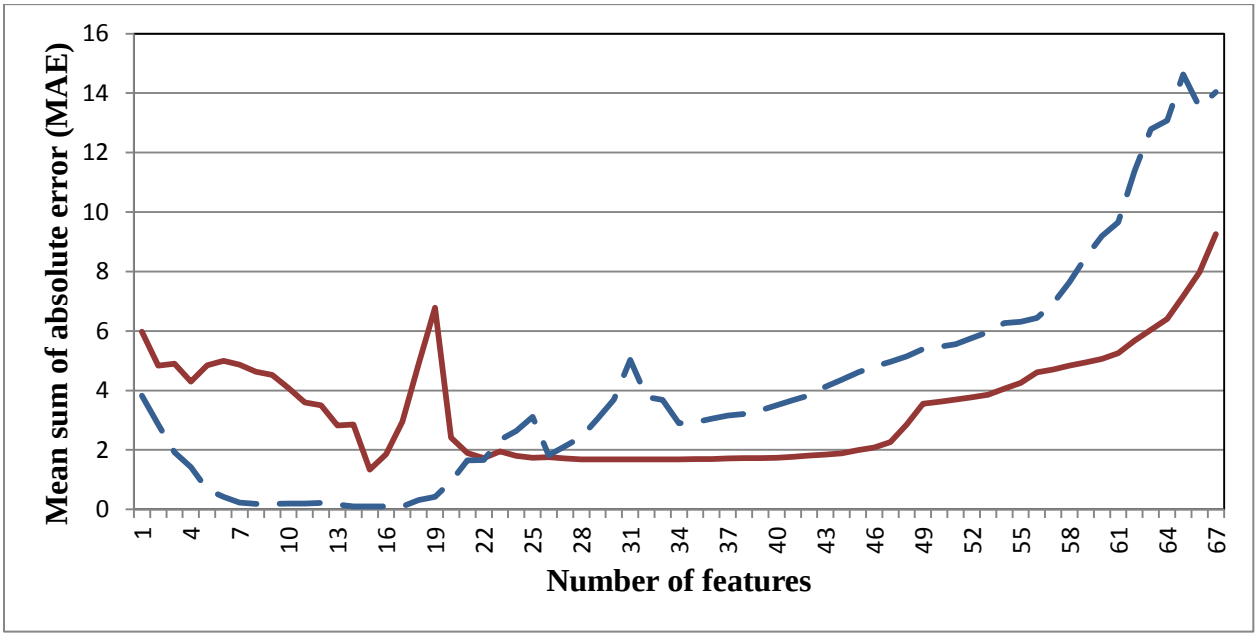


***Figure 4.19**: Characteristic plot of the SFS (blue line '--') and the SBS (red line) methods using the male reading speech from Database A to predict the HAMD scores*

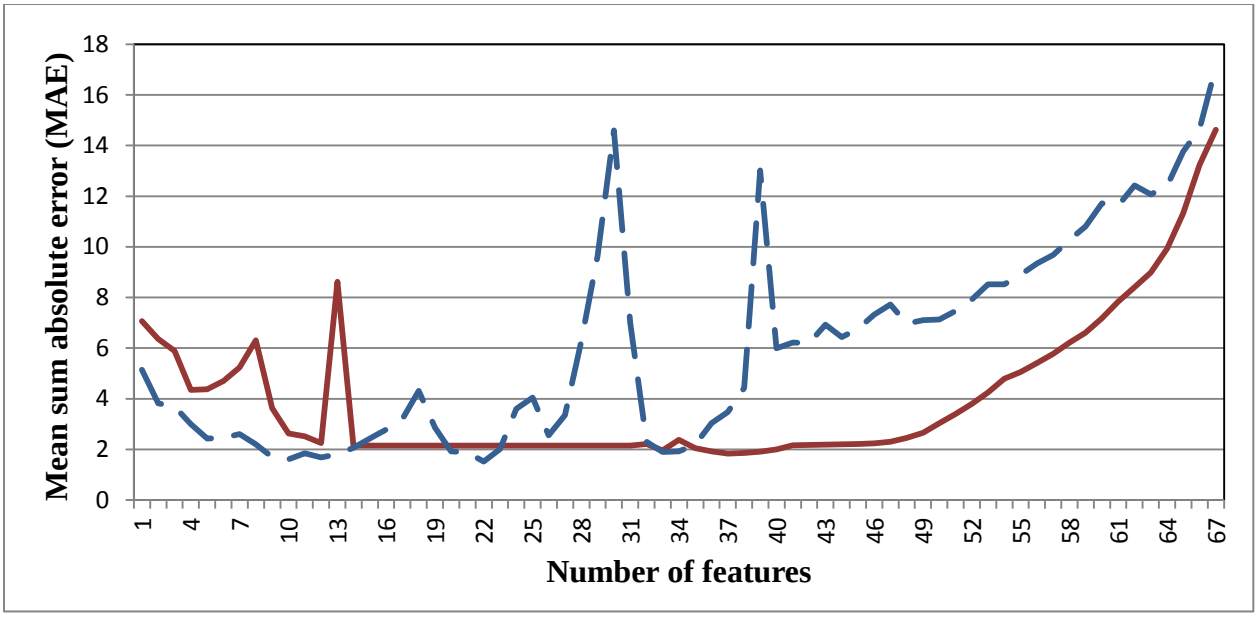


***Figure 4.20**: Characteristic plot of the SFS (blue line '--') and the SBS (red line) methods using the female reading speech from Database A to predict the HAMD scores*

The application of SBS method on the male and female reading speech demonstrated a similar characteristic as shown in figure 4.19 and figure 4.20. Keep in mind that the SBS graph reads from right to left because the initial set contains all 67 features and each feature was then discarded one at a time. The graph originally exhibits a decreasing trend which demonstrate that the discarded features were irrelevant thus improving the prediction performance. After reaching a certain point, the discarded features do not change the MAE significantly thus the approximately straight line was obtained for a number of feature combinations. After removing all irrelevant features, the graph exhibit an increment in MAE. This shows that the discarded features contains significant information thus removing these features decreases the performance of the prediction. For the female reading speech, the minimum MAE was obtained after removing all spectrum-based measure. However, for the male reading speech, the feature combination that yielded the minimum MAE contains a mixture of the spectrum- and timing-based measures.

Based on figure 4.19, the characteristic plot of the SFS method using male reading speech initially demonstrated a decreasing trend when adding one feature at a time. This shows that each added feature contains information that increases the accuracy of the prediction thus reducing the MAE. Beginning at the 5th added feature, the MAE continues to be less than one until the 20th

feature added. This demonstrates that the added features contain no significant information. The graph then exhibits an increase in MAE indicating that the additional information reduces the performance of the prediction. The combination of the first five features yielded a MAE of 0.6720. The 5th feature that was added to the combination is a spectrum-based measure. However, removing the 5th feature produced a MAE of 1.4282 which is still considered as an insignificant prediction error and plus, all four features are the timing-based measure. On the other hand, the combination of features using the SFS method that produced the minimum MAE (shown in figure 4.20) for the female reading speech consists of a combination of the spectrum- and timing-based measures.

Figures 4.21 to 4.22 display comparison plots of the actual and the predicted HAMD scores obtained using the selected SFS and SBS feature combination from the male reading speech.

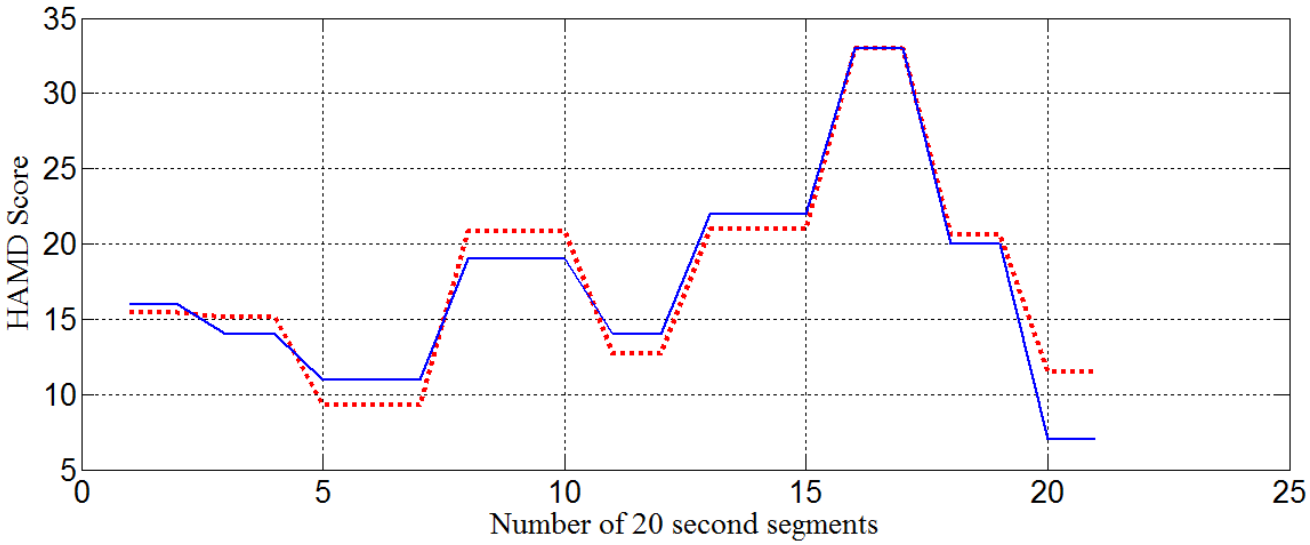


***Figure 4.21**: The actual (blue '—') and the predicted (red '--') HAMD scores for male reading patients in Database A using the SFS*

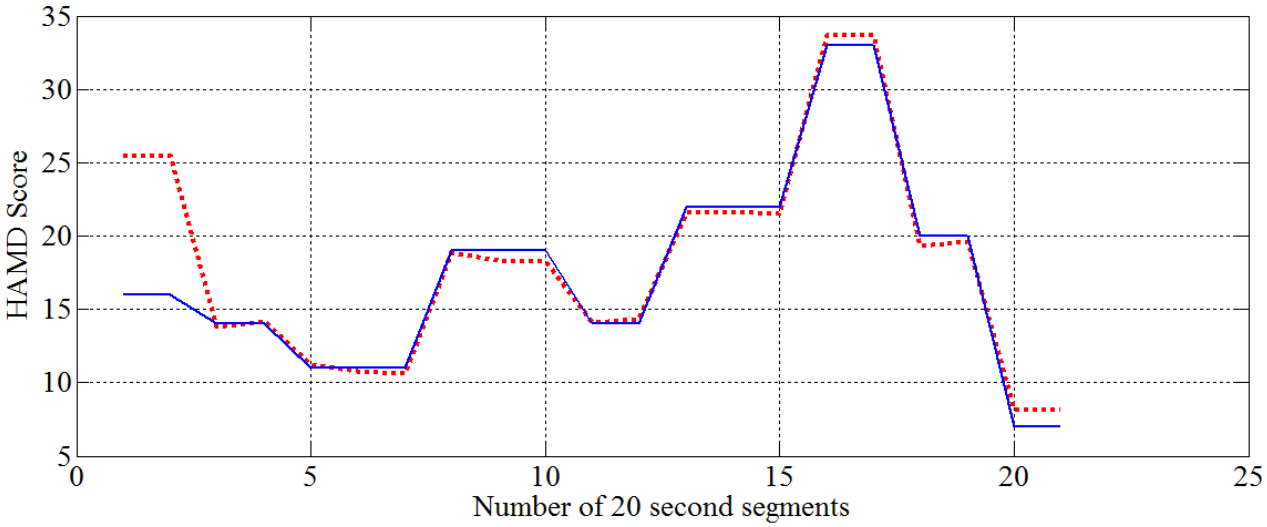


***Figure 4.22**: The actual (blue '—') and the predicted (red '--') HAMD scores for male reading patients in Database A using the SBS*

The predicted HAMD scores for the male patients using a feature combination obtained through the method of SBS were more accurate compared with the SFS method. However, the errors caused by the features selected using the SFS method are evenly distributed for each patient but the prediction errors produced by the features selected using the SBS method were mostly concentrated on the first patient (the first two 20 second segments) which is the reason why their MAE are almost equivalent.

Figures 4.23 to 4.24 display comparison plots of the actual and the predicted HAMD scores obtained using the selected SFS and SBS feature combination from the female reading speech.

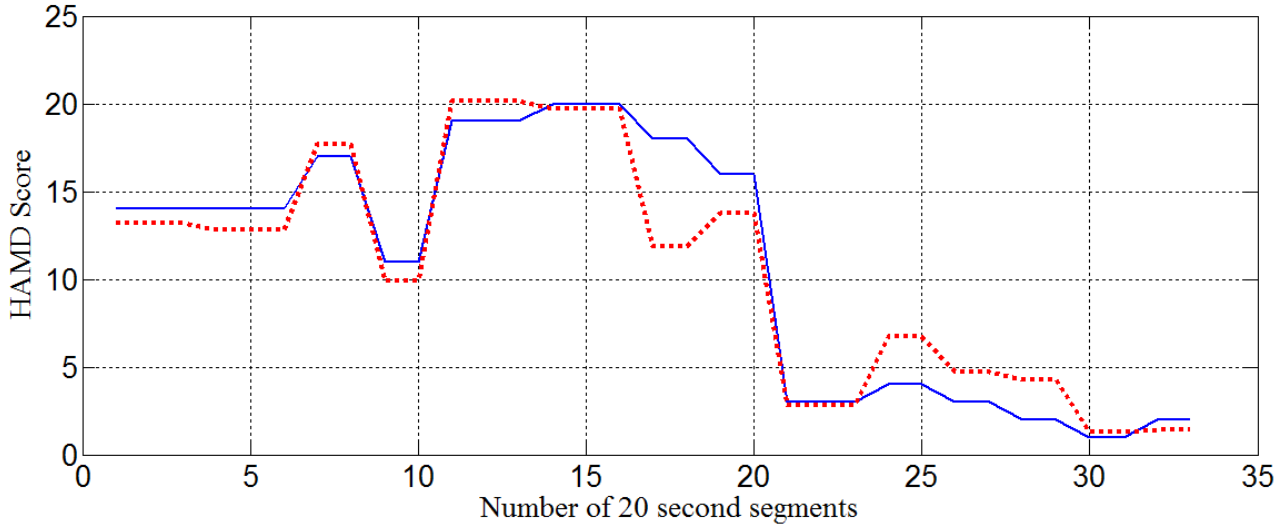


***Figure 4.23**: The actual (blue '—') and the predicted (red '--') HAMD scores for female reading patients in Database A using the SFS*

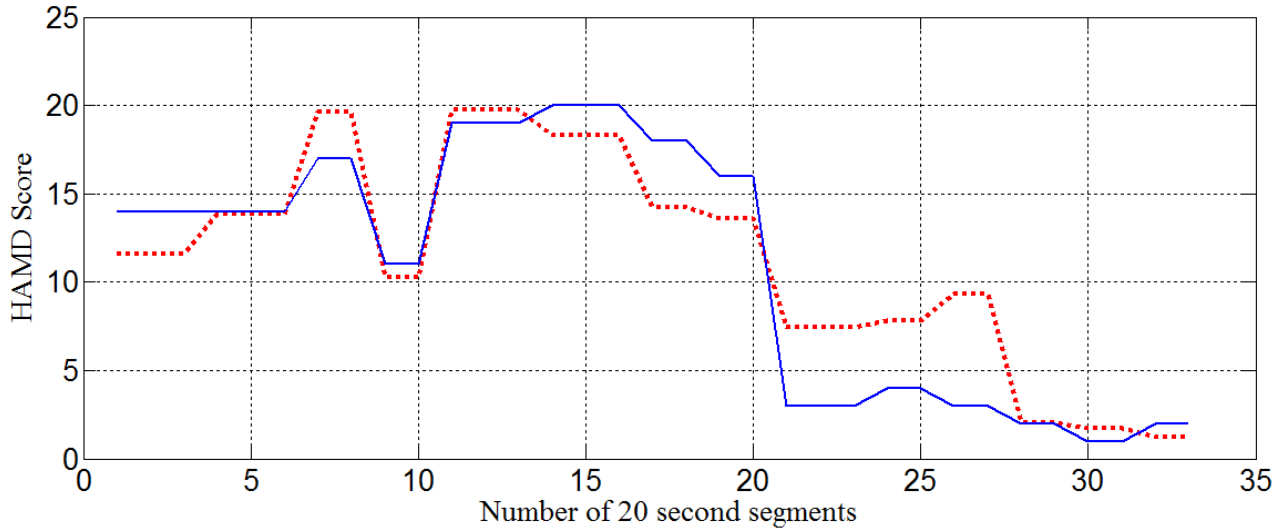


***Figure 4.24**: The actual (blue '—') and the predicted (red '--') HAMD scores for female reading patients in Database A using the SBS*

Comparison between the predicted HAMD score using a combination of features obtained by the method of SFS and SBS on female speech exhibited a similar characteristic in the direction of the errors. Overall, the prediction errors caused by the selected feature combination using the SBS method are slightly higher than the SFS method.

Table 4.6 Percentage of patients with an error prediction of the HAMD score of less than one, two or three for the male and female reading speech by methods of SFS and SBS

	SFS			SBS		
Percent Error (% Er)	< 1	< 2	< 3	< 1	< 2	< 3
Male Reading	28.57	90.48	90.48	80.95	90.48	90.48
Female Reading	39.39	72.73	87.88	42.42	51.52	72.73

Table 4.6 displays the comparison between the methods of SBS and SFS based on the percentage of the 20 second segment of speech that produced a HAMD score with prediction error of less than one, two and three for male and female reading speech. Error larger than three is assumed to be significant. Both sequential methods that were applied to the male reading speech successfully identified combination of features that effectively predicted the HAMD score with a performance of 90.48% error less than three. For the female patients, the method of SFS slightly outperformed the SBS in determining a combination of speech feature that could predict the HAMD score with a better accuracy and a higher percentage of prediction error less than three.

5.3 Regression Analysis on Speech Features with BDI-II using Database A

Table 4.7 Statistical comparison on the application of the forward (SFS) and backward (SBS) feature selection procedure using the reading speech from the male and female patients in Database A for predicting the BDI score

		Male Reading	Female Reading
SFS	# features	33	14
	MAE	4.8646	8.4502
	SDAE	1.1840	-0.6335
	MdAE	6.2891	11.6754
	MaxE	-18.6899	34.9431
	% >MAE	41 (22,19)	44 (20,24)
SBS	# features	34	38
	MAE	0.7284	8.5391
	SDAE	0.0075	0.1119
	MdAE	1.3116	12.1395
	MaxE	-5.4640	-47.1330
	% > MAE	22 (11,11)	37 (19,18)

Table 4.7 displays the statistical comparison between the implementation of the SFS and SBS procedure in identifying a combination of feature that can predict the BDI-II score for male and female patients using their reading speech. An acceptable 5% error for the BDI-II score would be an error of about 3.

Regression analysis on the female speech using the SFS and the SBS procedures failed to produce speech feature combinations that could predict the BDI-II score with an insignificant error. Results from the analysis revealed that the feature combinations obtained by the application of the SFS and SBS procedures to the male reading speech predicted the BDI-II score with a MAE of 4.8646 and 0.7284, respectively. Based on table 4.1, there are a total of 35 male patients and 69 segments of 20 second speech. The feature combination obtained by the SBS and SFS procedures for male speech are less than the total number of patients, thus the prediction is believed to be generalizable within similar population. 22% of the prediction errors were larger than the MAE with 11% of the error occurring in the positive direction and another 11% in the negative direction for the SBS method. On the other hand, 41% of the prediction errors were larger than the MAE with 22% of the error occurring in the positive direction and another 19% in the negative direction for the SFS method. In this case, the feature combination from the SBS method would be chosen over the SFS method due to the significant difference.

The plots of MAE BDI-II score prediction with respect to the total number of features obtained by the SFS and the SBS procedure for the male and female reading patients are illustrated in figures 4.25 to 4.26.

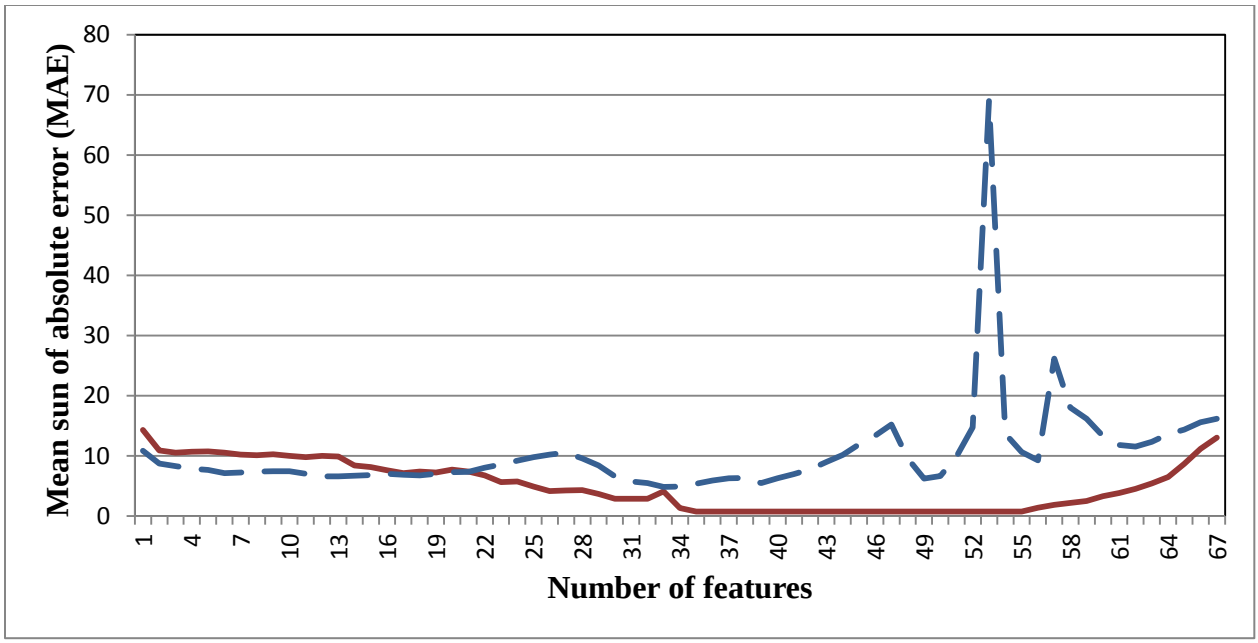


***Figure 4.25**: Characteristic plot of the SFS (blue line '--') and the SBS (red line) methods using the male reading speech from Database A to predict the BDI-II scores*

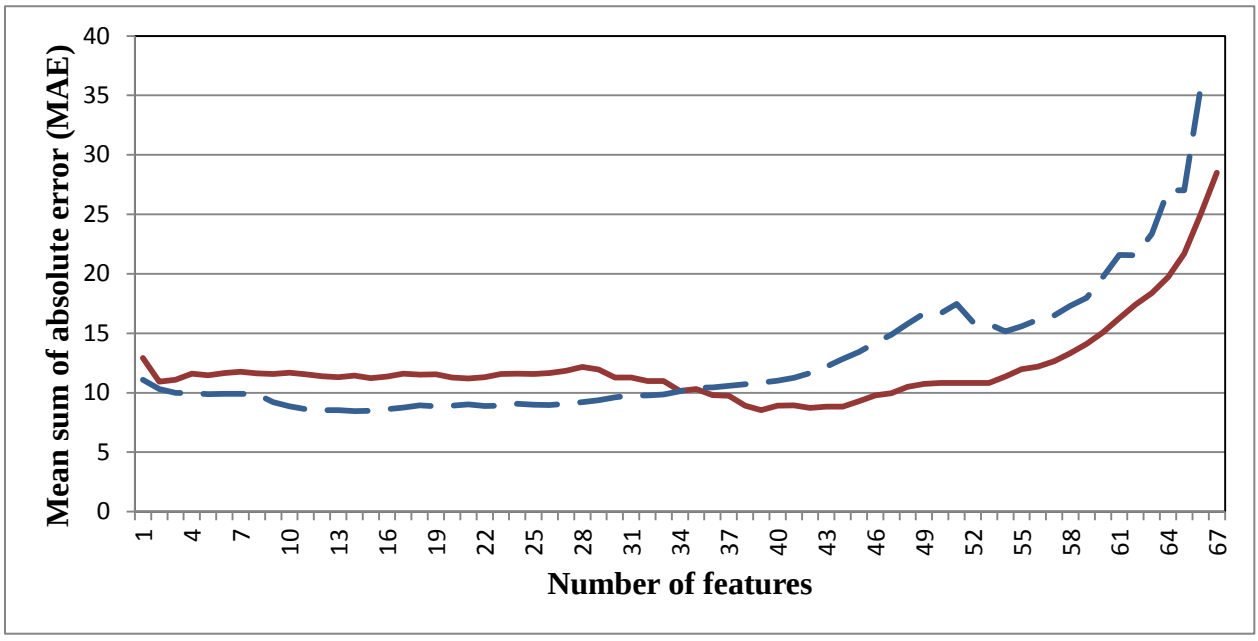


***Figure 4.26**: Characteristic plot of the SFS (blue line '--') and the SBS (red line) methods using the female reading speech from Database A to predict the BDI-II scores*

The analysis of regression demonstrated that the application of the SBS procedure on the male reading speech as shown in figure 4.25 lowered the MAE of the prediction as a feature being discarded one at a time. Starting about the 55^{th} feature until the next 20 features were discarded, the MAE remains unchanged. These 20 features have been identified as the spectrum based measures. The minimum MAE was achieved after removing all spectrum based measures. From here on, each discarded feature gradually increases the MAE or slightly decreases the MAE indicating that these features contain relevant information for predicting the BDI-II score.

On the other hand, prediction error caused by the feature combination obtained by the SFS procedure begins with a high MAE prediction and continued progressing with insignificant changes in the MAE but as more features being added, the MAE experiences a sudden increase and decrease thus represented by the spikes in the plot.

For the female reading data shown in figure 4.26, selecting features by the SBS method yielded an error of prediction that decreases exponentially as each feature being discarded. However, the error begins to stabilize but still unable to reach an appropriately low minimum MAE. Similar behavior was observed for regression analysis on the female reading speech using the SFS method as demonstrated in figure 4.26 except that the added features initially increases the MAE gradually and after approximately the 35^{th} feature, the MAE increases exponentially towards the end.

The selected feature combinations from the male and female reading speech that produced the MAE prediction of the BDI-II score based on the SFS and the SBS procedures were then compared to the actual score as illustrated in figures 4.27 to 4.30.

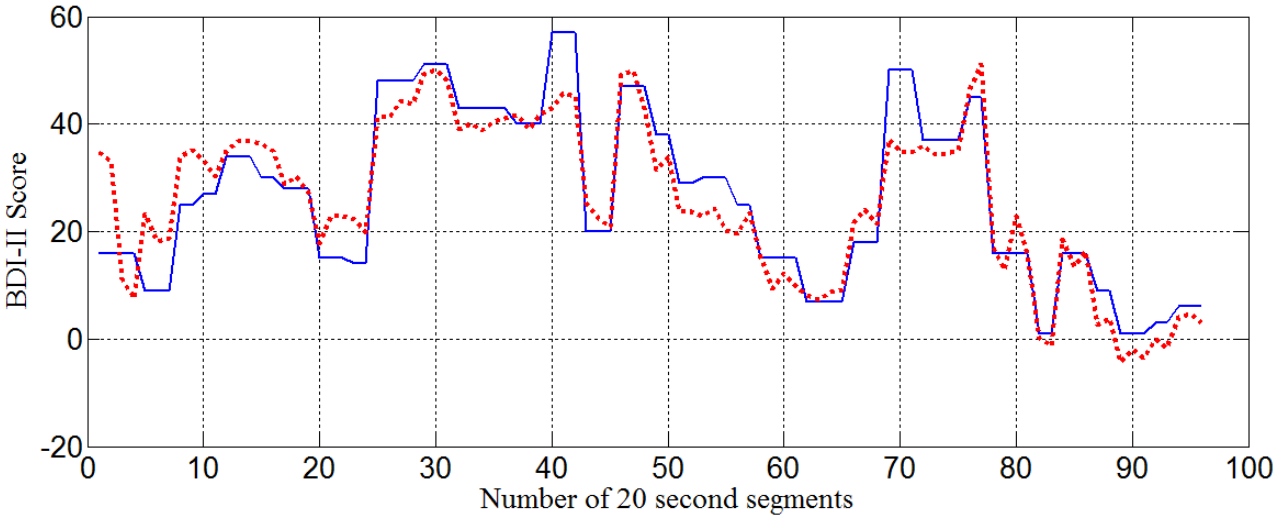


***Figure 4.27**: The actual (blue '—') and the predicted (red '--') BDI-II scores for male reading patients in Database A using the SFS procedure*

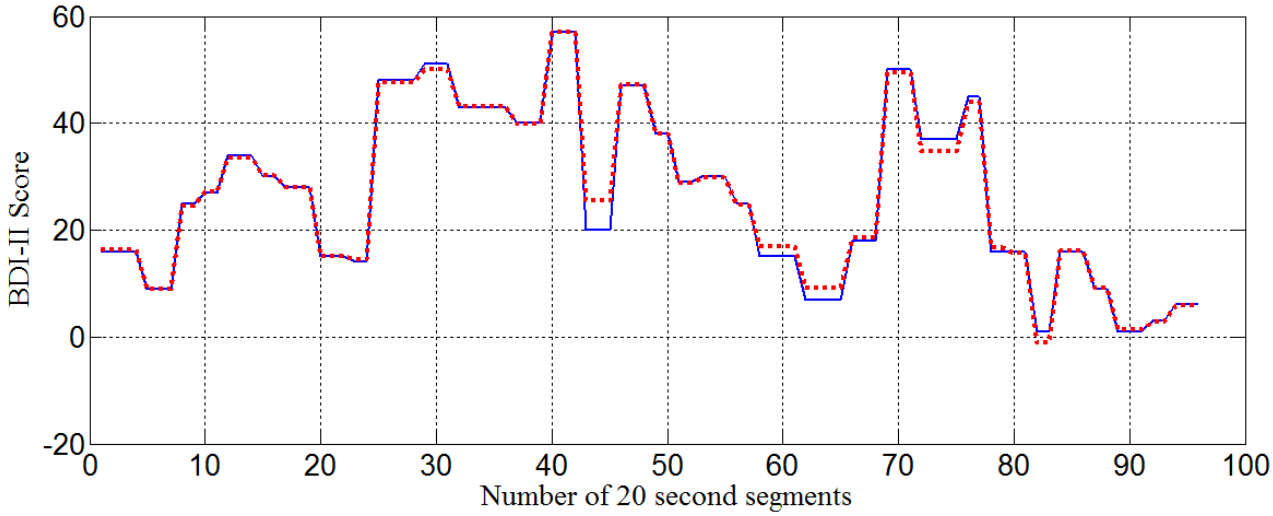


***Figure 4.28**: The actual (blue '—') and the predicted (red '--') BDI-II scores for male reading patients in Database A using the SBS procedure*

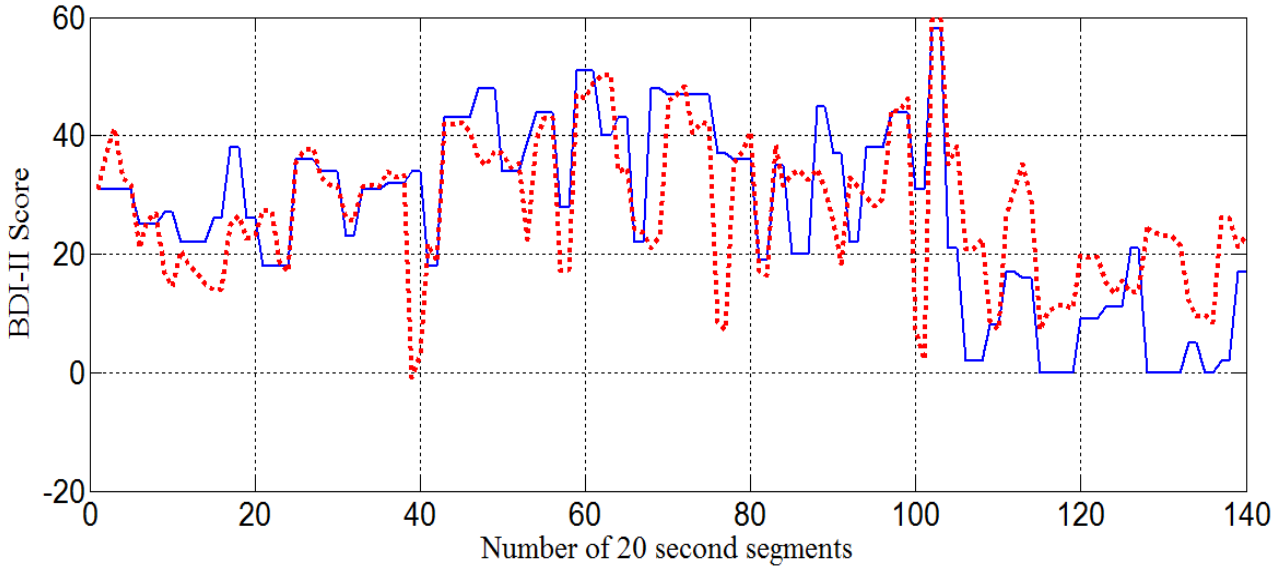


***Figure 4.29**: The actual (blue '—') and the predicted (red '--') BDI-II scores for female reading patients in Database A using the SFS procedure*

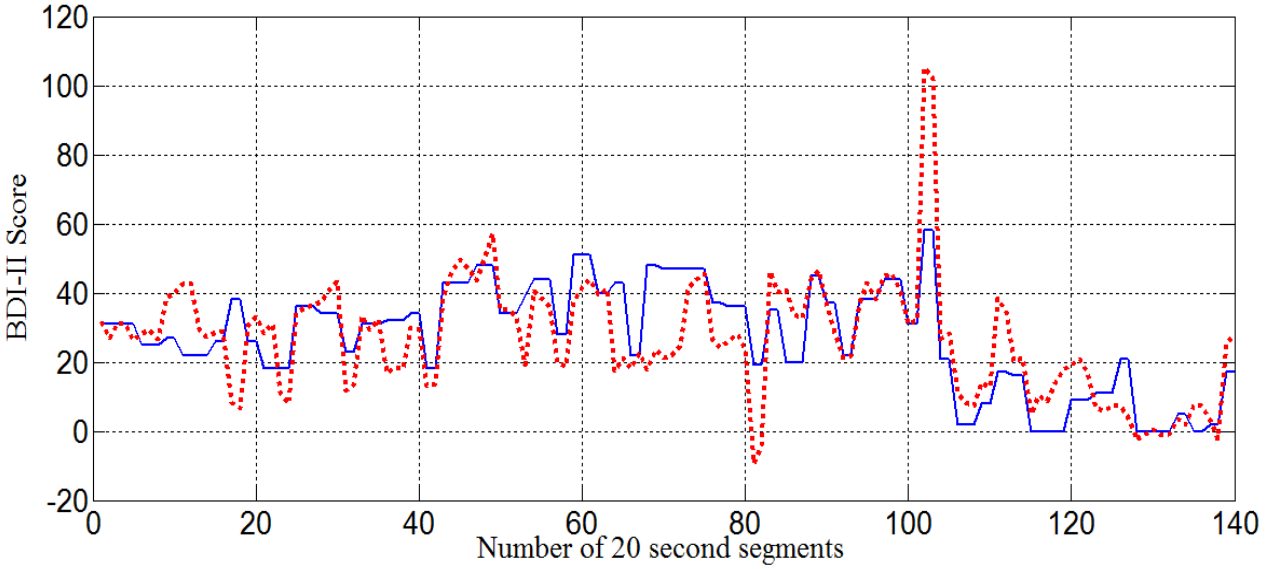


***Figure 4.30**: The actual (blue '—') and the predicted (red '--') BDI-II scores for female reading patients in Database A using the SBS procedure*

The predicted BDI-II score for the male speech using features obtained by the SFS method and for the female speech using features obtained by the SFS and SBS methods revealed that almost all of the patients' scores were predicted substantially incorrect. This is further

demonstrated in table 4.8 which shows the percentage of patients with a difference between the predicted and the actual BDI-II score of less than one, two and three. Using the SFS and SBS feature combinations from female speech, a poor performance of approximately 35.00% and 27.14% respectively were predicted with an error of less than three. The male reading feature combination from the SFS moderately predicted the actual BDI-II scores with an error of 44.79% less than three score.

Table 4.8: Percentage of patients with an error prediction of the BDI-II score of less than one, two and three for the male and female reading speech by methods of SFS and SBS

	SFS			SBS		
Percent Error (% Er)	< 1	< 2	< 3	< 1	< 2	< 3
Male Reading	12.50	22.92	44.79	80.21	86.46	96.88
Female Reading	14.29	27.86	35.00	12.14	24.29	27.14

By observing figure 4.28, it is clearly shown that the SBS feature combination predicted the male patient's BDI-II score fairly accurate. The combination of features successfully predicted the BDI-II score with a percentage of 96.88% errors less than six.

6.0 Discussion

This research demonstrated the effectiveness of using acoustic measurements as a possible means to predict the clinical scores. The results are based on method of linear regression and applying a feature selection procedure to increase the performance of predictions with fewer numbers of features. Previous related researches [23]-[29] have been studying the correlations between acoustic measurements and the clinical ratings. Regression and correlation are two powerful methods that describe and test associations between continuous variables. Although both methods are related, however their purpose is different. Correlation merely describes the relationship between two variables whereas regression predicts the value of a dependent variable (i.e., clinical scores) using one or more measurements of independent variables (i.e., acoustic features).

Analysis of the results revealed that the SBS method was able to identify a set of features that produced better predictions compared to the SFS method. However, the results also suggested that the SFS procedure more preferable when choosing a method on the basis of

smaller number of features. Backward and forward feature selection methods are popular choices for the use of dimensionality reduction of feature space and removing redundant, irrelevant or noisy data because the algorithms are simple and easy to implement. Also, feature selection selects a subset of features without transformation thus retaining the original physical interpretation whereas feature extraction reduces dimensionality by projecting onto a new dimension. Even though both methods performed well in this study, these algorithms have a tendency to produce error that will move in a downward direction and can sometimes become trapped in local minima. The drawback for the SFS procedure is the inability of replacing or eliminating selected features that have become redundant after the inclusion of new features while the SBS method does not allow the discarded features to be reexamined once removed from the set and thus eliminate the likelihood of finding a feature that works best on its own.

In this study, we implemented the jackknife analysis using a heuristic search algorithm which is the sequential feature selection method for searching a set of features that are close to an optimal solution. This procedure is simple but it only explores a limited number of structures. On the other hand, the brute force technique searches for all possible outcomes and it is also capable of producing an optimal solution, however, this procedure is considered unreasonable to be applied in this study. The reason is because of the extensive number of possible combinations to search for the 67 features. For example, finding all possible combinations of 10 features out of 67 features will require approximately 1.28×10^{12} jackknife analyses to be executed.

Evaluation of model performance error was based on the measure of mean absolute error (MAE). This quantity was preferred over mean squared error (MSE) because the nature of its calculation clearly describes the results. MAE weighs all individual differences in an equally manner and clearly measures how close prediction are to the actual value without considering their direction. Although MSE is a conventional method that has been used commonly, the error measures are considered unreliable for this study because the interpretations of the results are abstruse. MSE quantifies the variance of errors by measuring the difference between the prediction and the actual value. Also, by squaring the errors, greater emphasis is being put on to large errors thus allowing the total square error to be affected by a possible existence of outliers. Another error measurement that was also reported in this study is the maximum error (MaxE) which expresses the maximum estimated differences between the prediction and the actual value.

The results act as a bound that describes the maximum margin of an error for a certain set of predictions.

In cases where there is a small number of training data, non-linear model may not be appropriate to be used because of the insufficient information to represent the complexity of the model. Even for a linear model, least square regression also generally tends to perform poorly. One way to convince the assumption of linearity with a small sample size data was by incorporating the jackknife technique when performing linear regression. The reason is because jackknife technique creates new observation each time by removing one patient at a time and evaluating the ability of an individual to be predicted by the remainder thus, mimicking the process of having completely different training datasets.

Essentially, by looking at the results, we found that we were able to identify features that do generalize by demonstrating the ability of a certain population to predict the behavior of an individual through their clinical scores. However, the regression and the feature selection methods are design selections and may require further analysis by taking into account the different error trade-offs.

References

[1] Centers for Disease Control and Prevention [http://www.cdc.gov]

[2] DK Kenneth, X Jiaquan, LM Sherry, MM Arialdi, K Hsiang-Ching, Deaths: Final Data for 2009, National Vital Statistic Reports, vol. 60(3) (2011)

[3] JL McIntosh, U.S.A Suicide: 2009 Official Final Data, American Association of Suicidology (2010), http://www.suicidology.org. Accessed March 25, 2012

[4] R Burns, An impact: Suicides are surging among US troops (The Associated Press, 2012 in press)

[5] LR Wingate, TE Joiner, RL Walker, MD Rudd, DA Jobes, Empirically informed approaches to topics in suicide risk assessment. Behavioral Sciences & the Law (Suicide and the Law), pp. 651–665 (2004)

[6] EK Moscicki, Epidemiology of Completed and Attempted Suicide: Toward a Framework for Prevention, Clinical Neuroscience, 1, 310-23 (2001)

[7] M Hamilton, A Rating Scale for Depression, Journal Neurol. Neurosurg. Psychiat., 23, 56 (1960)

[8] GK Brown, A Review of Suicide Assessment Measures for Intervention Research with Adults and Older Adults, Technical report submitted to NIMH Bethesda, MD: National Institute of Mental Health (2002)

[9] DJ France, RG Shiavi, SE Silverman, MK Silverman, DM Wilkes, Acoustical Properties of Speech as Indicators of Depression and Suicidal Risk, IEEE Transaction on Biomedical Engineering, Vol 47, No 7 (2000)

[10] A Ozdas, RG Shiavi, SE Silverman, MK Silverman, DM Wilkes, Investigation of Vocal Jitter and Glottal Flow Spectrum as Possible Cues for Depression and Near-Term Suicidal Risk, IEEE Transaction on Biomedical Engineering, Vol 51, No 9 (2004)

[11] T Yingthawornsuk, "Acoustic Analysis of Vocal Output Characteristics for Suicidal Risk Assessment", Ph.D, Thesis, Vanderbilt University (2007)

[12] HK Keskinpala, Analysis of Spectral Properties of Speech for Detecting Suicide Risk and Impact of Gender Specific Differences, PhD Thesis, Vanderbilt University (2011)

[13] NH Nik Wahidah, Analysis of Power Spectrum Density of Male Speech as Indicator for High Risk and Depression Decision, MS Thesis, Vanderbilt University (2011)

[14] WS Wan Hasan, Acoustic Analysis of Speech Based on Power Spectral Density Features in Detecting Suicidal Risk Among Female Patients, MS Thesis, Vanderbilt University (2011)

[15] NH Nik Wahidah, DM Wilkes, RM Salomon, JS Meggs, Analysis of Features Based on the Timing Pattern of Speech as Potential Indicator of High Risk Suicidal and Depression, Manuscript submitted for publication (2012)

[16] E Moore, MA Clements, JW Peifer, L Weisser, Critical Analysis of the Impact of Glottal features in the Classification of Clinical Depression in Speech, IEEE Transaction on Biomedical Engineering, Vol 55, No 1 (2008)

[17] JHL Hansen, MA Clements, Evaluation of Speech Under Stress and Emotional Condition, Journal of Acoustic Society of America, Vol 81, pp.S17-S18 (1987)

[18] ME Ayadi, MS Kamel, F Karray, Survey on Speech Emotion Recognition: Features, Classification Schemes and Databases, Pattern Recognition, 44(3), pp.572-587 (2011)

[19] NG Bradley, AJ Rush, MH Trivedi, SR Wisniewski, GK Balasubramani, DC Spencer, T Petersen, M Klinkman, D Warden, L Nicholas, M Faza, Major Depression Symptoms in Primary Care and Psychiatric Care Settings: A Cross-Sectional Analysis, Annals of Family Medicine, vol 5, no 2 (2007)

[20] Y Conwell, Management of Suicidal Behavior in the Elderly, Psychiatric Clinics of North America, Vol 2, Issue 3 (1997)

[21] SE Silverman, "Vocal parameters as predictors of near-term suicidal risk", U.S. Patent 5 148 483, (1992)

[22] Interactive Screening Program [http://www.afsp.org/]

[23] P Hardy, R Jouvent, D Widlocher, Speech Pause Time and the Retardation Rating Scale for Depression (ERD), Towards a Reciprocal Validation, Journal of Affective Disorders (1984)

[24] H Ellgring, KR Scherer, Vocal Indication of Mood Change in Depression, Journal of Nonverbal Behavior, 20(2), 83-110 (1996)

[25] HH Stasen, S Kuny, D Hell, The Speech Analysis Approach to Determining Onset of Improvement Under Antidepressants, European Neuropsychopharmachology, 8, 303-310 (1998)

[26] M Cannizaro, B Harel, N Reilly, P Chappell, PJ Snyder, Voice Acoustical Measurement of the Severity of Major Depression, Brain and Cognition, 56(1), pp.30-35, (2004).

[27] JC Mundt, PJ Snyder, MS Cannizzaro, K Chappie, DS Geralts, Voice Acoustic Measures of Depression Severity and Treatment Response Collected Via Interactive Voice Response (IVR) Technology, Journal of Neurolinguistic, 20, 50-64 (2007)

[28] JC Mundt, AP Vogel, DE Feltner, WR Lenderking, Vocal Acoustic Biomarkers of Depression Severity and Treatment Response, Journal of Biological Psychiatry, 72, 580-587 (2012)

[29] AC Trevino, TF Quatieri, N Malyska, Phonologically-based Biomarkers for Major Depressive Disorders, Journal on Advances in Signal Processing (2011)

[30] 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 in press)

[31] S Malcolm, Auditory Toolbox version 2, Interval Research Corporation, Technical Report (1998)

[32] DW Aha, RL Bankert, A Comparative Evaluation of Sequential Feature Selection Algorithms, Lecture Notes in Statistics, Volume 112, pp 199-206 (1996)

[33] L Ladha, T Deepa, Feature Selection Methods and Algorithms, International Journal of Computer Science and Engineering, Volume 3, No. 5 (2011)

[34] MW Jeffrey, Introductory Econometrics: A Modern Approach, Fourth Edition, Chapter 3, p. 73-120 (2009) (http://www.swlearning.com/pdfs/chapter/0324289782_3.PDF)

[35] AT Beck, RA Steer and GK Brown, Manual for Beck Depression Inventory-II, San Antonio, TX: Psychological Corporation, (1996)

[36] S Levine, RJ Ancill, AP Roberts, Assessment of suicide risk by computer-delivered self-rating questionnaire: Preliminary findings. Acta psychiatrica Scandanavica, 80, 216-220, (1989)

[37] A Bowling, Mode of questionnaires administration can have serious effects on data quality, Journal of public health, Oxford, England Vol. 27, No. 3, p. 281-91, (2005)

[38] RM Salomon, HK Keskinpala, MH Sanchez, T Yingthawornsuk, NH Wahidah, WS Hasan, N Taneja, D Vergyri, BH Knoth, PE Garcia, DM Wilkes, R Shiavi, Analysis of Voice Speech Indicators in Suicidal Patients. Manuscript submitted for publication (2012)

[39] Ying Yang, F. Catherine, J. F. Cohn, "Detecting Depression Severity from Vocal Prosody", *IEEE Trans. On Affective Computing*, vol. 4, no. 2, 2013.

CHAPTER V

ANALYSIS OF CLASSIFICATION BASED ON AMPLITUDE MODULATION IN THE SPEECH OF DEPRESSED AND HIGH RISK SUICIDAL MALE AND FEMALE PATIENTS

Abstract

This analysis seeks to study the characteristics of root mean square amplitude modulation (RMS AM) in the speech of depressed and near-term suicidal patients in order to determine its potential for discriminating between the two groups. The study is a partial replication and extension of the work by France [1], who reported an effective overall classification score of 77% for depressed and near term suicidal speech in male patients. The current database consists of interviews and passage-based readings by male and female patients. Statistical RMS AM measures include the maximum, range, variation, average, skewness, kurtosis and coefficient of variation. Analyses were performed using linear (LDA) and quadratic (QDA) discriminant analysis with three resampling methods of equal-test-train, jackknife and cross-validation. France's RMS AM results were partially replicated: France reported a combination of RMS AM range and coefficient of variantion in male speech as significant features whereas the analysis in this paper identified a combination of RMS AM range and skewness as a significant discriminator for male reading speech. However, poor classification scores were demonstrated for male interview, female interview and female reading speech.

1.0 Introduction

Speech is a rich source of information that is not only used to convey spoken messages between a speaker and a listener, but it also contains hidden messages such as emotions, mental state and attitude of the speaker. Various studies have recognized the fact that there is a relationship between human speech and the individual's psychological state. One manner in which this relationship can be applied is in the detection and diagnosis of patients with psychiatric disorders. The primary method of psychiatric assessment requires extensive effort by the clinicians which involves gathering comprehensive information about the patient and

answering a series of questions in the presence of a trained clinician. Identifying a patient's mental condition from the speech signal is quite desirable since recording a stream of data and extracting features from it is comparatively easier and simpler than the conventional method that is often time-consuming. Psychological screening by means of human speech may be used as a secondary tool in helping clinicians to recognize psychiatric disorders and prevent misdiagnosis particularly in distinguishing between patients that are experiencing depression and near-term suicidal predisposition.

Among the speech features that have been investigated in the field of psychiatry are the Power Spectral Density (PSD), Mel-Frequency Cepstral Coefficient (MFCC), fundamental frequency (F0), formant, glottal flow spectrum, transition parameters and interval probability density function [2] - [13]. Another acoustic feature that has been suggested as one of the characteristics of the depressed *speech* is known as the root mean square amplitude modulation (RMS AM). This feature represents the envelope of a waveform. Patients with psychological disorders are often believed to speak with lack of variation in pitch and amplitude. This pattern of speaking is often dull and often described by trained listeners as monoloud and monotone.

Over the years, the amplitude modulation has not been studied systematically and understandably, the literature on the subject pertaining to the field of psychology is sparse. The earliest study on the effect of the depressed and the high risk suicidal speech on the amplitude modulation were performed by Silverman [7, 8]. He described the speech signal of a high risk suicidal patient as either having a low value of amplitude modulation or exhibiting a slow decay at the end of every utterance. Mundt [14] reported that a depressed speech pattern has often been characterized as dull, monotone, monoloud, lifeless and metallic. Kraepelin [15] described the severely depressed patients to "speak in a low voice, slowly, hesitatingly, monotonously, sometimes stuttering, whispering, try several times before they bring out a word, and become mute in the middle of a sentence". Clinical observations performed by Ostwald [16] on 30 acutely disturbed psychiatric patients revealed that a change in loudness level as the single most consistent measure of clinical improvement.

In the previous research, France [2] investigated RMS AM as one of the acoustic feature in distinguishing between the depressed and high risk suicidal speech in male patients. Six statistical measurements that are known as the range, mean, variance, skewness, kurtosis, and coefficient of variation were extracted for each 20 second segment of speech. It was reported that

there was a significant difference between groups of the depressed and high risk suicidal speech when a classification analysis was performed using a combination of RMS range and coefficient of variation. Classification by means of a leave one out (jackknife) procedure and quadratic discriminator yielded an *equally* effective performance of 77% correctly classified high risk suicidal patients and 76% correctly classified depressed patients. RMS AM range, skewness and kurtosis were also shown to be reduced in the high risk suicidal speech while the RMS AM coefficient of variation increased slightly higher.

This study attempts to continue the work by France, particularly on the analysis of classification using RMS AM and examine the relative contributions of amplitude variation on the depressed and high risk suicidal speech. He performed the analysis using a very diverse recording environment, often poor quality devices such as telephone messages and cheap recorders. France previously examined this feature only on the depressed and high risk suicidal male patients, however in this analysis, the RMS AM feature will be analyzed using a high quality interview and reading speech database containing both male and female patients. The objective of this study is also to investigate the RMS AM feature properties in the depressed and the high risk suicidal patients using the conventional analytical method used throughout this thesis.

2.0 Database

2.1 Database Collection

All recording sessions were conducted at the Vanderbilt Emergency Department or Psychiatric Hospital with patient's documented informed consent. Patients who volunteered were made aware of the aim of the study with an assurance of maximum identity protection procedures. Patients under the influence of alcohol, toxicity or experiencing respiratory problems such as shortness of breath were excluded. All recordings were made in a standard, empty psychiatric interview room without the benefit of soundproof or acoustically ideal environment, mimicking the real-world clinical environment. For this research, only the interview and reading speech from Database A that was gathered from the high risk suicidal and depressed patients were used. Group assignment was made according to assessment made by experienced clinicians using the Beck Depression Inventory (BDI-II), MINI International Neuropsychiatric Interview and Pierce Suicide Intent Scale (SIS) [17]. Patients were asked to read from a standardized

"rainbow passage" which contains every sound in the English language and is considered to be phonetically balanced with the ratios of assorted phonemes similar to the ones in normal speech [18].

All speech samples were digitized using a 32-bit analog to digital converter at 44.1 kHz sampling rate for both databases. Table 5.1 shows the number of patients for male and female interview and reading sessions from Database A. Recordings were collected once per patient and each recording was categorized as either high risk suicidal or depressed. Audio acquisitions were made using a high-quality Audix SCX-one cardioid microphone with a frequency response of 40Hz to 20kHz, Sony VAIO laptop with Pentium IV 2GHz CPU 512 Mb memory, Windows XP, a Digital Audio MBox for digital audio interface and recording software PROTools LE for the digital audio editor.

Table 5.1: Number of male and female patients for interview and reading sessions

Database A	**Male**				**Female**			
	HR*		DP*		HR		DP	
	Int	Read	Int	Read	Int	Read	Int	Read
Total number of patients	10	8	11	12	12	11	20	18
Total number of 20 second segments	252	58	194	67	163	59	400	95

*Int-Interview recordings, Read-Reading recordings, HR-high risk suicidal, DP-depressed

2.2 Data Pre-processing

In the preprocessing stage, recordings were edited using a free audio digital editor called Audacity 2.0.1 to remove any identifying information, to preserve patient privacy. Undesirable sounds such as long pauses that are present for more than 0.5 second, the interviewer's voice, voices other than the patient, sneezing, coughing and door slams were removed from the de-identified recordings. Each edited speech sample was detrended by subtracting the mean value to compensate for possible variability that exists during recording. For analysis purposes, the sampled signals were divided into 20 second segments. For one segment of 20 second voiced sample recording, the mean of each six RMS amplitude modulation statistical measurements were obtained. Therefore, each patient will have a certain set of the six RMS amplitude modulation statistical measurements depending on the number of 20 second segments it has.

3.0 Methodology

3.1 Feature Extraction

3.1.1 Root Mean Square Amplitude Modulation (RMS AM)

An envelope of a signal can be thought of as the outline of the signal that connects all the peaks. The amplitude changes of the signal carry the information we seek. The method used for this analysis is normally called 'the square-law envelope detector' as shown in figure 5.1. This method squares the input signal and sends it through an averaging represented by a lowpass filter. An averaging is a crude lowpass filter with a gain of 1. By squaring the signal, the input signal is demodulated by using itself as the carrier wave. The square root is then taken in order to reverse the scaling distortion from squaring the signal and to characterize a more accurate statistical measure.

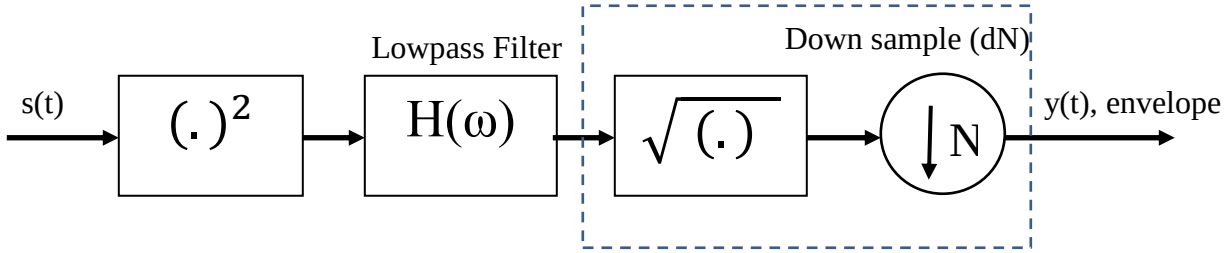


Figure 5.1: *Block diagram representing the square-law envelope detector*

The method of RMS can be easily adapted to obtain the amplitude envelope by applying it with a sliding window. Assuming the signal is zero mean, amplitude RMS is given by the equation:

$$Amplitude\ RMS(t) = \sqrt{\frac{1}{T}\sum_{i=0}^{T} w_i(t)\ {x_i}^2(t)}$$

where, $x_i(t)$ is the sampled signal with t number of samples through the analysis window,

$w_i(t)$ is the sliding window

and T is the window length

The amplitude modulation extraction procedure is outlined as follows:

1. Detrend and normalize the speech signal for each 20 second segment using the whole speech sample with voiced, unvoiced and silence segments.

2. Square the speech signal.

3. Convolve the squared signal with a 'sliding' 40ms window.

4. Calculate the mean and square-root for every frame.

5. Calculate the seven following statistical measurements for each frame (maximum, range, variance, average, skewness, kurtosis, and coefficient of variation).

6. Obtain a single mean vector containing the seven statistical measurements to represent the 20 second segment.

7. Repeat step #1 for the next 20 second segment.

8. Repeat steps #1 to #6 for the next patient

A complete description on what the six statistical measurements represent and how to obtain them is further discussed in [2].

3.2 Discriminant Analysis and Resampling Method

The pairwise classification was performed using the method of linear (LDA) and quadratic (QDA) discriminant analysis. Due to small sample size, a resampling method was necessary to be used when performing the classification. The resampling methods that were adopted in this research were Equal Test-Train, Jackknife (Leave-One-Out) and Cross-Validation. A further description on these statistical methods was discussed in chapter III sections 4.3 and 4.4.

3.3 Analysis of Classification

Each patient will have a vector of *m* rows by seven columns where *m* is the number of 20 second segment and the columns are the RMS AM maximum, range, variance, average, skewness, kurtosis and coefficient of variation. The classification analysis was performed on all possible combinations of the seven RMS AM statistical features. There are a total number of 127 possible combinations created from the seven features.

4.0 Results and Analysis

4.1 Statistical Analysis

A preliminary test for the equality of the variances indicates that the variances of the two samples were significantly different. Therefore, a two sample t-test with the assumption of unequal variances was done to compare the mean of each RMS AM statistical measurement extracted from the high risk suicidal and the depressed speech. Table 5.2 displays the statistical comparison between the six measurements (excluding the feature RMS AM maximum) for both male and female interview and reading speech. At the 0.01 significance level, we found that the mean skewness measurement from male reading speech is probably significantly different with a p-value of 0.0169 which is less than 0.05 but larger than 0.01 (0.05 < significance < 0.01). However, besides skewness for male reading, there are no other significant evidence between the mean of the two samples of high risk and depressed speech shown by the RMS AM features.

Table 5.2 Comparison between six RMS AM statistical measurements for male and female interview and reading speech

		Male Interview	Male Reading	Female Interview	Female Reading
Range	μ_{HR}	3.43 ± 0.66	3.14 ± 0.84	3.18 ± 0.35	3.10 ± 0.40
	μ_{DEP}	3.54 ± 0.49	2.59 ± 0.36	3.38 ± 0.72	2.97 ± 0.47
	p-value	0.3345	0.0558	0.1502	0.2041
Variance	μ_{HR}	0.26 ± 0.11	0.26 ± 0.10	0.27 ± 0.07	0.29 ± 0.06
	μ_{DEP}	0.32 ± 0.05	0.27 ± 0.04	0.28 ± 0.09	0.26 ± 0.07
	p-value	0.0852	0.4661	0.3053	0.1617
Average	μ_{HR}	0.85 ± 0.07	0.86 ± 0.06	0.85 ± 0.04	0.84 ± 0.04
	μ_{DEP}	0.82 ± 0.03	0.86 ± 0.02	0.84 ± 0.05	0.86 ± 0.04
	p-value	0.0970	0.4900	0.2828	0.1683
Skewness	μ_{HR}	1.58 ± 0.38	1.21 ± 0.46	1.24 ± 0.36	0.97 ± 0.20
	μ_{DEP}	1.38 ± 0.47	0.73 ± 0.42	1.40 ± 0.51	0.99 ± 0.34
	p-value	0.1392	0.0169	0.1533	0.4174
Kurtosis	μ_{HR}	4.49 ± 3.25	2.83 ± 3.60	2.41 ± 1.79	1.21 ± 0.87
	μ_{DEP}	2.99 ± 2.66	0.65 ± 1.43	3.30 ± 2.97	1.23 ± 1.17
	p-value	0.1330	0.0712	0.1499	0.4807
Coefficient of variation	μ_{HR}	0.60 ± 0.18	0.60 ± 0.17	0.61 ± 0.11	0.63 ± 0.09
	μ_{DEP}	0.69 ± 0.08	0.60 ± 0.06	0.63 ± 0.15	0.59 ± 0.11
	p-value	0.0959	0.4799	0.2950	0.1575

4.2 Classification Analysis for the High Risk Suicidal and the Depressed Group

The classification analysis that was calculated based on the percentage of all data represents the percentage of the total number of high risk suicidal and depressed patients that are correctly classified. The high risk percentage represents the percentage of high risk patients that are correctly classified and the depressed percentage represents the percentage of depressed patients that are correctly classified. A good performance classifier is evaluated based on its ability to classify both groups equally well with a high percentage of correctness and using the smallest number of feature (lower dimensionality). For our purposes, a classification score of 80% and above is considered excellent, 60% to 79% as moderate and less than 60% as poor.

4.2.1 Male Interview Results

Based on observing the all data percentage, the performance of the QDA slightly exceeds the LDA for all resampling methods. This result implies that more patients were classified correctly in the QDA compared with the LDA.

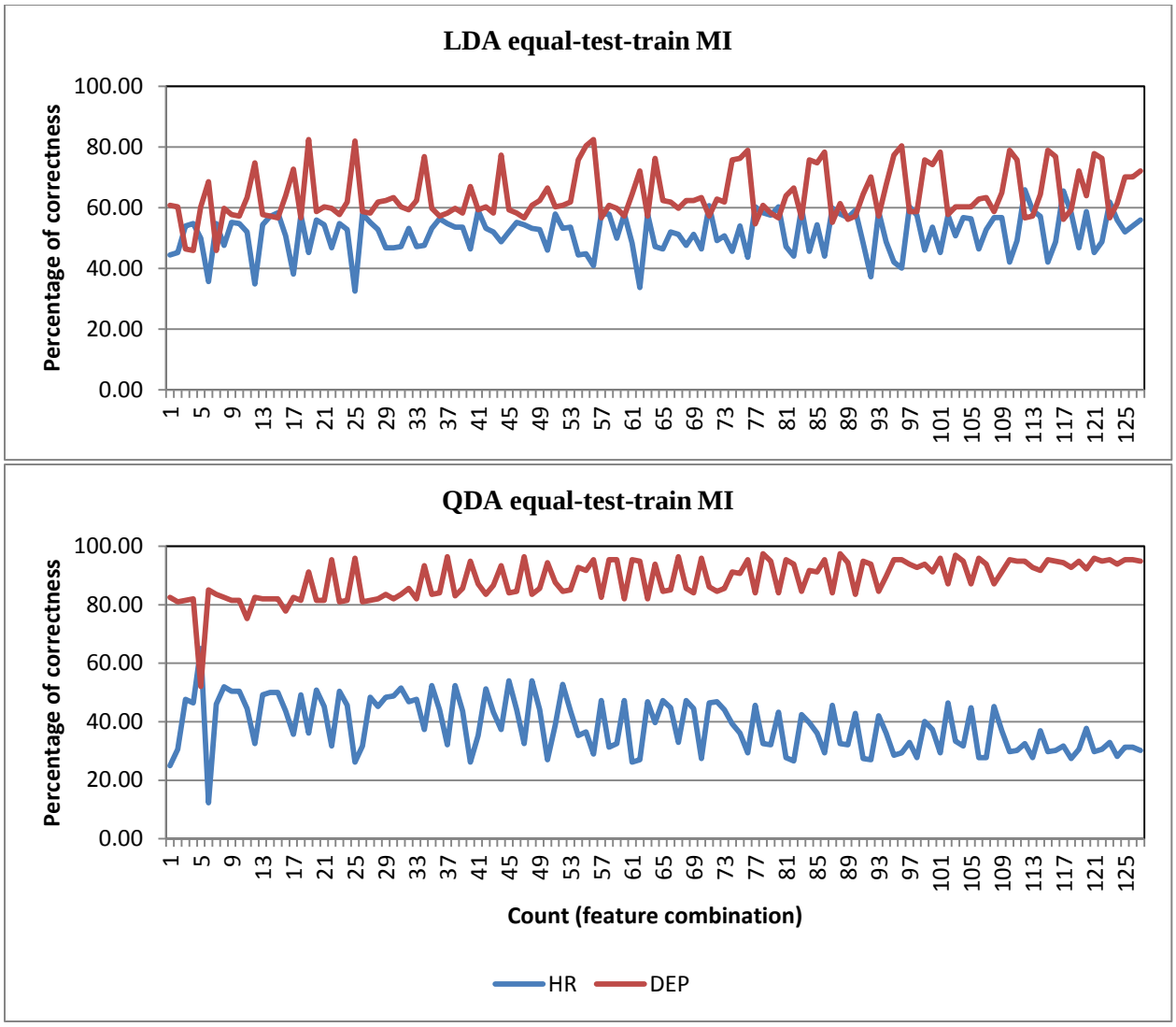


Figure 5.2: *Plot of the correctly classified high risk suicidal and depressed percentages over a number of feature combinations using the linear and quadratic classification with the equal-test-train procedure for the male interview speech*

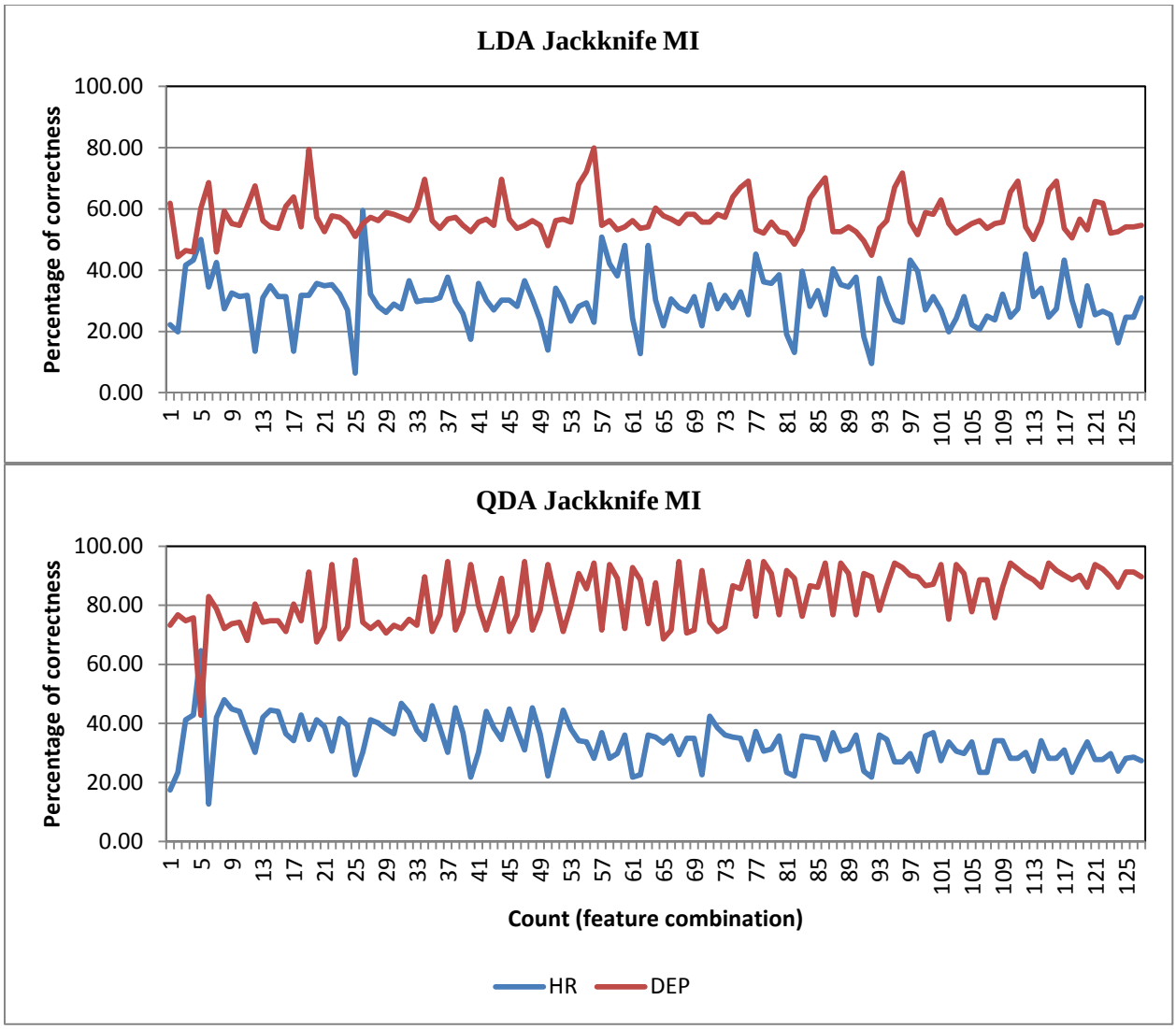


Figure 5.3: *Plot of the correctly classified high risk suicidal and depressed percentages over a number of feature combinations using the linear and quadratic classification with the jackknife procedure for the male interview speech*

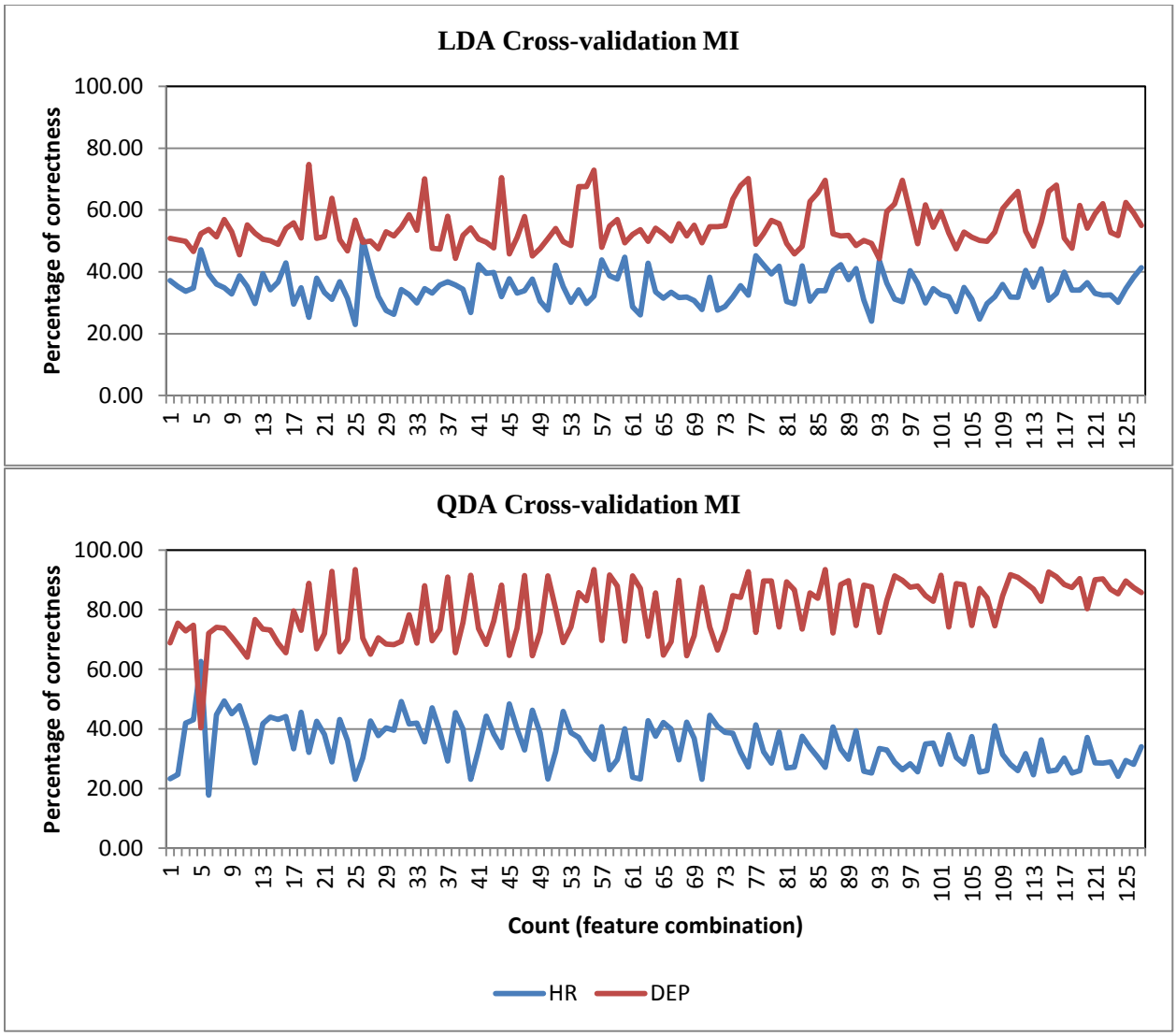


Figure 5.4: *Plot of the correctly classified high risk suicidal and depressed percentages over a number of feature combinations using the linear and quadratic classification with the cross-validation procedure for the male interview speech*

Figures 5.2 to 5.4 display detailed plots of the correctly classified high risk suicidal and depressed patients (in percentage) over all possible combinations of the seven features for male interview speech using the LDA and QDA with the equal-test-train, jackknife and cross-validation procedure, respectively. According to a rule of thumb for an adequate sample size, an appropriate number of samples per estimated feature are of the 5:1 ratio [19]. Therefore, for a number of 21 male interview patients, a maximum of four features will be considered adequate. Classification analysis for both classifiers and almost all feature combinations demonstrated a higher percentage of correctness for the depressed group compared with the high risk suicidal

group. However, these differences are observed to be larger for the analysis using the QDA classifier.

The classifiers were able to correctly identify the depressed patients with a percentage reaching up to 95%. On the other hand, poor performance classification of only about 20% to 55% of the high risk suicidal patients were classified correctly except for the 5th feature combination using QDA with jackknife and cross-validation procedure. Classification performance using the 5th feature combination identified about 65% of the high risk patients and only about 40% of the depressed patients. The majority of the feature combinations using the LDA with the equal-test-train and the cross-validation procedure yielded a small percentage difference between the high risk and the depressed group. However, only a slight number of patients were classified correctly.

Table 5.3 The selected classification result for male interview speech

Resampling method	All data percentage	High risk percentage	Depressed percentage	Feature combination	Classifier
Equal-test-train	67	54	84	Range, Variation, Skewness	QDA
Jackknife	58	60	55	Skewness, Kurtosis	LDA
Cross-validation	60	49	74	Maximum, Range	QDA

Table 5.3 displays the best results for male interview speech based on judgment analysis. Classification analysis using the QDA and the equal-test-train procedure with a combination of range, variation and skewness moderately classified 67% of the male interview speech. Approximately five out of 10 high risk patients and nine out of 11 depressed patients were identified correctly. For the jackknife procedure, using the LDA with a combination of skewness and kurtosis ineffectively classified 58% of the male interview speech. Approximately six out of 10 high risk patients and six out of 11 depressed patients were identified correctly. A moderate performance of 60% was demonstrated in the cross-validation procedure using the QDA with a combination of maximum and range for the male interview speech. Approximately five out of 10 high risk patients and eight out of 11 depressed patients were identified correctly. The selected result from the QDA with the equal-test-train and the cross-validation procedure demonstrated that the classifier was more effective in classifying the depressed group than the high risk with a

large percentage difference. Meanwhile, the LDA with the jackknife procedure almost evenly misclassified the two groups.

4.2.2 Male Reading Results

The all male reading data percentages using the LDA and the QDA display an almost equal range of correctly classified percentage in classifying the group of high risk and depressed for all resampling methods. However, the plots of the high risk and the depressed percentage shown in figures 5.5 to 5.7 reveal that the classifiers using all resampling methods were more effective in classifying the depressed patients. For 20 male reading recordings, a maximum of four features will be considered adequate.

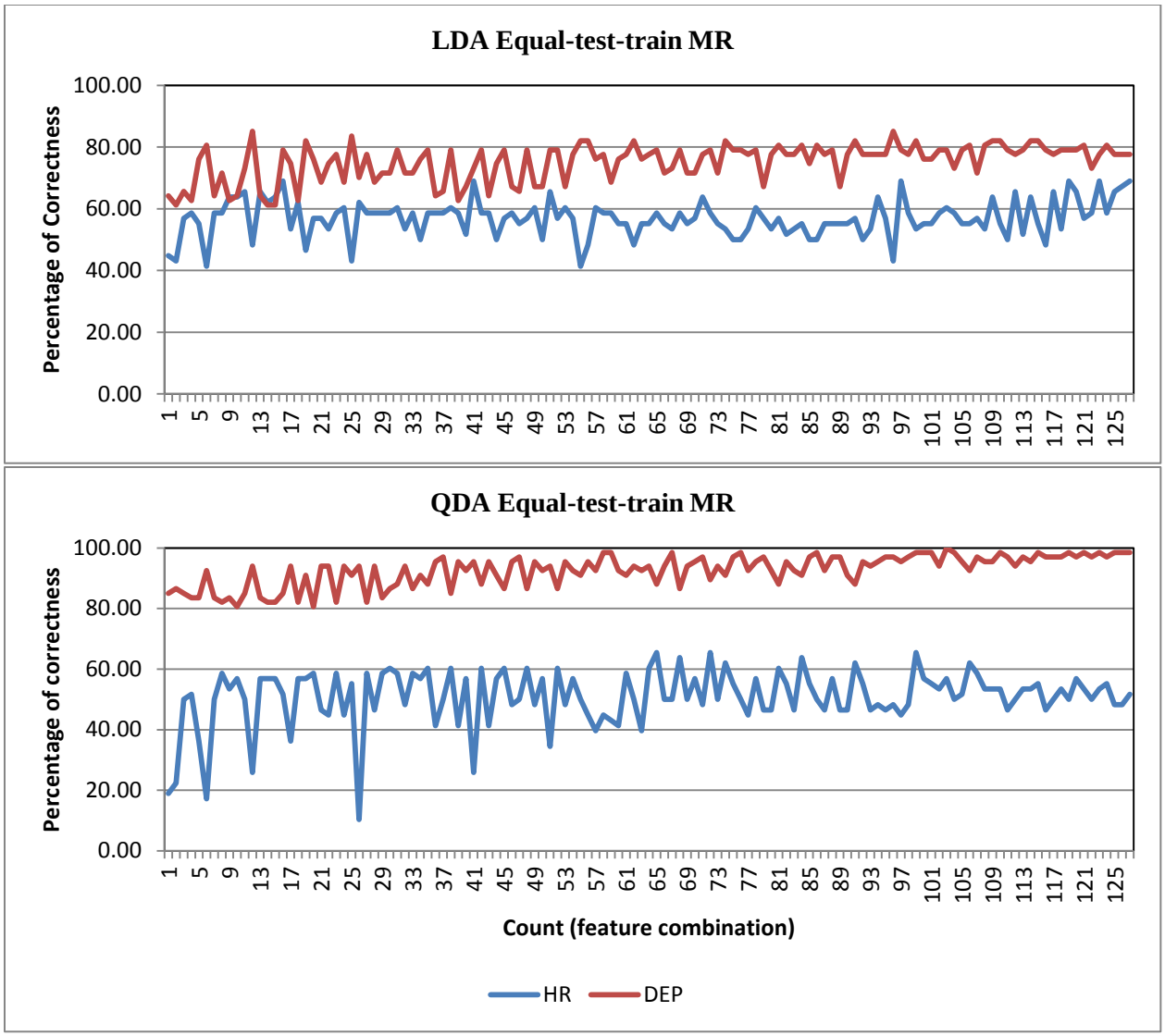


Figure 5.5: *Plot of the correctly classified high risk suicidal and depressed percentages over a number of feature combinations using the linear and quadratic classification with the equal-test-train procedure for the male reading speech*

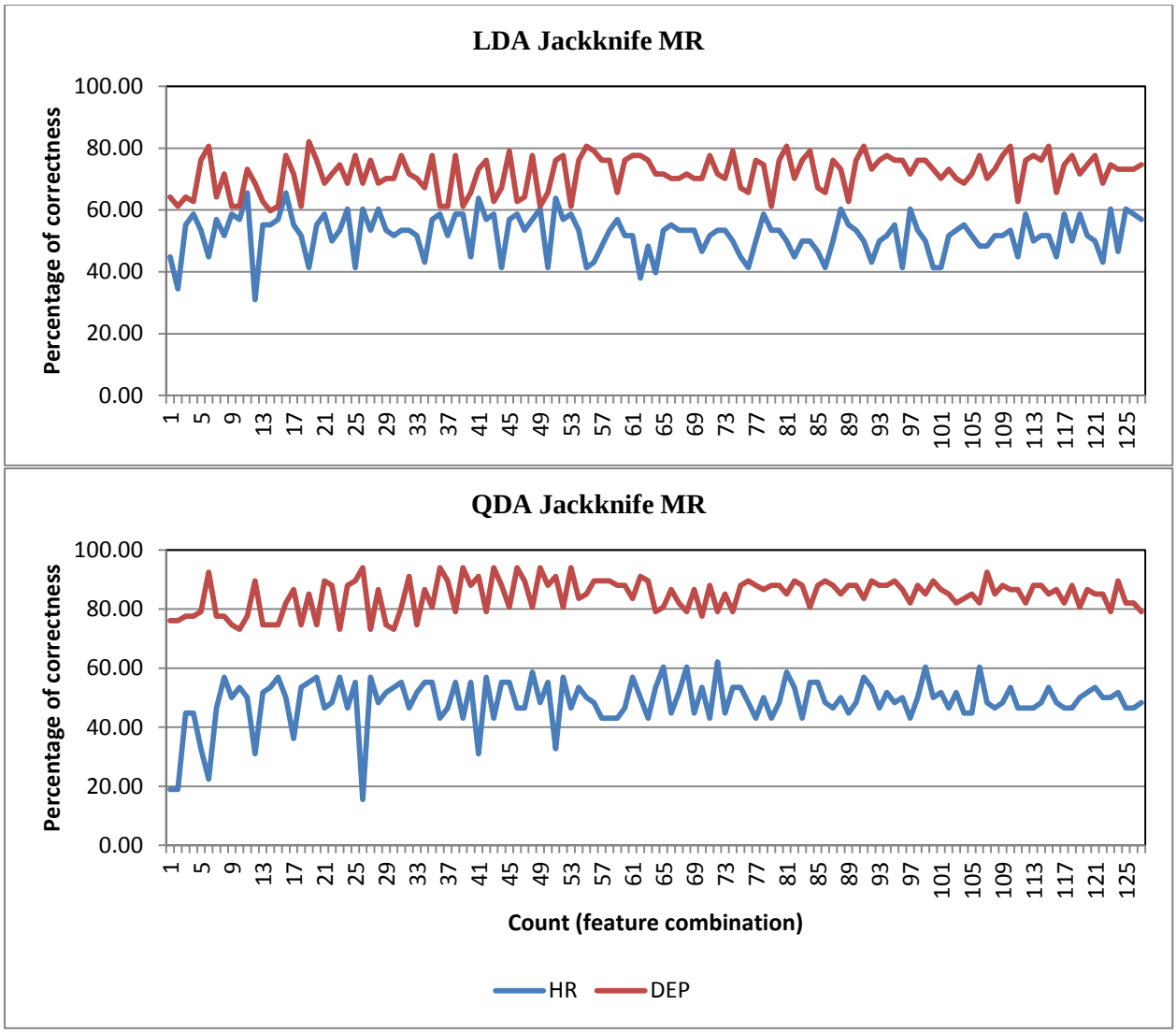


Figure 5.6: *Plot of the correctly classified high risk suicidal and depressed percentages over a number of feature combinations using the linear and quadratic classification with the jackknife procedure for the male reading speech*

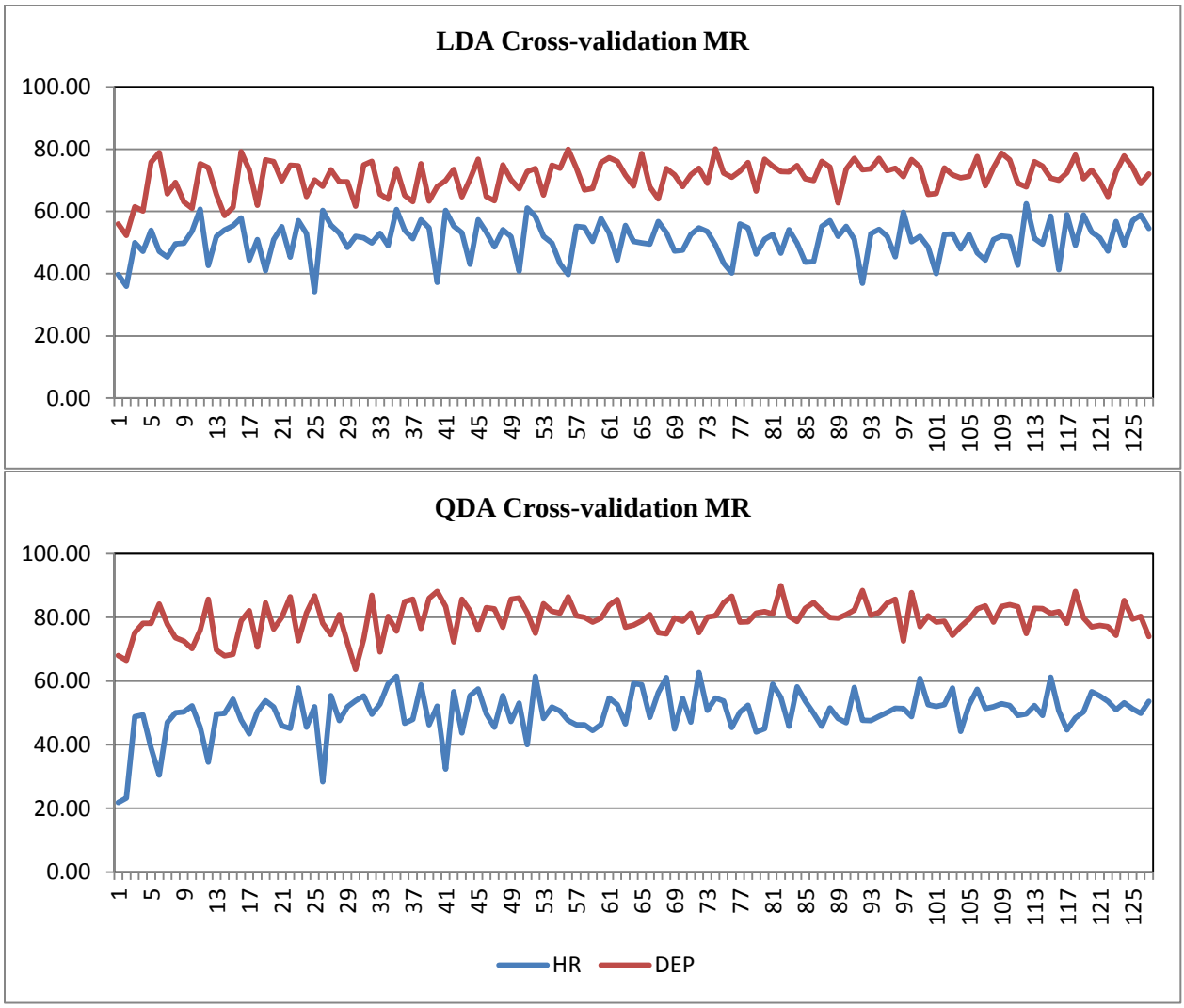


Figure 5.7: *Plot of the correctly classified high risk suicidal and depressed percentages over a number of feature combinations using the linear and quadratic classification with the cross-validation procedure for the male reading speech*

Comparing the two plots of LDA and QDA for all resampling procedures shown in figures 5.5 to 5.7, QDA demonstrated a larger difference between the classification percentage of the high risk patients and the depressed patients. The QDA plots display the classification percentage of the high risk patients ranging from approximately 20% to 60% and the percentage of depressed patients ranging from approximately 70% to 100%. The selected result from each resampling procedure is shown in table 5.4.

Table 5.4 The selected classification result for male reading speech

Resampling method	All data percentage	High risk percentage	Depressed percentage	Feature combination	Classifier
Equal-test-train	74	69	79	Range, Skewness	LDA
Jackknife	72	66	78	Range, Skewness	LDA
Cross-validation	67	61	75	Maximum, Skewness	LDA

Table 5.4 displays the selected results for male reading speech that were chosen based on the least number of feature combination with the highest high risk percentage. Classification analysis using the LDA and the equal-test-train procedure with a combination of RMS AM range and skewness effectively identified 74% of the male reading speech. An overall effective performance of 72% was reported for the jackknife procedure using the LDA with a combination of RMS AM range and skewness. Classification analysis using the LDA and cross-validation procedure with a combination of RMS AM maximum and skewness yielded an average performance of 67%. All methods correctly classified approximately five out of eight high risk patients and nine out of 12 depressed patients. The selected classifiers were more effective in classifying the depressed group compared with the high risk group.

4.2.3 Female Interview Results

All female interview data percentages demonstrated that the performance of the LDA slightly exceeds the QDA for the analysis using the equal-test-train with an average all data percentage of up to 60%. Classifiers using the jackknife and the cross-validation procedure demonstrated an evenly performance in classifying the two groups with an ineffective all data percentage of equal to or less than approximately 55%. The results indicate that not more than 19 out of 32 female interview patients were correctly identified. For a number of 32 female interview patients, a maximum of six features will be considered adequate. Figures 5.8 to 5.10 display detailed plots of the correctly classified high risk suicidal and depressed patients (in percentage) over all possible combinations of the seven features for female interview speech using the LDA and the QDA classifier with the equal-test-train, jackknife and cross-validation procedure, respectively.

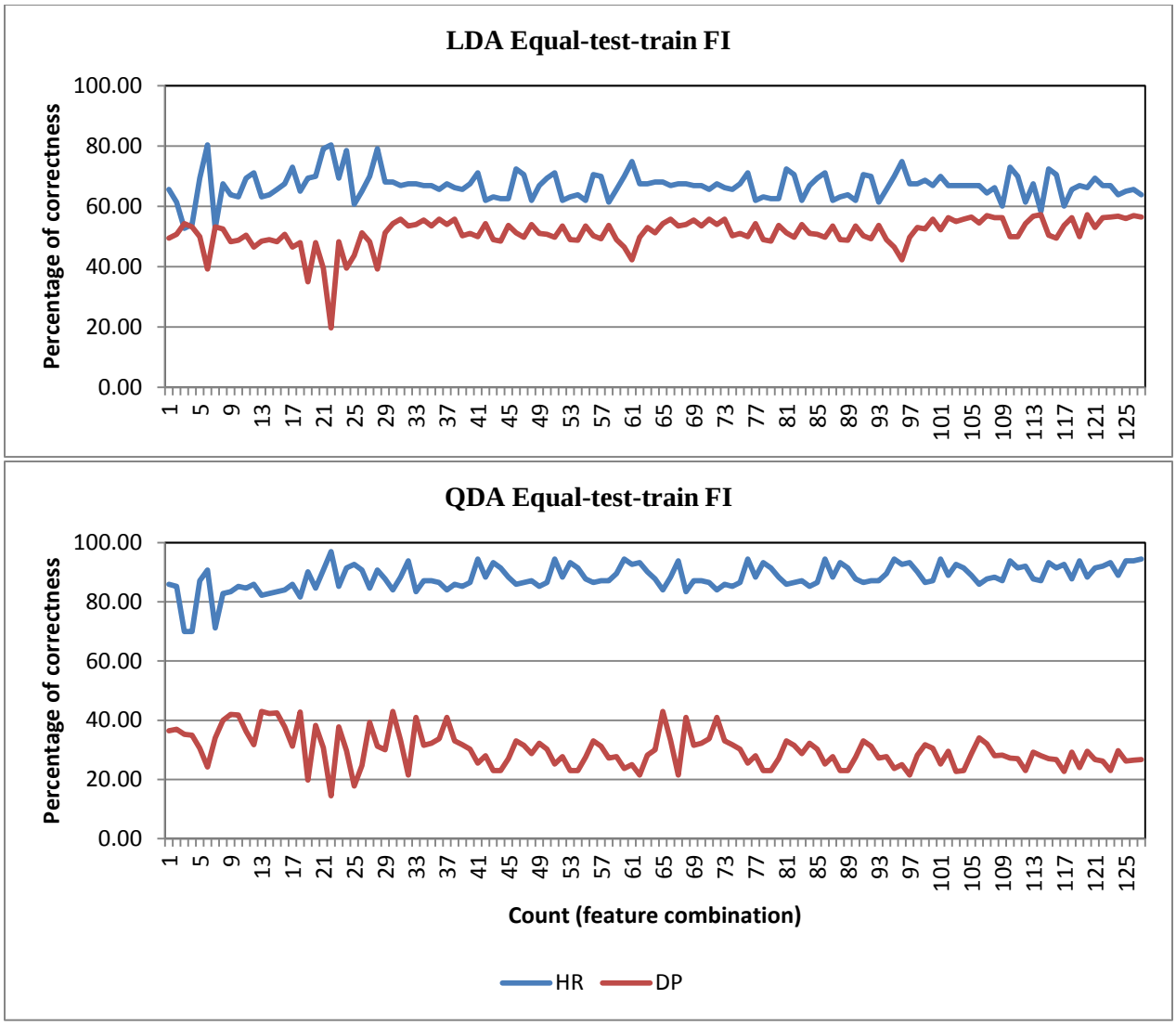


Figure 5.8: *Plot of the correctly classified high risk suicidal and depressed percentages over a number of feature combinations using the linear and quadratic classification with the equal-test-train procedure for the female interview speech*

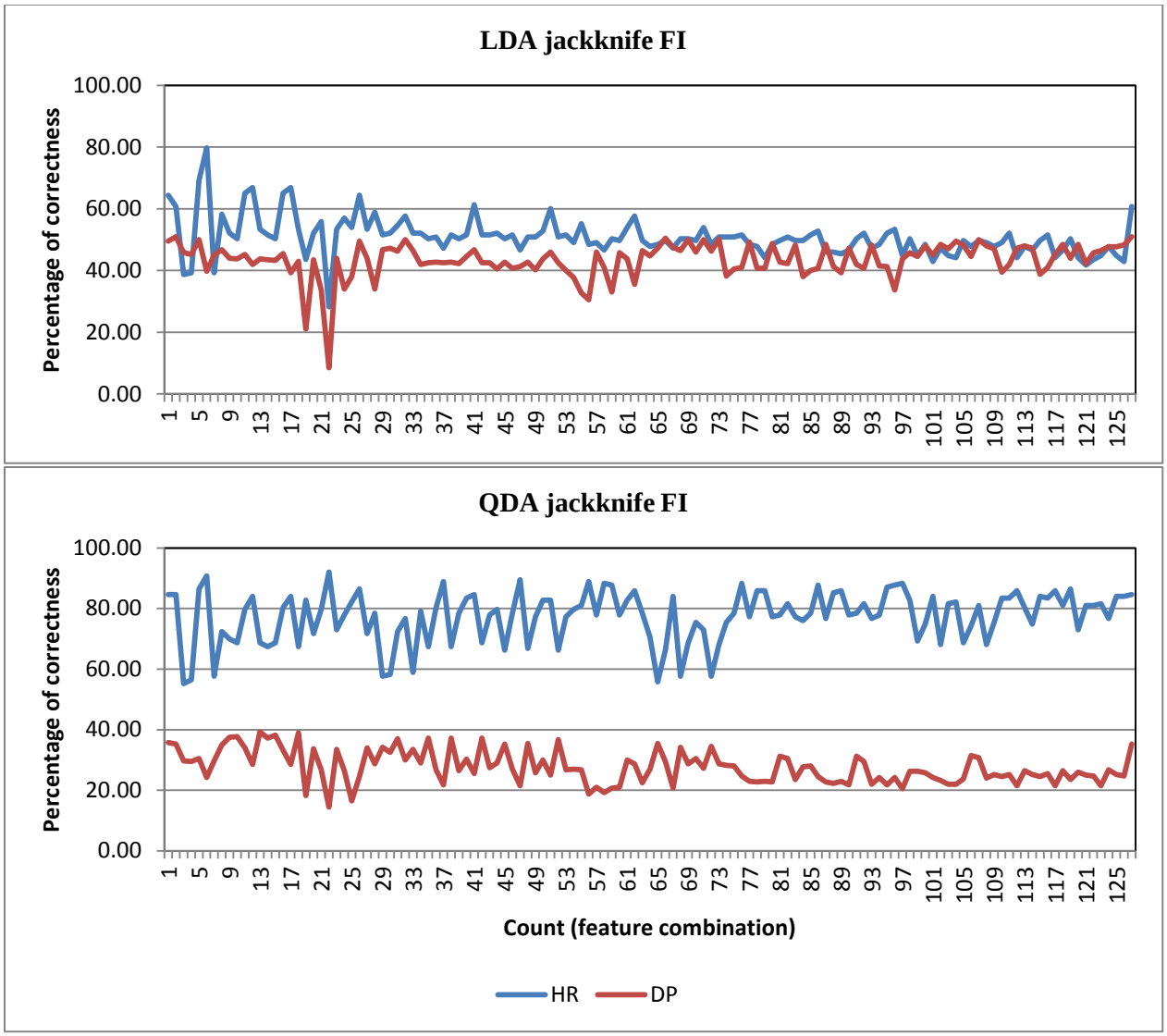


Figure 5.9: *Plot of the correctly classified high risk suicidal and depressed percentages over a number of feature combinations using the linear and quadratic classification with the jackknife procedure for the female interview speech*

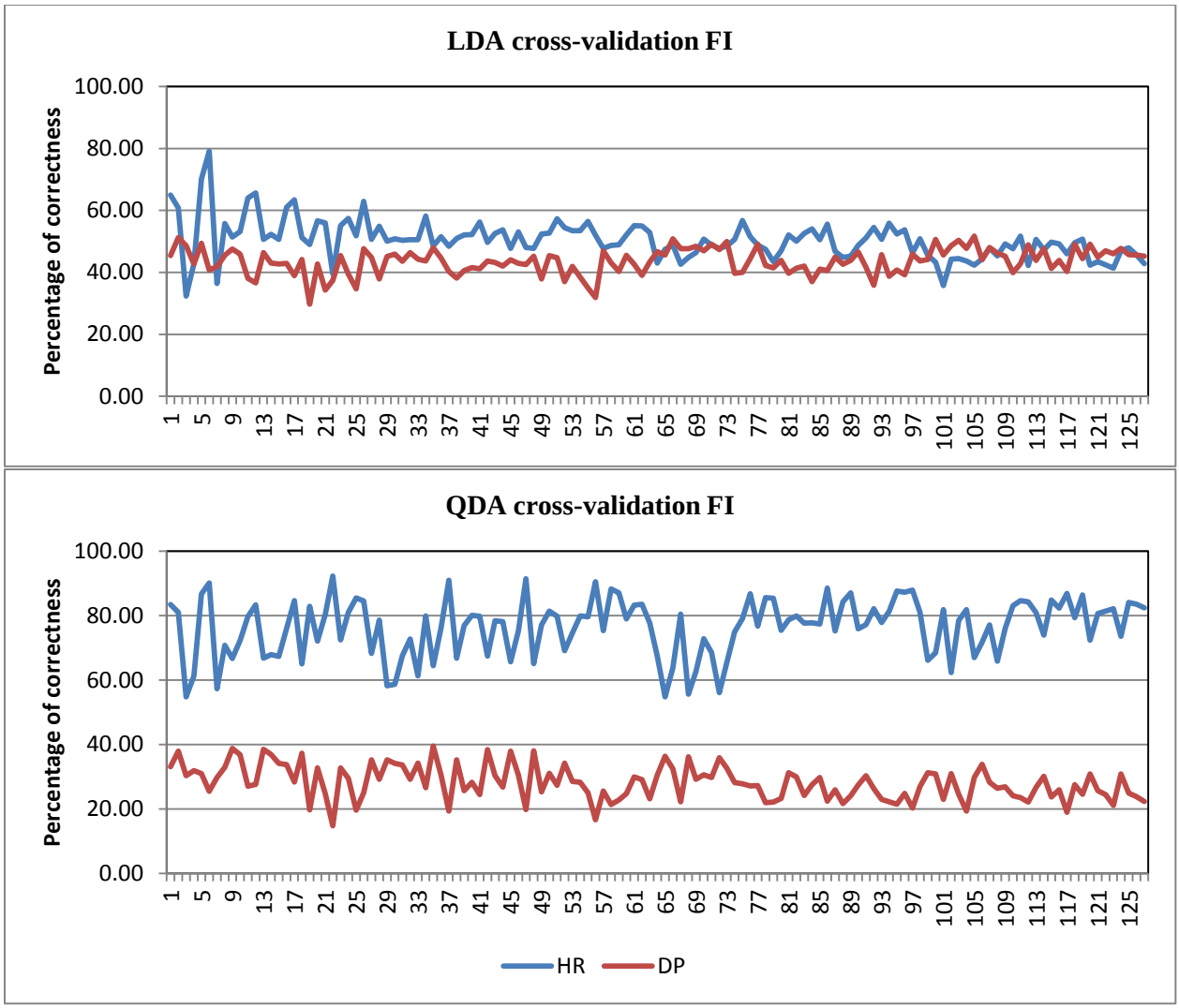


Figure 5.10: *Plot of the correctly classified high risk suicidal and depressed percentages over a number of feature combinations using the linear and quadratic classification with the cross-validation procedure for the female interview speech*

Based on figures 5.8 to 5.10, observing the plots of the high risk and the depressed percentages separately revealed that the classifiers performed more effectively in identifying the high risk suicidal group than the depressed group for female interview speech. However, classifications using the QDA with all resampling methods were observed to exhibit larger percentage differences between the two groups for all feature combinations. The QDA plots demonstrated percentages of correctly classified high risk speech ranging from approximately 60% to 95% while the percentages for the classified depressed speech ranged from

approximately 15% to 40%. The LDA plots revealed an almost equal classification performance between the two groups. Even though most percentages for the high risk group in the QDA are considered large, the error classification on the depressed group is considered too large, thus signifying the false alarm rate is too high. For LDA with the equal-test-train, jackknife and cross-validation procedure, the percentage of the classified high risk patients ranged from approximately 50% to 80%, 40% to 80% and 30% to 80% respectively whereas the percentage of the classified depressed patients ranged from approximately 20% to 50%, 10% to 50% and 30% to 50% respectively.

Table 5.5 The selected classification result for female interview speech

Resampling method	All data percentage	High risk percentage	Depressed percentage	Feature combination	Classifier
Equal-test-train	59	67	56	Maximum, Average, Skewness	LDA
Jackknife	56	69	50	Skewness	LDA
Cross-validation	58	70	49	Skewness	LDA

Table 5.5 displays the selected results for female interview speech that were chosen based on the highest (or near the highest value) all data percentage and with the smallest number of feature combination. Classification analysis using the LDA and the equal-test-train procedure with a combination of RMS AM maximum, average and skewness ineffectively identified 59% of the female interview subjects. Approximately eight out of 12 high risk patients and 11 out of 20 depressed patients were classified correctly. However, there are a few feature combinations that had similar performance but with only a slightly lower all data percentage. For the jackknife procedure, classification analysis using the LDA and the RMS AM skewness feature ineffectively identified 56% of the female interview subjects. For the cross-validation procedure, classification analysis using LDA and the RMS AM skewness feature ineffectively identified 58% of the female interview subjects. Approximately eight out of 12 high risk patients and 10 out of 20 depressed patients were classified correctly using the jackknife and cross-validation procedure.

4.2.4 Female Reading Results

The female reading data percentages produced by LDA ranges from 50% to 60% for the equal-test-train, 20% to 55% for the jackknife and 35% to 60% for the cross-validation procedure. On the other hand, the percentages produced by QDA ranges from 45% to 70% for the equal-test-train, 30% to 55% for the jackknife and 40% to 55% for the cross-validation procedure. These results demonstrated an almost equal classification performance by the LDA and QDA.

For 29 female interview subjects, a maximum of six features will be considered adequate. Figures 5.11 to 5.13 display detailed plots of the correctly classified high risk suicidal and depressed patients (in percentage) over all possible combinations of seven features for female reading speech using the LDA and the QDA with the equal-test-train, jackknife and cross-validation procedures, respectively.

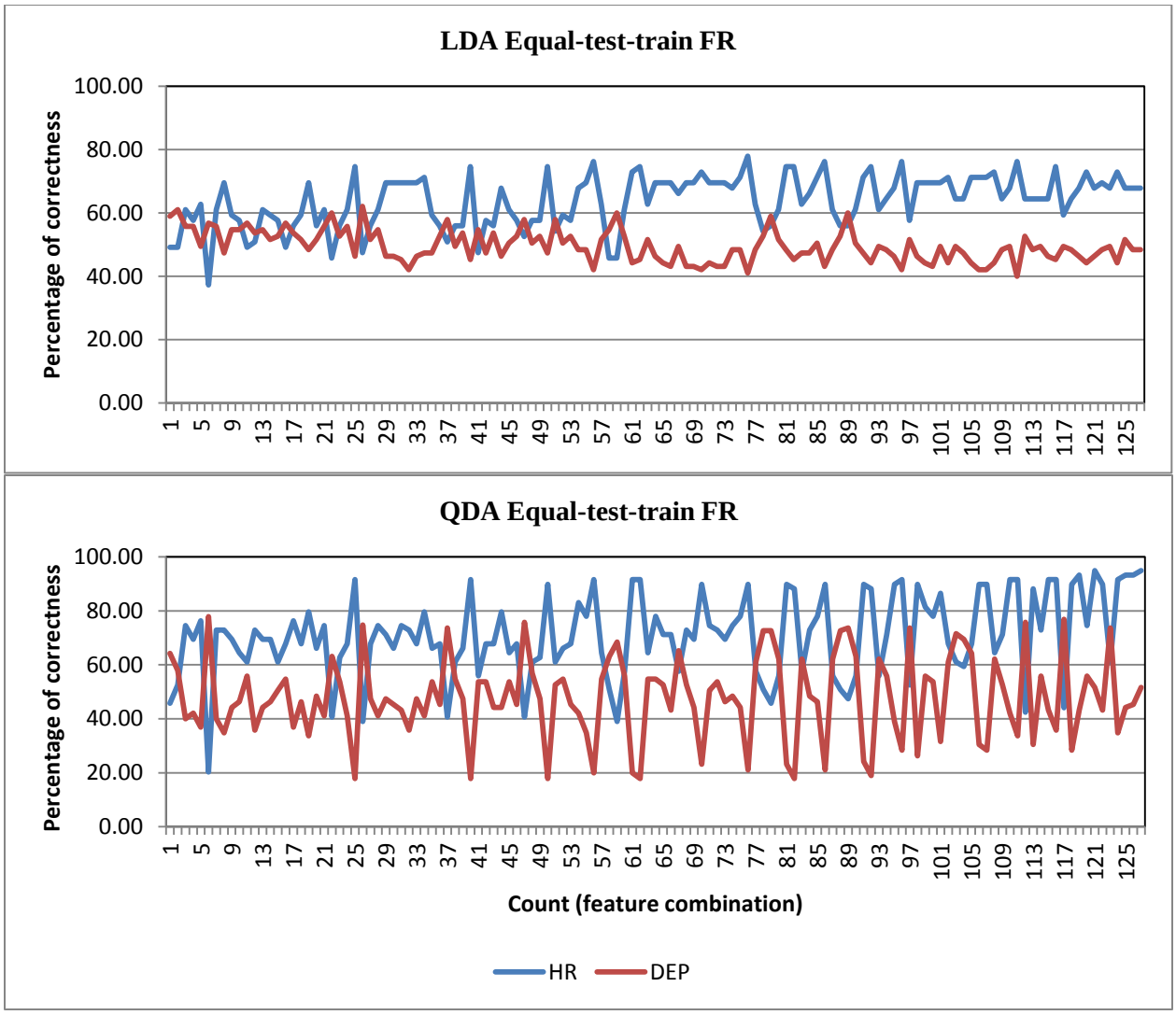


Figure 5.11: *Plot of the correctly classified high risk suicidal and depressed percentages over a number of feature combinations using the linear and quadratic classification with the equal-test-train procedure for the female reading speech*

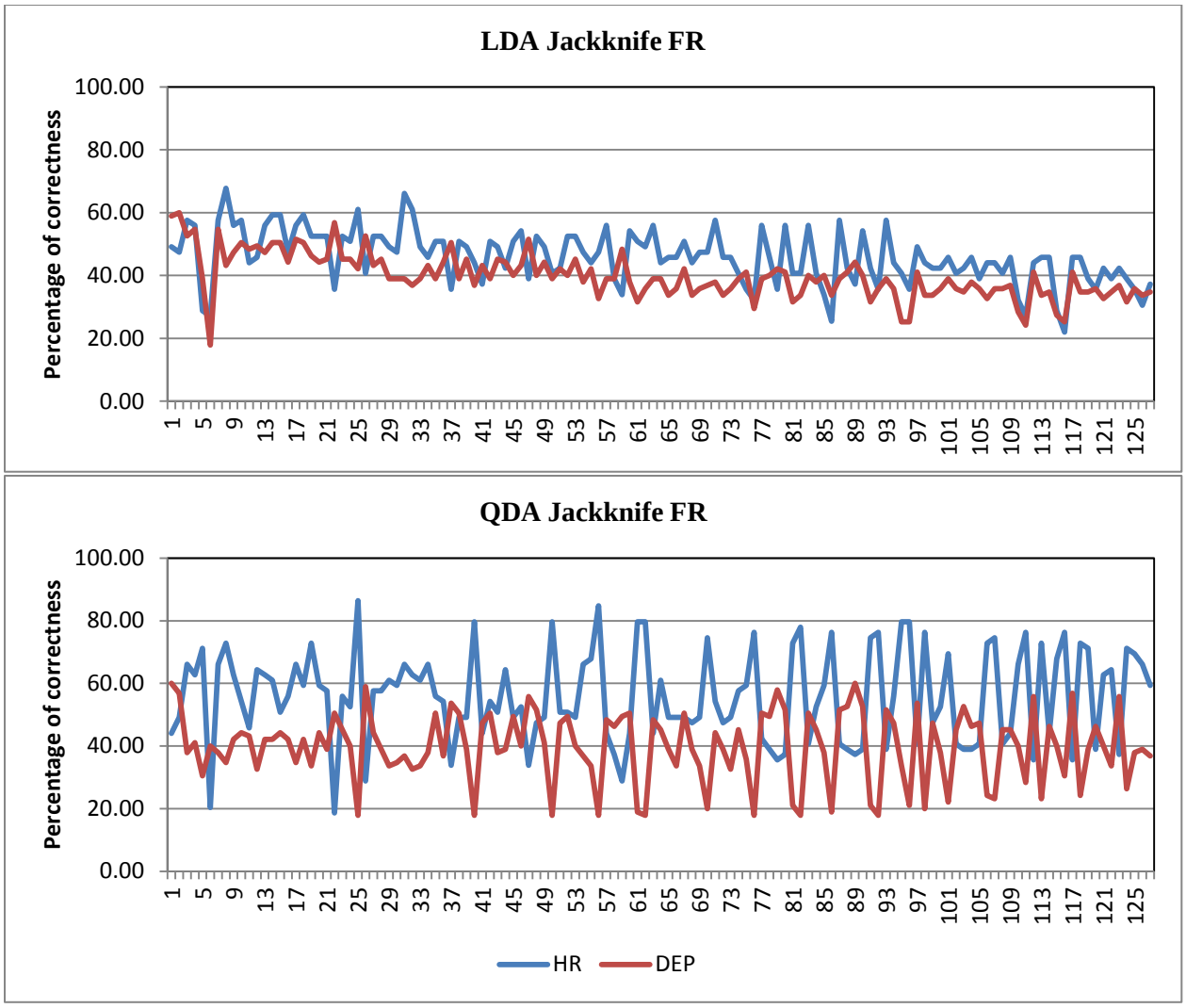


Figure 5.12: *Plot of the correctly classified high risk suicidal and depressed percentages over a number of feature combinations using the linear and quadratic classification with the jackknife procedure for the female reading speech*

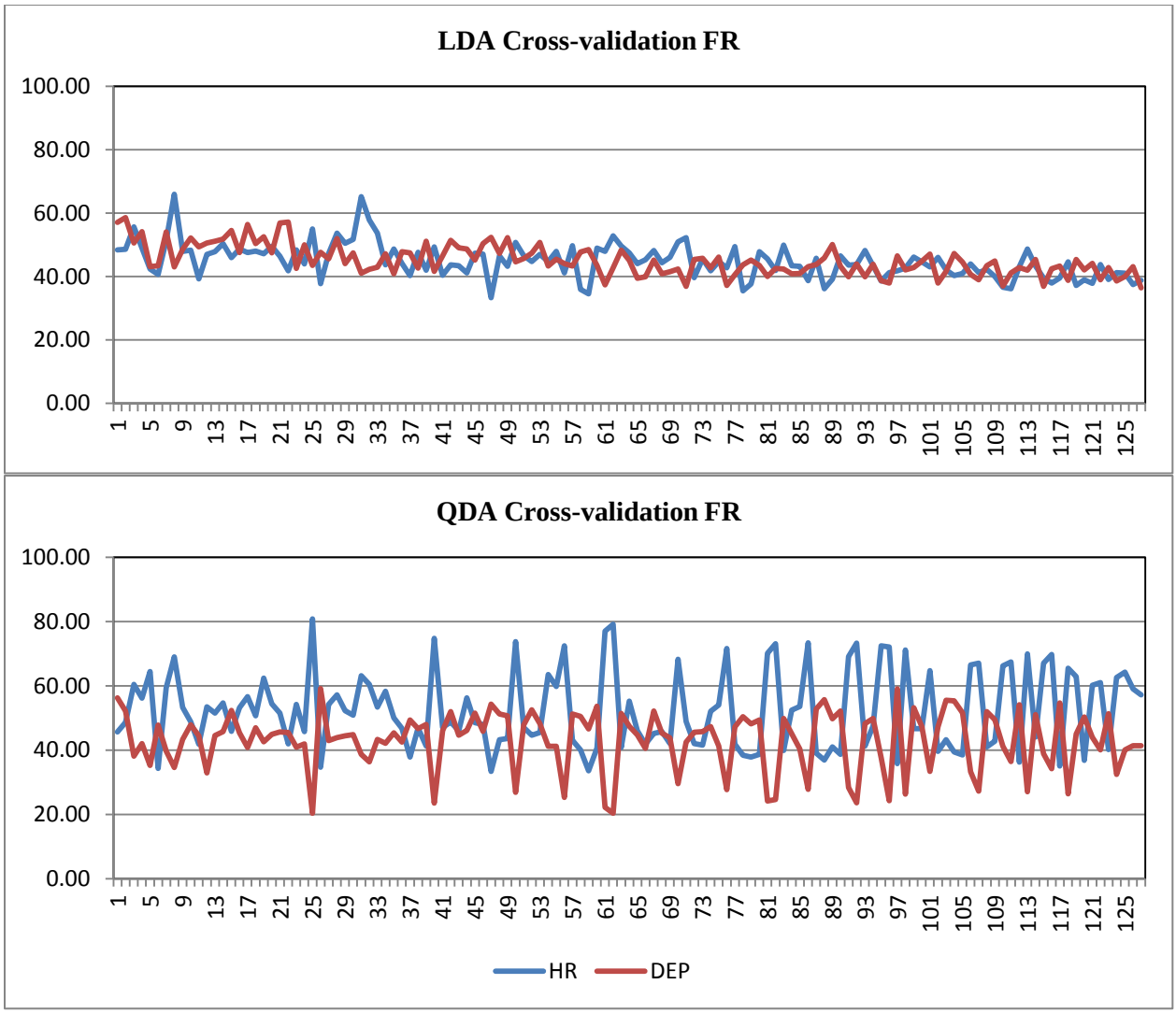


Figure 5.13: *Plot of the correctly classified high risk suicidal and depressed percentages over a number of feature combinations using the linear and quadratic classification with the cross-validation procedure for the female reading speech*

Similar to the female interview speech results, the performance of the LDA and the QDA in identifying the high risk group slightly exceeds the depressed group for all resampling procedures. Also, the large upward and downward spikes that are observed in the QDA plots demonstrated that the difference between the high risk percentage and the depressed percentage for a number of feature combinations are greater in the QDA as compared with the LDA.

Table 5.6 The selected classification result for female reading speech

Resampling method	All data percentage	High risk percentage	Depressed percentage	Feature combination	Classifier
Equal-test-train	68	61	72	Maximum, Range, Variance, Skewness, Kurtosis	QDA
Jackknife	56	58	55	Coefficient of variation	LDA
Cross-validation	53	56	51	Variance	LDA

Table 5.6 represents the selected classification results for female reading speech. Based on the second highest all data percentage for the quadratic classifier with the equal-test-train procedure, combination of RMS AM maximum, range, variance, skewness and kurtosis yielded seven out of 11 high risk patients and 13 out of 18 depressed patients that are correctly classified. The highest all data percentages are just slightly higher but it uses a combination of six features. The difference between this result and the result for the highest all data percentage is the lower 2% identified depressed patient and the additional feature (kurtosis) in the combination. Chosen based on the highest all data percentage, classification analysis with the LDA and jackknife procedure using the RMS AM coefficient of variation poorly identified 56% of the patients. Approximately six out of 11 high risk patients and 10 out of 18 depressed patients were classified correctly. However, a number of two feature combinations also yielded similar results with only a slightly lower percentage. Selected based on a high all data percentage with a balanced high risk and depressed percentages and smallest number of feature combination, classification with linear classifier and cross-validation procedure using the variance feature ineffectively identified 53% of the patients. Approximately six out of 11 high risk patients and 9 out of 18 depressed patients were classified correctly.

5.0 Discussion

For the purpose of analysis and reporting the results, the 'best' classifier and feature combination are selected on the basis of which classifier yielded the highest (or near the highest) percentage of all data classification with an almost equally balance percentage between the correctly classified high risk group and the depressed group. For example, one feature combination produced the highest all data percentage which is 59% but the correctly classified high risk is only 35% and the correctly classified depressed is 90%. Another feature combination produced a slightly lower all data percentage which is 57% but the correctly classified high risk is 57% and the correctly classified depressed is 57%. The latter would be chosen over the former due to the more balanced percentage between the high risk and the depressed.

The linear and quadratic classifier with all three resampling methods using speech from male patients demonstrated a higher performance in the depressed classification and a higher performance in identifying the high risk speech for female patients. However, results from this analysis suggest that there is no significant difference between the RMS AM speech features in the near term suicidal patients and the depressed patients that were collected from the male interview, female interview and female reading speech. The performance of the classifiers produced using the speech collected from the three groups mentioned above either exhibited an unbalance percentage of correctness or yielded a poor classification score for both high risk and depressed patients. However, the RMS AM speech feature might be able to distinguished between the near term suicidal and depressed male patients using the reading speech based on the results of the moderately balance classification score. The measure of the classification performance is evaluated based on the jackknife and cross-validation procedure. The equal-test-train procedure is a biased sampling with the testing data being a duplicate of the training data set.

For the male reading patients, the LDA using jackknife procedure with a combination of RMS AM range and skewness yielded a 72% overall classification score, 66% identified high risk and 78% identified depressed patients. This result partially agrees with the analysis reported by France [2]. He reported that the QDA using jackknife procedure with a combination of RMS AM range and coefficient of variation produced a 77% overall classification score, 77% identified high risk and 76% identified depressed for male speech. These recordings were obtained prior to a patient's suicide or suicide attempt. In this study, the classification

performance from male reading speech was better than the interview speech thus indicating that the RMS AM features extracted from the reading speech were more effective in classifying the two groups of high risk and the depressed compared with the interview speech. The current Databases were recorded with better recording devices (M Audio and TASCAM). On the other hand, France's study used speech samples from diverse recording environments and poor quality devices that were collected from therapy sessions, phone conversations, suicide notes and recordings provided by the Federal Bureau of Investigation. It is difficult to compare these two results. France's study had the advantage of ground truth database (the patients were known to have attempted suicide) and the current database has the advantage of a better recording device. His successful result on RMS AM may have been influenced by the differences in the recording devices and environment during the collection of the near-term suicidal speech. On the other hand, he stated that the recordings of the depressed patients were obtained using similar tape recording equipment, specifications and environments.

France also identified the important relationship between the depressed and the high risk suicidal speech with the characteristics of RMS AM range and skewness. However, the relationship did not reveal significant differences. In the current analysis, a similar outcome was observed. The RMS AM skewness appears to be the most frequent feature in the feature combination listed in tables 5.3 to 5.6 and then followed by the RMS AM range. Also, referring to table 5.2, the t-test performed on the male reading speech revealed that the RMS AM skewness feature exhibited the lowest p-value of 0.0169 followed by the RMS AM range from the male reading speech with a p-value of 0.0558. Both features show no significant difference at the level of 0.01, but the former suggests that there exist at least 95% confidence in the separation of the depressed and the high risk speech. In spite of the non-significant difference reported by the latter feature, the p-value for RMS AM range can still be considered close to the conventional level of 0.05 significance.

Literature in [20] – [22] related to the amplitude envelope revealed that this feature is essential for speech understanding and for providing perceptual information. However, these findings do not reject the possible contribution of amplitude envelope relating to speech discrimination. The poor to low moderate classification results may be due to the patient's ability to alter their speech. A study performed by Siegman and Boyle [23] reported that attenuating the way of speech in terms of speech rate and loudness can change the mood of the speaker,

independent of the speech content. Loudness can also be described by the amplitude envelope of speech and thus the intensity of the speech can either be amplified or attenuated using expressive vocal behavior and shown through the manifestation of cardiovascular, heart rate and blood pressure. The feelings of sadness and depression were associated with soft and slow speech. The blood pressure associated with the feeling of distress was altered when an event relating to sadness and depression were spoken in a normal and incompatible speech behavior. Another study by [24] demonstrated that a high percentage of syllables in a speech waveform were still recognizable in spite of removing all amplitude information by transforming them to either plus or minus one. Vocal intensity and loudness were also reported by Darby [25] to be highly uncorrelated with the depression. His findings on vocal intensity and loudness disagree with the outcome described by Ostwald [16] as 'the single most reliable criteria' of clinical improvement relating to depression. Therefore, the characteristic of monoloud and monotone in the depressed speech that were described through subjective analysis by trained physicians might be considered only suggestive in nature.

6.0 Conclusion

In summary, our findings discovered that the RMS AM range and skewness appear to be the distinctive properties of high risk suicidal and depressed speech in male reading patients but failed to identify any distinguishing characteristics when using male interview recordings. This result partially agrees with the findings by France where he reported a combination of RMS AM range and coefficient of variation in male speech as significant features. Our investigation of RMS AM as vocal correlates in the speech of high risk and depressed in female patients revealed poor classification performance for both interview and reading recordings. It is however difficult to conclude the different findings between this study and France's study because of the dissimilar database where he used recordings that were known to have attempted suicide whereas this study has the advantage of a better recording device.

References

[1] D. J. France, R. G. Shiavi, S. E. Silverman, M. K. Silverman, D. M. Wilkes, "Acoustical Properties of Speech as Indicators of Depression and Suicidal Risk", *IEEE Transaction on Biomedical Engineering*, vol. 47, no. 7, 2000.

[2] D. J. France, "Acoustical Properties of Speech as Indicators of Depression and Suicidal Risk", Ph.D, Thesis, Vanderbilt University, August, 1997.

[3] A. Ozdas, "Analysis of Paralinguistic Properties of Speech for Near-term Suicidal Risk Assessment", Ph.D, Thesis, Vanderbilt University, May, 2001.

[4] T. Yingthawornsuk, "Acoustic Analysis of Vocal Output Characteristics for Suicidal Risk Assessment", Ph.D, Thesis, Vanderbilt University, December, 2007.

[5] H. K. Keskinpala, "Analysis of Spectral Properties of Speech for Detecting Suicide Risk and Impact of Gender Specific Differences", PhD Thesis, Vanderbilt University, 2011.

[6] S. E. Silverman, "Method for Detecting Suicidal Predisposition" U.S patent 4 675 904, June 23, 1987.

[7] S. E. Silverman, "Method for Detecting Suicidal Predisposition" U.S patent 5 148 483, Sept 15, 1992.

[8] S. E. Silverman, "Method for Detecting Suicidal Predisposition" U.S patent 5 976 081, Nov 2, 1999.

[9] S. E. Silverman, "Method for Detecting Suicidal Predisposition" U.S patent 5 591 238, June 8, 2003.

[10] S. E. Silverman, M. K. Silverman, "Method and Apparatus for Evaluating Near-Term Suicidal Risk using Vocal Parameters" U.S patent 7 062 443, June 13, 2006.

[11] L. Campbell, "Statistical Characteristics of Fundamental Frequency Distributions in the Speech of Suicidal Patients", Master's Thesis, Vanderbilt University, 1995.

[12] W. A. S Wan Ahmad Hasan, "Acoustical Analysis of Speech Based on Power Spectral Density Features in Detecting Suicidal Risk among Female Patients", Master's Thesis, Vanderbilt University, 2011.

[13] N. H. Nik Nur Wahidah, "Analysis of Power Spectrum Density of Male Speech as Indicators for High Risk and Depressed Decision", Master's Thesis, Vanderbilt University, 2011.

[14] J. C. Mundt, A. P. Vogel, D. E. Feltner, W. R. Lenderking, "Vocal Acoustic Biomarkers of Depression Severity and Treatment Response", *Journal of Biological Psychiatry*, vol. 72, 580-587, 2012.

[15] E. Kraepelin, "Manic depressive insanity and paranoia", pp. 38, Livingstone, Edinburgh, 1921.

[16] P. F. Ostwald, "The Sounds of Emotional Disturbance", *Arch Gen Psychiatry*, vol. 5, no. 6, 1961.

[17] R. M. Salomon, H. K. Keskinpala, M. H. Sanchez, T. Yingthawornsuk, N. H. Nik Nur Wahidah, W. A. S. Wan Ahmad Hasan, N. Taneja, D. Vergyri, B. H. Knoth, P. E. Garcia, D. M. Wilkes, R. Shiavi, "Analysis of Voice Speech Indicators in Suicidal Patients", Manuscript submitted for publication, 2012.

[18] 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 in press)

[19] H. M. Kalayeh, and D. A. Landgrebe, "Predicting the Required Number of Training Samples", *Pattern analysis and machine intelligence, IEEE transaction*, pp. 664-667, 1983.

[20] R. V. Shannon, Fan-Gang Zeng and J. Wygonski, "Title of the Speech Recognition using Only Temporal Cues", from the book of The auditory processing of speech: From sounds to words, edited by MEH Schouten, 1992.

[21] J. Kubanek, P. Brunner, A. Gunduz, D. Poeppel and G. Schalk, "The Tracking of Speech Envelope in the Human Cortex", PLoS ONE, vol. 8, no. 1, e53398, 2013.

[22] H. Riquimaroux, "The Extent to Which Changes in the Amplitude Envelope Can Carry Information for Perception of Vocal Sound without the Fundamental Frequency or Formant Peaks", *Dynamics of speech production and perception*, 2006.

[23] AW Siegman and S Boyle, Voices of fear and anxiety and sadness and depression: The effects of speech rate and loudness on fear and anxiety and sadness and depression, Journal of Abnormal Psychology, Vol 102, No. 3, pp. 430-437, 1993.

[24] J. C. R. Licklider, I Pollack, "Effects of Differentiation, Integration and Infinite Peak Clipping on the Intelligibility of Speech", *J Acoust. Soc. Am*, vol. 20, pp. 42-51, 1948.

[25] J. K. Darby, and H. Hollien, "Vocal and Speech Patterns of Depressive Patients", *Folia. Phoniat.*, vol. 29, pp. 279-291, 1977.

Chapter VI

COMPARISON OF THE SIGNIFICANT MEAN DIFFERENCE FOR DIFFERENT SPECTRAL ENERGY BAND RANGES AND COMBINATIONS

Abstract

This small-scale study attempts to address two research questions. The first question asks whether the Power Spectral Density (PSD) feature is capable of significantly discriminating between the high risk suicidal (HR) patients and depressed (DP) patients. Second, the analysis tries to address the question of whether there is a statistically significant change in patients' vocal characteristics as they progress from the first recording session (labeled as high risk suicidal) to the next recording session (condition unknown to the researcher) that was collected a few days after receiving treatments. Within the rules of adequate number of features, multiple combinations of 8 PSD bands yielded statistically significant difference between the mean of HR and DP female patients. However, only probably statistical significant differences were observed between the HR and DP male patients. For the second analysis, patients' condition improved significantly during the second recording session as demonstrated by feature 4 PSD band 1 in male patients and features combination of 4 PSD band 2:3 in female patients.

1.0 Introduction

The assessment and distinguishing patient with imminent suicidal risk is considered to be one of the most complex and difficult task in psychiatry. Determining the patient's level of suicidal is crucial in order to select the appropriate treatment and safety management. Standardize predictors that were derived from classic clinical studies and other studies reported behavioral correlation in suicide. This comprehensive information is collected on the basis of organizing the risk factors for suicide. These factors are related to psychological milieu (life events, environmental factors and medical illnesses), presence of psychiatric disorders, biological factors, health records, family history and history of previous suicide attempts, if exists [1].

Patients are usually given antianxiety, antipsychotic and sleep medications. Apart from that, they are segregated from seclusion and reduce anxiety by encouraging them to get involve in psychosocial interventions such as group therapy and socializing with other patients. Traditionally, treatment for suicidal patient mainly focused on the psychiatric disorder that dominates the diagnosis. For example, suicidal patient with major depression is treated for their depressive disorder and treatment for psychotic patients with the presence of suicidal behavior is directed at diminishing the expressions of the psychosis [2].

The psychological state of a patient is expected to get better after receiving treatments. In certain cases, patient's suicidal condition improves rapidly after admission but in most cases, it often improves at a gradual pace. Development in the sense that the patient is no longer considered being acute high risk suicidal even though the condition still prolonged after the treatment. There are no laboratory tests or sophisticated diagnostic instruments that are available for psychiatrist to use to investigate whether the given treatments and medications have a positive effect on the patients. In addition, current environment imposes difficult challenges on psychiatrist. Patients that are facing severe risk of suicidal are admitted into inpatient unit for a short length of stay where both treatment time and duration are made limited by insurance requirements [3]. Due to this limitations and requirements, discharged patients that have gone through short-term treatments may experience an illusion effect of the medication where they seem to look better but in reality, may still be at an imminent risk of committing suicide. Because of the complexity, erroneous improvement is merely impossible to distinguish from the real improvement.

A number of researches have investigated speaking behaviors in patients with psychiatric disorders during the time of admission into the hospital and after improvement [11]-[17]. An early longitudinal study performed by Darby [22] reported that reduction in depression severity was not observed in voice and speech changes. A longitudinal study performed by [18]-[21] investigated the change in depression severity over the course of therapy by observing changes in speech acoustic measurements. Prominent features of speaking behavior and voice characteristics were demonstrated to correlate significantly with the time course recovery from depression. These features comprise of speech rate, pause duration, vocal timing, fundamental frequency, energy and pitch variability.

It has been reported that Power Spectral Density (PSD) feature appeared to be a distinguishing vocal feature when discriminating between suicidal and major depressed patients [4], [6]-[9]. The aim of this small-scale study was not to get a clear separation between two pairwise groups of high risk suicidal, depressed or remitted. The analyses were divided into two sections that were performed separately on Database A and Database B for both male and female patients using their interview and reading speech. Using the PSD feature analysis, the first objective of this study is to measure the significant difference between HR and DP for patients from Database A. The second objective is to measure the significance of patient's progression through different recording sessions and analyze the performance of the separating hyperplane in measuring severity from Database B.

The spectral density feature collected from voice speech act as an indicator that tracks the 'movement' of the patient's mental condition. This analysis allows researchers to observe the progression of patients through different recording sessions and analyze the performance of the separating hyperplane in measuring severity. A vector that is normal to the hyplerplane acts as a scale that shows the direction of the population. For example, the population is showing improvement in their psychological state if the high risk population is in the direction of either positive or negative along the line that is normal to the hyperplane and the 'unknown' population also moves further away in the same direction.

2.0 Database

Details on the database collection and pre-processing were explained in Chapter IV under section 3.0. However, for this study, only patients with three recording sessions were used for Database B. Table 6.1 displays the information on the databases used for this analysis.

Table 6.1 Information on the databases

Database A	Male				Female			
	HR		DP		HR		DP	
	Int*	Read	Int	Read	Int	Read	Int	Read
Total number of patients	9	8	11	12	12	10	20	18
Database B	Male				Female			
	Int		Read		Int		Read	
Number of patients with three recordings	5		4		8		7	

3.0 Methodology

3.1 Distance Measurement from the Separating Hyperplane

The decision hypersurface in the the *l*-dimensional space is a hyperplane that is represented by

$$g_{12}(x) = w^T x + w_o$$

where,

$$w = \sum^{-1}(\mu_1 - \mu_2)$$

$$w_o = \frac{1}{2}(\mu_2{}^T \sum^{-1} \mu_2 - \mu_2{}^T \sum^{-1} \mu_1)$$

w is known as the weight vector and w_o as the threshold. On one side of the hyperplane, $g_{12}(x) > 0(+)$ for vectors that are classified as class 1 takes the positive values and $g_{12}(x) < 0(-)$ for class 2 takes the negative values [23]. A geometry illustration of the decision hyperplane is shown in figure 6.1. The normalize Euclidean distance from each vector to the decision hyperplane can be represented by,

$$GNorm = \frac{g_{12}(x)}{\sqrt{w_1{}^2 + w_2{}^2}}$$

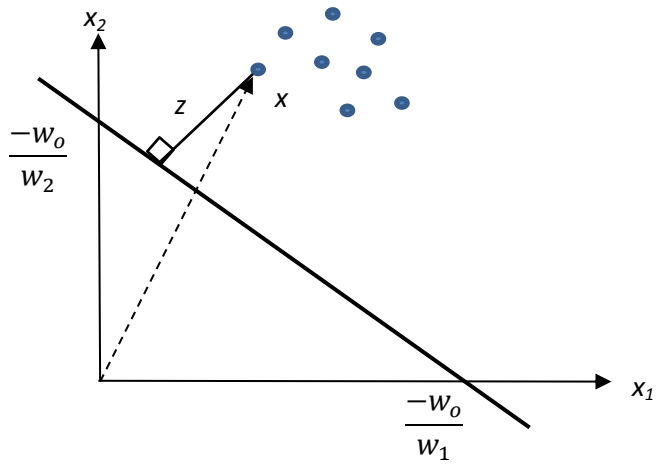


Figure 6.1: *Geometry of the decision hyperplane[76]*

3.2 Feature Extraction

The analysis was performed separately on Database A and Database B for both male and female patients using their interview and reading speech samples. Only the voiced speech segments were collected using the voice/unvoiced/silence detection [5] and the collected voiced signals were then split into 20-second segments. Energy distribution for four equal 500Hz bands, six equal 333Hz bands and eight equal 250Hz bands were extracted from each 20 second segments as demonstrated in [8, 9]. Patients will have varying numbers of 20-second segments depending on the length of the recordings. Finally, we measured the distance for each vector of the 20-second segment to the separating hyperplane and represent the measured distance as GNorm.

3.3 Analysis of Significant Measures

Database A

Median is considered to be a better measure of the center for representing an aggregate value for the dependent vectors per patients due to the skewed-shape data shown by the large standard deviation. The statistical analysis was then computed using 'ranksum' in Matlab which uses the Mann-Whitney U tests that compares two-independent population medians. It is a non-parametric test that does not require any specific form for the distribution of the population particularly for small sample based on the principle of ranking observations with the lowest value receives rank 1. The significance was measured on all combinations of 0-1500Hz (4 bands), 0-1666Hz (6bands) and 0-1750Hz (8bands).

Database B

Two-sample statistical analysis was performed separately on the estimated GNorm distance to identify the statistical significance between HR(session 1)-'others'(session 2), HR(session 1)-'others'(session 3) and 'others'(session 2)-'others'(session 3). The analysis was done on all combinations of 0-1500Hz (4 bands) for both interview and reading samples. Each patient has three recording sessions and each recording session have varying number of 20 second segments. All patients were identified to be in the high risk suicidal during session 1. After receiving treatment, medication and being hospitalized for a certain short period of time, patients continued for their second session and had their last session before getting discharged. In

order to keep the same number of observations across sessions for every patient, the mean of the estimated GNorm for each session was calculated. The significance of the group means and standard deviation were obtained using a paired two sample for means t-test analysis with one tailed p-value due to the expectation of one mean to be larger than the other.

4.0 Results

4.1 Results for the Comparison of Significance between Group of HR and DP for Database A

According to a rule of thumb for an adequate sample size, an appropriate number of samples per estimated feature are of the 5:1 ratio [24]. Therefore, for male patients' database, a maximum of four features for 20 interview recordings and four features for 20 reading recordings are considered adequate. On the other hand, for female patients' database, a maximum of six features for 32 interview recordings and five features for 28 reading recordings are suitable to satisfy the rule of generalization.

Table 6.2 displays results of the Mann-Whitney U test that was measured on all possible combinations of 4PSD band for male interview and reading and female interview and reading database. Male interview band 1:3, female interview band 2 and female reading band 2:3 are probably significantly different with a p-value less than 0.05 but larger than 0.01 ($0.05 <$ significance < 0.01). The rest of band combinations show no significant difference between the mean population of HR and DP given that their significance are larger than 0.05.

Table 6.2 Comparison of the independent two-tailed significance p-value for measuring the mean difference for all possible combinations of four PSD bands

Significance (*p*)	**Category**							
	Male Interview	**h**	**Female Interview**	**h**	**Male Reading**	**h**	**Female Reading**	**h**
4 PSD band 1	0.3233	0	0.0645	0	0.0587	0	0.0506	0
4 PSD band 2	0.2545	0	0.0450	1	0.0698	0	0.0506	0
4 PSD band 3	0.8197	0	0.1340	0	0.9079	0	0.3568	0
4 PSD band 1 to 2	0.7040	0	0.0703	0	0.0698	0	0.0506	0
4 PSD band 1 to 3	0.0334	1	0.0590	0	0.1137	0	0.0506	0
4 PSD band 2 to 3	0.5949	0	0.0765	0	0.0972	0	0.0366	1
4 PSD band 1 and 3	0.5433	0	0.0765	0	0.1137	0	0.0506	0

Table 6.3 shows results of the Mann-Whitney U test measured on all possible combinations of 6PSD band 1:5 for male and female interview and reading database. Based on the lowest number of band combination within an adequate feature size, the mean difference between HR and DP is probably significant for female interview 6PSD band 3, male reading 6PSD band 1,2,3,5 and female reading 6PSD band 1,2 with a p-value of 0.0121, 0.0339 and 0.0482 respectively. The male interview speech did not display any significance difference between the two population mean of HR and DP (p-value > 0.05).

Table 6.3 Comparison of the independent two-tailed significance p-value for measuring the mean difference for all possible combinations of six PSD bands

Significance (*p*)	Category							
	Male Interview	h	Female Interview	h	Male Reading	h	Female Reading	h
6 PSD band 1	0.1106	0	0.1922	0	0.0972	0	0.7419	0
6 PSD band 2	0.2545	0	0.2351	0	0.1770	0	0.8508	0
6 PSD band 3	0.8792	0	0.0121	1	0.4179	0	0.0666	0
6 PSD band 4	0.5433	0	0.1670	0	0.7285	0	0.4804	0
6 PSD band 5	1.0000	0	0.1149	0	0.7871	0	0.2589	0
6 PSD band 1.2	0.2875	0	0.0540	0	0.2030	0	0.0482	1
6 PSD band 1,3	0.1106	0	0.0134	1	0.1770	0	0.0739	0
6 PSD band 1,4	0.1965	0	0.1670	0	0.2633	0	0.4239	0
6 PSD band 1,5	0.2875	0	0.1793	0	0.0826	0	0.2793	0
6 PSD band 2,3	0.1715	0	0.0150	1	0.1535	0	0.0666	0
6 PSD band 2,4	0.2875	0	0.2351	0	0.2633	0	0.5724	0
6 PSD band 2,5	0.3233	0	0.2351	0	0.2633	0	0.2041	0
6 PSD band 3,4	0.3619	0	0.0279	1	0.5120	0	0.0997	0
6 PSD band 3,5	0.8197	0	0.0339	1	0.3749	0	0.0431	1
6 PSD band 4,5	0.8197	0	0.0590	0	0.4179	0	0.7067	0
6 PSD band 1,2,3	0.2875	0	0.0410	1	0.2030	0	0.0482	1
6 PSD band 1,2,4	0.1715	0	0.0307	1	0.1535	0	0.0599	0
6 PSD band 1,2,5	0.2545	0	0.0185	1	0.1137	0	0.0482	1
6 PSD band 1,3,4	0.1715	0	0.0307	1	0.2976	0	0.0904	0
6 PSD band 1,3,5	0.2875	0	0.0373	1	0.1535	0	0.0385	1
6 PSD band 1,4,5	0.1965	0	0.1793	0	0.1770	0	0.7067	0
6 PSD band 2,3,4	0.1715	0	0.0373	1	0.3348	0	0.0818	0
6 PSD band 2,3,5	0.2545	0	0.0373	1	0.1325	0	0.0538	0
6 PSD band 2,4,5	0.4033	0	0.2201	0	0.2318	0	0.7067	0
6 PSD band 3,4,5	0.3619	0	0.0307	1	0.4179	0	0.0997	0
6 PSD band 1,2,3,4	0.1286	0	0.0307	1	0.2318	0	0.0385	1
6 PSD band 1,2,3,5	0.1715	0	0.0373	1	0.0339	1	0.0739	0
6 PSD band 1,2,4,5	0.1489	0	0.0252	1	0.2633	0	0.0599	0
6 PSD band 1,3,4,5	0.1489	0	0.0373	1	0.2976	0	0.0818	0

6 PSD band 2,3,4,5	0.1715	0	0.0373	1	0.2318	0	0.0818	0
6 PSD band 1 to 5	0.1965	0	0.0279	1	0.0030	1	0.0739	0

Table 6.4 shows results of the Mann-Whitney U test measured on all possible combinations of 8PSD band 1:7 for male and female interview and reading database. Based on the lowest number of band combination within an adequate feature size and the smallest significant value, the mean difference between HR and DP is probably significantly different for male interview 8PSD band 1,6,7 and male reading 8PSD band 1,3 with a p-value of 0.0276 and 0.0409 respectively. On the contrary, significant difference (p-value < 0.01) was demonstrated between the mean of HR and DP in female interview 8PSD band 2,6,7 and female reading 8PSD band 1,2 with a p-value of 0.0077 and 0.0037 respectively.

Table 6.4 Comparison of the independent two-tailed significant p-value for measuring the mean difference for all possible combinations of eight PSD bands

Significance (*p*)	**Category**							
	Male Interview	**h**	**Female Interview**	**h**	**Male Reading**	**h**	**Female Reading**	**h**
8 PSD band 1	0.1965	0	0.1340	0	0.1137	0	0.8397	0
8 PSD band 2	0.3233	0	0.5463	0	0.7285	0	0.1009	0
8 PSD band 3	0.1965	0	0.0540	0	0.0698	0	0.0687	0
8 PSD band 4	1.0000	0	0.0339	1	0.5628	0	0.1321	0
8 PSD band 5	0.5433	0	0.1149	0	0.9692	0	0.4860	0
8 PSD band 6	0.8197	0	0.3603	0	0.6713	0	0.4860	0
8 PSD band 7	0.7040	0	0.0410	1	0.8471	0	0.2164	0
8 PSD band 1,2	0.2241	0	0.0167	1	0.0698	0	0.0037	1
8 PSD band 1,3	0.2241	0	0.0228	1	0.0409	1	0.0015	1
8 PSD band 1,4	0.1106	0	0.0279	1	0.2318	0	0.1210	0
8 PSD band 1,5	0.1715	0	0.1444	0	0.4179	0	0.4315	0
8 PSD band 1,6	0.2545	0	0.1149	0	0.2633	0	0.2164	0
8 PSD band 1,7	0.4033	0	0.0981	0	0.2633	0	0.1009	0
8 PSD band 2,3	0.1286	0	0.0252	1	0.0491	1	0.0015	1
8 PSD band 2,4	0.3619	0	0.0206	1	0.5628	0	0.0366	1
8 PSD band 2,5	0.4941	0	0.0493	1	0.6713	0	0.0455	1
8 PSD band 2,6	0.2545	0	1.0000	0	0.9079	0	0.0836	0
8 PSD band 2,7	1.0000	0	0.0765	0	0.6713	0	0.0366	1
8 PSD band 3,4	0.1965	0	0.0373	1	0.0698	0	0.0506	0
8 PSD band 3,5	0.4941	0	0.0540	0	0.1535	0	0.0328	1
8 PSD band 3,6	0.6485	0	0.0645	0	0.0339	1	0.0687	0
8 PSD band 3,7	0.9394	0	0.0540	0	0.0587	0	0.0561	0
8 PSD band 4,5	0.3233	0	0.0373	1	0.9692	0	0.2164	0

8 PSD band 4,6	0.8197	0	0.0339	1	0.7871	0	0.1704	0
8 PSD band 4,7	0.8197	0	0.0339	1	0.5628	0	0.0919	0
8 PSD band 5,6	0.8792	0	0.2844	0	0.6160	0	0.3568	0
8 PSD band 5,7	0.5433	0	0.0339	1	0.7285	0	0.2002	0
8 PSD band 6,7	0.4474	0	0.0061	1	0.6713	0	0.2909	0
8 PSD band 1,2,3	0.4941	0	0.0228	1	0.0698	0	0.0010	1
8 PSD band 1,2,4	0.2241	0	0.0150	1	0.0698	0	0.0010	1
8 PSD band 1,2,5	0.3233	0	0.0228	1	0.1325	0	0.0037	1
8 PSD band 1,2,6	0.1965	0	0.0167	1	0.0698	0	0.0043	1
8 PSD band 1,2,7	0.3619	0	0.0228	1	0.0826	0	0.0024	1
8 PSD band 1,3,4	0.2875	0	0.0252	1	0.0409	1	0.0015	1
8 PSD band 1,3,5	0.2241	0	0.0228	1	0.1137	0	0.0018	1
8 PSD band 1,3,6	0.1715	0	0.0339	1	0.0409	1	0.0007	1
8 PSD band 1,3,7	0.3619	0	0.0206	1	0.0587	0	0.0006	1
8 PSD band 1,4,5	0.1715	0	0.0493	1	0.7285	0	0.0919	0
8 PSD band 1,4,6	0.2545	0	0.0307	1	0.6160	0	0.1210	0
8 PSD band 1,4,7	0.4033	0	0.0450	1	0.4179	0	0.0687	0
8 PSD band 1,5,6	0.1965	0	0.0373	1	0.4636	0	0.3339	0
8 PSD band 1,5,7	0.0946	0	0.1242	0	0.6713	0	0.0758	0
8 PSD band 1,6,7	0.0276	1	0.0206	1	0.1137	0	0.1441	0
8 PSD band 2,3,4	0.1489	0	0.0206	1	0.0587	0	0.0011	1
8 PSD band 2,3,5	0.3619	0	0.0228	1	0.1325	0	0.0015	1
8 PSD band 2,3,6	0.3233	0	0.0307	1	0.0587	0	0.0013	1
8 PSD band 2,3,7	0.3233	0	0.0279	1	0.0587	0	0.0010	1
8 PSD band 2,4,5	0.1715	0	0.0307	1	0.4636	0	0.0328	1
8 PSD band 2,4,6	0.1965	0	0.0228	1	0.6160	0	0.0366	1
8 PSD band 2,4,7	0.5949	0	0.0206	1	0.5628	0	0.0232	1
8 PSD band 2,5,6	0.2241	0	0.1922	0	0.5628	0	0.0561	0
8 PSD band 2,5,7	0.4474	0	0.0645	0	0.4636	0	0.0206	1
8 PSD band 2,6,7	0.1715	0	0.0077	1	0.5628	0	0.0455	1
8 PSD band 3,4,5	0.3619	0	0.0765	0	0.1535	0	0.0328	1
8 PSD band 3,4,6	0.5949	0	0.0373	1	0.0491	1	0.0506	0
8 PSD band 3,4,7	0.9394	0	0.0540	0	0.0587	0	0.0561	0
8 PSD band 3,5,6	0.5949	0	0.0339	1	0.1535	0	0.0506	0
8 PSD band 3,5,7	0.2875	0	0.0645	0	0.2030	0	0.0455	1
8 PSD band 3,6,7	0.0946	0	0.0185	1	0.0339	1	0.0506	0
8 PSD band 4,5,6	0.8197	0	0.0410	1	0.7285	0	0.2336	0
8 PSD band 4,5,7	0.2241	0	0.0307	1	0.6713	0	0.1210	0
8 PSD band 4,6,7	0.3233	0	0.0042	1	0.4179	0	0.0293	1
8 PSD band 5,6,7	0.4474	0	0.0009	1	0.7871	0	0.0506	0
8 PSD band 1,2,3,4	0.3233	0	0.0307	1	0.0491	1	0.0008	1
8 PSD band 1,2,3,5	0.2545	0	0.0339	1	0.1325	0	0.0015	1
8 PSD band 1,2,3,6	0.1965	0	0.0077	1	0.0491	1	0.0008	1
8 PSD band 1,2,3,7	0.5433	0	0.0134	1	0.0587	0	0.0006	1
8 PSD band 1,2,4,5	0.1965	0	0.0185	1	0.1325	0	0.0013	1
8 PSD band 1,2,4,6	0.1965	0	0.0134	1	0.0972	0	0.0010	1

8 PSD band 1,2,4,7	0.3619	0	0.0252	1	0.1325	0	0.0011	1
8 PSD band 1,2,5,6	0.2545	0	0.0134	1	0.1325	0	0.0043	1
8 PSD band 1,2,5,7	0.1286	0	0.0228	1	0.1137	0	0.0018	1
8 PSD band 1,2,6,7	0.0402	1	0.0061	1	0.0698	0	0.0015	1
8 PSD band 1,3,4,5	0.2241	0	0.0228	1	0.1137	0	0.0015	1
8 PSD band 1,3,4,6	0.1965	0	0.0185	1	0.0491	1	0.0007	1
8 PSD band 1,3,4,7	0.4033	0	0.0252	1	0.0587	0	0.0006	1
8 PSD band 1,3,5,6	0.2241	0	0.0252	1	0.0826	0	0.0018	1
8 PSD band 1,3,5,7	0.1489	0	0.0185	1	0.1535	0	0.0010	1
8 PSD band 1,3,6,7	0.0334	1	0.0121	1	0.0491	1	0.0004	1
8 PSD band 1,4,5,6	0.2241	0	0.0185	1	0.6160	0	0.1210	0
8 PSD band 1,4,5,7	0.0682	0	0.0493	1	0.7871	0	0.0758	0
8 PSD band 1,4,6,7	0.0276	1	0.0026	1	0.2318	0	0.0506	0
8 PSD band 1,5,6,7	0.0185	1	0.0061	1	0.5628	0	0.0293	1
8 PSD band 2,3,4,5	0.1965	0	0.0167	1	0.0972	0	0.0015	1
8 PSD band 2,3,4,6	0.1965	0	0.0206	1	0.0698	0	0.0008	1
8 PSD band 2,3,4,7	0.3233	0	0.0185	1	0.0587	0	0.0008	1
8 PSD band 2,3,5,6	0.3233	0	0.0167	1	0.0972	0	0.0015	1
8 PSD band 2,3,5,7	0.1965	0	0.0206	1	0.1535	0	0.0010	1
8 PSD band 2,3,6,7	0.0334	1	0.0096	1	0.0587	0	0.0007	1
8 PSD band 2,4,5,6	0.2241	0	0.0279	1	0.6713	0	0.0328	1
8 PSD band 2,4,5,7	0.0806	0	0.0134	1	0.3348	0	0.0183	1
8 PSD band 2,4,6,7	0.1106	0	0.0026	1	0.6713	0	0.0143	1
8 PSD band 2,5,6,7	0.1489	0	0.0012	1	0.4636	0	0.0037	1
8 PSD band 3,4,5,6	0.6485	0	0.0252	1	0.1535	0	0.0561	0
8 PSD band 3,4,5,7	0.2241	0	0.0373	1	0.2318	0	0.0561	0
8 PSD band 3,4,6,7	0.0682	0	0.0068	1	0.0339	1	0.0328	1
8 PSD band 3,5,6,7	0.0575	0	0.0096	1	0.1325	0	0.0261	1
8 PSD band 4,5,6,7	0.1286	0	0.0054	1	0.6713	0	0.0561	0
8 PSD band 1,2,3,4,5	0.4474	0	0.0228	1	0.1325	0	0.0013	1
8 PSD band 1,2,3,4,6	0.2241	0	0.0134	1	0.0409	1	0.0008	1
8 PSD band 1,2,3,4,7	0.0227	1	0.0086	1	0.1325	0	0.0005	1
8 PSD band 1,2,3,5,6	0.2241	0	0.0150	1	0.0826	0	0.0015	1
8 PSD band 1,2,3,5,7	0.1106	0	0.0410	1	0.1137	0	0.0011	1
8 PSD band 1,2,3,6,7	0.0334	1	0.0054	1	0.0491	1	0.0004	1
8 PSD band 1,2,4,5,6	0.2241	0	0.0134	1	0.1137	0	0.0015	1
8 PSD band 1,2,4,5,7	0.0575	0	0.0185	1	0.1325	0	0.0011	1
8 PSD band 1,2,4,6,7	0.0402	1	0.0061	1	0.1137	0	0.0007	1
8 PSD band 1,2,5,6,7	0.0227	1	0.0037	1	0.0972	0	0.0018	1
8 PSD band 1,3,4,5,6	0.2241	0	0.0228	1	0.0972	0	0.0015	1
8 PSD band 1,3,4,5,7	0.0806	0	0.0134	1	0.1137	0	0.0011	1
8 PSD band 1,3,4,6,7	0.0276	1	0.0054	1	0.0587	0	0.0005	1
8 PSD band 1,3,5,6,7	0.0276	1	0.0033	1	0.0826	0	0.0008	1
8 PSD band 1,4,5,6,7	0.0185	1	0.0033	1	0.6713	0	0.0328	1
8 PSD band 2,3,4,5,6	0.2241	0	0.0206	1	0.1137	0	0.0013	1
8 PSD band 2,3,4,5,7	0.0682	0	0.0134	1	0.1137	0	0.0011	1

8 PSD band 2,3,4,6,7	0.0276	1	0.0077	1	0.0698	0	0.0005	1
8 PSD band 2,3,5,6,7	0.0276	1	0.0020	1	0.0826	0	0.0010	1
8 PSD band 2,4,5,6,7	0.0575	0	0.0033	1	0.5120	0	0.0057	1
8 PSD band 3,4,5,6,7	0.0575	0	0.0068	1	0.1770	0	0.0143	1
8 PSD band 1,2,3,4,5,6	0.1106	0	0.0134	1	0.0186	1	0.0013	1
8 PSD band 1,2,3,4,5,7	0.0098	1	0.0017	1	0.0339	1	0.0008	1
8 PSD band 1,2,3,4,6,7	0.0227	1	0.0061	1	0.0279	1	0.0005	1
8 PSD band 1,2,3,5,6,7	0.0227	1	0.0054	1	0.0826	0	0.0008	1
8 PSD band 1,2,4,5,6,7	0.0227	1	0.0068	1	0.0972	0	0.0008	1
8 PSD band 1,3,4,5,6,7	0.0227	1	0.0048	1	0.0698	0	0.0008	1
8 PSD band 2,3,4,5,6,7	0.0227	1	0.0054	1	0.0826	0	0.0008	1
8 PSD band 1 to 7	0.0098	1	0.0017	1	0.0014	1	0.0010	1

For graphical view of the separation, figures 6.2 to 6.6 represent plots of mean and standard deviation error bars in each category of male and female for interview and reading sessions.

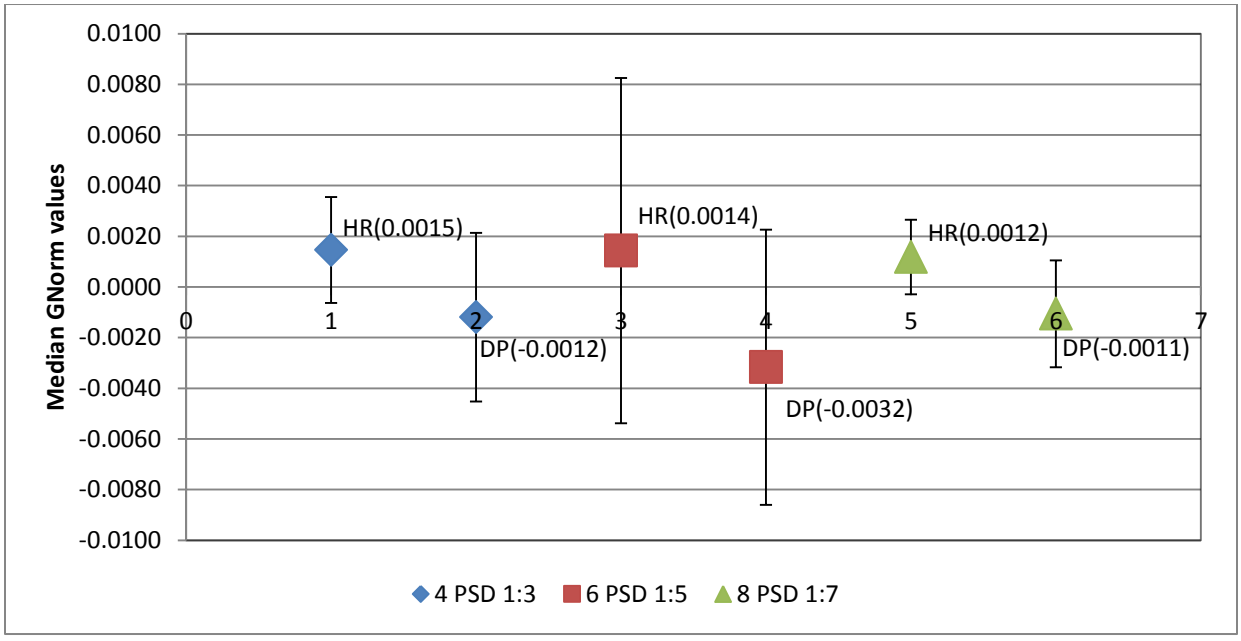


Figure 6.2: *Comparison of the mean and standard deviation for HR-DP male interview*

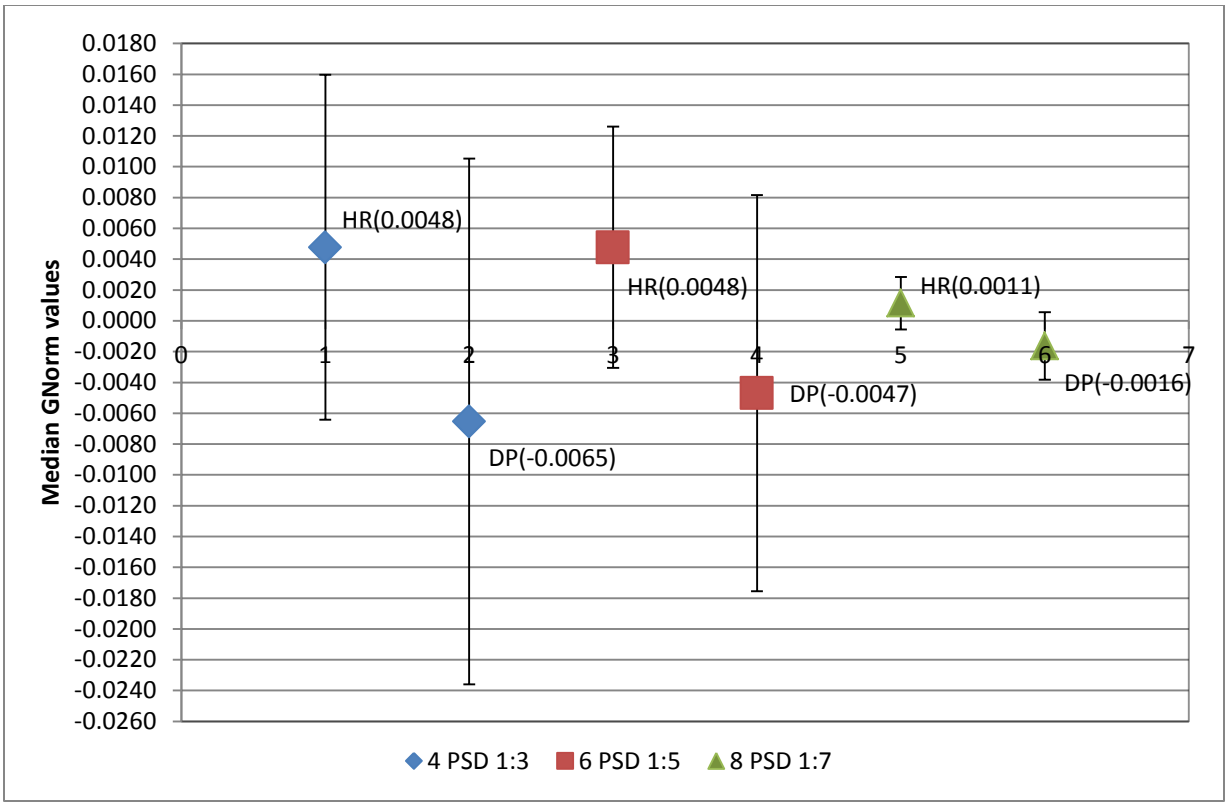


Figure 6.3: *Comparison of the mean and standard deviation for HR-DP female interview*

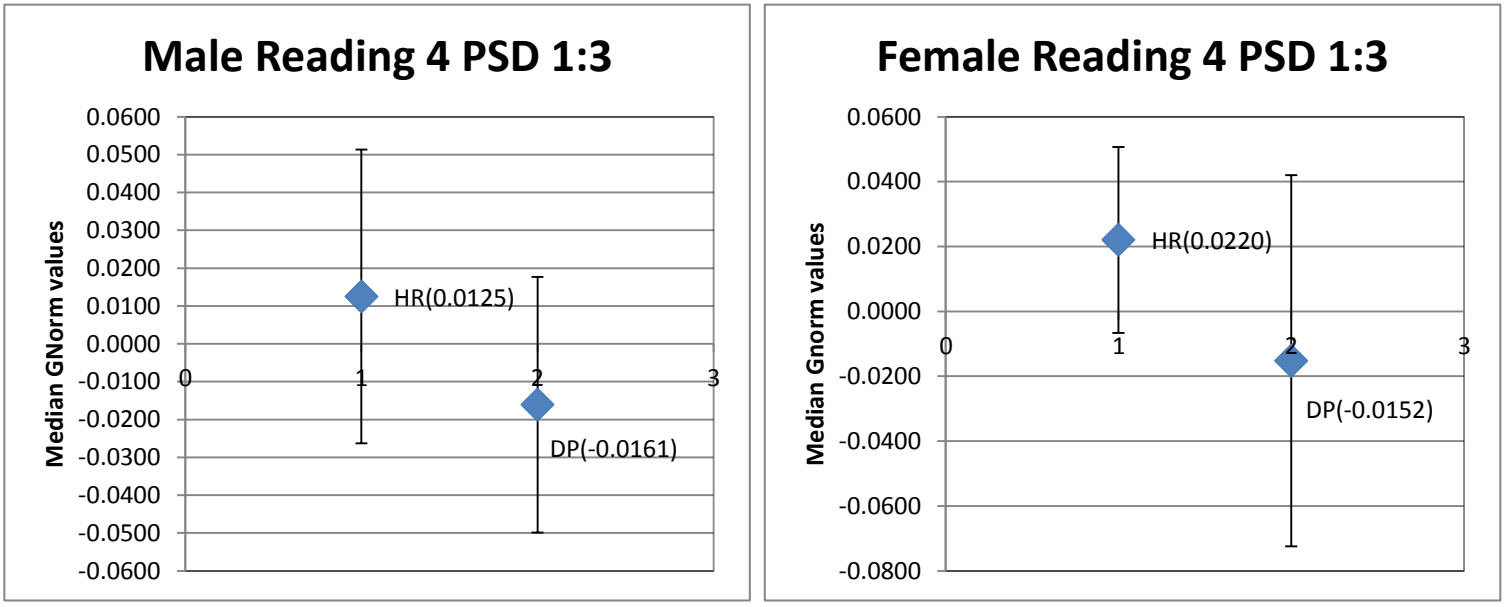


Figure 6.4: *Comparison of the mean and standard deviation for HR-DEP 4PSD 1:3 male and female reading*

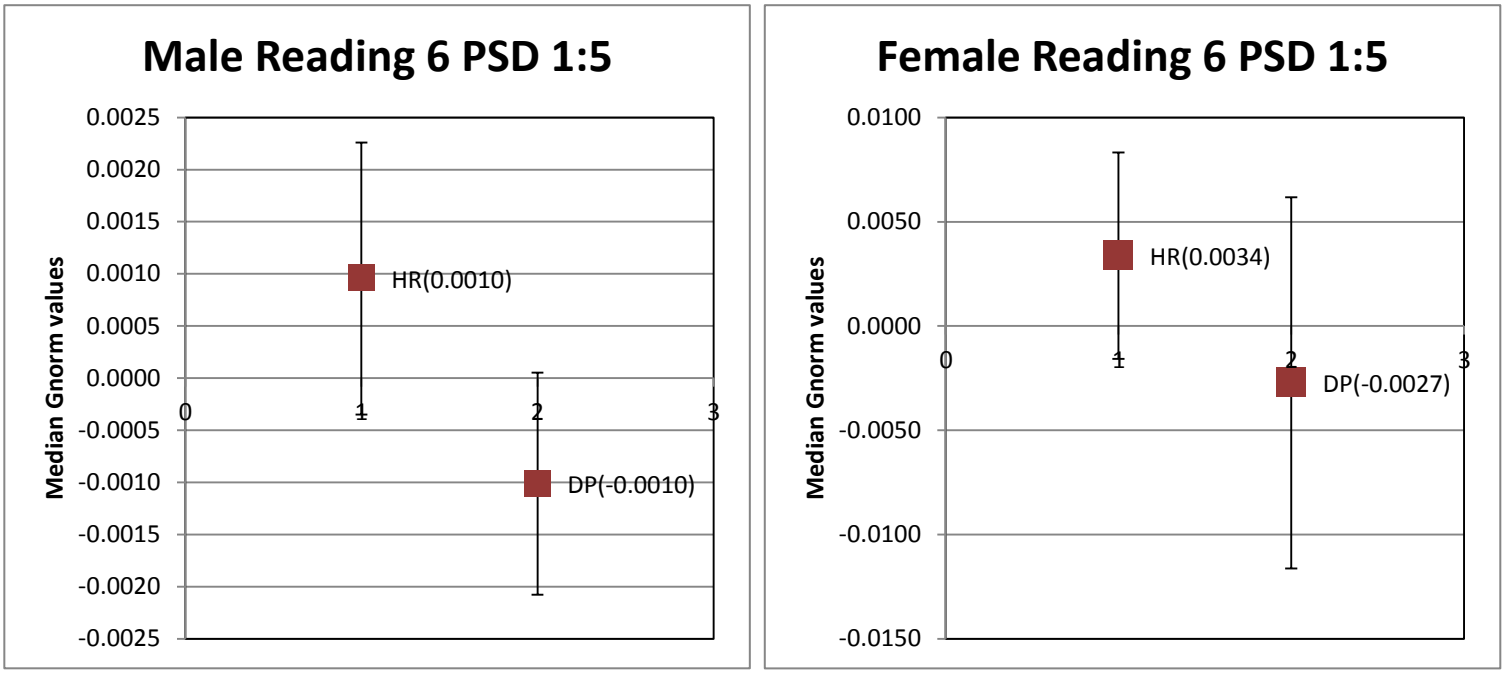


Figure 6.5: *Comparison of the mean and standard deviation for HR-DEP 6PSD 1:5 male and female reading*

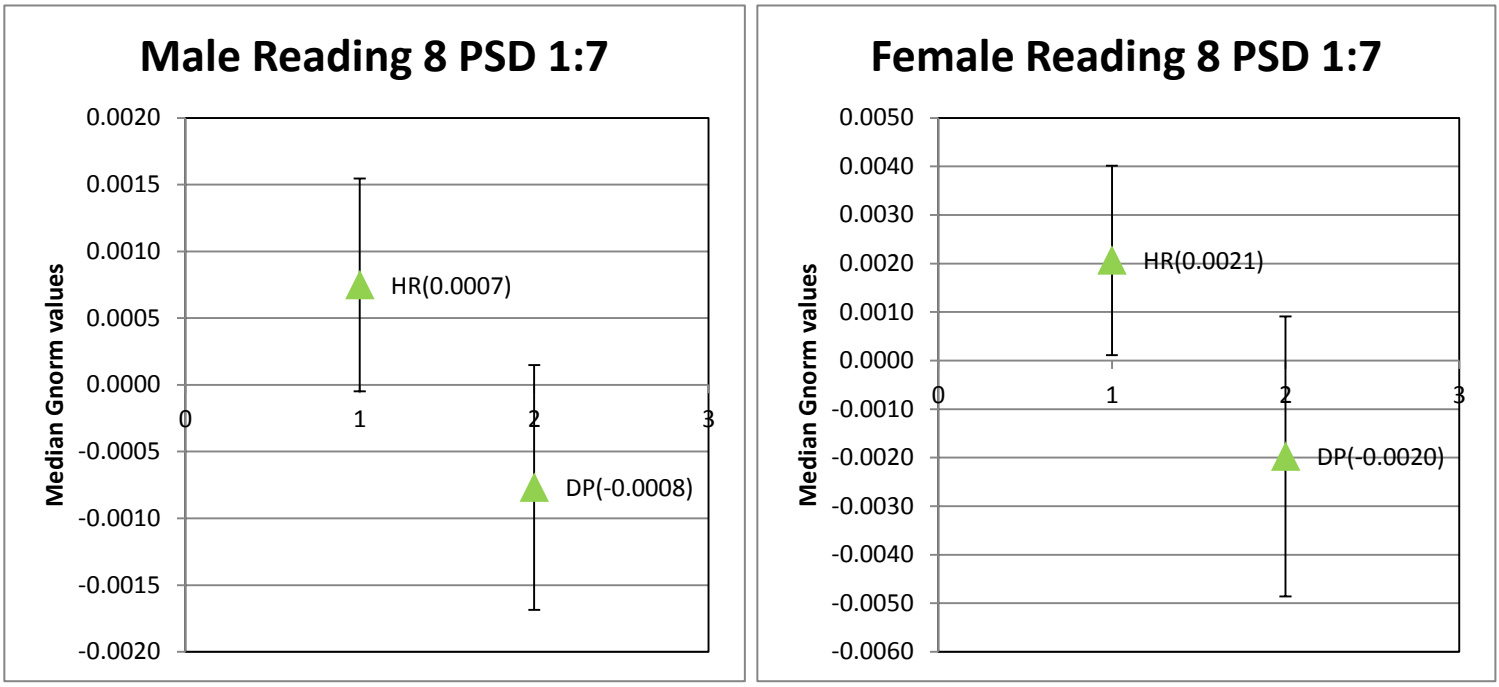


Figure 6.6: *Comparison of the mean and standard deviation for HR-DEP 8PSD 1:7 male and female reading*

4.2 Results for the Statistical Analysis and Significant Differences between Recording Sessions for Database B

4.2.1 Statistical Analysis on Male Interview

For Database B, in order for the features to be generalizable according to [24], a maximum of two features are considered adequate for male interview, one feature for male reading, three features for female interview and two features for female reading.

The estimated mean normalized Euclidean distance gathered from all combinations of 4PSD bands for male patient using an interview speech is shown in table 6.5. We can observe that during the first session, the distances were further away in the positive direction along the normal line signifying that they are in the high risk suicidal state. A larger number represents a greater level of high risk. The mean values demonstrate a general trend of progression in the negative direction during the second session which equates to patients getting better. An insignificant progression or regression towards the hyperplane was observed occurring from the second session to the third session.

Table 6.5: Estimated mean normalized Euclidean distances for all 4 PSD band combinations using the male interview speech

Patient		mi1	mi2	mi3	mi4	mi5	**Mean**	**Stdev**
4 PSD band 1	s1	0.0691	0.0784	0.0401	0.1142	0.0205	0.0644	0.0361
	s2	-0.0393	-0.2137	-0.1701	-0.1771	-0.1028	-0.1406	0.0694
	s3	-0.1038	0.0569	0.0843	0.0080	0.0776	0.0246	0.0778
4 PSD band 2	s1	0.1029	0.0768	0.0186	0.1109	0.0270	0.0672	0.0426
	s2	-0.0820	-0.1793	-0.0679	-0.1718	-0.1003	-0.1203	0.0518
	s3	-0.1272	0.0257	0.0285	0.0073	0.0604	-0.0011	0.0730
4 PSD band 3	s1	0.0197	0.0018	0.0172	0.0001	0.0038	0.0085	0.0092
	s2	-0.0308	-0.0248	-0.0793	-0.0055	0.0026	-0.0276	0.0320
	s3	-0.0068	0.0211	0.0422	0.0103	-0.0115	0.0111	0.0218
4 PSD band 1:2	s1	0.0732	0.0218	0.0256	0.0481	0.0080	0.0353	0.0256
	s2	-0.0638	-0.0342	-0.1142	-0.0743	-0.0132	-0.0599	0.0388
	s3	-0.0842	-0.0094	0.0591	0.0029	-0.0019	-0.0067	0.0511
4 PSD band 1:3	s1	0.0507	0.0236	0.0190	0.0144	0.0024	0.0220	0.0179
	s2	-0.0431	-0.0369	-0.0851	-0.0186	-0.0026	-0.0373	0.0311
	s3	-0.0596	-0.0103	0.0444	-0.0062	-0.0022	-0.0068	0.0369
4 PSD band 2:3	s1	0.0993	0.0249	0.0228	0.0607	0.0115	0.0438	0.0361
	s2	-0.0866	-0.0389	-0.0988	-0.0895	-0.0266	-0.0681	0.0328
	s3	-0.1141	-0.0109	0.0498	-0.0046	0.0065	-0.0147	0.0604

Results from a paired t-test analysis for male interview are shown in table 6.6. The progression from the first to the second session did not reveal any significance when using only band 3. However, band 1:3 revealed a probably statistically significant difference with a p-value of $0.0164 < 0.05$ and the rest of the combinations show significant differences in the mean of the normalized Euclidean distance with a p-value < 0.01. The latter results suggest that at least 99% confidence they are indeed distinctive and thus, providing evidence that there is positive changes in patient's condition after receiving treatment, medication or being hospitalized. There were no significant differences found between second and third session indicating that patient's condition were not noticeably improving or deteriorating during the last session.

Table 6.6: T-statistics, t and the one-tailed significance p-value, p for measuring the mean difference between each pairwise session for male interview using all possible combinations of 4 PSD bands

	session number		
significance (p)	**s1-s2**	**s1-s3**	**s2-s3**
4 PSD band 1	0.0033	0.2086	0.9741
4 PSD band 2	0.0036	0.1099	0.9718
4 PSD band 3	0.0536	0.5940	0.9171
4 PSD band 1:2	0.0082	0.1281	0.9042
4 PSD band 1:3	0.0164	0.1362	0.8489
4 PSD band 2:3	0.0073	0.1164	0.9268
t-statistic (t)	**s1-s2**	**s1-s3**	**s2-s3**
4 PSD band 1	5.2016	0.9038	-2.7438
4 PSD band 2	5.0474	1.4536	-2.6594
4 PSD band 3	2.0705	-0.2539	-1.6925
4 PSD band 1:2	3.9755	1.3236	-1.5694
4 PSD band 1:3	3.2029	1.2718	-1.1831
4 PSD band 2:3	4.1173	1.4044	-1.7989

Figure 6.7 illustrates the mean and standard deviation error bars representing a general trend of progression in the negative direction during the second session and a minor progression or regression along the normal line during the last session using 4 PSD band 1 that was selected due to its lowest significance for male interview patients.

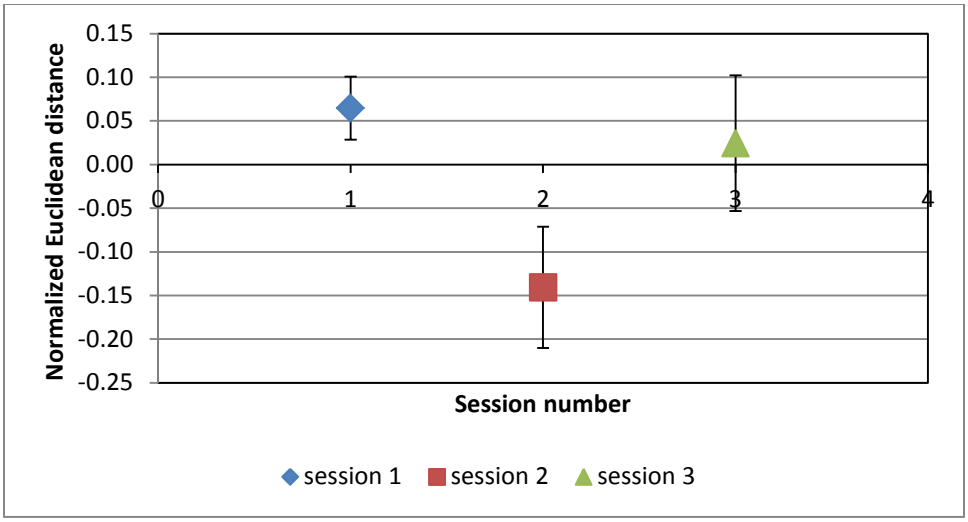


Figure 6.7: *Plot of the mean normalized Euclidean distance in three different sessions using 4 PSD band 1 for male interview*

4.2.2 Statistical Analysis on female Interview

The estimated mean normalized Euclidean distances gathered from all possible combinations of 4 PSD bands for female patients during an interview are given in table 6.7. By observation, the same general trend of progression displayed by male interview was also shown in female interview. All patients showed a progression in the negative direction along the normal line when going from the first session to the second session and showed no significant changes during the last session. Patient labeled fi4 and fi7 demonstrated consistent improvement in their consecutive sessions for all band combinations.

Table 6.7: Estimated mean normalized Euclidean distances for all 4 PSD band combinations using the female interview speech

Patient		fi1	fi2	fi3	fi4	fi5	fi6	fi7	fi8	**Mean**	**Stdev**
4 PSD band 1	s1	0.1372	0.0336	0.0127	0.0587	0.0102	0.1777	0.0465	0.0280	0.0505	0.0523
	s2	-0.0722	-0.0802	-0.2171	0.0178	0.0948	-0.1425	-0.0628	-0.1950	-0.0514	0.1171
	s3	-0.2208	0.0084	0.1009	-0.1353	-0.1583	-0.2481	0.0068	0.1372	-0.0810	0.1318
4 PSD band 2	s1	0.1381	0.0566	0.1072	0.0652	0.0105	0.1511	0.0706	0.0280	0.0755	0.0490
	s2	-0.0881	-0.0899	-0.2344	0.0063	0.0493	-0.0907	-0.0778	-0.2044	-0.0714	0.1093
	s3	-0.2024	-0.0266	-0.0366	-0.1367	-0.0949	-0.2720	-0.0475	0.1461	-0.0995	0.0729
4 PSD band 3	s1	0.0041	0.0187	0.0613	0.0017	0.0165	0.0202	0.0212	0.0029	0.0205	0.0240
	s2	0.0047	-0.0036	-0.0137	0.0097	0.0002	-0.0363	-0.0156	-0.0057	-0.0005	0.0089
	s3	-0.0153	-0.0323	-0.0877	-0.0131	-0.0402	0.0121	-0.0393	-0.0004	-0.0377	0.0302
4 PSD band 1:2	s1	0.0495	0.0271	0.0823	0.0436	0.0052	0.0983	0.0705	0.0377	0.0415	0.0285
	s2	-0.0394	-0.0278	-0.0752	0.0004	0.0039	-0.0916	-0.0775	-0.2669	-0.0276	0.0323
	s3	-0.0625	-0.0266	-0.0863	-0.0875	-0.0179	-0.1117	-0.0475	0.1889	-0.0562	0.0327
4 PSD band 1:3	s1	0.0299	0.0073	0.0636	0.0501	0.0133	0.0511	0.0207	0.0032	0.0328	0.0239
	s2	-0.0246	-0.0090	-0.0568	0.0000	-0.0154	-0.0485	-0.0238	-0.0070	-0.0212	0.0219
	s3	-0.0367	-0.0058	-0.0674	-0.1002	-0.0103	-0.0563	-0.0106	0.0006	-0.0441	0.0399
4 PSD band 2:3	s1	0.0651	0.0401	0.0959	0.0537	0.0155	0.0806	0.0733	0.0082	0.0541	0.0298
	s2	-0.0479	-0.0399	-0.0975	0.0000	-0.0044	-0.0704	-0.0793	-0.0445	-0.0379	0.0394
	s3	-0.0873	-0.0403	-0.0950	-0.1075	-0.0311	-0.1009	-0.0537	0.0275	-0.0722	0.0343

The results from a paired t-test analysis shown in table 6.8 quantify the significance of the progression or regression between pairwise sessions. The significance of $p=0.0188 < 0.05$ for band 3 suggested at least 95% confidence of the separation between recordings in session 1 and session 2. The rest of the combinations demonstrated a significant difference with p-value << 0.01 between the first and second session providing evidence there is a positive change in patient's condition after receiving treatment, medication or being hospitalized. Significant differences were observed for combinations of band 1:3 and 2:3 between the first and third session suggest a 99% confidence that they are indeed separable, thus also indicating that patients were not experience a relapse that occurred during their first session. Likewise, there were no significant differences observed between the second and third session thus indicating that the patients' condition were not noticeably improving or deteriorating during the last session.

Table 6.8: T-statistics, t and the one-tailed significance p-value, p for measuring the mean difference between each pairwise session for female interview using all possible combinations of 4 PSD bands

	session number		
significance (p)	**s1-s2**	**s1-s3**	**s2-s3**
4 PSD band 1	0.0073	0.0551	0.5902
4 PSD band 2	0.0025	0.0138	0.5422
4 PSD band 3	0.0188	0.0155	0.0815
4 PSD band 1:2	0.0044	0.0365	0.7370
4 PSD band 1:3	0.0031	0.0070	0.1792
4 PSD band 2:3	0.0010	0.0015	0.2518
t-statistic (t)	**s1-s2**	**s1-s3**	**s2-s3**
4 PSD band 1	3.2199	1.8285	-0.2369
4 PSD band 2	4.0437	2.7722	-0.1098
4 PSD band 3	2.5584	2.6905	1.5589
4 PSD band 1:2	3.5914	2.1086	-0.6673
4 PSD band 1:3	3.8604	3.2518	0.9829
4 PSD band 2:3	4.7864	4.4303	0.7049

Figure 6.8 illustrates the mean and standard deviation error bars representing a general trend of progression in the negative direction during the second session and a minor progression or regression along the normal line during the last session using 4 PSD bands 2:3 that was selected due to its lowest significance for female interview patients.

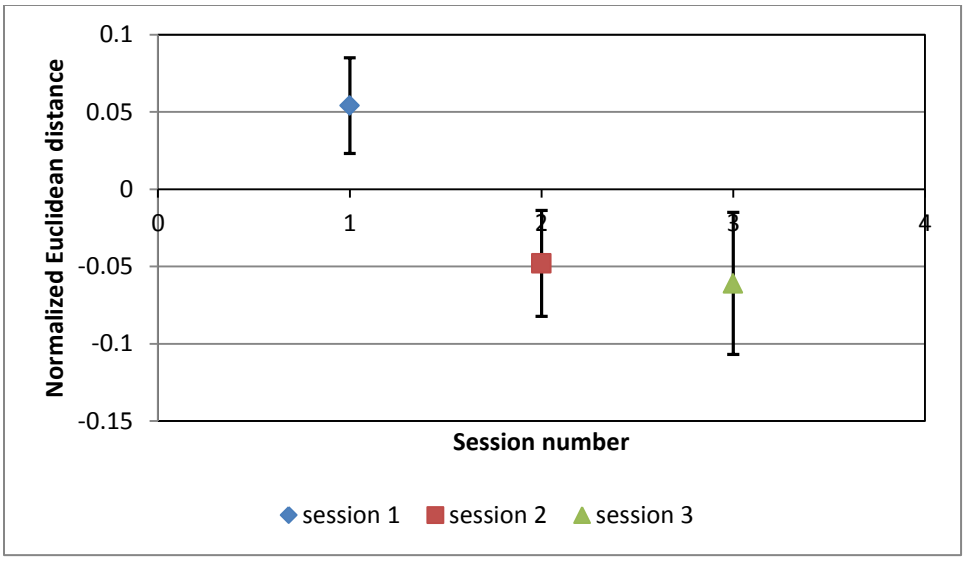


Figure 6.8: *Plot of the mean normalized Euclidean distance in three different sessions using 4 PSD bands 2:3 for female interview*

4.2.3 Statistical Analysis on Male Reading

The estimated mean normalized Euclidean distance gathered from all possible 4 PSD band combinations for male patients during a reading session are given in table 6.9. Based on the mean, results from the male reading also demonstrate the same general trend of progression as displayed by the results in the interview. All patients showed a progression in the negative direction along the normal line in the second session and experienced a minor regression in the third session.

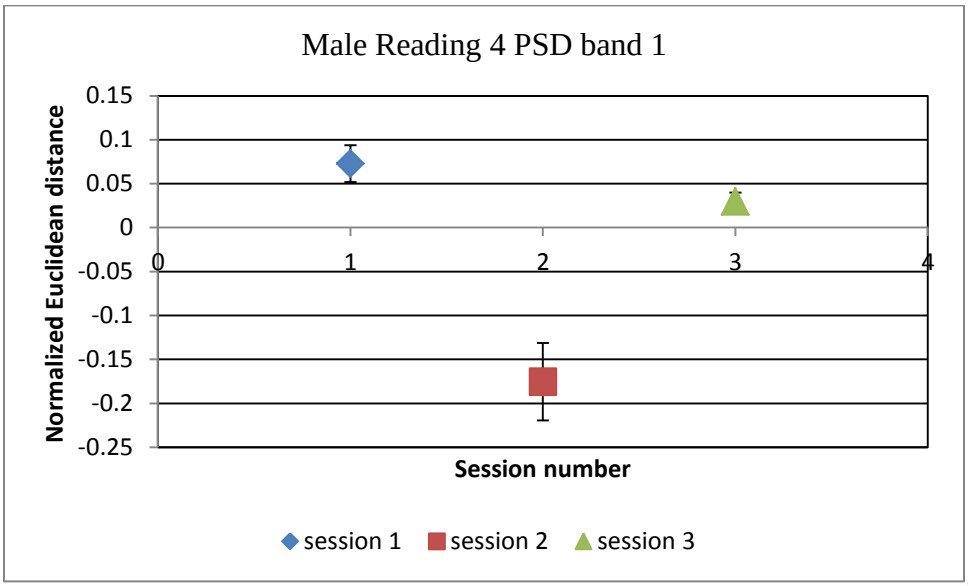
Referring to table 6.10, the progression from the first to the second session using band 1 demonstrated a significant difference between the mean of the normalized Euclidean distance with a one-tailed p=0.0023. This result suggests a 99% confidence that these two sessions were separable and provide evidence of an existing progression in the patients' condition. Likewise, there was no significant difference found between second and third session indicating that patients' condition were not noticeably improving or deteriorating during the last session. A significant difference of p=0.0007 << 0.01 between the first and third session using combination of band 1:3 indicate that patients were not experiencing relapse.

Table 6.9: Estimated mean normalized Euclidean distances for all 4 PSD band combinations using the male reading speech

Patient		mr1	mr2	mr3	mr4	Mean	Stdev
4 PSD band 1	s1	0.0438	0.0785	0.0939	0.0750	0.0728	0.0210
	s2	-0.1218	-0.1788	-0.2296	-0.1713	-0.1754	0.0441
	s3	0.0341	0.0218	0.0419	0.0212	0.0297	0.0100
4 PSD band 2	s1	0.0379	0.0859	0.0635	0.0927	0.0700	0.0248
	s2	0.0593	-0.1663	-0.1442	-0.1912	-0.1106	0.1149
	s3	-0.1350	-0.0055	0.0173	0.0057	-0.0294	0.0710
4 PSD band 3	s1	0.0533	0.0055	0.0222	0.0187	0.0249	0.0202
	s2	-0.0392	0.0060	-0.0655	-0.0151	-0.0284	0.0308
	s3	-0.0674	-0.0169	0.0211	-0.0223	-0.0214	0.0362
4 PSD band 1:2	s1	0.0579	0.0197	0.0296	0.0144	0.0304	0.0194
	s2	-0.0574	-0.0217	-0.0621	-0.0181	-0.0398	0.0231
	s3	-0.0584	-0.0177	0.0028	-0.0107	-0.0210	0.0264
4 PSD band 1:3	s1	0.0164	0.0227	0.0425	0.0255	0.0268	0.0111
	s2	-0.0166	-0.0254	-0.0889	-0.0267	-0.0394	0.0333
	s3	-0.0162	-0.0200	0.0039	-0.0242	-0.0141	0.0125
4 PSD band 2:3	s1	0.0479	0.0235	0.0431	0.0224	0.0342	0.0132
	s2	-0.0465	-0.0292	-0.0902	-0.0227	-0.0472	0.0304
	s3	-0.0494	-0.0177	0.0040	-0.0221	-0.0213	0.0219

Table 6.10: T-statistics, t and the one-tailed significance p-value, p for measuring the mean difference between each pairwise session for male reading using all possible combinations of 4 PSD bands

	session number		
significance (p)	**s1-s2**	**s1-s3**	**s2-s3**
4 PSD band 1	0.0023	0.0153	0.9983
4 PSD band 2	0.0398	0.0166	0.7783
4 PSD band 3	0.0485	0.0871	0.5951
4 PSD band 1:2	0.0193	0.0498	0.5951
4 PSD band 1:3	0.0291	0.0007	0.8279
4 PSD band 2:3	0.0141	0.0142	0.8291
t-statistic (t)	**s1-s2**	**s1-s3**	**s2-s3**
4 PSD band 1	7.6678	3.8621	-8.4870
4 PSD band 2	2.6122	3.7448	-0.8806
4 PSD band 3	2.3873	1.7737	-0.2628
4 PSD band 1:2	3.5289	2.3567	-0.2628
4 PSD band 1:3	2.9898	11.3375	-1.1204
4 PSD band 2:3	3.9873	3.9739	-1.1268



3.

Figure 6.9: *Plot of the mean normalized Euclidean distance in three different sessions using 4 PSD band 1 for male reading*

Figure 6.9 illustrates the mean and standard deviation error bars representing a general trend of progression in the negative direction during the second session and a minor progression or regression along the normal line during the last session using 4 PSD band 1 that was selected due to its lowest significance for male reading patients.

4.2.4 Statistical Analysis on Female Reading

The estimated mean normalized Euclidean distance gathered from all possible 4 PSD band combinations for female patients during a reading session are given in table 6.11. By observation, the same general trends of progression displayed by previous results were shown in the female reading speech. The mean value for all band combinations demonstrated a consistent improvement in their consecutive sessions.

Table 6.11: Estimated mean normalized Euclidean distances for all 4 PSD band combinations using the female reading speech

Patient		fr1	fr2	fr3	fr4	fr5	fr6	fr7	**Mean**	**Stdev**
4 PSD band 1	s1	0.0621	0.0731	0.0740	0.0708	0.0534	0.1948	0.0059	0.0667	0.0088
	s2	0.0308	-0.1700	-0.2003	0.0415	0.0674	-0.2064	-0.0046	-0.0461	0.1281
	s3	-0.1550	0.0722	0.0523	-0.1831	-0.1742	-0.1832	-0.0073	-0.0776	0.1282
4 PSD band 2	s1	0.0476	0.0625	0.1317	0.0744	0.0308	0.1858	0.0383	0.0694	0.0385
	s2	0.0249	-0.1575	-0.2354	0.0083	0.0407	-0.1866	-0.0907	-0.0638	0.1247
	s3	-0.1201	0.0798	-0.0280	-0.1570	-0.1022	-0.1849	0.0140	-0.0655	0.0938
4 PSD band 3	s1	0.0132	0.0108	0.0328	0.0078	0.0169	0.0063	0.0282	0.0163	0.0098
	s2	0.0061	-0.0154	-0.0355	-0.0375	0.0045	-0.0077	-0.0650	-0.0156	0.0209
	s3	-0.0326	-0.0037	-0.0301	0.0219	-0.0382	-0.0048	0.0087	-0.0166	0.0253
4 PSD band 1:2	s1	0.0111	0.0230	0.0545	0.0279	0.0234	0.0415	0.0286	0.0280	0.0161
	s2	0.0046	-0.0464	-0.0541	-0.0140	0.0314	-0.0356	-0.0728	-0.0157	0.0355
	s3	-0.0267	0.0122	-0.0549	-0.0419	-0.0781	-0.0474	0.0157	-0.0379	0.0338
4 PSD band 1:3	s1	0.0018	0.0049	0.0561	0.0171	0.0027	0.0416	0.0110	0.0165	0.0230
	s2	0.0004	-0.0096	-0.0559	-0.0135	-0.0025	-0.0400	-0.0212	-0.0162	0.0228
	s3	-0.0039	0.0023	-0.0564	-0.0208	-0.0029	-0.0432	-0.0008	-0.0163	0.0240
4 PSD band 2:3	s1	0.0132	0.0220	0.0347	0.0440	0.0196	0.0400	0.0262	0.0267	0.0124
	s2	0.0061	-0.0440	-0.0388	-0.0283	0.0081	-0.0388	-0.0605	-0.0194	0.0249
	s3	-0.0326	0.0109	-0.0305	-0.0596	-0.0472	-0.0412	0.0080	-0.0318	0.0267

Referring to table 6.12, combinations of band 1:2 and band 2:3 display significant differences between the first and second session with p=0.0079 and p=0.0020, respectively and also between the first and third session with p=0.0038 and p=00022, respectively. There was no significant difference found between second and third session indicating that patient's condition were not noticeably improving or deteriorating during the last session. These results provide evidence of a progression in patient's condition after treatment and also suggest that patients did not experience a relapse that occurred during their first session.

Figure 6.10 illustrates the mean and standard deviation error bars representing a general trend of progression in the negative direction during the second session and a minor progression or regression along the normal line during the last session using 4 PSD bands 2:3 that was selected due to its lowest significance for female reading patients.

Table 6.12: T-statistics, t and the one-tailed significance p-value, p for measuring the mean difference between each pairwise session for female reading using all possible combinations of 4 PSD bands

	session number		
significance (p)	**s1-s2**	**s1-s3**	**s2-s3**
4 PSD band 1	0.0328	0.0145	0.4068
4 PSD band 2	0.0155	0.0101	0.5815
4 PSD band 3	0.0106	0.0187	0.7187
4 PSD band 1:2	0.0079	0.0038	0.4240
4 PSD band 1:3	0.0228	0.0339	0.7216
4 PSD band 2:3	0.0020	0.0022	0.5122
t-statistic (t)	**s1-s2**	**s1-s3**	**s2-s3**
4 PSD band 1	2.2485	2.8545	0.2464
4 PSD band 2	2.8016	3.1354	-0.2147
4 PSD band 3	3.0978	2.6622	-0.6127
4 PSD band 1:2	3.3268	3.9467	0.2001
4 PSD band 1:3	2.5138	2.2240	-0.6221
4 PSD band 2:3	4.5431	4.4503	-0.0318

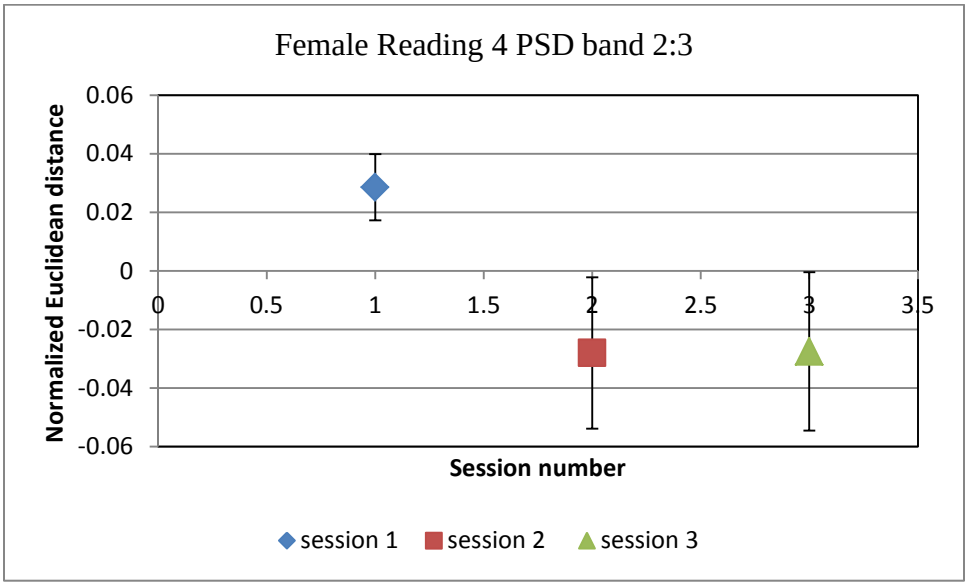


Figure 6.10: *Plot of the mean normalized Euclidean distance in three different sessions using 4 PSD bands 2:3 for female reading*

5.0 Discussion and Conclusion

5.1 Database A

The statistical test performed on the two population means demonstrated that PSD features exhibit different characteristic that can be used to distinguish between high risks suicidal and major depressed populations. Mean values from the depressed group were shown to be lower than the high risk group and moving in the opposite direction. The fact the two groups are indeed separable as illustrated above. If they clustered up in one cloud and are not separable, it destroys our argument and would also mean that this feature is unusable. The t-test results quantifying the measure of separability present a significance difference between these two populations for certain combination of bands.

5.2 Database B

According to the assessment made by psychiatrist, patients were in high risk during the first session and the other two sessions were kept unknown to the researcher. With the type of data in hand, the aim was not so much in making a clear decision between pairwise group of HR-DEP, DEP-REM or HR-REM. Instead, the hyplerplane was used as a 1-D scale, measuring progression versus regression. By observation, patients seem to look better after each session and as shown by the results in this study, their condition does seem to correlate in the measured Euclidean distance (z-value). In addition, the significance of improvement and deterioration in their condition were represented in the statistical t-test analysis.

There are many other bands combination that can be explored for the statistical analysis besides the 4 PSD band combinations (i.e. using 6 PSD and 8 PSD band combinations) but since we were able to demonstrate a significant difference using only combinations from 4 PSD bands and thus we did not proceed with the analysis on smaller band combinations. Plus, we are only looking at the progression or regression along the normal line between pairwise sessions. Choosing other combinations of bands only change the threshold of where zero is but it doesn't change the direction of the 'movement'.

A linear separation was chosen as the hyperplane because it gives a unique normal line. The direction orthogonal to the hyperplane can be used to establish a direction indicating whether patients are getting better or worse as they go through the normal line. As

illustrated in the results, they are strikingly moving in the same direction (in the negative direction) during the second session with the exception of a few cases.

Even though most patients did not display a significant improvement when going from the second to third session, there is at least a significant of improvement from the first session (HR) to the other two sessions with the exception of a few cases. The fact that the second and third session did not demonstrate a significant difference from each other, it shows that more or less, there is only minor progression or regression along the normal line and importantly, providing evidence that their condition does not return back to being high risk even after the third session.

The hyperplane was chosen in a way that maximizes the number of HR vectors in one and the rest on the other. By doing this, it guarantees that the second and third session will have a smaller distance than the first session. This is simply due to the orientation of the hyperplane. But it does not guarantee that the third session will be more negative than the second session and the possibility of the existence of outlier in either direction was neglected. The fact that they are indeed separable in most cases definitely shows that overtime with treatment, their distances to the hyperplane do tend to move in one direction.

References

[1] K. A. Busch, J. Fawcett, D. G. Jacob, "Clinical Correlates of Inpatient Suicide", Psychiatry Ann, 1993.

[2] S. J. Blumenthal, D. J. Kupter , "Suicide over the Life Cycle: Risk Factors, Assessment, and Treatment of Suicidal Patients", pg.111.

[3] R. I. Simon, "*Preventing Patient Suicide: Clinical Assessment and Management"*, Washington, DC, American Psychiatric Publishing, 2011.

[4] D. J. France, *"Acoustical Properties of Speech as Indicators of Depression and Suicidal Risk"*, PhD Thesis, Vanderbilt University, 1997.

[5] A. Ozdas, *"Analysis of Paralinguistic Properties of Speech for Near Term Suicidal Risk Assessment"*, PhD Thesis, Vanderbilt University, 2001.

[6] T. Yingthawornsuk, *"Acoustic Analysis of Vocal Output Characteristic for Suicidal Risk Assessment"*, PhD Thesis, Vanderbilt University, 2007.

[7] H. K. Keskinpala., *"Analysis of Spectral Properties of Speech for Detecting Suicide Risk and Impact of Gender Specific Differences"*, PhD Thesis, Vanderbilt University, 2011.

[8] N. N. Wahidah, *"Analysis of Power Spectrum Density of Male Speech as Indicators for High Risk and Depressed Decision"*, Master Thesis, Vanderbilt University, 2011.

[9] W. A. Hasan, *"Acoustic Analysis of Speech Based on Power Spectral Density Features in Detecting Suicidal Risk Among Female Patients"*, Master Thesis, Vanderbilt University, 2011.

[10] S. Theodoridis, K. Koutroumbas, "*Pattern Recognition, Fourth Edition*" Ac-ic Press, 2008

[11] M. Alpert, Encoding of feelings in voice.P. J. Clayton & J. E. Barrett Edition, New York: Raven Press, pp. 2 17-228, 1983.

[12] W.A. Hargreaves, W. A., J. A. Starkweather, "Voice quality changes in dcprcssion", *Language and Speech*, vol. 7, pp. 84-88, 1946.

[13] S. Newmann, V. G. Mather, "Analysis of spoken language of patients with" *Journd of Psychiatry*, vol. *94, pp.912~ 942, 1932.*

[14] A. Nilsonne, "Acoustic analysis of speech variables during depression and after improvement",*, vol. 76, pp. 235-245, 1987.*

[15] A. Nilsonne, J. Sundberg. S. TcrnstrBm, A. Askcnfelt, “Measuring the rate of change of fundamental frequency in fluent speech during mental deprcssion”. *J. of the Acoustical Society of America,* vol. 83, no. 2, pp. 716 728, pp. 1988.

[16] F. Tolkmitt, H. Helfrich, R. Standke, K. R. Scherer, “Vocal indicators of psychiatric treatment effects in depressives and schizophrenics”, *J. of Common Disorders,* vol. 15, pp. 209-222, 1982.

[17] E. Szabadi, C. M. Bradshaw, J. A. Besson, “Elongation of pause-time in speech: a simple objective measure of motor retardation in depression”, *British J. of Psychiatry,* vol. 129, pp. 592-597, 1976.

[18] H. Ellgring and K.R. Scherer, “Vocal Indicators of Mood Change in Depression,” J. Nonverbal Behavior, vol. 20, no. 2, pp. 83-110, 1996.

[19] S. Kuny and H. Stassen, “Speaking Behavior and Voice Sound Characteristics in Depressive Patients during Recovery,” J Psychiatr Res., vol. 27, no. 3, 1993.

[20] J.C. Mundt, P.J. Snyder, M.S. Cannizzaro, K. Chappie, and D.S. Geraltsa, “Voice Acoustic Measures of Depression Severity and Treatment Response Collected via Interactive Voice Response (IVR) Technology,” J. Neurolinguistics, vol. 20, pp. 50-64, 2007.

[21] Ying Yang, F. Catherine, J. F. Cohn, “Detecting Depression Severity from Vocal Prosody”, *IEEE Trans. On Affective Computing,* vol. 4, no. 2, 2013.

[22] J. K. Darby, H. Hollien, “Vocal and Speech Patterns of Depressive Patients”, vol. 29, pp. 279-291, 1977.

[23] Theodoridis S., Koutroumbas K., "*Pattern Recognition, Fourth Edition*" Ac-ic Press, 2008

[24] H. M. Kalayeh, D. A. Landgrebe, “Predicting the required number of training samples, Pattern analysis and machine intelligence”, *IEEE transaction on Pattern Analysis and Machine Learning,* vol. 5, no.6, pp. 664-667, 1983.

CHAPTER VII

SUMMARY AND CONCLUSIONS

Current objective metrics for differentiating between patients with suicidal predisposition and patients experiencing major depression rely solely on subjective clinical observation. The evaluation of psychological disorders has relied on the clinical judgment by trained listeners based on perceptual parameters such as loudness, pitch and the articulatory precision which indicates the timing speech pattern. A quantitative assessment might improve the diagnostic precision and thus allowing the high risk suicidal patients to be hospitalized and treated. More recently, the analysis of speech acoustic measures has allowed the possibility of performing the psychological assessment using a more quantitative and objective comparisons of speech patterns from different level of severity. Two main analysis performed in this thesis were the classification between high risk suicidal and depressed patients and the study of regression which has not been looked at before in the field of suicidality where we attempted to predict the HAMD and BDI-II ratings by means of speech acoustic measures.

The first manuscript presented in Chapter III is entitled "Analysis of Features Based on the Timing Pattern of Speech as Potential Indicator of High Risk Suicidal and Depression". The timing based features known as Transition Parameters and Interval PDF that are related to pauses and phonation in speech were analyzed for its ability to distinguish the high risk suicidal patients from the depressed patients. Since this feature is derived from a pattern based speech, it should not be affected by the acoustic content of the speech. This studies only used datasets that were collected from readings of a standard "rainbow passage" essay. Use of the leave-one-out procedure as a means to measure the classifier performance for all-data classification revealed a single speech timing based measure (transition from silent to voice) to be a significant discriminator with 74% and 72% correct classification for male and female speech from Database A, respectively. Multiple combinations of voiced Interval PDF and silence Interval PDF increase the accuracy up to 70%-90% for male patients and 79% for female patients. For male patients, using the trained features and classifiers on Database A and performed testing on Database B, results revealed up to100% detection of high risk speech in Database B. However,

analysis of classification using female automatic speech only worked well within subpopulations. Using only a single and/or two combinations of features (both Transition Parameters and Interval PDF) yielded the best classifier performance. The advantage of a small number of features suggests that the classification can be generalizable even when using small data set. These features were shown to be robust across data sets despite the less than ideal recordings conditions and different equipment used for each database.

The second manuscript presented in Chapter IV in entitled "Investigation on Acoustic Measures of Speech as a Potential Predictor for the Hamilton Depression Scale (HAMD) and Beck Depression Inventory (BDI-II)". In this study, a consistent pattern of significant and predictive validity of the regression model was shown and demonstrated through the feasibility of using acoustic speech features as a potential means to predict the clinical HAMD and BDI-II scores. Application of Multiple Linear Regression was effective for generating the model prediction on our databases. For Database B, predicting the HAMD ratings revealed minimal mean sum of absolute error (MAE) of approximately two scores or less for interview and reading speech from male and female patients. Model predictions were mostly found to significantly improve through the method of Sequential Backward Selection as opposed to Sequential Forward Selection when attempting to fit the regression with a suboptimal number of features. However, there are some cases where the Sequential Backward Selection method performed better than the Sequential Forward Selection method. For Database A, HAMD score predictions also yielded MAE of approximately two score or less using suboptimal speech features found by the method of SFS and SBS for male and female reading patients. However, the SFS method identified smaller number of suboptimal features compared to the SBS method. Also on Database B, regression analysis was performed between the speech features and BDI-II scores. Only the male patients' BDI-II scores were successfully predicted with a MAE of less than one. The encouraging performance of the predictors makes this research worthy of further investigation on a larger sample population. Above all, this procedure is practical for use in real applications and can be used during a standard clinical interview by having the recordings done in a normal closed room and without strict control on the recording environment.

In Chapter V entitled "Analysis of Classification based on Amplitude Modulation in the Speech of Depressed and High Risk Suicidal Male and Female Patients", the analysis of the characteristics of root mean square amplitude modulation (RMS AM) in the speech of depressed

and near-term suicidal patients was performed in order to determine its potential for discriminating between the two groups. The study is a partial replication and extension of the work by France [IEEE T-BME 47(7) (2000)], who reported an effective overall classification score of 77% for depressed and near term suicidal speech in male patients. The current database consists of interviews and passage-based readings by male and female patients. Statistical RMS AM measures include the maximum, range, variation, average, skewness, kurtosis and coefficient of variance. Analyses were performed using linear (LDA) and quadratic (QDA) discriminant analysis with three resampling methods of equal-test-train, jackknife and cross-validation. France's RMS AM results were partially replicated: France reported a combination of RMS AM range and coefficient of variance in male speech as significant features whereas the analysis in this paper identified a combination of RMS AM range and skewness as a significant discriminator for male reading speech. On the other hand, poor classification scores were demonstrated for male interview, female interview and female reading speech. It is however difficult to conclude the different findings between this study and France's study because of the dissimilar database where he used recordings that were known to have attempted suicide whereas this study has the advantage of a better recording device.

The small-scale study in Chapter VI in entitled "Comparison of the Significant Mean and Difference for Multiple Spectral Energy Band Ranges and Combinations". The analyses were divided into two sections according to the database. The first analysis used database where patients were labeled as high risk suicidal and depressed by the trained clinicians. Results show that multiple combinations of 8 PSD bands yielded statistically significant difference between the mean of HR and DP female patients. However, only probably statistical significant differences were observed between the HR and DP male patients. The second analysis used database that were labeled as high risk suicidal during the patients' initial recording session and the patients' condition were made unknown to the researcher for their next two recording sessions that were each collected a few days after receiving treatments. The patients' condition were demonstrated to improve significantly during the second recording session when using feature 4 PSD band 1 in male patients (interview and reading) and a combination of 4 PSD band 2:3 in female patients (interview and reading).

For future research, this dissertation offers a promising methodology and results that might be applicable for use in a clinical trial. These analyses could also be performed on patients

that were diagnosed with suicidal ideation and remitted (one who has been entirely free of the depression symptom). Different psychometric properties of measures in suicidal predisposition could also be tested using the regression analysis and speech features. Considering the difficulty to obtain these high risk suicidal speech databases, however more effort could be put in to collecting more data on suicidal predisposition speech. Besides that, a longitudinal study on a patient or a group of patients recovering from high risk suicidal to healthy and control state might be done as an alternative to distinguishing features in a cross-sectional study.