FATIGUE DETECTION USING VOICE ANALYSIS

A Dissertation submitted in fulfillment of the requirements for the Degree
of

MASTER OF ENGINEERING

in

Electronic Instrumentation & Control Engineering

Submitted by

Sonika
801351028

Under the Guidance of

Dr. M. D. Singh
Assistant Professor, EIED

2015

Electrical and Instrumentation Engineering Department
Thapar University, Patiala
(Declared as Deemed-to-be-University u/s 3 of the UGC Act., 1956)
Post Bag No. 32, Patiala – 147004
Punjab (India)

DECLARATION

I hereby certify that the work which is presented in dissertation entitled, **"FATIGUE DETECTION USING VOICE ANALYSIS"** in partial fulfillment of the requirements for the award of the degree of **Master of Engineering in Electronics Instrumentation And Control**, submitted to Electrical & Instrumentation Engineering Department of Thapar University, Patiala is as authentic record of my own work carried under the supervision of **Dr. M. D. Singh.** It refers others researcher's work which are duly listed in the reference section. The matter contained in this dissertation has not been submitted, neither in part nor in full to any other degree to any other university or institute except as reported in text and references.

Sonika
801351028

Place: Patiala
Date: 11/7/15

It is certified that the above statement made by the student is correct to the best of my knowledge and belief.

Date: 11/7/15

Dr. M. D. Singh
Assistant Professor
Electrical & Instrumentation Engineering Department
Thapar University, Patiala

Countersigned by:

Dr. Ravinder Agarwal
Head
Electrical & Instrumentation Engineering Department
Thapar University, Patiala

Dr. S. S. Bhatia
Dean (Academic Affairs)
Thapar University, Patiala

ACKNOWLEDGEMENT

First of all I wish to express my gratitude to **Dr. M. D. Singh, Assistant Professor,** Electrical and Instrumentation Engineering Department, Thapar University Patiala, who has been a tremendous mentor and teacher. This thesis would not have been possible without his constant encouragement and support. His constructive criticism and passion for perfection has had a profound impact in my life. I have always enjoyed his way of explaining things in a simple and elegant manner. I am truly very fortunate to have the opportunity to work with him. I found this guidance to be extremely valuable.

I am also thankful to our Head of Department, **Dr. Ravinder Agarwal** as well as PG coordinator, **Mr. Nirhowjap Singh, Assistant Professor**, Electrical and Instrumentation Engineering Department. I would like thank the entire faculty and staff of Electrical and Instrumentation Engineering Department and my friends who devoted their valuable time and help me in all possible ways towards successful completion of this work .I thank all those who have contributed directly or indirectly to this work.

Lastly, I would like to thank my parents for their years of unyielding love and encourage. They have always wanted the best for me and I admire their determination and sacrifice.

Sonika

Sonika

TU, PATIALA

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ANN	Artificial Neural Networks
AIFF	Audio interchange file format
ASR	Automatic speech recognition
CoV	Coefficients of variances
CD	Compact disc
cm	Centimetre
CNS	Central nervous system
CoV	Coefficient of variation
CSL	Computerized Speech Laboratory
CoMIRVA	Collection of Music Information Retrieval and Visualization Applications
CLAM	C++ Library for Audio and Music
dB	Decibels
DCT	Discrete cosine transform
DFT	Discrete fourier transform
ECG	Electrocardiogram
EEG	Electroencephalogram
EMG	Electromyogram
EOG	Electro-oculogram
EMA	Electro-magnetic articulography
EPG	Electropalatography
f0	Fundamental Frequency

F1	First formant frequency
F2	Second formant frequency
F3	Third formant frequency
F4	Fourth formant frequency
FEAPI	Feature Extraction plugin application programming interface
FDR	Fisher discriminant ratio
GSR	Galvanic skin response
GB	Gigabyte
HMM	Hidden Markov Model
HNR	Harmonic-to-Noise ratio
hr	Hours
Hz	Hertz
IDFT	Inverse discrete fourier tansform
kHz	Kilohertz
KSS	Karolinska Sleepiness Scale
KNN	K-Nearest Neighbours
Log	Logarithm
LTAS	Long term average spectrum
LPCM	Linear pulse-code modulation
LPCC	Linear predictive cepstrum coefficients
LFCC	Linear frequency-cepstral coefficients
LPC	Linear predictive coding
MRI	Magnetic resonance imaging

MFCC	Mel-frequency cepstrum coefficients
MP3	MPEG-1, audio layer 3
MIR	Music information retrieval
MATLAB	Matrix laboratory
mm	Millimetre
mL	Millilitre
Max	Maximum
Min	Minimum
NLD	Non-linear dynamics
PC	Personal computer
PPG	Photoplethysmography
PERCLOS	Percent Eye Closure
PCM	Pulse code modulation
PSD	Power spectral density
RIFF	Resource interchange file format
RMS	Root mean square
SVM	Support Vector Machine
SD	Standard deviations
s	Seconds
SPL	Sound pressure level
USB	Universal Serial Bus
$\bar{x}$	Mean

ABSTRACT

Fatigue is harmful to human health as it impairs the maximal cognitive and physical performance. Fatigue is a critical element in many professions. Human voice has inevitable dependence on fatigue. The main objective of this study is to estimate the fatigue in an individual by speech analysis. Non-intrusive fatigue measurement systems are required to accurately examine the attentiveness and concentration of a person prior to and in an ongoing critical mission or during life threatening activities (e.g. for pilots, drivers, neurosurgeons etc., fatigue is a critical element in their profession). The speech based fatigue measurement system is non-intrusive and has many advantages over other measurements techniques.

This study discusses fatigue and its affect on the human speech, factors affecting acquisition and analysis of speech data. In this study speech samples have been acquired individually from 14 healthy adults during 24 hours of sustained wakefulness. Speech samples were recorded using digital voice recorder in a quiet room. From each participant 14 samples were acquired and hence a database having a total of 196 samples was created. The different feature sets and different classifiers have been studied in order to improve the detection of fatigue.

Dedicated to

My Parents

CHAPTER 1

INTRODUCTION

1.1 FATIGUE

Fatigue is a state of extreme tiredness that induces changes in psychological and physiological functioning. It reduces the efficiency and willingness to work and harmful to human health as it impairs the maximal cognitive and physical performance. It is caused due to mental or physical exertion or illness.

Causes of Fatigue are: (a) sleep deprivation, (b) intense physical activity, (c) prolonged mental activity, (d) prolonged durations of mental stress and anxiety, (e) poor sleep quality, and (f) health troubles [1].

The affects of Fatigue are: (a) lack of energy and motivation, (b) reduced cognitive ability, (c) weakened communicate skills, (d) reduced performance level, and (e) reduced level of alertness [2,3].

Fatigue is broadly classified as mental fatigue and physical fatigue:

1. Mental fatigue

"Mental fatigue is a transient decrease in maximal cognitive performance resulting from strong or prolonged periods of cognitive activity" [4,5]. Numerous damaging chemicals are produced by the brain of a fatigued person, which blocks the nutrition channel. Consequently the neurons get suppressed and there is a decrease in flow of information and chaos occurs [6]. It is the temporary inability to maintain optimal cognitive performance [7]. It depends upon the cognitive ability of an individual, and also upon other factors, such as sleep quantity, sleep quality and overall health. It is also seen that mental fatigue decreases physical performance [8]. This can be risky while performing tasks that require a certain level of focus, such as a doctor operating on a patient.

2. Physical fatigue

"The Physical fatigue is the transient inability of a muscle to maintain optimal physical performance, and is made more severe by intense physical exercise" [9]. It depends on the level of physical fitness of an individual, and also upon sleep quantity, sleep quality and overall health. The mentally and physically fatigued state reduces the activity of central nervous system that adversely affects the cognitive information processing and attention level [10].

Many existing techniques for detecting fatigue and are broadly classified into contact and non-contact techniques:

1. Contact

Electrocardiogram [11,12], Electroencephalogram [1,3,13], Electromyogram [2], Electro-oculogram [2,12], Electro-magnetic articulography [14], Electropalatography [14], Photoplethysmography [12], Galvanic skin response [11].

2. Non-Contact or Non-obstructive

Visual features (Percent Eye Closure, eye blinking, yawning, head pose, facial expression etc.) [2,12], Speech features (Frequency, loudness, harmonic to noise ratio, cepstral coefficient, speech quality, formant etc.) [1,2], Ultrasound [14], Magnetic resonance imaging [15], Reflexes Analysis (e.g. Key-striking, Gripping etc.) [16].

Advantages of speech based fatigue measurement over other measurement techniques are as follows: (a) utilization of already existing hardware and software, (b) un-obstructive (c) free from sensors application and calibration efforts, (d) cost efficient, durable, and maintenance free, (e) even possible in darkness or where mobile devices cannot provide adequate visual feedback, (f) robust against different conditions of the environment and person-specific variations (e.g. luminous light, high humidity and temperature, wearing correction glasses, angle of face), and (g) advancement in technology enhances the speech recognition ability even in noisy environments [17].

1.2 MOTIVATIONS

Fatigue state is dangerous to human well-being and can lead to fatal accident. Non-intrusive fatigue measurement systems are required to accurately examine the attentiveness and concentration of a person prior to and in an ongoing critical mission or during life threatening activities. The results of the study done by researches in this field are significant to pilot, drivers, doctors, workers, employers, war fighters, public safety officials, air traffic control personnel and military officers who are concerned with managing fatigue over long duration assignments. Hence there is a raising interest in developing a non-invasive system that can be used to detect and manage fatigue in both health and workplace settings.

A great part of literature is based on electrode (i.e. intrusive) and visual based fatigue measurement techniques. Electrocardiogram [11,12], Electroencephalogram [1,13] and Visual based fatigue measurement (Percent Eye Closure, eye blinking, yawning, head pose, facial expression etc.) are the main techniques used to determine fatigue. Small empirical research has been done to detect the fatigue by speech characteristics. Fatigue is a critical element in many professions. Human voice has inevitable dependence on fatigue. The speech based fatigue measurement is of great importance and has many advantages over other measurement systems.

This thesis discusses the effects of sustained wakefulness which induces fatigue and describes the effects of fatigue on the central nervous system which can be revealed by analysing speech. The main objective of this study is to estimate an individual's fatigue by speech analysis. The previous work supports that the voice characteristics are directly related to the participant's level of attentiveness and most studies have focused on discrete characteristics of the participant's voice for the detection of fatigue and have examined only small phonetic feature sets.

In this study different features like Pitch, Energy, Formants, Mel-frequency cepstrum coefficients, Linear predictive coding, Mean, Standard deviation, Amplitude, Energy, Duration has been analysed. Subsequently the most significant features have been used in different classifiers like Support Vector Machine, K-Nearest Neighbours, and Artificial Neural Networks in order to detect the fatigue with high level of accuracy.

1.3 OBJECTIVES

Below are the main objectives of this thesis:

1. Building a database of active and fatigued samples to estimate an individual's fatigue by speech analysis. Experimental stimuli were selected carefully which included few words and phrases, each having its own significance.

2. Extracting different features (e.g. Frequency, loudness, harmonic to noise ratio, cepstral coefficient, speech quality, formant etc.) from voice database for analysis.

3. Reducing the features to further improve the accuracy and efficiency of fatigue detection system.

4. Applying different classifiers to obtain the best accuracy of the fatigue measurement system.

1.4 CHALLENGES

1. Most of the datasets used by other research groups are not publicly available. Hence, there was need to build a new dataset. Building a dataset is the first challenge for this project.

2. To build a database volunteers are required and to get their fatigue samples, participants and experiment assistants have to stay awake up to 24 hrs, which is a difficult task and this proves to be the second challenge for this project.

3. Feature extraction from each sample is a time taking and tedious process which proves to be the third challenge. It is quite helpful in finding the most useful set of features. The system's efficiency depends on various features used to represent the signals. These features extracted are required to be representative, reliable, and robust. Also, the features should be organized according to the classifier.

4. The classifier is the central part of the system. The classifier should be chosen very carefully as all the other modules need to be designed according to it.

1.5 BRIEF METHODOLOGY

Speech samples have been acquired individually from healthy adults for 24 hours of sustained wakefulness. Experimental stimuli will be explained to the participants verbally and provided in written form which included few words and phrase. Samples will be recorded using the recorder in a quiet room and each speech sample was segmented. Then feature extraction will be done to extract useful information from the input data. Afterwards the feature reduction will be done to further improve the accuracy. Finally, the classifiers will be applied and the accuracy of the fatigue measurement system will be calculated.

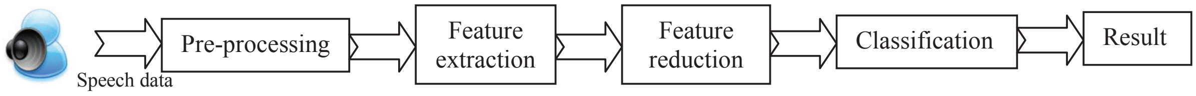


Figure 1.1: Block diagram of brief methodology

1.6 THESIS OVERVIEW

This thesis is organized as follows:

- Chapter 2 describes the basics of speech and its production and perception mechanism. It also describes how voice can be used as an indication to physical and mental state.
- Chapter 3 introduces the literature review in detail, describes the summary of the work done by the researchers in detecting the level of fatigue by speech analysis. It also includes the recording regime, stimuli and features for the detection of fatigue.
- Chapter 4 describes methodology in order to create a database which includes the process of data acquisition, hardware, software, stimuli, feature extraction, feature reduction and classifiers selection process and their detail description. It also includes the factors affecting acquisition and analysis of speech data for fatigue detection.
- Chapter 5 outlines the results and discussion of the fatigue detection using voice analysis.
- Chapter 6 discusses the conclusion and future scope.

CHAPTER 2

BASICS OF SPEECH SIGNAL

2.1 SPEECH

Speech sound is a wave of air that arises from complex mechanisms in the human body. It is assisted by three functional unit's viz. generation of air pressure, regulation of vibration and control of resonators. Basically, speech is the ability to communicate thoughts and feelings by articulating sounds.

The speech production mechanism is divided into two parts:

1. Phonation

The phonatory organs consist of lungs and larynx. Phonation acts as a voice production system, as it creates the voice source sounds. This is done by adjusting the air pressure in the lungs and vocal cords vibration at the larynx. The two organs collectively adjust the loudness, prosody, pitch and quality of the voice of speech.

2. Articulation

The articulatory organs consist of the lips, tongue, lower jaw and the velum. They give modulations or resonances to the voice source and also produce additional sounds for some consonants. The properties of the acoustic resonator depend on the position of the articulatory organs. The larynx also takes a part in distinctions of voiced/voiceless articulation [18]. The phonatory and articulatory systems regulate each other mutually in sequential manner for producing voice. The vocal tract can be viewed as an acoustic filter on sounds originating at the larynx as it enhances some frequencies and attenuates others [19].

Figure 2.1 describes the schematic diagram of human speech production system and the schematic representation of the complete physiological mechanism of speech production is shown in Figure 2.2 and the Figure 2.3 describes the descriptive signal level analogy.

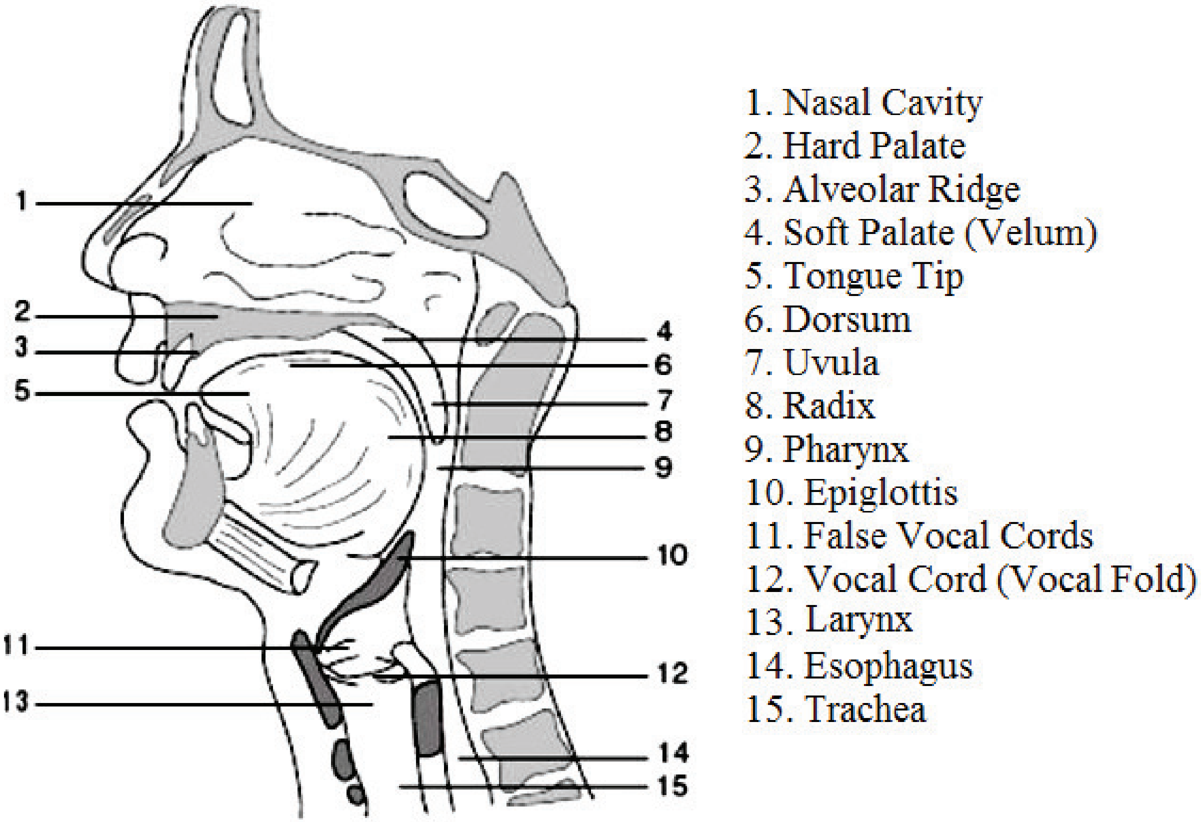


Figure 2.1: Schematic diagram of human speech production system [20]

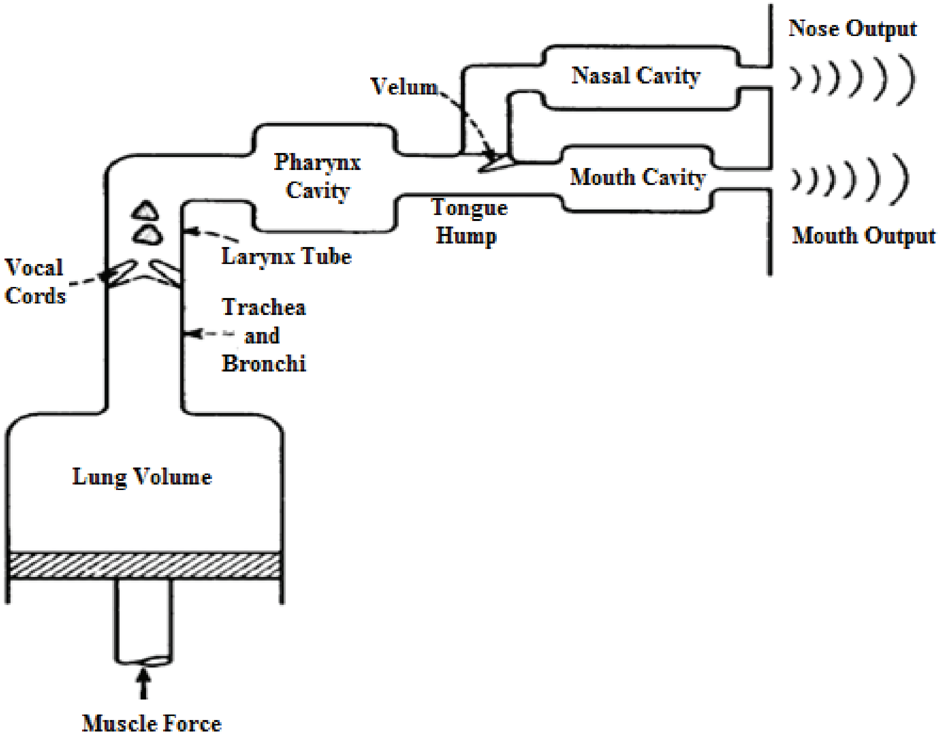


Figure 2.2: Schematic representation of the complete physiological mechanism of speech production [21]

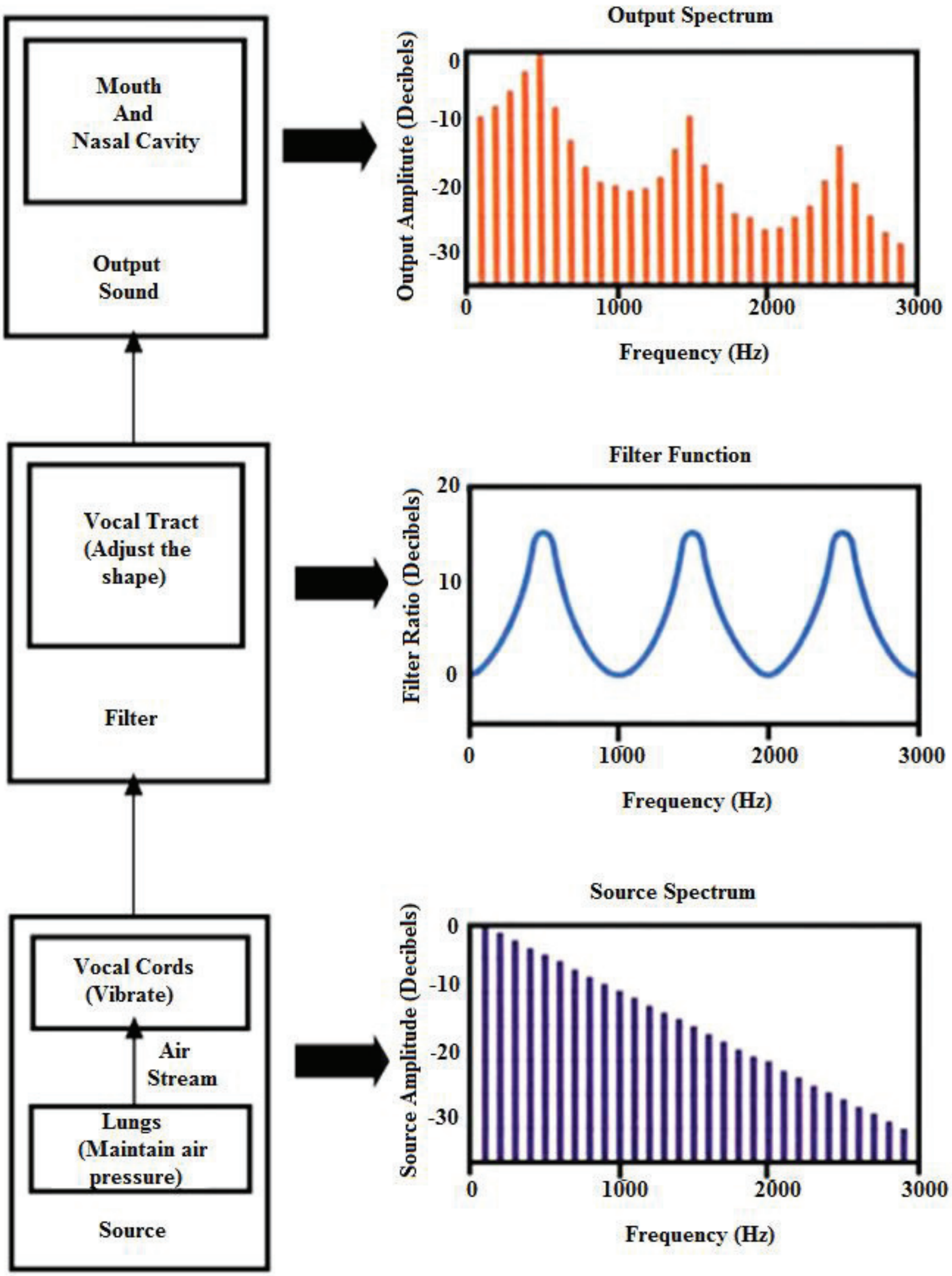


Figure 2.3: Descriptive Signal level analogy [19]

2.2 SPEECH COMMUNICATION PATHWAY

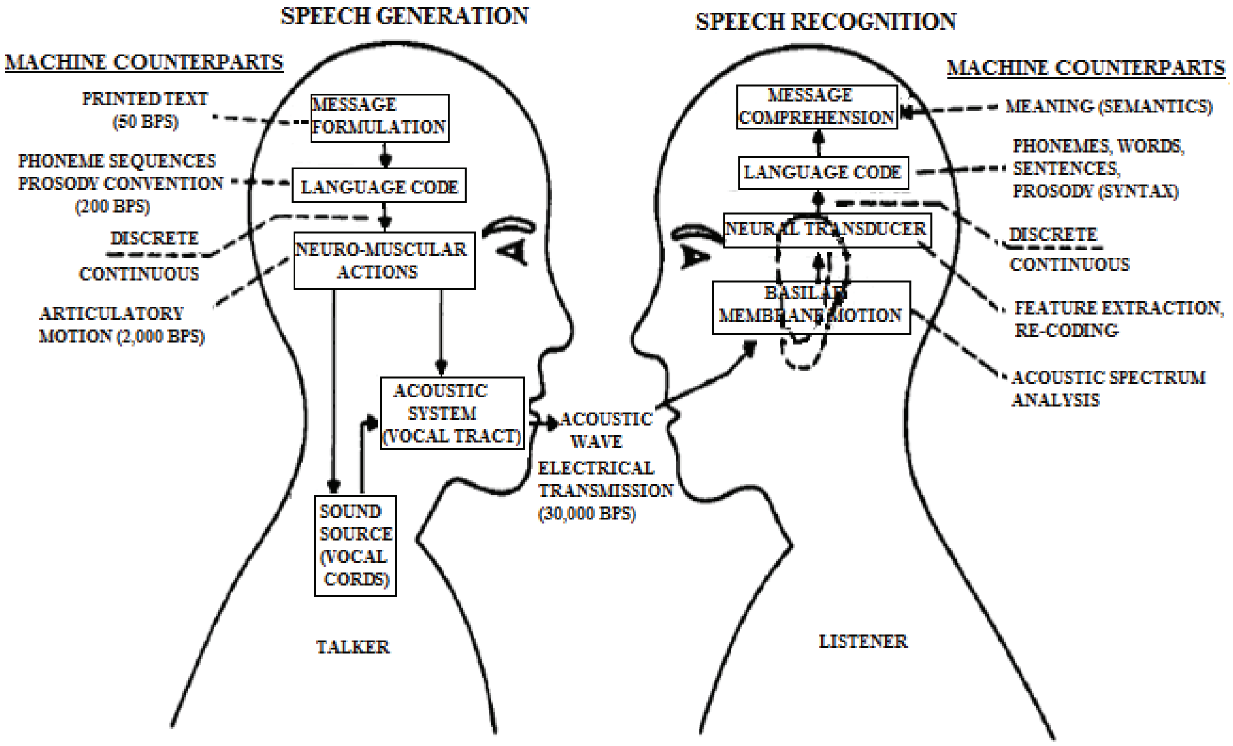


Figure 2.4: Speech production and perception mechanism [21]

Speech communication pathway is divided into two parts:

1. Speech production mechanism

The talker initiates the speech generation process by articulating a message in his brain that he wants to communicate to the listener through speech. In the next step the message is converted into a set of phoneme sequences corresponding to the sounds that further make up the words, along with prosody convention representing features like loudness, pitch and duration of the sound. Then the talker executes a series of neuromuscular commands which vibrate the vocal cords and simultaneously control the position of articulatory organs. Then the proper sequence of speech sounds is created and spoken by the talker, which provides acoustic signal as the final output.

2. Speech perception mechanism

The speech perception mechanism begins when the speech signal is generated by the speaker and transmitted to the listener. The acoustic wave propagates along the baseline membrane in the inner ear. This provides acoustic spectrum analysis of the input acoustic wave. The spectral signal is converted into activity signals on the auditory nerve by neural transduction. A neural transduction is a features extraction process. The activity signals are further converted into language codes as shown in Figure 2.4. The brain processes the language code into messages and message comprehension is achieved [20].

2.3 CLASSIFICATION OF SPEECH SOUNDS

Speech sounds are classified as:

1. Voiced

The voiced region in speech is nearly periodic in nature. The air coming out of lungs gets interrupted by the vibrating vocal cords periodically, this result into glottal wave. The glottal wave (airflow) is modulated by the articulatory organs resulting in the voiced speech. It has relatively high energy, less no. of zero crossings and more correlation among successive samples than unvoiced and silence regions in speech. The voiced region has periodic nature in time domain and harmonic structure in frequency domain. It has more energy in low frequency region.

2. Unvoiced

The unvoiced region in speech has random noise like nature (i.e. non-periodic). The air coming out of lungs does not get interrupted by the vibrating vocal cords. The partial or total closure occurs along the length of vocal tract and results in obstruction of airflow narrowly or completely. The airflow results in stop or frication excitation and modulated by the articulatory organs resulting in the unvoiced speech. The unvoiced speech has non-periodicity and noise-like waveform in the time domain and don’t have the harmonic structure in the frequency domain. The spectrum has more energy in the high frequency region.

3. Silence

The voiced and unvoiced speech is separated by the silence region. There is no air supplied to the vocal tract and hence no speech output during silence region as shown in Figure 2.5. Silence is the integral part of speech signal with low energy, more no. of zero crossing and no correlation among successive samples than voiced and unvoiced regions in speech. The silence region doesn’t have signal in the time domain and as well as no spectral information in the frequency domain [22].

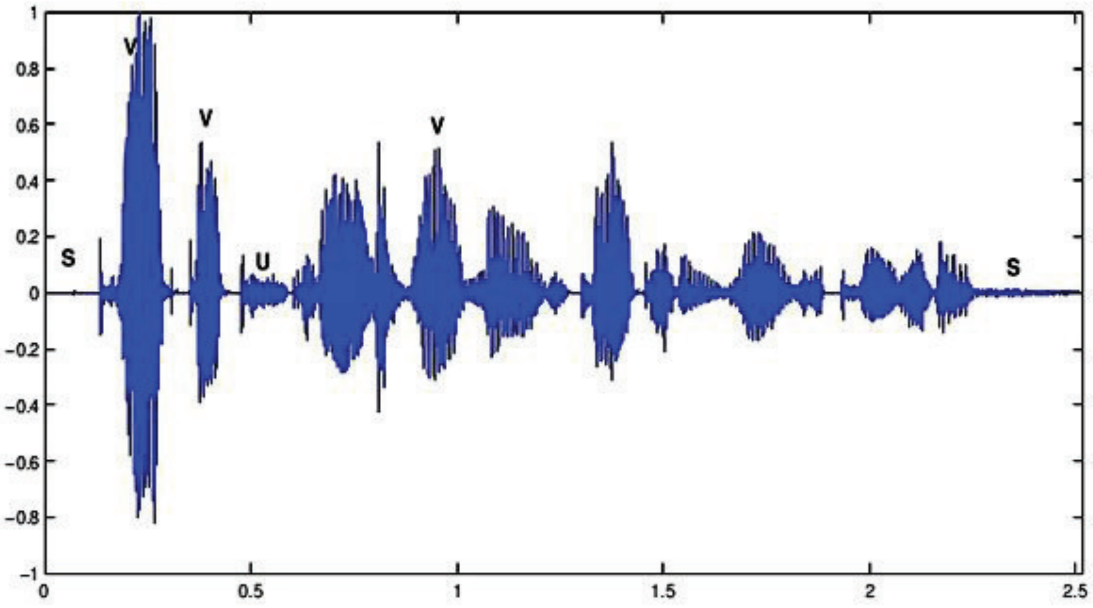


Figure 2.5: Amplitude waveform of the speech signal labelled with V (Voiced), U (Unvoiced) and S (Silence) regions [20]

2.4 VOICE TO DETECT FATIGUE

Voice is an indication of physical and mental state. Table 2.1 is the sleepiness hypothesis relating physiological changes with the psychological state.

Table 2.1: Fatigue induced physiological and psychological changes [2]

Effects on physical state	**Effects on mental state**
Respiration: 1. Reduced muscle stress causes: ● Decreased subglottal pressure ● Sluggish and periodic respiration Phonation: 1. Decreased muscle tension causes: ● Decrease in vocal fold tension, stiffness and viscosity ● Increase in vocal fold elasticity	Reduced cognitive ability: ● Adversely affects the speech planning and neuromuscular motor coordination processes

2. Decreased body temperature: ● Change in viscosity and elasticity of vocal fold Articulation/ resonance: 1. Reduced muscle stress causes: ● Lowering velum ● Softening of vocal tract walls and hardening of pharynx 2. Reduced body temperature: ● Decreased heat conduction changes the laminar flow and turbulence Radiation: 1. Reduced muscle stress causes: ● Reduced lip spreading and facial expressions	

CHAPTER 3

LITERATURE REVIEW

3.1 INTRODUCTION

“Mental fatigue is a psychobiological state caused by prolonged periods of demanding cognitive activity”[5]. It is also seen that mental fatigue decreases physical performance [5,6]. A number of psycho physiological parameters have been used to detect fatigue but EEG has been the most accurate technique [3,7]. But due to the sensor application and need of calibration, EEG proves to be complex, uncomfortable and obstructive technique. Therefore, techniques are required which are non-obstructive and free from sensor applications to accurately examine the individual’s fatigue state [17]. Human voice has inevitable dependence on fatigue. Acoustic features extracted from voice contain significant amount of information about the participant’s fatigue state [2,3,6]. Rogado *et al.* [11] used fatigue detection technique to analyse the decreasing attentiveness prior to a driver falling asleep while driving a car.

The previous work supports that the voice characteristics are directly associated to the participant’s level of attentiveness and most studies have focused on discrete characteristics of the participant’s voice for the detection of fatigue.

Whitmore *et al.* [23] discussed that the results of the voice analysis are quite similar to results of the cognitive and subjective tests of alertness. They also stated that quality of speech follows the circadian trend, as it is at its best during normal working hours and the worst in usual sleeping hours and found the significant changes in fundamental frequency and word duration.

Bard *et al.* [24] concluded that the speech duration and work performance measures indicate repercussions of medication and sleep deprivation.

Harrison *et al.* [25] found a loss of intonation (i.e. monotonic and flattened voices) between sleep and no sleep condition.

Greeley *et al.* [26] found that the formant frequencies are related to participant’s attentiveness, which is directly associated to his/her fatigue level and MFCC value changes with the participant’s fatigue level.

Greeley *et al.* [27] noted that the Subject's voice for speech sounds which need a greater average air flow varied in synchrony with both level of fatigue and the time of sustained wakefulness. It implies that the sounds which need a greater average airflow are more sensitive to fatigue. They also analyzed the changes in the mathematical representation of the entire speech (i.e. the Cepstral components).

Krajewski *et al.* [28] achieved 80.0% accuracy with simple linear classifier (LDA) and 79.4% with artificial neural network classifier and after using an ensemble classification strategy recognition rate of 88.2% was achieved.

Krajewski *et al.* [29] achieved recognition rate of eighty three percent in use of [a:] vowel by studying the different levels of performance between an active and fatigue state of a person.

Krajewski *et al.* [30] concluded the SVM is the best model for the static acoustic feature vector and it had a recognition rate of eighty six percent in predicting fatigue states (i.e. micro-sleep endangered sleepiness stages).

Dhupati *et al.* [31] analysed that the duration of pauses between words seemed to increase and response time increased with increase in fatigue.

Vogel *et al.* [32] found the effect of fatigue on speech to be strongest just before sunrise (after 22 hours) and analyzed that the formant patterns remained invariant despite increasing levels of fatigue, with the exception of F4 and F4 variation (SD/CoV). They also found that total sample duration, total speech time and mean pause length increased significantly when levels of fatigue were amplified.

Zhang *et al.* [33] discussed that MFCC focused on the auditory mechanism and LPCC focused more on the sound channel mode and each has their own advantages. They also found that MFCC is superior to LPCC.

Krajewski *et al.* [34] had shown that when the level of fatigue increased in the subject, it influenced the speed production mechanism to generate nonlinear aerodynamic phenomena. They also concluded that the non-linear dynamic features provide additional information regarding the dynamics and structure of speech of fatigue person in comparison to the Speech emotion recognition feature set.

Krajewski *et al.* [35] had shown that the best performance for male and female speakers on the phonetic and the NLD feature set were achieved by bagging procedure.

Rashwan *et al.* [36] had shown that the Hidden Markov Model classifier was experimentally proven to be a good solution for car driver fatigue monitoring and also performed better than SVM.

Table 3.1: Summary of literature review in chronological order

Author	Recording regime	Stimuli	Features
Whitmore *et al.* [23]	• 36 hours of sustained wakefulness. • Recordings were made after every 3 hours.	Two sentences: "Futility Magellan, this is (x y); The time is hr:min Zulu". x : participant's rank, y: name, and hr:min: time.	Fundamental Frequency (f0);Word duration
Bard *et al.* [24]	64 hours of sustained wakefulness.	Dialogue recorded during map task.	Speech length; Pause length
Harrison *et al.* [25]	• 36 hours of sustained wakefulness. • Recordings were made between 8-9 hours and 32-33 hours.	Reading a passage for approximately 3 minutes.	Intonation; Pitch
Greeley *et al.* [26]	• 34 hours of sustained wakefulness. • Recordings were made 6 times in a span of 34 hours (10:00, 16:00, 22:00, 04:00, 10:00 and 16:00).	1] list of 37 words 2] list of 31 words	Formant frequencies; Mel-frequency cepstrum coefficients (MFCC)
Greeley *et al.* [27]	Participants are divided into 3 groups: Group 1: 1] 34 hours of sustained wakefulness. 2] Recordings were made after every 6 hours. Group 2: The testing period consisted of 3 nights, where the participants were	Group 1: subjects recited 31 unrelated words. Group 2: subjects recited eight fixed phrases. Group 3: subjects recited eight fixed phrases	Cepstral Coefficient

	allowed to sleep 2 hours each on their second and third nights. Group 3: Recordings were made after every 2 hours during a normal workday.		
Krajewski *et al.* [28]	Sustained wakefulness during normal sleeping hours (8.00 p.m. to 4.00 a.m.).	German phrase, in form of a statement: “Ich suche die Friesenstraße“ [“I´m searching for the Friesen Street”]”.	Frequencies, bandwidths, and amplitudes of the F1-F5 formants; f0; Intensity; Jitter; Shimmer; Short-term fluctuations in energy; Mean; Standard deviation; Maximum; Minimum; Range, Positions and values of Maxima and minima; Harmonic-to-Noise ratio (HNR); Frequencies and amplitudes of the first 2 harmonics; MFCC.
Krajewski *et al.* [29]	• Sustained wakefulness during normal sleeping hours (8.00 p.m. to 4.00 a.m.). • Recordings were made 4 times in a span of testing period (8.30 p.m., 9.00 p.m., 3.00 a.m., and 3.30 a.m.).	Sustained phonation of the “German vowel [a:]” for 2 second.	Intensity;f0; Linear predictive coding (LPC); Formants (Position and bandwidth); MFCC ; Linear frequency-cepstral coefficients (LFCC); Harmonics-to-noise ratio; Voiced segments duration; unvoiced segments duration.
Krajewski *et al.* [30]	Sustained wakefulness during normal sleeping hours (01.00 - 08.00 am).	Pilot-air traffic controller communication: "Cessna nine three four five Lima, County tower, runway two four in use, enter traffic pattern, report left base, wind calm, Altimeter three zero point zero eight”.	Perceptual and signal processing: f0, Formants, Cepstral Coefficients; Prosody: Pitch, Intensity, Rhythm, Pause Pattern, Speech Rate; Articulation; Speech Quality.

Dhupati *et al.* [31]	36 hours of sustained wakefulness.	“*Now the time is* ______”.	Voiced duration; Unvoiced duration; Response time; MFCC; EEG Based Parameters: Alpha and Theta band energy.
Vogel *et al.* [32]	• 24 hours of sustained wakefulness. • Recordings were made every 4 hours.	Automated and extemporaneous tasks, sustained vowel and a passage.	Timing, Intensity and spectral tilt; Frequency: f0, Formants(F1-F4), Standard deviations (SD) and Coefficients of variances(CoV) of frequency, Power (alpha ratio)
Zhang *et al.*[33]	Recordings were taken at four different times (4:00 a.m., 10:00 a.m., 4:00 p.m., and 10:00 p.m.).	Six Chinese vowels	MFCC; LPCC
Krajewski *et al.* [34]	Sustained wakefulness during normal sleeping hours (8.00 p.m. to 4.00 a.m.).	A long vowel [o:] extracted from a German phrase: “Rufen Sie den N[o:]tdienst” (“Please call the ambulance”).	Non-linear dynamics (NLD) features: State space features, Fractal features and entropy features; Phonetic features.
Krajewski *et al.* [35]	Sustained wakefulness during normal sleeping hours (8.00 p.m. to 4.00 a.m.).	Sustained phonation of the vowel /a:/ for three to five second.	NLD features: State space features, Fractal features and entropy features; Phonetic features.
Rashwan *et al.* [36]	Recording were made two times, one in the early morning and another in the late evening after a working day.	The participant is asked to make 2 phone calls.	MFCC; Statistical: Mean Variance, Median, Max, Min; Heart rate; Steering wheel, gas, Clutch, and brake pedals positions.

3.2 SUMMARY OF LITERATURE REVIEW

The previous work supports that the speech characteristics are directly related to the participant’s level of attentiveness. Whitmore *et al.* [23] discussed that quality of speech follows the circadian trend, as it is at its best during normal working hours and the worst in usual sleeping hours. Harrison *et al.* [25] found a loss of intonation (i.e. monotonic and flattened voices) between sleep and no sleep condition. Greeley *et al.* [27] noted that the subject's voice for speech sounds which need a greater average air flow varied in synchrony with both level of fatigue and the time of sustained wakefulness. Vogel *et al.* [32] found the effect of fatigue on speech to be strongest just before sunrise (after 22 hours). Krajewski *et al.* [34] had shown that when the level of fatigue increased in the subject, it influenced the speed production mechanism.

Most studies have focused on discrete characteristics of the participant's voice for the detection of fatigue. Whitmore *et al.* [23] found significant changes in f0 (fundamental frequency) and word duration with the participant's fatigue level. Greeley *et al.* [26] found that the MFCC value changes with the participant's fatigue level. Greeley *et al.* [27] analyzed the changes in the mathematical representation of the entire speech (i.e. the Cepstral components). Dhupati *et al.* [31] analysed that the duration of pauses between words seemed to increase and response time increased with increase in fatigue. Vogel *et al.* [32] found that the total sample duration, total speech time and mean pause length increased significantly when levels of fatigue were amplified.

Krajewski *et al.* [28] achieved 80.0% accuracy with simple linear classifier (LDA) and 79.4% with artificial neural network classifier and after using an ensemble classification strategy recognition rate of 88.2% was achieved. Krajewski *et al.* [29] achieved recognition rate of 83% in use of [a:] vowel by studying the different levels of performance between an active and fatigue state of a person. Krajewski *et al.* [30] concluded the SVM is the best model for the static acoustic feature vector and it had a recognition rate of 86% in predicting fatigue states (i.e. micro-sleep endangered sleepiness stages). The present thesis discusses different feature sets and different classifiers in order to improve the detection of fatigue and accuracy as noted from literature review discussed above.

CHAPTER 4

MATERIAL AND METHODOLOGY

4.1 FACTORS AFFECTING ACQUISITION AND ANALYSIS OF SPEECH DATA

The researcher needs to ensure that each stage and each component for acquisition and analysis of speech data are appropriate, so that they can adequately address their objective and optimize their result for accuracy and precision.

The factors that affect the acquisition and analysis of speech data for fatigue detection are:

- Hardware selection
- Recording software
- Microphone preference
- Impact of noise
- Sampling rate
- Number of speech samples
- File format
- Hours awake and time of recording
- Word spotting
- Overall system features viz. portability, cost efficiency, easy to use etc.

4.1.1 HARDWARE SELECTION

Various hardware configurations can be used to record the speech samples having low to high quality. The hardware selection for the most apt configuration depends on a number of features (e.g. Quality of recording, portability etc.) which decide the level of fidelity. Table 4.1 describes various hardware specifications of the recorders used frequently and are arranged in order of high to low quality. Digital voice recorder proves to be the best suitable choice for recording among other recording devices in terms of quality and portability.

Table 4.1: Specifications of recording devices [37]

Recording Device	Quality of recording	Portability
Hard disk recorder	Highest	Low portability (require additional equipments)
Digital voice recorder	High	High
Flash Recorder	High	Medium
Mini disc	Medium	High
Computer /Laptop	Medium	Low
MP3 recorders	Low	High
Telephone	Low	Medium
Voice over internet	Dependent on internet connection quality	Low (Requires a computer)

4.1.2 RECORDING SOFTWARE

The different recording software provide different features (e.g. recording, analysing files and trimming) and different options viz. input mode (i.e. mono or stereo), mic sensitivity, sampling rate, file format, noise cut, low cut and output display for controlling the hardware setting. Table 4.2 describes different recording software used for analysis and recording purpose. Most of the recording software mentioned in the table are available to us free of cost except MATLAB, Computerized speech laboratory, Dr Speech and TF32 which are commercially available. Praat and MATALB are the most dominant and significant software used frequently for analysis of speech among all other software as mentioned below.

Table 4.2: Specifications of recording software [37]

Software	Analysis	Recording
Audacity		✓
Sonic Visualiser	✓	
EMU	✓	
CoMIRVA	✓	
jMIR	✓	
Praat	✓	✓
Speech Filing System	✓	✓
CLAM	✓	
Wavesurfer	✓	
FEAPI	✓	
TF32	✓	
Computerized Speech Laboratory	✓	✓
MATLAB	✓	
Dr Speech	✓	✓

4.1.3 MICROPHONE PREFERENCE

The microphone specifications and configurations are the most powerful features that determine the quality and reliability of the speech signal. The microphone offers different specifications viz. polar pattern (cardioids (Unidirectional), omnidirectional), impedance (low, medium, high), sensitivity (low, medium, high), type (condenser, dynamic), frequency response (wide, narrow), connection (XLR, USB, 3.5mm), and positioning (i.e. distance, angle and type (viz. head mounted, Table top and lapel)). Table 4.3 describes specifications of different microphones. Different options are available from which the most optimal microphone can be chosen according to the specification and configuration required.

Table 4.3: Specifications and optimal configuration for microphones [37]

Specifications	Options	Optimal configuration
Directionality (i.e. Polar pattern)	Cardioid, omnidirectional	Cardioid microphones (as it receive less ambient noise)
Impedance	Low (<600), medium (600-10000), high (>10000)	Low impedance (provides good quality speech signal)
Sensitivity	Low, medium , high	High sensitivity (picks up more quieter voices)
Frequency response	Few Hz to thousand Hz	Should be in the range of human voice (i.e. 20 Hz to 20 kHz)
Power supply	Dynamic (electret), condenser	Condenser (more sensitive and appropriate for speech analysis)
Connectors	XLR, 3.5mm, USB	USB (bypasses the sound card)
Positioning	Table top, head-mounted, lapel	Head-mounted (prevents the changes in amplitude due to head movement)

4.1.4 IMPACT OF NOISE

The signal can be altered due to external factors like environmental or additive noises. The advancement in technology improves the speech recognition ability even in noisy environments.

4.1.5 SAMPLING RATE

Analog to digital conversion involves sampling and quantization to convert a continuous physical quantity into digital number that represents the amplitude of physical quality at different instant of time. “The sampling rate is the number of the samples per second”. “The quantization level is the number of discrete levels of signal amplitude corresponding to the number of binary bits in each digital number". The sampling rate and quantization level determine the quality of the speech signal recorded. The Nyquist theorem is the main principle which helps obtaining the optimal sampling rate. It states that "the number of samples needed to faithfully represent a signal is twice the highest frequency of interest present in the signal" [38].

4.1.6 FILE FORMAT

File format (i.e. the mode of data storage) also plays a very important part in the recording process. The Resource interchange file format (RIFF)(e.g., .wav) and Audio interchange file format (AIFF)(e.g., .aif) save data in uncompressed form using pulse code modulation (PCM) format. PCM format has a sampling rate of 44.1 kHz and 16 bit quantization. The quality and the fidelity of the recorded signal are ensured by this method [37].

4.1.7 HOURS AWAKE AND TIME OF RECORDING

Whitmore and Fisher [23] and Roth *et al.* [39] have observed a strong circadian trend, as the speech production system offer best performances during regular working hours, and worst during usual sleeping hours. "Circadian means processes occurring periodically approximately in a 24-hour interval." Sleep cycle follows the circadian trend and if sleep is interrupted it causes fatigue. That is why strong fatigue is noticed during usual sleeping hours. Therefore, the time of recording is quite vital in analysis of speech data for fatigue detection.

4.1.8 WORD SPOTTING

All the phonemes in the human speech are not affected equally by fatigue. Sounds which need a greater average airflow are more sensitive to fatigue. Airflow is directly related to driving pressure and resistance. The lungs generate the driving pressure and respiratory tract produces the resistance. The following equation describes the relationship of airflow to driving pressure and airway resistance [27]:

$$F = P / R \tag{1}$$

In the above equation F represents airflow, P represents driving pressure, and R represents airway resistance. Subject's voice varies in synchrony with both level of fatigue and the time of sustained wakefulness [27].

Table 4.4 describes the average airflow required to generate the speech sounds. Sounds /t/ (as in tea) and /p/ (as in pea) are more sensitive to fatigue as they require greater average airflow for their pronunciation. Greater average airflow can be achieved to generate speech sounds when the lungs generate high driving pressure and respiratory tract produces less airway resistance.

Table 4.4: Average airflow required to generate the speech sounds [27]

Average air flow required to generate the speech sounds	
Sound	**Average Airflow(mL/s)**
/t/	968
/p/	933
/d/	525
/g/	372
/l/	133
/m/	168
/z/	159

4.2 METHODS

4.2.1 DATA ACQUISITION

Speech samples were recorded using SONY 4GB UX series digital voice recorder in a quiet room. The recorder used proved to be the best suitable choice for recording among other recording devices in terms of quality and portability as also mentioned in the Table 4.1.

Digital voice recorder hardware settings are:

- Recording format: LPCM (44.1 kHz, 16-bit)
- Frequency response: 50- 20,000 Hz
- Playback format: WAV
- Sensitivity: Medium
- PC connectivity: Direct USB
- Mic sensitivity: Medium
- Functionality: Noise cut Filter

Recorder was positioned at specific distance from mouth during the experiment. Recorder had a USB port through which data was transferred to the PC for further processing. Figure 4.1 shows the Digital voice recorder used during the experiment.

Figure 4.1: 4GB UX Series Digital Voice Recorder

4.2.2 PARTICIPANTS

Speech samples were acquired individually from 14 healthy adults over 24 hours of sustained wakefulness. From each participant 14 samples were acquired (i.e. 7 active and 7 fatigued samples). In this manner the total database of 196 samples was created. They were asked to study during the experiment. They were advised not to perform any high arousal activity and were not allowed any intake of alcohol, tea or coffee during the experiment. They were not on any medication also.

4.2.3 STIMULI

Experimental stimuli were explained to the participants verbally and provided in written form. Experimental stimuli were selected carefully which included few words and phrase, each having its own significance.

The experimental stimuli were as follows:

- Hello
- My name is
- The quick brown fox jumps over the lazy dog
- Teeth
- Tata
- Papa
- Pen

‘Hello’ was chosen as a stimulus as it is the most commonly used vocabulary word when we start a conversation. ‘My name is.’ was used for the identification purpose of the participant and it is also commonly used phrase during conversations. ‘The quick brown fox jumps over the lazy dog’ phrase was chosen as it contains all of the letters of the English alphabet. ‘Teeth, Tata, Papa, Pen’ were used as stimuli as these words are formed by the sounds of /t/ and /p/ which need a greater average airflow and hence are more sensitive to fatigue [27].

4.2.4 FEATURE EXTRACTION

Samples were recorded using the recorder in a quiet room and each speech sample was segmented and analyzed using PRAAT and MATLAB (using MIR toolbox). Praat and MATALB are the most dominant and significant software used frequently for analysis of speech among all other software as mentioned in the Table 4.2.

1. MATLAB

“MATLAB® is a high-level language software and interactive tool for numerical computation, visualization, and programming. MATLAB can be used to analyze data, develop algorithms, and create models and applications”.

2. PRAAT Software

Praat is Dutch word for “talk”. It is a software package that can be used to record, visualize and analyse the speech signal. Figure 4.2 describes the Praat software.

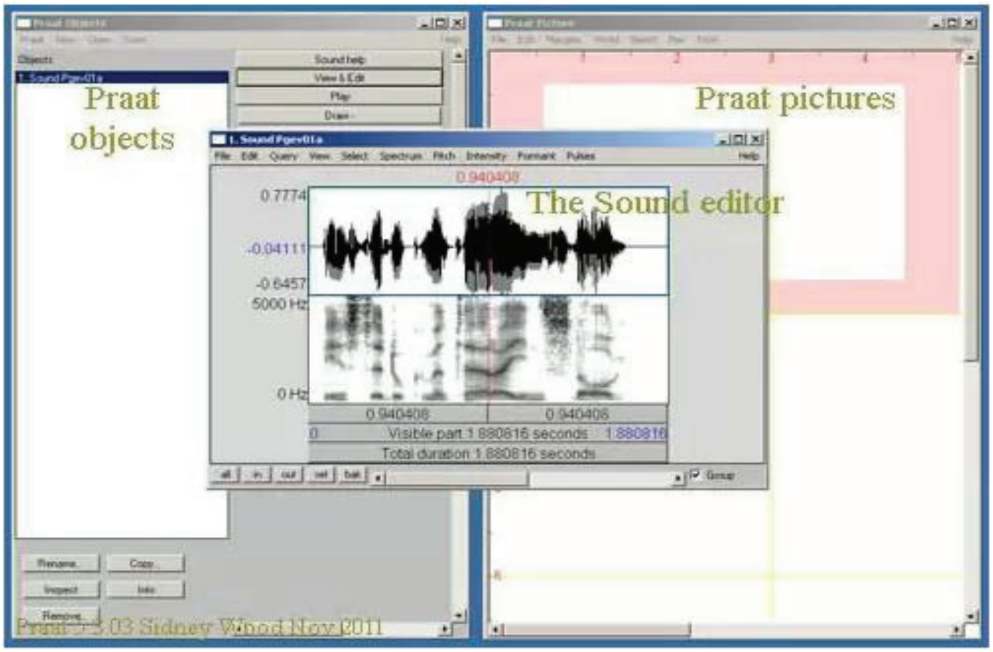


Figure 4.2: PRAAT Software [40]

Silences were removed from the start and end of the samples. Feature extraction is a procedure that extracts the useful information and discards the irrelevant and redundant information from the input data. It is necessary to extract those particular features from the speech signal that would reveal useful information regarding fatigue detection. It also improves the performance of classifiers. To further improve the accuracy, feature reduction was done.

Features extracted from the speech samples were: Intensity (loudness)-(Max., Min., Mean), Fundamental frequency (pitch) -(Max., Min., Mean), Formants and bandwidth (F1-F4; B1-B4), Speech Rate, Speech duration, Mean, Root mean square, Standard deviation, Variance, Energy, Power, Low energy, Roll-off, Spectral Centroid, Spectral Spread, Spectral Skewness, Spectral kurtosis, Spectral flatness, Entropy, Average Silence Ratio, Pulses, Median, Maximum and minimum value, Speech Quality, Voiced/unvoiced duration, Harmonics-to-noise ratio, Zero crossing Rate, Jitter, Shimmer, Mean autocorrelation, Mel frequency cepstrum coefficients, No. of voiced breaks, Pause Pattern, Long term average spectrum, Power spectral density, Centre of gravity, Spectral Tilt, Sound pressure level, Band energy, Band density, Band energy difference, Band density difference, Linear predictive cepstrum coefficients. Table 4.5 describes different features of the speech signal that are analysed in the literature with their reference.

Table 4.5: Features of the speech signal with reference

Feature	Reference
Fundamental frequency (f0)	[23,25,28,29,30,32]
Formant position (F1-F6)	[26,28,29,30,32]
Mel frequency cepstrum coefficients (MFCCs)	[26,28,29,31,36]
Formant bandwidth (Fbw1–Fbw6)	[28,29,30,32]
Duration of voiced–unvoiced segments	[24,30,31,33]
Loudness(Intensity)	[28,29,30,32]
Cepstral coefficients	[27,30]
Harmonics-to-noise ratio (HNR)	[28,29]
Mean	[28,36]
Variance	[32,36]
Standard Deviation	[28,32]
Non-linear dynamics (NLD) features	[34,35]
Linear predictive cepstrum coefficients (LPCCs)	[33]
Shimmer	[28]
Jitter	[28]
Short-time energy	[28]
Speech duration	[23]
Entropy	[28]
Maxima and minima	[28]
Median	[36]
Speech Rate	[30]

In this thesis those particular features are extracted from the speech signal that could reveal useful information regarding fatigue detection. Apart from the features that are analysed in literature some other features like Root mean square, Power, Low energy, Roll-off, Spectral Centroid, Spectral Spread, Spectral Skewness, Spectral kurtosis, Spectral flatness, Average Silence Ratio, Pulses, Zero crossing Rate, Mean autocorrelation, No. of voiced breaks, Pause Pattern, Long term average spectrum, Power spectral density, Centre of gravity, Spectral Tilt, Sound pressure level, Band energy, Band density, Band energy difference, Band density difference are also analysed for fatigue detection from voice.

FEATURES AND THEIR DESCRIPTION

1. Fundamental frequency (f0)

Fundamental frequency is defined as the lowest frequency of a periodic waveform. Fundamental frequency is acoustic equivalent to [30]:

- Pitch
- Rate of vocal fold vibration
- Speech melody indicator
- Maximum of the autocorrelation function

2. Loudness (Intensity)

Loudness is an attribute of sound. It is defined as the intensity of auditory sensation produced [30].

3. Harmonics-to-noise ratio

Harmonic-to-noise is the ratio between harmonic (voiced segment) and aperiodic (unvoiced segment) signal energy. It is an indicator of breathiness [30].

4. Formant frequency

Formant frequencies are the resonant frequencies of the vocal tract. It depends on the actual shape of the vocal tract [30]. It represents:

- Spoken content

- Speaker characteristics
- Spectral maxima

5. Formant bandwidth

Formant bandwidth represents [30]:

- Vocal tract shape
- Energy loss of speech signal due to physiological changes (viz. vocal tract elasticity etc.)

6. Voiced/unvoiced duration

Voiced/unvoiced duration represent temporal speech rhythm characteristics such as [30]:

- Speech rate
- Pause structure

7. Cepstral coefficients

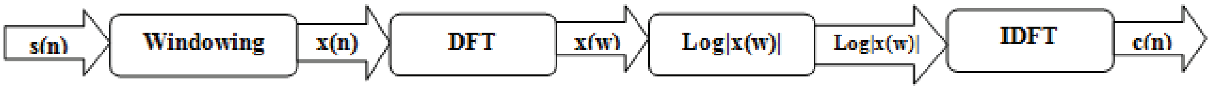


Figure 4.3: Block diagram representing computation of cepstrum

Figure 4.3 describes the steps for computing the cepstrum.

s(n): Speech sequence

x(n): Windowed frame

x(w): Spectrum of the windowed sequence x(n)

Log|x(w)|: Log magnitude spectrum obtained by computing logarithm of the |x(w)|

c(n): Spectrum for the voiced sequence s(n)

Speech sequence can be expressed as

$$s(n) = e(n) * h(n) \tag{2}$$

e(n): Excitation sequence

h(n): Vocal tract filter sequence

Frequency domain representation of s(n) is S(w) and it can be expressed as

$$S(w) = E(w) . H(w) \tag{3}$$

Spectrum for the voiced sequence s(n) and can be expressed as

$$c(n) = IDFT(\log|S(w)|) = IDFT(\log|E(w)|+\log|H(w)|) \tag{4}$$

Speech is composed of excitation source and vocal cord components. Cepstrum analysis leads to computation of a discrete number of coefficients known as cepstral coefficients. The changes in the entire speech production system can be examined with the analysis of these coefficients. It can be used to analyse filter (system) and excitation (source) components independently in time domain without any prior knowledge about source or system [41].

8. Long-term average spectrum

LTAS represents the logarithmic power spectral density as a function of frequency. LTAS provides:

- Formant info
- Speech quality
- Relative amount of energy within selected frequency bands[30]

9. Power Spectral Density

A speech signal generally has limited average power and can be characterized by an average power spectral density. The area under the square of the magnitude of its Fourier transform is equal to the total energy and is called power spectral density [43].

The average power P_{avg} of a signal s(t) is the following time average:

$$P_{avg=} \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} s(t)^2 \, dt \tag{5}$$

10. Mel frequency cepstrum coefficients

MFCC is the spectrum of the spectrum. It has a great significance in speech emotion recognition and speech recognition tasks [30]. MFCC is based on characteristics of human ear's hearing. The block diagram in Figure 4.4 represents the computation of MFCC.

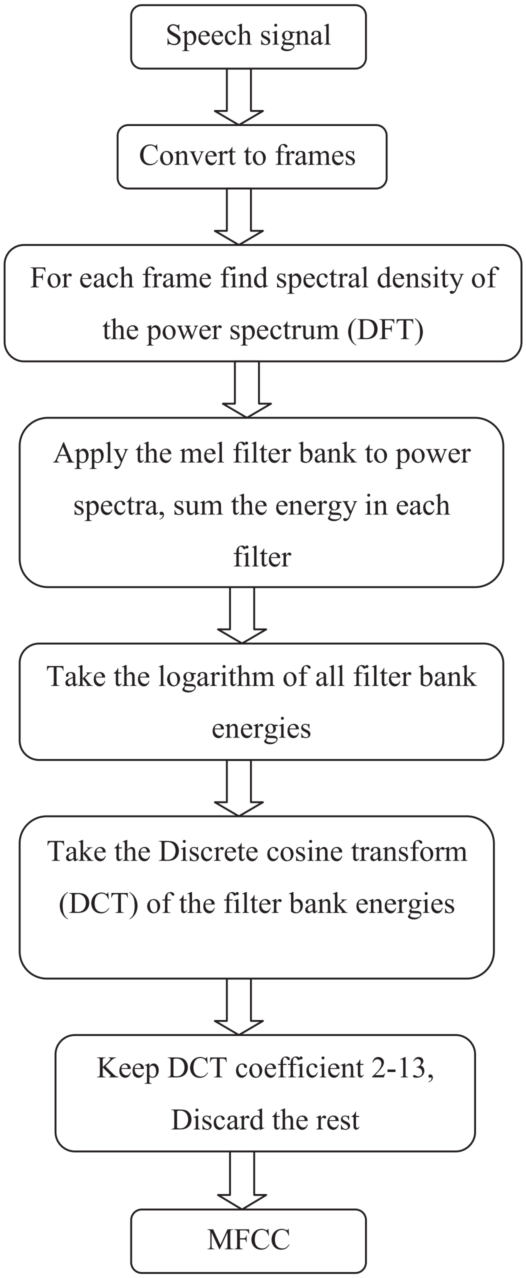

Speech signal
Convert to frames
For each frame find spectral density of the power spectrum (DFT)
Apply the mel filter bank to power spectra, sum the energy in each filter
Take the logarithm of all filter bank energies
Take the Discrete cosine transform (DCT) of the filter bank energies
Keep DCT coefficient 2-13, Discard the rest
MFCC

Figure 4.4: Block diagram representing computation of MFCC

11. Linear predictive cepstrum coefficients

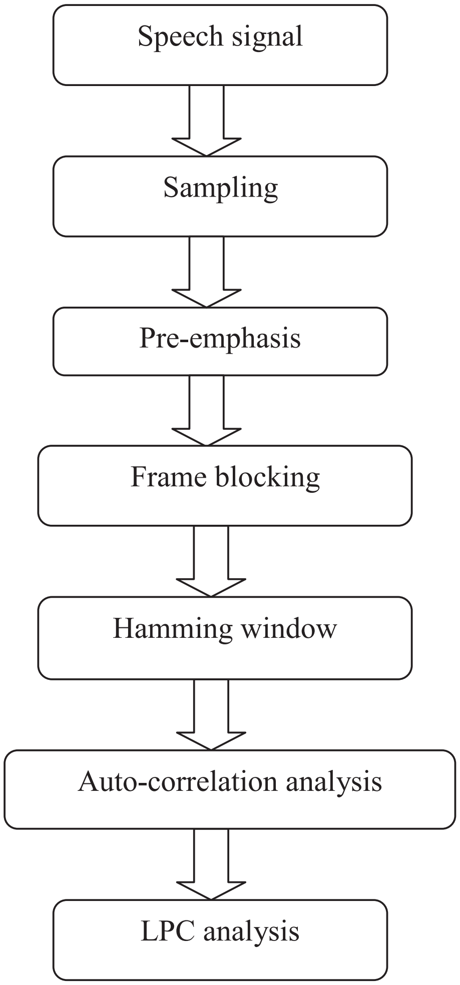


Figure 4.5: Block diagram representing computation of LPCC

The block diagram in Figure 4.5 represents the computation of LPCC and is explained below step by step.

- Speech sampling: Higher recognition accuracy is achieved by increasing the frequency of sampling.

- Pre-emphasis: The sampled speech signal suffers from additive noise and has a high dynamic range. So pre-emphasis is done to remove the noise from the speech signal.

- Frame blocking: The speech signal changes with time or is dynamic in nature. When it is analyzed over quite a small period of time, it is assumed as stationary. Speech signal is analyzed by blocking it into frames of L samples, with each neighbouring frame separated by K samples. If K=L, then the estimation of LPC spectral will be quite smooth from frame to frame and if K>L, there will be no overlapping between adjacent frames.

- Windowing: Signal discontinuities are decreased by windowing each frame. The frame length L should be long enough to avoid tapering effects of the window on the result.
- Autocorrelation analysis: Fundamental frequency or pitch can be found using auto correlation analysis. Repeating patterns can be found or missing fundamental frequency can be identified using this technique.
- LPC analysis: Levinson-Durbin recursive algorithm is used to convert each frame of autocorrelation coefficients into LPC coefficients.

The convolution of excitation source and time varying vocal tract system components produce the speech signal. Each component has to be separated to study them independently. The cepstral analysis is used for the deconvolution of the given speech into excitation and vocal tract system components by traversing through frequency domain. Linear prediction analysis finds the source and system components from time domain and hence is used to reduce computational complexity. LPC analysis leads to the computation of a discrete number of coefficients called linear predictive cepstrum coefficients [42].

12. Speech duration

The time taken by the speaker to complete its speech is known as speech duration. The duration of a sound is one of its most basic characteristics, and measuring duration is therefore part of virtually every acoustic analysis.

13. Short-time energy

The short time energy is the energy of short speech segment [44]. The sum of squares of the samples in a frame is called short-time energy and expressed in the equation below:

$$\mathrm{E(n)} = \sum_{\mathrm{m}=-\infty}^{\infty} (\mathrm{x(m)}.\, \mathrm{k(n-m)})^2 \tag{6}$$

k(n) represent the windowing function of finite duration

x(n) represent the speech signal

n is the shift or rate in number of samples, required to find the short term energy.

14. RMS Energy

The global energy of the speech signal can be found by calculating the root average square of the amplitude, also called root-mean-square (RMS). RMS of a continuous function (or waveform) f(t) defined over the interval $T_1 \leq t \leq T_2$ is:

$$f_{rms} = \sqrt{\frac{1}{T_2 - T_1} \int_{T_1}^{T_2} [f(t)]^2 dt} \tag{7}$$

15. Shimmer

Shimmer is defined as the short term perturbation in amplitude [45]. It is used to describe the pathological voice quality.

(a) Shimmer (absolute) can be expressed as [45]:

$$\text{Shimmer (absolute)} = \frac{1}{N-1} \sum_{i=1}^{N-1} \left| 20 \log \left(\frac{A_{i+1}}{A_i}\right) \right| \tag{8}$$

A_i: peak-to-peak amplitude data

N: No. of extracted fundamental frequency periods

(b) Shimmer (relative) can be expressed as [45]:

$$\text{Shimmer (relative)} = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A_i - A_{i+1}|}{\frac{1}{N} \sum_{i=1}^{N} A_i} \tag{9}$$

(c) Shimmer (apq3) can be expressed as [45]:

$$\text{Shimmer (apq3)} = \frac{\frac{1}{N-2} \sum_{i=2}^{N-1} \left| A_i - \frac{(A_i + A_{i-1+} A_{i+1})}{3} \right|}{\frac{1}{N} \sum_{i=1}^{N} A_i} \tag{10}$$

(d) Shimmer (apq5) can be expressed as [45]:

$$\text{Shimmer (apq5)} = \frac{\frac{1}{N-4} \sum_{i=3}^{N-2} \left| A_i - \frac{(A_i + A_{i-2} + A_{i-1+} A_{i+1+} + A_{i+2})}{5} \right|}{\frac{1}{N} \sum_{i=1}^{N} A_i} \tag{11}$$

(e) Shimmer (apq11) can be expressed as [45]:

$$\text{Shimmer (apq11)} = \frac{\frac{1}{N-10}\sum_{i=6}^{N-5}\left|A_i - (\sum_{k=i-5}^{i+5}\frac{A_k}{11})\right|}{\frac{1}{N}\sum_{i=1}^{N}A_i} \tag{12}$$

16. Jitter

Jitter is defined as the short-term perturbation in the fundamental frequency of the voice [45]. It is used to describe the pathological voice quality.

(a) Jitter (absolute) can be expressed as [45]:

$$\text{Jitter (absolute)} = \frac{1}{N-1}\sum_{i=1}^{N-1}|T_i - T_{i-1}| \tag{13}$$

T_i: extracted f0 period length

N: no. of extracted f0 period

(b) Jitter (relative) can be expressed as [45]:

$$\text{Jitter (relative)} = \frac{\left(\frac{1}{N-1}\right)\sum_{i=1}^{N-1}|T_i - T_{i-1}|}{\frac{1}{N}\sum_{i=1}^{N}T_i} \tag{14}$$

(c) Jitter (rap) can be expressed as [45]:

$$\text{Jitter (rap)} = \frac{(\frac{1}{N-2})\sum_{i=2}^{N-1}\left|T_i - \frac{(T_i + T_{i-1} + T_{i+1})}{3}\right|}{\frac{1}{N}\sum_{i=1}^{N}T_i} \tag{15}$$

(d) Jitter (ppq5) can be expressed as [45]:

$$\text{Jitter (ppq5)} = \frac{(\frac{1}{N-4})\sum_{i=3}^{N-2}\left|T_i - \frac{(T_i + T_{i-2} + T_{i-1} + T_{i+1} + T_{i+2})}{5}\right|}{\frac{1}{N}\sum_{i=1}^{N}T_i} \tag{16}$$

17. Zero crossing Rate

Zero crossing rate is the rate at which a signal changes from positive value to negative value or vice-versa. It is also used to determine voiced or unvoiced segment of speech.

18. Roll-off

Roll-off determines the highest frequency in the signal that contains a certain fraction of the total energy below it.

19. Speech Rate

The speech rate is the measure of syllables per second. A reduced speaking rate may also be the result of several factors including cognitive demand and fatigue induced changes to motor functioning.

20. Spectral Kurtosis

The spectral kurtosis is a statistical tool. It can be used to determine the presence of transients and their position in the frequency domain [46].

21. Sound pressure level (SPL)

Sound Pressure is the difference between the pressure produced by a sound wave and the barometric (ambient) pressure at the same point in space.

4.2.5 FEATURE REDUCTION

Feature reduction was done using FDR (fisher discriminant ratio) and Correlation to further improve the accuracy.

1. FDR (Fisher's Discriminant Ratio)

FDR is the statistical technique that is often used in reducing dimensionality of a collection of unstructured random variables for analysis and interpretation. FDR can be expressed as:

$$\mathrm{FDR} = \left| \frac{\mathrm{Avg.}\,1 - \mathrm{Avg2}}{\sqrt{(\mathrm{Stddev.}\,1)^2 + (\mathrm{Stddev.}\,2)^2}} \right| \tag{17}$$

2. Correlation coefficient

Correlation coefficient determines the relationship between two properties. Correlation coefficient can be expressed as:

$$\mathrm{Correl}(X,Y) = \frac{\sum(x-\bar{x})\,(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2\,\sum(y-\bar{y})^2}} \tag{18}$$

4.2.6 CLASSIFICATION

Classifiers applied with standard parameter settings were: Artificial Neural Networks (i.e. Multi layer Perceptron), Support Vector Machines, k-Nearest Neighbours (Nearest Neighbours; k = 1, 3), Logistic regression, Naive bayes.

1. Support vector machine

SVM can be used for two class problem. It classifies data by determining the best hyperplane. The hyperplane divides the plane into two i.e. all the data points which belong to first class lie in one side of plane and the data points which belong to second class lie on other side of plane. The best hyperplane is the one which provides largest margin between the two classes.

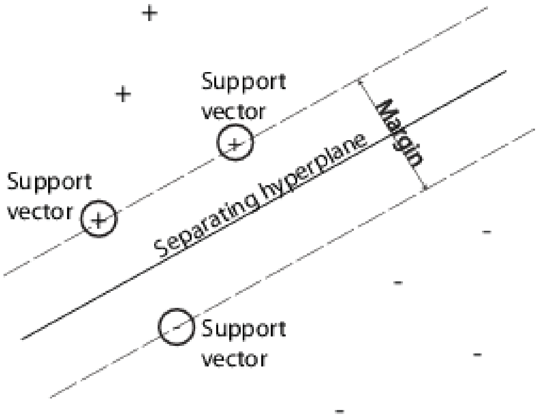


Figure 4.6: Support vector machine description [47]

Figure 4.6 describes the support vector machine, with + indicating data points of first class, and – indicating data points of second class. The support vectors are the data points which are closest to the hyperplane and lies on the boundary of the slab [47].

Parameters used for classification:

- Kernel: Linear kernel

- C (Soft margin constant): 1
- Epsilon(For round-off error): 1.0E-12

2. Artificial neural network (i.e. Multi layer Perceptron)

It is a feed-forward neutral network. It is inspired by the human brain. It combines several perceptron (i.e. network of simple neurons) to create non-linear decision boundary. It has three types of layers as shown in Figure 4.7. The first layer is the input layer, last layer is the output layer and layer between the input and output layer is called hidden layer. Hidden layer can be one or more in number depending on the complexity of the system. Each layer can have any number of perceptron. The perceptrons are also called nodes or artificial neurons. Each perceptron receives an input (i.e. stimuli) for processing and provides the output via its related links to the neighbouring perceptron.

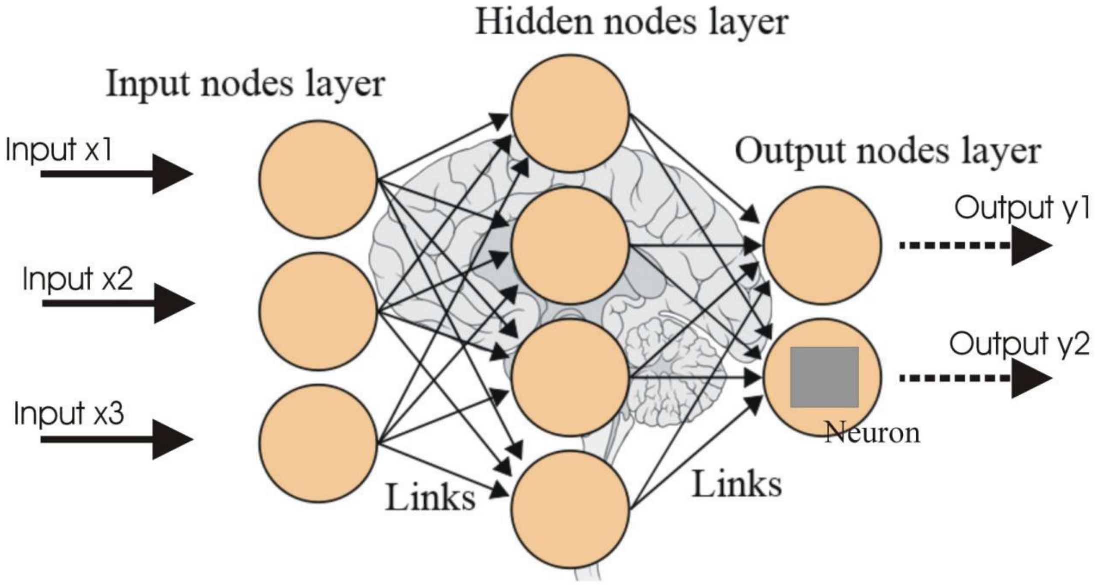


Figure 4.7: Basic structure of Artificial Neural Network [48]

The ANN has four main sections:

1. A perceptron as a unit that activates after receiving the stimuli

2. Interconnections between perceptrons

3. Learning function for managing weights between input and output

4. An activation function that converts input into output inside the perceptron

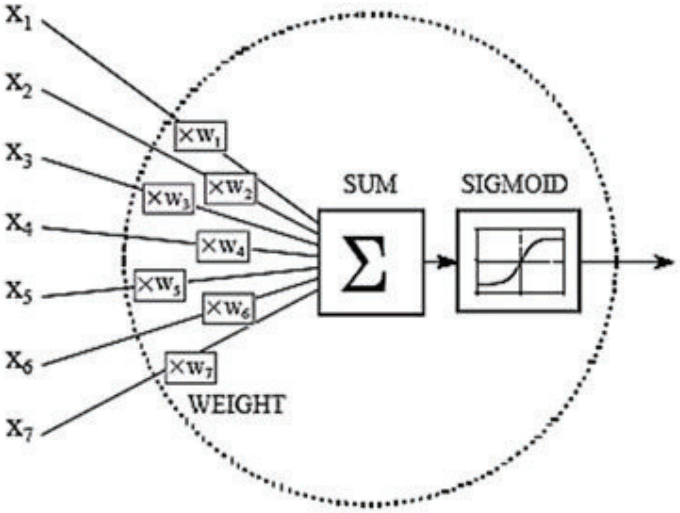


Figure 4.8: Internal structure of the perceptron [48]

Figure 4.8 describes the internal structure of the perceptron with its input X_i , weighted input XW_i and Activation function (i.e. Sigmoid)

The perceptron first calculates the weighted sum of inputs and then passes it to the activation function for the output. Activation function generates the output as 1 if it is greater than the threshold value and 0 otherwise.

Parameters used for classification:

- Learning rate (The amount the weights are updated): 0.3
- Momentum (Momentum applied to the weights during updating): 0.2
- Hidden layer: Sum of attributes and classes
- Training type (0: Train by epoch; 1: Train by Min error): 0
- Epoch (The number of epochs to train through): 500
- Activation function: Sigmoid

3. K-Nearest neighbours

The data points are classified based on the class of their nearest neighbours. The classification is done based on the class of only one nearest neighbour so the technique is known as nearest neighbour classifier. When it is done based on the class of more than one neighbour the technique is known as K-Nearest neighbours. K determines the no. of neighbours is to be consideration during the classification.

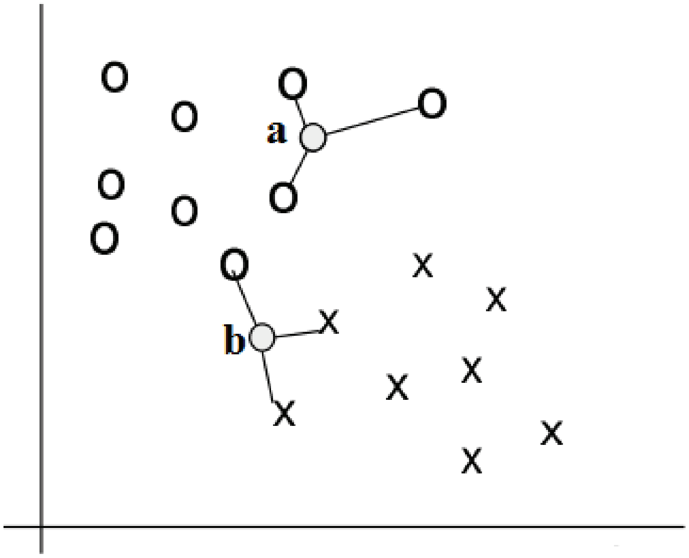


Figure 4.9: A basic example of KNN [49]

Figure 4.9 describes the 3- nearest neighbour classifier based on a two class problem in a two dimensional feature space. In the above example decision for '**a**' is simple as all the three nearest neighbour belong to the same class '**O**' so '**a**' is classified as '**O**'. The decision for '**b**' is a bit difficult as it has two nearest neighbour of class '**X**' and one of class '**O**'. This can be decided by checking the majority of the data points. Therefore '**b**' belongs to class '**X**'.

Parameter used for classification:

- K=1
- Search algorithm: Linear nearest neighbour search algorithm
- Distance function: Euclidean distance

4. Logistic regression

Weighted logistic regression is a memory-based classifier. It is used to find the probability P ($b_q|S_p$, a_q) approximately i.e. for an underlying system $S_{p,}$ the output of the system shall be b_q for a defined input of a_q.

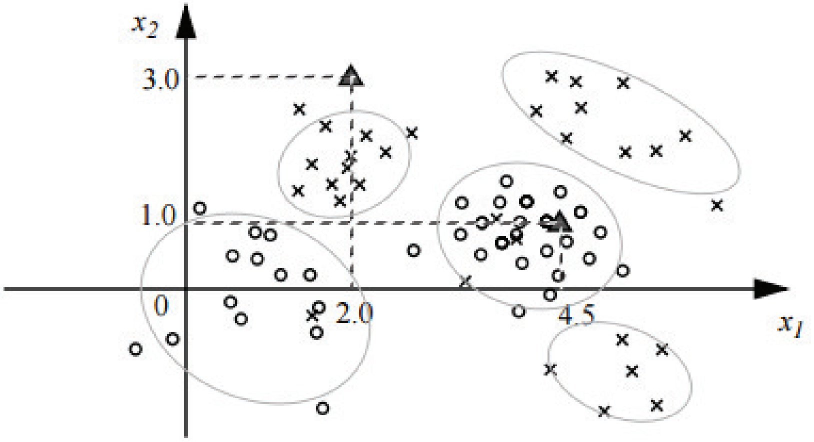
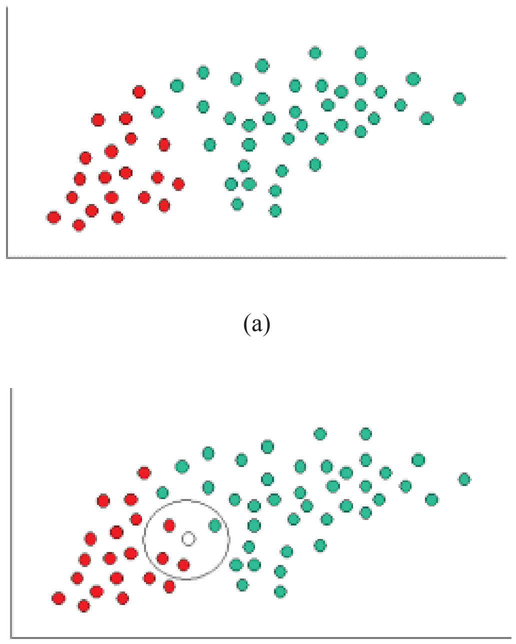


Figure 4.10: An example of logistic regression [50]

Figure 4.10 describes a 2-dimensional trained system S_p with Boolean output. In memory based systems knowledge is derived from previous observations, like in this case the crosses and circles denote the training data points of the system. To find P ($b_q|S_p$, a_q), of an unknown point (a_q, b_q), knowledge of the system is required. Circles correspond to data points with output 0 while crosses denote output 1. So the point (2.0, 3.0) corresponds to output 1 as it is majorly surrounded by crosses and similarly the point (4.5, 1.0) corresponds to output 0.

5. Naive bayes

Naive bayes classifier is based on the Bayes theorem. Though it looks quite simple but sometimes it can outperform other classifiers. Figure 4.11 (a) shows two clusters of green and red balls. The task is to classify a new ball which enters the space as green or red. By observation it can be seen that the chances of the new ball to be green are more than being red as green balls are almost twice in number than red. There are total 60 balls out of which 40 are green and 20 are red. The prior probability of the ball being red is (2/6) while for green it is (4/6).



(a)

(b)

Figure 4.11: An example of Naive Bayes theorem [51]

A white ball enters the space and has to be classified. In naive bayes a circle is drawn about the white ball (X) encircling a specific number of balls (decided before) irrespective of its class (green or red). Now the likelihood probability of the ball being green or red is found.

$$\text{Likelihood of X being green} = \frac{\text{Number of green in the vicinity of X}}{\text{Total number of Green}} \qquad (19)$$

Similarly likelihood of X being red is calculated. Probability of X being green is (1/40) and being red is (3/20). Then finally probability is calculated by multiplying prior probability and likelihood probability which comes out to be (1/60) for green and (1/20) for red. Finally X is classified as RED since its class membership achieves the largest posterior probability.

4.2.7 SUMMARY OF THE METHODOLOGY

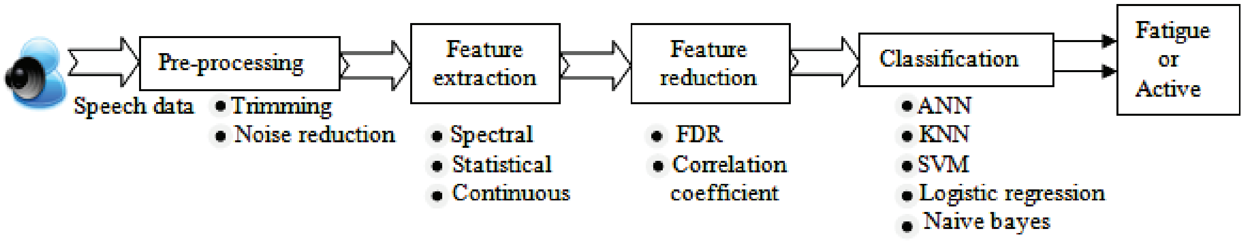


Figure 4.12: Block diagram of the methodology

The Figure 4.12 describes the summary of the methodology. Samples were recorded using the recorder in a quiet room and each speech sample was segmented and analyzed using PRAAT and MATLAB. Silences were removed from the start and end of the samples. Feature extraction is done to extract those particular features from the speech signal that would reveal useful information regarding fatigue detection. It also improves the performance of classifiers. To further improve the accuracy, feature reduction was done. The feature set was obtained which was used for classification purpose. Different classifiers were applied to the selected feature set table and accuracy was noted.

CHAPTER 5

RESULTS AND DISCUSSION

Each of the participants was asked to repeat the experimental stimuli at each stage of the experiment at an interval of 4 hr. The participants were asked about their own fatigue assessment at the time of experiment by marking the Karolinska Sleepiness Scale (KSS) as shown in Figure 5.1. The experiment assistant was also asked to assess the participants on KSS. Samples during Alert state (i.e. scores between 1 to 3) were considered as active samples and during inactive state (i.e. scores between 7 to 9) were considered as fatigue samples.

Karolinska Sleepiness Scale (KSS)

Check mark the ONE statement that best describes your/ participant's state during the experiment. You may also use the intermediate steps.

---1. Very alert
---2.
---3. Alert- normal level
---4.
---5. Neither alert nor sleepy
---6.
---7. Sleepy, but no effort to keep awake
---8.
---9. Very sleepy, great effort to keep awake

Figure 5.1: Karolinska Sleepiness Scale (KSS)

The active and fatigue speech samples were used to find the features discussed above for each participant. FDR was calculated for each experimental stimulus and features were arranged from highest to lowest FDR. Best twenty features having the highest FDR were separated and further feature reduction was done using the correlation i.e. the feature having the highest FDR value was used to find the correlation coefficient with rest of the features one by one. If the correlation coefficient value with the rest of the features came out to be greater than or equal to +/-9 then it was removed from the feature set and then the next feature with highest FDR value was used to find the correlation coefficient with rest of the features and so on. After completing the above described procedure feature set was obtained which was used for classification purpose. Different classifiers were applied to the selected feature set table and accuracy was noted.

Spectrogram: A spectrogram is a visual representation of sound. It displays the amplitude of the frequency components of the signal over time. It also shows changes in frequency values of the component of the signal over time. It is especially useful with complex signal which contain more than one frequency component. Those frequency components cannot be discerned from the visual examination of the waveform, but can be visualised by the spectrogram.

Waveform: A waveform is another visual representation of sound. It displays the variation in air pressure over time. The upper half of the window shown in Figure 5.2 and Figure 5.3 depicts the waveform and the greyish image in the lower half of the window shows the spectrogram of the sound wave.

The spectrogram is a graphic representation of the three dimensions of sounds in terms of their component frequencies. The horizontal direction of the spectrogram corresponds to time, the vertical direction corresponds to frequency and the degree of shading corresponds to amplitude. The time scale is common for both the spectrogram and the waveform. The spectrogram extends form 0 Hz to 5000 Hz on the vertical axis. The components that makes up a complex signal do not share the same amplitude (i.e. refers to the loudness of the components) value, differences in the amplitude are shown on a spectrogram by shading. The frequency components with the highest amplitude values are shown in dark black and the components with lower amplitude values are displayed in lighter shades of grey up to white. The white signifies very low amplitude or silence.

The dark shade of grey around a particular frequency in the spectrogram of sound wave corresponds to the formant. It is the concentration of acoustic energy and distinctly seen in a wideband spectrogram. The darker the shade of grey, the stronger it is (i.e. with more energy, or it is more audible). It is shown by the red dots in the spectrogram. The blue vertical line in the upper part of the window is the representation of pulses (i.e. glottal closures). Figure 5.2 and Figure 5.3 is the waveform and spectrogram of a participant uttering the words tea, ten, teeth, tata, papa, pen, pear and paper.

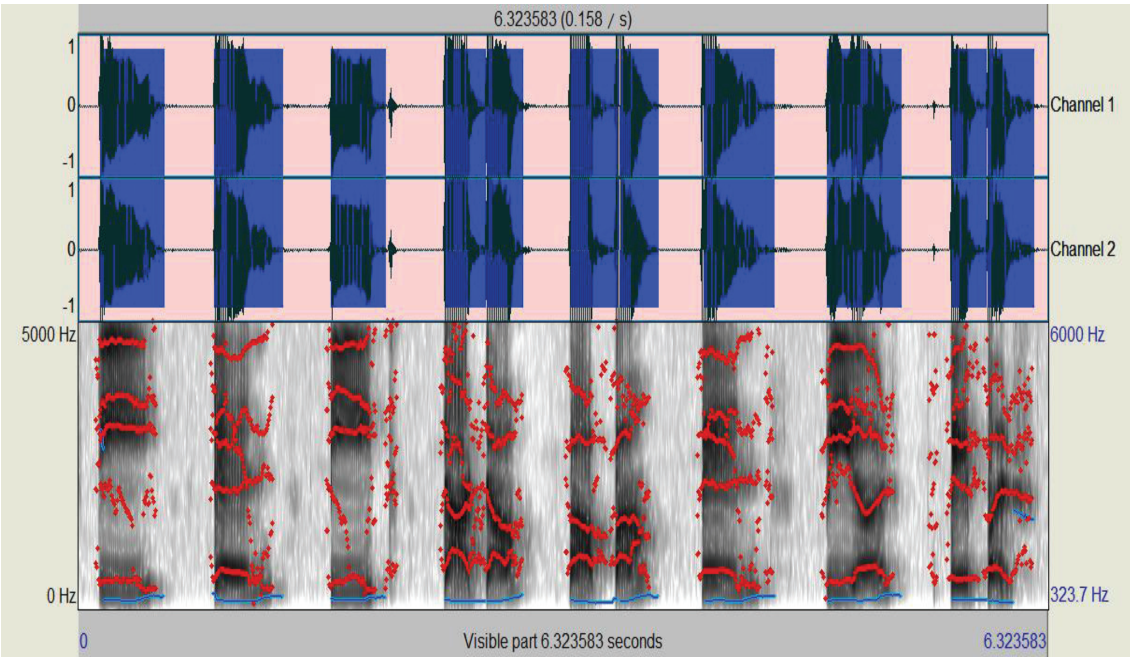


Figure 5.2: A waveform and spectrogram of an active participant

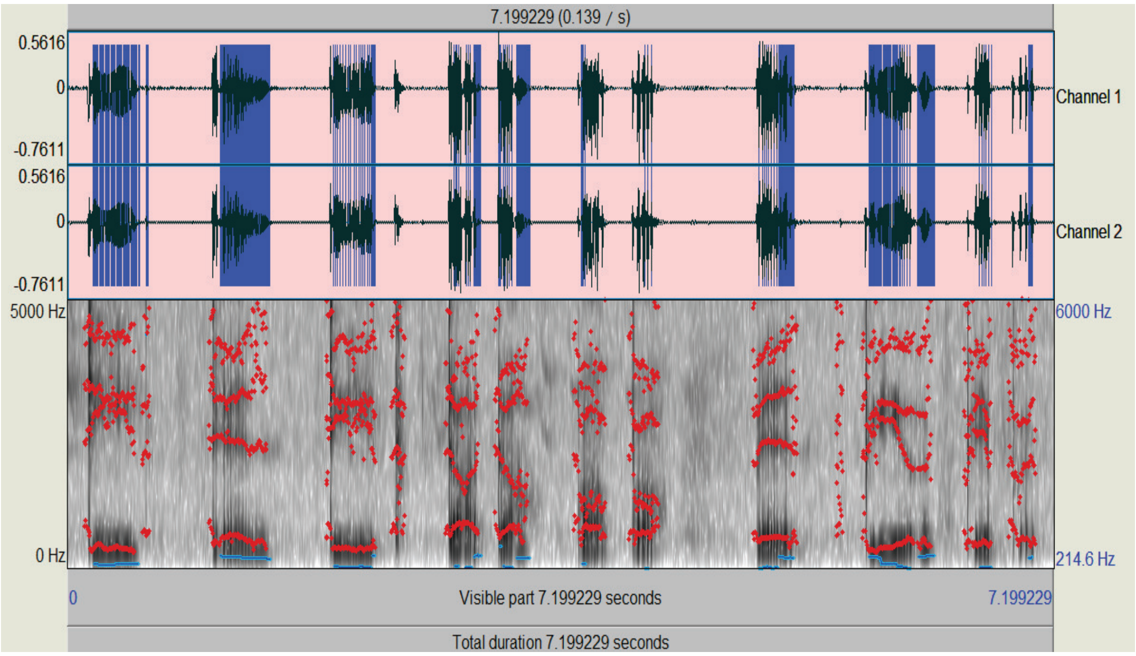


Figure 5.3: A waveform and spectrogram of a fatigue participant

From Table 5.1 it was concluded that the most significant feature set was different for different stimuli and were arranged in the order of high to low rank. The features that were most common for different stimuli were RMS, mean, band density, min intensity and unvoiced frames are the features which were independent to the stimuli and showed prominent change with increase in fatigue.

Table 5.1: Most significant features in each stimuli

STIMULI/ FEATURE	**HELLO**	**MY NAME IS**	**THE QUICK BROWN FOX..**	**TEETH**	**TATA**	**PAPA**	**PEN**
1	**RMS**	**RMS**	Shimmer (apq3)	**Min intensity**	**RMS**	**RMS**	**RMS**
2	**Min intensity**	Power	Shimmer (local)	LPCC1	Band energy diff.	Band energy diff	Energy
3	Max	**Band density**	Power	**RMS**	**Min intensity**	Energy	**Min intensity**
4	Low energy	Total SPL	**RMS**	Max	Energy	**Min intensity**	Min. pitch
5	Speech rate	Min pitch	Shimmer (local, dB)	LPCC2	**Mean**	MFCC0	Centre of gravity
6	**Band density**	**Mean**	Shimmer (apq5)	Min pitch	**Band density**	Centre of gravity	MFCC0
7	LPCC1	Formant 4	**Band density**	Speech rate	Max	**Band density**	Formant4
8	Skewness	No. of voice breaks	Jitter(rap)	**Band density**	Centre of gravity	MFCC1	Skewness
9	Band energy diff.	Shimmer (dda)	**Unvoiced frames**	MFCC0	Degree of voice breaks	**Mean**	**Mean**
10	**Mean**	Band energy diff	Speech rate	Band density diff.	LPCC9	**Unvoiced frames**	Bandwidth 4
11	No. of voice breaks	Speech rate	**Mean**	**Mean**	**Unvoice d frames**	Kurtosis	**Unvoiced frames**
12	Formant2	**Min intensity**	**Min intensity**	**Unvoiced frames**	Max intensity	Formant4	**Band density**
13	**Unvoiced frames**	**Unvoiced frames**	Jitter(local)	Bandwidth 2	Min pitch	Min pitch	Shimmer (local)

The participant's speech analysis showed that there are features whose value prominently changes (i.e. increase or decrease) with fatigue as described in Table 5.2 and can also be seen from Figure 5.2 and Figure 5.3. Voice is an indication of physical and mental state. Fatigue induced physiological and psychological changes. Fatigue reduced cognitive ability due to which it adversely affects the speech planning and neuromuscular motor coordination processes, which in turn slows down the speed due to which speed duration and no. of voice break increase, and speech rate decrease. Reduced muscle stress and body temperature due to fatigue affects the respiration, phonation, articulation, and radiation, which in turn decrease the loudness, sound pressure level, power spectral density, energy, power, RMS, Standard deviation, band energy difference and band density difference. This also decreases the voice quality due to which jitter and shimmer increases. The softening of vocal tract walls, decrease in vocal fold tension, stiffness and viscosity, and lowering of velum due to fatigue increase the formant frequency (especially in lower formant), fundamental frequency and no. of pulses.

Table 5.2: Features whose value prominently changes with fatigue

Features whose value decrease in fatigue	**Features whose value increase in fatigue**
• Loudness • Standard variation • Root mean square (RMS) • Sound pressure level (SPL) • Power spectral density (PSD) • Band energy difference • Band density difference • Speech Rate • Energy • Power	• Formant(Acoustic Resonance) • Fundamental frequency(pitch) • Speech duration • Jitter and shimmer • No. of voice break • No. of pulses (glottal closures)

Table 5.3: Classifiers applied on different stimuli and accuracy is calculated

Stimuli/ Classifiers	**ANN**	**SVM**	**KNN**	**Logistic Regression**	**Naive bayes**
Hello	100%	100%	96.42 %	100%	96.42 %
My name is…	96.42 %	100 %	96.42 %	96.42 %	96.42 %
The quick brown fox..	96.42 %	100 %	96.42 %	96.42 %	96.42 %
Teeth	96.42 %	100 %	96.42 %	96.42 %	96.42 %
Tata	96.42 %	96.42 %	96.42 %	96.42 %	96.42 %
Papa	96.42 %	96.42 %	96.42 %	96.42 %	96.42 %
Pen	96.42 %	100 %	96.42 %	100 %	96.42 %

From Table 5.3, it was concluded that the Support vector machine was proven to be the best classifier among the other classifiers used.

Table 5.4: Classifiers applied on different stimuli and over all accuracy is calculated with confusion matrix

Classifiers	**Accuracy**	**Confusion Matrix**
Artifical neural network	96.9%	a b <-- classified as 92 6 | a = NO 0 98 | b = YES
Support vector machine	98.9%	a b <-- classified as 96 2 | a = NO 0 98 | b = YES
k- nearest neighbour	96.4%	a b <-- classified as 91 7 | a = NO 0 98 | b = YES
Logistic Regression	97.4%	a b <-- classified as 93 5 | a = NO 0 98 | b = YES
Naive Bayes	96.4%	a b <-- classified as 91 7 | a = NO 0 98| b = YES

Table 5.5: Accuracy noted from literature review

Author	**Accuracy**
Krajewski *et al.* [28]	88.2%
Krajewski *et al.* [29]	83%
Krajewski *et al.* [30]	86%

This thesis discussed different feature sets and different classifiers in order to improve the detection of fatigue and accuracy is improved as noted from literature review discussed in the Table 5.5. SVM was found to be the best classifier with the accuracy of 100% in most cases and overall accuracy was found to be approximately 99%.

CHAPTER 6

CONCLUSION AND FUTURE SCOPE

6.1 CONCLUSION

Fatigue is a critical element in many professions. Human voice has inevitable dependence on fatigue. Most of the literature deals with electrode and visual based fatigue measurement techniques and only small research has been done to detect fatigue by voice analysis. The main objective of this particular study is to analyze different feature sets in order to improve the detection of fatigue. In the present study a total of 45 features have been extracted using different voice stimuli. It was concluded that the most significant feature set was different for different stimuli. The features that were independent of the stimuli were RMS, mean, band density, min intensity and unvoiced frames. They also showed prominent change with increase in fatigue. It was found that Loudness, Standard variation, RMS, SPL, PSD, Band energy difference, Band density difference, Speech Rate, Energy and Power were the features whose value decreased in fatigue. Formant, Fundamental frequency, Speech duration, Jitter and shimmer, Number of voice breaks and Number of pulses are the features whose value increases with fatigue. To classify the testing data set, five state-of-the-art classifiers ANN, KNN, SVM, Logistic regression and Naive bayes have been used. After the analysis, SVM was found to be the best classifier with the accuracy of 100% in most cases and overall accuracy was found to be approximately 99%. This achieved accuracy is better than the previous methods, due to the fusion of different features.

6.2 FUTURE SCOPE

There is always a scope for improvements. In the present study, the main limitation was of smaller database. Though every effort was made in order to detect the fatigue with high level of accuracy, it would definitely be better to use larger database for training. A database with large numbers and variety of participants would improve its value. The experiments were conducted in a quiet room, to make the system more realistic the effects of noise could be taken into consideration.

In future, a more generalized system to detect fatigue can be developed which would not need prior knowledgebase of the subject. Moreover, features from other domains like facial images and reflexes of the subject can be used to further improve the accuracy.

Further, a numerical value of fatigue can be proposed which would tell the fatigue percentage of the subject and accordingly action could be taken.

LIST OF PUBLICATIONS

1. Sonika, Mandeep Singh, “Fatigue Detection Using Voice Analysis: A review”, Accepted at International Journal of Applied Engineering Research (IJAER).

2. Sonika, Mandeep Singh, “Fatigue Detection Using Voice Analysis”, Communicated to Biomedical Signal Processing and Control.

HARDWARE SPECIFICATIONS

- Model no: **ICD-UX533F/B**
- Built-In memory: 4GB
- PC Connectivity: Direct USB
- Recording Format: Linear PCM/MP3
- Playback Format: MP3/AAC/WMA/WAV
- Earphone Jack: Yes
- Battery type: AAA NiMH Rechargeable Battery
- Dimension (W x H x D): 36.6 x 102.0 x 13.9 mm
- Weight (including batteries): 58g
- Mic-In Jack: Yes
- Supplied software:
    - 1 x NH-AAA Rechargeable Battery
    - Application Software 'Sound Organizer‘
- Functionality:
    - Low cut filter
    - Noise cut Filter
    - Track Mark
    - Scene Select
    - Mic sensitivity selection: Low, medium and high
- OVERALL FREQUENCY RESPONSE:
    - LPCM (44.1kHz, 16-bit): 50 - 20,000 Hz
    - MP3 8kbps : 60 - 3,400 Hz
    - MP3 48kbps : 50 - 14,000 Hz
    - MP3 128kbps: 50 - 16,000 Hz
    - MP3 192kbps: 50 - 20,000 Hz
- Recording time:
    - LPCM (44.1kHz, 16-bit):6Hrs 0 Min
    - MP3 8kbps : 1073 Hrs 0 Min
    - MP3 48kbps: 178 Hrs 0 Min

Digital Voice Recorder

- MP3 128kbps: 67 Hrs 05 Min
- MP3 192kbps: 44 Hrs 40 Min

REFERENCES

[1] Akerstedt T., “Consensus Statement: Fatigue and Accidents in Transport Operations”, Journal of Sleep Research, vol. 9, pp. 395, 2000.

[2] Krajewski J., Trutschel U., Golz M., Sommer D., and Edwards D., “Estimating Fatigue From Predetermined Speech Samples Transmitted By Operator Communication Systems”, Fifth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Yellowstone Conf. Centre, Montana, USA, pp. 468-474, 2009.

[3] Lal S. K., and Craig A., “A Critical Review of The Psychophysiology of Driver’s Fatigue”, Biological Physiology, vol. 55, pp.173-194, 2001.

[4] Chen L.L., Sugi T., and Shirakawa S., “Comfortable Environments for Mental Work by Suitable Work-Rest Schedule: Mental Fatigue and Relaxation,” Proc. 6th IEEE International Conference on Industrial Informatics, pp. 365-370, 2008.

[5] Marcora S.M., Staiano W., and Manning V., “Mental Fatigue Impairs Physical Performance in Humans”, Journal of Applied Physiology, vol. 106 (3), pp. 857-864, 2009.

[6] Li Z., Li W., Yao Z., and Li Y., “Speech-based Five Features Extraction and Hardware Trade-off Design for Mental Fatigue Monitoring” , International Conference of Information Technology, Computer Engineering and Management Sciences, pp. 68-71, 2011.

[7] Shen K.Q., Ong C.J., and Li X.P., “A Feature Selection Method for Multilevel Mental Fatigue EEG Classification”, IEEE Biomedical Engineering, vol.54 (7), pp.1231-1237, 2007.

[8] Marcora S. M., Staiano W., and Manning V., “Mental Fatigue Affects Endurance Performance in Humans”, Journal of Applied Physiology, vol. 106, pp. 857 – 864, 2009.

[9] Hagberg M., “Muscular Endurance and Surface Electromyogram in Isometric and Dynamic Exercise”, Journal of Applied Physiology, vol. 51, pp. 1–7, 1981.

[10] Murata A., and Uetake A., “Evaluation of Mental Fatigue in Human-Computer Interaction Analysis Using Feature Parameters Extracted from Event Related Potential”, Proc.10th IEEE International Workshop on Robot and Human Interactive Communication, pp.630-635, 2001.

[11] Rogado E., García J. L., Barea R., and Bergasa L.M., “Driver Fatigue Detection System”, Proceedings of the 2008 IEEE International Conference on Robotics and Biomimetics Bangkok, Thailand, 2009.

[12] Kang H., “Various Approaches for Driver and Driving Behavior Monitoring: A Review,” Computer Vision Workshops (ICCVW), IEEE International Conference, pp. 616-623, 2013.

[13] Golz M. and Sommer D., “Detection of Strong Fatigue During Overnight Driving”, Proceedings Annual Congress of the German Society for Biomedical Engineering, vol. 39, pp. 479-480, 2005.

[14] Tabain M., “ Research Methods in Speech Production”, in The Bloomsbury Companion To Phonetics, Book Edited by Mark J. Jones, Rachael-Anne Knight, 1st ed. , ch. 3, Sec. 2, pp. 39-54, 2013.

[15] Baer T., Gore J.C., Gracco L.C., and Nye P.W., “Analysis of Vocal Tract Shape and Dimensions Using Magnetic Resonance Imaging: Vowels”, Journal of Acoustical Society of America, vol. 90, pp. 799–828, 1991.

[16] Kabi B., Samantaray A., Patnaik P., and Routray A., “Voice Cues, Keyboard Entry and Mouse Click for Detection of Affective and Cognitive States: A Case for Use in Technology- Based Pedagogy”, IEEE Fifth International Conference on Technology for Education (T4E), pp. 210-213, 2013.

[17] Krajewski J., Wieland R., Sommer D., and Gloz M., “Fatigue in Air Traffic Communication-combining Acoustic features within a Computational Intelligence Approach”, 28th conference of the European Association for Aviation Psychology, pp.193-197, 2009.

[18] Sondhi M. M., “Production, Perception, and Modeling of Speech”, Springer handbook of speech processing, Benesty J., Sondhi M. M., and Huang Y., Eds, Springer-Verlag Berlin Heidelberg, pp.7-96, 2008.

[19] Gramley V., “Acoustic Phonetics”, [Online]: Available: http://www.uni-bielefeld.de/lili/ personen/vgramley/teaching/HTHS/acoustic_2010.html, June 2014.

[20] IIT GUWAHATI Virtual Lab, “Identification of Voice/Unvoiced/Silence regions of Speech”, Retrieved 13 Jan 2015, Available: iitg.vlab.co.in/?sub=59&brch=164&sim=613&cnt=1, 2011.

[21] Rabinar L. R. and Schafer R. W., “Digital Processing of Speech Signals”, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1978.

[22] Atal B. S. and Rabiner L. R., “A Pattern Recognition Approach To Voiced-Unvoiced-Silence Classification With Applications To Speech Recognition,” IEEE Trans. Acoust., Speech, and Signal processing, vol. 24 (3), pp. 201-211, 1976.

[23] Whitmore J., and Fisher S., “Speech during Sustained Operations”, Elsevier, Speech Communication, vol. 20, pp. 55-70, 1996.

[24] Bard E.G., Sotillo C., Anderson A.H., Thompson H.S., and Taylor M.M., “The DCIEM Map Task Corpus: Spontaneous Dialogue under Sleep Deprivation and Drug Treatment”, Speech Communication, vol. 20, pp. 71-84, 1996.

[25] Harrison Y. and Horne J.A., “Sleep Deprivation Affects Speech”, Sleep, vol. 20, pp. 871–877, 1997.

[26] Greeley H. P., Friets E., Wilson J. P., Raghavan S., Picone J., and Berg J., “Detecting Fatigue From Voice using Speech Recognition,” IEEE International Symposium on Signal Processing and Information Technology, Vancouver, pp. 567–571, 2006.

[27] Greeley H.P., Berg J., Friets E., Wilson J., Greenough G., Picone J., Whitmore J., and Nesthus T., Fatigue Estimation Using Voice Analysis”, Behaviour Research Methods, vol. 39, pp. 610-619, 2007.

[28] Krajewski J. and Kroger B., “Using Prosodic and Spectral Characteristics for Sleepiness Detection”, Inter Speech Proceedings, 10th European Conference on Speech Communication and Technology, Antwerp, Belgium, pp.1841-1844, 2007.

[29] Krajewski J., Wieland R., and Batliner A., “An acoustic framework for detecting fatigue in speech based human-computer interaction”, Miesenberger K., Klaus J.,

Zagler W., Karshmer A., Eds, Computers Helping People with Special Needs, Springer, Heidelberg, pp. 54-61, 2008.

[30] Krajewski J., Batliner A., and Golz M., “Acoustic sleepiness detection: Framework and validation of a speech-adapted pattern recognition approach,” Behavior Research Methods, vol. 41 (3), pp. 795–804, 2009.

[31] Dhupati L., Kar S., Rajaguru A., and Routray A., “A Novel Drowsiness Detection Scheme Based on Speech Analysis With Validation Using Simultaneous EEG Recordings”, Proc. IEEE Conference on Automation Science and Engineering (CASE), Toronto, pp. 917–921, 2010.

[32] Vogel P., Fletcher J., and Maruff P., “Acoustic Analysis of the Effects of Sustained Wakefulness on Speech”, Journal of the Acoustical Society of America, vol. 128, pp. 3747–3756, 2010.

[33] Zhang X., Gu J., and Tao Z., “Research of Detecting Fatigue from Speech by PNN”, Dept. of Phys. Sci. and Tech., Soochow University, SuZhou, International Conference on Information, Networking and Automation (ICINA), Kunming, China, Soochow University, SuZhou, 2010.

[34] Krajewski J., Heinze C., Sommer D., Schnupp T., and Laufenberg T., “Applying Nonlinear Dynamics Features for Speech-based Fatigue Detection”, Proceedings of Measuring Behavior, pp. 322-325, 2010.

[35] Krajewski J., Schnieder S., Sommer D., Batliner A., and Schuller B., “Applying Multiple Classifiers And Non-Linear Dynamics Features For Detecting Sleepiness From Speech”, Neurocomputing, Special Issue “From neuron to behavior: evidence from behavioural measurements”, vol. 84, pp. 65–75, 2012.

[36] Rashwan A.M., Kamel M.S., and Karray F., “Car Driver Fatigue Monitoring Using Hidden Markov Models and Bayesian Networks”, IEEE Connected Vehicles and Expo (ICCVE) International Conference, 2013.

[37] Vogel A., “Factors Affecting the Quality of Sound Recording For Speech And Voice Analysis”, International Journal of Speech- Language Pathology, 2009.

[38] Roth T., Roehrs T.A., Carskadon M.A., and Dement W.C., "Daytime Sleepiness and Alertness", M. H. Kryger, T. Roth, and W. C. Dement (Eds.), Principles and Practice of Sleep Medicine, 2nd ed., pp. 14-23.

[39] Oppenheim A., Schafer R., and Buck J., "Discrete-Time Signal Processing", 3rd edition, Prentice Hall, 1999.

[40] Boersma P. and Weenink D., "Praat: doing phonetics by computer", 5.0.46 ed, p. Computer program, 2009.

[41] IIT GUWAHATI Virtual Lab, "Cepstral Analysis of Speech", Retrieved 16 June 2015, from iitg.vlab.co.in/?sub=59&brch=164&sim=615&cnt=1107, 2011.

[42] Venkateswarlu R. L. K, Raviteja R., and Rajeev R., "The Performance Evaluation of Speech Recognition by Comparative Approach, Advances in Data Mining Knowledge Discovery and Applications", Associate Prof. Adem Karahoca (Ed.), ISBN: 978-953-51-0748-4, InTech, DOI: 10.5772/50640, 2012.

[43] Millers S. and Childers D., "Probability and Random Processes", Academic Press, pp. 370–5, 2012.

[44] Yang X., Tan B., Ding J., Zhang J., and Gong J., "Comparative Study on Voice Activity Detection Algorithm," International Conference on Electrical and Control Engineering, ICECE, 2010.

[45] Farrus M., Hernando J., "Using Jitter and Shimmer in Speaker Verification", IET signal Process, vol. 3, pp. 247-257, 2009.

[46] Antoni J, "The Spectral Kurtosis: A Useful Tool for Characterising Non-Stationary Signals", Mechanical Systems and Signal Processing, Elsevier, vol. 20, pp. 282-307, 2004.

[47] MathsWorks, "Support Vector Machines (SVM)", [Online], Available: http://in.mathworks.com/help/stats/support-vector-machines-svm.html?refresh=true, Dec 2014.

[48] Tadiou K. M., "Artificial Neural Networks", [Online], Available: http://futurehumanevolution.com/artificial-intelligence-future-human-evolution/artificial-neural-networks, Dec 2014.

[49] Cunningham P. and Delany S. J., “K-Nearest Neighbour Classifiers”, Technical Report UCD-CSI-2007-4, 2007.

[50] Carnegie Mellon university, “Logistic Regression as a Classifier”, [Online], Available: https://www.cs.cmu.edu/~kdeng/thesis/logistic.pdf, May 2015.

[51] Dell software, “Naive Bayes Classifier”, [Online], Available: http://documents.software.dell.com/Statistics/Textbook/Naive-Bayes-Classifier, May 2015.