E. YILMAZ

DETECTING AIR TRAFFIC CONTROLLERS' STRESS LEVELS USING MACHINE LEARNING TECHNIQUES

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
ATILIM UNIVERSITY

EVRIM YILMAZ

A MASTER OF SCIENCE THESIS
IN
THE DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

AUGUST 2020

ATILIM UNIVERSITY 2020

DETECTING AIR TRAFFIC CONTROLLERS' STRESS LEVELS USING MACHINE LEARNING TECHNIQUES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
ATILIM UNIVERSITY

BY

EVRİM YILMAZ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
THE DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

AUGUST 2020

Approval of the Graduate School of Natural and Applied Sciences, Atilim University.

_____

Prof.Dr. Ali KARA
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of **Master of Science in Electrical and Electronics Engineering, Atilim University**.

_____

Assoc.Prof.Dr. Kemal Efe ESELLER
Head of Department

This is to certify that we have read the thesis DETECTING AIR TRAFFIC CONTROLLERS' STRESS LEVELS USING MACHINE LEARNING TECHNIQUES submitted by EVRİM YILMAZ and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

_____

Asst.Prof.Dr. Uğur TURHAN
Co-Supervisor

Asst.Prof.Dr. İbrahim Baran USLU
Supervisor

Asst.Prof.Dr. Birsen AÇIKEL
Department of Aviation Management, Kastamonu University _____

Assoc.Prof.Dr. Reşat Özgür DORUK
Electrical and Electronics Engineering, Atilim University

Asst.Prof.Dr. Hakan TORA
Department of Avionics, Atilim University

Asst.Prof.Dr. Uğur TURHAN
Department of Air Traffic Control, Eskişehir Technical Uni.

Asst.Prof.Dr. İbrahim Baran USLU
Electrical and Electronics Engineering, Atilim University

Date: *07/08/2020*

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name : Evrim Yılmaz

Signature :

ABSTRACT

DETECTING AIR TRAFFIC CONTROLLERS' STRESS LEVELS USING MACHINE LEARNING TECHNIQUES

Yılmaz, Evrim

MSc., Department of Electrical and Electronics Engineering

Supervisor : Dr. İbrahim Baran USLU

Co-Supervisor : Dr.Uğur TURHAN

AUGUST 2020, 75 pages

Accurate understanding of stress detection with machines will allow preventive measures to be taken for undesirable situations, such as in air traffic control, where communication is mostly through sound and intense stress can directly affect the quality of work and hence human life. Within the scope of this thesis, it was aimed to measure the stress levels of Air Traffic Controllers', which are considered to be under occupational stress, from their speeches on duty. For this purpose, a unique data set was created for the thesis, sound features were extracted and classification studies were carried out with artificial neural networks. As a result of the tests, the average performance for 26 features was 31.2% for NN and 25.9% for SVM.

Keywords: Air Traffic Controller, ATCO, Stress detection, Speech, Machine Learning, ANN, SVM

ÖZ

HAVA TRAFİK KONTROLÖRLERİNİN STRES SEVİYELERİNİN MAKİNE ÖĞRENME TEKNİKLERİYLE ALGILANMASI

Yılmaz, Evrim
Master, Elekrik Elektronik Mühendisliği Bölümü
Tez Yöneticisi : Dr.Öğr.Üyesi İbrahim Baran USLU
Ortak Tez Yöneticisi : Dr.Öğr.Üyesi Uğur TURHAN

AĞUSTOS 2020, 75 sayfa

Hava Trafik Kontrolünde olduğu gibi, iletişimin çoğunlukla ses üzerinden sağlandığı ve yoğun stresin iş kalitesini ve dolayısıyla da insan hayatını doğrudan etkileyebildiği koşullarda, stresin tespitinin makinalarla doğru bir şekilde anlaşılması, istenmeyen durumlar için önleyici tedbirler alınabilmesine olanak tanıyacaktır. Bu tez kapsamında, ses üzerinden özellikle mesleki stres altında olduğu düşünülen Hava Trafik Kontrolörlerinin, stres seviyelerinin ölçülmesi amaçlandı. Bu amaçla tez için benzersiz bir veri seti oluşturulup ses özellikleri çıkarıldı ve yapay sinir ağları ile farklı stres düzeylerinin algılanması için sınıflandırma çalışmaları gerçekleştirildi. Yapılan testler sonucunda 26 özellik seçilerek yapılan testlerde ortalama başarım, yapay sinir ağları kullanıldığında %31,2, destek vektör makinaları kullanıldığında ise %25,9'dır.

Anahtar Kelimeler: Hava Trafik Kontrolörleri, Stres, Stres tespiti, Konuşma, Ses, Makine Öğrenmesi, Yapay Sinir Ağları, Destek Vektör Makineleri

To my mom and dad …

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my academic advisor, Dr. İbrahim Baran USLU. His tremendous kindness, support, and academic guidance have been great value to me.

I shall also thank to my co-adviser Dr. Uğur TURHAN For its sincerity, important guidance and for allowing me to enter the world of ATC throughout this thesis. His academic works inspired me a lot.

I would also like to thank the members of the jury, First to Dr Hakan TORA for his kindness, support, guidance throughout my studies at Atilim University

Furthermore, I thank the members' of the jury, Dr. Reşat Özgür DORUK and Dr Birsen Yörük AÇIKEL.

Finally, last but not least, I thank to my parents for their love and support throughout my life.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF SYMBOLS/ABBREVIATIONS

ANN	Artificial Neural Network
ATC	Air Traffic Control
ATCO	Air Traffic Control Operator
ATFM	Air Traffic Flow Management
ATM	Air Traffic Management
GMM	Gaussian Mixture Model
HMM	Hidden Markov Model
HNR	Harmonic to Noise Ratio
LLD	Low Level Descriptor
LPCC	Linear Predictive Cepstral Coefficients
MFCC	Mel Frequency Cepstral Coefficients
MOD	Module file format representing music
SVM	Support Vector Machine
WAV	Windows Audio file format

CHAPTER 1

INTRODUCTION

Human-machine interaction technologies are growing rapidly in parallel with algorithmic developments. Speech technologies are one of the most popular technologies in this field. The future of speech applications will be better in analysing the data behind the words. Perception of human psychology through speech through computers is one of the turning points. The main reason for this is to improve the quality of life by making computers understand people better than any person can. With the help of such innovations, creating a reliable environment to eliminate negativity, meeting human needs will improve the quality of work.

Determining stress, which is one of the parameters that affect job quality in every area, makes this situation very difficult due to the complex biology of the human. The presence of conditions that can cause stress, the manifestation of visible reactions as a result of their effects on humans, opens a door for measurements.

In areas where communication is intense and critical decisions are made frequently due to the job, errors can lead to decreases in job performance and even fatal consequences. Aviation is at the top of these areas.. Since the accident, error prevention models are applying to the eliminate these failures in aviation industry, it is the purpose of to close the gaps [1][2].

The human factor has a great impact on aviation accidents [3][4]. For example , the accident known as the Tenerife disaster on 27 March 1977, is known as the deadliest accident in aviation history . Two Boeing 747 of KLM and Pan Am airlines, collided on the runway of the small airport of Spanish Island of Tenerife, killing 583 people. It turned out that one of the reasons for the collision was a communication disruption

and misunderstanding between the KLM pilot and the air traffic controller. In the accident, the KLM captain both used non-standard phraseology and took off over the runway, thinking that he received ATC clearance [4][5][6][7]. This accident showed the importance of communication and using standard phraseology in aviation.

Another accident was on 1 July 2002, over Überlingen. The Tupolev Tu -154 of Bashkirian Airlines Flight 2937 and a Boeing 757 cargo jet of DHL Flight 611 collided in mid-air. All 69 passengers and crew of Bashkiriian Airlines and DHL Flight crew were killed. After the investigation, it is indicated that one of the reasons for the collision of the Swiss Air Traffic Control crew workload and instruction error. There were shortcomings of the controller in the shift, traffic congestion, and traffic collision avoidance system error was also the case [4][8].

These are some examples showing the result of the communication errors and the stress effects due to workload on human error. The correct communication is the main part of aviation safety. By taking preventative measures it is predicted that such risks can be eliminated. If it is succeeded to observation automatically through a non-invasive technology on the controller, it will help to minimize such accidents and incidents.

1.1. Scope

The scope of this thesis is the stress detection in the actual speech of Air Traffic Controllers'. On this aim, it is to prepare a substructure for stress and emotion recognition studies consisting of Turkish-English conversations now and in the future and to make classification studies to detect occupational stress in the field of Air Traffic Control.

1.2. Methods

In these studies, stress levels were classified as low (L), medium (M) and high (H) according to air traffic complexity levels [9,10]. The speech dataset, which is the subject of stress assessment, was created for this study. It consists of non-artificial

expressions. It has unique features in terms of spoken languages and Air Traffic Control content.

Aerodrome simulator environment was used to create the required data set in the studies. Natural speeches selected from the voice and video recordings of the students of the Air Traffic Control Department of Eskişehir Technical University, who have met the professional criteria and will be working as real air traffic control operators, were used.

1.3. Motivation

Since there are people in every field in the aviation field, the human factor has become very important. Simple mistakes, especially in areas where communication is intense and critical decisions are often made due to business reasons, can lead to serious financial and time losses or fatal consequences. As the preference of air transportation becomes widespread as a result of the technological development of air transportation, considering the undesirable situations experienced in the past, reliable and effective air traffic management has gained importance in this field.

While the share of ATMs in aviation accidents was quite low, it was stated that the effects of these human factors that caused accidents and incidents were 90% and above, considering the incidents related to ATM. It was stated that human factor studies are very important in technology-intensive areas such as ATMs and studies are carried out to determine the factors that may adversely affect the performance of the responsible person. Although a decrease in controller performance does not always cause accidents, it has been stated that it causes delays in air transport and inefficiency of the desired efficiency. Therefore, air traffic controller performance is desired to be at the desired level. It is aimed to benefit from technological developments that can improve standards and performance positively [11].

The environments in which air traffic controls work are highly dynamic, noisy and where attention should be focused. Therefore, the measurements to be made are expected to be in a structure that will not disturb the motion and attention factors of the controller. Since it is a business where speech is at the forefront, the most appropriate method is thought to be analysis over speech.

Researches in this area are given great importance in the world. In the future, it is thought that automation will become more prominent in aviation as in many other fields. Studies in the context of artificial intelligence in aviation are supported and researches are carried out. One of these supports is in the air traffic control area. In order to increase the reliability of air traffic, studies are supported by SESAR Joint Venture within the scope of Eurocontrol and European Union Horizon 2020 research and innovation program. According to the European Master Plan, the Air Traffic Management modernizations are planned in the long term sustainable solutions are predicted [12]. The motivation of this study is to contribute air traffic control field and create a basis for future studies.

CHAPTER 2

DEFINITIONS /CONCEPTS/ LITERATURE REVIEW

2.1. Stress in Air Traffic Management (ATM) and Environment

Sources state that the emergence of the term stress has a history dating back to the 16th and 17th centuries in the Oxford English Dictionary, and its use has become frequent since the Second World War [13, 14]. Hans Selye, who is an Austrian endocrinologist, carries out the most influential research in this field. The theory of his known as General Adaptation Syndrome states that reactions are predictable according to the source of stress. He developed this theory based on the functioning of the nervous and endocrine system, and accordingly divided the response to stress into alarm, resistance and fatigue phases, respectively [13, 15]. Contrary to this theory, reactions to stress are personal and difficult to understand from the outside [13, 16].

Within the definition of stress in the field of psychology, it has been introduced in various ways over the years. In terms of meaning, very broad and very special definitions have been made. For example, in a definition made by McGrath, if stress is low in anticipation, it is expressed as an imbalance, whereas in another definition by Lazarus it is considered as a way of evaluating the relationship between the person and his environment by risking himself and his resources [13, 17, 18]. Over time, there are critical thoughts stating that stress cannot be defined and measured, as stress is not adopted by everyone. Nevertheless, it is stated that stress against them is a psychological condition reflected in the relationship of people with each other and their environment [13, 19].

Although the definition remains uncertain, the view that stress is not a single aspect, but a concept with multiple parameters and processes is a more accepted approach [13, 18].

As it is listed in the Table 2.1, there are 3 approaches that are taken into account in determining stress [13]. According to these, environmental events, workload factors can cause stress. Personal differences, attitudes can also mean different responses to stress. Therefore, the effects of stress can be observed as physical and psychological consequences.

Table 2.1 Stress determination [13]

Measures	Examples
stressors	environmental events, situations, past experiences on life, workload factors
intervening variables	personal differences, personality traits, specific reactions to stressful occasions
strain outcomes	anxiety or physical symptoms

Beside these approaches, there are two proposed concepts of stress by Hans Selye. According to this concept, two stress definitions can be made, one is positive and the other is negative. Eustress, which causes positive changes, is beneficial to the person. Distress, on the other hand, is stated as the type of stress that has harmful physical and mental effects on the person and causes incompatible behaviour [20] [21].

One of the business areas where the quality of the work done can affect the result in a short time is Air Traffic Control. Air Traffic controllers are responsible for the management of air traffic. They ensure that a large number of planes within the

airspace are managed to operate safely and efficiently. The smallest numerical error or wrong instruction they make can cause human life and material damage. [11]

In aviation environment, safety and effectiveness are the most important issue directly affected by stress. In a study made in Human Factor Analysis in Air Traffic Control Environment, it is pointed that in a general framework, conditions of humans such as the adverse physiological and mental states, physical and mental limitations of the human are base for the unsafe acts conclusions [3]. Inter-personal differences are another aspect of these results.

2.1.1. Physiological Effects of Stress on Human in ATM

It is accepted that the effects of stress on people vary according to the person and cause disease in general. Accordingly, there are changes that occur in the human body during stress. These are changes in cardiovascular symptoms such as blood pressure, serum and cholesterol level, and biochemical factors such as catecholamine, cortisol, and uric acid [13]. To determine these changes, some medical tests required. Nevertheless, it is not possible to monitor these chemical changes in real time. Also, since the complex biology of humans can mislead the interpretation the validation of the results requires a medical expert.

Typical distress symptoms are state of high arousal, increased heart rate, excessive adrenaline production, failure of coping mechanism, feeling of strain, exhaustion and inability to concentrate, decreased quality of speech, skin perspiration increase muscle tension of the vocal cord and vocal tract [20, 22]. Nevertheless these reactions are quite personal. It is not expected to see all these reactions in each person under the same distress.

2.1.2. Psychological Effects of Stress on Human in ATM

Adverse mental conditions, such as stress, anxiety, motivation, mental fatigue directly affects the performance of Air Traffic Controller. Stress causes some behavioural results. Especially the behaviours that cause stress in professional life can be seen with effects such as change in business volume, accidents, and absenteeism. However, being able to determine that the source of these results is stress, is may be related to unbiased accurate measurements within organizations. Furthermore, psychiatric symptoms such as anxiety, depression presence might be taken as a sign to stress. Individual differences, such as genetic and trend factors, are the subject of stress measurement research. Such factors cause different responses to a stressor among individuals [13].

2.1.2. Emotion and Stress Relation

While stress and emotion are the result of similar biologic and psychological processes, emotions have sharper lines and more understandable images. The varying response types of stress make us think that it contains complex emotions. In some description, it is indicated that emotion covers psychological arousal, expressive behaviours and conscious experience [10].

There are growing studies in different disciplines of social and behavioural sciences in order to define, distinguish and measure emotions. These are showed that emotions are depending on neurophysiologic activation, motor expression and subjective feeling but possibly also action tendencies and cognitive processes. In those studies, there is 4-dimensional space which categorizes the emotions on closeness of their meanings. Those 4-dimensional elements are evaluation-pleasantness, potency-valence, activation-arousal and unpredictability. Since the complexity of emotion appraisal wheels are difficult to apply into computer science, they are simplified into 2-D or 3-D wheels. In Figure 2.1, it is shown a 2-dimensional wheel which is mainly used by the speech and video processing research groups [22][23]. According to this graph, there is a main two axis that reveal the emotion.

Arousal axis defines the activation level and valance axis locates emotion between pleasant-unpleasant evaluations. High arousal is accepted as the high activated region and low valance is accepted around the unpleasant region. Relation with stress in this graph is the distress location. This is defined in between high arousal and low pleasant placement in the middle of the upset and nervous emotions. Similarly; Alerted status also in the activated region and the valance level is at the bottom of the pleasant region. So, since the emergency situations in ATM could create some of these emotions, we might consider that stressful conditions and emotions are interrelated.

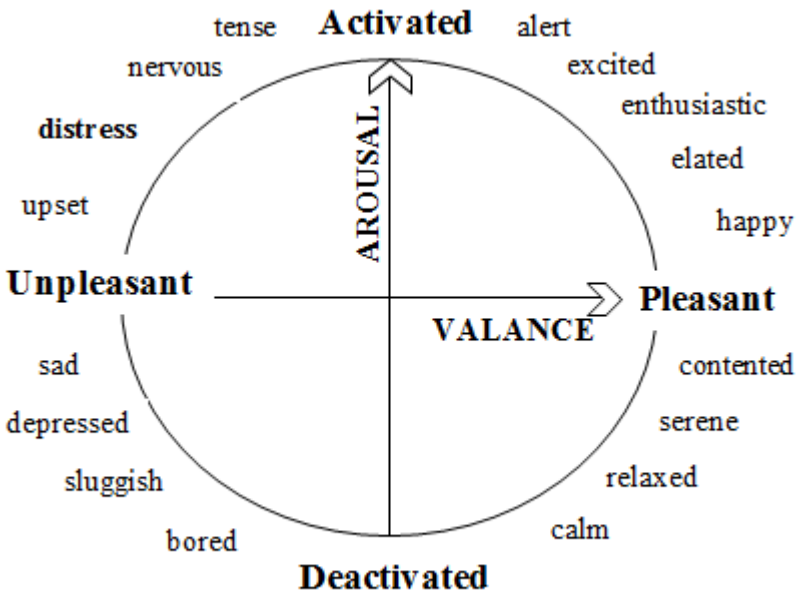

Figure 2.1 An emotion wheel on arousal-valance dimensions [22, 23]

2.2. Air Traffic Control (ATC) Environment

ATM is an integral part of advanced air transport. It is through ATM activities that pilots can complete their flights safely and efficiently in complex air traffic. It consists of ATS (Air Traffic Services), ATFM (Air Traffic Flow Management) and ASM (Airspace Management) parts [11].

In ATS part, control of aircraft is carried out by controllers with different responsibilities working in these 3 main units [11].

Tower Control Unit: The movement of aircraft on the ground and in the airport area is controlled. In this unit, controllers are divided in two groups as aerodrome controllers and ground controllers. The aerodrome controller is responsible for the work on the runway and for the control of flights in the area of responsibility. The ground controller is responsible for traffic in the manoeuvring area outside the tracks.
Approach Control Unit: Planes are climbed to cruise levels after take-off and guided by direction of departure. Sometime this is conducted by aerodrome control tower or area control unit.
Area Control Unit: The flights in the air corridor between the terminal areas where the departures and departure controls are carried out are checked.

2.2.1. ATC Operator Performance

ATCO performance may affect the efficiency of the air traffic management system positively or negatively. The conditions that affect their performance are depending on their communication and teamwork, stress and workload, individual differences, computing abilities, environment (automation and workplace design), selection and training, and organizational climate and job satisfaction [11].

2.2.1. ATC Stressors

Air Traffic Control Operator have to conduct critical decision making required works for long hours under high vigilance. Their main part of the job is depending on speech communication. By using their specific phraseology they give clear guidance to pilots. Therefore, a small error or unreliable tone in their voice may cause irreversible conclusions. Although controllers consider stress as part of their job, under very intense stress or when they are not at all stressed they are more likely to

make mistakes. Workload is such an important factor that affects the stress [1]. Stress effects in short periods may cause operation errors, and the long term effects can be irreversibly serious [16]. There are some other different sources of stress in this field as given in the Figure 2.2.

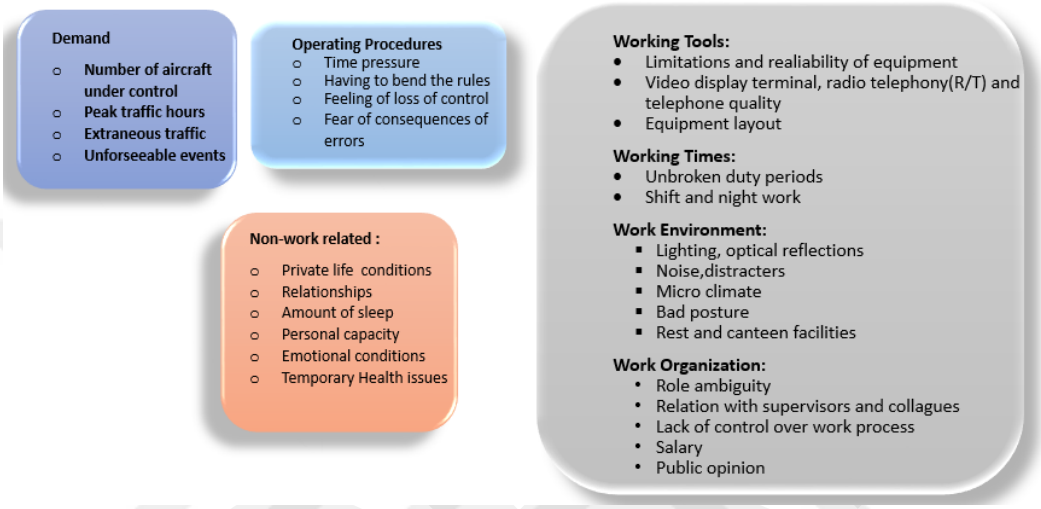

Figure 2.2 Possible sources of stress for ATC, modified from source [24]

If the effect response scheme is presented according to these factors, the biological manifestations of the stress on the controller are seen in the following Figure 2.3.

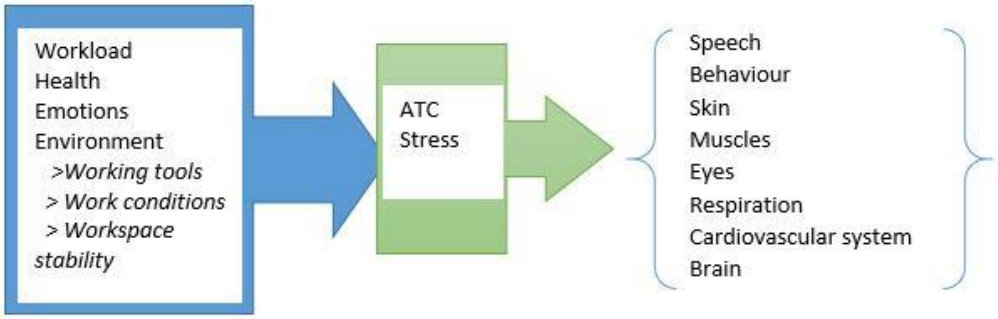

Figure 2.3 ATC stressors and effects, modified from source [24]

Stress has an effect on human body parts and as consequence on speech. It causes variations in acoustical properties on voice.

2.3. Voice and Stress Relation in ATC

The speech signal is created at the vocal cords, travels through the vocal tract, and produced at speaker's mouth. It gets to the listeners ear as a pressure wave. It is non-stationary, but can be divided to sound segments which have some common acoustic properties for a short time interval. There are two major classes as vowels and consonants [25]. The digital model of speech production is defined as discrete time model. It is simply depicted in the Figure 2.4 below.

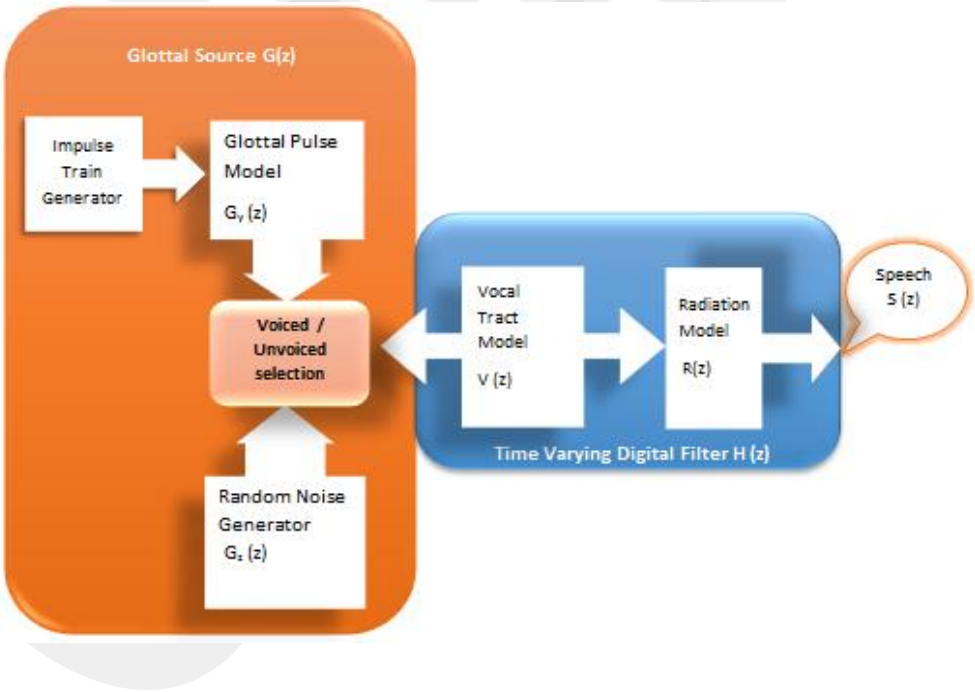

Figure 2.4 Digital model of speech production, modified from sources [26, 27]

The system function of the digital model of speech is obtained in the formula below (1).

$$H(z) = G(z).V(z).R(z) \quad (1)$$

Vocal tract model formula (2) corresponds to the creation of periodic sounds, noisy sounds impulsive sounds. The sources of the speech sounds created in this area are between vocal cords to the lips.

$$V(z) = \frac{G}{1 - \sum_{k=1}^{N} \alpha_k z^{-k}} \quad (2)$$

Radiation model formula (3), gives the effect of the pressure at the lips.

$$R(z) = R_0(1 - z^{-1}) \quad (3)$$

2.3.1. Voice Features

Speech is conducted between 20Hz and 4 kHz. It is assumed as stationary consist of linguistic and paralinguistic information. Linguistic information is qualitative patterns of speech such as words or phrases that are directly express emotions. Paralinguistic information is the quantitative feature of these linguistic patterns. It does not contain linguistic information but the way it is pronounced.

2.3.1.1. Non-Linguistic Features

Speech processing is followed by the feature acquisition, feature selection then classification. Important speech features to be consider in the speech emotion recognition and their descriptions are listed in the Table 2. below.

Table 2.2 Speech features and their description [26]

Features	Description
Mel-frequency cepstral coefficients (MFCCs), Linear prediction cepstral coefficients (LPCCs)	Derive from cepstrum, which is the inverse spectral transform of the logarithm of the spectrum
Formants (spectral maxima or spectral peaks of the sound spectrum of the voice),Log-filter-power-coefficients (LFPCs)	Derive from Spectrum
Noise-to-harmonic ratio, Jitter ,Shimmer, Amplitude quotient, spectral tilt, spectral balance	Measurements of Signal (voice) quality
Energy, Short energy	Measurements of intensity
Fundamental frequency (pitch)	Measurements of frequency
Temporal features (duration,time stamps)	Measurements of time

Speech vector features are categorized according to temporal structure namely suprasegmental and segmental. Suprasegmental features are calculated over long time duration of utterance. Segmental features are are calculated over small time frames (milliseconds) using windowing techniques.

Furthermore, as it is listed in the Table.3. , speech vector features also classified in under two other parameters, Low Level-Descriptors (LLDs) and functionals. LLDs contain prosodic features, which are suprasegmental and spectral features and their derivatives that are segmental. The second class (functionals) includes statistical features that derive from LLDs and therefore, they are suprasegmental features.

Table 2.3 Speech feature categorization according to their temporal structure (suprasegmental vs segmental) and parameterization (LLDs, vs functionals) [26]

Low-level descriptors (LLDs)	**Functionals (applied to LLDs)**
Suprasegmental features (Long Time, duration of utterance)	Extreme values (maximum, minimum), Means(arithmetic, quadratic, geometric), Moments (standard deviation, variance, kurtosis, skewness), percentiles and percentile ranges, quartiles, centroids, offset, slope, mean squared error, sample values, time/durations
Fundamental frequency (Pitch), energy , intensity, harmonics-to-noise ratio (HNR), shimmer, jitter, speech rate, normalized amplitude quotient, spectral tilt, spectral balance	
Segmental features (Short Time, msec)	
Mel frequency cepstral coefficients (MFCCs), Linear prediction cepstral coefficients (LPCCs), Log-filter power coefficients (LFPCs), formant amplitude, formant bandwidth, formant frequency, line spectral pairs, short(Frame) energy, frame intensity	

In early studies, prosodic features especially pitch, duration and intensity was considered with small feature sets. Later, researches tend to enhance the feature sets in order to create novel features. By performing statistical analysis with functionals, voice quality features of LLDs, HNR, jitter, shimmer, and spectral and cepstral measurement have been derived and extensively used. Rhythm and sentence duration were included, along with classical measurements, such as pitch, energy and formants, as a classification feature [26].

In order to analyse the harmonicity of the sound, there are some methods to look at. These parameters can be used to describe the voice quality. The autocorrelation, which is a statistical method, is to compare a sound signal with the delayed version

of itself. According to the formula at (4) given by the Boersma [31] the total change is defined according to the sampled speech signal at t and multiplying of the discrete times at which the same signal is defined. τ is defined as time lag and x is the sampled speech signal.

$$r_x(\tau) \equiv \int x[t].x[t+\tau]\,.dt \qquad (4)$$
$$x = sampled\ speech\ signal\ ,\ \tau = time\ lag$$

Harmonicity of the speech gives the acoustic periodicity and known as HNR or NHR. HNR is described as the ratio of the speech signal to the noise in logarithmic scale and emphasizes the signal part. HNR value is infinite for perfect periodic sounds according to the formula at (7) given by Boersma [28]. There is a relation in HNR and the autocorrelation function. Autocorrelation at the zero lag gives the signal and the complement of it gives the noise part. So NHR is easily obtained, when the noise factor in the sound needs to be emphasized.

$$r_x(\acute{\tau}_{max}) = \frac{r_H(0)}{r_x(0)} \qquad (5)$$

$$1 - r_x\acute{(}\tau_{max}) = \frac{r_N(0)}{r_x(0)} \qquad (6)$$

$$HNR = 10.\ \log_{10}\frac{r_x\acute{(}\tau)}{1-r_x\acute{(}\tau)} \qquad (7)$$

According to Boersma [28], HNR value of a healthy speaker will be between 20-40dB and below the 20dB it will be a hoarse sound.

In studies, it is pointed that suprasegmental features for the input feature vector which identify emotions, has a better achievement than segmental features. Through extensive statistical analysis, suprasegmental parameters can be obtained from

segmental features. For example, traditional segmental features MFCCs and LPCCs are transformed to suprasegmental parameterizations through statistical processing to obtain prosodic feature vector [26].

In speech processing, there are free software programs for labelling, spectrographic analysis and pitch analysis in phonetics available for researches. In this thesis, PRAAT software is preferred [29].

2.3.2. Speech under Stress

According to the speech production model at the left hand side in Figure 2.5, the stressors and their sources are given in the Table 2.4.

The third level stress effect is mentioned as psychological stress factors. These factors are at the initial stage of speech production, where conscious thinking is affected. At this stage, mental abstract thoughts are at the forefront.

In the second stage, stress factors intuitively affect the neuromuscular and cause changes in the muscle group that control speech, which may affect the way of conversation. At this stage, para-linguistic features come to the fore.

The first level stress effect, physical changes on the voice change speech. This effect can be accompanied by other factors such as insomnia, medication use, various diseases and thirst.

There may be conditions such as zero level stress effect, vibration and acceleration. The stress effect acts as soon as the sound appears and the change is evident.

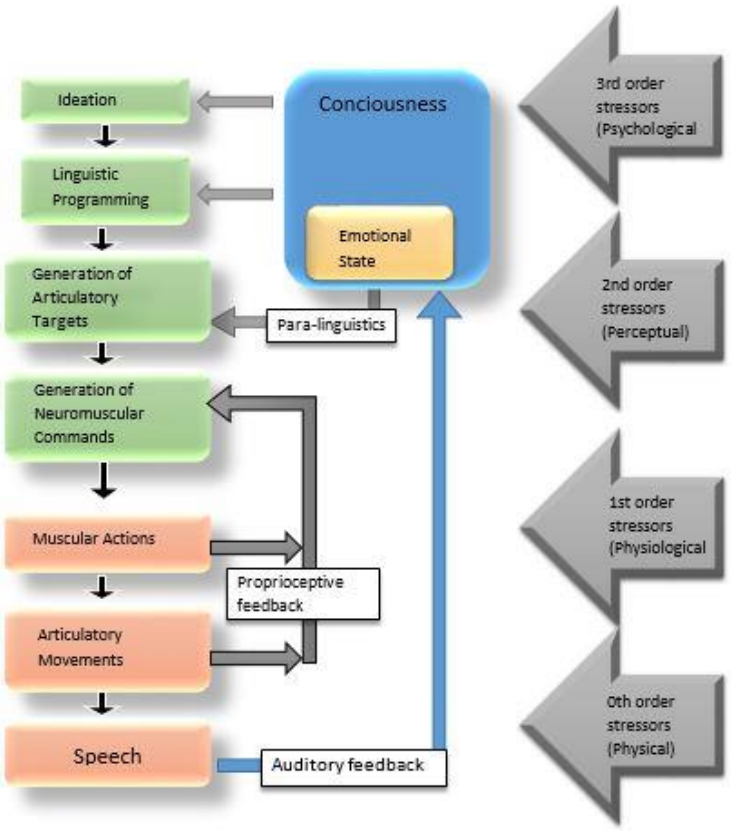

Figure 2.5 Model of Speech Production Under Stress [24, 30]

Table 2.4 Speech Production and Stressor Sources

Speech Production Stage	Stressor	Stressor Sources
Consciousness, Emotional Stage, Ideation	3rd order Stressor Psychological	Emotion, workload, anxiety, depression
Linguistic Programming, Generation of Articulatory Targets	2nd order Stressor Perceptual	Speech quality, noise,
Generation of Neuromuscular Commands, Muscular Actions	1st order Stressor Physiological	Illness, medicine, fatigue, dehydration, nicotine,
Articulatory Movements, Speech	0th order Stressor Physical	Vibration, acceleration, physical workload, personal equipment for breathing or speaking

Speech and stress productions are both complex processes of human brain. In some studies considered features for stress analysis are available. These are focused on the

change of non-verbal vocal characteristics of speech. The most related features are pitch, intensity, duration, glottal source properties, vocal tract spectrum and Teager energy operator (TEO).

Stress is a combination of various human emotions. There are several studies on emotion recognition. Nevertheless there is not an exact definition of stress. However, stress causes anxiety and mental issues. In a study, it is assessed that stressed speech is parallel to the fear/ anxiety emotion [27]. In the studies related to speech under stress, pitch, articulation rate, energy, Mel frequency cepstral coefficients (MFCCs) are considered. Mean energy; mean intensity is used to distinguish five different emotions. These are happiness, disgust, sadness, anger and fear/anxiety. MFCC, mean energy, mean intensity are used to classification of emotions [27].

Speech features affected by the stress are such as pitch, energy, duration, jitter, shimmer and individuals' glottal source and vocal tract properties are descripted below.

2.3.2.1 Pitch

Pitch as known as fundamental frequency (F_0) is worked in many emotional recognition and stress related studies. It is basically the average vibration of the sound, perceived by the human ear. Variation in the compression and rarefactions of the sound wave affects the perceived frequency. The source of the speech sound is the vocal tract and vocal cords. These biological organs have different structural variations in each person. Because of this, the created sound has a characteristic in its' period. Estimating the average alternation in time provides valuable material in speech processing.

In general acceptance women and men speak in different frequency range. According to this, while womens' fundamental frequency may differ in the range of 120 and 200 Hz, mens' differ in the range of 100-150Hz [31]. In some sources this

ranges are accepted for women from 155 Hz to 334 Hz and for men from 85 Hz to 196 Hz [32]. And the mean frequency in each person can also be affected by environmental factors, mental and health conditions.

A study made on aviation environment, pointed out that when dangerous situations increase, F_0 level of the voice increases as well [33]. Stress is also causes increase in F_0 range and F_0 contour. A psychiatric study also pointed out that, depressive patients speak with higher F_0 and larger proportion of the high frequency components [33]. In an early experiment by listening stressful speeches showed that a change in the mean pitch and maximum pitch has been an indication of stress [34] .In a different study, mean pitch values and pitch variance together used to determine the several stressed speeches [34].

Pitch related features used in the studies are listed in the Table 2.5 below [34];

Table 2.5 Pitch Related Features.

Pitch (F_0)	Pitch range
Pitch contour	Pitch variance
Jitter (short-term pitch variations)	Maximum Pitch

2.3.2.2 Energy

Teager energy operator (TEO) feature is designed by Teager and Kaiser based on the hearing air flow in the vocal tract system where the sound is produced [35]. In stressful conditions, a tension in the biologic structure of speaker effects the physical oscillation process thus air flow changes [35]. TEO is based on the observation of energy stored in a physical oscillation process [34]. It is a non-linear feature of speech basically designed to identify the stressed emotions [35]. The TEO is usually applied to voiced sound.[34] In a study it is observed that emotions such as happiness

and anger have high energies at the high frequencies. Nevertheless sadness has low energy [27].

Simple form of TEO is at (8);

$$\psi[x(t)] = \dot{x}^2 - x\dot{x} \tag{8}$$

$$\psi[.] : \ Teager\ Energy\ Operator$$
$$x(t) : continuous\ speech\ signal$$

Discrete Time Form of TEO is at (9);

$$\psi\{x[t]\} = x^2[n] - x[n+1]x[n-1] \tag{9}$$

$$x[n] : sampled\ speech\ signal$$

In the studies it is found that TEO to be useful for amplitude modulation (AM)-frequency modulation (FM) energy tracking . The acknowledged model for speech is assumed as the sum of AM-FM signal and each is centered at a formant frequency . In an early study, the TEO to track F0 and formats are used and some relations found between them [34].

In some studies, combination of features proposed to identify stressed emotions. Variations of FM Component (TEO-FM-Var), Normalized TEO Autocorrelation Envelope (TEO-Auto-Env), Critical Band Based TEO Autocorrelation Envelope (TEO-CB-Auto-Env) techniques used SUSAS database [27] The TEO-FM-Var and TEO-Auto-Env features are dependent on the pitch estimation accuracy, so they are not effective ways for stress classification due to their dependence on pitch estimation accuracy[27]. Later TEO-MFCC-SER technique used for speech emotion recognition and it improved the performance [35].

2.3.2.3 Duration

Some early studies are conducted based on classification of duration both of phonemes and words in speech under stress, nevertheless these results have not been shared [34]. In recent experiments which psychomotor tasks involved, show that difficulty of tasks leads to a rise in the pitch, average amplitude and the word duration [33]. Surprisingly word duration decreases if the difficulty level continues to rises up [33].

2.3.2.4. Glottal Source Properties

Due to the differences in the glottal structure among humans, speaking styles differ from person to person. Nevertheless, there are similarities and differences in those speaking styles in terms of the reaction to some emotion, in the velocity of speaking, or some environmental conditions could trigger similar outcomes in the glottal waveform.

To find out those parameters some experiments have been conducted. Different speaking styles are analysed in three stages of speaking and it is found out that there is a waveform for each speaking style. Also, workload effect has an impact on the style, nevertheless these results found rather confusing. In the conclusion there is a clear difference in the stressed and unstressed speech in speaker dependence. Also, natural , angry speech has a significant difference [26].

2.3.2.5. Vocal Tract Spectrum

The vocal tract spectrum consists of cepstrum based features, fourier transform coefficients or linear prediction coefficients (LPCs). It is a widely studied area. A study made to discriminate the speaking styles from the emotional speeches using SUSAS database. In the conclusion Log frequency power coefficients (LFPC) involved sets, got the higher achievements in classification. Another study conducted

for to test the seperability of phoneme groups and stress conditions. Cepstrum based features autocorrelation of mel-cepstrum and excitation based feature pitch gave the best results [34]

The other features which are jitter and shimmer, given below with their definitions [36, 37].

2.3.2.6 Jitter and Shimmer

Jitter is a measure of frequency instability, while shimmer is a measure of amplitude instability. In a study made in aviation simulation environment showed that, increasing of the fundamental frequency decreases the vocal jitter significantly [33]. So the decreasing of the fundamental frequency increases the frequency instability. In the Figure 2.6 , in a sample sound cut from a utterance, indicating the jitter and shimmer in sound wave. In the utterance the speaker have an hesitation and saying 'a' sound. Shimmer appears in such long sustained sounds. There is an increase and decrease in the volume of the sound in time. And jitter effect is like a shivering . The horizontal placed orange arrows shows the jitter, and the vertical placed green arrows shows the shimmer in the frequency.

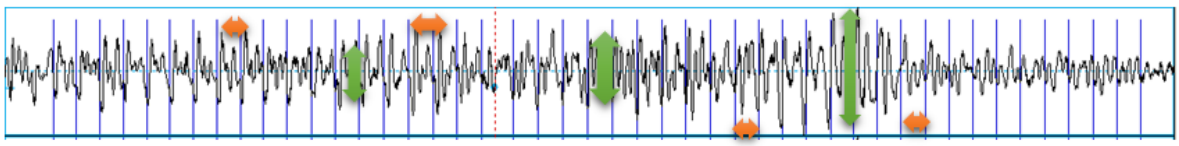

Figure 2.6 Jitter and Shimmer representation in a sound wave

In general form of jitter and shimmer formula are given below (10)(11) [33].

$$Jitter = \frac{1}{N-1}\sum_{n=1}^{N-1}(T_n - T_{n-1}) \quad (10)$$

$$Shimmer = \frac{1}{N-1}\sum_{i=1}^{N-1} 20.\log\left(\frac{A_i}{A_{i-1}}\right) \quad (11)$$

According to those formulas N is the number of periods extracted from the fundamental frequency, T is the duration in seconds which is the fundamental period, A is the peak to peak amplitude in the fundamental frequency.

Jitter and shimmer has variations. Those are given with the definitions along with the formulas below [38][39].

Jitter (abs): This is the average absolute difference between consecutive periods, in seconds. According to the formula if the number of period N is lower than 2, the result will be undefined.

$$\text{Jitter (abs)} = \frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i-1}| \tag{12}$$

Jitter (local): This is the average absolute difference between consecutive periods, divided by the average period. This value is between 0 and 2.

$$\text{Jitter (local)} = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i-1}|}{\frac{1}{N} \sum_{i=1}^{N} T_i} \cdot 100 \tag{13}$$

Jitter (rap): This is the Relative Average Perturbation, the average absolute difference between a period and the average of it and its two neighbours, divided by the average period. This value is between 0 and 2.

$$\text{Jitter (rap)} = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} \left| T_i - \frac{1}{3} \sum_{n=i-1}^{i+1} T_n \right|}{\frac{1}{N} \sum_{i=1}^{N} T_i} \cdot 100 \tag{14}$$

Jitter (ppq5): This is the five-point Period Perturbation Quotient, the average absolute difference between a period and the average of it and its four closest neighbours, divided by the average period. This value is beween 0 and 4.

$$\text{Jitter (ppq5)} = \frac{\frac{1}{N-1}\sum_{i=2}^{N-2}\left|T_i - \frac{1}{5}\sum_{n=i-2}^{i+2} T_n\right|}{\frac{1}{N}\sum_{i=1}^{N} T_i} \cdot 100 \qquad (15)$$

Jitter (ddp): This is the average absolute difference between consecutive differences between consecutive periods, divided by the average period. The extension of the abbreviation is difference of differences of periods. The result is the three times the Jitter(rap) measurement. The value is between 0 and 6.

$$\text{Jitter (ddp)} = \frac{\frac{1}{N-1}\sum_{i=2}^{N-1} |(T_{i+1} - T_i) - (T_i - T_{i-1})|}{N/2} \qquad (16)$$

Shimmer (local): This is the average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude.

$$\text{Shimmer (local)} = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|(A_{i+1} - A_i)|}{\frac{1}{N}\sum_{i=1}^{N}(A_i)} \qquad (17)$$

Shimmer (dB): This is the average absolute base-10 logarithm of the difference between the amplitudes of consecutive periods, multiplied by 20.

$$\text{Shimmer (dB)} = \frac{1}{N-1}\sum_{i=1}^{N-1}|20.\log(A_{i+1} - A_i)| \qquad (18)$$

Shimmer (apq3): This is the three-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of its neighbours, divided by the average amplitude.

$$\text{Shimmer (apq3)} = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|A_i - \left(\frac{1}{3}\sum_{n=i-1}^{i+1} A_i\right)\right|}{\frac{1}{N}\sum_{i=1}^{N}(A_i)} . 100 \quad (19)$$

Shimmer (apq5): This is the five-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of it and its four closest neighbours, divided by the average amplitude.

$$\text{Shimmer (apq5)} = \frac{\frac{1}{N-1}\sum_{i=2}^{N-2}\left|A_i - \left(\frac{1}{5}\sum_{n=i-2}^{i+2} A_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N}(A_i)} . 100 \quad (20)$$

Shimmer (apq11): This is the 11-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of it and its ten closest neighbours, divided by the average amplitude.

$$\text{Shimmer (apq11)} = \frac{\frac{1}{N-1}\sum_{i=5}^{N-5}\left|A_i - \left(\frac{1}{11}\sum_{n=i-5}^{i+5} A_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N}(A_i)} . 100 \quad (21)$$

Shimmer (dda): This is the average absolute difference between consecutive differences between the amplitudes of consecutive periods.

2.4. Machine Learning in Air Traffic Management

Machine Learning is an Artificial Intelligences' area of application that helps to create a model from the previous experiments to estimate the future events. The main aim of machine learning is to find out the most valuable data by developing and improving the algorithms. The algorithms use inductive method. The purpose of the inductive method is to reach the whole by using the separate parts. In this process generated inferences is used to make a prediction or identification in the next step [28].

Here there are five steps to perform of machine learning [40]. These are data collecting, data preparation, model training, evaluating the trained model and performance improvement stage. All these are defined in Table 2.6. across the headings below. These steps are data collecting, data preparation, model training, model evaluation and performance improving steps.

Table 2.6 Machine Learning Steps

Collecting Data	Collecting the raw data from the various sources. Old data is providing a basis for future learnings. The better the variety, density and volume of data the better the learning for machine learning
Preparing the Data	Covers the pre-processing operations for creating a learning model before data being given to the system
Training the Model	Covers the selection of the appropriate algorithm for creating the model and the process of creating a suitable model for the representation of the data. For supervised learning, data is divided into training and test sets. With the training set, the system will be run to create a suitable model for the data to be used in the testing phase.
Evaluating the Model	The model developed in the previous step for testing the accuracy of the system is applied to the test set and the results are recorded. The best way to test the accuracy of the model is to measure the success of the model on data that it has never seen before.
Improving the Performance	This step may require selecting a completely different model or providing more variables to increase productivity. Therefore, it may be necessary to pre-process a significant amount of data to be collected then modelled

2.4.1. Machine Learning Methods

Machine learning algorithms try to create the best model by using the available data. They try to analyse the new data by using the model created with the previous data in the new data collection. The creation of meaningful and useful information through the processing of the available data is called data mining. Data mining is classified under four main title in Figure 2.7.

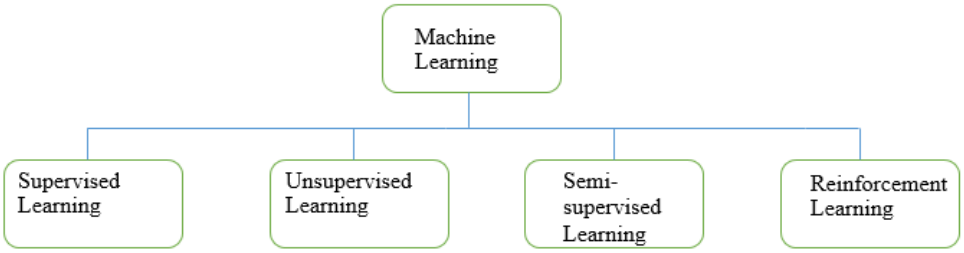

Figure 2.7 Machine learning branches [40]

2.4.1.1. Supervised Learning

Supervised learning has a supervisor to perform the learning process. The role of the supervisor here is, to feed the system with the training samples for the model to be learned. By matching the input and output, system is learned or the related model is to be created [40].

Supervised learning taken as a classification problem. Estimation and recognition is performed over test sets via the created model by trained system. The common supervised learning algorithms are Support Vector Machines, Artificial Neural Networks, Logistic Regression, Naïve Bayes, Multinomial Naïve Bayes, k-Nearest Neighbour, Random Forest and Decision Trees [40]. Basic structure diagram of supervised learning is given in Figure 2.8.

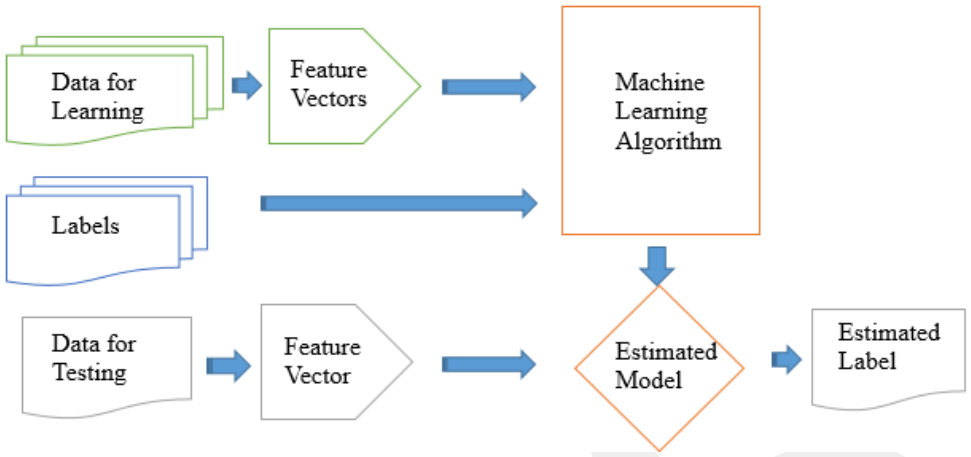

Figure 2.8 Supervised Learning Structure [40]

2.4.1.2. Unsupervised Learning

As it is seen in Figure 2.9, the difference of this method from the supervised learning is that there is no supervisor. There is only inputs feed into the system. The algorithms are aimed to create a model that will find a relation of inputs. There is not any pre-information of the output of the system.

The purposes of the unsupervised learning are such as clustering, probability density estimation, finding relation between attributes and size reduction.

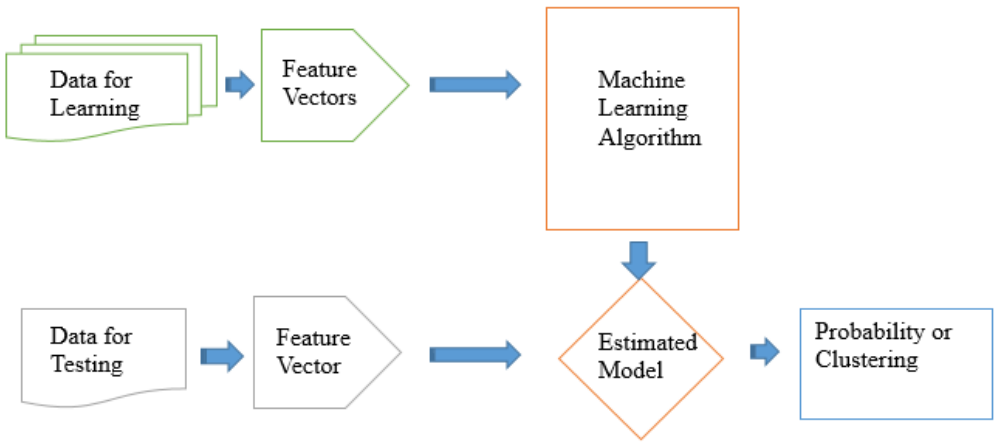

Figure 2.9 Unsupervised Learning Structure [40]

2.4.1.3. Semi-supervised Learning

There are unlabelled inputs and a few number of labelled inputs feed into system. This method is preferred when there are less labelled data sets and when it is easy to obtain unlabelled data. In Figure 2.10, the basic learning structure diagram is given.

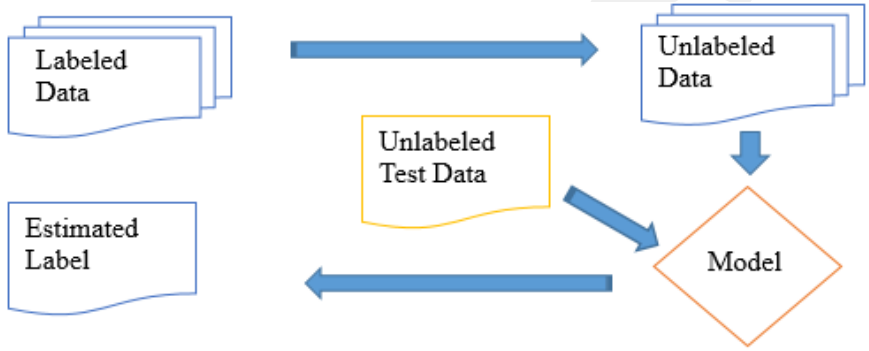

Figure 2.10 Semi-supervised Learning Structure [40]

2.4.1.4 Reinforcement Learning

This is like the supervised learning. There is a supervisor and it compares the expected output of the input feed and the calculated output, then returns this result to the system as true or false. If the calculated output is true, system is awarded otherwise it is punished. By doing this system is forced to learn. In Figure 2.11 this structure is given.

The most common reinforcement learning algorithm is Q-Learning algorithm. This algorithm is designed to find to learn the most convenient way in a random medium.

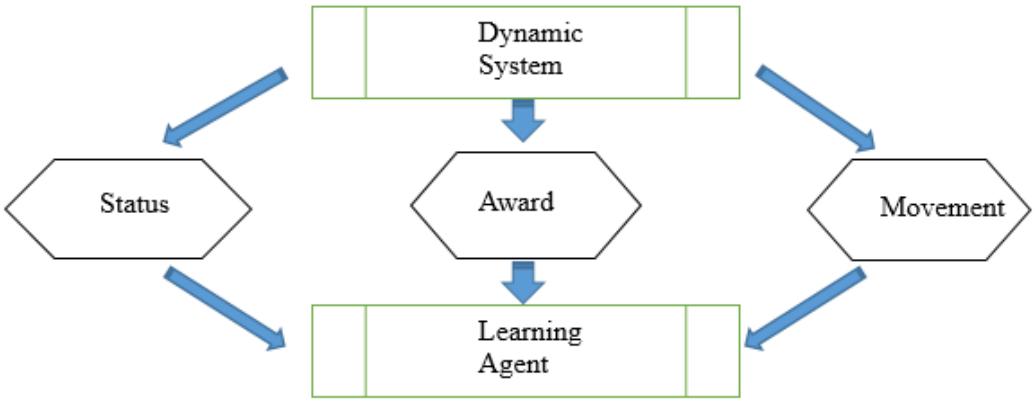

Figure 2.11 Reinforcement learning structure [40]

2.5. Most Used Classifiers of Stressed Speech

2.5.1. Support Vector Machines (SVMs)

In many papers from 2005 to 2010, SVMs have been assessed as a promising classification schema for emotion recognition in speech. SVM classifier is basically Transforming the original set of feature to a higher dimensional feature space by using the kernel function, which required to get optimum classification in this new feature space [26].

2.5.2. Hidden Markov Models (HMMs)

Hidden Markov Models (HMMs) are another widely used classification technique for speech emotion recognition and especially for stress state recognition. Hidden Markov Model is a statistical Markov model.

2.5.3. Gaussian Mixture Models (GMMs)

GMMs are the most studied among all classifiers. GMM is an unsupervised learning model. They are probabilistic models for density estimation representing by normally distributed subpopulations within an overall population.

2.5.4. Artificial Neural Networks (ANNs)

Artificial Neural Networks have various architectures such as Multi-Layer Perceptron's (MLPs), Probabilistic Neural Networks, Vector Quantization Networks, Deep Neural Network.

CHAPTER 3

METHODOLOGY

The information in this section is compiled from classification studies covering stress, emotion and speech. The methods and results related to the emotional context were taken to give an idea

Basically, the general speech processing framework has pre-processing, feature extraction and classification phases. To analyse these are given in the Figure 3.1.this structure can be modified according to the study.

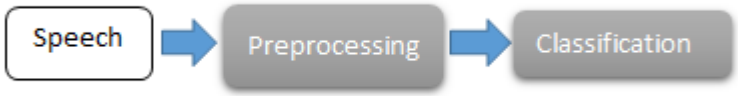

Figure 3.1 General framework for speech processing

3.1. Databases related with Stressed Speech

Various databases found during literature searches were mostly created in emotional speech for psychological studies. Among these, databases that may be related to this thesis were examined. Some of these are no longer available, some are not free, and some do not meet the needs in terms of size and content. The information and evaluations of the datasets that are most used in the market and the closest to our subject are explained under the headings below.

SUSAS (Speech under Simulated and Actual Stress)

In addition to the 35 terms used in the aviation field, this database contains natural and simulated records taken from 32 men and women in a wide range of ages, in various stressful situations, environments, and speaking styles. This database requires payment [41].

VOCE Corpus

This corpus consists of public speaking data, heart rate measurements and metadata file which contain information about the speaker such as gender, age, health information, experience in public speaking STAI (State-Trait-Anxiety Inventory) test scores and information about the recording quality. There are 135 recordings of ages between 19-49 years of 45 people, 21 men, 17 women and 7 unidentified from the University of Porto [42].

UT-SCOPE

This dataset is for physical task stress speech researches. Physical stress was induced with an elliptical, stair stepper machine. There are 35 sentences of 72 speakers, in neutral state and in 3 stages of stress. It is not publicly available [43].

Berlin EMO Database

This database is emotional speech database in German. It contains about 800 utterances spoken by actors in a happy, angry, anxious, fearful, bored and disgusted way as well as in a neutral version. The utterances are from 10 different 5 male and 5 female actors and 10 different texts [44].

3.1.1. Databases related with Air Traffic Control

With the development of technology, especially voice recognition applications started to be applied in the field of Air Traffic Control. To this end, there are a number of databases in this area. Real-life ATC databases nnMTAC and VOCALISE are not available [45].

AIRBUS-ATC is funded for speech recognition to process ATC communication and their data creation methods were published. It consist of real life, French accented speech. It consists of 59 hours to 100 hours transcribed English audio and related metadata [45].

ATCOSIM

This database is for Civil Air Traffic Control. There are 10 non-native speakers in 3 accents, in both gender. It does not contain a real life data. Records are made in simulated environment. There is not a metadata available with the recordings. This database is available [46].

HIWIRE

This is a database of military air traffic control. It contains 8100 utterances, read or prompted words and sentences of radio phraseology, pronounced in 4 different accents by 81 non-native speakers. The recordings are not from real life, they have made in a studio setting, and cockpit noise was artificially added. Additive noise levels have also indicated as low, mid, and high conditions. This database is available according to the request [46].

NIST AIR TRAFFIC COMPLETE CORPUS

This corpus is created by National Institute of Standards and Technology (NIST) in 1994 for speech recognition applications in such domains with using small vocabulary set, several speakers and noisy channels. It consists of 70 hours of real-life ATC radio recordings from native US speakers, from 3 different civil US airports. It contains controllers and pilots radio communications, transcriptions. This corpus is commercially available [46].

3.2. Pre-processing

This stage is an important part of speech processing. It includes some optional process such as noise reduction, data selection, normalization, transformation and a necessary step of feature extraction.

To look at the feature extraction aspect, in Emotional Speech Recognition applications, the most frequently used parameters are; formant frequencies, bandwidths, Pitch, Log-Energy, Normalized 1.st Order Autocorrelation Coefficients On the other hand for stressed speech recognition Mel Frequency Cepstral Coefficients (MFCC) and prosodic features, pitch(F0), energy, word duration, Formants) are frequently used features. Emergency, emotional load, cognitive load, sleep deprivation voicing and voiceless transients, relative average perturbation, jitter, shimmer, MFCC are the related speech features.

3.3. Stressed Speech Classification Methods

Stressful speech classification methods are as follows in the literature [22].

- Based on Feature Parameters
- Based on Linear Features
- Based on Nonlinear Features
- Based on MFCC
- Based on Sub bands

3.4. Classifier Comparisons

3.4.1. Support Vector Machines (SVMs)

In many papers from 2005 to 2010, SVMs have been assessed as a promising classification schema for emotion recognition in speech. SVM classifier is basically Transforming the original set of feature to a higher dimensional feature space by using the kernel function, which required to get optimum classification in this new feature space [26]. It is more efficient for the speaker dependent classifications

comparing with speaker independent classifications. Advantages of SVMs over GMM and HMM are the global optimality of the training algorithm and the existence of excellent data-dependent generalization bounds. Disadvantage is over fitting problem due to perfect separation of training data. So, SVMs success in non-separable cases is relatively heuristic. Over fitting problem, studied in 2009 using Twins Support Vector Machines (TWINsSVM). Comparing with standard SVMs, TWINsSVM gives better performance [26].

3.4.2. Hidden Markov Models (HMMs)

Comparing with the GMMs, HMMs have more training and testing requirements. In the application, emotions are individually modelled by a single state HMM and by a scaling factor emotions distinct from each other. Unlike GMMs, HMMs are random processes and their states are hidden from observer. HMM can be discrete type like single state HMM or continuous HMM. Creating a HMM structure has these critical steps: Determining optimum number of state and the type of the observations. For discrete HMM, the optimal number of observation symbols and for continuous HMM, the optimum number of Gaussian components are other important factors.

3.4.3. Gaussian Mixture Models (GMMs)

GMMs are more efficient for global feature extraction (functionals: extreme values, means, centroids, etc.) of emotion from the training utterances. The algorithm is based on expectation-maximization or Maximum a Posterior (MAP) Parameter Estimation. All the vectors are independent therefore GMM cannot form temporal structure of the training data.[24] GMM achieved maximum efficiency of 78.77% using the accurate features of speech signal. In speaker dependent system calculated 89.12% for recognition performance using GMM and obtained typical performance of 75% using speaker independent recognition system. [47].

3.4.4. Artificial Neural Networks (ANNs)

ANNs have better classification performance than HMM and GMM when the number of training and testing examples relatively small and better efficiency in modelling nonlinear mappings. Mapping is the possible optimization for classification of two separate data class points linearly. For nonlinear values, there is another solution set in upper space. Optimizing and ANN is a hard task since there are various parameters which effects the output performance and there is no definitive procedure to create an ANN topology. In the literature, classifier performances for emotional database namely Berlin Emo–DB are given below.

In Table 3.1 Classifier performances for Berlin Emo Database among different research papers are given. According to these SVM has better performance in overall. The tables below were modified from source [26.]

Table 3.1. Classifier performances for Berlin Emo DB

Database	Author	SVM	GMMs	HMM	ANN	Hybrid
Emo-DB	Schuller et al. (2005a)	87.50%				
Emo-DB						
Emo-DB						80.5% (SVM, K-NN, Naïve Bayes, C4.5, ANN)
Emo-DB	Vlasenko et al. (2007)	~90%				
Emo-DB	Wu et al. (2009)	88.60%				
Emo-DB	Yang et al. (2009a)	89%				
Emo-DB	Luengo et al. (2010)	78%				
Emo-DB speaker independent	Atassi and Esposito (2008)		81%			
Emo-DB speaker independent	Lugger and Yang (2007a)		74.60%			
Emo-DB	Mishra and Sekhar (2009)		~63%			
Emo-DB	Yun and Yoo (2009)			89%		
Emo-DB speaker independent	Fu et al. (2008b)			78.40%		

Emo-DB	Fu et al. (2008a)				63.30%	
Emo-DB speaker dependent but utterance independent	Anagnostopoulos and Vovoli (2010)				47%	
Emo-DB speaker dependent	Iliou and Anagnostopoulos (2009)				83.20%	
Emo-DB speaker independent	Iliou and Anagnostopoulos (2009)				55%	

Table 3.2 Classifier performances for 3 emotions in unknown datasets

Database	Author	GMMs	C4.5	RF	Hybrid
3 emotions (neutral, emphatic, negative)	Neiberg et al. (2006)	90%			
3 emotions in two unknown datasets	Rong et al. (2007)		76.50%		
3 emotions in two unknown datasets	Rong et al. (2007)			80.60%	
3 emotions in two unknown datasets	Rong et al. (2007)				78.1% C4.5, RF

For only 3 emotion used classification studies show that GMMs have better rates .And increasing of emotion sets gave better and promising results with hybrid classifiers. In Table 3.2, Table. 3.3, Table 3.4. , Table 3.5 and Table 3.6 gave the results from these examples. In those, unknown datasets are used for 3, 4, 5 and 6 emotions and along with the unknown dataset. Berlin Emo DB used for 7 emotions.

Table 3.3 Classifier performances for 4 emotions in unknown datasets

Database	Author	SVM	GMMs	HMM	ANN	Hybrid
4 emotions in 2 unknown datasets, (speaker independent)	Wu and Liang (2011)	75.33%				
4 emotions in 2 unknown datasets, (speaker independent)	Wu and Liang (2011)		68.73%			
4 emotions in 2 unknown datasets, (speaker independent)	Wu and Liang (2011)				69.86%	
4 emotions in 2 unknown datasets, (speaker independent)	Wu and Liang (2011)					80% SVM, GMM, MLP
4 emotions in 2 unknown datasets (speaker independent with linguistic information),	Wu and Liang (2011)	78.16%				
4 emotions in 2 unknown datasets (speaker independent with linguistic information),	Wu and Liang (2011)		72.61%			
4 emotions in 2 unknown datasets (speaker independent with linguistic information),	Wu and Liang (2011)				71.87%	
4 emotions in 2 unknown datasets (speaker independent with linguistic information),	Wu and Liang (2011)					83.55% SVM, GMM, MLP
4 emotions (unknown dataset)	Firoz Shah et al. (2009)			68.50%		
4 emotions (unknown dataset)	Wenjing et al. (2009)				71.40%	

Table 3.4 Classification Performances for 5 emotions in unknown dataset

Database	Author	GMM	HMM	k-NN	Hybrid
5 emotions (unknown dataset)	Pao et al. (2007a)	70.30%			
5 emotions (in two unknown datasets)	Yu (2008)		~87%		
5 emotions (unknown dataset)	Ijima et al. (2009)		~81%		
5 emotions (unknown dataset)	Pao et al. (2007a)		62.50%		
5 emotions (unknown dataset)	Pao et al. (2007a)			72.20%	
5 emotions (unknown dataset)	Pao et al. (2007b)				83.8% SVM, K-NN

Table 3.5 Classifier performances for 6 emotions in unknown dataset

Database	Author	SVM	ANN	RF	k-NN	Hybrid
6 emotions (unknown dataset)	Morrison et al. (2007)	71.85%				
6 emotions (unknown dataset)	Morrison et al. (2007)		65.37%			
6 emotions (unknown dataset)	Morrison et al. (2007)			67.36%		
6 emotions (unknown dataset)	Morrison et al. (2007)				61.83%	
6 emotions (unknown dataset)	Morrison et al.(2007)					73.3%, (SVM, MLP,K-NN, RF)

Table 3.6 Classifier performances for 7 emotions in unknown and Berlin(Emo-DB)

Database	Author	SVM	HMM	C4.5	Hybrid
7 emotions (unknown dataset)	Schuller et al. (2003)		86%		
7 emotions (large unknown dataset)	Schuller et al. (2005c)	70.30%			
7 emotions (large unknown dataset)	Schuller et al. (2005c)			50%	
7 emotions (large unknown dataset)	Schuller et al. (2005c)				71.62% SVM, K-NN, Naïve Bayes, Boosted C4.5
several cross-corpus experiments with varying number of classes	Schuller et al. (2010)	up to 81%			
7 emotions (Emo-DB)	Schuller et al. (2005a)	87.50%			
7 emotions (Emo-DB)	Schuller et al. (2005a)			61.50%	
7 emotions (Emo-DB)	Schuller et al. (2005a)				80.5% SVM, K-NN, Naïve Bayes, C4.5, ANN

CHAPTER 4

EXPERIMENTAL STUDIES

4.1. Aerodrome Simulator Environment

Aerodrome Simulator is located in the Air Traffic Control Department of Eskisehir Technical University. It is the latest technology used in Air Traffic Controller education. It has a 360 degree of vision through 10 ceiling projectors, able to monitor pre-installed six large Turkish airports and a generic airport. Simulator software itself has a voice recognition technology that enables to self-practice for 12 work positions [9]. The software is also enabled to simulate several aircraft types, air carriers, weather conditions, and various emergency problems. The environment has a realistic effect, which gives the operators similar instincts.

It is a circular room with 360 degree projection, reached by stairs. Simulator controls were done by the instructor in the middle. Students can settle in 4 positions in the tower where they can control the landing, take-off and ground positions of the planes on both sides of the square. In the Figure 4.1 Aerodrome simulator is three different windows were shown and in Figure 4.2. the bird's eye view of the placements were depicted.

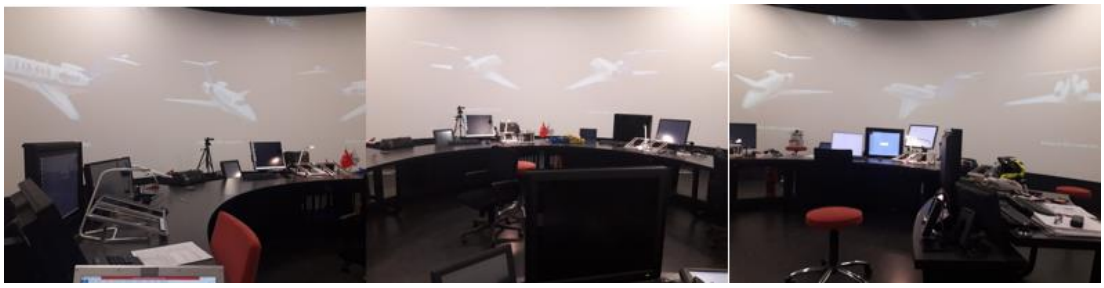

Figure 4.1 Aerodrome simulator at Eskisehir Technical University, (left to right, taken in 2019)

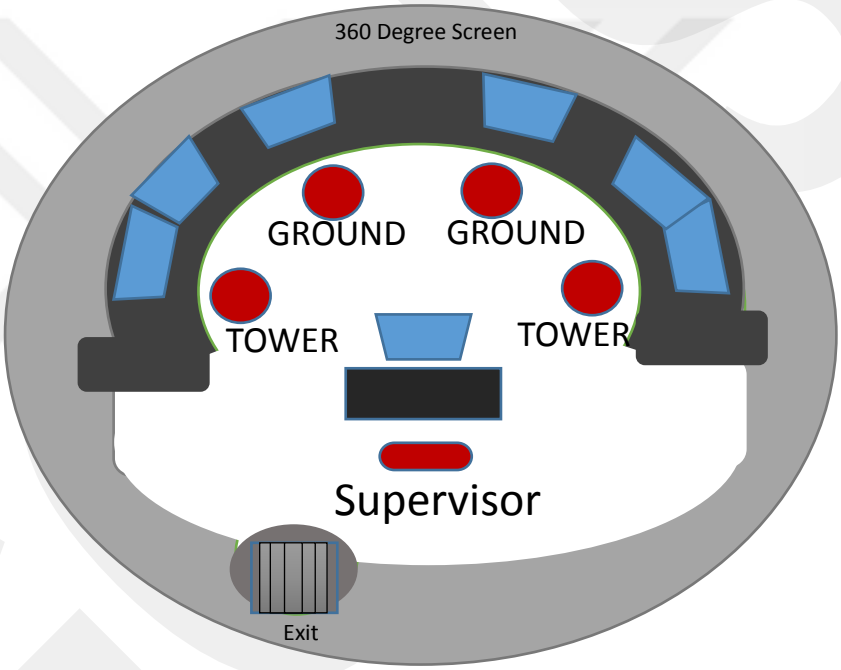

Figure 4.2 Sketch of 3D Aerodrome Simulator and working positions

4.2. Dataset Preparation

The process in this study was tailored according to the general framework given in the Chapter 3. This flow chart is given in the Figure 4.3 below. The data preparation stage is between recording and raw dataset in this figure.

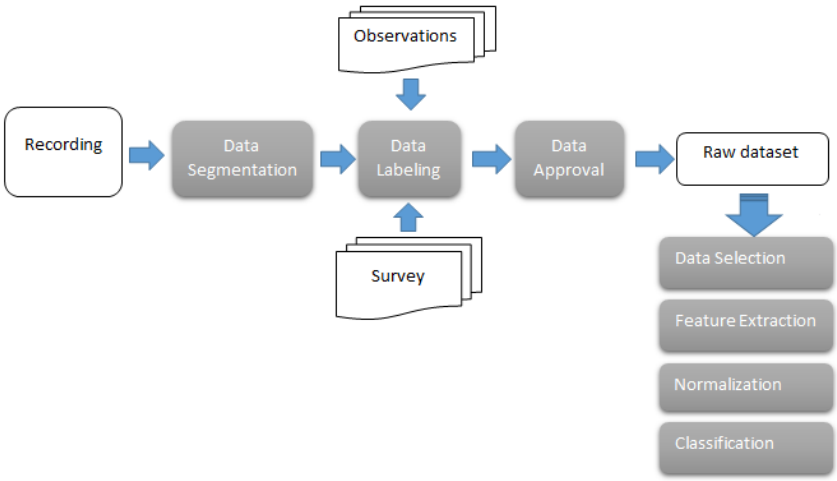

Figure 4.3 Flow chart of the process

In this thesis, a novel dataset is created for Air Traffic Control Environment for Turkish-English users. During the literature review it is seen that, there is a lack of datasets by means of language combination and air traffic control environment. The real Air Traffic Control students who were in their sophomore year have attended this experiment. All have informed and signed the consent form approved by the Atılım University ethical committee.

Among all, 16 of those records' are the subject of this thesis. 4 female and 11 male controllers, age between 20 and 23 have participated at Aerodrome Simulator of Air Control Department at Eskisehir Technical University. Before the final record, there have been some visits in order to understand the environment.

4.2.1. Voice Recording in ATC Environment

Voice samples were gathered in two ways for guarantee the quality of records. In first way, by using a microphone and Audacity 2.3.1 software, controllers' speeches were recorded as close as possible without causing any distraction. In this version, microphone and computer were located in near side of the Tower position, at the left side of the simulator.

Another way is to use one of the two cameras to place one of them at the closest angle to the microphone assembly and see the behaviour of this controller clearly, and the other is to see the teammate, where this operator is most communicated and sometimes comes to speak. In this way, double control of the sound recordings and review of the observations are provided. Considering Figure 4.2, one of these two cameras with sound recording was placed on the left side of the simulator, and the other one was placed in a spotlessly distracted way to see the ground position.

Air Traffic Controller trainees' training sessions are 20 minutes long in the same position. After 20 minutes, students changing their positions to practice different duties. These recordings were conducted without a break. In order not to cause any confusion each student and work positions were labelled. Each students have an ID number that indicates the session, first working position and after the shift.

During the sessions, students were observed, and along with the scenario, their behaviours, stress, voice levels considered. When the emergency situations raised, their reactions were noted and rated. Rates were given by means of low to high as indicated as on the 1-5 scale. After the practice, a 10 words list of single-word vocalization was collected for 16 students. These records have stress free content.

Audacity[48] recordings are conducted in the sampling rate 44100Hz, stereo channel and WAV file format. Camera recordings were not adjustable and have been saved in MOD format. To view and convert the video file with the desired extension, Video Pad Professional Editor V7.32 was used [49]. Trial version of this program was used in a limited time.

4.2.2. Survey Data

Each of the students in the near recording setup, have given set of questions before and after the ATC session. Questionnaires were for self-evaluations.

In the questionnaire given before the session, there are 31 expressions that students can evaluate themselves according to their current situation. In the Figure 4.3 these 31 statuses are given as quick look. Students were evaluated themselves from low to high level, as they are familiar with scale 1-5. This survey contains mostly one-word queries, including moods, stress levels, and factors that could possibly affect stress. The full survey forms are available at APPENDIX A and APPENDIX B.

In order not to negatively affect the working performance of the controllers in the simulation, it has been paid attention to the questionnaire questions to be short and concise. However, a blank is left below the questionnaire in case anyone wants to add additional information.

No	Status	1	2	3	4	5	No	Status	1	2	3	4	5
1	Happiness						17	Concentration					
2	Joy						18	Loneliness					
3	Anger						19	Stress					
4	Sadness						20	Disease					
5	Confused						21	Voice					
6	Fear						22	Hearing					
7	Disdain						23	Seeing					
8	Disgusted						24	Sleepiness					
9	Concern						25	reflexes					
10	Neutral						26	Physical aches, tensions					
11	Grudge						27	Hunger					
12	Frustration						28	Thirst					
13	Anxiety						29	Alcohol consumption					
14	Shame						30	Cigarette consumption					
15	Depression						31	Caffeine consumption					
16	self-confidence						32						

Figure 4.4 Queries before the Session (See APPENDIX A)

The post-test questionnaire where the list of the questions are given in Figure 4.4 also in APPENDIX B, is now a more comfortable assessment form for the student who is free from the stress of the exam. With this questionnaire, the students were asked to evaluate themselves. The questionnaire consists of 17 questions about how the exam passed and also about the stress situation in the exam and the situations that may have caused stress. In the later assessment of these surveys, positive –negative question sentences' transformed and answers are arranged, in order to prevent

confusion. Choices for assessments are like 1-5 scale as indicated as “I do not agree” – “I agree”.

No	Status	1	2	3	4	5
1	The exam passed as I expected					
2	The simulator environment was as I expected					
3	I was not excited in the exam.					
4	I was not stressed on the exam.					
5	I was comfortable in the exam.					
6	I did not have any health problems that negatively affected my speech during the exam.					
7	During the exam, I did not have problems that negatively affect my eyesight.					
8	I did not have any health problems that negatively affected my decision-making during the exam.					
9	During the exam, I did not have problems that negatively affect my hearing.					
10	During the exam, I did not have environmental problems that negatively affected my exam.					
11	My conversations were fluid and straightforward.					
12	The test environment was not disturbing. (Cold, hot, noisy, light, dusty... etc)					
13	I had no problem using the equipment.					
14	I had no problems communicating with the pilots.					
15	I had no problem communicating with my colleague					
16	I used the physiology without any problem. I spoke loud and clear.					
17	My English was enough, I did not have difficulty.					

Figure 4.5 Queries after the Session (See APPENDIX B)

All these survey questions were used to re-evaluate the observed stress while examining the audio and video recordings in the subsequent processes. And all of these questionaries’ are collected and added in the dataset.

4.2.3. Observations

During the air traffic control simulation session, qualitative observations are made and controller behaviours’ were noted by the author of this thesis. Before the session, information was given about the simulation scenario. The scenario in this simulation was 20 minutes long with 10 minute segments. Emergencies were starting to emerge in the second 10-minute segment. Since the author herself is not an air traffic controller, her observations were sometimes verified during the session by

consultants who examined the controllers. Before the final recording, he was present as an observer in other sessions where the same controllers worked.

The observation chart for each session consists of controllers' ID in the session, simulation start time and finish time and emergency indication in the scenario. Each controller's working positions are marked and they observed simultaneously as much as possible .Their stress, loudness of speech and movement are rated from 1 to 5. Additionally, other behavioural observations could be related to stress indication also noted.

These ratings were then re-examined in the data labelling phase, from the video and audio recordings to avoid missing parts that could not be examined.

4.3. Analysis of Collected Data

4.3.1. Data Segmentation

The data recorded from both the microphone and the cameras were evaluated based on the background noise and less noisy recording was chosen. The less noisy recording was a microphone-computer assembly, but in some cases, these recordings were found to have low volume. For this reason, these deficiencies were eliminated with quality sections taken from video recordings.

The data were separated on the basis of utterances in two types. According to the specific Air Traffic radio phraseology and Turkish speeches of the controllers.

Segmentation is conducted with Praat software version 6.0.50. The next feature extraction applications were conducted with this software too. In Figure 4.5 and Figure 4.6. part of this process is available.

At first, segregation was kept lengthy at times of intense communication to make it easier to assess emergencies that might involve stress. In this record, the in-team and radio talks and instructor voices were left as they were.

Later, while examining the sound features with Praat software, sounds longer than 10ms were also torn because it was necessary to divide the speeches into 10 ms parts according to the software's requirement. These separated pieces were also numbered and sound files were not mixed.

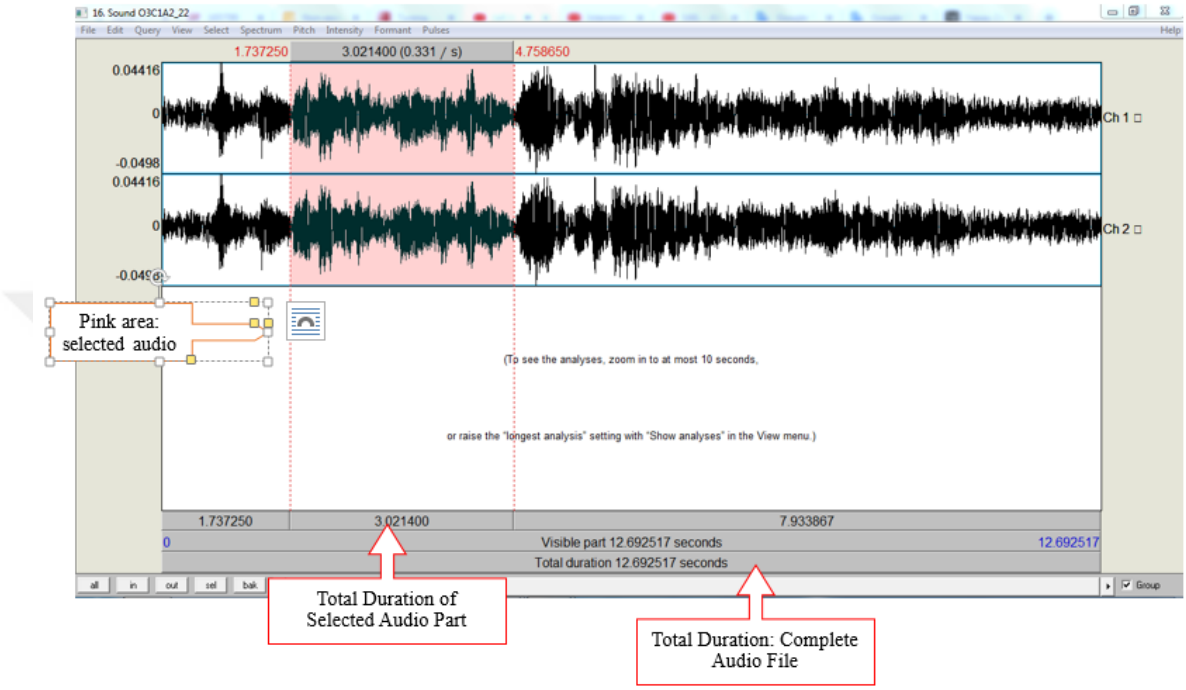

Figure 4.6 Partitioning an audio file in Praat

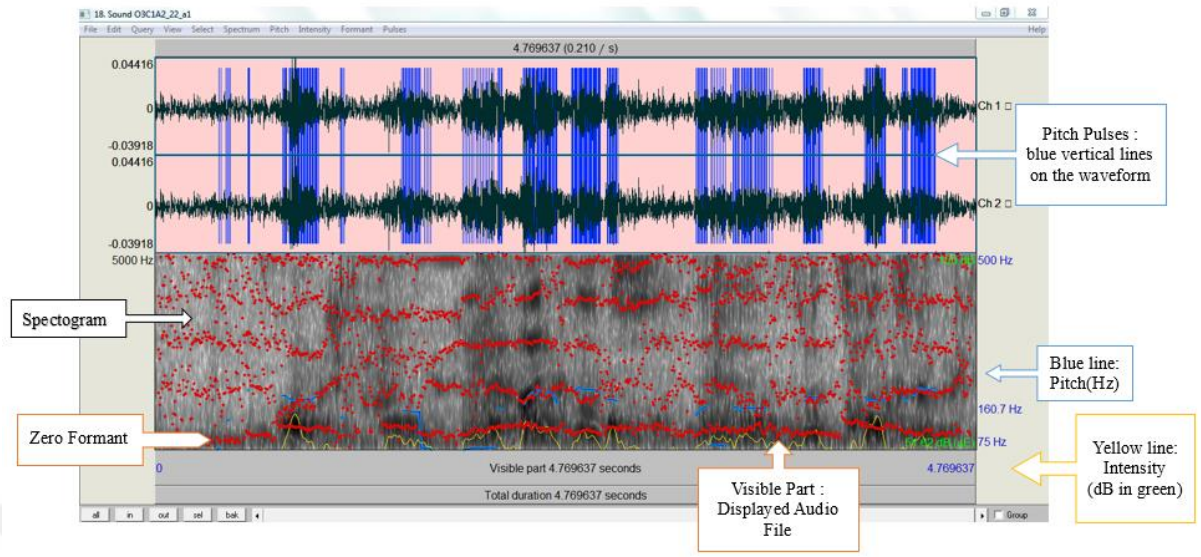

Figure 4.7 Partitioned audio file and voice properties on spectrogam

4.3.2. Data Labelling

Partitioned data has a unique ID number, each manually labelled in an excel file. Each speaker has his identity, gender, utterances, start and end times of his speech. In addition to these, based on the observation notes taken during emergencies and session recordings, the recordings were later re-evaluated over videos, sounds and expressions.

Among these evaluations, the stress, loudness, frequency of movement, intra-team communication within the tower, conversations with pilots and emergency response teams, pause while talking and positive and negative evaluations of the situation are given priority.

Self-assessment questionnaires that were conducted before and after the session were also included in these data and consistencies were observed.

Gender (1 F) (2 M)	ATCO ID	RECORD ID	Emergency	Observed Stress (1-5)	Start Time	Duration	Phraseology + Turkish conversations	Partitioned Records and Details
2	O3C1A2	O3C1A2_22	x 4	5	35.29	12.69	**O3C1A2:** AEROFLOT 032, Are you able to land on runway 36.left, because runway 36 center is closed due to EMERGENCY TRAFFIC	**O3C1A2_22_a1:** O3C1A2: AEROFLOT 032, **O3C1A2_22_a2:** O3C1A2: AEROFLOT 032, Are you able to land on runway 36.left, because runway 36 center is closed due to EMERGENCY TRAFFIC

Figure 4.8 A section of a labeled data4.

4.3.3. Data Validation by an Expert

The labelled data was evaluated by the authority of this area and consultant of this thesis, Dr. Uğur Turhan By looking at the expressions that were transcribed manually from the recorded speeches; a number was given to the emergencies according to the scale of 1 to 5. This value was evaluated as a number indicating the level of the emergency. According to the scale, the number 1 represents the low value and the number 5 the high value.

4.3.4. Speech Processing

In the sound processing stage, Praat software was used for feature extraction. This software is a very useful program that allows getting detailed numerical reports of sounds of 10ms length. The voice menu of the Praat software gives features details in such as pitch, pulses, voicing, jitter, shimmer and harmonicity.

In Table 4.1., the voice report the parameters with the values are provided as an example .Also in the Figure 4.8 and Figure 4.9 the pattern and the spectrogram of the natural speech and a record of emergency speech given respectively in Figure 4.10.

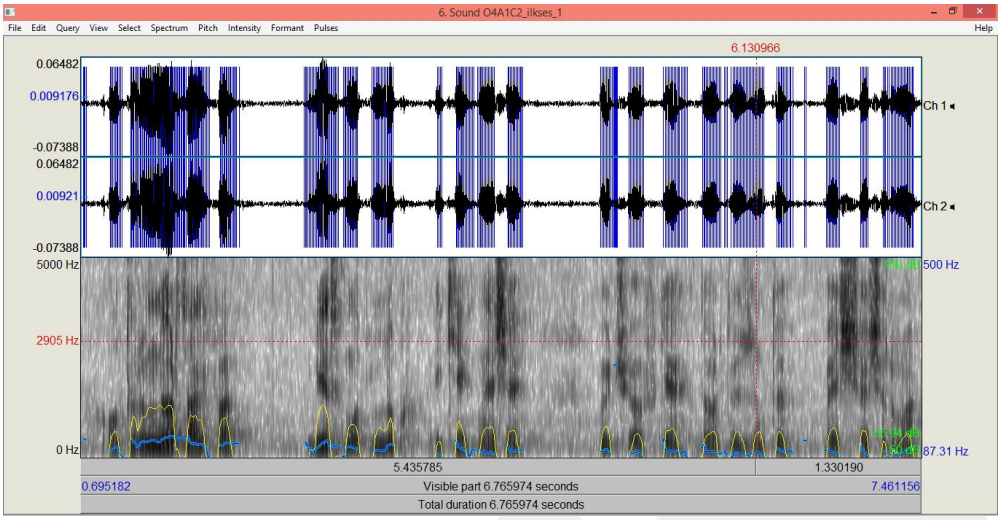

Figure 4.9. Spectrogram view of the neutral speech

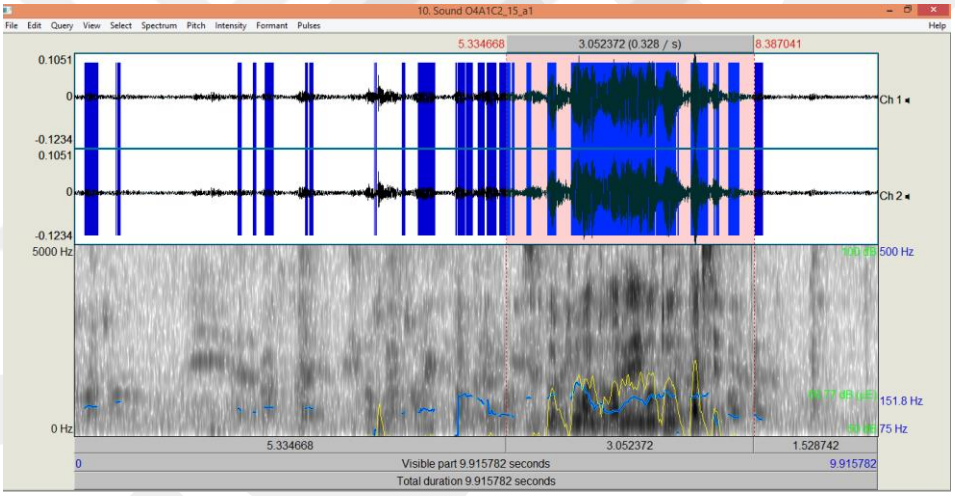

Figure 4.10. Spectrogram view of the speech in an emergency situation

In Praat software, some of the threshold levels are accepted by the other softwares to discriminate the pathologic speech .These features in Table 4.1, can be an easy way to understand the voice report values compared in the Table 4.5. Pathologic level does not in the scope, yet this helps to understand the numerical values.

Table 4.1 Threshold values and nominal range for some features [29] [38][39][50]

Feature	**Threshold**	**Nominal Values**
Jitter(local)	< = 1.040 %	a value between 0 and 2
Jitter (loca, abs)	< = 83.200 μs	-
Jitter (rap)	< = 0.680 %	a value between 0 and 4
Jiter (ddp)	< = 2.04 %	a value between 0 and 6 3 times the Jitter (rap)
Shimmer (local):	< = 3.810 %	-
Shimmer (dB)	< = 0.350 dB	-
Shimer (Apq 3)	-	-
Shimmer (Apq5)	-	-
Shimmer (Apq11)	< = 3.070 %	-
Shimmer(ddp)	-	3 times the value of Shimmer (Apq 3)
HNR:	< 7dB	-

In Table 4.5. there are voice reports of the records taken from the Controller ID of O4A1C2. The neutral speech record which is abbreviated O4A1C2_neutral , had been taken after the air traffic control session finished in order to get true relaxed and neutral tone in the sound. In this example it is in two parts as 6.6 ms and 5.3ms. It consists of , reading of some specific words in the before session survey list.

A part of record taken from the same controllers' in simulation, O4A1C2_15, is taken from the emergency situation, rated as 3 and it is observed stress level is rated as 5. This record is consists of two utterances and it is divided in two parts. Voice report of each part is compared in Table 4.5. The record details are given in Table 4.2

Table 4.2 An example record detail of an emergency situation

Gender F/M	**Controller ID**	**Record ID**	**Emergency**	**Observed STRESS (1-5)**	**Partitioned Records (Phraseology and other speeches in English or Turkish)**
M	O4A1C2	O4A1C2_15	X 3	5	**O4A1C2_15_a1:** Pilot : *We had a bird strike, we lost our two engines..* **O4A1C2 :** SPEEDBIRD 212 !! Are you able to taxi! **O4A1C2_15_a2:** **O4A1C2 :** SPEEDBIRD 212 ! ... If you able to taxi , press button twice!!

The observation chart given in Table 4.3., filled in the air traffic control session and then later re-checked and editing in the data labelling phase, for this controller. In this chart observations made subjectively. These data helped to assign a value observed stress level.

Volume of the speaker is one of the subjective measurements in this observation. Controllers' natural speaking loudness is considered and highest points were given mostly in the emergency and stressful situations for some controllers. In this example, the controllers' voice was loud and got the highest rate.

Coordination one of the important aspects of the duty. In this chart, coordination with pilots and emergency first response team has 3 points. The controller is trying to contact the pilot nevertheless he has not got a response in return and there is not mutual converse. Hesitation of the controller is the perceived speech in accordance with his behaviours.

In team communication meaning is talking to the other controllers in the simulation. Movement indicates if the controller standing or sitting in his workposition and the rate is the movement intensity. In this example, the controller is standing which is abbreviated as *std* and his movements are between moderate and not at all.

Table 4.3 A sample observation notes in the sessions

Observation Notes	Volume (1-5)	Positive (1-5)	Negative (1-5)	Coordination with Pilots and EMFR team (1-5)	Hesitation (1-5)	In team Communication (1-5)	Movement (1-5)
He sounds angry	5	2	4	3	2	1	2 std

The positive and negative observation meaning is the overall interpretation of the behaviours and emotions that the controller was exhibiting. For example, if there was an anger, or anxiety in behaviour or in speech, this adds to the negative section. Similarly, the positive section corresponds to the positive performance such as concentration, confidence and so on.

Similar clustering made with the survey filled by the controllers. In this questionnaire list, there were some emotions and situations that were thought to affect performance negatively or positively. Controllers rated these before the ATC session. The value were given in the range of between 1 and 5. In Table 4.4., the average of these values is calculated and those markings as they considered corresponding to the scope of negative status are found 2, while the positive status scope corresponds are found 2.8. Similarly stress statement ranked by the controller, before the ATC session was 2 and after the ATC session it was 3. These levels are between low and moderate.

Table 4.4 Stress status of the controller before and after the ATC session and. the averages of the surveys

I am stressful (before session)	I was stressful (after session)	Stress, anxiety, anger, fear , sadness, sleepiness, neutral AVERAGE	Joy, happiness, concentration, confidence, neutral AVERAGE
2	3	2	2,8

If we examine the voice reports of both situation, the see there are clear differences in the features appear in the emergency situations. Compared with the neural speech; pitch, shimmer , the total number of pulses and periods increased while voicing parts, the mean period, the standard deviation of the period, and jitter decreased in the emergency speech. Autocorrelation and NHR did not changed significantly, while HNR increased. HNR value is not low than 7dB which is accepted as pathological by Boersma [28]. Due to the noise factor has an increase in emergency speech, decrease in this value is normal. .For example mean pitch value is increasing in the range of male voice range from 96Hz to 153Hz. This is the acceptable range and it is an obvious change

Table 4.5 Comparing voice features of neutral and emergency speeches, PRAAT voice report example

Voice Features	**Voice Report Feature Details**	**Voice report for Sound O4A1C2_neutral_1**	**Voice report for Sound O4A1C2_neutral_2**	**Voice report for Sound O4A1C2_15_a1**	**Voice report for Sound O4A1C2_15_a2**
Duration	**Time range of SELECTION , seconds**	From 0.866150 to 7.461156 seconds (duration: 6.595006 seconds)	From 8.144574 to 13.402650 seconds (duration: 5.258076 seconds)	From 5.573135 to 8.224883 seconds (duration: 2.651748 seconds)	From 0.162134 to 4.753527 seconds (duration: 4.591393 seconds)
Pitch	**Median pitch, Hz**	90.952 Hz	90.451 Hz	160.139 Hz	138.390 Hz
	Mean pitch, Hz	96.352 Hz	92.191 Hz	153.013 Hz	138.418 Hz
	Standard deviation, Hz	19.955 Hz	9.984 Hz	21.040 Hz	21.986 Hz
	Minimum pitch , Hz	74.026 Hz	75.619 Hz	83.414 Hz	86.234 Hz
	Maximum pitch , Hz	271.827 Hz	132.426 Hz	195.673 Hz	288.720 Hz
Pulses	**Number of pulses**	333	240	286	510
	Number of periods	307	215	276	487

	Mean period , seconds	10.461161E-3 seconds	10.918807E-3 seconds	6.527026E-3 seconds	7.237603E-3 seconds
	Standard deviation of period, seconds	1.664882E-3 seconds	1.152137E-3 seconds	0.942549E-3 seconds	0.995963E-3 seconds
Voicing	**Fraction of locally unvoiced frames,%**	38.754% (255 / 658)	43.321% (227 / 524)	18.868% (50 / 265)	13.943% (64 / 459)
	Number of voice breaks	22	22	6	14
	Degree of voice breaks	49.114% (3.239043 seconds / 6.595006 seconds)	52.166% (2.742909 seconds / 5.258076 seconds)	29.190% (0.774057 seconds / 2.651748 seconds)	18.262% (0.838497 seconds / 4.591393 seconds)
Jitter:	**Jitter (local) %**	3,890%	4,034%	2,483%	2,302%
	Jitter (local, absolute) , seconds	406.921E-6 seconds	440.452E-6 seconds	162.057E-6 seconds	166.619E-6 seconds
	Jitter (rap) %	1,892%	1,661%	1,214%	1,170%
	Jitter (ppq5) %	1,987%	1,887%	1,433%	1,421%
	Jitter (ddp) %	5,677%	4,984%	3,642%	3,511%
Shimmer:	**Shimmer (local) %**	14,143%	13,343%	16,697%	16,069%
	Shimmer (local, dB)	1.255 dB	1.137 dB	1.509 dB	1.482 dB
	Shimmer (apq3) %	6,338%	5,089%	7,197%	7,841%
	Shimmer (apq5) %	8,031%	7,466%	10,337%	11,638%
	Shimmer (apq11) %	15,305%	14,171%	19,082%	20,116%
	Shimmer (dda) %	19,015%	15,268%	21,591%	23,524%
Harmonicity of the voiced parts only	**Mean autocorrelation**	0.823671	0.812474	0.810203	0.795236
	Mean noise-to-harmonics ratio	0.272432	0.287328	0.283530	0.308277
	Mean harmonics-to-noise ratio	8.678 dB	8.217 dB	7.444 dB	7.104 dB

CHAPTER 5

RESULTS & DISCUSSION

5.1. Classification Performances

In the Figure 5.1 all data labels are given. There are 455 pieces of voice records in the dataset that are 260 minute long in total. In the experiments, %80 of data is taken as training and %20 of data are taken as test group. Target levels are separated from low, medium and high. Voice records were segmented into pieces shorter than 10ms, long. due to pre-processing needs.

In the left hand side of the Figure 5.1, datasets listed according to the emergency labels. . Emergency labelled data have 5 different level of label. These rates given according to the scenario and validated by an ATC expert. In an emergency there are such situations like mayday calls, bird strikes, engine failures of an aircraft etc. Nevertheless manageability of the situation defines the risk level. X11 means the lowest level and the X55 is the highest level of emergency. Matrix size across this data label name column shows the quantity of the voice record with the selected amount of feature set. In this example, there are 26 features selected. All the data have common 26 features. X11 has total 52 amount of voice record. This notation is given in the matrix form, 26x52. Total amount of emergency data is 305 which is the sum of all emergency levels from 1 to 5.

In the right hand side of the table there is data, labelled with observed stress levels. In this set, the emergency situations are not considered. Observations made by this thesis author, and they are subjective. The stress assessment is rated according to the reaction of the controller. Observed stress levels have the same assessment rate as the emergency situations. The labels are given from 1 to 5. Nevertheless in this dataset stress level 3, stress level 4 and stress level 5 are included. Total amount of labelled

stress is 150 in the scope of this experiment. The rest of the dataset will be available for future tests.

For machine learning applications, each data set is separated for training and testing sets. The nearly %80 of the data is selected for training and the nearly %20 of the data is selected for testing. For example, consider the all emergency situation. 249 amounts of the data of 305 are reserved for training and the rest of the 59 are reserved for testing. All observed stress data here is selected in the same notion. Training data amount is 120 while the testing data remained 30.

ALL EMERGENCY DATA				**ALL OBSERVED STRESS DATA (EMERGENCY SITUATIONS ARE NOT INCLUDED)**			
DATA Label name	**matrix SIZE**	**Training Data**	**Testing Data**	**DATA Label name**	**matrix SIZE**	**Training Data**	**Testing Data**
X11	26 x 52	26x43	26x9	-	-	-	-
X22	26 x 37	26x29	26x8	-	-	-	-
X33	26 x123	26x98	26x25	**STRESS 33**	26x 60	26x48	26x12
X44	26 x 65	26x53	26x12	**STRESS 44**	26 x 54	26x43	26x11
X55	26x 28	26x23	26x5	**STRESS 55**	26 x 36	26x29	26x7
Total Emergency data	**26x 305**	**26x246**	**26x59**	**Total Stress data**	**26 x 150**	**26x120**	**26x30**

Figure 5. 1 Data amounts for 26 features

To increase the reliability of the training and test results, all the emergency labelled data and observed stress labelled data are combined. In Figure 5.2, the data amounts

in details are given. In the machine learning experiments the results were evaluated in three main levels. These are LOW; MEDIUM and HIGH. With the available data, the labels were assumed in the range of these levels. For example HIGH level, consists of emergency labelled data at the level of 4 and 5 and observed stress labelled data at the level of 4 and 5. MEDIUM level is consists of the emergency level 3 and observed stress level 3. Finally, LOW level is consisting of emergency level 1 and 2. The amount of the training and test data, across each of the labels and the sum of these corresponding to the related levels are given in Figure 5.2.

Data Labels	target	Training DATA (26x..)	training data subtotal	TEST DATA (26x..)	test data subtotal
X11	LOW	43	72	**9**	17
X22		29		**8**	
X33	MEDIUM	98	146	**25**	37
stress 33		48		**12**	
X44	HIGH	53	148	**12**	35
X55		23		**5**	
stress 44		43		**11**	
stress 55		29		**7**	

Figure 5. 2 Data details

There are 3 set of test conducted. All data were normalized. In the first test all 26 amount of speech features in the Table 4.1, were used. ANN is applied with different number of neurons. In the Figure 5.2, there are 366 training data and 89 test data were selected.

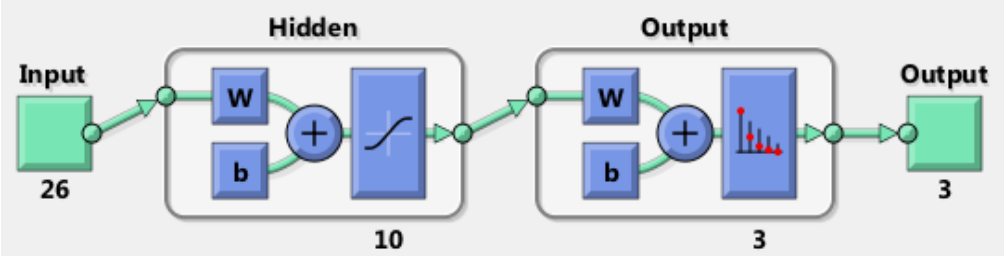

Figure 5. 3 NN for 10 neuron with 26 Features

Table 5.1 Test Results with 26 features for different amount of neurons

Neuron	Training Performance	Test performance
10	93.7%	34.83%
20	100 %	38.20%
50	100 %	22.47%
100	100 %	25.84%
1000	100 %	34.83%

After these results another dataset used with same amounts but different features. In this set only jitter an shimmer features were remained.

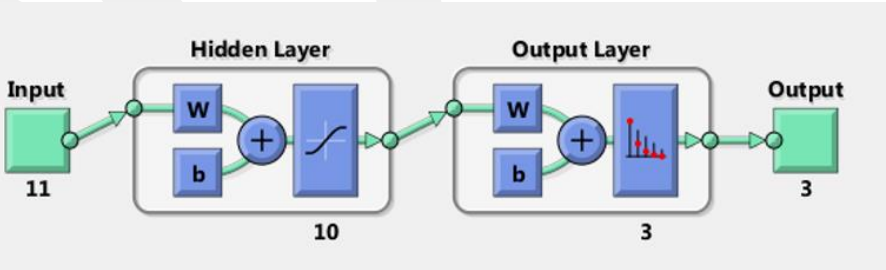

Figure 5.4. NN for 10 neuron with 11 Features

Table 5.2 Test Results with 11 features for different amount of neurons

Neuron	Training Performance	Test performance
10	69.9 %	26.97%
20	85.8 %	26.97%
50	100 %	26.97%
100	100 %	26.97%
1000	100 %	39.33%

The same dataset with 26 features, trained in SVM and the Fine Gaussian SVM training performance was %100. In Figure 5.5, the model predictions are seen. The dots are the correctly predicted data. Dots are color coded. The blue color indicates the level 1, red color indicates level 2 and yellow color indicates level 3.

The normalized data test performance was %29.2.Testing without a normalized data giving the performance as %39.3. Without a validation training performance is achieved at 100%. According to these results with 26 features, normalized data gave best performance with Fine Gaussian SVM classifier. Cross validation decreases the training performance.

Table 5.3 SVM Test Results for data with 26 feature set

Dataset	Validation	SVM type	Training	Test
Not normalized	No validation	Fine Gaussian SVM	100%	29.2%
Not normalized	5 fold	Fine Gaussian SVM	44.8%	29.2%
Normalized	No validation	Fine Gaussian SVM	100%	39.3%
Normalized	5 fold	Fine Gaussian SVM	43.4%	38.2%

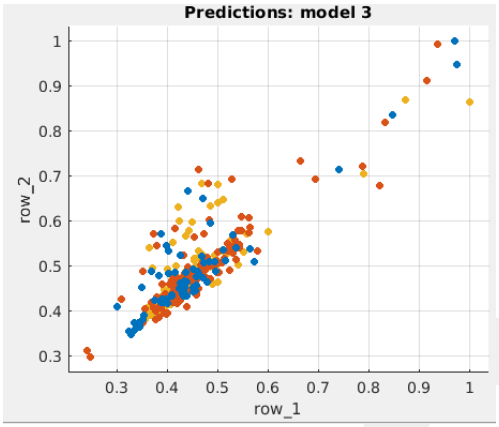

Figure5.5 Training performance for Fine Gaussian SVM

Table 5.4 SVM Test Results for data with 11 feature set

Dataset	Cross Validation	SVM type	Training Performance	Test Performance
Not normalized	5 fold	Quadratic SVM	43,2%	%35,96%
Not normalized	no	Quadratic SVM	68.2%	%42.7%
Not normalized	5 fold	Fine Gaussian SVM	38%	%38,2%
Not normalized	no	Fine Gaussian SVM	92.5%	%39.3%

In the Table 5.4, only jitter and shimmer features are classified. The validation in training made in two options and when no validation is selected, raw data gave better training performances with Fine Gaussian SVM classifier; nevertheless Quadratic SVM got better test performance result. When cross validation applied, Quadratic SVM gave better training performance and Fine Gaussian SVM gave the better test performance.

Similarly normalized data were trained and tested with the same classifiers. In Table 5.5. these results are given. Without validation, Fine Gaussian SVM gave best training and test performances. By using cross validation, both classifiers gave same training performance but , Fine Gaussian gave the best test performance.

If normalized and non-normalized data with 11 feature are compared, it will be seen that there is no clear difference between them. Nevertheless, non-normalized data appear to have better results. Besides, 26 feature training performances are better than 11 feature set, with no validation. Yet, the test results seem to have similar levels.

Table 5.5 SVM Test Results for normalized data with 11 feature set

Dataset	Cross Validation	SVM type	Training Performance	Test Performance
Normalized	5 fold	Quadratic SVM	39.6%	%25,8%
Normalized	no	Quadratic SVM	61.7%	%24.7%
Normalized	5 fold	Fine Gaussian SVM	39.6%	40.4%
Normalized	no	Fine Gaussian SVM	92.3%	40.4%

5.2. Discussions

In this thesis, a data set of a new Turkish-English speaking Air Traffic Controller was created. The data collection process was carried out with limited resources and in a limited time frame, but the data set part was the most time and effort-consuming part of the thesis. In future possible studies, it may be more beneficial to reconsider this process, to increase the number of records over different scenarios and to take

the records from the data recorded by the simulator itself. Of course, since the simulator is actively used for training, it will be necessary to obtain special permissions from authorized places and perhaps work within the simulator system. If the voice recording of each operator is not recorded separately, it will also require a separate study to automate the separation of simulator recordings.

The classification experiments were made with different set of feature sets. In the experiments, classifications are made according to feature vectors. In each feature sets, the training and test sets consist of same amount of controller. However, the controller ID's are different in majority. In Table 5.6. the controllers' distribution are given.

As it is seen, there are two common controllers in both sets. These are male controllers. In training set there is one female controller and in the test set there is two female controllers. In the possible future works, this distribution can be adjusted more balanced along with the enhancements. This may improve success.

Table 5.6 Controller ID distribution in datasets

Training Set	Testing Set
O2C1A2	**O4C1A2**
O3A1C2	O5C1A2
O4A1C2	O6A1C2
O4C1A2	**O6C1A2**
O5A1C2	O7A1B2
O6C1A2	O7B1A2

Since the data labels cannot be considered to be error-free, it may be considered normal that the results are not at the expected levels. It is possible to bring these results to desired levels by conducting more detailed studies on the methods of separation of speech data. All of the survey queries did not apply in order not to

decrease the amount of dataset. The queries were beneficial to decide the stress levels, nevertheless, there needs to be a larger database to apply all the options in those.

In addition, there is another part in the records collected but not included in the thesis and approximately the size of the data set used in the thesis.In the future ,with the addition of this part, classification can give more successful results.

If the results are compared with the emotion classification results given in section 3, it can be seen that the values of SVM and ANN results are below these levels. First of all, the database used in these studies is not the database used in the thesis. And these results will not give a healthy comparison since they differ in terms of the features selected in emotion classification studies. It would be much more meaningful to compare the values in this thesis with stressful sound classification studies made with various data sets. In research, studies made using GMM and HMM and combined hybrid models with ANN. In the future works hybrid classifiers will be used and compared with the single classifiers.

REFERENCES

[1] J.Reason, E. Hollnagel, J Paries , "Revisitiong The Swiss Cheese Model Of Accidents",Eurocontrol Experimental Center , France, EEC Note No13/06, October 2006

[2] S.T Shorrock, B.Kirwan, "Development and application of a human error identification tool for air trafficcontrol." Appl. Ergon. 2002, 33, 319–336

[3] D.A Wiegmann, S.A. Shappell, "A Human Error Approach to Aviation Accident Analysis: The Human Factors Analysis and Classification System", Routledge: Abingdon-on-Thames, UK, 2017.

[4] T .Lyu, W. Song; K.Du, "Human Factors Analysis of Air Traffic Safety Based on HFACS-BN Model." Appl. Sci. 2019, 9, 5049.

[5] John McCreary , Pollard, Michael; Stevenson, Kenneth; Wilson, Marc B.; "Human Factors: Tenerife Revisited", Journal of Air Transportation World Wide ,Vol3 No1,pp. 2-3, 1998

[6] Gerard M. Bruggink, Remembering Tenerife, Internet: https:// web.archive.org/web/20060513193139/http://cf.alpa.org/internet/alp/2000/aug00p18.htm,,August 2000,[Sept.24,2020]

[7] Aviation Safety Network (ASN) 1977 database Tenerife, Internet: https://aviation-safety.net/database/record.php?id=19770327-0, [Sept.24,2020]

[8] Wikipedia contributors. "2002 Überlingen mid-air collision.", Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 29 Jul. 2020. Web. 24 Sep. 2020.

[9] G.Borghini, G. Di Flumeri, P. Aricò, et al. "A multimodal and signals fusion approach for assessing the impact of stressful events on Air Traffic Controllers." Sci Rep 10, 8600 (2020).

[10] Ling He , Stress and Emotion Recognition in Natural Speech in the Work and Family Environments, RMIT University , PhD Thesis, Nov.2010

[11] Uğur Turhan, "Hava Trafik Kontrolorü Adaylarının Seçimi ve Türkiye'deki Uygulama", Anadolu University, PhD Thesis, Sept.2007

[12] SESAR Joint Undertaking, "The Roadmap For Delivering High Performing Aviation For Europe, European Master Plan", European Union and Eurocontrol, Edition 2015

[13] F. Jones, J. (Eds.) Bright, , “Stress: Myth, theory and research.”, Prentice Hall/Pearson Education ,2001, p 3-45

[14] T .Newton,(1995).”Managing Stress: Emotion and Power at Work”. London : Sage.

[15] H.Selye, (1993). “History of the stress concept.” , In L. Goldberger and S.Breznitz(Eds), Handbook of Stress : Theoretical and Clinical Aspects (2nd Edition) .New York: The Free Press

[16] L.E. Hinkle, (1973). “The concept of stress in the biological and social sciences.”,Science, Medicine and Management, 1, 31-48

[17] J.E. McGrath, (1976), “Stress and behaviour in organizations. In M.Dunentte(Ed),Handbook of Industrial and Organizational Physchology.Chicago: Rand Mc Nally

[18] R.S .Lazar,S.Folkman,(1984), “Stress Appraisal and Coping”.New York:Springer.

[19] T.Cox, (1993). “Stress Research and Stress Management: Putting Theory to Work. (Health and Safety Executive Contract Research Report No 61/1993)

[20] Bo Zang, “Stres Recognition from Heterogenous Data”. Human Computer Interaction [cs HC].Université de Lorraine, 2017. English.

[21] H.Selye.” The Stress of life”. New York: MacGrawHill, 1956.

[22] Yao Xiao, “Classificaition of Speech Under Stress Based on Physical Characterisitics of Vocal Folds Vibration”, Dissertation, Nagoya University, August, 2013

[23] R.Plutchik, (1980). “A general psychoevolutionary theory of emotion”. In R. Plutchik, & H. Kellerman (Eds.), Emotion: Theory, Research, and Experience (pp. 3-33). New York: Academic Press.

[24] M. Hagmüller,G. Kubin,E Rank. “Evaluation of the Human Voice for Indications of Workload Induced Stress in the Aviation Environment.” ., 2005.

[25] Lawrence R. Rabiner and Ronald W. Schafer (2007), "Introduction to Digital Speech Processing", Foundations and Trends® in Signal Processing: Vol. 1: No. 1–2, pp 1-194. http://dx.doi.org/10.1561/2000000001

[26] Christos-Nikolaos Anagnostopoulos, Theodoros Iliou, and Ioannis Giannoukos. 2015.” Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011.” Artif. Intell. Rev. 43, 2 (February 2015), 155–177. DOI:https://doi.org/10.1007/s10462-012-9368-5

[27] K. Tomba, J.Dumoulin ,E. Mugellini, O.Abou Khaled, S.Hawila , (2018). “Stress Detection Through Speech Analysis”, In Proceedings of the 15th International Joint Conference on e-Business and Telecommunications - Volume 1: ICETE, ISBN 978-989-758-319-3, pages 394-398.

[28] Paul Boersma , “Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound”, IFA Proceedings 17: 97-110.,1993

[29] Paul Boersma, David Weenink:“Praat: doing phonetics by computer” [Computer program],Version 6.0.37, retrieved 14 March 2018 from, http://www.praat.org/

[30] J.Hansen, C.Swail,A.South, R.Moore, H. Steeneken, E. Cupples, T. Anderson, C.Vloeberghs, I.Trancoso, and P. Verlinde. (2000). “The impact of speech under ‘stress’ on military speech technology.” Technical Report AC/323(IST)TP/5 IST/TG-01, NATO Research & Technology Organization RTO-TR-10

[31] İ.Baran Uslu, EE519, "Speech Processing and Applications", 2023, Faculty of Electrical and Electronics Engineering, Atılım University, Ankara, Sept.24.2018

[32] G. Williamson ,Acoustic Measures (Norms) , https://www.sltinfo.com/acoustic-measures-norms/, Feb. 1 ,2014 [Agust 17, 2020]

[33] L. Rothkrantz,J.W.Wees, R.Vark, (2004). “Voice Stress Analysis”. Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science). 3206. 449-456. 10.1007/978-3-540-30120-2_57.

[34] H. Traunmüller, (2000). “Evidence for demodulation in speech perception”. In Proceedings of the 6th ICSLP, volume 3, pages 790–793, Beijing, China.

[35] S.R. Bandela . “Stressed speech emotion recognition using feature fusion of teager energy operator and MFCC"; 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT):1-5 (2017)

[36] G. Williamson ,Pitch , https://www.sltinfo.com/acoustic-measures-norms/, Jan. 31 ,2014 [Agust 17, 2020]

[37] G.Williamson, (2006) Human communication: a linguistic introduction (2 ed.) Billingham: Speech-Language Services.

[38] J.Teixeira, C. Oliveira, C.Lopes, (2013). “Vocal Acoustic Analysis – Jitter, Shimmer and HNR Parameters.” Procedia Technology. 9. 1112-1122. 10.1016/j.protcy.2013.12.124.

[39] Mireia Farrús, Javier Hernando, Pascual Ejarque , "Jitter and Shimmer Measurements for Speaker Recognition",INTERSPEECH 2007, 8th Annual Conference of the International Speech Communication Association, Antwerp, Belgium, August 27-31, 2007

[40] Dr. Metin Bilgin, “Makine Öğrenmesi Teorisi ve Algoritmaları”, Papatya Yayıncılık, 2018

[41] J. H Hansen, S. E Bou-Ghazale, R. Sarikaya ,B. Pellom (1998). “Getting started with the SUSAS: Speech under simulated and actual stress database.” Technical Report RSPL-98-10, Robust Speech Processing Laboratory, Duke University

[42] A.Aguiar, M.Kaiseler,M. Cunha,H. Meinedo, P.R. Almeida, J. Silva, “ VOCE Corpus: Ecologically Collected Speech Annotated with Physiological and Psychological Stress Assessments.” In Proceedings of the LREC 2014: 9th International Conference on Language Resources and Evaluation, Reykjavik, Iceland, 26–31 May 2014

[43] A. Ikeno, V. Varadarajan, S. Patil and J. H. L. Hansen, "UT-Scope: Speech under Lombard Effect and Cognitive Stress," 2007 IEEE Aerospace Conference, Big Sky, MT, 2007, pp. 1-7, doi: 10.1109/AERO.2007.352975.

[44] F. Burkhardt, A. Paeschke, M. Rolfes, W.F. Sendlmeier, B. Weiss , “A database of german emotional speech” ,Ninth European Conference on Speech Communication and Technology , 2005

[45] Estelle Delpech, Marion Laignelet, Christophe Pimm, Céline Raynal, Michal Trzos, “A Real-life, French-accented Corpus of Air Traffic Control Communications”.Language Resources and Evaluation Conference (LREC), May 2018, Miyazaki, Japan

[46] K.Hofbauer, S.Petrik, “ATCOSIM Air Traffic Control Simulation Speech Corpus.” 2007.

[47] R.P. Gadhe, R.A. Shaikh Nilofer, V.B. Waghmare, P.P. Shrishrimal, “Emotion Recognition from Speech: A Survey”,International Journal of Scientific & Engineering Research 6 (4), 632-635

[48]Audacity Software 2.3.1 , internet: https://www.audacityteam.org/download/windows/ , 8 March 2019,

[49]Video Pad Professional Editor V7.32, Trial version, internet: https://www.nchsoftware.com/videopad/index.html, June 2019

[50] G. Williamson ,Acoustic Measures (Norms) , https://www.sltinfo.com/acoustic-measures-norms/, Feb. 1 ,2014 [Agust 17, 2020]

APPENDIX A

BEFORE SESSION SURVEY

Controller ID	Age	Gender (F/M)	Date
			……../ 05 /2019

Evaluate the following feelings and situations according to your current situation
NO :1 ← → YES: 5

No	Status	1	2	3	4	5	No	Status	1	2	3	4	5
1	Happiness						17	Concentration					
2	Joy						18	Loneliness					
3	Anger						19	Stress					
4	Sadness						20	Disease					
5	Confused						21	Voice					
6	Fear						22	Hearing					
7	Disdain						23	Seeing					
8	Disgusted						24	Sleepiness					
9	Concern						25	reflexes					
10	Neutral						26	Physical aches, tensions					
11	Grudge						27	Hunger					
12	Frustration						28	Thirst					
13	Anxiety						29	Alcohol consumption					
14	Shame						30	Cigarette consumption					
15	Depression						31	Caffeine consumption					
16	Self-confidence						32						

In this box, you can write your thoughts you want to add.

APPENDIX B

AFTER SESSION SURVEY

Controller ID	Age	Gender (F/M)	Date
			……../ 05 /2019

Evaluate the following situations to suit you best. **No :1 ← → Yes:5**

No	Status	1	2	3	4	5
1	The exam passed as I expected					
2	The simulator environment was as I expected					
3	I was not excited in the exam.					
4	I was not stressed on the exam.					
5	I was comfortable in the exam.					
6	I did not have any health problems that negatively affected my speech during the exam.					
7	During the exam, I did not have problems that negatively affect my eyesight.					
8	I did not have any health problems that negatively affected my decision-making during the exam.					
9	During the exam, I did not have problems that negatively affect my hearing.					
10	During the exam, I did not have environmental problems that negatively affected my exam.					
11	My conversations were fluid and straightforward.					
12	The test environment was not disturbing. (Cold, hot, noisy, light, dusty… etc)					
13	I had no problem using the equipment.					
14	I had no problems communicating with the pilots.					
15	I had no problem communicating with my colleague					
16	I used the physiology without any problem. I spoke loud and clear.					
17	My English was enough, I did not have difficulty.					

In this box, you can write your thoughts you want to add

APPENDIX C

PARTICIPANT CONSENT FORM WITH SIGNATURE

Title of Research: Speech Analyzing using Machine Learning Techniques
Graduate Student: Evrim Yılmaz
Supervisors: Dr. Ibrahim Baran Uslu, Dr. Uğur Turhan

You are being asked by Evrim Yılmaz to participate in her graduate research project continuing under Electrical and Electronics Department of Atilim University. For you to be able to decide whether you want to participate in this project, you should understand what the project is about, as well as the possible risks and benefits in order to make an informed decision. This process is known as informed consent. This form describes the purpose, procedures, possible benefits, and risks of the research project. It also explains how your personal information/bio specimens will be used and protected. Once you have read this form and your questions about the study are answered, you will be asked to sign it. This will allow your participation in this study.

Summary of Study

You are asked to participate in a study analyzing human voice in Air Traffic Control environment. You have been asked to take part in this study because you are an Air Traffic Controller candidate selected through various tests. You are a healthy young candidate with no history of neurologic or psychiatric disturbance, a fluent English speaker, have normal hearing and vision, and are physically capable of key press responses. Before agreeing to participate in this study, it is important that you read and understand the following explanations, so you can make an informed decision about taking part in this study.

Explanation of Study

This study is designed to analyze the human voice in Air Traffic Control environment. If you agree to participate, your voice will be recorded and stored. Your participation in the study will last during the simulation practices.

Risks and Discomforts

No risks or discomforts are anticipated.

Benefits

The benefit associated with participating in this study is the satisfaction to contribute to an original work in the area of Air Traffic Control and Electrical and Electronics Engineering. Your data will be used in scientific papers.

Confidentiality and Records

Data collected will remain strictly confidential. All data will be used for research purposes and to write a scientific paper about the speech processing applications in Air Traffic Control environment. Only people who are associated with the study will see your responses. Also, responses will not be associated with your name; instead, your name will be converted to a code number when the researchers store the data. No names or identifying information will be used in any publication or presentation.

Future Use Statement

Identifiers might be removed from data/samples collected, and after such removal, the data/samples may be used for future research studies or distributed to another researchers for future research studies without additional informed consent from you or your legally authorized representative.

Contact Information

If you have any questions regarding this study, please contact the graduate student ***[Evrim Yılmaz,*** yilmaz.evrim@student.atilim.edu.tr ***, phone]*** or the supervisors ***[Dr.Uğur Turhan ,*** uturhan@eskisehir.edu.tr ***, +90 (222) 335 0580 / 6848 ; Dr.Ibrahim Baran Uslu ,*** baran.uslu@atilim.edu.tr ***, +90 (312) 586 8253].***

By signing below, you are agreeing that:

- you have read this consent form (or it has been read to you) and have been given the opportunity to ask questions and have them answered;
- you have been informed of potential risks and they have been explained to your satisfaction;
- you understand Atilim University has no funds set aside for any injuries you might receive as a result of participating in this study;
- you are 18 years of age or older;
- your participation in this research is completely voluntary;
- you may leave the study at any time; if you decide to stop participating in the study, there will be no penalty to you and you will not lose any benefits to which you are otherwise entitled.

Signature_____Date_____

Name /Surname_____

APPENDIX D

ATILIM ÜNİVERSİTESİ İNSAN ARAŞTIRMALARI ETİK KURULU

DEĞERLENDİRME FORMU

Atılım University Human Research Ethics Board
Evaluation Form

PROJE-ARAŞTIRMA BAŞLIĞI /*Project- Research Title*:

SORUMLU ARAŞTIRMACI/*Responsible Researcher*:

DEĞERLENDİRME TARİHİ/ *Date of Evaluation*:

Başvuru Evrak Kontrol Listesi /*Application File Checklist*:

☐ Etik Kurul Başvuru Formu /*Ethics Board Applicaion Form*

☐ Gönüllü Katılım Formu / *Informed Consent Form*

☐ Katılım Sonrası Bilgilendirme Formu (varsa)/ *Debriefing Form (if applicable)*

☐ Veli Onay Formu (varsa) / *Parental Approval Form (if applicable)*

☐ Kullanılacak yazılı veri toplama araçlarının (anket soru formu, ölçek, test vb.) birer örneği / A *Copy of Each Data Collection Tool (eg. tests, questionaire, test etc.)*

1.	☐ **Kabul / *Accepted***
2.	☐ **Düzeltme gereklidir / *Correction Requested*** **Açıklama/*Explanation*:**
3.	☐ **Red / *Rejected*** **Gerekçe / *Rationale*:**

Atılım Üniversitesi İnsan Araştırmaları Etik Kurul Üyesi
Member of Atılım University Human Research Ethics Board
Ad- Soyad /*Name- Last name*

İmza/ *Signiture*