EXTRACTING EMOTIONS FROM AUDIO SIGNALS FOR EFFECTIVE REQUIREMENT ENGINEERING

by

Sara Azeem

Supervised By

Dr. Sajid Saleem

Co-Supervised By

Dr. Basit Shehzad

Submitted for partial fulfillment of the requirements of the degree of MSCS to the Faculty of Engineering and Computer Science

NATIONAL UNIVERSITY OF MODERN LANGUAGES, ISLAMABAD

JUNE 2020

EXTRACTING EMOTIONS FROM AUDIO SIGNALS FOR EFFECTIVE REQUIREMENT ENGINEERING

by

Sara Azeem

Supervised By

Dr. Sajid Saleem

Co-Supervised By

Dr. Basit Shehzad

Submitted for partial fulfillment of the requirements of the degree of MSCS to the Faculty of Engineering and Computer Science

NATIONAL UNIVERSITY OF MODERN LANGUAGES, ISLAMABAD

JUNE 2020

NATIONAL UNIVERSITY OF MODERN LANGUAGES

FACULTY OF ENGINEERING AND COMPUTER SCIENCE

THESIS AND DEFENSE APPROVAL FORM

The undersigned certify that they have read the following thesis, examined the defense, are satisfied with overall exam performance, and recommend the thesis to the Faculty of Engineering and Computer Sciences.

THESIS TITLE: EXTRACTING EMOTIONS FROM AUDIO SIGNALS FOR EFFECTIVE REQUIREMENT ENGINEERING

Submitted By: Sara Azeem

Registration #: 14/MS/MCS/F16

Master of Science
Degree Name in Full

Computer Science
Name of Discipline

Dr. Sajid Saleem
Name of Research Supervisor

Signature: ______________________

Dr. Basit Shehzad
Name of Co-Supervisor

Signature: ______________________

Col. (R) Jamil Ahmed Zia
Name of Dean (FE&CS)

Signature: ______________________

Brig. Muhammad Ibrahim
Name of Director General (NUML)

Signature: ______________________

11th June, 2020
(Date)

CANDIDATE DECLARATION

I declare that this thesis entitled "*EXTRACTING EMOTIONS FROM AUDIO SIGNALS FOR EFFECTIVE REQUIREMENT ENGINEERING*" is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature	:	
Name	:	Sara Azeem
Date	:	June 11th, 2020

ABSTRACT

Requirements engineering (RE) is a fundamental process in software development project. From the social point of understanding, emotions play a significant role in behavior and also act as developmental motivators. Thus, if RE is considered as a presentation on a set of knowledge-intensive everyday tasks, which include specification, acceptance, rejection and negotiation activities, then the emotional influence, characterizes as a fundamental element in RE. However, the emotional dynamics in RE has not received the consideration it deserves. This research provide collective devotion towards the learning of the emotional content of communication during requirement gathering, and hence, a concrete method has been projected to recognize the emotions within a spoken statement. As speaker emotion recognition is accomplished through processing means that contain isolation of the speech signal, extraction of features, formation of databases and suitable classifiers. Hence this research is an analysis on speaker emotions classification method, addressing central aspects of design of a speaker emotion recognition system. To accomplish this research, a speech emotion recognition (SER) system, which is based on classifier and various techniques for features extraction, is constructed. Mel Frequency Cepstrum Coefficients (MFCC) features are extracted from the audio signals, it was applied in order to catch the most appropriate feature category. Machine learning pattern was used for the emotion classification task. GMM-UBM classifier is used to classify seven emotions. Various databases such as TESS, RAVDEES and SERAD are used for experimental resolution. This research indicates that for TESS database, proposed method achieves an accuracy of 99.63%, for RAVDESS, the accuracy is about 84.37% and for SREAD, the accuracy is 94.05%.

Keywords: Requirement engineering, emotion recognition, feature extraction, classification.

DEDICATION

This thesis work is dedicated to my parents and my teachers throughout my education career who have not only loved me unconditionally but whose good examples have taught me to work hard for the things that I aspire to achieve.

ACKNOWLEGEMENT

First of all, I wish to express my gratitude and deep appreciation to Almighty Allah, who made this study possible and successful. This study would not be accomplished unless the honest espousal that was extended from several sources for which I would like to express my sincere thankfulness and gratitude. Yet, there were significant contributors for my attained success and I cannot forget their input, especially my research supervisors, Dr. Sajid Saleem and Dr. Baist , and my teacher Dr. Fazli Subhan, who did not leave any stone unturned to guide me during my research journey.

I would like to say a special thanks to my parents and family for their support throughout my educational career and for their loving forbearance during the long period it has taken me to conduct the research and write up this thesis, specially my father Maj. Muhammad Azeem. I would like to thank my colleague Mr. Mohsin Abbas who collaborated with me on research and provided much needed support. Thanks also to all the people who participated in my research and provided valuable feedback.

I shall also acknowledge the extended assistance from the administrations of Department of Computer Sciences who supported me all through my research experience and simplified the challenges I faced.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

LPC	Linear Predictive Coding
LPCC	Linear Predictive Cepstral Coefficient
MFCC	Mel-frequency Cepstral Coefficient
GMM	Gaussian Mixture Model
GMM-UBM	Gaussian Mixture Model-Universal Background Model
DCT	Discrete Cosine Transform
EER	Equal Error Rate
EM	Expectation Maximization
RE	Requirement Engineering
TESS	Toronto Emotional Speech Set
RAVDEES	Ryesion Audio-Visual Database of Emotional Speech
SREAD	Software Requirement Engineering Audio Dataset

CHAPTER 1

INTRODUCTION

1.1 Overview

Requirement explains what the customer, user, developer, businessman and supplier need and also how system should work in order to satisfy that need. Requirement are considered as foundation for every project because it first explains the scope of project and then add all succeeding development details and data in it. Without a respective strong requirement base the project development and growth can only be a thrash. Project development without Requirement (Engineering) is like setting out for a journey without having an idea about destination and without having navigation system/chart, therefore process of (Requirement Engineering)/ Requirement gathering provides both the ”navigation chart” and guide towards the desired destination.

Requirement provide foundation for arranging data, planning for developing project, acquiring on completion and accepting it on execution. On the other hand some stakeholder’s or customers accept the project and system without any potential risk management blueprint, this may lead them towards difficulty. Even when prospective answers are established one should always have the risk of failing in order to get a moderate and acceptable solutions, therefore it is considered that Requirements enables the risk to be managed and solved in earliest possible stage in development.

To well understand by everyone, requirements are conveyed in natural language and at this place challenge and provocation lies. As, needs of the customer may be varied frequently and may there lie a conflict. All these needs may not be explained clearly or may be inspired, infused and change by other design which themselves have direct effect on software development. Hence Requirement forms the foundation and basis for:

i. Project Planning

ii. Arrange Schedule/data

iii. Risk Management

iv. Change control

v. Execution testing

vi. Completion testing

vii. Acceptance testing

Generally the most usual cause for project success and failure are said to be not technical. A Standish Group is an individualistic global IT research consulting firm, established in 1985 which is known for their report on the information about project implementation in private and public sector. This firm focus on success rate, failures and potential improvements in IT project by publishing their report on project failure and success since 1994. Table 1.1 reveal a 1994 chart that identify the dominant reason why systems/project fails.

Table 1.1: Reason for project failure

	Description	**Percentage**
∗∗	Insufficient requirements	13.3%
∗∗	Absence of user participation	12.3%
	Deficiency of resources	10.7%
∗	Realistic expectation	9.8%
	Absences of administrative support	9.4%
∗∗	Uncertain requirement	8.8%
	Lack of arrangement	8.2%
∗∗	Didn't need requirement any longer	7.5%

Those having ∗∗ are directly associated with requirements and those with ∗ have a limited association. In 2015, Standish Group studied fifty thousand softwaree projects worldwide, starting from minute amplification to huge project enhancement along with System re-configration. Table 1.2 shows the results of the software projects over last few years.

Table 1.2: The Standish Group Report (2015)

	2011	2012	2013	2014	2015
Successful	28%	28%	32%	27%	28%
Modified	48%	55%	50%	54%	53%
Failed	22%	17%	18%	19%	18%

A key point Standish Group analysis over the 20 years is that, Requirement engineering and Emotional Maturity makes the project more successful. Table 1.3 demonstrate the factors which makes project more successful [1].

Table 1.3: Reason for project success (Standish Group 2014)

	Description	Percentage
**	End user involvement	15.8%
**	Clear declaration of Requirement	13.1%
	Correct planning	9.5%
*	Real expectation	8.3%
	Clear vision and objectives	2.9%

1.2 Introduction to Requirement Engineering

Software requirements means ”to figure out what to build” which prompt the essential requirements and limitation which are employed on software product that contributed towards the solution of some realistic world problem [2, 3]. Requirement Engineering is considered as heart of any federation’s capability to mentor the craft and keep tempo with the increasing tide of convolution and complexity. In general, requirement engineering is the subdivision of software engineering, as dealing with the needs/requirements is an imperative and crucial chunk of all engineering trails. The term requirement engineering involves all the activities related to requirement such as understanding, analyzing, formulating, managing and developing requirements. One of the most well-established and broader definition, which is derived from DoD (software strategy document) in 1991 is;

> ”Requirement engineering: includes all life-cycle activities which are said to be constant towards th identification of customs needs, requirements investigation to manage additional modification, documentation preresquisite as a specification, and document validation against user

requirements, as well as a method which support all these activities" (DoD 1991)

The indicated definition explains all the key activities which are directly related to requirement engineering. In simple words, requirement engineering is the subdivision of software engineering process which deals with identifying , blooming, deleting, developing, discovering, analyzing, conveying , managing and organizing requirements of software system [4]. It also reflects the connection between indicated factors to pinpoint the general specifications of a software [5]. It will be the worst nightmare when a stakeholder looks in the eye of software developer and say,

"I know you believe that you recognize what I told you but what you don't understand, is what I say is not what I mean."

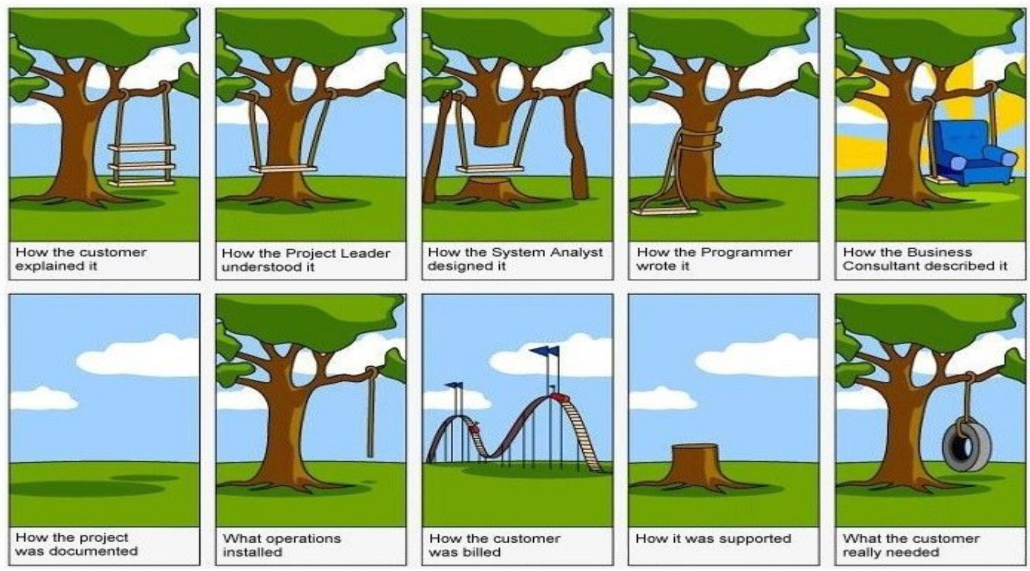

Figure 1.1: Example of Project Failure

To avoid this problem requirement engineering is the best solution, as this process clearly identify the needs of stockholder. This process includes same activities which are enumerated by many researchers. Requirement engineering activities are catalogued in [4, 6] as, (i) Inception, (ii) Elicitation, (iii) Elaboration, (iv) Negotiation, (v) Specification/Documentation, (vi) Validation, (vi) Management

1.2.1 Inception

Inception is the opening stage of requirement engineering process which focuses about problem, stockholder and desired solution. The introductory collaboration and communication between software developer and stockholder is also planned in this stage. At project inception, depth and breadth of the project is identified by creating a rough feasibility inspection or analysis, by defining an absolute answer for one of the most basic question that ”How does a project get started?”

1.2.2 Elicitation

The elicitation in the Requirement Engineering process focus on the objectives of the product that how product fix according to the needs of stockholder, how product will be used frequently, how product match the standard needs of the business and how it will be accomplished, all this is collected simply by asking the accurate questions [4]. It seems simple but in realistic way it is very hard because during elicitation several customers dispense unrelated requirements which may cause conflict for the implementation, therefore it is compulsory to have a phase to examine, analyze and frame the requirement that are obvious for implementation. During elicitation stage, problems that confront and encountered are:

i. Problem of Scope

ii. Problem of understanding

iii. Problem of volatility (changing of requirements over time)

1.2.3 Elaboration

During elaboration process all the data/ information which is obtained during elicitation and inception activity, from the stakeholder is expanded, polished and refined. This process handle all the refinements of user’s storyline which makes it much clear about how end user will interact with the project. Elaborating the requirements is necessary for establishing and developing a firm foundation for project design, because a supplemental view of the project is produced, by identifying the relationship between requirements.

1.2.4 Negotiation

Conflicts in project requirement is reconcile through the process of negotiation. In general, users, stakeholder and customer are asked to rank and designate the requirements which are obvious for development. This help the stakeholder to evaluate cost, risk, depth and breadth of the system which automatically gives some measure of satisfaction to each party. Negotiation is important for the delivery and transmission of all information from stockholder to project developer.

1.2.5 Specification

All the information obtained during different activities of requirement engineering are documented, for developing project in an ideal manner. Requirement specification is defined in [7] as;

> ”A requirement specification is a documentation which includes systematical representation of gathered requirements, typically for a software project or component, and this documentation specifies given criteria in a standard form”

Requirements has direct and indirect influence on project success and failure, therefore they should be documented in order to overcome the conflict and interdependencies. According to the project size, specifications can be a written document, a prototype, a graphic model, a set of usage framework or scenarios for example, in case of large projects written document, graphic model and natural language illustration are considered to be a best approach.

1.2.6 Validation

Specifications are again examined in validation process, to ensure that all the requirements of a project are stated. Validation detect and then correct the errors, omissions, conflicts and inconsistencies of requirements, which may occur during documentation. It also provide a guideline for the standard establishment of project. Requirements are validated by a team which includes customers, stockholders, end-users and software engineers. All these parties review the requirements documentation in order to find any area where collection and clarification is required. This activity provide a guarantee that all predefine requirements are documented and meet the quality criteria of the project.

1.2.7 Management

In computer-based system, change in the requirements linger throughout the life of a project, hence requirement management activity help to maintain consistency of the project after changes. It reviews the implementation of the requirements after change by helping the project team in tracking the needs of the customers at any time during project development. In general, number of graphical models are built which depict the basic informational, functional and statistical outlook of a system.

1.3 Emotional Requirements Engineering

The importance of human emotions cannot be underestimated in fields such as medicine, psychology, neuroscience, and in some areas of computer science like human computer interaction (HCI), affective computing and software engineering (requirement engineering). One should know the importance of emotions in requirement engineering, as for software success it is essential to examine the vibes of stakeholder like what stakeholder would like to perceive, observe and feel about product? Hence emotions provide basis for software development in a case when client is unable to explain the requirements. But unfortunately little deliberation and consideration is given in the discipline of requirement engineering . In order to find out the reason behind this, it is absolutely necessary to understand the basic concept of human emotion and have to know the answer of some questions like what is meant by emotion?, how they are expressed and classified. In this section, a brief sketch on human emotion and its integration in software development is provided which goose motivation towards research.

1.3.1 Emotions

Emotion regulates the thoughts and impression of human being and how human react at a particular situation. Defining an emotion is a challenging concept which attracted many researcher to explore the concept of emotion and their impact in various domains. A number of definitions has been advocated by researchers' overtime. In [8] emotion is defined as a complex form of feeling that results as psychological and visible expression which influence the person way of thinking and reacting. For example, a student who was not expecting bad result will express a strong emotion like "shocking". Every theorist viewed emotion with different perspective, as [9, 10] highlights twenty dominant definitions of emotion from the year 1884 to 2000. In [4, 11] explain that emotion have four main components, (i) cognitive response, (ii) behavioral response, (iii) affect/subjective feeling, and (iv) physiological arousal.

Cognitive response: People express their emotions in the way they learn, memorize and perceive things. For example, thinking of a best friend can provoke emotion of happiness

Behavioral response: The behavior, action and response of a person after experiencing the emotion. For example, someone moves to and fro and drinks water continuously, gives impression of muddle.

Physiological arousal: It is the emotion that provokes some changes and reflect informal conversion. For example changes facial expressions and increased heartbeat gives the impression anger.

Affect/subjective response: Study on affect/subjective feeling implies that people relate their internal feelings as emotions. For example an activity is given to participants and they were asked to describe their emotion after experiencing that activity. They gave answers in words such as anger, confused, happy, and depressed.

In [12], it is pinpoint that emotion is not only an expression of a person but actually it is a complex chain that initiate with the feeling and boost with the change in expression, mood and behavior goal of a person, which indicates that emotions are the outcomes of a particular action which effect the feeling state of a person. Some theorists consider that emotions are directly related to the individuals' internal frame of perceiving things. For example, hearing a creepy voice in dark could be thrilling for someone but could also be scary/frightening for someone else. Hence,

> ”Emotions are expressed in the state that shows some visible physiological change in facial expression, action or mood of a person.”

1.3.2 Classification of Emotions

Many theorists defend that emotions should be categorized into some basic/primary groups and all other emotions should be assemble under those basic emotions. The classification of emotions goes back to ancient Romans and Greeks. A theorist Cicero classified emotions into four primary groups; Letitia (pleasure), metus (panic), libido (love), and aegritudo (pain). A number of classifications of emotions have been proposed because theorist disagree with the number of basic groups and also with the number of emotions lies in those groups. Different definitions present by theorists, about basic groups and their classification are shown in Table 1.4 [11]. The different views about basic/fundamental emotions leads toward confusion regarding to their classification. Out of all these definitions the most extensively used and supported definition is Ekman's theory which includes six emotions namely; happily surprise, sadness, horror, joy, disgust and anger.

Table 1.4: Classification of Emotions

Theorist	**No. of Emotions**	**Emotions**	**Year**
Arnold	11	Anger, hate, daring, depression, desire, aspiration, fear, dislike, anticipate, love, and sadness	1960
Frijda	6	Desire, happy, attract, surprise, wonder, reject	1986
McDougall	7	Anger, disgust, ejoy, fear, oppression, caring, wonder	1926
Panksepp	4	Excitement, panic, rage, fear	1982
Watson and Rainer	3	Fear, love, rage	1920
Ekman, Friesen and Ellsworth	6	Anger, disgust, fear, joy, sadness, surprise	1982
Matsumoto	22	delight, hope, annoy, disgust, sad, shock, fear, acceptances, reserved, pride, admire, calm, respect, disrespect, love, happy, thrill, regret, ease, pain, polite, likeness	2005

Emotions are classified by keeping in view different theories on emotions. Out of all 3 most important theories are discussed in this section: [13] classified emotions into 6 basic sets which are communicated and delivered by facial expression. These emotions are horror, surprised, angry, disgust, sad and happy. Secondly, most complex taxonomy of emotions are classified into 22 basic emotions, which depends upon the reactions against event, object and person. This taxonomy of emotions is said to be as OCC model proposed by Ortony, Collins and Clore. The OCC model distinguish and perceive emotion in term of their basic conditional and environmental meaning. The third one is Norman's model which classified emotions by placing them into three layers; (a) Visceral layer, (b) Behavioral layer, (c) Reflective layer

Applying Norman's model in RE it is visualized that ”Visceral layer” concerned with overall appearance or quality of the product and its aim is to get inside the mind of user in order to improve the visual interface of the product. It is an automatic response result in fast pre-conscious judgment, based on the overall look of software, such as size or color. It is programmed as ”look&feel” in humans mind and have a consistent effect on different people. ”Behavioral layer” refers to the emotions that someone expressed on the completion of a product and after experiencing the functionality of software. This expression results in the acceptance or rejection of a product, the third layer is ”Reflective layer” in which conscious emotions are justified and rationalized by knowing the significance of software [14, 15]. Table 1.5 describes the Norman's emotional design model.

Table 1.5: Norman's simplified model of emotional design

Layer	**Software Characteristics**
Visceral layer	Project appearance
Behavioral layer	The satisfaction and effectiveness during use
Reflective layer	Personal experience

1.3.3 Ways of Expressing Emotions

In early $1960's$, it was claimed that human express their emotions through verbal and non-verbal means [16]. This study was further supported by many researcher [17, 18, 19]. From literature on emotions it is found that the total effect of the message/speech, tone of voice accounted 38%, words accounted 7% and non-verbal language accounted 55%. Hence in case of determining the emotions of a human the non-verbal statement have greater impact over verbal communication. In [20], the most regular means through which emotions are demonstrated are anticipated as:

i. **Body posture**: while walking, bending head all the time indicates nervousness and stand so close to someone indicates love or dominance.

ii. **Gesture**: pointing, moving and waving hand while communication, indicates emotions experiencing by a person.

iii. **Facial expression**: to detect emotional state of a person, face is the strongest medium. For example, smile on face indicates that the person is happy.

iv. **Physiological cues**: moment of specific part or movement of different parts of body at the same time termed as Physiological cues. For example increased heartbeat, increased temperature, deep breath, narrowing of the eyebrows etc.

v. **Speech patterns**: human also express their emotions through words/speech patterns. For example if a person use bad/negative words it shows that the person is unhappy. Vocal information like pauses, tone, stress, frequency and pitch also conveyed emotional state of person.

1.3.4 Emotions in Requirement Engineering

In the domain of requirement engineering specialization, it is believed that stakeholder perceptions about requirements are manipulated, influenced or affected by emotional responses moreover functionally accomplished system will not be accepted by the stakeholder unless it will not appeal to emotional state. Generally Requirement engineering have much developed methodologies which fully grasp the functional and non-functional requirements of stakeholder but till now, questions like how stakeholder wants to feel? And what are the emotional needs of users? Are not fully investigated. This concept is clearly understand by keeping in view the statement of Hartmut Esslinger;

> ”Even if a design is fully elegant and functional, it will never have a place in our life unless it can attract and appeal at a deepest level by realizing our emotional needs and satisfaction.”

It suggest that insufficient attention to requirement may cause software failure as software engineers are supposed to construct systems with desired functional and non-functional requirements. These requirements may not be declared by stakeholder or clearly gathered by developer during RE process, which results in the form of failure of software as it actually fails to fulfill the thirst of its users [15].

One of the reason for system rejection is due to the misconception that the emotional requirement can be labeled after system's deployment by simply fixing them. From this perspective, eliciting the emotional needs of user should be addressed as first class citizen during requirement gathering process. The study of emotions in requirement engineering enables and facilitate the developer to predict feedback, about a product, through human expressions which will help to eliminate the pitfall and downside of a product. In RE emotions may also be used by a developer as a tool of thought for successfully gathering requirement of a product and its relation with human responses such as trust, satisfaction, anger and disappointment. The response in the form of emotions can be tracked back to the sourced feature and once the source feature is known the requirements can be specified by removing the offending text, event or image. Emotional responses can be tracked back to the source feature which bring disapproval. Disapproved feature indicates missing requirements or imperfect design. For example smiling face indicates satisfaction whereas aggressive face communicates disapproval about the product [14].

Research shows that during collaboration tasks, emotional awareness is influential as it allows team members (developers) to act appropriately for better end result of product, by keeping in view how someone feel and perceive the product. As requirement engineering is highly collaborative process in which developers exercise mailing, communications, reports and software code stocks to conduct and manage their tasks. These shared artifacts are the communication medium where they express their feelings/emotions regarding to projects. By studying the literature on emotions in RE it is found that emotional attitudes of developers and stakeholder's can have direct influence on the creativity, inspiration, originality, quality and productivity of the product [14, 21]. Many scholars such as Daniel at el, suggest that for better understanding quality, productivity and effective job performance it is critical to flourish and bloom happiness among stakeholders/developers. As happiness leads towards the success of system. Recent study on "what happen when developers are happy/unhappy while developing software?" shows that happiness (positive emotions) can improve performance and quality of software [22]. The summary of study is shown in Figure 1.2.

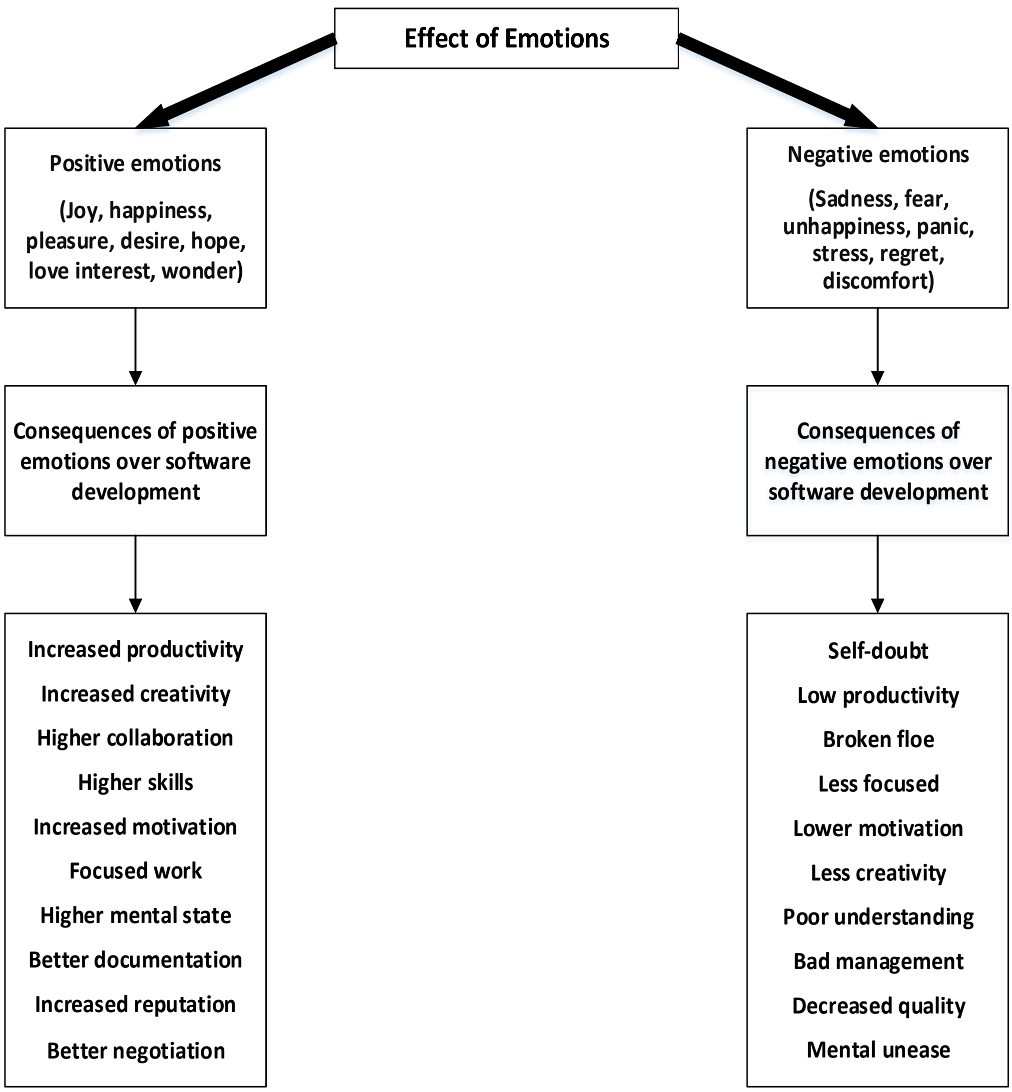

Figure 1.2: Effect of Emotions on Software

1.3.5 Implication of Emotions during RE phases

In Requirement Engineering (RE), the human dimensions sometimes have more significance than that of technical dimensions therefore the supervision of people during projects development became critical and researchers specified with abundant experiments, which verifies that human outlook and aspects became a source of central complications linked with RE process. It is because of the fact that RE process is fundamentally established on social and intellectual activities [2, 3], hence human aspects specially emotions of stakeholder should be addressed in an accessible way so that it will provide an outline about people's attitude and opinion in order to grip on the context of social-political misunderstanding about product requirements that may affect software development [11]. Keeping track of human emotions during requirement elicitation helps developer to read the mind of stakeholder regarding to their needs, expectation, values and excitement which will provide guideline for changes and specifications during project configuration [23]. Understanding emotions during elicitation process clearly interpret the functional and nonfunctional requirement which goose motivation towards development and customization. As some people didn't react directly towards the end product but they do illustrate their reactions through patterns, stories, gesture, body posture and expressions hence, [24] consider scenarios schemes and story depicting audio's as tool for illustrating the participant's responses during elicitation, elaboration and negotiation phases in RE. It is also essential to record the emotions of a developer because software development have that potential to change the emotional state and motivation of a developer during product customization. For example the secured information by participants become available or a system may change the power balance during product configuration may result in frustration and may block the overall process. Requirements are not static due of their progression with time. During the process of the requirements gathering, emotions may exists from ”elicitation to negotiation” and from ”modeling to prioritizing” [3].

In [24], it is considered that monitoring emotions in RE process help developer to understand the requirement in a proper manner as many analyst reported that during management and product design several stakeholder's didn't speak clearly about their needs hence analyst had to use their common sense to understand what might be compulsory for stakeholder in desired product. The survey on the consequences of emotions in negotiation phase during RE revealed that there need an awareness regarding to socio-political issues as many analyst mentioned that they didn't give much attention towards stakeholder's VMEs (Values, Motivation and Emotions) unless they got stuck during designing and configuring the project requirements. The significance of human's emotional aspects in requirement gathering is reflected by the statement that, irrespective of the procedure, method and tools employed, the realization of software requirement exploration depends on exactly how skillfully analysts and stakeholder connect, collaborate and negotiate with one another [2]. Table 1.6 mentioned some of the important VMEs that should be kept in mind by developers during requirement gathering.

Table 1.6: Values, Motions and Emotion

VME's occurred in RE PROCESS		
Sentiments	**Related terms**	**Implication phases**
Trust/Pleasure	Happiness, joy, hope, love, reliability	Inception and specification
Collaboration	Sympathy, joy	Elicitation and management's positivity
Motivation	Wonder, expectation, surprise	Negotiation and specification may leads towards clashes
Fear	Threat, worry	Negotiation and specification may be conflicted
Frustration	Anxiety, disgust, depression	During design phase and configuration of product

This research aims to create and test a technique for the prediction of emotion in the process of Requirement Engineering. The eventual objective of this research is to implement emotional measure of stakeholder's necessities for more significant requirement's elicitation, elaboration, management and greater appropriateness of requirements specification.

1.4 Extraction of Emotions from Audio Signals

Increasing communication interaction between digital media and human made emotion recognition from audio signals, as one of the attractive research topic. For more natural and clear interaction among human and digital world, computer system should be able to recognize emotions as same as human do. Human speech is a very complex signal which also carries information about gender, age, feelings and emotions of a speaker hence for human computer interaction, it is essential to conceive emotion so that we can response accordingly. Communication is a primary mean, which is constructed by linguistic explanation along with emotional state of mind, and that emotional state of mind, have direct influence on person's facial expressions, gesture, posture, linguistic context and voice characteristics. In our research we are going to focus on the extraction of emotion through voice, which is a fundamental faculty of human communication. Emotion recognition from audio signal is a complex process which includes following dominant steps:

i. Isolation of audio signals(linguistic/paralinguistic)
ii. Pre-processing
iii. Feature selection reduction
iv. Database design
v. Classification

1.4.1 Isolation of audio signals

As human speech carries paralinguistic information along with linguistic information that articulates emotions of a speaker. Linguistic part pinpoints qualitative statement articulated by speaker whereas paralinguistic information identifies the variations in the way words are pronounced. These variations includes pitch, patterns, pauses, vocal quality, stress and intensity. Sometime linguistic information is used for classification by using Automatic Speech Recognition (ASR) systems that recognize pre-defined speech or words which do have some connection with emotions of a person. For the victorious administration of linguistic information in emotion recognition systems, a well-organized and update dictionary is required in order to get reliable identification of emotions. In [25], an approach was developed for emotion's identification by using linguistic information as input. The emotional salience words were distinguished by two hyper-classes (positive and negative). The words were defined on the basis of relationship between emotion and that of specific word. For example the word ”good” was placed into the positive salience word class whereas the word ”wrong” would place into the negative word class, which indicates negative emotions, hence the statement which have more number of negative words identifies negative emotion [26].

In case of non-linguistic information, features are categorised in two groups namely as functions and Low Level Descriptors (LLDs). Functions are derived from (LLDs) and these functions are also known as statistical features of LLDs like mean, variance, change rate, minimum/maximum, zero-crossing rate and so on. LLDs are prosodic features includes frequency, Mel-frequency cepstral coefficients (MFCC), pitch, formats, energy, shimmer, jitter, vocal quality and speaking rate parameters. Table 1.7 summarizes various speech features and provide their description [26].

Table 1.7: Speech features with description

Features	**Features Description**
Mel-frequency cepstral coefficients (MFCCs)	Derive through cepstrum, shadowy
Linear prediction cepstral coefficients (LPCCs)	transform of spectrum logarithm
Formants (spectral maxima or spectral peaks of the sound spectrum of the voice)	Derive from cepstrum
Jitter, Noise-to-harmonic ratio, shimmer, spectral balance,amplitude quotient	Measurements of Speech signal
Energy, low energy	Measurements of power and intensity
Fundamental frequency and pitch	Measurements of pitch and frequency
Log-Filter-Power-Coefficients (LFPCs)	Measurements of audio signals for voice quality
Temporal features such as time and duration)	Measurements/Calculation of time

1.4.2 Pre-processing

The pre-processing step includes audio/speech stabilization, pre-emphasis filtration, normalization and deletion of still intervals. These silence intervals are detached by employing logarithmic process for splitting, segmenting and separating speech signals from noisy background.

1.4.3 Feature selection/reduction

A significant complication in emotion recognition framework is the nomination of a supreme set of features, to pervade and characterize the audio signals. The intention of feature segmentation, is to accurately specify feature's subset from indigenous set to improve the classification accuracy and reduce the computational cost with respect to time. Hence we can say that in order to reduce required time to complete recognition process and to limit the computational complexity, it is crucial to minimize the number of features. Usually it is expected to employ more number of features to improve classification accuracy. Nevertheless, performance can be decreased when number of features are increased, when pattern's numbers are too compact, this muddle is known as ”Curse of Dimensionality” hence to speed-up the task a process known as variable selection/feature selection is done. Several methods can be applied in order to get the optimal subset of variable features. Out of most well-known methods of feature extraction some are Sequential Floating Forward Selection (SFFS algorithm), SVM Sequential Forward Flouting Search algorithm, Mutual Information (MI), Genetic algorithm, Correlation-based Analysis and BestFirst [26, 27].

1.4.4 Classification

Generally classification is accomplish by using an isolated dataset/database. In this prototype, various testing frames developed based on content (dependent/independent) along with speaker (dependent/independent). According to literature, most commonly used frameworks are that of ”independent”, as they provide more reliable and well-grounded evaluations includes Leave-One-Speaker-Group-Out (LOGOS) and Leave-One-Speaker-Out (LOGO). For any audioo database there are some fixed recording conditions such as room acoustic, noise level and language (data should be recorded only in one language).Moreover, system must be trained and then tested by using same dataset/database. Cross-Corpora evaluation method was proposed in [28] to simplify the training phase of the system and to increase independence between testing and training sets. For emotional state classification/modelling researchers used single classifier scheme as well as Hybrid classifier scheme. In single classifier scheme only one classifier is used for classification of emotions such as Support Vector Machine Mechanism (SVM), Artificial Neural Network (ANN), Gaussian Mixture Method (GMM),Hidden Markov Method, k-Nearest Neighbor and Decision Trees. As some emotions were not well evaluated by single classifier scheme like neutral emotion may not be recognized well by GMM, therefore in order to achieve higher performance Hybrid-classifier scheme was introduces. Table 1.8 indicates the performance differences between single and hybrid classifiers schemes.

1.5 Motivation

Emotion is an essential feature in communication that’s why emotion appreciation from speech has emerged as a significant research field in the recent past. Human speech is a complex signal that encompasses information about speaker, statement, language and emotions. Emotions make communication more real, expressive and effective. Emotion recognition means discovery of emotional state of speaker through features extracted from his or her recorded audio signal. As modern innovations in neurosciences, along with emotional intelligence leads towards the emergence of a new framework in the field of ”Machine Learning” and ”Affective computing”, which objective is to construct machines that identify, communicate, recognize and then respond according to user emotions. Hence emotion recognition through audio processing is an area that escalates the attention towards human-machine collaboration [31]. The motivation for ”extracting emotions from audio signals for effective requirement engineering” is to build a system which automatically identify human’s emotion through their communication during requirement gathering which may help a developer to shape a system according to the needs of customer. Literature suggests that among others like facial expressions or impersonator, gesture, posture, recognizing emotions from audio is one of the most promising and encouraging

Table 1.8: A brief review of execution differences between single and hybrid classifiers schemes

Classifiers	Performances Results	Year
SVMs	70%, seven emotions (comprehensive dataset) ˜63.1% in a dataset which includes clear and misleading audio Upto 80% in various cross-databases experiments with different amount of classes 78% in Berlin EMO dataset	2005 2006 2010 2010
GMMs	˜62% in a corpus that includes deceptive and non-deceptive speech 70.4%, five emotions in a dataset 86% in Chinese-LDC	2006 2007 2009
HMMs	62.6%, five emotions in constructed database) 78.1% in Berlin EMO dataset which is a speaker independent database ˜81.5%, five emotions in a dataset	2007 2008 2009
ANN	63.4% in Berlin EMO dataset and 61% for a unknown dataset 68%, four emotions (unknown database) 71.5%, four emotions (unknown database) 47.1% in Berlin EMO dataset which is a statement independent but speaker dependent database	2008 2009 2009 2010
K-NN	67%, six emotions 72.3%, five emotions	2007 2007
SVM, GMM, MLP	80.1% for speaker independent database , four emotions in 2 datasets	2011
SVM, K-NN	89%, five emotions for self constructed dataset	2007
SVM, MLP, K-NN, RF	73.4%, six emotions for unknown dataset	2007

research field.

1.6 Research Questions

i. What is the finest machine learning technique for emotion recognition from audio signals?

ii. Does automatic emotion recognition from audio with high accuracy, can increase the efficiency of requirement gathering process?

1.7 Contribution

Emotion related requirements are critical element for the acceptance and success of any software system therefore emotions should be preserved and treated with equivalent significance as that of functional goals during requirement engineering. Addressing emotional goals of user during requirement engineering process of software project expands the likelihoods of these goals being reflected in the innovative period of the software development life cycle. This thesis, focus on the importance of user emotions in requirement engineering phases and presented a technique which extract emotions from customer’s communication for an effective requirement gathering process. More precisely, the foremost contributions of this thesis is summarized as follows:

i. This thesis highlight the significance of emotional goals as being a serious factor for the success of software. Given that emotional objectives refer to the user feelings, opinions and emotions generated from users' involvement with software. Study suggested that emotional goals do not relate to functional requirement and drop into the group of non-functional requirement hence it is supported by thesis that ”Quality goals are properties of the system while emotional goals are the properties of the user”. Investigation also specified the prominence of user emotional goals with respect to the approval or rejection of a product, so emotional goals cannot be ignored and should be specified in requirements documentation.

ii. Finally this research presented a method to generate emotion from audio speech of a customer for effective requirement engineering. This method was designed to build a bridge, in order to reduce a gap between the requirements stage and the design stage.

1.8 Thesis Organization

Chapter 1 illustrated the brief description on Requirement Engineering, steps in requirement engineering, emotions, ways of expressing emotions, classification of emotions, modelling of emotions in requirement engineering, and extraction of emotions from audio signals. Chapter 2 presents literature review and also explains the baseline structures for an automatic emotion recognition system. It gives an outline, of state of the art audio features extraction techniques and classifiers for emotion recognition from audio signals. Chapter 3 give out a through discussion on speech features, classifiers, emotion classification and components of emotion recognition system. In Chapter 4, the proposed methodology is illustrated that how emotions are extracted and classified from audio conversation. Chapter 5 debate about the experimental setup and results along with the supporting datasets and their application on requirement engineering. Thesis is concluded in last Chapter which includes conclusion, future work and recommendation.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

In [14], a review is provided on the emotional theories applied in the scenario based procedure of requirement engineering. The paper provide a mini tutorial demonstrated by case study providing a difference between moods and emotions along with their importance in RE. The study concludes that moods are temporary as they expressed good or bad feeling which define personality traits of a person whereas emotions are the responses against event, object or person which have direct influence on functional and non-functional requirements. In [11], emotional goals are assessed, evaluated and represented by applying emotional-oriented requirement gathering techniques. By applying proposed approach researcher develop a sofiHub - which is a smart home for old people having sensors/devices for sensing movement, communication, resistless monitoring and exigency activities. Emotion oriented approaches are used to identify and evaluating the reactions of participants about installed devices through their emotions. By keeping track of participant's emotions, analyst observe that emotion-oriented system develop a strong bound between the product and the user. For example, if the product (sofiHub) didn't response as expected by the participants they get frustrated and foiled. Study shows that for the success of any software capturing, realizing and identifying emotions are critical and play a vital role for adapting such systems where goals are personal.

A tutorial is provided in [21], for enhancing emotional appreciation among developers through quantitative emotional encapsulation and summarization. The applied approached automatically extract emotions from collaboration stock of sentences, mostly used by developer by applying Lexical Sentiment Analysis and Latent Dirichlet allocation (LDA). This approach favors the idea that human produce negative and positive emotions at same time. Hence SentiStrength assign +score to positive words and -score to negative words used with in a sentence then TMT (Topic modeling toolbox) is used for evaluating final emotions. This technique was applied on

1000 artifacts generated by three software development teams and the outcomes of this technique shows that it could be helpful to detect the emotional state of the developer during any software development project. [23] highlights the importance of emotion, motivation and values of human psychology during elicitation phase in requirement engineering. The paper highlights the fact that understanding user's values and beliefs is mandatory for the success of any software, as it help to understand requirements during system configuration and customization. The technique mention by researcher is known as HyperText advisor which interpret the psychology of stakeholder during RE.

In [24] values, motivations and emotions are cited as matter of contention in the process of requirement engineering. In this context, an approach for the investigation of social-political issue is introduced, which is supported by a website (VBRE).The website illustrate the problems originated during elicitation like poor understanding of requirements and values conflicts. VBRE is evaluated by requirement engineering experts and by trainee. A case study was also done in the support of this approach by negotiating with stakeholders and expects .This case study provide evidence that how VMEs are identified and analyzed during elicitation phase in RE.

The impact of emotional consequences (happiness and unhappiness) on the productivity of developers during any software development project are studied in [22]. The survey done by Daniel at el, discovered 32 consequences of happiness and 42 consequences of unhappiness which are identified/declared by participants. The spectrum of participants includes 75% of software developers, 14% of nominee and 9% were other characters like (CEO, manager or academic researcher). The results shows that the consequences of happiness and unhappiness have direct impact on developer's productivity, creativity, ability, performance and collaboration. This paper clearly differentiate the outcomes of happiness/unhappiness among developers and illustrated that happiness results in the form of high motivation. Higher skills, increased productivity and improved collaboration whereas unhappiness consequences may affect the mental health result in anxiety, depression, stress and burnout which footprint low production and bad performance.

The case study [15], argued that the emotional needs of stakeholder's must be considered as first class citizen in requirement gathering. Miller at el, proposed a PORE (people-oriented requirement engineering) model in this scenario. The results of case study indicates that emotional understanding seems to be a positive step as many system such as (Emergency system for old people) were failed to address the emotional needs of their users, hence applying PORE model they make significance improvements in the quality and performance of emergency system. By keeping in account the emotional responses they design, evaluated and implemented a new prototype which was adopted by participants without any hesitation.

In [29], a system to recognize emotions from audio signals is proposed. The researcher used around 120 audio files of ".wav" format for analysis of training and testing algorithms. The dataset were collected from different group of people which were differentiated with age, gender, etc. There were eight types of emotions evaluated which includes fear, sad, nervous, anxiety, bored, happy, depressed and angry. Preprocessing and Feature Extraction is performed on audio signals to filter the noises and to compute the Duration, ZCR (zero crossing per unit time) and Energy by some specific mathematical formulas respectively. After that Pitch detection, Format frequency and Feature combination techniques are performed to make the sounds uniform and resize. After these are performed, the audio files are further classified by using statistical (K Nearest Neighbor KNN) classifier and Neural Network based classifier. These classifiers were implemented on MATLAB. In Neural Network based classifier, emotions like bored, angry and sad were recognized accurately whereas in KNN based classifier, emotions like anxiety, angry and fear were recognized accurately. The rest of the emotions were not as such accurate. Statistics shows that Feature length of the audio files were 102, in which KNN got 85.48 and Neural Network got 80.0. Therefore, this concludes that KNN classifier is better than Neural Network classifier due to more accurate results.

Audio emotion recognition is accomplished via processing approaches which contain isolation of audio signals and abstraction of nominated features for classification. In the course of sound quality, communication processing methods suggest tremendously appreciated paralinguistic data resulting mostly from spectral and prosodic features. In certain instances, the procedure is supported by audio recognition systems, which is added to classification accuracy by experiencing linguistic evidence. Both structures contract with a stimulating issue, as emotions do not have specific boundaries and frequently vary from individual to individual. [26] explored emotion recognition from different acoustic frequencies, which are then measured and classified, by following the methodology of construction through feature selection and classification procedure. Imperative issues from various classification approaches, such as datasets accessible for investigation, suitable feature abstraction and feature selection procedures, classifier's performances are also explained. The study emphasis the attention towards the research done in last era. The review also delivers an argument on open tendencies with a proper guidelines for upcoming research in this area.

In [27] emotions and gender of speaker is recognized from speech signals by using SVM. The proposed system is able to recognize human emotions formulated by two sub-structures that are Emotion Recognition (ER) and Gender Recognition (GR). Support vector based emotion classifier implement gender data as an input which provide a priori knowledge about the gender of that particular speaker. This algorithm is works on the pitch extraction method. Principal component analysis (PCA) is then investigated throughly and applied for feature selection. For classification, database known as Enterface Database (ED) is employed for testing and training the SVM-based classifier. The result indicated that prior knowledge about gender of speaker increases overall performance of the emotion recognition system's accuracy from 80.4% to 84.5%. [30] highlights the three fundamental aspects for designing Emotion Recognition System from speech signals. The basic step is the choice of appropriate features for speech depiction. The next step is the design of a relevant classification scheme and the third step is the formal construction of emotional speech database for the evaluation of system's productivity. The survey demonstrated suggestions about improving emotion recognition system's performance. It is concluded from literature that the recognition accuracy for speaker-dependent classification is unto 90% and that of speaker-independent is as low as 50%.

To analyze the behavior and characteristics of real time speech, a methodology is proposed in [31], which extracted acoustic features from speech signals. The system validate 4 basic emotions; happy, angry, sad and neutral. For feature extraction, MFCC, energy and formats are used. English database is employ along with SVM classifier for recognition of emotions.in order to map data to an intense dimensional feature's attribute space without misplacing/losing the originality, the Kernel Function of SVM is implemented. The paper concluded that MFCC minimize the frequency data into small chunks of co-efficient, which will executed fast and easily. 3-stage SVM classifier is implemented to classify 7 emotions; angry, disgust, fear, happiness, bore, neutral, and sad which are available in "Berlin Emotional Database (BED)". The methodology of [32] extracts MFCC features from all 535 files present in database. The hierarchical support vector machine along with linear kernel function is applied.Cross-validation is then used for training and testing the data. The accuracy achieved is about 68%.

Pitch, formats, energy and MFCC parameters are selected by[33], in order to extract features from speech signals for emotion recognition. Extracted features are investigated by applying Quadratic Discriminant Analysis (QDA) and then classify by using Support Vector Machine. The result indicated that pitch and energy are most useful feature factors. It also presents a comparison between various classifiers such as; Linear Discriminant Analysis, Quadratic Discriminant Analysis, Support vector machine and hidden markov model. The database SUSAS which is text-independent is used and accuracy achieved is about 70% for SVM and 64% for HMM. [34] recommends a set of research methods for emotion recognition, to be applied in the domains of: software engineering, education, website customization and gaming. The objective of implementing the setups is to evaluate the possibility of exhausting emotion recognition methods. It also highlights the complications of outlining sets of emotions to be predicted in various applications, expressing the clear emotions, gathering the records and training/testing data. The paper also presented number of potential applications of emotion recognition and highlights some limitations such as emotion illustration prototypes not always exactly designate the authentic emotional state of a person hence developer not always realize the actual possible responsive states. Moreover training data's quality and assigned emotion labels intensely influence the outcomes of the classifiers. The researcher pointed out that the precision of recognition method not always accomplish the requirements of real time systems. Hence, this should not be ignore as it is an open research problem which requires exploration.

In [35] experimental investigations on emotion recognition and expression for a call center was presented. The main study deals with 700 statements which expresses 5 emotions: happy, fear, anger, sadness and neutral state, which were expressed by 30 non-professional artists. This evaluated corpus is then used for feature extraction and training neural network classifier. By using feature selection methods, some statistics of energy, formants, pitch and speaking rate were selected as appropriate features. Neural Network classifier is used for classification of emotions which established the subsequent precisions: ”happiness: $61-70\%$, anger: $70-81\%$, sadness: $71-86\%$, neutral: $60-75\%$, and fear: $36-55\%$.” The accuracy of the research was about 70%. The additional study uses an amount of fifty-six telephone written communication of span (from 16 to 90 seconds) articulating commonly angry, happy and neutral emotions which were expressed by 18 random actors. These expressions were then used for constructed (training and testing) database. The recognizer classifier differentiate between two conditions: ”calm” (includes neutral and sad states respectively) and ”agitation” (includes anger, fear and happiness). The average accuracy of proposed system is about 77.1%. The purpose of such system is to use as a fragment of an assessment backing method for placing voice messages, emails in order to assign a suitable executive to response appropriately.

In [36] features are extracted by applying Principal component analysis (PCA) due to which measurement of the processing becomes less and also compares the outcomes of Gaussian Mixture Model (GMM) classifier with other classifiers. The paper proposed a Boosted-GMM algorithm, which inserts the Expectation Maximization (EM) algorithm into a refined boosted framework, which is used to evaluate the class-conditional probabilistic distributions accurately, in any pattern recognition obstacle constructed by training data set. The Boosted-GMM algorithm is applied into speech emotion recognition and research illustrated that ”emotion recognition processes are successfully and significantly increased” by employing Boosted GMM algorithm as compared to that of EM-GMM algorithm. This is due to the statistics that boosting can lead to more accurate estimation of the distributions of acoustic features. In [37] a system is proposed that is used to extract and recognize emotions from audio signals produced by speaker. The system engaged two statistical prototypes one is Support Vector Machine (SVM) and other is Hidden Markov Model (HMM). For emotions recognition four acoustic features are extracted such as spectral centroid, spectral projection, spectral spread, and spectral flatness. System is distributed into various stages including audio pre-processing (used to remove noise exist in audio signal), feature extraction (extraction of four acoustic features), segmentation (employed to divide audio clips in to unvoiced and voiced category), training and testing of database (implemented to train prototypes for classification purpose) and classification. Database is developed by generating audios which implicates the speech expressions belongs to 10 different emotions. This method was claimed to be very operative for the classification of human emotions with worthy recognition rate. This technique is beneficial for psychiatrics to regulate the emotional position of a person as well as it is also operational in E- learning environment to identify the temperament of listener. This method was implemented using MATLAB software.

Emotions are significant subjects in human’s behavior. The significance of emotions and their management in software development is obvious as taking into account by description the requirement engineering is a human principal intensive activity. In this research field, [2] suggests the integration of people’s sentiments in RE procedure by the mean of employing affect grid which is classic device aimed to analyze and evaluate the emotions which are incorporated into the organization of the requirements. Outcomes of this implementation demonstrate that emotions are the key factor that should take into account while establishing requirements steadiness, as understanding stakeholder's emotions also implicates the knowledge about reliability and stability of user's requirements and needs. In addition, stakeholder should accept that ”there are no magical procedure of solving RE issues because complexity is indispensable in a software development process where their is no proper modelling of emotions” (Brooks).

A comparative study on different procedures for feature selection from audio recognition system is given in []. The study explore some of the feature extraction method such as Mel frequency Cepstral Co-efficient (MFCC), Zero Crossings with Peak Amplitudes (ZCPA), Dynamic Time Wrapping (DTW), Relative Spectra Processing (RASTA) and Linear Predictive Coding (LPC) Analysis along with their pros and cons. This investigation indicates that hybrid methods should be developed in order to improve the overall performance of the recognition systems. The concluding highlights shows that MFCC is easy to execute moreover it is relative fast to extract features from audio signals because MFCC reduces the frequency data into small numbers of co-efficient and the logarithmic function interpreted the loudness according to the human auditory system.

2.2 Summary

This chapter initially supported different outlooks of emotion, containing theories of emotion, ways of expressing emotion and concept about the extraction of emotions from audio signals for effective requirement engineering. Based on outcomes from the literature on emotion, it is suggested that emotion is a subjective and complex subject and various concepts have appeared to define this topic. An essential conclusion from the literature surrounding the integration of emotion in software systems is that, even though emotion has been comprehensively investigated in areas such as Human Computer Interaction, game design, and affective computing insignificant attention has been given to emotions in the field of Requirement Engineering. Study on emotions in RE, advocates that user emotional expectations act as a serious factor in the acceptance or rejection of any software. Even though some investigations have highlighted the compulsion, to introduce emotions within the requirement engineering phases specifically during elicitation and validation. Evidence also indicates that insufficient attention on user emotional need is a foremost reason why end-users refuse to accept end product.

CHAPTER 3

A REVIEW OF FEATURES AND CLASSIFIERS

3.1 Overview

This chapter gives a comprehensive review of essential features and classifier that are used for emotion recognition from audio signals. As to find out the statistically appropriate data from incoming audio records, it is essential to have mechanisms which reduce the data of every individual fragment from audio signals into comparatively compact number of frames or parameters known as features. These features defines every fragment in such a distinctive way that the same fragments are joined/grouped together by associating their structures. Generally, there exist number of stimulating ways that describe the audio signal in terms of their constraints, which are known as feature extraction methods [38]. Feature extraction is a process which is used for dimensionality reduction that shrink a data which is oversized to be processed accurately, by an algorithm. In feature reduction the input raw data is converted into a finest set of features, which do have all that relevant information which is required to perform a desired task by using the reduced set instead of using full lengthy raw data. The audio recognition method has a background of Digital signal processing (DSP), which is a center of evolution in any audio processing throughout the whole development and construction of speech recognition systems. The leading aim of feature selection method is to interwine the speech signals into the diverse acoustically recognizable components and to find out a standard set of features with low rate of alter, to get the computational accuracy in less time. Numerous feature extraction techniques are existing which include; First order derivative (DELTA), Linear predictive, Mel scale cepstral analysis (MEL), Perceptual linear predictive coefficients (PLP), analysis (LPC), Power spectral analysis (FFT), Relative spectra filtering of log domain coefficients (RASTA), Linear predictive cepstral coefficients (LPCC),and Mel-frequency cepstral coefficients (MFCC) [39].

In this chapter MFCC, LPC, LPCC techniques which are used in our research for feature extraction and GMM-UBM classifier, used for emotion classification, are described in detail.

3.2 Mel-Frequency Cepstral Coefficients

The human speech contains frequent discriminative structures/features that are used to recognize speaker's emotions and identity. The objective of speaker's emotions recognition is to take out, illustrate, characterize and then recognize the information about human's state of mind. As, human speech contains momentous energy from 0 to 5 kHz frequency and the possessions of speech signal alters significantly as a function of time therefore the concept of time varying Fourier representation is used to review the spectral properties of audio signals. Conversely, the sequential properties of audio signals like pitch, zero crossing rate, frequency, intensity and energy are supposed constant having short-time fixed characteristics. Therefore, Speech signal is distributed into blocks of short duration by applying hamming window. The procedure of audio classification includes extraction of biased or discriminatory features out of the audio data and serving them to a classifier. The features can be extracted either straight from the time sphere or from transformation sphere depending upon the approach selected for signal analysis. Different approaches for features extraction were proposed with erratic success rate [40]. Out of all the most widely used feature extraction approaches in automatic emotion recognition the finest and accurate technique is Mel-Frequency Cepstral Coefficients (MFCC).

To abstract a feature vector which holds all the relavent data about the communication, MFCC duplicates/mimics some fragments of human speech. MFCC also copycats the logarithmic sensitivity of pitch, loudness, intensity of human acoustic scheme and attempts to exclude speaker dependent features by eliminating the ultimate frequencies along with their harmonics. To represent the dynamic nature of audio, MFCC also accept the modification of selected features set over time [41]. MFCC is constructed on the model of human marginal acoustic mechanism. The human observation of the sound frequency for any audio does not follow linear scale. Therefore for respectively tone with a definite frequency $'f'$ measured in Hz, and a particular pitch is measured on a particular scale named a 'Mel Scale'

$$f_{mel} = 2595 log_{10}(1 + f/700) \tag{3.1}$$

where f_{mel} represents particular pitch in (Mels) analogous to the frequency in Hz and indicates the description of MFCC, which is a standard acoustic feature vector for emotion recognition systems [40, 42]. The standard implementation for extracting features using Mel-Frequency Cepstral Coefficients is presented in Figure 3.1 and process is described below [41, 43].

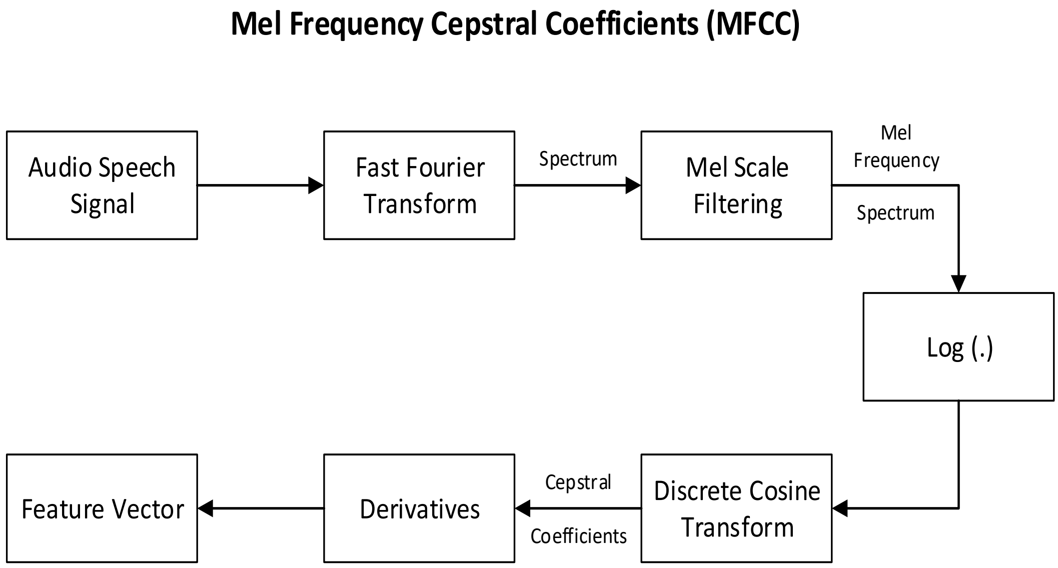

Figure 3.1: MFCC algorithm block diagram

3.2.1 Fourier Transform

The initial action step is the computation of frequencies of input audio signals which is accomplished by calculating the Discrete Fourier Transform. It is defined in Equation 3.2 as:

$$C_{T,k}^{(1)} = \left| \frac{1}{N} \sum_{j=0}^{N-1} f_j exp\left[-\iota 2\pi \frac{jk}{N} \right] \right| \qquad k = 0, 1, \ldots, (N/2) - 1 \qquad (3.2)$$

where N is the number of sampling points within a speech frame and the time frame T.

3.2.2 Mel-Frequency Spectrum

The next step is calculation of Mel-frequency spectrum. The spectrum can be clarified with various band-pass filters and then the power of every frequency band is calculated. This filtering process mimics the human speech as, human acoustic system practices the power above that of frequency band as signals for processing. Generally the filter set with the band-pass strainer cannot imitate the auricle as similar to human ear which custom any frequency as central frequency. For any ASR filters, Mel-scale is cast-off. The scale is basically a non-linear measurer which is constrated, on the idea of non-linear pitch sensitivity similar to human hearing system. It is defined in

Equation 3.3 as:

$$C_{T,j}^{(2)} = \sum_{k=0}^{N/2-1} d_{j,k} C_{T,k}^{(1)} \qquad j = 0, 1, \ldots, N_d \tag{3.3}$$

where d is the magnitude of the band-pass filters having index and frequency k.

3.2.3 Logarithm

This step compute the log of audio signal, in order to copy or mimic the human discernment of sound intensity as investigations indicated that human recognize sound intensityy on logarithmic scales. It can be given in Equation 3.4 as:

$$C_{T,j}^{(3)} = \log(C_{T,j}^{(2)}) \qquad j = 0, 1, \ldots, N_d \tag{3.4}$$

3.2.4 Cepstral Coefficients

In this step, speaker-dependent attributes are excluded by computing the cepstral coefficients. It is identified through literature, that the audio signals are the density of utterer dependent foundation signal and filter signal. Hence cestrum is computed to conquer the source signal. The inverse transformation of high cepstral coefficients illustrated the frequencies spectrum of the basis signal and the inverse transformation of the low cepstral coefficients demonstrated the frequencies respond of that particular speakers vocal tract. Inverse of Fourier Transform can be computed using Equation 3.5 as:

$$F^{-1}\{\log(F\{f_n\})\} \tag{3.5}$$

where f is the input signals and F represented Fourier Transformation. Other then Fourier Transform, discrete cosine transform can also be used because the absolute and fixed estimate of spectrum, as respect to the periodic continuation of the signals, is symmetric and real. The cepstral coefficients are calculated in following Equation 3.6 by:

$$C_{T,j}^{(4)} = \sum_{j=1}^{N_d} C_{T,j}^{(3)} \cos\left[\frac{k(2j-1)\pi}{2N_d}\right] \qquad k = 0, 1, \ldots, N_{mc} < N_d \tag{3.6}$$

where N_{mc} represent the number of selected cepstral coefficients which are used further processing.

3.2.5 Derivatives

The preceding steps contains information or finest data about audio signal frames. In order to symbolize the dynamic and vibrant nature of audio, first and second order derivatives for cepstral coefficients are computed by encompasses the feature set. It can be computed using Equation 3.7 as:

$$C_T = \left[C_{T,j}^{(4)}, \Delta C_{T,j}^{(4)}, \Delta\Delta C_{T,j}^{(4)}\right] \tag{3.7}$$

where $\Delta C_{T,j}^{(4)} = C_{T+1,j}^{(4)} - C_{T-1,j}^{(4)}$ and $\Delta\Delta C_{T,j}^{(4)} = \Delta C_{T+1,j}^{(4)} - \Delta s C_{T-1,j}^{(4)}$.

3.3 Linear Predictive Coding

LPC is known as one of the supreme and influential audio analysis procedures which is helpful for encoding quality speech signals at low-bit rate. The general principle behind Linear Predictive Coding (LPC) is that the speech illustration ccooul be estimated as a linear union or combination of previous audio illustrations. The idea behind tthiiss tecchnique is to minimize the sum of squared differences between the actual speech signals and that of estimated speech signals during a short interval of time [39]. In LPC, audio signals are examined with the support of formant approximation. The formant effects are reduced from that of audio signals, then frequency and strength is calculated for extra background humming sounds [39]. Audio signal illustrations are procured as linear fusion of LPC prior illustrations and the calculation accomplishment is recognized as direct or straight predictor hence it is known as Linear Predictive Coding.

3.3.1 Cepstrum computation

Cepstrum is calculated for LPC. Cepstrum is fictional parallel to that of auto-correlation order, which is calculated by power spectrum by applying Wiener-Khinchin theorem [44]. It is defined as follow, $x(n)$ is characterized as time domain distinct and discrete signals, where n represents index moreover $X(k)$ is said to be as complex spectrum where $k = 1, 2, ...N$ whereas number of samples are represented as N. $A(n)$ means auto-correlation sequence of $x(n)$. $P(k)$ is power spectrum of $x(n)$. For complex spectrum value, (DFT) of $x(n)$ is calculated by applying Equation 3.8 as:

$$DFT(x(n)) \rightarrow X(k) \tag{3.8}$$

$$X(k) = \sum_{n=1}^{N} x(n)e^{-i2\pi kn/N} \tag{3.9}$$

where N is sample number and $k = 0 \ldots\ldots N$. $e^{-i\theta}$ can be expanded using Equation 3.10

$$e^{-i\theta} = \cos\theta - i\sin\theta \tag{3.10}$$

where $\theta = 2\pi kn/N$. Equation 3.9 can be inscribed as Equation 3.11 and Equation 3.12:

$$X(k) = \sum_{n=1}^{N} x(n)(\cos(2\pi kn/N) - i\sin(2\pi kn/N)) \tag{3.11}$$

$$X(k) = \sum_{n=1}^{N} x(n)\cos\left(\frac{2\pi kn}{N}\right) - i\sum_{n=1}^{N} x(n)\sin\left(\frac{2\pi kn}{N}\right) \tag{3.12}$$

Inverse of DFT gives the value of $x(n)$ using Equation 3.13 as:

$$IDFT(X(k)) \rightarrow x(n) \tag{3.13}$$

The $x(n)$ obtained from inverse of DFT can be represented as shown in Equation 3.14

$$x(n) = \frac{1}{N}\sum_{n=1}^{N} X(k)e^{i2\pi kn/N} \tag{3.14}$$

Power spectrum ($P(k)$) is obtained by taking the square root of the absolute of the signal domain's ($x(n)$) DFT as shown in Equation 3.15.

$$|DFT(x(n))|^2 \rightarrow P(k) \tag{3.15}$$

P(k) is said to be as the power spectrum of frame. Periodogram spectral approximate holds statistics that is not desired for audio recognition hence this information is discarded. Clump of periodogram bins support to find out the appraisal of energy exist in various frequency zones. Then Complex Fourier transform complete and absolute value is calculated and end product is squared. Normally 512 opinion for Fast Fourier transform (FFT) are executed and out of these first 257 coefficients are stored. Periodogram based audio frame power spectral approximation is calculated in Equation 3.16 as:

$$p(k) = \frac{1}{N}|S(k)|^2 \tag{3.16}$$

where $S(k)$ is DFT value and N is sample number. Complex Discrete Fourier Transform (DFT) provides $S_i(k)$, where i signifies the number of frame equivalent. DFT of frame is computed by multiplying framed signal with that of hamming window. It is achieved by using Equation 3.17 as shown below:

$$S_i(k) = \sum_{n=1}^{N} s_i(n)w(n)e^{-j2\pi kn/N} \qquad 1 \leq k \leq K \tag{3.17}$$

where $w(n)$ is Hamming window, K is the DFT length, analysis window $s(n)$ is framed signal and $n = 1 \dots N$. Auto-correlation order is accomplished by captivating the IDFT of the power spectrum $(P(k))$ by using Equation 3.18 as:

$$IDFT(P(k)) \rightarrow A(n) \tag{3.18}$$

Correlation has the consistency with in the signals. When x and y are parallel, then $x(i)$ is positive and vice versa. Signals with same attributes achieve high connection; whereas signals having different attributes attain low connection. Signals having both resemblances and variances at identical level accomplish a correlation score somewhere in between. Signals are said to be analogous when there is the great number of negative correlations exist [44]. In that case one signal is overturned with respect to other signal. It is symbolized in Equation 3.19 as:

$$A(n) = \sum_{i=0}^{N} x(i)y(i) \tag{3.19}$$

In the end by taking the power spectrum logarithm earlier the IDFT, Cepstrum is attained by:

$$IDFT(\log(P(k))) \rightarrow C(n) \tag{3.20}$$

Generally, Cepstrum is defined as an auto-correlation arrangement of compressed logarithm. From the power spectrum of log alternate to the standard power spectrum, Cepstrum is derived. It is shown in Equation 3.21 as:

$$c_x(n) = \left(\frac{1}{2\pi}\right) \int_{-\pi}^{\pi} ln(|X(k)|)e^{jwn} d\omega \tag{3.21}$$

where $X(k)$ is the signified as a Fourier transform of the sequence $x(n)$.

3.4 Linear Prediction Cepstral Coefficient

Linear Prediction Cepstral Coefficients (LPCCs) are calculated in two methods, first is the similar technique as Cepstral and other technique is computation from LPC. The only variation in LPCC is computed by means of interruption free power spectrum whereas Cepstral is calculated from that of periodogram approximation of power spectrum [44]. For computing LPC, Auto-correlation coefficients are calculated. Figure 3.2 shows a block diagram for LPCC feature extraction [45]. LPCC feature extraction procedure is enlightened in following steps:

3.4.1 Input Signal

The input audio signal is digitized spectrally then the flatten speech signal is placed through a low order digital system in order to make it less susceptible to finite precision effects during signal processing.

3.4.2 Frame Blocking

To calculate LPC features, primarily the audio signal is blocked into N number of frame samples. The production of the pre-emphasizer system is associated to the input to the system.

3.4.3 Windowing

The next step is to box each respective individual frame in order to minimize the signal discontinuities in the start and end of each frame. Usually, this boxing is said to be as Hamming window.

3.4.4 Auto-correlation Analysis

In this step each frame of windowed signal is auto-correlate, and the maximum autocorrelation value is the command of that of LPC analysis.

3.4.5 LPC Analysis

The next phase is LPC exploration, that changes each and every frame of auto-correlations into a finest set of LPC parameter by the mean of Durbin's technique. .LPC cepstral coefficients, is said to be a very essential set of LPC parameter, that could be derived directly from the existing set of LPC coefficient. Linear prediction is a classical model which is works on the basis of human audio process. It follows a conventional source-filter illustration, in which the glottal, lips radiation and vocal tract transfer functions are combined together into a delicate filter that mimics the auditory of speakers vocal tract. LP is a mathematical procedure which delivers the approximation of the existing samples of discrete signals in a linear combination of numerous preceding samples. The norm behind the LPC process is to reduce the aggregate of the squared differences between the original audio signals and that of estimated audio signal over a determinate time which gives an exclusive set of predictor co-efficient. These predictor coefficients valued each frame, which is usually $20ms$ long. The transfer function of time varying digital filter is given in Equation 3.22 as follows:

$$H(z) = \frac{G}{1 - \sum_{a}^{k} z - k} \tag{3.22}$$

Where $k = 1top$, which will be 10 for the LPC-10 algorithm and 18 for the improved algorithm. The predictor coefficients are symbolized by ak and gain (G) is another central parameter [38, 46].

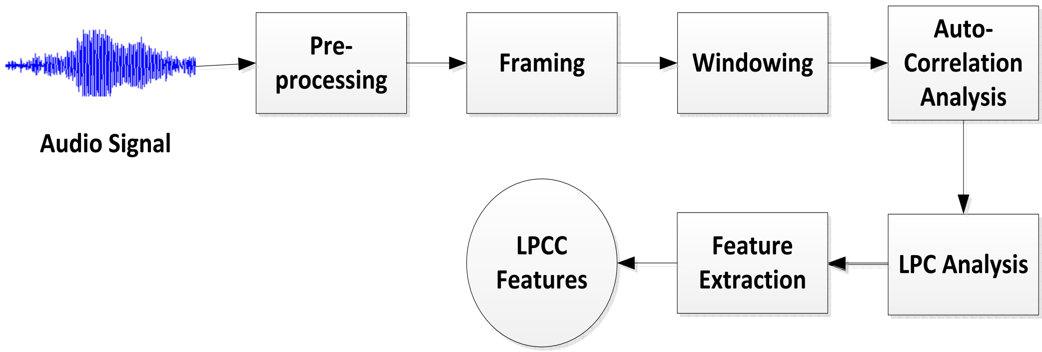

Figure 3.2: LPCC feature extraction process block diagram

3.5 GMM-UBM

Over the last few years, the Gaussian Mixture model (GMM) has become a well-known standard classifier for speaker's emotion recognition. Gaussian Mixture model (GMM) is frequently used for emotion verification because this approach has decent ability of verification, appreciation and recognition. Generally, when there is no prior knowledge about what the speaker will says, then in this case most successful likelihood function will be GMM because this method is used for generative unorganized clustering of learning data. It is also known as Expectation Maximization (EM) clustering based optimization scheme. The idea behind this is very simple that it is for given data set in which each point is originated by the combination of multiple Gaussians.

> **Gaussian is a type of mathematically convenient distribution which is a listing of outcomes of an experiment and the probability is associated with each outcome.**

It is an even distribution through which half of the data falls right and half of the data falls left of it. In order to understand, Gaussian distribution, two variables are required namely mean and standard deviation.

i. Mean (which is the average of all the data points and finds center of the curve).

ii. Standard deviation (which describes how spread out the data is).

Sometime data have multiple peaks or distributions then in that case the Gaussian function, of a vector ’x’, having ’d’ dimensions will be as following:

$$b_i(x) = 1/(2\pi)^D/2|\sum_i|1/2\exp -1/2(x-u_i)'\sum_i^{-}1(x-u_i) \tag{3.23}$$

Where covariance represent the measurements indicating that how the variation in one variable are interconnected with the variation in other variable and it also exactly measure the degree to which variables are linearly correlated and associated with one another. The comprehensive GMM is specified by covariance matrices, mean vectors and fusion of weight from all constituent densities [47, 48].

Maximum Likelihood Parameter Estimation: For GMM configuration, we have to evaluate the constraints of GMM, the finest matches, for the distribution of training feature vectors. The supreme and recognized technique is ”Maximum Likelihood Estimation” which is also known as ML Estimation. The central purpose of ML approximation is to discover the perfect parameters, that will maximize/increase the likelihood, feasibility and probability of GMM. One of the most standard procedure to maximize the likelihood of GMM is to employ the Expectation-Maximization (EM) algorithm. The simple knowledge behind the EM algorithm is, starting with an original model and then evaluate a new model. This new and up to date model then becomes the initial model for the succeeding iteration and the progression is repeated until some conjunction threshold is grasped.

GMM with a universal background model: UBM is an enhancement in the Gaussian mixture modell technique and one of the most influential quality of GMM-UBM is its ability to formulate a smooth calculations of randomly shaped distributions [47]. GMM-UBM have distinctive supremacy over other modeling methods because its training is relatively agile and models can be scaled and updated easily, even by adding new data. Universal Background Model, intention is to model the inclusive data distribution which will be composed of 100 of Gaussians [49]. For a D-dimensional features vector ”x”, the mix density which is used for the probability function is stated as a subjective sum of Gaussian models and given in Equation 3.24

as follows:

$$Ssp(x|\lambda) = \sum_{i=1}^{M} \omega_i p_i(x) \tag{3.24}$$

Where M is the number of Gaussian components. The UBM is accomplished using an enormous and descriptive set of data by employing Expectation Maximization (EM) algorithm [49]. This EM algorithm is utilized in order to train the models, correspondingly in case of adaptation base method, which are trained by maximizing the posteriori estimation (MAP). This algorithm is a 2 step procedure:

1. An information about the constraints are obligatory to adapt UBM in order to estimate the current existing class.
2. New information about the constraints is mixed up with thar of old constraints and then by applying data-dependent mixing co-efficient the models of UBM are updated.

Due to closse-fitting association between UBM and that of trained models the overall performance of UBM rooted techniques are better than that of decoupled methods. Due to implication of close-fitted association method, the performance is not affected by unseen auditory activities as whenever an unseen auditory incident arises the mixture constraints of that particular unseen auditory class are straightly derivatives from UBM which means during testing phase, the unseen audotory episode produces nearly zero likelihood ratio [48].

3.6 Summary

This chapter gives a detailed review on feature extraction techniques which includes MFCC, LPCC and LPC. It also covers a comprehensive discussion on a newest and finest classifier which is used for emotion classification named as GMM-UBM.

CHAPTER 4

METHODOLOGY

4.1 Overview

This chapter demonstrate the proposed method for emotion recognition from audio signals. The suggested method is constructed by using MFCC feature extraction technique and for the classification purpose GMM-UBM classifier is used.

4.2 Proposed Method

The block diagram of suggested method is shown illustrated in Figure 4.1

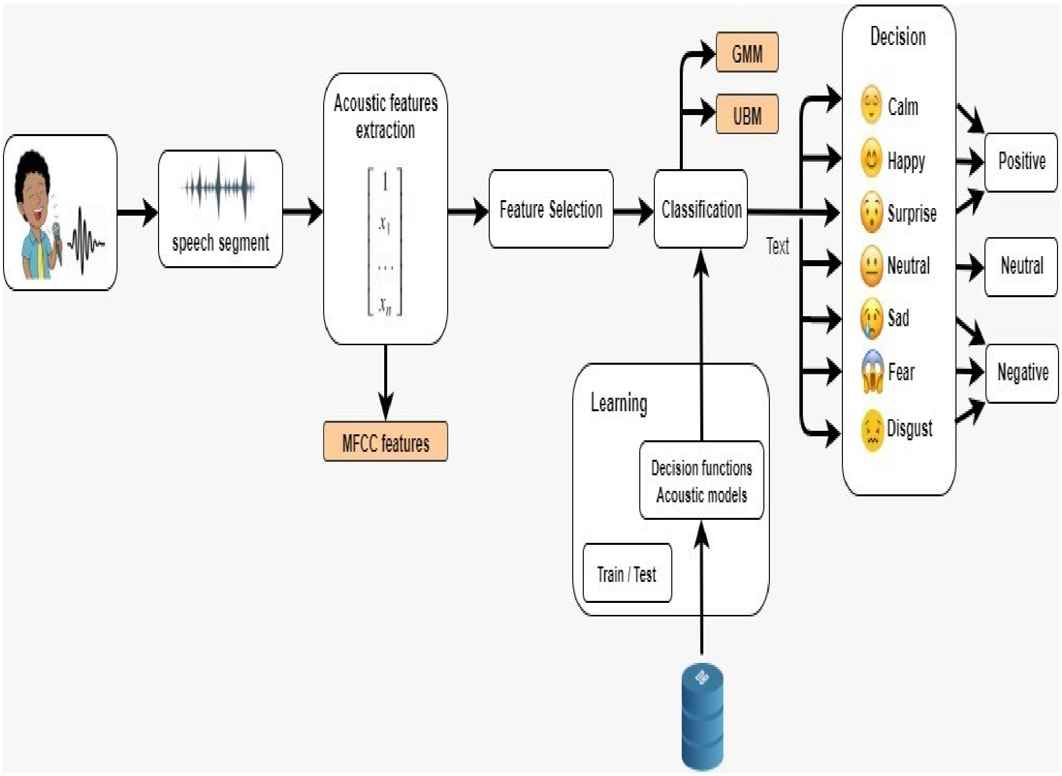

Figure 4.1: Effect of Emotions on Software

First of all, an Emotion based dataset is assembled. A feature extraction technique is applied in order to decrease the amount of resources which are required for processing without losing relevant information. The application of feature extraction technique in dataset reduces the aggregate of redundant data and help in dimensionality reduction so, it became easier to visualize the data and this method help to reduce the time and storage space as well.

In proposed method MFCC feature extraction technique is used. These MFCC features are extracted in such a way that the sampling rate of all audio recordings are kept fixed at $16000Hz$ or $16kHz$. The dimension of MFCC feature vector is 13. Using this configuration, MFCC features for train and test set have been calculated. For classification, GMM-UBM is used to identify emotional classes. The training of GMM-UBM is achieved by various mixture components i.e. $32, 64, 128, 256$ and 512. The result shows that, MFCC features perform better as compared to other features extraction methods on emotion based dataset. The datasets used for this research are TESS, RAVDEES and SREAD.

4.3 Speech segmentation

The constructed audio conversation of requirement gathering and project discussing was about 2-3 minutes long. As audio signals are not static even in inclusive logic, hence it is common in audio processing to split an audio signal into small fragments called frames, so the given audio is segmented into small frames of 3 seconds and within each interval the signal is considered to be static.

4.4 Pre processing

The pre-processing step was conducted for speech stabilization, pre-emphasis ?ltering, normalization and removal of stillness intervals, then these silence intervals are detached by employing logarithmic method for splitting, segmenting and separating speech signals from noisy background.

4.5 Feature extraction

To optimize the classification accuracy and time, feature extraction method is used. As it is crucial to minimize the amount of features, which helps in the reduction of required time to complete recognition process and to limit the computational complexity. In proposed method features are extracted from the audio files by using MFCC feature abstraction technique.

4.6 Classification

For classification of emotion with in an audio conversation GMM-UBM classifier is trained. 70% of the audios are used for training and 30% are used for testing. The emotions in the constructed databases were angry, happy, calm, surprise, sad, disgust and neutral. These emotions are divided into three categories positive, neutral and negative. Out of above emotion emotions happy, calm and surprised were categorized as positive. Similarly angry, sad and disgust were categorized as negative emotion and neutral remains neutral.

4.7 Summary

This chapter sums up the methodology used for emotion recognition from audio signals. It also gives a brief review of proposed method used for feature extraction and how classifier is applied on different databases for emotion classification. It describes the process of constrution of audio databases. The constructed conversation databases are then divided into two datasets that are testing dataset and training dataset. These datasats are then delivered though MFCC for feature extraction. The extracted features are trained on classifier and tested audio recordings are verified through trained features for outputs.

CHAPTER 5

EXPERIMENTAL SETUP AND RESULTS

5.1 Overview

This chapter presents all the emotion recognition results of three datasets having different actors, statements, intensity and accents. It concisely discuss the supporting datasets which are used for emotion recognition from audio conversation. It also illustrates the performance evaluation of the proposed system in term of accuracy and equal error rate.

5.2 TESS Audio Dataset

Toronto Emotional Speech Set (TESS) is an emotion based dataset that is transferred from TSpace which is a free and protected research repository established by university of Toronto in order to preserve and publicize the scholarly data of university. A set of 200 objective words were articulated in the phrase like ”Say the word ______” by two performers (aged 26 and 64 years respectively). The audio recordings were constructed in a set, depicting seven emotions which includes happiness, anger, sadness, fear, disgust,neutral and pleasant surprise. There are total 2800 stimuli in this particular dataset that is 2 actors x 200 lines x 7 emotions = $2,800$ audio files in total. The dataset is structured in such a way that each of the two female performer and their targeted emotions are comprehend within its specific file in which 200 target audio records can be found. The format of the audio record is WAV format. The motivation behind the assembly of this sort of dataset is to implement speaker and text independent emotion recognition. The dataset is randomly divided into two disjoint datasets such as the training set and the testing set. The training set contains 80% audio files that is the total number of files in training set are 2240 whereas, test set contain 20% files which are about 560. Table 5.1 sum up the TESS emotional speech corpus

Table 5.1: TESS Emotional audio corpus

Emotions	Number of Actors per Category	Number of Samples per Actor	Training Samples per Actor	Test Samples per Actor	Nature of Speech Samples
Angry	2	200	160	40	Speaker and text independent
Disgust	2	200	160	40	
Fear	2	200	160	40	
Happy	2	200	160	40	
Neutral	2	200	160	40	
Sad	2	200	160	40	
Surprise	2	200	160	40	

5.2.1 Confusion Matrix for TESS database

It is a matrix which evaluate the overall performance of a specific classifier on test dataset. It is generally identified as summary of prediction. The total numbers of accurate and inaccurate predictions are calculated category wise. Every column of the matrix is related to an actual category whereas every row of the confusion matrix is related to that of predicted category. Every computation of precise and imprecise classification are filled into a table. The entire values of exact likelihood for a category falls into that of estimated row for that category value and the predicted column for that category value. Similarly, the entire number of imprecise likelihood for a category falls into the estimated row for that category and the predicted column for that category value. Basic expressions used in a confusion matrix are:

i. True positive (TP): A predicted emotion is similar as actual emotion.

ii. True negative (TN): A predicted emotion is false and the actual emotion is also false.

iii. False positive (FP): A predicted emotion is true however, actual emotion is false.

iv. False negative (FN): A predicted emotion is false however, actual emotion is true.

Table 5.2: TESS dataset confusion matrix for 32 and 64 mixtures

Emotion	Positive	Neutral	Negative
Positive	119	1	0
Neutral	0	40	0
Negative	0	0	120

All the diagonal values demonstration in Table 5.2 are true positive means the predicted emotion is same as that of actual emotion, whereas all the remaining values are true negatives (TN), false positives (FP) and false negatives (FN). The diagonal values are TP, therefore $119 + 40 + 120 = 279$ hence accuracy is equal to (TP)/total = $(279)/280 = 99.64\%$. Accuracy is articulated in percentage.

Table 5.3: TESS dataset confusion matrix for 128, 256 and 512 mixtures

Emotion	**Positive**	**Neutral**	**Negative**
Positive	120	0	0
Neutral	0	40	0
Negative	0	0	120

All the diagonal values demonstration in Table 5.3 are true positive means the predicted emotion is same as that of actual emotion, whereas all the remaining values are true negatives (TN), false positives (FP) and false negatives (FN). The diagonal values are TP, therefore $120 + 40 + 120 = 280$ hence accuracy is equal to (TP)/total = $(280)/280 = 100\%$. Accuracy is articulated in percentage.

5.3 RAVDEES Emotional Speech Dataset

The RAVDESS is a corroborated multimodal emotional database having speech and song audios. This set of audio recordings were estimated by 319 entities having characteristic of untrained research contributors of America. RAVDEES stands for (Ryerson Audio-Visual Database for Emotional Speech and Songs) which holds 7356 files and the total size is about 24.8 GB. . This set of 7356 audio files can freely be downloaded. The dataset covers 24 professional actors in which 12 are female and 12 are male. These 24 actors vocalizing two lexically coordinated sentences in American accent. Audio contains sad, angry, calm, surprise, happy, fearful, and disgust expressions. Each emotion is produced in two attributes of emotional intensity such as normal and powerful/strong with a further neutral communication. Audio-files of all speakers $(01-24)$ are easily accessible and freely available in a separate zip file which is about 215 MB in size. Emotional Audio file (Speech_Audio_Speakers_01-24.zip, of 215 MB) holds 1440 audio files having 60 trials per speaker x 24 audios $= 1440$. Each of the 7356 audio set of RAVDESS database has a distinctive filename which contains seven-part statistical identifier such as $03-01-01-01-01-01-01.mp4$. These identifiers express the incentive properties:

Table 5.4: Statistical identifier for audio file

Modality (01 = full-AV, 02 = video, 03 = audio).
Vocal form (01 = audio, 02 = songs).
Emotion (01 = neutral, 02 = calmness, 03 = happiness, 04 = sadness, 05 = panic, 06 = fear, 07 = disgust, 08 = happily surprised).
Emotion intensity (01 = normal, 02 = powerful/loud).
Declaration (01 = ”Kids are talking by the door”, 02 = ”Dogs are sitting by the door”).
Repetitions (01 = 1st , 02 = 2nd).
Speakers (01 to 24, even numbered speakers are female and Odd numbered speakers are male).

For research experiment, we constructed a dataset from RAVDEES having following characteristics: emotions in constructed dataset are angry, calm, disgust, happy, neutral, sad and surprise. Total samples in each emotions are 120, train samples in each emotion are 90, and test samples in each emotions are 30. There are 15 speakers in each emotion category and each speaker has 8 Recordings. Table 5.5 summarizes the TESS emotional speech corpus

Table 5.5: RAVDESS Emotional speech database

Emotions	**Number of Actors per Category**	**Number of Samples per Actor**	**Training Samples per Actor**	**Test Samples per Actor**	**Nature of Audio Samples**
Angry	15	8	90	30	Speaker and text independent
Disgust	15	8	90	30	
Fear	15	8	90	30	
Happy	15	8	90	30	
Neutral	15	8	90	30	
Sad	15	8	90	30	
Surprise	15	8	90	30	

5.3.1 Confusion Matrix for RAVDEES database

All confusion matrix below are said to be as a summary of predicted results which are based on proposed classification technique. The number of precise and imprecise predictions are sum up with total values which are broken down by respective category. The confusion matrix gives a means in which proposed classification technique is confused, when it creates predictions. It also provides understanding not about the errors existence which may constructed by a classifier but more significantly the nature of errors that are being made. The investigation on RAVDEES database is made by using MFCC feature extraction technique and the training of GMM-UBM is attained by various mixture components i.e. $32, 64, 128, 256$ and 512. The results are as following:

Table 5.6: RAVDEES dataset confusion matrix for 32 mixtures

Emotion	**Positive**	**Neutral**	**Negative**
Positive	75	4	11
Neutral	3	22	5
Negative	18	6	66

Values demonstrated in Table 5.6 in diagonal shows true positive (TP). Therefore TP $= 75 + 22 + 66 = 163$. Whereas other values are either TN, FP or FN. Accuracy for 32 mixture component is equal to (TP)/total $= (163)/210 = 77.61\%$.

Table 5.7: RAVDEES dataset confusion matrix for 64 mixtures

Emotion	**Positive**	**Neutral**	**Negative**
Positive	74	4	12
Neutral	1	25	4
Negative	16	1	73

Values demonstrated in Table 5.7 in diagonal shows true positive (TP). Therefore TP $= 74 + 25 + 73 = 172$. Whereas other values are either TN, FP or FN. Accuracy for 64 mixture component is equal to (TP)/total $= (172)/210 = 81.90\%$.

Table 5.8: RAVDEES dataset confusion matrix for 128 mixtures

Emotion	**Positive**	**Neutral**	**Negative**
Positive	75	6	9
Neutral	1	29	0
Negative	13	1	76

Values demonstrated in Table 5.8 in diagonal shows true positive (TP). Therefore TP $= 75 + 29 + 76 = 180$. Whereas other values are either TN, FP or FN.

Accuracy for 128 mixture component is equal to (TP)/total $= (180)/210 = 85.71\%$.

Table 5.9: RAVDEES dataset confusion matrix for 256 mixtures

Emotion	**Positive**	**Neutral**	**Negative**
Positive	78	4	8
Neutral	0	30	0
Negative	14	1	75

Values demonstrated in Table 5.9 in diagonal shows true positive (TP). Therefore TP $= 78 + 30 + 75 = 183$. Whereas other values are either TN, FP or FN. Accuracy for 256 mixture component is equal to (TP)/total $= (183)/210 = 87.14\%$.

Table 5.10: RAVDEES dataset confusion matrix for 512 mixtures

Emotion	**Positive**	**Neutral**	**Negative**
Positive	82	2	6
Neutral	0	30	0
Negative	14	0	76

Values demonstrated in Table 5.10 in diagonal shows true positive (TP). Therefore TP $= 82 + 30 + 76 = 188$. Whereas other values are either TN, FP or FN. Accuracy for 512 mixture component is equal to (TP)/total $= (188)/210 = 89.52\%$.

5.4 Construction of SREAD (Software Requirement Engineering Audio Database)

SREAD database is constructed by real requirement gathering conversations between software developers and client. There are ten audio conversation having three emotional categories positive, negative and neutral. The duration of these audio conversations are from 2-3 minutes. First of all the requirement gathering conversation between software developer and client are noted down, then 2 speaker record these conversations by expressing motioned emotional categories against each sentence. These audio files are grouped together in order to built SREAD database. The ground truth for each audio is calculated. Each audio in SREAD is then slpited into small chunks for 3 sec duration and three emotional categories (positive, negative and neutral) are labelled against each chuck. SREAD database is first trained by supporting TESS database and then by RAVDEES database.

5.4.1 Confusion Matrix for SREAD database

The investigation on SREAD database is made by using MFCC feature extraction technique and the training of GMM-UBM by using TESS database and then by RAVDEES database and is attained by various mixture components i.e. $32, 64, 128, 256$ and 512. Results demonstrated in Table 5.11 and Table 5.12 are obtained by training GMM-UBM model with TESS dataset. And results of Table 5.13 and Table 5.14 are obtained by training GMM-UBM model with RAVDEES dataset.

Table 5.11: SREAD dataset confusion matrix of test file 05 for 32 and 64 mixtures using TESS based GMM-UBM

Emotion	**Positive**	**Neutral**	**Negative**
Positive	1	3	0
Neutral	0	25	0
Negative	0	1	1

Values depicted in diagonal shows true positive (TP). Therefore TP $= 1 + 25 + 1 = 27$. Whereas other values are either TN, FP or FN. Accuracy for 32 and 64 mixture component is equal to (TP)/total $= (27)/31 = 87.09\%$.

Table 5.12: SREAD dataset confusion matrix of test file 05 for 128, 256 and 512 mixtures using TESS based GMM-UBM

Emotion	**Positive**	**Neutral**	**Negative**
Positive	1	3	0
Neutral	0	25	0
Negative	0	0	2

Values depicted in diagonal shows true positive (TP). Therefore TP $= 1 + 25 + 2 = 28$. Whereas other values are either TN, FP or FN. Accuracy for 128, 256 and 512 mixture component is equal to (TP)/total $= (28)/31 = 93.33\%$.

Table 5.13: SREAD dataset confusion matrix of test file 05 for 32 and 64 mixtures using RAVDEES based GMM-UBM

Emotion	**Positive**	**Neutral**	**Negative**
Positive	0	1	3
Neutral	0	23	2
Negative	0	0	2

Values depicted in diagonal shows true positive (TP). Therefore TP $= 0+23+2 = 25$. Whereas other values are either TN, FP or FN. Accuracy for 32 and 64 mixture component is equal to (TP)/total $= (25)/31 = 80.64\%$.

Table 5.14: SREAD dataset confusion matrix of test file 05 for 128, 256 and 512 mixtures using RAVDEES based GMM-UBM

Emotion	**Positive**	**Neutral**	**Negative**
Positive	0	0	4
Neutral	0	24	1
Negative	0	0	2

Values depicted in diagonal shows true positive (TP). Therefore TP $= 0+24+2 = 26$. Whereas other values are either TN, FP or FN. Accuracy for 128, 256 and 512 mixture component is equal to (TP)/total $= (26)/31 = 83.87\%$.

5.5 Equal Error Rate Comparison

Equal error rate (EER) is a technique which helps to select the threshold values for false rejection and false acceptance rate. Whenever the rates are corresponding, then the common value is mentioned as that of *equal error rate*. The value identifies that the percentage of false acceptances rate is equivalent to the percentage of false rejections rate. It means that when the value of equal error rate is low, then the accuracy of that particular system is said to be as higher. The system's accuracy and equal error rate is described in detailed. The equal error rate of TESS and RAVDEES is show in Table 5.15. While Table 5.16 and Table 5.17 shows the equal error rate comparison for all test files of SREAD dataset.

Table 5.15: Equal Error Rate based comparison between databases using GMM-UBM with various mixture components for emotion recognition

GMM-UBM components	32	64	128	256	512
TESS	0.36	0.36	0	0	0
RAVDEES	22.39	18.1	14.29	12.86	10.48

Table 5.16: Equal Error Rate based comparison of SREAD dataset using TESS based GMM-UBM with various mixture components

GMM-UBM components	32	64	128	256	512
Test File 01	28.57	24.49	24.49	22.45	20.41
Test File 02	25.00	21.43	21.43	21.43	21.43
Test File 03	28.57	32.14	28.57	28.57	25.00
Test File 04	20	16.67	16.67	16.67	16.67
Test File 05	12.91	12.91	6.67	6.67	6.67
Test File 06	32.14	28.57	28.57	25.00	25.00
Test File 07	23.33	20.00	20.00	20.00	20.00
Test File 08	25.00	25.00	25.00	25.00	25.00
Test File 09	26.47	23.53	23.53	17.65	14.71
Test File 10	8.11	8.11	5.41	5.41	2.71

Table 5.17: Equal Error Rate based comparison of SREAD dataset using RAVDEES based GMM-UBM with various mixture components

GMM-UBM components	32	64	128	256	512
Test File 01	34.69	32.66	30.62	28.57	24.49
Test File 02	28.57	28.57	25.00	21.43	21.43
Test File 03	28.57	28.57	28.57	25.00	25.00
Test File 04	30.00	30.00	26.67	23.33	23.33
Test File 05	19.36	19.36	16.13	16.13	16.13
Test File 06	39.29	32.14	28.57	25.00	25.00
Test File 07	33.33	33.33	30.00	23.33	23.33
Test File 08	37.5	37.50	37.50	34.37	31.25
Test File 09	38.24	32.35	32.35	29.42	29.42
Test File 10	35.14	35.14	32.43	29.73	24.33

5.6 Accuracy Comparison

Accuracy is machine learning is said to be as one metric for evaluating and weighing the classification simulations. Generally, it is the fraction of predictions our simulation of the classification model that got right. Accuracy of the classification model is calculated as: Accuracy = Number of true predictions / Total number of predictions. The accuracy comparison of proposed method for TESS and RAVDEES are given the Table 5.18. While Table 5.19 and Table 5.20 shows the equal error rate comparison for all test files of SREAD dataset.

Table 5.18: Accuracy (%) based comparison between different databases by applying GMM-UBM with arious mixture components for emotion recognition

GMM-UBM components	**32**	**64**	**128**	**256**	**512**
TESS	99.64	99.64	100	100	100
RAVDEES	77.61	81.90	85.71	87.14	89.52

Table 5.19: Accuracy (%) based comparison of SREAD dataset using TESS based GMM-UBM with various mixture components

GMM-UBM components	**32**	**64**	**128**	**256**	**512**
Test File 01	71.43	75.51	75.51	77.55	79.59
Test File 02	75.00	78.57	78.57	78.57	78.57
Test File 03	71.43	67.86	71.43	71.43	75.00
Test File 04	80.00	83.33	83.33	83.33	83.33
Test File 05	87.09	87.09	93.33	93.33	93.33
Test File 06	67.86	71.43	71.43	75.00	75.00
Test File 07	76.67	80.00	80.00	80.00	80.00
Test File 08	75.00	75.00	75.00	75.00	75.00
Test File 09	73.53	76.47	76.47	82.35	85.29
Test File 10	91.89	91.89	94.59	94.59	97.29

Table 5.20: Accuracy (%) based comparison of SREAD dataset using RAVDEES based GMM-UBM with various mixture components

GMM-UBM components	**32**	**64**	**128**	**256**	**512**
Test File 01	65.31	67.34	69.38	71.43	75.51
Test File 02	71.43	71.43	75.00	78.57	78.57
Test File 03	71.43	71.43	71.43	75.00	75.00
Test File 04	70.00	70.00	73.33	76.67	76.67
Test File 05	80.64	80.64	83.87	83.87	83.87
Test File 06	60.71	67.86	71.43	75.00	75.00
Test File 07	66.67	66.67	70.00	76.67	76.67
Test File 08	62.50	62.50	62.50	65.63	68.75
Test File 09	61.76	67.65	67.65	70.58	70.58
Test File 10	64.86	64.86	67.57	70.27	75.67

5.7 Modelling of Emotions in Requirement Engineering

The manner in which people feel about tools and technology can define whether the technology is adapted or rejected by its expected users. People's moods, feelings and emotions are mainly significant for the acceptance of any socio-technical system which involves social behavior. Software engineering procedure usually focused on the functional and non-functional requirements of a system. The emotional requirements of customers (i.e. the feelings that characterize their state of mind) in relation to the adaptation of a system are often ignored. In a result systems does not fulfil the needs of their customers. This result suggest that one tactic to construct better systems is to explore the emotions of the users, in order to uncover hidden needs and requirements that may otherwise be ignored. As requirement engineering plan depends expressively on team performance because software project is shaped by people and for people that linking human cooperation.

This research shows that without having an ability to cultivate and grasp definite emotions, human's capability to cooperate with others is impossible. Research shows that people are more efficient and creative towards a problem solving, when they are happy [10, 21],as their is an influence of emotions on software productivity. People's emotion, positive or negative, not only affect the overall performance - containing creativity, decision making, teamwork, efficiency, teamwork, leadership abilities and negotiation - but also behavior of team members and customers. Positive emotions boost the creativity and enhance the ideas towards better solution whereas negative emotions bring frustration and anxiety.

Emotion in requirements engineering does have significant role. As we all know that some people are not very much expressive about what they are feeling and what they actually need. So through emotional requirement engineering, it became easier to gather the actually needs of the customers which otherwise can be neglected. Requirement engineering is all about logics, about measurements and that of digital words. Emotions and requirement engineering seems so different from one another but by bring them together it became superlative experience for consumer. Hence emotional requirement engineering is a methodology with which software developer can figure out what emotional needs and requirements does a customer want to put in a project which will result in the form of a perfect superlative experience and acceptance of the project.

Traditionally product design start with a look-at generally known as need or requirement analysis. In any requirement engineering the overall blue print of product's acceptance depends not only on rational hook but also on emotional hook. When we talk about the emotional requirement engineering, the engineering part has to be sorted that we cannot say that we can make substandard product and then connect to the emotional hook because that does not work. For better response from customer, the product has to be with the power of what are the requirements and then additionally has an emotional hook hence we can say that emotions are the very powerful tool in terms of being able to retain customers. Emotional requirement engineering is a process by which we consciously identify what emotion we want to trigger or identify in consumer and then design them into the product. The emotions can be predicted from different means basically asking from the customer, through audio/speech, through facial expression, through vital signals and through gesture. For our research we select to recognize emotions through audio signals because negotiation is one of the central aptitude in requirement engineering process, which is not only constructed on verbal announcements but also on the emotional hook.

5.8 Summary

This chapter shows all the findings and results of this research. This chapter accomplishes the data gathering, databases construction and clearly explains the databases used in this research. It describes how the data is collected, composed and then structured within the databases. A blueprint which describes the overall performance measures and the modeling of emotion in requirement engineering is also given in the end of this chapter.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Overview

This chapter summarizes the result of emotion recognition from audio signal for effective requirement engineering. It demonstrated the conclusion and recommendation.

6.2 Conclusion

This thesis investigated the emotion recognition from audio signal for effective requirement engineering for English language. We have also deliberated some important features extraction techniques along with the classi?ers used for the classification purpose by highlighting their properties. An innovative method for extraction of emotion during requirement gathering discussion, is presented. The emotions of a person is extracted from the audio conversation between client and the software engineer which contributes an opportunity that give people feedback on what they are feeling and what they are expressing during requirement gathering. Some of the information can be used for sensing the flexibility of emotions inside a person. For the implementation of proposed technique, through critical literature review on various audio features, MFCC feature extraction technique is selected due to its suitability and popularity, while GMM-UBM was selected as a classifier due to its higher accuracy. Comprehensive and inclusive emotion databases TESS, RAVDEES and self-constructed SERAD were used. Research includes implementation of the proposed method using MATLAB, performance evaluation and benchmarking. The vision behind this research is to show respect for the users or customer's feelings. The foremost objective of emotion recognition from audio signal, is to automatically pinpointing the emotional or physical state of client from his or her conversation during requirement gathering, project development and demonstration. The benefit of the emotion recognition is defined by the experimental results.

6.3 Future work

This thesis has some limitations such as it covers only English language. Other languages can be incorporated and the emotion recognition can be evaluated. In future, we will progress this proposed model to increase the competence, efficiency, and productivity, besides this, the investigation on the recognition of emotional intensity with in an audio, will be executed by exploration of different speech features. Classifier used in this research is also limited hence we will try to apply different classifiers to intensify the process. Also, we observe that some new dimensionality reduction techniques, pattern classification methods, hybrid classifier techniques are proposed recently. We recommend multiple classifier methods as a further research direction, which need to be explore in future for better results and productivity.

REFERENCES

[1] The Standish Group. The Standish group: the chaos report. *Project Smart*, 2014: 16.

[2] Colomo-Palacios, R., Hernández-López, A., García-Crespo, Á. and Soto-Acosta, P. A study of emotions in requirements engineering. *Communications in Computer and Information Science*, 2010. 112 CCIS(PART 2): 1–7.

[3] Colomo-Palacios, R., Casado-Lumbreras, C., Soto-Acosta, P. and García-Crespo, Á. Using the affect grid to measure emotions in software requirements engineering. 2011.

[4] Nuseibeh, B. and Easterbrook, S. Requirements engineering: a roadmap. 2000, vol. 1. 35–46.

[5] Zave, P. and Jackson, M. Four dark corners of requirements engineering. *ACM Transactions on Software Engineering and Methodolog*, 1997. 6(1): 1–30.

[6] Melegati, J., Goldman, A. and Paulo, S. Requirements Engineering in Software Startups : a Grounded Theory approach. *International Workshop on Software Startups, Trondheim, Norway*, 2016.

[7] Lauenroth, K. and Pohl, K. Towards automated consistency checks of product line requirements specifications. *IEEE/ACM International Conference on Automated Software Engineering*. 2007. 373–376.

[8] Cambria, E., Livingstone, A. and Hussain, A. The hourglass of emotions. In: *Cognitive Behavioural Systems*. 144–157. 2012.

[9] Plutchik, R. *Emotions and life: Perspectives from psychology, biology, and evolution.* American Psychological Association. 2003.

[10] Nesse, R. M. and Ellsworth, P. C. Evolution, emotions, and emotional disorders. *American Psychologist*, 2009. 64(2): 129.

[11] Curumsing, M. K., Fernando, N., Abdelrazek, M., Vasa, R., Mouzakis, K. and Grundy, J. Understanding the Impact of Emotions on Software: A Case Study in Requirements Gathering and Evaluation. *Journal of Systems and*

Software, 2018: 223.

[12] Plutchik, R. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, 2001. 89(4): 344–350.

[13] Ekman, P. What emotion categories or dimensions can observers judge from facial behavior? *Emotions in the Human Face*, 1982: 39–55.

[14] Sutcliffe, A. Emotional requirements engineering. *IEEE International Requirements Engineering Conference*. 2011. 321–322.

[15] Miller, T., Pedell, S., Lopez-Lorca, A. A., Mendoza, A., Sterling, L. and Keirnan, A. Emotion-led modelling for people-oriented requirements engineering: The case study of emergency systems. *Journal of Systems and Software*, 2015. 105: 54–71.

[16] Mehrabian, A. Some referents and measures of nonverbal behavior. *Behavior Research Methods & Instrumentation*, 1968. 1(6): 203–207.

[17] Hsee, C. K., Hatfield, E. and Chemtob, C. Assessments of the emotional states of others: Conscious judgments versus emotional contagion. *Journal of Social and Clinical Psychology*, 1992. 11(2): 119–128.

[18] Adolphs, R. Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Reviews*, 2002. 1(1): 21–62.

[19] Joung, H. Y. and Do, E. Y.-L. Tactile hand gesture recognition through haptic feedback for affective online communication. *International Conference on Universal Access in Human-Computer Interaction*. 2011. 555–563.

[20] Bernardi, N. F., De Buglio, M., Trimarchi, P. D., Chielli, A. and Bricolo, E. Mental practice promotes motor anticipation: evidence from skilled music performance. *Frontiers in Human Neuroscience*, 2013. 7: 451.

[21] Guzman, E. and Bruegge, B. Towards emotional awareness in software development teams. *Proceedings of the Joint Meeting on Foundations of Software Engineering*. ACM Press. 2013. 671–674.

[22] Graziotin, D., Fagerholm, F., Wang, X. and Abrahamsson, P. What happens when software developers are (un)happy. *Journal of Systems and Software*, 2018. 140: 32–47.

[23] Sutcliffe, A. and Thew, S. Analysing ”people” problems in requirements engineering. *IEEE International Conference on Software Engineering*, 2010. 2(0): 469.

[24] Thew, S. and Sutcliffe, A. Value-based requirements engineering: method and experience. *Requirements Engineering*, 2018. 23(4): 443–464.

[25] Lee, C. M., Narayanan, S. S. *et al.* Toward detecting emotions in spoken dialogs. *IEEE Transactions on Speech and Audio Processing*, 2005. 13(2): 293–303.

[26] Anagnostopoulos, C.-N., Iliou, T. and Giannoukos, I. Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011. *Artificial Intelligence Review*, 2015. 43(2): 155–177.

[27] Kumar, S. S. and Rangababu, T. Emotion and Gender Recognition of Speech Signals Using SVM. *International Journal of Engineering Science and Innovative Technology*, 2015. 4(3): 128–137.

[28] Schuller, B., Vlasenko, B., Eyben, F., Wollmer, M., Stuhlsatz, A., Wendemuth, A. and Rigoll, G. Cross-corpus acoustic emotion recognition: Variances and strategies. *IEEE Transactions on Affective Computing*, 2010. 1(2): 119–131.

[29] Rajni and Das, D. N. N. Emotion Recognition from Audio Signals. *International Journal of Advanced Research in Computer Engineering & Technology*, 2016. 5(6): 2105–2107.

[30] El, M., Kamel, M. S. and Karray, F. Survey on speech emotion recognition : Features , classification schemes , and databases. *Pattern Recognition*, 2011. 44(3): 572–587.

[31] Shah, R. D., Anil, D. and Suthar, C. Speech Emotion Recognition Based on SVM Using MATLAB. *International Journal of Innovative Research in Computer and Communication Engineering*, 2016: 2916–2921.

[32] Milton, A. and Roy, S. S. SVM Scheme for Speech Emotion Recognition using MFCC Feature. *International Journal of Computer Applications*, 2013. 69(9): 34–39.

[33] Kwon, O.-w., Chan, K., Hao, J. and Lee, T.-w. Emotion Recognition by Speech Signals. 2003: 125–128.

[34] Kołakowska, A., Landowska, A., Szwoch, M., Szwoch, W. and Wróbel, M. R. Emotion Recognition and Its Applications. In: *Human-Computer Systems Interaction: Backgrounds and Applications 3*. vol. 300. 51–62. 2014.

[35] Petrushin, V. A. Emotion in speech: Recognition and application to call centers. *Proceedings of artificial neural networks in engineering*, 1999. 710: 22–30.

[36] Patel, P., Chaudhari, A., Kale, R. and M.A. Pund. Emotion Recognition From Speech With Gaussian Mixture Models & Via Boosted Gmm. *International Journal of Research In Science & Engineering*, 2017. 3(2): 294–297.

[37] Tandel, S., Patil, S. and Kadam, V. Recognition of emotions from audio signals. 2018. 5(2): 31–35.

[38] Shrawankar, U. and Thakare, V. M. Techniques for feature extraction in speech recognition system: A comparative study. *arXiv preprint arXiv:1305.1145*, 2013.

[39] Gupta, D., Bansal, P. and Choudhary, K. The state of the art of feature extraction techniques in speech recognition. In: *Speech and language processing for human-machine communications*. Springer. 195–207. 2018.

[40] Tiwari, V. MFCC and its applications in speaker recognition. *International journal on emerging technologies*, 2010. 1(1): 19–22.

[41] O'Shaughnessy, D. Automatic speech recognition: History, methods and challenges. *Pattern Recognition*, 2008. 41(10): 2965–2979.

[42] Hossan, M. A., Memon, S. and Gregory, M. A. A novel approach for MFCC feature extraction. *IEEE International Conference on Signal Processing and Communication Systems*. 2010. 1–5.

[43] Khulage, A., Pathak, P. *et al.* Analysis of speech under stress using Linear techniques and Non-Linear techniques for emotion recognition system. *arXiv preprint arXiv:1207.5104*, 2012.

[44] CR, R. Review of Algorithms and Applications in Speech Recognition System. *International Journal of Computer Science and Information Technologies*, 2014. 5(4): 5258–5262.

[45] Tahir, F., Saleem, S. and Ahmad, A. Extracting accent information from Urdu speech for forensic speaker recognition. *Turkish Journal of Electrical Engineering & Computer Sciences*, 2019. 27(5): 3763–3778.

[46] Kurzekar, P. K., Deshmukh, R. R., Waghmare, V. B. and Shrishrimal, P. P. A comparative study of feature extraction techniques for speech recognition system. *International Journal of Innovative Research in Science, Engineering and Technology*, 2014. 3(12): 18006–18016.

[47] Bhattacharjee, U. and Sarmah, K. GMM-UBM based speaker verification in multilingual environments. *International Journal of Computer Science Issues*, 2012. 9(6): 373.

[48] Kumar, V. R., Vydana, H. K. and Vuppala, A. K. Significance of GMM-UBM

based Modelling for Indian Language Identification. *Procedia Computer Science*, 2015. 54: 231–236.

[49] Charbuillet, C., Tardieu, D., Peeters, G. *et al.* Gmm supervector for content based music similarity. *International Conference on Digital Audio Effects, Paris, France*. 2011. 425–428.

APPENDIX A

CONVERSATION

Audio 1
Emotion: Confuse/neutral/calm
Intensity: normal/neutral

Customer: Hello Mr. John, Mr. Steve sent me. He might have told you about me.
Software Developer: Hello Mr. Aden. Yes Mr. Steve have informed your arrival. And told me that you wanted to discuss about your business website.
Customer: Yes, I am here for the website.
Software Developer: Sir, Can you explain me your business?
Customer: Yes, my business is related to sports items. I had a shop nearby and also I import and export sports items outside the city.
Software Developer: So sir, you wanted to have a software or online store or point of sale kind of thing?
Customer: Ah, yes, maybe. But I am not sure what to do. Could you suggest me anything which is within my budget?
Software Developer: Sir, point of sale or online store are costly in developing than a static website with display of your products. That is not much costly.
Customer: Oh ok. I got it. But I am not so much interested in online store right now. Due to budget and also I may work on it in future. But right now I am interested in having something that would help me in advertising my products to other city suppliers. So that I would have more customer to exports my products to.
Software Developer: Understood. So you wanted a static website with a bit description about your business, products samples, prices, images, contact information and some fancy revelant content. That would be enough?
Customer: I am confused right now. As you have told me much. Things are revolving around my head now. But I may be clear after thinking about the things you told me.

Software Developer: Ok sir, you may take your time. Whenever you come up with decision, we will talk about it further.
Customer: That would do it. Thank you for your guidance.
Software Developer: Your welcome sir.
Customer: Well take care. Bye.

Audio 2
Emotion: Sad/Disappointed
Intensity: strong

Customer: Hello Mr. Ali. How are you?
Software Developer: Hello sir. I am fine how about you?
Customer: I am fine. Thank you. What is the status of the milestone?
Software Developer: Sir it is still under development.
Customer: There is a meeting coming up with my clients at the end of this week. I wanted to show some revelant workings. I know the application is going towards more complex things but we wanted to finish the current milestone by the end of this week.
Software Developer: Sir I may be not sure that we could do that. This was a big module. And also after completion, we need to forward it to the QA team.
Customer: Can we just manage to complete some modules of this milestone, so that I would be able to show some contents to clients.
Software Developer: Sir we will try our best to do what we could. Even if we had to work on weekend, we would try to complete some modules, so that it you can show some possible completed modules.
Customer: That sounds good. But I am disappointed by the results right now. But I also understand the complexity of this milestone.
Software Developer: Sir you just don't worry about it. Me and my team is working on it. Hope we won't disappoint you by the end of this week.
Customer: Glad to hear that Ali. Keep working and do update me timely.
Software Developer: Sure sir. I will keep you updated.
Customer: Ok then. Take care. Bye
Software Developer: Bye.

Audio 3
Emotion: Happy/Satisfied/Neutral
Intensity: strong

Customer: Hello Mr. Waqar. How are you doing?
Software Developer: Hello sir. I am fine how about you?
Customer: I am fine. Thank you. Yesterday I had the meeting with my client. I showed him the latest changes of the system.
Software Developer: Oh ok sir, that one. How did it went?
Customer: He was so much happy to see that the things are making real sense now. I showed him the demo of the new modules. And he was also surprised and happy to see the animations. That were very helpful. I am glad by your work. That was some really good stuff.
Software Developer: Thank you sir. And what about the graphs and reports?
Customer: ”It is awesome” was his comment on that. The color scheme that you choose was outstanding. It was attractive on the first look. Also the reports were very revelant. I had some trouble with the dropdowns runtime data but the rest things covered that so well that it doesn’t mattered to him at all.
Software Developer: Oh I will have a look on that. It was working fine. There might be some issue in production side, but I will go through it.
Customer: That’s fine. So what tasks are you working after this?
Software Developer: Sir, I would be going towards more reports as some sections were incomplete.
Customer: Very well. I will discuss with you about further thing next week.
Software Developer: Understood sir.
Customer: Thank you again for all the effort and time. See you again. Bye
Software Developer: Ok sir. Bye. Take care

Audio 4
Emotion: Angry/Dissatisfied/Sad/Intensed
Intensity: strong

Customer: Hello Mr. Hassan. How are you doing?
Software Developer: Hello sir. I am fine how about you?
Customer: I am fine. Thank you. I need to talk to you about the latest changes in the system?
Software Developer: Yes sir?
Customer: Everything you mentioned in the slides were not what the system had. The system had some missing modules. Some modules were incomplete. Some were not working properly. And the main thing is that. The system hangs a lot and had a lot of errors as it pops up on browser. What is this? Can you explain?
Software Developer: No sir it can't be. I have worked on it. How it is possible.
Customer: Well I am not the developer. Could you please verify your working? I am surprised how the QA team handling the system if it had so much bugs and irrelevant errors. How am I going to show this to my clients? I wanted the complete modules as mentioned in slides by the end of 1st weeks of next month.
Software Developer: Sir I have figured it out, it was database issue that why everything was not working properly.
Customer: Please fix all the issues right now. I have a meeting with another client right now. I had to show him some demo regarding the old modules which were working fine.
Software Developer: I will go through that issue sir.
Customer: Thank you so much. I am sorry if it seems that I was harsh with you, but you had to understand.
Software Developer: No problem sir. I can understand that.
Customer: Good. Please continue your work. And do inform me when it is ready to use.
Software Developer: Sure sir.
Customer: Ok then. Bye
Software Developer: Bye sir

Audio 5
Emotion: Neutral/Satisfied/calm
Intensity: normal/neutral

BDM: Business Development Manager

Customer: Hello Mr. Ahmed speaking?
BDM: Hello. Yes sir. It's Mr. Ahmed here.
Customer: How are you doing?
BDM: I am fine thank you. How about you?
Customer: I am fine too. I heard your company develops websites for businesses?
BDM: Yes sir. Our company works on that. Have you called up for website purpose?
Customer: Yes, I was interested to have one for my business.
BDM: Very well sir. What sort of business you are up to?
Customer: Actually, my business is regarding to stationary items. I have book stores in 3 different cities. But I just wanted to have a static website for my business.
BDM: Ok sir. You just wanted to have a static website for your business?
Customer: Yes indeed. I knew a bit about websites and development. That's why I came up to point that I need just a static website with little information regarding to business, some images or gallery and contact information. All the exact requirement will be sent to you. I just needed a five to six page website with all the required information.
BDM: Very well sir. Our company also deals with marketing and advertising of websites and businesses. Do you want to us to also work on digital marketing?
Customer: Would that be costly?
BDM: Not so much sir. But it will help you improve your business.
Customer: I will think about it later. First we get through the website.
BDM: Ok sir. I will transfer your call to my finance department for all your information and discussion related to finance.
Customer: Very well. Thank you.
BDM: Thank you sir.

Audio 6
Emotion: Angry/Disappointed
Intensity: strong

Customer: Hello Mr. Fahad. It's Mr. Waqar speaking. How are you?
Software Developer: Hello sir. I am fine. How about you sir?
Customer: Yes good. I wanted to inform that the website that your team has developed for my business it not working. It is down for some reason. Can you please verify?
Software Developer: Sure sir. Wasn't that the graphics website?
Customer: Yes you are right. It was a graphics business.
Software Developer: Ok sir.... Sir you domain has been expired. After end of every year you had to renew your domain. All the costs regarding domain and servers were informed to you.
Customer: What costs? Do i need to pay every month for domain?
Software Developer: Yes sir. It was already informed.
Customer: I don't remember. But what kind of service is this. I had to pay each year for my website.
Software Developer: It is how it works sir. You have to pay each year for this service that your website has been deployed to their servers.
Customer: It is of no use then. I was only using my website occasionally. And I had to pay for what I am not using that much?
Software Developer: Sir if you won't pay. Then the website won't work. And we cannot do anything as it is not in our hands.
Customer: I am disappointed. Let me first talk to my partners. I need to involve them too due to business partners.
Software Developer: Sure sir. Whenever you are ready, I will be up again.
Customer: Fine. Give me some time. Thank you
Software Developer: Sure sir. Thank you. Have a nice day.
Customer: Thank you. Bye
Software Developer: Bye

Audio 7
Emotion: Calm/Satisfied/Happy/neutral
Intensity: strong

Customer: Hello Naveed. How are you?
Software Developer: Hello sir. I am fine. How about you sir?
Customer: I am fine too thank you. What's the status of our module?
Software Developer: Yes sir currently working on graphs module. Rest is completed.
Customer: Can you show me a demo of what's graphs we are having.
Software Developer: Yes sir... As you can see, this multi stacked vertical bar graph would be perfect for viewing timeline of each user.
Customer: This is very good and attractive. Add some more filters for searching purpose.
Software Developer: Yes sir. I will be working on the filters too.
Customer: Good. Everything is working well I see.
Software Developer: Yes sir just to let you know that I couldn't finish up yesterday's tasks for your meeting.
Customer: Oh don't worry about that. I will handle it somehow. You just focus on the work you are currently doing. And don't put yourself into much trouble. If you find out that anything would be done eaily and you would be searching some hard way, just go for the simplicity.
Software Developer: Ok sir. Thank you.
Customer: Just try to make things simple, brief and attractive. Rest modules were working fine. I will handle something in the next meeting. For now, just relax.
Software Developer: Hahaha. Sure sir. I will just inform you when this will be completed. Hope so it will be completed before your next meeting.
Customer: Don't worry about that. If it completes then that's good. But if it isn't completed before that, then don't worry about it. All you need is to take some rest. As we needed your calm brain working perfectly on this product.
Software Developer: Haha. Got it sir. Thank you
Customer: You're welcome. Now I had to leave. Take care. Bye
Software Developer: Bye sir.

Audio 8
Emotion: Dissatisfied/Disgust
Intensity: strong

Customer: Hello Shahid. How are you?
Software Developer: Hello sir. I am good. How about you sir?
Customer: I am good thanks. How is the work going?
Software Developer: Everything is working fine sir. I have updated the latest release on the server. Have you checked it?
Customer: Yes I have seen that. Everything was good but I wanted to talk about the procedure of sale form if customer is not added in the system. I have seen the workflow, that the customer must be added before sale. So it means that every time when a sale is performed and a customer is standing on the counter and if he is not registered, does he had to wait for like 2 to 3 mins to register himself, rather than just buy the products and leave?
Software Developer: Sir right now it is happening in this way. But we are trying to manage the walk-in customers too. Currently my team is working on it. Soon it will be released but for that case, you won't be able to see the ledgers of that customer as it will be added with the walk-in customer information.
Customer: That is rough I guess?
Software Developer: Yes sir. There is no other way I guess. But if we found something else we would go with it.
Customer: Sure. Just think about getting around from somewhere else. It is good for now. But i might find it difficult later on.
Software Developer: Sure sir. We will figure it out.
Customer: I don't want my customers to dissatisfy my progress. Everything is working well.
Software Developer: Sure sir i understood.
Customer: Thank you for understanding. Please keep on your work. You are doing well. Take care. Bye
Software Developer: Thank you sir. Take care. Bye

Audio 9
Emotion: Worried/Tensed
Intensity: strong

Customer: Hello Zafar. How are you?
Software Developer: Hello sir. I am good. How about you sir?
Customer: I am good thanks. What happened to the server?
Software Developer: Everything was working fine. But suddenly all the services stopped right now. Let me check the server status.
Customer: Yes please verify the server.
Software Developer: Sir I have verified the server status. Server is not responding. Let me contact the providers.
Customer: Please verify. I am worried.
Software Developer: Sir I talked to the providers. They say that our server is down. Someone powered it off.
Customer: Oh yes I powered it off a moment ago. But I didn't know that we shouldn't do that. Now what should we do?
Software Developer: I will contact the providers. They will open that up but we will have to pay for the charges.
Customer: Oh don't worry about the charges. I am only worried about my data. It shouldn't loss at any cost.
Software Developer: I am not sure about that. This situation has come up for the first time. I am not sure about the data loss
Customer: Please confirm that from the providers. And also it there is issue regarding data loss, then please upload some backups. We might not have the latest but at least we have some backup's right?
Software Developer: We will have to look to it.
Customer: I am more worried now. Please update me as soon you get something up. I am online right now.
Software Developer: Sure sir. I will give you a call.
Customer: Thank you. I am waiting. Bye
Software Developer: Bye sir

Audio 10
Emotion: Anxiety/Happy/Surprised/Worried
Intensity: strong

Customer: Hello Saad. How are you?
Software Developer: Hello sir. I am good. How about you sir?
Customer: I am good thanks. Is everything ready? Are we all set?
Software Developer: Yes sir. Everything is ready and working. My QA team has verified everything on production side. It is all good.
Customer: Good. How much time is left for the Christmas?
Software Developer: Sir, 1:30 more minutes.
Customer: I am very anxious and curious.
Software Developer: Yes sir. We all are anxious here.
Customer: I am sure the server won't be down this time.
Software Developer: Yes sir we have worked on that and tested with hundred thousand requests per second. It didn't went down. Because this time, we are using hybrid cloud infrastructure.
Customer: This is a good news. But hope that it won't worry when we are live.
Software Developer: Sure sir. Sir we are getting live now in 3 2 1... Opened.
Customer: Good. Let me open the graph stats
Software Developer: Yes sir we are watching that graph.
Customer: Yes. Yes. A hundred requests in first ten seconds. Yes that is really good.
Software Developer: Yes sir. Let's hope for the best.
Customer: Hope my goods in shops in all cities would be enough.
Software Developer: Thousand request in ten second. And this is working pretty well.
Customer: Yes. That is a good handle. I am anxious more and more. What will happen till the end of the day? That's a lot of sales man.
Software Developer: Yes sir. Hope it handle hundred thousand request by the end of the day. We need to add backup small server backup too just in case if anything goes down.
Customer: Sure, whatever you think is the best for current moment.
Software Developer: Yes sir. It will be installed within 10 mins. Sales reached to five thousand. That's incredible.
Customer: Oh yes. That's awesome. Ok you guys just view the things and enjoy. Do perform quick actions if anything goes down. I am going to have a look on the shops status and check if everything is going well.
Software Developer: Ok sir. You just don't worry about it. We will handle this.
Customer: Thank you Saad. Keep me updated.

Software Developer: Sure sir.
Customer: Ok then. Take care and good job everyone. Bye
Software Developer: Thank you sir. Have a nice day. Bye

APPENDIX B

EMOTIONAL SPEECH DATABASES

English: ”Yeah right” corpus by Tepperman et al. (2006). **Emotions:** Sarcastic, neutral. **Elicitation:** ”Yeah right” utterances taken from the Switchboard and Fisher corpora of spontaneous telephone dialogues. **Size:** 131 utterances.
English: ISL Meeting corpus (Neiberg et al. 2006). **Emotions:** Negative, positive, neutral **Elicitation:** Recordings of 18 meetings with a total of 92 persons and an average duration of 35 min accompanied by orthographic transcription. **Size:** 12,068 utterances, thereof 424 negative, 2,073 positive and 9,571 neutral.
English: Emotional speech corpus by Kumar et al. (2006). **Emotions:** Inappropriateness, lack of clarity, uncertainty, neutral **Elicitation:** Recordings of participants interacting with an SLDS in terms of a customer survey about grocery stores plus answering of a questionnaire. **Size:** , 257 utterances, 17 participants (10 females, 7 males).
English: Speech database by Lee et al. (2006). **Emotions:** Angry, happy, sad, neutral **Elicitation:**Recordings and magnetic resonance images of a male speaker uttering a set of four sentences. **Size:** , 80 utterances (four sentences, five repetitions, four emotions), one male speaker.
English: Castaway database (Devillers et al. 2006). **Emotions:** Wide range of emotions **Elicitation:**Audio-visual recordings of a reality TV show. **Size:** , 10 recordings (30 min each), 10 speakers.
English: Situation Analysis in a Fictional and Emotional (SAFE) corpus (Clavel et al., 2006). **Emotions:** Fear, other negative emotions, positive emotions, neutral **Elicitation:**Audio-visual excerpts taken from movie DVDs, abnormal contexts. **Size:** , 400 sequences, total length 7 h, 4,724 segments of speech (up to 80 s).
English: Expressive spoken corpus of children’s stories modified by Alm and Sproat (2005). **Emotions:** Angry, disgusted, fearful, happy, sad, surprised, neutral **Elicitation:**: Recordings of a semi-professional female speaker reading two children’s stories in an extremely expressive mode. **Size:** , Approx. 10 min of speech, 128 sentences, one female speaker.

English: HMIHY speech database (Liscombe et al. 2005). **Emotions:** Positive/neutral, somewhat angry, somewhat frustrated, somewhat other negative, very angry, very frustrated, very other negative **Elicitation:** Recordings of callers interacting with an automated agent concerning account balance, explanation of bill charges, AT&T rates and calling plans, etc. **Size:** , 5,690 dialogues, 20,013 user turns.
English: Speech database by Lee et al. (2005).. **Emotions:** Angry, happy, sad, neutral. **Elicitation:** Recordings of a male speaker producing sentences with non-emotional content in the respective emotional states (including biosensor data). **Size:** ,280 utterances, 14 sentences, one male speaker.
English: WOZ data corpus (Zhang et al., 2004). **Emotions:** Confidence, puzzle, hesitation **Elicitation:** Audio-visual recordings of children interacting with an intelligent tutoring system for learning basic concepts of Mathematics and Physics. **Size:** 714 students' utterances (approx. 50 min of clean speech), 4.2 s of speech and 8.1 words per utterance, 17 speakers.
English: Modified LDC CallFriend corpus prepared by Yu et al. (2004. **Emotions:** Boredom, happy, hot anger, interest, panic, sadness, no emotion plus numerical values (on a discretized scale from 1 to 5) for each of arousal, valence and engagement **Elicitation:** Recordings of social telephone conversations between friends. **Size:** 1,888 utterances (1,011 utterances from female speakers, 877 utterances from male speakers, eight speakers (four females, four males).
English: Emotional speech synthesis database (Tsuzuki et al., 2004). **Emotions:** Anger, happiness, sadness, neutral **Elicitation:** Recordings of a non-professional male speaker uttering short declarative sentences with emotional content **Size:** ,363 utterances, one male speaker.
English: Speech database by Lee et al. (2004) and Yildirim et al. (2004). **Emotions:** Angry, happy, sad, neutral. **Elicitation:** Recordings of a semi-professional actress uttering 112 unique sentences in four emotions **Size:** 880 utterances, one female speaker.
English: Sensitive Artificial Listener (SAL) database by Cowie et al. (2004). **Emotions:** Wide range of emotions or emotion related states **Elicitation:** Audio-visual recordings of speakers interacting with an artificial listener with different personalities **Size:** Recordings of approx. 10 h, 20 speakers.

English: LDC Emotional Prosody Speech and Transcription used by Liscombe et al. (2003) and Yacoub et al. (2003). **Emotions:** Anxiety, boredom, cold anger, contempt, despair, disgust, elation, happy, hot anger, interest, panic, pride, sadness, shame, neutral. **Elicitation:** Professional actors reading short (4-syllables each) dates and numbers **Size:** eight actors (five females, three males), 44 utterances used by Liscombe et al. (2003), 2,433 utterances used by Yacoub et al. (2003).
English: Speech database by Lee and Narayanan (2003). **Emotions:** Negative, non-negative. **Elicitation:** Users interacting with a machine agent in a call center. **Size:** 1,367 utterances (776 utterances of female speakers, 591 utterances of male speakers).
English: Emotional speech database by Fernandez and Picard (2003). **Emotions:** Stress **Elicitation:** Recordings of speakers solving mathematical problems while driving a car simulator. **Size:** four speakers, four situations, 598 utterances, length varying from 0 to 6 s.
English: Capital Bank Service and Stock Exchange Customer Service (Devillers et al. 2002). **Emotions:** Anger, excuse, fear, satisfaction, neutral. **Elicitation:** Human-human interaction in a stock exchange customer service (call) center. **Size:** 100 dialogues, 5,229 speaker turns.
English: DARPA Communicator corpus (Walker et al. 2001). **Emotions:** Annoyance, frustration. **Elicitation:** Users making air travel arrangements over the phone **Size:** Recordings of simulated interactions with a call center, 13,187 utterances in total (1,750 emotional utterances).
English: Database by France et al. (2000). **Emotions:** Depression, suicidal state, neutrality **Elicitation:** Recordings of spontaneous dialogues between patients and therapists in therapy sessions, phone conversations and post therapy evaluation sessions. **Size:** 115 speakers (48 females, 67 males).
English: Belfast Naturalistic Database (Douglas-Cowie et al. 2000). **Emotions:** Wide range of emotions. **Elicitation:** Audio-visual recordings of people discussing emotive subjects with each other/the research team plus recordings of extracts from television programs, i.e., members of the public interacting in a way that appears essentially spontaneous. **Size:** 239 clips (209 from TV Recordings, 30 from interview recordings.

English: Database produced by Polzin and Waibel (2000). **Emotions:** Anger, sadness, neutrality (other emotions as well, but in insufficient numbers to be used). **Elicitation:** Audio-visual data, i.e., sentence-length segments taken from movies. **Size:** 1,586 angry segments, 1,076 sad segments, 2,991 neutral segments.
English: Emotional speech database by Pereira (2000). **Emotions:** Cold anger, happiness, hot anger, sadness, neutral. **Elicitation:** Recordings of actors uttering two sentences in different emotional states. **Size:** 80 utterances (two repetitions of 40 utterances), two speakers.
English: Emotional speech database by McGilloway et al. (2000). **Emotions:** Anger, fear, happiness, sadness, neutral. **Elicitation:** Recordings of speakers reading emotional texts in appropriate style. **Size:** 40 speakers, five texts (100 syllables each).
English: Reading/Leeds Emotional Speech Corpus (Greasley et al. 2000). **Emotions:** Anger, disgust, fear, happiness, sadness, neutral. **Elicitation:** Recordings of interviews on radio/television, speakers asked by interviewers to relive emotionally intense experiences. **Size:** Approx. 5 h of samples of emotional speech.
English: Emotional speech database by Robson and Mackenzie-Beck (1999). **Emotions:** Smiling, neutral **Elicitation:** Recordings of speakers uttering sentences in a neutral state and while smiling **Size:** 66 utterances, 11 speakers, three sentences.
English: Emotional corpus by Cowie et al. (1999b). **Emotions:** Wide range of emotions as defined in the FEELTRACE tool **Elicitation:** Video tape recordings of groups of three friends each discussing about issues they strongly felt about. **Size:** Recordings of 1 h per group, nine speakers (three groups of three friends).
English: Emotional speech database by Whiteside (1998). **Emotions:** Cold Anger, elation, happiness, hot anger, interest, sadness, neutral. **Elicitation:** Recordings of actors uttering sentences in different emotional states. **Size:** 70 utterances, two speakers (one female, one male), five different short sentences.

English: Emotional speech database by Li and Zhao (1998). **Emotions:** Anger, fear, happy, sad, surprised, neutral. **Elicitation:** Recordings of actors uttering 20 sentences with emotional content and three sentences without emotional content in different emotional states. **Size:** 5 untrained speakers (two females, three males), 23 sentences per speaker.
English: SUSAS database (Hansen et al. 1998). **Emotions:** Talking styles (angry, clear, fast, loud, question, slow, soft), single tracking task (high stress, Lombard effect, moderate), dual tracking task (highstress, moderate), actual speech under stress (anxiety, fear, G-force, Lombar effect, noise), psychiatric analysis (angry, anxiety, depression, fear **Elicitation:** Recordings of isolated-word utterances under simulated or actual stress in several scenarios, e.g., amusement park roller-coaster, helicopter cockpit, patient interviews. **Size:** Approx. 16,000 utterances of 36 speakers (13 females, 23 males) in total.