A Survey of Information Technology Applications to Treat Fear of Public Speaking

Qays Algahreeb[1,*] , Florica Moldoveanu[1] and Alin Moldoveanu[1]

[1]University Politehnica of Bucharest Faculty for Automatic Control and Computers Department of Computer Science and Engineering

*Corresponding Author: Qays Algahreeb

Keywords: fear, phobias, treatment, computerbased toolsPublic speaking started to gain much attention when it comes to phobias, which is anxiety for new presenters. In some cases, specialists consider that avoiding the phenomenon which causes the phobia is sufficient treatment; in others, the exact opposite, being gradually exposed to the object of fear may lead to a cure. We have to start looking for other psychotherapeutic methods, innovative ones, to help people surpass their immense fears and improve their ability to give presentations. The current article presents a survey on discovering fear and anxiety when preventing and treating it and analyses their utility as tools for learning how to overcome this type of phobias, thus improving presentation ability. Using IT-based solutions for treating presented this fear, especially anxiety for new presenters. The current methods of dealing with the fear of public speaking will be reviewed, as well as Clarify the technology (tools, systems, and applications) based used for detecting and treatment. We will analyze research that studies how to detect fear and the ways to treat it, the concept behind their mechanism and the possibility of exploiting them in presentations. therefore, the paper debates these IT instruments and applications in this field. Based on the results of the survey, we will propose an appropriate mechanism for detecting degrees and types of fear when presenting presentations and their treatment.

1. INTRODUCTION

Modern life often involves situations where we are required to speak in public, both in our professional lives, for instance when presenting results of our work in front of colleagues, and in our working life, such as when teaching and giving presentations.

Given the prevalence of public speaking situations in modern professional and personal life, it is natural that some individuals want to improve their ability to speak in public. Additionally, anxiety about public speaking is very common, and some people experience an uncontrollable amount of stress when preparing or public speaking. These two cases require the development of some methods and tools to support assessing people's ability to speak to the public, training them in public speaking skills and reducing anxiety and stress during public speaking [1, 2].

Emotion is defined as a conscious mental reaction subjectively experienced and directed towards a specific object, accompanied by physiological and behavioral changes in the body. The field of affective computing aims to enhance the interaction between the human and the machines by identifying emotions and designing applications that automatically adapt to these changes. [2]

Effective computing is a study of systems or devices that can identify and simulate emotions and their treatment methods. This field is applicable to education, medicine, social sciences, entertainment and so on. The purpose of emotional computing is to improve user experience and quality of life, which is why various emotional models have been proposed over the years and effective mathematical models have been applied to extract, categorize and analyze emotions. [3]

It has drawn the attention of researchers from interdisciplinary domains, being at the confluence of psychology, medicine, and computer science. With applications in education, cognitive-behavioral sciences, healthcare and entertainment, affective computing deals with recognizing and modeling human emotions in a way that would improve overall user experience. To classify emotions, some separate models and dimensions have been proposed and applied over the years.

Discrete patterns of influence depend on the existence of a set of fundamental feelings from which the most complex feelings are derived. Dimensional models rely on a multidimensional space where each axis represents the values of the emotional component [4, 5].

Public speaking started to gain much attention when it comes to phobias, which is anxiety for new presenters. Depending on the survey results we will use new integrated artificial intelligence techniques, we will propose computational models that detect models and levels of emotion and voice of those suffering from a phobia. As a goal, develop a phobias treatment system that will automatically determine fear levels and adjust exposure according to the current effective user situation.

Examination and designing of human conduct are basic for human-driven systems to anticipate the result of communication between social, and to improve the connection between people or among people and PCs. Human conduct is communicated and saw regarding verbal and viewable prompts (for example hand and body signals, facial expression). These conduct signs can be caught and handled to anticipate the result of social connections. Open talking is a significant part of the human correspondence. A decent speaker is lucid, has persuading non-verbal communication, and regularly, can significantly influence individuals. While the accomplishment of open talking generally relies upon the substance of the discussion, and the speaker's verbal conduct, non-verbal (viewable) signals, for example, motions and physical appearance play asignifi control in public speaking.

Our paper consists of two sections: the first section is the introduction, in the second part we Show and analyzes a group of research that discusses the IT applications to detect and treat fear of public speaking, and the last section summarizes conclusion.

2. SHOW AND ANALYZES GROUP IMPORTANT RESEARCH THAT DISCUSSES THE IT SYSTEMS AND APPLICATIONS TO DETECT AND TREAT FEAR OF PUBLIC SPEAKING

2.1 DETECTING EMOTIONS USING VOICE SIGNAL ANALYSIS

The application examines speech more specifically to determine the feeling of using statistics and neurological systems to describe the parameters of the speech signal through the feelings the systems are directed to perceive.

The application detects a passionate state in a voice signal. The system contains a speech pickup and one computer associated with the speech pickup tool.

The present development identifies with the investigation of discourse and all the more specially to distinguishing feeling utilizing measurements and neural systems to arrange discourse signal parameters as per feelings the systems have been educated to perceive. These outcomes give a significant understanding of human showing and can avail in as a benchmark for correlation with a computer. the new field of research in Artificial Intelligence (AI) known as affective computing has as of late been recognized. Affective computing investigates around IT and emotional states, joining data about human feelings with registering capacity to improve human-computer connections. furthermore, as research on perceiving feelings in the discourse, AI scientists have made passionate discourse combination, acknowledgment of feelings, and the utilization of specialists for unraveling and communicating feelings.

A more intensive see how well individuals can perceive and depict feelings in discourse is uncovered Thirty subjects of both genders recorded to four short sentences with five distinct feelings (joy, outrage, bitterness, dread, and equal state or typical). shows an exhibition perplexity, wherein just the numbers on the expected (true) feeling with the distinguished (estimated) feeling.

This demonstrates (11.9%) of expressions that were depicted as cheerful were assessed as equal (unemotional), (61.4%) as really upbeat, (10.1%) as irate, (4.1%) as miserable, and (12.5%) as apprehensive. The most effectively conspicuous class is outrage (72.2%) and the least unmistakable classification is dread (49.5%). There is significant disarray among pity and dread, bitterness and apathetic state, and satisfaction and dread. The mean precision of (63.5%) concurs with the consequences of another trial examines.

These outcomes give a significant understanding of human execution and can fill in as a pattern for correlation with PC execution.

the system moreover contains memory operably associated with the computer, and a computer program including a neural net for separate the Voice signal into several portions,

and for analyzing down the Voice signal Based on the merit of fragments to identify the passionate state in the Voice signal. The system involves a database of speech signal discernment and the PC for examination with highlights of the Voice signal, and output device coupled to the PC for informing a client of the emotional State identified in the Voice signal.

Another epitome of the method is regulation for characterizing speech. The method involves a PC system having a (CPU), an inputting device, one memory for saving information indicative of a talking signal, and an outputting device. The PC method likewise contains a rationale for getting and examining the discourse signal, the rationale for partitioning the discourse sign, and rationale for taking one characteristic from the discourse signal. The system likewise includes a database of discourse signals and statistics attainable of a computer for comparing the Voice signal and an outputting tool coupled to the PC for advising a user of the passionate situation that disclosed in the Voice signal.

2.2 MULTIMODAL EXPRESSIONS OF STRESS DURING A PUBLIC SPEAKING TASK

"Emotion" means to a set of psychic situations including subjective experiment, expressive attitude(e.g., verbal, facial, bodily,), and circumferential physiological responses (e.g., heart rate, breathing). In emotional Computing, the differentiation or the molding of these compound states requires the set of the human dataset.

To collect data on emotional features naturally. Previously, the research relied on sentiment representatives to create examples of traditional passionate cases for humans.

The current approach it's by collects natural data in two ways. attention to emotional expressions, its role in communication, and putting of a communication model protocols (GEMEP Corpus). then proposing new protocols to represent emotions that are a reaction to some events.

There are various materials and tasks, that full of feeling as pictures, or games. Stress states are models' instances of affectional states that individuals in regular daily existence with potential emotional affective applications.

public speaking is a task referred to the experience as a strain state. the protocol proposed gathering a multimodal database that will empower futurity researches to How individuals' contrast regarding how they can adapt to this difficult circumstance? What feelings are cause this state? and Which statistics ought to be utilized to evaluate users' pressure-related feelings?

there are Databases of multimodal explanations of feeling states occurring through tasks are but are few.

to present a protocol for evoking worry in a talking task. Practices of 19 members were recorded through a multimodal arrangement including discourse, video of the outward appearances and body developments, balance utilizing a power plate, and physiological measures. Surveys were utilized to attest sentimental states, character profiles and pertinent adapting practices to contemplate how members adapt to upsetting circumstances and some personal emotional behavior assessed. Results show a noteworthy effect on the general situations and conditions on the members' passionate influence. The future utilization of this new multimodal passionate corpus is represented.

introduced a protocol for gathering multimodal non-acted emotional articulations in a distressing circumstance. This protocol was a copy adaptation of the generally utilized Social Stress Test (Trier) known to instigate moderate worry in research settings.

To empower futurity searches a total examination of feeling articulations, guidelines and adapting right now, convention incorporated a wide assortment of measures: surveys about character and adapting patterns to have singular profiles, surveys about passionate and uneasiness situations, evaluation of the state and personal performance to have the self-reported practice of the user, multimodal conduct measures to catch non-verbal looks of members (voice, face, body) and physiological factors to have markers of the activity level of the member. adjectives statistics are provided for users.

Another point of view is to utilize the gathered information to analyze various algorithms for multimodal combination for feeling recognition. At last, investigations of explicit coincide portions can plot precis parts of passionate articulation and organization. This incorporates, for instance, the investigation of connection style between the user and the assessors. These outcomes urge to expand the size of the current database to empower examinations that think about individual contrasts. Incited pressure doesn't generally debilitate execution as it relies upon the status.

2.3 PRESENTATION TRAINER, YOUR PUBLIC SPEAKING MULTIMODAL COACH

"Practice does not make perfect. Only perfect practice makes perfect." It is a well-known expression by Vince Lombardi, one of the best mentors throughout the entire existence of Professional Football [6].

A key factor to accomplish this "perfect practice" required for the progress and perfection aptitudes is feedback, which has additionally been distinguished as one of the most persuasive involvements in learning [7]. Having a human coach furnishing us with criticism at whatever point they have to rehearse their abilities is neither nor a possible. In an effort to study an affordable solution for this feedback availability challenge, they explored the topic 'public speaking skills. keep track of a plan-based discussion design [8] creating various models of the Presentation Trainer (PT). The PT is an idea of a mechanized feedback apparatus that tracks the users ' voice and body. It gives them feedback around their nonverbal correspondence, with the goal to help them with the advancement of their public talking abilities. characterize the present form of the PT and present the learner experience assessment of examination, where users needed to set themselves up

for a lift pitch. This examination followed a semi exploratory set-up where they investigated the learning impacts of the feedback gave by the.PT.

The Presentation Trainer.

is a multimodal instrument intended to help the exercise of public talking abilities, on granting the learner the notes about various parts of nonverbal correspondence? It tracks the client's voice and body to explain his present showing. In light of this showing the Presentation Trainer chooses the kind of interference that be displayed as notes to the client. feedback was developed system thinking about the outcomes from past examinations that show how troublesome it is for users to see and effectively decipher continuous notes while rehearsing their talking. to introducing the learner experience assessment of users who utilized the Presentation Trainer to rehearse for a lift pitch, demonstrating that the input gave by the Presentation Trainer important impacts on education. studies have confirmed that notes given from a mentor impacts the improvement of public talking aptitudes [9] and that the extent of this impact relies upon how these notes were given to the user. A significant factor that influences the improvement of these aptitudes is the time where notes are given. like the nonverbal connection of the user, quick notes have demonstrated to be effective and productive [10]. Thusly the adaptation of the PT depicted here can examine the user's presentation, and manner select nonverbal connection to be displayed as notes.

2.4 SELF-SPEECH EVALUATION WITH SPEECH RECOGNITION AND GESTURE ANALYSIS

Two fundamental procedures assist an interlocutor to convey a significant discourse. The methods are voice change which communicates a verbal message and movement signals of body that pass on the message to the public.

There are known associations to help and improve discourse like (Toastmasters international), (Australian Rostrum) and (Association of speakers). [11]. Their systems of evaluation are such as, Tracking Filler words, Usage of Redundant words and Phrases, Checking Grammar and Pronunciation.

Their systems of assessment as, the pursuit the filler words, use of plus words and Sentences, Check regulation and spoke, Use and check Gestures, pursuit of Vocal changes, and Time administration. anyone that needs to self-assess for his discourse, he must be a member of these associations.

With Utilizing this method, individuals can assess their discourse without relying upon these associations.

All the previously aforesaid criteria in manual assessment content are Founded in this methodology. and Based on the widespread use of the mobile phone, the proposal will be based on a system Android Platform. Many technical and processes are utilized to interconnect with Android as OpenCV, Microsoft Cognitive Services and MATLAB to realize the goals of the usage. vocal model, Support Vector Model (SVM), Hidden Markov Model (HMM), are a few models used to fabricate the application all the more proficiently by giving around exact outcomes.

Six standards should be looked at the self-evaluation of speak. check vocal variety, filler words, use of repetitive words and expressions, check sentence structure and elocution, use of body movement, and time administration.

the application will recognize physical signals, move, and places, for example, rise hands, move hands, the situation of hands. These little moves will be extremely valuable to make the self-evaluator mindful of what kind of signs they made and whether they have any moves of discourse. In this manner by creating a report that includes how often the user made proper motions and improper motions.so that it would be simple for the speaker to address the wrong signals later.

The mechanism of work is as follows: The client transfers the video to enforcement. The enforcement will investigate the body motions and signals utilizing the Model made by the enforcement. so that to give the outcomes,

A) **hand signal acknowledgment comprises of three essential preparing stages [9].**

1. Hand/Body Segmentation: The essential strategy of hand division is to identify and understand the hand area in the picture which is got hand motions, and subtract them from the surroundings.

2. Gesture Modeling: In the period of hand movement assessment, sundry hand moves and motions are assembled and registered to send them as traineeship and testing information to make a model that will be utilized during order.

3. Gesture Classification: hand motion estimate, which will be utilized for preprocessing and advantages conclusion.

B) **Tracking Filler words. Pace and Time administration.**

the application will take the client's recorded sound and transform it into a sound transcript. [12], [13], [14], [15]. The system is created to recognize sentence stops and filler words from the outcomes. The primary functionalities of unit advantages are below.

1. Identifying Filler words: In request to recognize the filler words, the application should transform talking to write, and the procedure is done through the sound copy [16].

2. Pace: The app shows the pace of words every minute.

3. Time administration: Time administration is assessed by color cards in the interface.

C) Tracking Grammarian Functionalities.

The client needs to choose the sound document. The app will get the sound segment and transform it into a written document [17]. At long last, the app will grant a report about the assessment of grammarian operation [18].

D) Tracking Vocal Variations.

The app of the vocal varieties is partitioned into three primary practical units. They are:

1. Detecting vocal varieties.

2. Sound segment of the discourse in '.wav' form is embed into the app. Volume variety is specified. [19], [20], [21], [22], [23], [24].

3. Generating charts for vocal varieties:

4. Distinguishing the vocal variety and make a chart of volume variation. [25].

5. Show variety in the discourse transcript:

6. The system converts audio to text format to show differences in sound and size [26].

2.5 AUTOMATIC SPEECH EMOTION RECOGNITION USING MACHINE LEARNING

The Comparative study of speech emotion recognition (SER) systems are presents from Leila Kerkeni et al. [27]. Theoretical definition, categorization of affective state and the modalities systems. Theoretical definition, categorization of affective state and the modalities of emotion expression are presented. To achieve this study, a SER system, based on different classifiers and different methods for features extraction, is developed. Mel-frequency cepstrum coefficients (MFCC) and modulation spectral (MS) features are extracted from the speech signals and used to train different classifiers. Feature selection (FS) was applied in order to seek for the most relevant feature subset. Several machine learning paradigms were used for the emotion classification task. A recurrent neural network (RNN) classifier is used first to classify seven emotions. Their performances are compared later to multivariate linear regression (MLR) and support vector machines (SVM) techniques, which are widely used in the field of emotion recognition for spoken audio signals. Berlin and Spanish databases are used as the experimental data set. This study shows that for Berlin database all classifiers achieve an accuracy of 83% when a speaker normalization (SN) and a feature selection are applied to the features. For Spanish database, the best accuracy (94 %) is achieved by RNN classifier without SN and with FS.

The researchers present a system for the recognition of «seven acted emotional states (anger, disgust, fear, joy, sadness, and surprise)». To do that, they extracted the MFCC and MS features and used them to train three different machine learning paradigms (MLR, SVM, and RNN). They demonstrated that the combination of both features has a high accuracy above 94% on the Spanish database. All previously published works generally use the Berlin database. To our Social Media and Machine Learning knowledge, the Spanish emotional database has never been used before. For this reason, they have chosen to compare them. In this chapter, they concentrate to improve accuracy; more experiments have been performed. This chapter mainly makes the following contributions:

- The effect of speaker normalization (SN) is also studied, which removes the mean of features and normalizes them to unit variance. Experiments are under a speaker-independent condition.

- Additionally, a feature selection technique is assessed to obtain good features from the set of features extracted in.

The rest of the chapter is organized as follows. In the next section, they researchers start by introducing the nature of speech emotions. Section 3 describes features they extracted from a speech signal. A feature selection method and machine learning algorithms used for SER are presented. Section 4 reports on the databases they used and presents the simulation results obtained using different features and different machine learning (ML) paradigms. Section 5 closes this chapter by analyses and conclusion.

2.6 DETECTION AND ANALYSIS OF EMOTION FROM SPEECH SIGNALS

Assel Davletcharova et al. [28] they are presents Experimental study on recognizing emotions from human speech. The emotions considered for the experiments include neutral, anger, joy and sadness. The distinguishability of emotional features in speech were studied first followed by emotion classification performed on a custom dataset. The classification was performed for different classifiers. One of the main feature attributes considered in the prepared dataset was the peak-to-peak distance obtained from the graphical representation of the speech signals. After performing the classification tests on a dataset formed from 30 different subjects, it was found that for getting better accuracy, one should consider the data collected from one person rather than considering the data from a group of people.

For studying the basic nature of features in speech under different emotional situations, Researchers used data from three subjects. As part of the data collection, they recorded the voice of three different female subjects. The subjects were asked to express certain emotions when their speech was recorded. The subjects were Russians and they spoke Russian words under different emotional states. A mobile phone was used to record the speech and was kept at a distance about 15cms away from the mouth. The experiments were conducted in an ordinary bedroom having an area of 25m2. For extracting features from the recorded speech segments, MATLAB functions were used.

2.7 MULTIMODAL EXPRESSIONS OF STRESS DURING A PUBLIC SPEAKING TASK COLLECTION, ANNOTATION AND GLOBAL ANALYSES

Databases of spontaneous multimodal expressions of affective states occurring during a task are few proposed by Tom Giraud et al. [29]. This research presents a protocol for eliciting stress in a public speaking task. Behaviors of 19 participants were recorded via a multimodal setup including speech, video of the facial expressions and body movements, balance via a force plate, and physiological measures. Questionnaires were used to assert emotional states, personality profiles and relevant coping behaviors to study how participants cope with stressful situations. Several subjective and objective performances were also evaluated. Results show a significant impact of the overall task and conditions on the participants' emotional activation. The possible future use of this new multimodal emotional corpus is described.

19 participants were recruited from University of Paris-Sud (male n=7, 37%; female n=12, 63%). 7 of the participants were doctoral students (37%), 11 were master students (58%), and one was undergraduate student (5%). The average age of participants was 26 years (SD= 6.1). All participants were volunteers and signed an informed consent designed in collaboration with the administrative heads of the partners' laboratories.

Researchers selected several personality questionnaires which feature potentially relevant dimensions for stress studies. they considered the personality profiles that might have a positive impact on the performance (e.g. extroversion, agreeableness, conscientiousness and functional copings), but also those that might be unfavorable for the performance (e.g. neuroticism, alexithymia, trait anxiety, vulnerable narcissism, and dysfunctional copings). they selected the following questionnaires: The Big Five, the State Trait Anxiety Inventory, the Toronto Alexithymia Scale and the Hypersensitive Narcissism Scale. The Big Five is the most widely used and extensively researched model of personality . It is a hierarchical model of personality traits with five broad factors: Extraversion (E), Agreeableness (A), Conscientiousness (C), Neuroticism (N), and Openness to experience (O). The French version (Big Five Inventory Francais; (BFI-Fr)) is a 45-item was evaluated by The Spielberger's State Trait Anxiety Inventory; STAI (French version). The trait subscale includes 10 items on a 4-point Likert scale (STAI-T) which assess the feelings of stress and worry on a day to day basis. Alexithymia was measured by the 20- item Toronto Alexithymia Scale (TAS-20; French version). It is an inventory consisting of 20 items on a 5-point Likert scale which assess a general deficit in experiencing and processing emotions. The scale measures three dimensions of the construct: Difficulty Identifying Feelings (DIF), Difficulty Describing Feelings (DDF) and External Oriented Thinking (EOT). Vulnerable narcissism, also referred to as covert narcissism, was measured by the HSNS . It consists of 10 affirmations on a 5-point Likert scale.

2.8 PRESENTATION TRAINER, YOUR PUBLIC SPEAKING MULTIMODAL COACH

Jan Schneider et al. [30] present the Presentation Trainer is a multimodal tool designed to support the practice of public speaking skills, by giving the user real-time feedback about different aspects of her nonverbal communication. It tracks the user's voice and body to interpret her current performance. Based on this performance the Presentation Trainer selects the type of intervention that will be presented as feedback to the user. This feedback mechanism has been designed taking in consideration the results from previous studies that show how difficult it is for learners to perceive and correctly interpret real time feedback while practicing their speeches. In this paper researchers present the user experience evaluation of participants who used the Presentation Trainer to practice for an elevator pitch, showing that the feedback provided by the Presentation Trainer has a significant influence on learning.

A key factor to achieve "perfect practice" required for the development and improvement skills is feedback, which has

also been identified as one of the most influential interventions in learning .Having a human tutor providing us with high quality feedback whenever they have time to practice our skills is neither an affordable nor a feasible solution. In our effort to study an affordable solution for this feedback availability challenge, they explored the topic 'public speaking skills'. Where they followed a design-based research methodology developing different prototypes of the Presentation Trainer (PT). The PT is an example of an automated feedback tool that tracks the learners' voice and body. It provides them with feedback about their nonverbal communication, with the purpose to support them with the development of their public speaking skills.

In this article researchers describe the current version of the PT, and present the user experience evaluation of a study, where participants had to prepare themselves for an elevator pitch. This study followed a quasi-experimental set-up where they explored the learning effects of the feedback provided by the PT.

2.9 RECOGNITION OF HUMAN EMOTION FROM A SPEECH SIGNAL BASED ON PLUTCHIK'S MODEL

Machine recognition of human emotional states is an essential part in improving man-machine interaction proposed from Dorota Kami´nska and Adam Pelikant [31]. During expressive speech the voice conveys semantic message as well as the information about emotional state of the speaker. The pitch contour is one of the most significant properties of speech, which is affected by the emotional state. Therefore, pitch features have been commonly used in systems for automatic emotion influence on pitch features have been studied. This understanding is important to develop such a system. Intensities of emotions are presented on Plutchik's cone-shaped 3D model. The k Nearest Neighbor algorithm has been used for classification. The classification has been divided into two parts. First, the primary emotion has been detected, then its intensity has been specified. The results show that the recognition accuracy of the system is over 50% for primary emotions, and over 70% for its intensities.

2.10 SELF-SPEECH EVALUATION WITH SPEECH RECOGNITION AND GESTURE ANALYSIS

S. Shangavi et al. [32] present MPIIEmo system for identifying the emotions of a person which are relevant to the body movements. This system is required more than four room. Therefore, this is not a portable device. Classification is used for identifying facial emotions the system is not handled by one person for an evaluation. Help is needed when a person is evaluating in our proposed system, the product minimizes above drawbacks of MPIIEmo and gives more reliable usages to identify the gestures and movements evaluated by the speaker. MPIIEmo is limited to space, hence Researchers decided to develop their system in a Mobile application where it is portable and usable at any time. Accordingly, they are using Android Studio, OpenCV with Classification algorithms and Android Sensors In the initial developing process, the proposed system will identify physical gestures, movements and positions such as lifting hands, waving hands, position of hands, whether they are in back or crossed and movements of body. These small gestures will be very useful to make the self-evaluator aware of what type of gestures they made and whether they have made any movements during their speech. Therefore, by generating a report that contains how many times the speaker made appropriate gestures and inappropriate gestures, they can correct any mistakes by themselves. In each appropriate and not appropriate categorization, the type of gestures and movements are classified, so that it would be easy for the speaker to correct inappropriate gestures next time.

2.11 VOICE EMOTION RECOGNITION USING CNN AND DECISION TREE

The use of decision tree and CNN as classifier to classify the emotions from the English and Kannada audio data has propose from N. Damodar et al. [33]. The performance of CNN and DT are potential for various emotions. Comparative study of the classifiers using various parameters is presented. The performance of CNN has been identified as the best classifier for emotion recognition. Emotions are recognized with 72% and63% accuracy using CNN and Decision Tree algorithms respectively. MFCC features are extracted from the audio signals and Model is trained, tested and evaluated accordingly by changing the parameters. Speech Emotion Recognition system is useful in psychiatric diagnosis, lie detection, call Centre conversations, customer voice review, voice messages. To achieve this work in this paper, features are extracted using Mel-frequency cepstrum coefficients (MFCC) and classified using Decision Tree and Convolutional Neural network.

3. CONCLUSION

In this paper, we are presenting a survey of information technology applications to treat the fear of public speaking. There are many types of applications were used different methods to treat fear of public speaking that are presented in this survey. which were many types of methods used that have yielded satisfactory results. Through our in-depth study,

we note that some of these methods need to be further developed in order to obtain accurate results. We suggest that data mining techniques that are more appropriate be used in order to avoid defects, one research [31] used the KNN algorithm and that the main disadvantage of this algorithm is that it is a lazy learner, that is, it does not learn anything from the training data and simply uses the same training data for classification. This can be addressed using the deep learning algorithm for better and accurate results.

FUNDING

None

ACKNOWLEDGEMENT

None

CONFLICTS OF INTEREST

The author declares no conflict of interest.

REFERENCES

[1] F. Dermody, A. Sutherland, and M. Farren *A Multi-modal System for Public Speaking Pilot Study on Evaluation of Real-Time Feedback*, vol. 1, pp. 499–501, 2015.

[2] M. Chollet, W. Torsten, L. Morency, and S. Scherer *A Multimodal Corpus for the Assessment of Public Speaking Ability and Anxiety*, pp. 488–495.

[3] O. Bălan *Emotions classification based on biophysical signals and machine learning techniques*, 2019.

[4] W. Torsten and S. Scherer *Automatic Assessment and Analysis of Public Speaking Anxiety : A Virtual Audience Case Study*.

[5] F. Moldoveanu *Dimensions and Machine Learning Techniques*, pp. 1–18, 2019.

[6] *Greatest.coaches.in.NFL.History*, 2015.

[7] J. Hattie and & H Timperley, "The power of feedback"," *Review of Educational Research Journal*, pp. 81–112, 2007.

[8] T. Anderson and J. Shattuck, "Design-Based Research A Decade of Progress in Education Research?," *Educational Researcher Journl*, pp. 16–25.

[9] D. Kerby and & J Romine, "Develop Oral Presentation Skills Through Accounting Curriculum Design and Course-Embedded Assessment," *Business*, vol. 85, pp. 172–179, 2009.

[10] P. King, J. Young, and & R Behnke, "Public speaking performance improvement as a function of information processing in immediate and delayed feedback interventions," *Communication Education Journal*, vol. 49, pp. 365–374, 2000.

[11] "Every Toastmasters Journey Starts with A Single Speech," *Available: [Accessed*, 2017.

[12] K. Lee, H. Hon, and R. Reddy, "An overview of the SPHINX speech recognition system," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, pp. 35–45, 1990.

[13] J. Goodman, *A bit of progress in language modeling*. 2001.

[14] W. Walker, P. Lamere, P. Kwok, B. Raj, R. Singh, E. Gouvea, P. Wolf, and J. Woelfel, "Sphinx-4: A flexible open source framework for speech recognition," *Sun Microsystems Inc*, 2004.

[15] . . Microsoft and B. Speech 2017.

[16] G. Bohouta, *Comparision Speech Recognition Systems*. 2017.

[17] 2017.

[18] . . Spellchecker 2017.

[19] " Mathworks, Matlab, and Mathworks 2017.

[20] D. Wood, "Sound: Definition, Influences, Pitch & Volume," *Study.com*, 2017.

[21] O. C. M. L. Six *TarsosDSP, a Real-Time Audio Processing*, 2014.

[22] A. D. Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *The Journal of the Acoustical Society of America*, vol. 111, 2002.

[23] P. Mcleod and G. Wyvill, "A Smarter Way to Find Pitch," *Proceedings of the International Computer Music Conference (ICMC 2005)*, 2005.

[24] M. J. Ross, H. L. Shaffer, A. Cohen, R. Freudberg, and H. J. Manley, "Average Magnitude Difference Function Pitch Extractor," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 22, pp. 353–362, 1974.

[25] P. Jahoda *MPAndroidChart*, 2017.

[26] "microsoft Cognitive Services 2017.

[27] L. Kerkeni, Y. Serrestou, M. Mbarki, K. Raoof, M. A. Mahjoub, and C. Cleder *Automatic Speech Emotion Recognition Using Machine Learning, Social Media and Machine Learning, Alberto Cano, IntechOpen*, 2019.

[28] A. Davletcharova, S. Sugathan, B. Abraham, and A. P. James, "Detection and Analysis of Emotion from Speech Signals," *the Second International Symposium on Computer Vision and the Internet*, vol. 58, 2015.

[29] T. Giraud, "Multimodal Expressions of Stress during a Public Speaking Task: Collection, Annotation and Global Analyses," in *Humaine Association Conference on Affective Computing and Intelligent Interaction*, pp. 417–422, 2013.

[30] J. Schneider, D. Börner, P. V. Rosmalen, and M. Specht, "Presentation Trainer, your Public Speaking Multimodal Coach," in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (ICMI '15)*, pp. 539–546, Association for Computing Machinery.

[31] D. K. nska and A. Pelikant, "Recognition of Human Emotion from a Speech Signal Based on Plutchik's Model"," *International Journal of Electronics and Telecommunications*, vol. 58, no. 2, pp. 2300–1933, 2012.

[32] S. Shangavi, S. Jeyamaalmarukan, A. Jathevan, M. Umatharsini, and P. Samarasinghe, "Self-Speech Evaluation with Speech Recognition and Gesture Analysis," in *2018 National Information Technology Conference (NITC)*, pp. 1–7, 2018.

[33] N. Damodar, H. Y. Vani, and A. M. A, "Voice Emotion Recognition using CNN and Decision Tree," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, pp. 2278–3075, 2019.

[34] V. A. Petrushin and B. Grove *Detecting emotions using voice signal analysis*, 2007.

[35] T. Giraud, J. Hua, and A. Delaborde, "Multimodal Expressions of Stress during a Public Speaking Task: Collection, Annotation and Global Analyses," in *Humaine Association Conference on Affective Computing and Intelligent Interaction*, pp. 417–422, 2013.

[36] J. Schneider, D. Boerner, P, and M. Specht *Presentation Trainer, your Public Speaking Multimodal Coach*.

[37] S. Shangavi, S. Jeyamaalmarukan, A. Jathevan, and M *Umatharsini Pradeepa Samarasinghe, Self-Speech Evaluation with Speech Recognition and Gesture Analysis*, 2018.