

# SPEECH EMOTION RECOGNITION USING MACHINE LEARNING

D. Shalini Sharma and Mrs N.Gowri

Final year PG Student, Assistant Professor,

Department Of Information Technology,

G. Venkataswamy Naidu College, Kovilpatti, Tamil Nadu, India

## ABSTRACT

Communication is the key to express one's thoughts and ideas clearly. Amongst all forms of era of the Internet of Things (IoT) is rapidly advancing in bringing more intelligent systems available for everyday use. These applications range from simple wearables and widgets to communication, speech is the most preferred and powerful form of communications in human. The complex self-driving vehicles and automated systems employed in various fields. Intelligent applications are interactive and require minimum user effort to function, and mostly function on voice-based input. This creates the necessity for these computer applications to completely comprehend human speech. A speech percept can reveal information about the speaker including gender, age, language, and emotion. Several existing speech recognition systems used in IoT applications are integrated with an emotion detection system in order to analyze the emotional state of the speaker. The performance of the emotion detection system can greatly influence the overall performance of the IoT application in many ways and can provide many advantages over the functionalities of these applications. This project presents a speech emotion detection system with improvements over an existing system in terms of data, feature selection, and methodology that aims at classifying speech percepts based on emotions, more accurately. The developed data model is linked with android app which helps to recognize the speech from anywhere through android mobile. As per the variation in tone, the value get detected and mapped via mobile app.

Keywords: Emotion recognition, Feature extraction, Emotions, Modelling, Machine Learning , deep neural network, Dataset.

\*\*\*\*\*

## I INTRODUCTION

One of the fastest and natural methods of communication between humans is a speech signal. For interaction between human and machine use of speech signal is the fastest and most efficient method. To maximum awareness of

received message, all available senses are used by human's natural ability. For machine emotional detection is a very difficult task, on the other hand, it is natural for humans. So, knowledge related to emotion is used by an emotion recognition system

in such a way that there is an improvement in communication between machine and human. The female or male speakers emotions find out through speech in speech emotion recognition. Linear prediction cepstrum coefficient (LPCC), Fundamental frequencies and Mel+ frequency cepstrum coefficient (MFCC) are some of the studied speech features. These features make a base for speech processing. In differentiating between various emotions particularly speech features are more useful is not clear is the reason that makes emotion recognition from speakers' speech very difficult. There is an introduction of accosting variability due to the existence of different speaking rates, styles, sentences and speakers that affects the features of speech. Different emotions may be shown by the same utterance and there are different portions of the spoken utterance of each correspond emotion that makes it difficult to differentiate these portions of utterance .The emotion expression depends on the culture and environment of the speaker that creates another problem as there is variation in the style of speaking by the variations in environment and culture. Transient and long terms emotion are two types of emotions and it is not clear about the type of emotion detected by recognizer. Speech information recognized emotions may be speaker independent or speaker dependent. Various classifiers like K-nearest neighbors (KNN), Support vector machine (SVM), CNN, etc are available for classification. In this paper brief

introduction about speech emotion recognition is given along with the speech emotion recognition system block diagram description in the second section of the paper. Various work has been done on different datasets so some of the existing datasets are covered in the third section along with modeling of emotions speech and different types of speech. The fourth section gives brief details about various feature extraction mechanism for speech emotion recognition and then focus is given on review on the classification part. In this section, we have covered KNN, SVM, CNN, recurrent neural network, etc. The sixth section gives brief about the use of deep learning for speech emotion recognition.

## **II RELATED WORK**

There is vigorous debate about what exactly individual can express nonverbally. Humans can express their emotions through many different types of nonverbal communication including facial expressions, quality of speech produced, and physiological signals of the human body. In this section, we discuss each of these categories.

### **A.FACIAL EXPRESSIONS**

The human face is extremely expressive, able to express countless emotions without saying a word. And unlike some forms of nonverbal communication, facial expressions are universal. The facial expressions for happiness,

sadness, anger, surprise, fear, and disgust are the same across cultures.

## **B. SPEECH**

In addition to faces, voices are an important modality for emotional expression. Speech is a relevant communicational channel enriched with emotions: the voice in speech not only conveys a semantic message but also the information about the emotional state of the speaker. Some important voice feature vectors that have been chosen for research such as fundamental frequency, mel-frequency cepstral coefficient (MFCC), prediction cepstral coefficient (LPCC), etc.

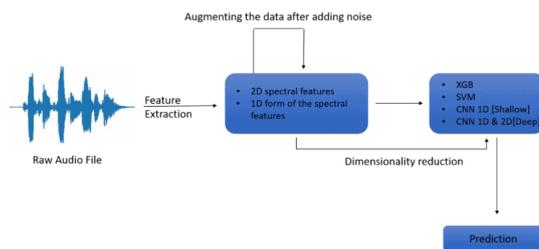
## **C. PHYSIOLOGICAL SIGNALS**

The physiological signals related to autonomic nervous system allow to assess objectively emotions. These include electroencephalogram (EEG), heart rate (HR), Social Media and Machine Learning electrocardiogram (ECG), respiration (RSP), blood pressure (BP), electromyogram (EMG), skin conductance (SC), blood volume pulse (BVP), and skin temperature (ST). Using physiological signals to recognize emotions is also helpful to those people who suffer from physical or mental illness thus exhibit problems with facial expressions or tone of voice.

## **III. PROPOSED WORK**

Identifying the emotion expressed in a speech percept has several use cases in the modern day applications. Human-Computer Interaction (HCI) is a field of research that studies interactive applications between humans and computers. For an effective HCI application, it is necessary for the computer system to understand more than just words. On the other hand, the field of Internet of Things (IoT) is rapidly growing. Many real world IoT applications that are used on a daily basis such as Amazon Alexa, Google Home and Mycroft function on voice-based inputs. The role of voice in IoT applications is pivotal. The study in a recent article foresees that by 2022, about 12% of all IoT applications would fully function based on voice command System mated voice calls can be efficiently transferred to customer service agents for further discussion. Other applications of using a speech emotion detection system can be found in lie detecting systems, criminal department analysis, and in humanoid.s only. These voice interactions could be monodirectional or bi-directional, and in both cases, it is highly important to comprehend the speech signal. Further, there are Artificial Intelligence (AI) and Natural Language Processing (NLP) based applications that use functions of IoT and HCI to create complex systems. Self-driving cars are one such application that controls many of its functions using voice-based commands. Identifying the emotional state of the user comes

with a great advantage in this application. Considering emergency situations in which the user may be unable to clearly provide a voice command, the emotion expressed through the user's tone of voice can be used to turn on certain emergency features of the vehicle. A much simpler application of speech emotion detection can be seen in call centers, in which automated voice calls can be efficiently transferred to customer service agents for further discussion. Other applications of using a speech emotion detection system can be found in lie detecting systems, criminal department analysis, and in humanoids.



#### IV CONCLUSION

In the paper brief introduction about speech emotion recognition is given along with the speech emotion recognition system block diagram description. In the field of affect detection, a very important role is played by a suitable choice of speech database. For good emotion recognition system mainly three databases are used. On the basis of ability, they have to recognize a speech recognition system can be separated in different classes are isolated, connected, spontaneous and continuous words. Relevant emotional features extraction from the speech is the second important

step in emotions recognition. To classify features there is no unique way but preferably acoustic and linguistic features taxonomy is considered separately. There are a number of methods for feature extraction like Linear predictive cepstral coefficients (LPCC), Power spectral analysis (FFT), First order derivative (DELTA), Linear predictive analysis (LPC), Mel scale cepstral analysis (MEL), perceptual linear predictive coefficients (PLP) and Relative spectra filtering of log domain coefficients (RASTA) and some of them are briefly covered in this paper. Another important part of speech emotion recognition system is the use of classifier. In the paper, the detailed review on KNN, SVM, CNN, Naive Bayes, and recurrent neural network classifier for speech emotion recognition system. The last section of the paper covers the review on the use of the deep neural network to make speech emotion recognition system. To further improve the efficiency of system combination of more effective features can be used that enhances the accuracy of speech emotion recognition system.

#### V REFERENCE

1. "Speaker Independent Connected Speech Recognition- Fifth Generation Computer Corporation". Fifthgen.com. Archived from the original on 11 November 2013. Retrieved 15 June 2013.
2. P. Nguyen (2010). "Automatic classification of speaker characteristics". International Conference on Communications and

Electronics 2010. pp. 147–152.  
doi:10.1109/ICCE.2010.5670700. ISBN 978-1-4244-7055-6. S2CID 13482115.

3. "British English definition of voice recognition". Macmillan Publishers Limited. Archived from the original on 16 September 2011. Retrieved 21 February 2012.

4. "voice recognition, definition of". WebFinance, Inc. Archived from the original on 3 December 2011. Retrieved 21 February 2012.

5. "The Mailbag LG #114". Linuxgazette.net. Archived from the original on 19 February 2013. Retrieved 15 June 2013.

6.Y. Kumar, N. Singh, —Automatic Spontaneous Speech Recognition for Punjabi Language Interview Speech Corpus, I.J. Education and Management Engineering, pp. 64-73, 2016.

7. C. M. Lee, S. S. Narayanan, —Toward detecting emotions in spoken dialogs, IEEE Transactions on Speech and Audio Processing, pp. 293-303, 2005.