

SPEECH EMOTION RECOGNITION USING MACHINE LEARNING

A PROJECT REPORT

Submitted by

VYOM DALAKOTI(19BSC1001)

AKSHITA DHIMAN(19BSC1003)

JOSHUA CHINGANGBAM(19BSC1005)

IN

BACHELOR'S OF SCIENCE IN COMPUTER SCIENCE



**CHANDIGARH
UNIVERSITY**
Discover. Learn. Empower.

Chandigarh University

FEBRUARY, 2022

I. Introduction to the Project

In today's world of Siri and Alexa, there is an increasing need for computers to understand human emotions and speech. Home automation systems such as Alexa or google home listen to our voices and extract meaning and commands out of them. An often under looked part of this interaction is how they also listen to minor intonations and inflections in our speech in order to better understand our 'language', this is a part of Natural Language Processing or NLP in Artificial Intelligence.

Speech emotion recognition belongs to a larger group in computing referred to as Affective Computing. It revolves around systems in computing that interpret human affects and then analyze or simulate them. Affective computing

Human affect is a field in psychology that refers to human emotions, feelings and mood. It has three major principals i.e. **Valence** which refers to then good-bad spectrum of a perceived event or situation, **Arousal** which is the activation of the Sympathetic nervous system that can be measured objectively and finally **Motivational intensity** which is the process which compels or motivates an individual to take some action on any particular event.

In this project we will be using machine learning and Natural language processing to analyze human speech and classify it. There are two main type of classifications for human speech, continuous and what we will be using for this project, Categorical. We will create neural networks that analyze audio of human speech and then categorize it into different emotions i.e. angry, happy, sad etc. This will allow us to understand human speech and is a step towards giving computers emotional intelligence, which make computers capable of understanding human emotions and personalize their behavior accordingly. There are various machine learning algorithms employed in creating this project such as discriminant classifiers (LDC), k-nearest neighbor (k-NN), Gaussian mixture model (GMM), support vector machines (SVM), artificial neural networks (ANN), decision tree algorithms and hidden Markov models (HMMs) depending on our need and specific classifier is used for different needs to achieve the highest accuracy

II. Client Identification and Recognition of need.

Speech emotion recognition is useful in many fields such as Home automation, medicine, marketing, education, human psychology and many further niche uses.

- **Home automation:**

Home automation devices such as Alexa or Google home use SER (Speech Emotion Recognition) in order to have natural conversations with human beings. These devices use SER to analyze the user's emotions and give a proper response.

Furthermore, using SER these devices can be programmed to use proper tone, inflections and speech patterns in order to properly imitate a human conversation and convey certain emotions to give the users the feeling of conversation with a 'person' and not a robot.

- **Marketing.**

SER has plenty of uses in the field of marketing. It is about more than just touting speeds and feeds, features, and benefits.

Marketing involves eliciting an emotional response from consumers in the hopes that the product, service, or brand will resonate deeply. SER can be useful in gauging customers response using video and audio data and create marketing strategies accordingly.

- **Health:**

In today's world of wellbeing apps and digital assistants SER can be utilized in understating the user's emotions better in order to give a more personalized response than a robotic one. Plus, Affective computing is useful in medicine in analyzing and monitoring patient's psychological health.

- **Education:**

Certain learning apps exist that require only the interaction between a learner and a computer. For example, there are many language learning apps that ask you to give vocal or written responses in order to learn a language and imitate the tone and

pattern in that particular language, using SER these apps can properly analyze the speech and judge the learner's tasks much more accurately.

III. Recognition & knowledge of relevant contemporary issues.

Understanding of human affects by computers is of an ever-rising need ever since COVID and the lockdowns many interactions that occurred in real life between humans have to be digitized and there is a middle layer of technological devices between human interactions as everything eventually moves to the cloud and gets into our smartphones and computers. A small example of the need for this technology is something we've all experienced online in the form of feedback forms. Many services ask us to give our text or speech-based feedback, however ever increasingly they just enquire "how does this make you feel?" and take emotions as input, with SER this emotion-based feedback can also be taken using speech which will be very helpful with marketing and other service-based fields. Furthermore, there are many people who for one reason or another cannot communicate very expressively either due to a physical disability such as blindness or a learning disability, for them this type of technology can bridge the gap towards seamless communication with computers.

SER technology, like many other emerging cutting-edge technologies is prone to being used in many nefarious ways. There is certainly a use of this technology in things such as surveillance and other invasions of privacy. These technologies alongside other sub-fields of affective computing can be used to monitor people's emotions without their consent or in a larger-scale lead to mass surveillance by governments or other institutions that unfortunately already takes place in the world over.

IV. Project Identification

The Basically, Emotion Recognition deals with the study of inferring emotions, methods used for inferring. Emotion can be recognized from facial expressions, speech signals. Various techniques have been developed to find the emotions such as signal processing, machine learning, neural networks, computer vision. Emotion analysis, Emotion

Recognition are being studied and developed all over the world. Emotion Recognition is gaining its popularity in research which is the key to solve many problems also makes life easier.

The main need of Emotion Recognition from Speech is challenging tasks in Artificial Intelligence where speech signals is alone an input for the computer systems. Speech Emotion Recognition (SER) is also used in various fields like BPO Centre and Call Centre to detect the emotion useful for identifying the happiness of the customer about the product, IVR Systems to enhance the speech interaction, to solve various language ambiguities and adaption of computer systems according to the mood and emotion of an individual.

A Need for inferring emotion from spoken utterances increases exponentially. Since there is an enormous development in the field of Voice Recognition. There are many voice products has been developed like Amazon Alex, Google Home, Apple Home Pod which functions mainly on voice-based commands. It is evident that Voice will be the better medium for communicating to the machines.

Emotion recognition is the part of speech recognition which is gaining more popularity and need for it increases enormously. Although there are methods to recognize emotion using machine learning techniques, this project attempts to use deep learning method to recognize emotion and classify the emotion according to the speech signals.

V. Task Identification

In this project, we have discovered all the tasks that we need to completed and perform, before starting the project in the processing period.

- As it is a machine learning project, the first and foremost task is collecting the database on which we'll perform our machine learning algorithm.
- Then comes the feature extraction process, in which we will classify the audio using the visual and numerical representation of the audio. Using the **Librosa package** we will extract all the features present in the audio, such as, Mel-spectrogram, Mel-Frequency Cepstral Coefficients, Chroma frequencies.
- Now we will split the data set and filter it into two parts, 80% of it for the training and the 20% for the testing.

- The main task, where we will train our model using machine learning algorithms will be executed now.
- In the last we will test our model and resolve the problems if any arrives.

VI. Timeline of the project

Considering all the tasks and time, we have mapped out the best plan to execute our project. We have assigned the best task fit to all the team members and considered that no one will feel left alone.

	February	March		April	May
19BSC1001	Project title selection	collecting dataset	machine learning model creation	working on final report	testing the model
19BSC1003	Project title selection	research on project	report making	working on GUI	working on final report
19BSC1005	Project title selection	research on project	feature extraction	data processing	working on final report
					finalization

Fig 1. Timeline of the project

VII. Gantt chart

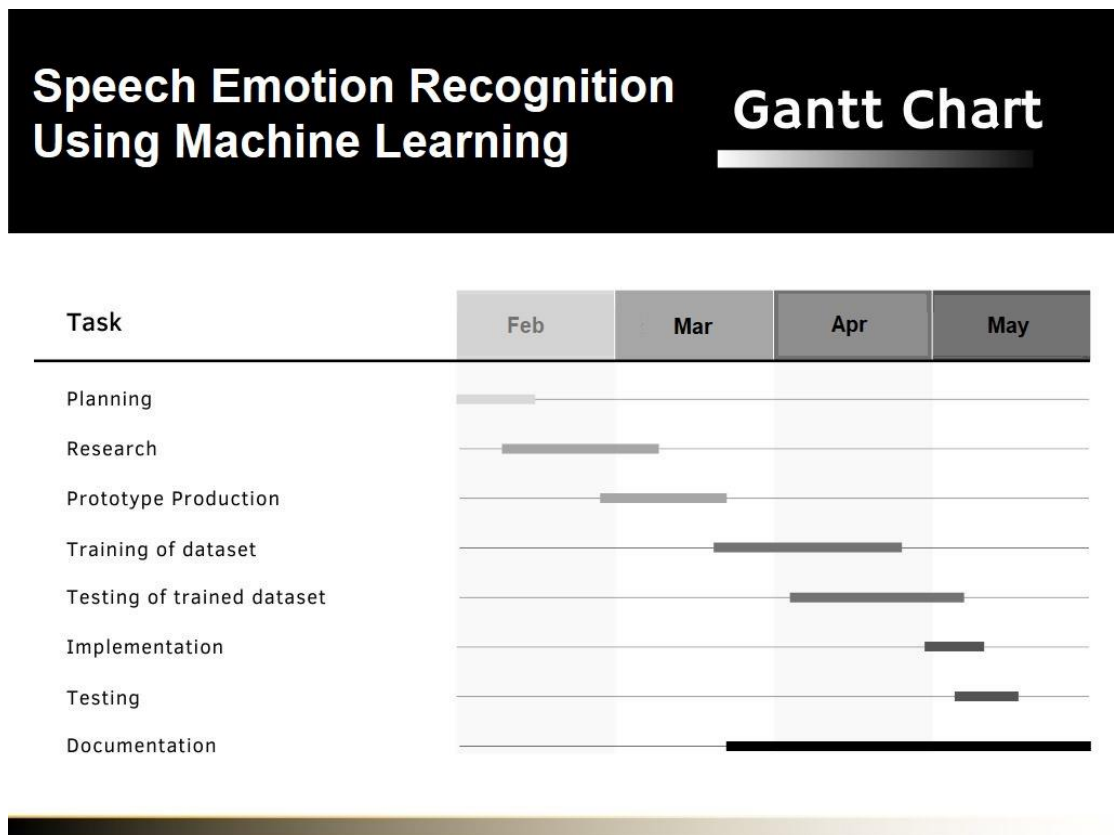


Fig 2. Gantt chart