

# Speech Emotion Recognition Using Machine Learning Algorithm

Sayali Borkar<sup>1</sup> Raj Anand Singh<sup>2</sup> Sweety Gaikwad<sup>3</sup> Pranita Chikte<sup>4</sup>

<sup>1,2,3,4</sup>Department of Computer Technology

<sup>1,2,3,4</sup>YCCE, Nagpur, Maharashtra, India

**Abstract**— The field of text mining has evolved over the past few years to assist, analyze the immense quantity of matter resources available on-line. Text Mining, however, may be used conjointly in various different applications. During this analysis, we are significantly interested in playing text mining techniques over transcribed audio recordings so as to find the speakers' emotions. Our work is originally actuated by use cases arising from decision centers, however can even have applications in different areas. We describe our overall methodology and gift our experimental results for speech-to-text transcription, text classification and text clump. We have a tendency to conjointly concentrate on analyzing the consequences of mistreatment different options choice way.

**Key words:** Text Mining, speech-to-text transcription

## I. INTRODUCTION

Speech may be an advanced signal containing information concerning message, speaker, language, feeling so on. Most existing speech systems method recognizes studio recorded, neutral speech effectively, however their performance is poor within the case of emotional speech. This can be because of the issue in modeling and characterization of emotions in speech. Presence of emotions makes speech a lot of natural. During a language, non-verbal communication carries vital info like intention of the speaker. Additionally to the message sent through text, the style within which the words square measure spoken conveys essential non-linguistic information. Constant matter message would be sent with totally different linguistics (meaning) by incorporating applicable emotions. Spoken text could have many interpretations, betting on however it's same.

For instance, the word 'OKAY' in English is employed to precise admiration, disbelief, consent, tolerance or assertion. Thus understanding the text alone isn't sufficient to interpret the linguistics of a spoken vocalization. However, it's vital that, speech systems ought to be ready to method the non-linguistic info like emotions, at the side of the message. Humans perceive the meant message by perceiving the underlying emotions additionally to phonetic info by victimization multi-modal cues.

Non-linguistic info is also ascertained through

- 1) Facial expressions within the case of video,
- 2) Expression of emotions within the case of speech,
- 3) Punctuation within the case of written communication.

The discussion during this paper is confined itself to emotions or expressions associated with speech.

Basic goals of emotional speech process square measure -

- a) Understanding emotions present in speech and
- b) Synthesizing desired emotions in speech consistent with the meant message.

From machine's perspective understanding speech emotions is viewed as classification or discrimination of emotions. Speech is one in every of the natural modalities of human machine interaction. Today's speech systems could

reach human equivalent performance only; they'll method underlying emotions effectively. Purpose of subtle speech systems mustn't be restricted to mere message processing; rather they must perceive the underlying intentions of the speaker by detection expressions in speech. Within the recent past, process speech signal for recognizing underlying emotions is emerged jointly of the vital speech analysis areas. Embedding the part of feeling process into existing speech systems makes them a lot of natural and effective. Therefore, whereas developing speech systems, one ought to fitly utilize the data of emotions. Speech feeling recognition has many applications in everyday life.

Emotion recognition system is also employed in an aboard automotive driving system, wherever info concerning status of a driver is also accustomed keep him alert throughout driving. This helps avoiding some accidents, caused because of stressed status of the motive force. Center language is also accustomed to analyze activity study of decision attendants with their customers, and helps to enhance quality of service of a decision attendant. Medical doctors could use emotional contents of a patient's speech as a designation tool for varied disorders. Feeling analysis of conversation between criminals would facilitate crime investigation department for the investigation. Language with robotic pets and automaton partners would be a lot of realistic and pleasant, if they're ready to perceive and categorical emotions like humans do. Automatic feeling analysis is also helpful in automatic speech to speech translation systems, wherever speech in language x is translated into alternative language y by the machine. Here, each feeling recognition and synthesis square measure is used. The emotions gift in speech square measure recognition, and also the same emotions square measures tend to be synthesized within the target speech, because the translated speech is anticipated to represent the spirit of the initial speaker.

The word feeling is inherently unsure and subjective. The term feeling has been used with totally different discourse meanings by different folks. It's troublesome to outline feeling objectively, because it is a private status that arises impromptu instead of through aware effort. Therefore, there's no common objective definition and agreement on the term feeling. This can be the elemental hurdle to proceed with scientific approach toward analysis. Most emotional speech systems square measure developed victimization full blown emotions, however world emotions square measure pervasive and underlying in nature. Some databases square measure recorded victimization full-fledged artists, whereas another square measure recorded victimization semi-experienced or inexperienced subjects.

The analysis on feeling recognition is restricted to 5–6 emotions, as most databases don't contain wide selection of emotions feeling recognition systems developed victimization varied options is also influenced by the speaker and language dependent info. Ideally, speech feeling

recognition systems ought to be speaker and language freelance. A vital issue within the development of speech feeling recognition systems is identification of appropriate options that with efficiency characterize totally different emotions.

Along with options, appropriate models square measure to be known to capture feeling specific info from extracted speech options. Speech feeling recognition systems ought to be sturdy enough to method real-life and streaky speech to spot emotions.

This paper provides a review of literature on speech feeling recognition, visible of various forms of feelingly speech corpora accustomed develop the emotion recognition systems, feeling specific options extracted from totally different aspects of speech, classification models used for recognizing the emotions. Some directions for more analysis on speech feeling recognition also are mentioned at the top of the paper.

## II. OVERVIEW

Speech emotion recognition aims to mechanically establish the spirit of a person's being from his or her voice. It supports in-depth analysis of the generation mechanism of speech signal, extracting some options that contain emotional information from the speaker's voice, and taking applicable pattern style to acknowledge the emotions, recognition strategies to spot emotional states. Like typical pattern recognition systems, our speech feeling recognition system contains four main modules: speech input, feature extraction, SVM primarily based classification, and feeling output. The overall design for SER system-

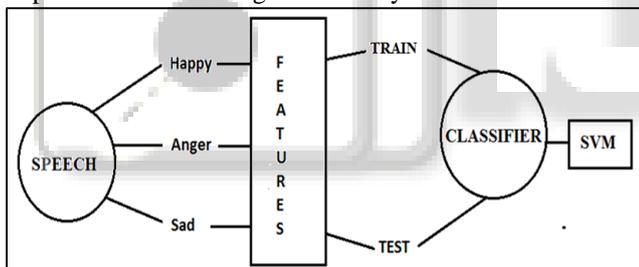


Fig. 1: Overall design for SER system

## III. DATA SET

For our dataset we have taken wav file, which is having different audio emotions fluctuating from happy to sad and other considerable emotions. One thing that should be taken into account is that only those audio clips which are in US English ascent is recognizable by our system. So selection of audio clip is made accordingly.

## IV. METHODOLOGY

There are many different feeling sets exists within the literature together with basic emotions, universal emotions, primary and secondary emotions, neutral vs. emotional, and for a few cases the matter is reduced to a 2 category classification downside (Sentiment Analysis) exploitation the Positive and Negative values as category labels. Simple classification sets offer higher performance than dilated sets of emotions that need psychological feature data and deeper understanding of the topic. In our analysis study, we've got used five feeling categories (anger, tender, scared, sad,

happy and excited) that type the intersection between the dataset and also the SemEval check set. Therefore, the amount of feeling categories  $s=6$ .

For the classification, we've used Support Vector machines and HMM. We have considered the result of the stemming, negation and intensity of emotions on classification performance.

### A. Stop Word Removal

A study additionally showed that some words solely appeared in specific sentences happiness to one feeling, thus stop-word removal supported minimum term frequencies isn't appropriate for emotion detection. Stop words square measure typically the foremost frequent words together with articles (a, an, the), auxiliary verbs (be, am, is, are), prepositions (in, on, of, at), conjunctions (and, or, nor, when, while) that don't give further improvement for search engines however increase the procedure complexness by increasing the dimensions of the lexicon. The necessary side of stop-word removal in feeling detection is that the words, not their frequencies. There square measure many in public offered stop-word lists offered wherever these lists accommodates close to 400-500 most frequent words during a given language. However, public stop-word lists contemplate the data retrieval and that they don't consider words carrying emotional content. Thus we tend to 1<sup>st</sup> need to take away a number of the emotional words from the stop-wordlist together with negative verbs (not, is not, does not, do not, should not, etc.). Additionally, we tend to replace the word "very" with blank and the word "blank not blank" is replaced by "blank not". We tend to neglect the part of speech tagging on input text as a result of its result of reducing the classification accuracy as delineate in. Since non-alpha tokens square measure mechanically removed by TMG, the exclamation marks and question marks square measure replaced by descriptive new words "XXEXCLMARK" and "XXQUESMARK" severally. Negative short forms are replaced by negative long forms specified "doesn't" is replaced by "does not".

Initially we've got used stemming for locating morphological root of a given word. Stemmers in linguistic square measure wide employed in search engines and question primarily based systems to boost the efficiency of those systems. For feeling classification, stemming additionally removes the emotional that means from the words. We found that tense data additionally affects the emotional meaning of the words. For instance the words "marry" and "love" is often shown in joy sentences whereas the words "married" and "loved" square measure appeared in unhappy sentences.

### B. Training and Testing

During the coaching section of machine learning algorithms, a model from the received input is build. At last it provides a hypothesis operate that may be wont to predict result for the further input file. There has been an excellent increase within the use of SVM's for prime performance pattern classification.

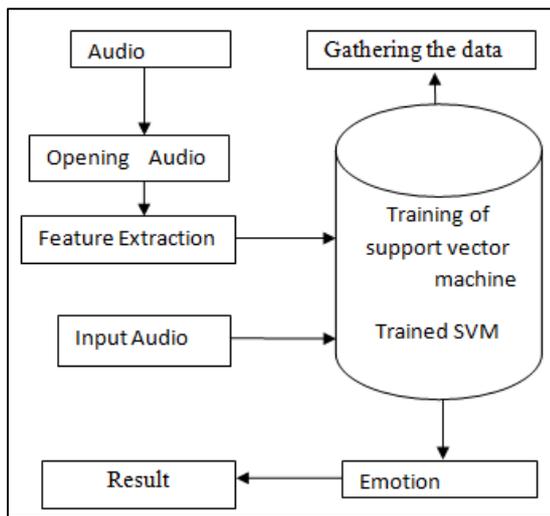


Fig. 2: Emotion detection using SVM

### V. RESULT

After doing regressive analysis following results as shown in the screenshot is generated. To be descriptive, once the data is loaded and is heard, and then the sequence vector is generated. After this string is concatenated. The concatenated string can be seen in the “concatenation string” box. Then stop words (words that do not play any role in depicting emotions) are removed.



Fig. 3: Screenshot-1

Once the check emotion button is clicked it checks for the emotion by making comparison with svm testing string. The result is generated likewise.



Fig. 4: Screenshot-2

Now considering the detail aspect of our result, we have obtained a successful depiction of result through the bar graph, pie chart and table representing the percentage outcome of respective emotion present in the dataset.

#### A. Bar Graph

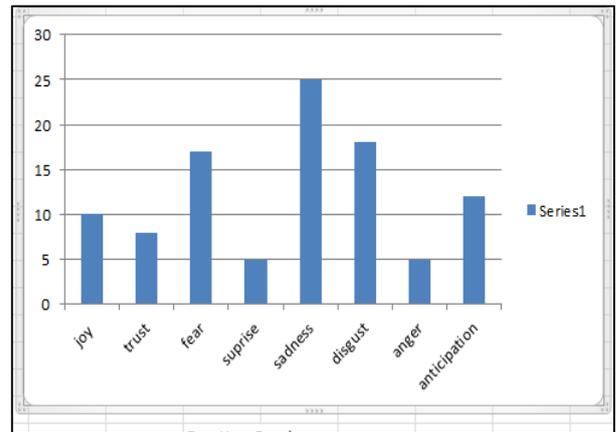


Fig. 5: Emotion Graph

The bar graph above shows the percentage of respective emotion present in the taken input dataset. The maximum being that of sadness while minimum is shared by anger and surprise.

#### B. Pie Chart

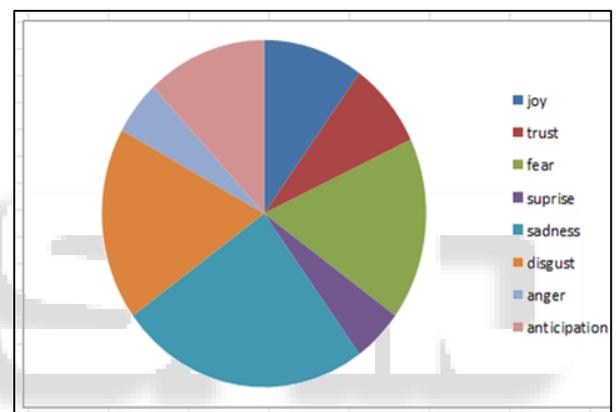


Fig. 6: Emotion Pie Chart

The obtained result is more diagrammatically represented by the above pie chart showing sadness being the key emotion present in the considered dataset.

#### C. Emotion Percentage Table

Joy	Trust	Fear	Surprise	Sadness	Disgust	Anger	Anticipation
10	8	17	5	25	18	5	12

Table 1: Emotion Percentage Table

The table above shows the percentage of various emotion present in the complete dataset. This representation helps in keeping a tabular record of our calculated data.

#### REFERENCES

- [1] K. Oatley, J.M. Jenkins, Understanding Emotions, Blackwell Publishers, Cambridge, USA, 1996.
- [2] P. Ekman. Basic emotions. In T. Dalgleish and T. Power, editors, The Handbook of Cognition and Emotion. JohnWiley & Sons, Ltd..
- [3] Philipp Michel and Rana El Kaliouby. Real Time Facial Expression Recognition in Video using Support Vector Machines. ICMI'03, November 5–7, 2003, Vancouver, British Columbia, Canada. Copyright 2003 ACM..

- [4] P. Ekman and W. Friesen. Facial Action Coding System (FACS): Manual. Consulting Psychologists Press, Palo Alto, CA, USA.
- [5] Philipp Michel. Support Vector Machines in Automated Emotion Classification.
- [6] Thorsten Joachims. Making large- scale SVM learning practical. Cambridge, MA: MIT Press, 1998.
- [7] Roget's new millennium thesaurus, 1st ed. Lexico Publishing Group, LLC (2007)
- [8] N. Jrad, M. Congedo, R. Phlypo, R. Flamary, F. Yer, A. Rakotomamonjy, sw-SVM : sensor weighting Support Vector Machines for EEG-based brain computer interfaces, Journal of Neural Engineering, Vol. 8, N. 5, 056004, 2011
- [9] A. Rakotomamonjy, Analysis of SVM regression bound for variable ranking, Neurocomputing, Vol 70, pp 1489-1491, 2007.
- [10] Chang and Lin, 2001; Chih-Chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines (version 2.3). 2001.

