IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

PREDICTING HUMAN BEHAVIOUR BY VOICE MODULATION

Chandana N [*1], Deeksha L [*2], Janani G M[*3], Komal Suthar [*4], Mrs. Swathi Darla[*5]

[*1]Student, Department of CSE, KSSEM, Bengaluru, Karnataka, India

[*2]Student, Department of CSE, KSSEM, Bengaluru, Karnataka, India

[*3]Student, Department of CSE, KSSEM, Bengaluru, Karnataka, India

[*4]Student, Department of CSE, KSSEM, Bengaluru, Karnataka, India

[*5]Assistant Professor, Department of CSE, KSSEM, Bengaluru, Karnataka, India

ABSTRACT

The emotional speech signals research has been boosted in human machine interfaces due to availability of high computation capability. There are many systems proposed in the literature to identify the emotional state through speech. Selection of suitable feature sets, design of a proper classifications methods and prepare an appropriate dataset are the main key issues of speech emotion recognition systems. The human voice is very versatile and carries a multitude of emotions. Emotion in speech carries extra insight about human actions. Through further analysis, we can better understand the motives of people, whether they are unhappy customers or cheering fans. Humans are easily able to determine the emotion of a speaker, but the field of emotion recognition through machine learning is an open research area. In this proposed project, we perform speech data analysis on speaker discriminated speech signals to detect the emotions of the individual speakers involved in the conversation. We are analyzing different techniques to perform speaker discrimination and speech analysis to find efficient algorithms to perform this task.

Keywords: Speech, MLP, emotion, Feature extraction, Feature Selection.

I. INTRODUCTION

The recognition of emotional speech aims to recognize the emotional condition of individual utterer by applying his/her voice automatically. Speech is the fast and best normal way of communicating amongst human. This reality motivate many researchers to consider speech signal as a quick and effective process to interact between computer and human. Speech emotion recognition is mostly beneficial for applications, which need human-computer interaction such as speech synthesis, customer service, education, forensics and medical analysis. Recognizing of emotional conditions in speech signals are so challengeable area for several reason. Speech processing is a unique discipline of signal processing. Study of speech signal and its processing method are the principles of speech processing. The speech processing application plays a major part in day-to-day life of commercial applications like Bank, Travel, Telecommunications and Voice Dialing. Some of the major growing applications are Language Identification, Speech Enhancement, Spoken Dialog System, Speaker Recognition and Verification, Speech Coding, Emotion and Attitude Recognition, Speech Segmentation and Labelling, Speech Recognition, Prosody, Text-to-Speech Synthesis, and Audio-Visual Signal Processing. Input speech is given to the machine which accepts the command and translates into text format known as Speech Recognition System or Automatic Speech Recognition or Computer Speech Recognition or Speech to Text. Speech recognition systems analyze and train an individual speech that exploit to tune the recognition of specific voice which produces a more accurate result. The extracted speech signal is trained by HMM model. Finally, the output result is compared with connected and continuous speech.

II. METHODOLOGY

Methodology 1: An in-built microphone in the computer system is used for taking voice as an input

Methodology 2: The Support Vector Classification Model (SVM) classifier analyses data and recognize patterns for the given input.

Methodology 3: librosa and sklearn are the python libraries to build a model using an MLP classifier for analyzing the speech and recognize human emotions.

Methodology 4: After MLP classifiers analyze the input speech, the output will be displayed to the client

III. MODELING AND ANALYSIS

The System takes the Input as an audio signal to that it performs feature extraction to enhance the feature and classifies it according to the trained data sets and detects the emotion of the audio and displays it in the form of text.

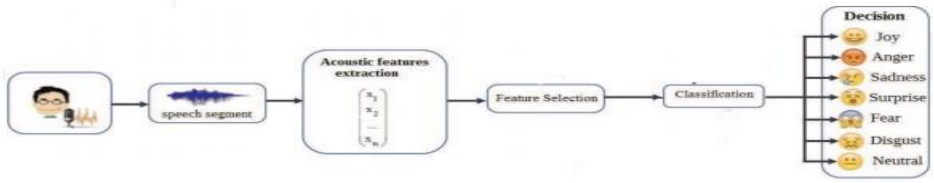

Figure 1: Design Architecture.

Audio Signal: The user speaks via the inbuilt microphone in the system, this audio signal is recognized by and the same is displayed in the form of text.

Some pre-requisite parameters required are:

- The audio signal should not have contained environment noise.
- The audio signal had to contain human speech in a form of few words or a full sentence.

Feature Extraction: Relevant emotional features extraction from speech is the second important step in emotions recognition. There are number of methods for feature extraction like , mel-frequency cestrum coefficients (MFCC),MEL Spectrogram Frequency (MEL).Mel frequency cepstral coefficients (MFCC): It is considered as one of the standard methods for feature extraction and in ASR most common is the use of 20 MFCC coefficients. Although for coding speech use of 10-12 coefficients are sufficient and it depend on the spectral form due to which it is more sensitive to noise. This problem can be overcome by using more information in speech signals periodicity although aperiodic content is also present in speech.

Feature Selection: Feature selection in ML is to "reduce the number of features used to characterize a dataset so as to improve a learning algorithm's performance on a given task." The objective will be the maximization of the classification accuracy in a specific task for a certain learning algorithm; as a collateral effect, the number of features to induce the final classification model will be reduced. Feature selection (FS) aims to choose a subset of the relevant features from the original ones according to certain relevance evaluation criterion, which usually leads to higher recognition accuracy.

TYPES OF SPEECH:

On the basis of ability, they have to recognize a speech recognition system can be separated in different classes. Following are the classification:

1. Isolated words: In this type of recognizers sample window both sides contain low pitch utterance.
2. Connected words: In this separate utterance can run together with minimal pause between them otherwise it is like isolated words.
3. Continuous words: It allows users to speak naturally and content are determined by computer.

Spontaneous words: It can be thought of as speech at basic level that is natural sounding and not rehearsed.

Many machine learning algorithms have been used for discrete emotion classification. The goal of these algorithms is to learn from the training samples and then use this learning to classify new observation. In fact, there is no definitive answer to the choice of the learning algorithm; every technique has its own advantages and limitations. For this reason, here we have used classifiers.

- **MLP**: MLP Classifier stands for Multi-layer Perceptron classifier which in the name itself connects to a Neural Network. Unlike other classification algorithms such as Support Vectors or Naive Bayes Classifier, MLP Classifier relies on an underlying Neural Network to perform the task of classification

Classification of Feature: After all the feature extraction and selection, the feature is selected according to highest accuracy which suits among the data sets.

- Datasets: In the field of affect detection, a very important role is played by suitable choice of speech database. Three databases are used for good emotion recognition the system as given below
  1. Elicited emotional speech database: In this case emotional situation is created artificially by collecting data from the speaker
  2. Actor based speech database: Trained and professional artists collect this type of speech dataset.
  3. Natural speech database: Real world data is used to create this database
- Emotion Detection: These databases(datasets) contain 5 types of human emotion which are trained to the model. According to the best accuracy score, the model identifies the feature.
  Types of Emotion choose are:
  1. Happy
  2. Disgust
  3. Surprised
  4. Angry

Display: The emotion detected will be displayed on the screen with the speech in the form of text.

IV. RESULTS AND DISCUSSION

This project is successfully completed by displaying the predicted emotion by taking the voice input with the accuracy percentage.

```
In [1]: runfile('C:/Users/galij/.spyder-py3/test.py', wdir='C:/Users/galij/.spyder-py3')
[*] Training the model...
Accuracy: 39.77%
              precision    recall  f1-score   support

        calm       0.57      0.26      0.35        47
     disgust       0.32      0.60      0.42        40
       happy       0.46      0.75      0.57        48
     neutral       0.38      0.11      0.17        28
         sad       0.30      0.45      0.36        47
   surprised       0.69      0.17      0.27        54

    accuracy                           0.40       264
   macro avg       0.45      0.39      0.36       264
weighted avg       0.47      0.40      0.37       264

[[12 15  4  0 16  0]
 [ 1 24  9  1  4  1]
 [ 0  6 36  2  2  2]
 [ 1  9  2  3 12  1]
 [ 5  6 13  2 21  0]
 [ 2 15 14  0 14  9]]
Please talk
PREDICTED  EMOTION is :
surprised
```

Figure 2: Predicted emotion(surprised)

```
In [1]: runfile('C:/Users/galij/.spyder-py3/test.py', wdir='C:/Users/galij/.spyder-py3')
[*] Training the model...
Accuracy: 35.98%
              precision    recall  f1-score   support

        calm       0.44      0.30      0.35        47
     disgust       0.31      0.65      0.42        40
       happy       0.63      0.35      0.45        48
     neutral       0.24      0.61      0.35        28
         sad       0.39      0.38      0.39        47
   surprised       0.75      0.06      0.10        54

    accuracy                           0.36       264
   macro avg       0.46      0.39      0.34       264
weighted avg       0.49      0.36      0.34       264

[[14 10  1 17  5  0]
 [ 2 26  4  4  4  0]
 [ 0 14 17 13  3  1]
 [ 5  4  0 17  2  0]
 [ 7  7  2 13 18  0]
 [ 4 24  3  6 14  3]]
Please talk
PREDICTED  EMOTION is :
happy
```

Figure 3: Predicted emotion(happy)

```
In [4]: runfile('C:/Users/galij/.spyder-py3/test.py', wdir='C:/Users/galij/.spyder-py3')
[*] Training the model...
Accuracy: 37.12%
              precision    recall  f1-score   support

        calm       0.67      0.26      0.37        47
     disgust       0.37      0.55      0.44        40
       happy       0.59      0.27      0.37        48
     neutral       0.20      0.07      0.11        28
         sad       0.27      0.68      0.39        47
   surprised       0.46      0.31      0.37        54

    accuracy                           0.37       264
   macro avg       0.43      0.36      0.34       264
weighted avg       0.45      0.37      0.36       264

[[12  9  3  2 21  0]
 [ 1 22  2  2  9  4]
 [ 0  4 13  2 19 10]
 [ 0 10  1  2 14  1]
 [ 5  3  2  0 32  5]
 [ 0 11  1  2 23 17]]
Please talk
PREDICTED  EMOTION is :
disgust
```

Figure 4: Predicted emotion(disgust)

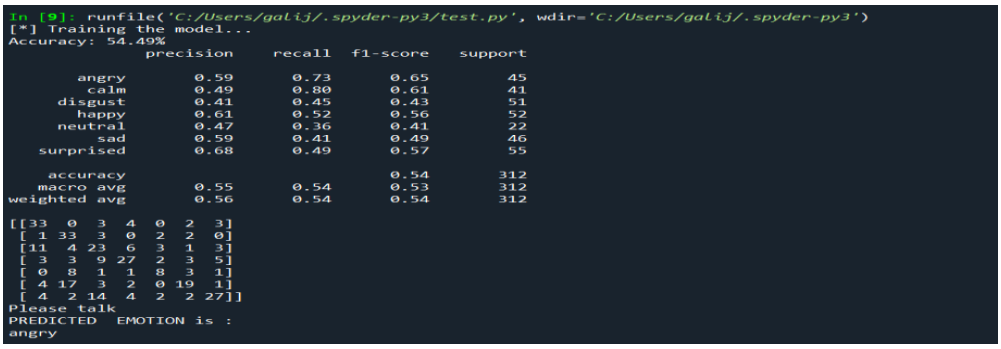

Figure 5: Predicted emotion(angry)

V. CONCLUSION

This project 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. Based on 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.

VI. REFERENCES

[1] H. Cao, R. Verma, and A. Nenkova, "Speaker-sensitive emotion recognition via ranking: Studies on acted and spontaneous speech," Comput. Speech Lang., vol. 28, no. 1, pp. 186–202, Jan. 2015.

[2] L. Chen, X. Mao, Y. Xue, and L. L. Cheng, "Speech emotion recognition: Features and classification models,"Digit. Signal Process., vol. 22, no. 6, pp. 1154–1160, Dec. 2012.

[3] T. L. Nwe, S. W. Foo, and L. C. De Silva, "Speech emotion recognition using hidden Markov models," Speech Commun., vol. 41, no. 4, pp. 603–623, Nov. 2003.

[4] S. Wu, T. H. Falk, and W.-Y. Chan, "Automatic speech emotion recognition using modulation spectral features," Speech Commun., vol. 53, no. 5, pp. 768–785, May 2011.

[5] J. Rong, G. Li, and Y.-P. P. Chen, "Acoustic feature selection for automatic emotion recognition from speech," Inf. Process. Manag., vol. 45, no. 3, pp. 315–328, May 2009.

[6] S. S. Narayanan, "Toward detecting emotions in spoken dialogs," IEEE Trans. Speech Audio Process., vol. 13, no. 2, pp. 293–303, Mar. 2005.

[7] B. Yang and M. Lugger, "Emotion recognition from speech signals using new harmony features," Signal Processing, vol. 90, no. 5, pp. 1415–1423, May 2010.

[8] E. M. Albornoz, D. H. Milone, and H. L. Rufiner, "Spoken emotion recognition using hierarchical classifiers," Comput. Speech Lang., vol. 25, no. 3, pp. 556–570, Jul. 2011.