SPEECH AND FACIAL EXPRESSION BASED REAL TIME EMOTION DETECTION USINGPYTHON

Khushali K. Trivedi[1] **,** Prof. Amee Chhaya[2] (**Assistant Professor**)**,** Electronics & Communication Engineering, Shantilal Shah Engineering College Bhavnagar,Gujarat Technological University, Ahmedabad, Gujarat.

Abstract

With the growing lifestyles, the problem of depression, anxiety, and suicidal issues are increasing day by day. The problem covers a large section of adults and teens. Thus, here is the 'Speech and Facial Expression based Emotion Detector'. The idea behind this research work is to build a machine-learning model that could detect emotions from speech & facial expressions. Nowadays personalization is something that is needed in all the things we experience every day. So why not have a real-time emotion detector that will gauge your emotions? This research work proposes Python-based deep learning. Here some predefine datasets are used to predict the proper emotion of the person. It gives a better performance than simple image processing. It may help the doctors to identify the real emotions and moods of the patient and treat them well. It also helps to predict nervousness, sadness, anger, happiness, and other such emotions and thus helps in making a proper judgment.

Keywords: Speech recognition, Face recognition, Emotion detection

Introduction

Speech-based Emotion Prediction is a task of speech processing and computational paralinguistics that aims to predict and categorize the emotions expressed in spoken words[1]. Facial Emotion Recognition (FER) is the technology that analyses facial expressions from both static images and videos in order to reveal information on one's emotional state[2]. The goal is to determine the emotional state of a speaker, such as happiness, anger, sadness, or frustration, from their speech & facial expression[3]. Speech-based emotion predictors along with Facial Gesture based emotion predictors can provide great accuracy[4].
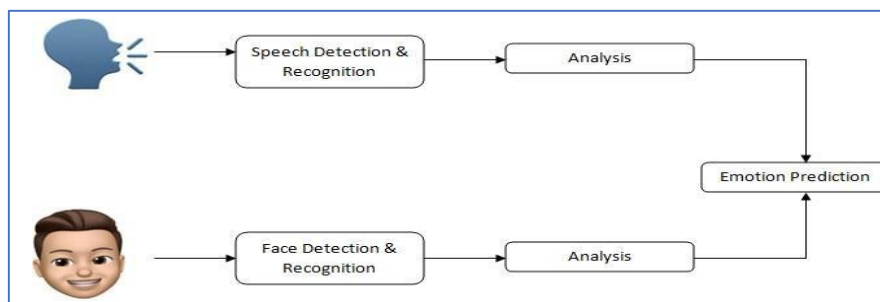

Figure 1 Speech & Face detection and Recognition

Emotion Detection Techniques

Emotion recognition or emotion detection software is a technology that uses artificial intelligence (AI)and machine learning algorithms to analyze and interpret facial expressions and emotions [3].
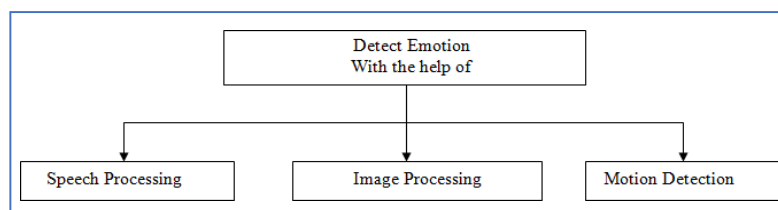

Figure 2 Emotion Detection Technique

Speech Processing

Speech Emotion Recognition is a task of speech processing and computational paralinguistic that aims to recognize and categorize the emotions expressed in spoken words[4]. There are two types of speech recognition. One is called speaker–dependent and the other is speaker–independent. Speaker–dependentsoftware is commonly used for dictation software, while speaker–independent software is more commonly found in telephone applications[6].

Speaker–dependent software works by learning the unique characteristics of a single person's voice, in a way similar to voice recognition. New users must first "train" the software by speaking to it, so the computer can analyze how the person talks. This often means users have to read a few pages of text to the computer before they can use the speech recognition software[6].

Speaker–independent software is designed to recognize anyone's voice, so no training is involved. This means it is the only real option for applications such as interactive voice response systems — where businesses can't ask callers to read pages of text before using thesystem. The downside is that speaker–independent software is generally less accurate than speaker–dependent software[6].
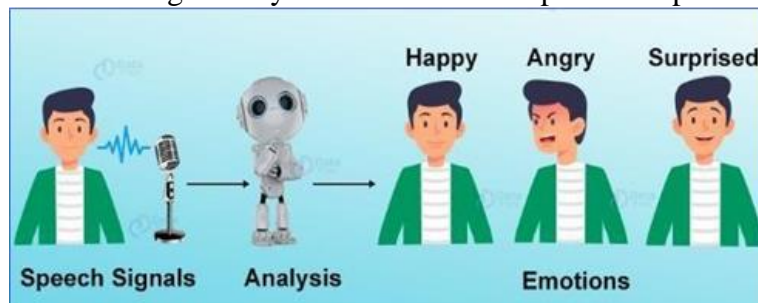

Figure 3 Emotions with speech [5]

Image Processing

Image processing is the process of transforming an image into a digital form and performing certain operations to get some useful information from it like emotion, gender, age of the person etc. The image processing system usually treats all images as 2D signals when applying certain predetermined signal processing methods [7]. We can use different Programming languages for Image processing and to recognize the emotion from the human face like C, C+, C++, Java, Python,MATLAB, SCILAB, etc. Using OpenCV libraries in Python for image processing functions is faster when compared to MATLAB. This is mainly because OpenCV libraries are written in C/C++ therefore the is only a small amount of time needed to execute the code. MATLAB is built on a lot of wrappers, which consumes more time when a code is run.
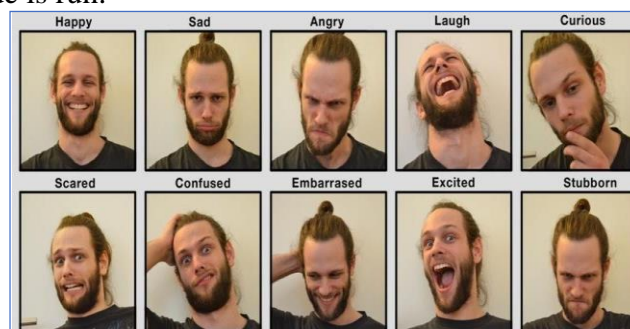
Motion Detection

Motion detection is the process of detecting a change in the position of an object relative to its surroundings or a change in the surroundings relative to an object. It can be achieved by either mechanical or electronic methods. When it is done by natural organisms, it is called motion perception [9]. Gesture recognition is a means of human-machine interaction using only body actions without the aid of voice. The concept of recognizing gestures using hands and/or other bodyparts is based on three layers: Detection, Tracking, and Recognition. [10]

Figure 5 Motion Detection [11]

Applications

It helps to provide more reliable psychological treatment; hence it is useful for doctors.

One can use this system continuously for monitoring purposes; hence it is useful for parents to monitor their child's behavior.

It helps to determine the potentially tired mood of a person.

Determine sounds differently because being in a certain mood; so it is useful for Investigation purpose.

Software Requirements

Programming Language: Python

Tools: PyCharm (Community), python3 IDLE (3.8), Anaconda(3.8)

Python Libraries: NLTK, NumPy, OpenCV, Matplotlib, Tensorflow, DeepfaceDataset: emotions.txt, haarcascade_frontalface_default.xml

NLP (Natural Language Processing)

Natural language processing (NLP) is a field that focuses on making natural human language usable by computer programs. NLTK, or Natural Language Toolkit, is a Python package that can be used for NLP. [14]

It helps computers to communicate with humans their own language and scales other language- related tasks. For example, NLP makes it possible for computers to read text, hear speech, interpret it, measure sentiment and determine which parts are important and also can analyse large amounts of natural language data.

Challenges in natural language processing frequently involve speech recognition, natural language understanding, and natural-language generation.

Deepface

Deepface is a lightweight face recognition and facial attribute analysis (age, gender, emotion and race)framework for python. [15]

It is essentially a wrapper for state-of-the-art models trained to recognize faces.

Literature Review

The primary objective of this paper is to improve man-machine interface. It can also be used to monitor the psycho physiological state of a person in lie detectors. In recent time, speech emotion recognition also finds its applications in medicine and forensics. It can also be used to recognize different emotions like sad, happy, angry, fear, etc. Basically, it can be used to detect positive and negative emotions. [11]

Sometimes DSP processor (Raspberry Pi 3) is used to recognize the emotions through speech. It canalso be implemented using complete software that is MATLAB. [11]

MFCC (Mel-Frequency Cepstrum Coefficients) & SVM (Support vector machines) technology is commonly used to detect the human emotions from speech. But it can provide approx 72% accuracy only. [12]

Also the main disadvantage of MFCC is its poor robustness to noise signals, as noise signals change all MFCCs if at least one frequency band is skewed.

Similarly an artificial intelligence (AI) system capable of emotion detection through facial expressions. It discusses about the procedure of emotion detection, which includes basically three main steps: face detection, features extraction, and emotion classification. [24]

Image processing and CNN (Convolutional Neural Network) is commonly used for Facial emotion recognition. And it can provide approx 72.16% to 80% of accuracy. [24]

Proposed Work:

Proposed work is based on Python Programming. We have used Speech Recognition package to detect the words spoken by the human.

Spoken words are stored in a text file from which we can detect the real emotion of a person by comparing the same with the emotions dataset.

Emotions Dataset is nothing but the predefine library which is having different emotions and classification of the emotions.

Also we have used OpenCV and DeepFace packages to detect the facial expression of a person in Real time.

To classify the different facial expression here, we have used haarcascade_frontalface_default dataset. This dataset is also used to detect the age, race & gender of the person. But here we are using this to detect the emotion of the person only.

Results

4.1 Speech Based Emotion DetectionHappy Emotion is predicted.

Figure 5 Happy Emotion-1
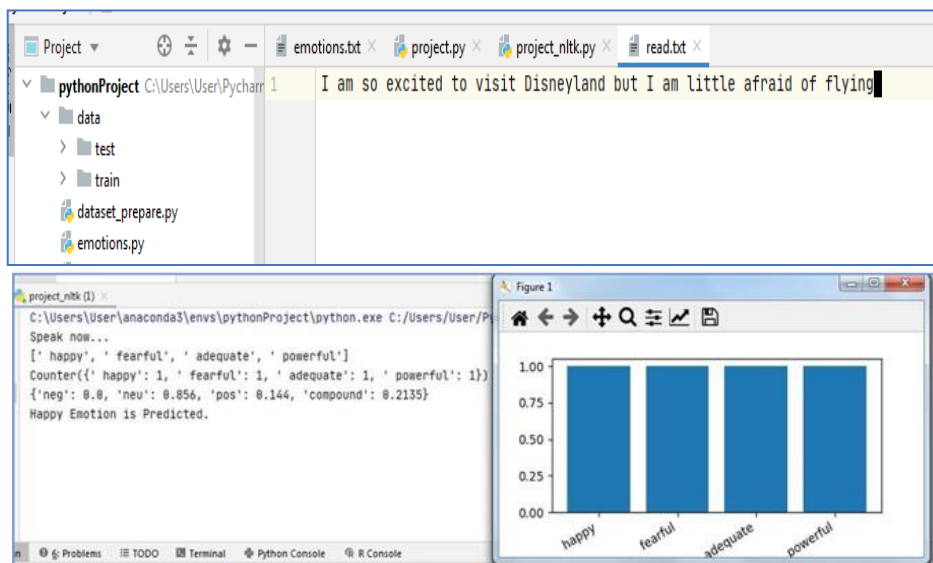
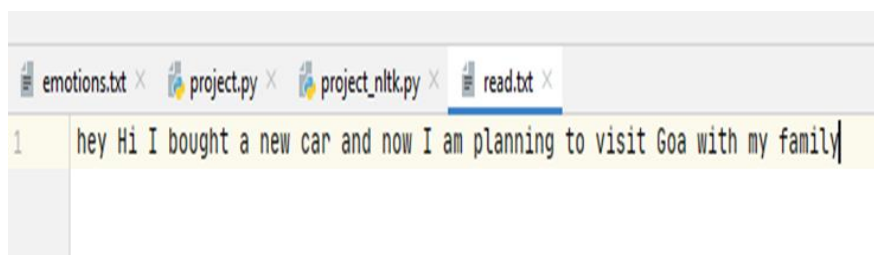Neutral Emotion is predicted. Figure 6 Happy Emotion-2

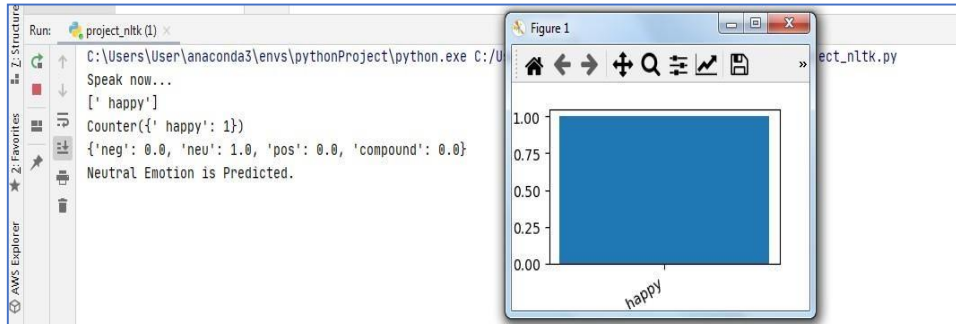Figure 7 Neutral Emotion-1

Figure 8 Neutral Emotion-2
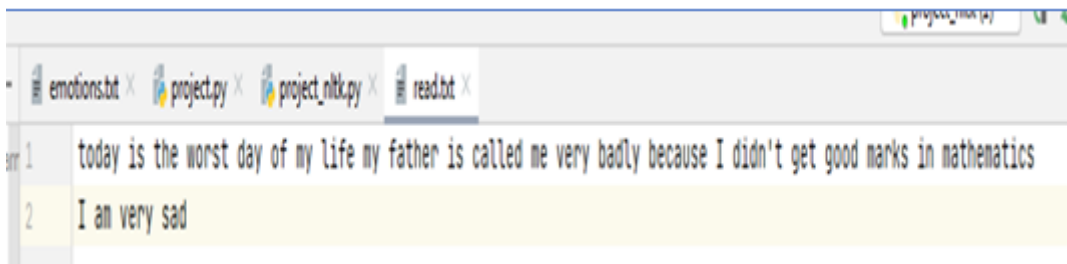Sad Emotion is predicted.
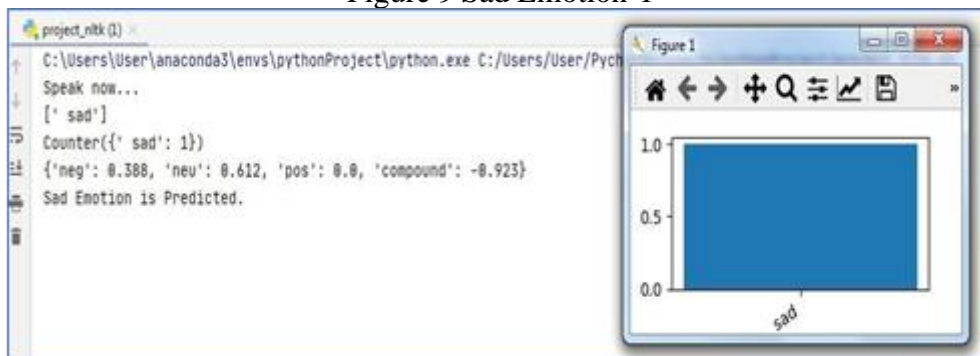

Figure 9 Sad Emotion-1

Figure 10 Sad Emotion-2

5.1 Facial Expression Base Emotion Detection [12]

Future Scope

To provide the easy evaluation for daily monitoring of child's mood using smart watch.We can use this project as a further option of NARCO TEST.

To provide the online treatment for psychological issues.To provide reliable lie detector.

To provide protection against cybercrime.

Limitations

It can only predict the emotion for the English language. Other regional languages cannot be used

topredict the emotion as of now.
Internet connectivity is required. Without proper internet connection it will not work and cannot provide an output.

Conclusion

This research work is very useful for society to understand mental and psychological condition of a person. With such pro-activeness we can reduce the stress of a person and can treat them well. This can be also usedfor Self-Monitoring purpose to avoid unwanted self-harm. With the help of Speech based emotion predictor along with Facial Gesture based emotion predictor, we can create a good & stress-free environment for future generation and also can reduce the crime.

References

[1] "B. Fasel and J. Luettin. Automatic facial expression analysis: a survey. Pattern recognition, 36(1):259–275,2003."

[2] "Speech Emotion Recognition | Papers With Code." https://paperswithcode.com/task/speech-emotion-recognition (accessed May 23, 2023).

[3] "E. Correa, A. Jonker,M.Ozo, andR.Stolk, 'Emotion recognition using deep convolutional neural networks,'Tech. Report IN4015, 2016".

[4] "Depressive disorder (depression)." https://www.who.int/news-room/fact-sheets/detail/depression (accessedMay 23, 2023).

[5] "C. Busso, S. Lee, and S. Narayanan, 'Analysis of emotionally salient aspects of fundamental frequency foremotion detection,' IEEE Trans. Audio Speech Language Process., vol. 17, no. 4, 2009, pp. 582–596.".

[6] "M. You, C. Chen, J. Bu, J. Liu, and J. Tao, 'Emotion recognition from noisy speech,' in IEEE InternationalConference on Multimedia and Expo, 2006, 2006, pp. 1653–1656l.".

[7] "Moataz El Ayadi, Mohamed S. Kamel, and Fakhri Karray, Survey on speech emotion recognition: features, classification, schemes and databases, 2010, Elsevier Ltd., Pattern Recognition, vol. 44, 2011, pp. 572–587".

[8] "D. Morrison, R. Wang, and L. De Silva, 'Ensemble methods for spoken emotion recognition in call-centres,'Speech Commun., vol. 49, no. 2, 2007, pp. 98–112".

[9] "M. M. H. El Ayadi, M. S. Kamel, and F. Karray, 'Speech emotion recognition using Gaussian mixture vectorautoregressive models,' in ICASSP 2007, vol. 4, 2007, pp. 957–960.".

[10] S. Dwijayanti, M. Iqbal, and B. Y. Suprapto, "Real-Time Implementation of Face Recognition and Emotion Recognition in a Humanoid Robot Using a Convolutional Neural Network," IEEE Access, vol. 10, pp. 89876–89886, 2022, doi: 10.1109/ACCESS.2022.3200762.

[11] A. Mishra, D. Patil, N. Karkhanis, V. Gaikar, and K. Wani, "Real time emotion detection from speech using Raspberry Pi 3," in 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai: IEEE, Mar. 2017, pp. 2300–2303. doi: 10.1109/WiSPNET.2017.8300170.

[12] X. Liu, Y. Mou, Y. Ma, C. Liu, and Z. Dai, "Speech Emotion Detection Using Sliding Window Feature Extraction and ANN," in 2020 IEEE 5th International

[13] Conference on Signal and Image Processing (ICSIP), Nanjing, China: IEEE, Oct. 2020, pp. 746–750. doi: 10.1109/ICSIP49896.2020.9339340.