IJARCCE

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 12, December 2021

DOI: 10.17148/IJARCCE.2021.101212

Speech Emotion Recognition in Machine Learning and IoT

Prathamesh Shinde[1], Sufiyan Gawandi[2], Atharva Baxi[3], Aman Pathan[4]

Student, Department of Computer Engineering, Trinity College of Engineering and Research, Pune, India [1,2,3,4]

Abstract: In the past decade plenty of analysis has gone into Automatic Speech feeling Recognition (SER). The first objective of SER is to boost man-machine interface. It also can be accustomed monitor the psychotic state of an individual in lie detectors. In recent time, speech feeling recognition conjointly finds its applications in drugs and forensics. During this paper seven emotions square measure recognized mistreatment pitch and prosody options. Majority of the speech options utilized in this work square measure in time domain. Support Vector Machine (SVM) classifier has been used for classifying the emotions. Berlin emotional info is chosen for the task. A decent recognition rate of 81 was obtained. The paper that was thought of because the reference for our work recognized four emotions and obtained a recognition rate of 94.2%. The reference paper conjointly used hybrid classifier so increasing complexes however will solely acknowledge four emotions.

Keywords : Speech Emotion Recognition, Machine Learning, IoT Automation, Graphical User Interface

1.INTRODUCTION

Human emotions square measure terribly troublesome to grasp from a quantitative perspective. Facial expressions square measure one in all the simplest ways that of dead reckoning the spirit of an individual. Speech is another modality which will be used. Speech may be an advanced signal that contains info regarding the message, speaker, language and emotions. There square measure numerous varieties of emotions which might be articulated victimization speech. Emotional speech recognition may be a system that essentially identifies the spirit of creature from his or her voice; speech is extremely dishonest even for humans to gauge the feeling of the speaker. A serious motivation comes from the need to boost the naturalness and potency of human-machine interaction. The reference paper that was chosen has been ready to with success acknowledge solely four emotions. The work given here has classified seven emotions with an overall sensible recognition rate. In general, the systems for speech analysis uses numerous techniques for the extraction of characteristics from the raw signal. Regarding emotions, the relevant info is within the Pitch, Prosody and within the Voice quality. Consecutive step during this strategy is to get the options that discriminate the speech information (to the coaching labels) and to discard the non-discriminative options. This is often achieved by calculative the cross validation between parameters once that grid of parameters is created; the one with the highest cross validation is chosen. The Emotional profiles (EP) square measure made victimization SVM with Radial Basis operate (RBF). Emotion-specific SVMs square measure trained for every category as self-versus others classifiers. Every EP contains n-components, one for the output of every emotion-specific SVM. The profiles square measure created by coefficient every of the n-outputs by the gap between the individual purpose and also the hyperplane boundary. The ultimate feeling is chosen by classifying the generated profile. This is typically done by one vs one comparison of every feeling to the present profile of the feeling. Fig.1 comprehensively explains the methodology followed during this paper. Feeling recognition is finished victimization 2 modules. The primary module is that the feature extraction module and also the second is that the classifier module. Within the feature extraction module, we've got used a feature set comprising pitch, prosody and voice quality options. Many classifiers exist for the task of feeling recognition. The various classifiers square measure SVM, MLP (Multilayer Perceptron), HMM (Hidden mathematician Model), GMM (Gaussian Mixture Model), ANN (Artificial Neural Networks) etc. The SVM classifier yields sensible results even from tiny check samples and thence it's wide used for speech emotional recognition. The SVM classifier is thus used for the planned work. Attributable to the Structural Risk minimization, SVM classifiers typically have higher performance than others.

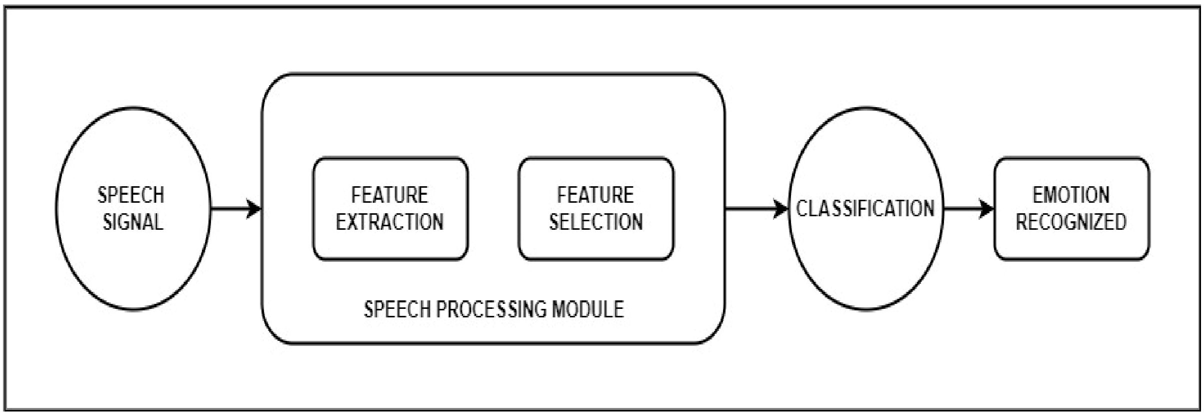
2.LITERATURE SURVEY

Emotion recognition and its production square measure easy skills of men, as against machines. Obtaining machines endued with the power to acknowledge emotions could be a tremendous challenge, which, if achieved can confirm the naturalness of human-machine interaction. The primary and foremost demand for achieving automatic feeling recognition is that the handiness of databases. Info assortment of emotional speech needs naturalness and genuineness, as a result of emotional speech is sometimes exhibited once someone isn't in Neutral state. Typically once an info is collected, it's sometimes during a controlled atmosphere, therefore the emotional speech tends to sound contrived. Thanks to this issue, speech recordings from conversations of decision centers and interviews of TV shows square measure typically used, as in. Basic feeling states like Anger, Happy, Sad, Fear, and surprise square measure the foremost common case of studies.

DOI: 10.17148/IJARCCE.2021.101212

Automatic determination of those emotions from a speech signal has been tried and productive in many researches in today's quick advancing speech technology. Classification of human speech into four emotional states (Neutral, Anger, Happy, Sad) are seen in, wherever sub-segmental options like loudness, energy of excitation, detection of voiced and. diagram for example the fundamental define of feeling recognition through speech signal. unvoiced region, fast F0, and strength of excitation were used for analysis. Spectral band energy magnitude relation and strength of excitation were examined for characteristic Anger and Happiness of 2 databases, with accuracy of seventy fifth for IIITH Telugu info and sixty eight for German Emotional info. For recognition of seven emotions, associate overall recognition rate of ninety-one.6% was obtained in, victimization modulation spectral options (MSF), short-run spectral options (MFCC and PLP) and delivery options. Fast harmonic (F0) is generally used for strength feeling recognition. Formats square measure quantitative characteristics of the vocal tract and square measure characterized by distinctive center frequencies and information measure. Estimation of Formats may be done employing a technique referred to as Linear Prediction Analysis (LPA). For feeling recognition, classifiers like Support Vector Machine (SVM), mathematician Mixture Model (GMM), Hidden Andre Mark off Model (HMM), Artificial Neural Network (ANN), k-Nearest Neighbor (kNN), etc. square measure ordinarily used. Excluding the fundamental emotions, many distinguished studies associated with non-verbal speech that depict emotions have conjointly been seen. Few distinguished mentions embody detection and analysis of yelled speech (indicating high-arousal or angry speech), and analysis of laughter. Another study has created intensive experimentation and analysis on the non-verbal speech sounds that were examined into 3 distinct classes - para linguistic sounds, emotional speech and communicator voices.

3.EXISTING SYSTEM



Phase 1 – Training phase

System learns reference patterns which represent different speech sounds (e.g. phrases, words, phones) that constitute the vocabulary of the application.

Phase 2 – Recognition phase

Unknown input pattern is identified using set of references.

Speech Recognition System works in following stages -

- Speech Analysis

Speech data is analyzed which includes speaker specific information due to vocal tract, excitation source and behavior feature which is important for speaker recognition.

- Feature Extraction

Different individual characteristics of speech embedded in utterances are extracted.

- Modelling

Hidden Markov Model (HMM) is used to create models for each letter.

- Testing

Feature testing of the dataset is done.

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 12, December 2021

DOI: 10.17148/IJARCCE.2021.101212

4.PROPOSED SYSTEM

Emotion recognition systems supported digitized speech comprises 3 basic components: signal preprocessing, feature extraction, and classification. Acoustic preprocessing like denoting, in addition as segmentation, is dispensed to see significant units of the signal. Feature extraction is employed to spot the relevant options obtainable within the signal. Lastly, the mapping of extracted feature vectors to relevant emotions is dispensed by classifiers. The system design given below depicts a simplified system used for speech-based feeling recognition. Within the initial stage of speech-based signal process, speech sweetening is dispensed wherever the clattering parts square measure removed. The second stage involves 2 components, feature extraction, and have choice. The specified options square measure extracted from the preprocessed speech signal and also the choice is created from the extracted options. Such feature extraction and choice square measure typically supported the analysis of speech signals within the time and frequency domains. Throughout the third stage, varied classifiers square measure used for classification of those options. Lastly, supported feature classification completely different emotions square measure recognized. The system planned here can have an added stage wherever lights are machine-controlled on the premise of feeling recognized within the earlier stage.

Phase 1 – Training phase

System learns reference patterns which represent different speech sounds (e.g. phrases, words, phones)

Phase 2 – Recognition phase

Unknown input pattern is identified using set of references.

Speech Recognition System works in following stages -

- Speech Analysis

Speech data is analyzed which includes speaker specific information due to vocal tract, excitation source and behavior feature which is important for speaker recognition.

- Feature Extraction

Different individual characteristics of speech embedded in utterances are extracted.

- Modelling

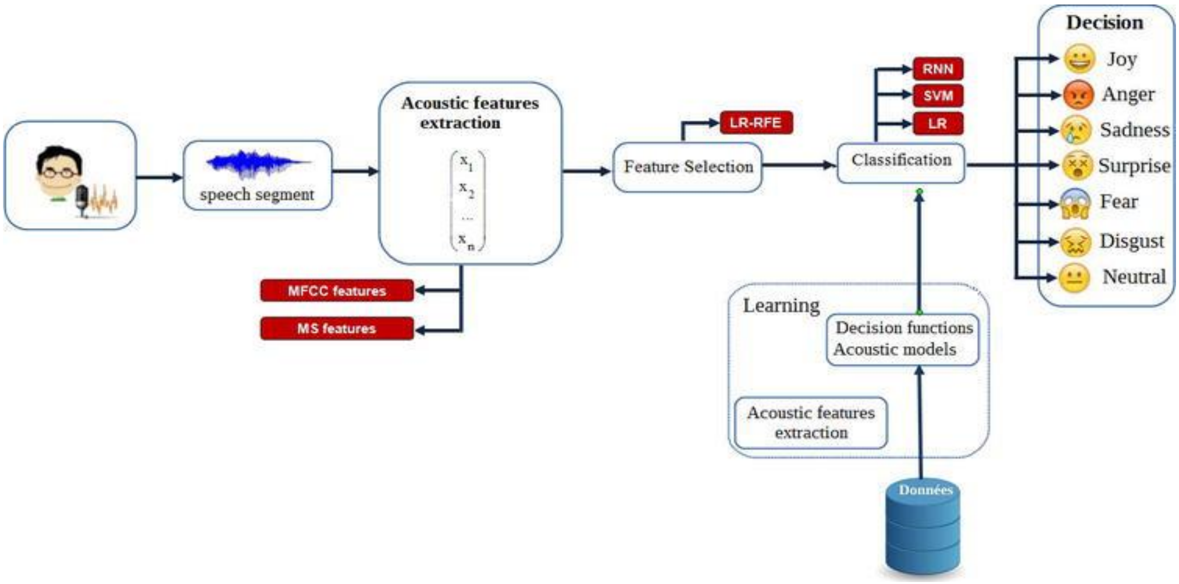
Then we test the model and recognize the emotion

- Testing

Feature testing of the dataset is done.

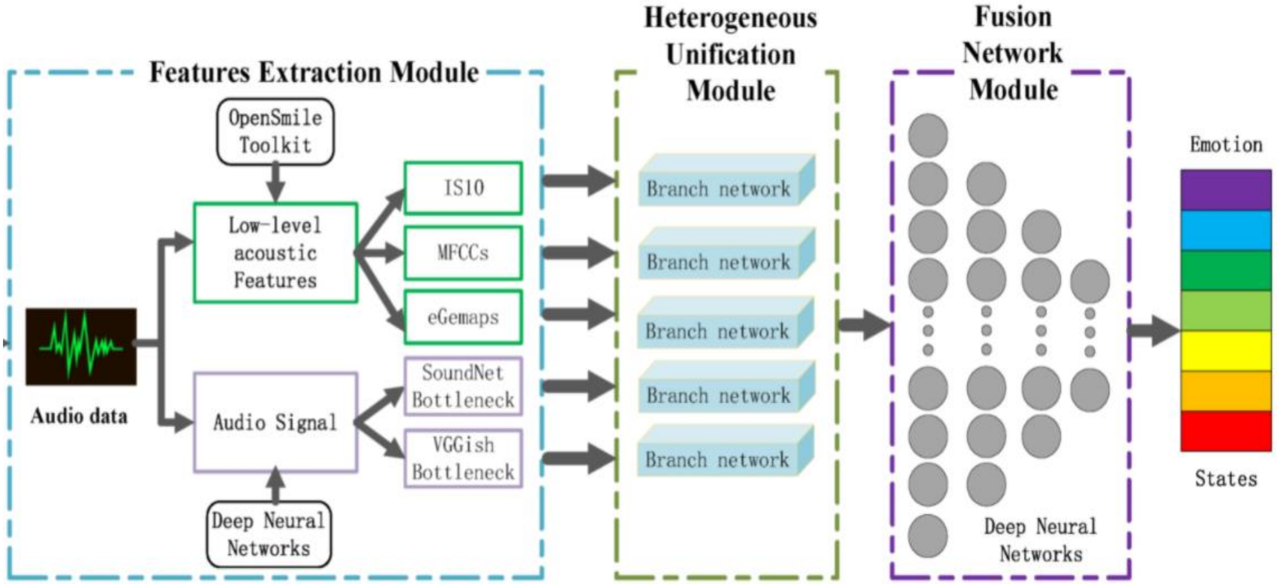
Lights are changed according to Emotions recognized

- According to the emotions recognized we change the lights of the room using the recognized emotion as field input
- Then the decision control is done and signal is sent to the sensors and lights are automated accordingly

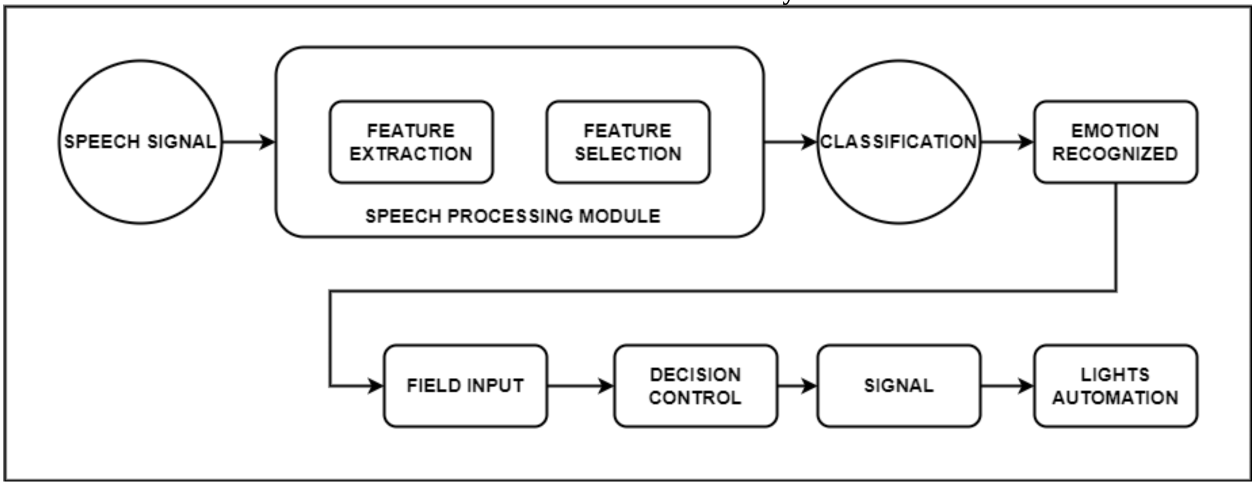


5.ML SYSTEM ARCHITECTURE

For Emotion Detection



Our IoT addition funtionality



6.CONCLUSION

This paper gives a descent approach for Speech Emotion Recognition after studying various researches done by multiple researchers in this field. We have given a brief idea about how our system is going to work. Our proposed system aims to be very useful for the people suffering from Alexithymia or for the people suffering mild depression and understands one's emotion in a better way and deal with in a best way possible and create a ambience to improve or enhance a person's emotion.

We are hoping to do more research in this field and try and implement this system with more functionalities for helping more and more people we can.

7.REFERENCE

[1] P. Gangamohan, S. R. Kadiri, and B. Yegnanarayana, "Analysis of emotional speech at subsegmental level," in INTERSPEECH, pp. 1916– 1920, 2013.

[2] S. R. Kadiri, P. Gangamohan, V. Mittal, and B. Yegnanarayana, "Naturalistic audio-visual emotion database," in 11th International Conference on Natural Language Processing, 2014, pp. 206.

[3] P. Gangamohan, S. R. Kadiri, and B. Yegnanarayana, "Analysis of emotional speecha review," in Toward Robotic Socially Believable Behaving Systems, vol. 1, Springer, pp. 205–238, 2016.

DOI: 10.17148/IJARCCE.2021.101212

[4] D. Neiberg, K. Elenius, and K. Laskowski, "Emotion recognition in spontaneous speech using gmms," in Ninth International Conference on Spoken Language Processing, 2006.
[5] D. Ververidis and C. Kotropoulos, "Emotional speech recognition: Resources, features, and methods," Speech communication, vol. 48, no. 9, pp. 1162–1181, 2011.
[6] M. El Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," Pattern Recognition, vol. 44, no. 3, pp. 572–587, 2011.
[7] P. Gangamohan, S. R. Kadiri, S. V. Gangashetty, and B. Yegnanarayana, "Excitation source features for discrimination of Anger and Happy emotions," in Fifteenth Annual Conference of the International Speech Communication Association, 2014.
azq
[17] J. J. Hopfield, "Artificial neural networks," IEEE Circuits and Devices Magazine, vol. 4, no. 5, pp. 3-10, 1988.
[18] V. K. Mittal and B. Yegnanarayana, "Production features for detection of shouted speech," in Consumer Communications and Networking Conference (CCNC), 2013 IEEE. IEEE, 2013, pp. 106–111.
[19] V. K. Mittal and B. Yegnanarayana, "Effect of glottal dynamics in the production of shouted speech," The Journal of the Acoustical Society of America, vol. 133, no. 5, pp. 3050–3061, 2013.
[20] V. K. Mittal and B. Yegnanarayana, "An automatic shout detection system using speech production features," in International Workshop on Multimodal Analyses Enabling Artificial Agents in Human-Machine Interaction. Springer, 2014, pp. 88–98.
[21] V. K. Mittal and A. Vuppala, "Changes in Shout Features in Automatically Detected Vowel Regions," in Proc. IEEE 11th International Conference on Signal Processing and Communication (SPCOM 2016), 12-15 Jun. 2016, IISc, Bangalore.
[22] V. K. Mittal and A. K. Vuppala, "Significance of Automatic Detection of Vowel Regions for Automatic Shout Detection in Continuous Speech," in Proc. IEEE/ISCA 10th International Symposium on Chinese Spoken Language Processing (ISCSLP 2016), Tianjin, China, 17-20 Oct. 2016.
[23] V. K. Mittal and B. Yegnanarayana, "Analysis of production characteristics of laughter," Computer Speech & Language, vol. 30, no. 1, pp. 99–115, 2015.
[24] V. K. Mittal and B. Yegnanarayana, "Study of changes in glottal vibration characteristics during laughter," in INTERSPEECH, 2014, pp. 1777–1781.
[25] V. K. Mittal and B. Yegnanarayana, "Study of characteristics of aperiodicity in Noh voices," The Journal of the Acoustical Society of America, vol. 137, no. 6, pp. 3411–3421, 2015.
[26] V. K. Mittal and B. Yegnanarayana, "An impulse sequence representation of the excitation source characteristics of nonverbal speech sounds," in Proc. SLPAT 2016 Workshop on Speech and Language Processing for Assistive Technologies, 2016, pp. 69–74.