https://doi.org/10.1051/itmconf/20224403071

# Regional language Speech Emotion Detection using Deep Neural Network

*Sweta* Padman[1,*], *Dhiraj* Magare[2]

[1]Ramrao Adik college of Engineering, Electronics & Telecommunication Department, Nerul, Navi Mumbai, India
[2] Ramrao Adik college of Engineering, Electronics Department, Nerul, Navi Mumbai, India

**Abstract.** Speaking is the most basic and efficient mode of human contact. Emotions assist people in communicating and understanding others' viewpoints by transmitting sentiments and providing feedback.The basic objective of speech emotion recognition is to enable computers to comprehend human emotional states such as happiness, fury, and disdain through voice cues. Extensive Effective Method Coefficients of Mel cepstral frequency have been proposed for this problem. The characteristics of Mel frequency ceptral coefficients(MFCC) and the audio based textual characteristics are extracted from the audio characteristics and the hybrid textural framework characteristics of the video are extracted. Voice emotion recognition is used in a variety of applications such as voice monitoring, online learning, clinical investigations, deception detection, entertainment, computer games, and call centres.

## 1 Introduction

The purpose of emotional speech recognition is to use a person's voice to automatically assess their emotional or physical state. During speech, air moves from the lungs to the larynx through the trachea, vibrating the vocal cords and producing speech signals [9][7]. People transmit their underlying intention through paralinguistic features such as emotions, intonation, and style through the interaction of human speech. The purpose of emotional speech recognition is to use a person's voice to automatically assess their emotional or physical state. This technology has a bright future and is critical for natural language comprehension [4]. Empathic and natural human–computer interactions necessitate the ability to perceive emotions [11][12]. Speech emotion recognition (SER) [12] has garnered a lot of academic interest in recent years, thanks to the rapid growth of conversational agents like Siri, Alexa, and Cortana. Emotions aid communication and understanding by transmitting sentiments and providing feedback to others [13]. Human voice provides a natural and instinctive interface for robot communication, and it is thus commonly used in robots that interact with humans [2]. The ability of computers to understand human emotional states such as joyful, angry, and disgust from speech signals is a fundamental goal of speech emotion recognition [12]. In recent years, a variety of viable solutions to this problem have been offered [14]. [15][6]. Speech emotion recognition is utilised in many applications, including voice surveillance, e-learning, clinical studies, lie detection, entertainment, computer games, and call centres [7].

### 1.1 Autonomous speech emotion recognition

In essence, autonomous speech emotion recognition systems employ a computer to mimic human emotions, including traits like accentuation, intonation, and pause, and match them to the target emotions using spectrum-based properties. To match their desired emotions, pause uses spectrum-based characteristics. Then, for accentuation, intonation, and pause, spectrum-based traits are used to match them to the intended emotions. A voice emotion recognition system is made up of three phases at its core: speech data preprocessing, emotion feature extraction, and emotion categorization [16][7][28]. As a result, two critical components of emotion detection are a sophisticated categorization architecture and speech emotion characteristics incorporating crucial information [7]. There are now numerous models for audio emotion identification that involve machine learning and deep learning [17][18][19][7]. The categorization process begins with the extraction of features. The quality and amount of characteristics employed determine how well a categorization system performs. Feature engineering is a key stage in categorization in this regard [8]. Speech emotions have been classified using hidden Markov models, support vector machines, deep belief networks, convolutional neural networks (CNN), and long shortterm memory networks (LSTM) [12] [1]. Acoustic characteristics of speech are extracted to identify emotions in speech. For understanding the relationship between retrieved speech data and preset emotion tags, many types of machine learning approaches are used [7]. The act of recognising emotions between people is called recognition. The effort of trying to design classifiers that generalise across application situations and acoustic

[*] Corresponding author: sweta.parkhedkar@gmail.com

settings is particularly crucial for the building of successful and practical systems of speech emotion recognition. [20][2].

Emotion detection has been used in a variety of industries, including smart homes, travel suggestion systems, and health monitoring, thanks to the rapid growth of artificial intelligence. Externally, by sight, verbal, and gestures; internally, through heart rate, respiration, blood pressure, body temperature, EEG signals, and so on. Because building speech and visual datasets is straightforward and intuitive, speech and visual characteristics are commonly employed in emotion identification. [3]. Intelligent services such as chatbots, psychological diagnosis aid, intelligent healthcare, sales advertising, and intelligent entertainment are examples of intelligent services that address not only the fulfilment of services but also the humanization of the human-computer link. [5]. It's difficult to model human emotions in words. The following are the key reasons: 1) Human emotions may be seen as noise and rejected by many existing speech recognition algorithms due to their abstraction. 2) Throughout general, human emotion can only be recognised at certain points in a protracted speech [21][4]. Nonverbal noises can effectively assist the brain in determining the difference in emotion expression when the human brain analyses emotional speech [22][5]. The automated identification and appraisal of human emotions is one of the most current research areas in fields spanning from biomedical engineering and psychophysiology to computer engineering and artificial intelligence [23][7].

## 2 Literature Review

Weighted Fusion and Consistent & Random fusion algorithms were suggested by Sheng Zhang et al.[1]. Adaptable and appropriate for activities requiring several modalities. Wisha Zehra et al.[2] investigated the Ensemble learning approach, which proved to be highly useful in the development of an emotion identification system for robots that deal with consumers from all over the world. Chen Guanghui et al.[3] adopt a Multi-modal emotion identification approach that successfully distinguishes similar classes and fuses speech and visual information, resulting in improved multi-modal emotion recognition performance. Chenghao Zhang et al.[4] employed an emotion embedding autoencoder capable of learning strong emotional information from labels. Jia-Hao Hsu et al.[5] Speech emotion detection in affective discussions with nonverbal vocalisation. Not only is it useful for recognising pleasant emotions, but it's also useful for recognising bad emotions. Transfer subspace learning was utilised by Na Liu et al.[6] to solve the unsupervised cross-corpus speech emotion recognition (SER) issue. Mehmet Bilal Er and his colleagues [7] The use of a novel hybrid architecture based on acoustic and deep features improves classification accuracy. Sofia Kanwal et al.[8] utilised a clustering-based genetic algorithm that can distinguish between different emotions.

## 3 Emotional Speech Databases

The naturalness of the information determines the success of speech emotion recognition. The Danish Emotional Speech corpus (DES) and Berlin Emotional info (EMO-DB) are two public databases, and four databases are available in Spanish, Slovenian, French, and English emotional speech. Only a few databases are authentic, and the majority of them include performed emotional speech. There appear to be three types of databases used in the SER research in terms of relevancy credibility: type one is performed emotional speech by human labeling. This information is gathered by having an actor talk with a predetermined emotion. Recently, strong challenges to the use of performed emotions have surfaced. The read of alternatives and accuracies differs between performed and spontaneous samples [15], and type two is genuine emotional speech with human labeling. Type 3 involves induced emotional speech using self-report instead of labeling, and this database comes from real-world systems (for example, contact centers). Anger and self-report are employed for labeling management whenever emotions are a unit.

## 4 Speech Emotion Recognition

The methods for emotion identification from Marathi speech, as well as the databases used to evaluate it, are presented in this chapter. The objective is to uncover acoustic emotion units that are suited for real-time applications first, and then to identify potential acoustic characteristics for emotion detection that can be extracted fast and automatically second. as well as evaluating a reasonable technique for picking the most relevant characteristics for a certain goal, and lastly selecting a quick yet accurate classification algorithm. As a result, for both the training and test phases, the approach utilized for all three processes as outlined in the overview in this chapter is detailed. Evaluation tests on the Marathi Language database with performed emotions are undertaken in order to be able to make as broad claims as feasible. The outcomes of the experiments will be discussed in the next chapter.

A fundamental challenge in speech emotion detection is defining a set of core emotions that can be classified by an automatic emotion recognizer. Languages have compiled a list of the most common emotional states we face in our everyday lives. There are 300 emotional states in a typical set. Classifying such a large variety of emotions, on the other hand, is incredibly difficult. Emotion is commonly broken down into core emotions, much as any colour may be broken down into a few basic hues. The basic emotions are anger, contempt, fear, pleasure, sadness, and surprise [2]. These are the most visible and recognisable feelings in our life. Table 1 depicts a strong relationship between mood and a few speech features.

**Table 1.** A Summary of the most common spoken emotion correlations.

| | Anger | Happiness | Sadness | Fear | Disgust |
|---|---|---|---|---|---|
| Speech Rate | Slightly Faster | Faster/ slower | Slightly slower | Much Faster | Very much slower |
| Pitch Average | Very much higher | Much higher | Slightly slower | Very much higher | Very much lower |
| Pitch Range | *Much wider* | *Much wider* | Slightly narrower | *Much wider* | Slightly narrower |
| Intensity | *Higher* | *Higher* | lower | normal | Lower |
| Voice quality | *Breathy chest tone* | *Breathy blaring* | Resonant | Irregular voking | *Grumbled chest tone* |
| Pitch change | *Abdrupt on stressed* | Smooth, Upward inflection | Downward inflections | Normal | Wide, downward terminal inflections |
| Articulation | *Tense* | Normal | Slurring | Precise | Normal |

### 4.1. Automatic Speech Emotion Recognition

Speech emotion recognition systems use a person's speech to automatically detect his or her emotional state. Analysis of the speech signal's generating process, as well as the extraction of some aspects that include emotional identification methods for identifying emotional states. The components of the speech emotion system are shown in Figure 1. The pattern recognition technology is similar to spoken emotion recognition. This demonstrates that the steps seen in the pattern recognition system are also found in the Speech emotion recognition system. The speech emotion recognition system has five main modules for training and testing: emotional speech input, preprocessing, feature extraction, feature normalization, classification, and recognised emotional output [2].

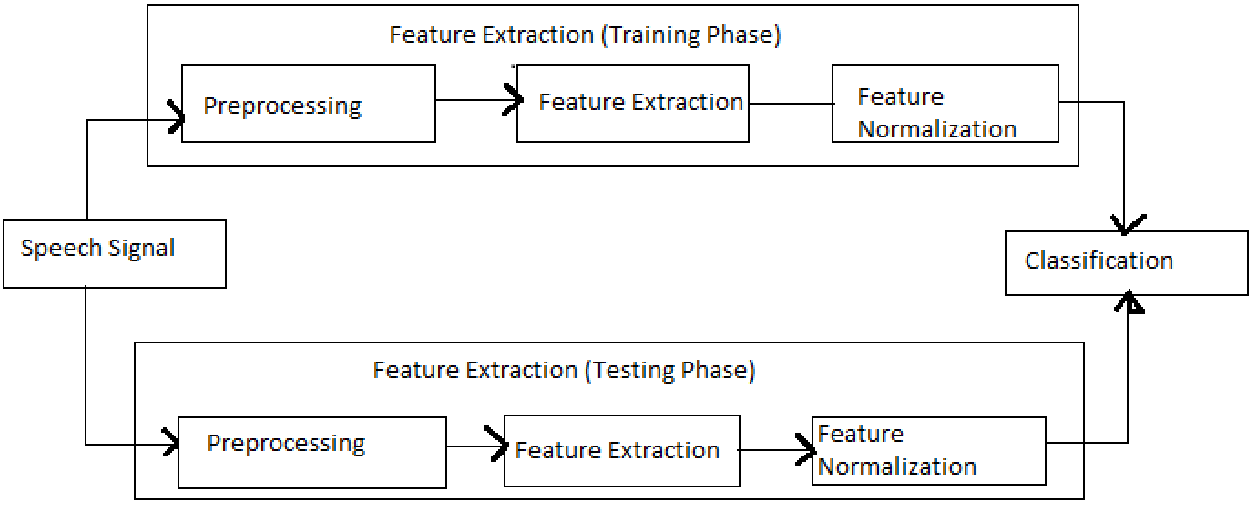

**Fig. 1.** Basic Emotion Recognition system.

#### 4.1.1 Database Creation

The naturalness of the database used as an input is used to evaluate the speech emotion recognition system. If the system is given a bad database as an input, the system may come to incorrect conclusions. The database used as an input to the voice emotion detection system might comprise real-life or staged emotions. It is more feasible to employ a database that has been compiled from real-world scenarios [15]. The first step is to Emotion identification from speech is derived from acoustic measurements of those units, which are then used to determine the real characteristics from the audio input signal. The units are frequently phrases or utterances, which are linguistically motivated medium-length time periods. Despite the fact that choosing whatever type of unit to join is clearly vital, it has gotten little attention. Neither the division into utterances nor the expectation of a continuous feeling over each speech are straightforward. A good emotion unit, in general, must meet a set of conditions.

The voice emotion recognition system is estimated based on the naturalness of the database that is utilised as an input. If a bad database is used as an input to the method, it is possible that incorrect conclusions will be made. The database, which is used as an input to the spoken emotion detection system, may contain actual or acted emotions. It is more feasible to employ a database that has been compiled from real-world scenarios [15]. The first step in speech emotion recognition is to make the audio input signal meaningful, after which acoustic measurements of those units are used to extract the real features. The units are generally medium-length linguistically inspired units. phrases or utterances are examples of temporal intervals. Despite the fact that choosing whatever type of unit to join is clearly vital, it has gotten little attention. Neither the division into utterances nor the expectation of a consistent feeling over a speech are correct. In general, a good emotion unit must adhere to a set of guidelines. It should be, in particular:

1. To be consistently extracted, it must be well-defined.

2. Long enough that statistical functions may be used to calculate characteristics convincingly.

3. Short enough to maintain consistent emotional acoustic characteristics throughout the Segment.

4. Consistent with the training database's labeling.

Speech samples used in training, testing, and applications should all follow the same set of criteria, i.e., they must have the same properties. The Marathi Database is created using these guidelines.

### 4.2. Feature Evaluation

A multi-algorithm technique for detecting emotion from audio signals is given. The suggested MFCC and Discrete Wavelet Transform-based algorithms will be utilised to extract emotional information from speech data, using characteristics derived from pitch and formant frequency to support them. Pitch contour features such as local maxima, local minima, frequency distance, temporal distance, and slope between nearby local extrema are determined for each frame of speech sample. In addition, the first four formant frequencies are determined. The MFCC technique is a time-honored way of analysing speech signals. It is based on a linear cosine transform of a log power spectrum on a nonlinear Mel frequency scale of frequency and depicts the short-term power spectrum of a sound. DWT is used to breakdown the input speech signal and provide approximation and detail coefficients as an alternate approach. 4th level decomposition using db4 wavelets will be used to derive wavelet characteristics for each input spoken sound. The SVM classifier will be used to determine the similarity between the recovered features and a set of reference features. The database will have Marathi speech samples for each of the six emotions.

### 5 Database:

The Emotional Prosody Speech corpus provided us with our data. This corpus comprises Marathi continuous utterances created by 6 speakers in 6 emotions: happy, rage, neutral, fear, sorrow, and boredom (3 female, 3 male). For the examination of distinct emotions, a corpus of 180 utterances of phrases was recorded. For each of the six emotions, each speaker is given a set of five sentences to utter.

**Recording:** The recording was done using an electric microphone in a partly sound 16 kHz/16 bit format, with the distance between the lips and the microphone set to about 30 cm.

**Listening Test:** We initially randomized all of one speaker's continuous sentence files, which were then shown to ten naive listeners who were asked to rate the process was repeated for all 10 speakers, and the feelings were divided into six categories: neutral, pleased, angry, grief, fear, and surprise. All of the listeners were educated and aged 18 to 28 years old. For this study, only those statements were picked that had at least 80% of all listeners recognising their sentiments.

## 6 Acoustic Analysis of Emotions:

The acoustic characteristics of the voice, such as intensity, pitch, and length, are also influenced by emotion. Sentence acoustic analysis is performed. The spectrograms of one of the sentences are shown in Figures (4.1-4.6). Both prosody-related and spectral variables were taken into account in our research of emotions.

| **Emotional Class** | **Neutral** | **Anger** | **Boredom** | **Fear** | **Happiness** | **Sadness** |
|---|---|---|---|---|---|---|
| **Neutral** | 20 | 20 | 60 | 0 | 0 | 0 |
| **Anger** | 0 | 100 | 0 | 0 | 0 | 0 |
| **Boredom** | 0 | 0 | 40 | 0 | 60 | 0 |
| **Fear** | 0 | 20 | 20 | 20 | 40 | 0 |
| **Happiness** | 0 | 20 | 20 | 0 | 60 | 0 |
| **Sadness** | 0 | 0 | 40 | 40 | 20 | 0 |

Table 2. Confusion Matrix for recognition using pitch based features

| **Emotional Class** | **Neutral** | **Anger** | **Boredom** | **Fear** | **Happiness** | **Sadness** |
|---|---|---|---|---|---|---|
| **Neutral** | 80 | 0 | 20 | 0 | 0 | 0 |
| **Anger** | 0 | 100 | 0 | 0 | 0 | 0 |
| **Boredom** | 0 | 0 | 100 | 0 | 0 | 0 |
| **Fear** | 0 | 0 | 0 | 80 | 20 | 0 |
| **Happiness** | 0 | 0 | 0 | 0 | 100 | 0 |
| **Sadness** | 0 | 0 | 0 | 0 | 0 | 100 |

Table 3. Confusion Matrix for recognition using MFCC based features

4.3.1 **Pitch**:

Tone height is the acoustic equivalent of pitch, which is a fundamental frequency. Pitch estimate by machine is a difficult problem. Effects of vocal tract resonance and short-term disruptions in the spoken stream can obscure pitch detection [17]. Figures shown below indicate pitch contour curves for specific emotion for one of Marathi sentence "KOKILACHA AWAZ KHUP MADHUR ASTO".

Both prosody-related and spectral variables were taken into account while studying emotions. Pitch contour curves in utterances of angry feeling increase and decrease towards the beginning of the phrase and descend towards the conclusion of the sentence, according to the findings (fig 2)[28]. For dread, the pitch increases at the start of the phrase and then stays the same before falling at the end (fig 3). Pitch contour curves of utterances in a happy mood exhibit a hold pattern at the start of the phrase and rise and fall at the conclusion of the sentence (fig 4)[28]. Pitch lowers at the end of the sentence and rises and falls at the beginning of the sentence in a neutral mood (fig 5.). The intensity of the sad emotion's F0 curve falls and rises in the start of sentences (fig. 6), whereas the intensity of the boring emotion's F0 curve falls and rises in the commencement of sentences and declines at the end position (fig. 7). With the exception of surprise, it has been determined that the intensity curve varies in proportion to the pitch in most emotions. The length of phrases pronounced in different emotions has varied values, according to the pitch contours of the sentences. As can be seen from the pitch contour (fig. 6), boredom emotions have the longest length of 2.8 seconds, while rage emotions have the shortest duration of 1.6 seconds. To evaluate single features, the information gain for all characteristics with relation to emotion classes was estimated in a database[28].

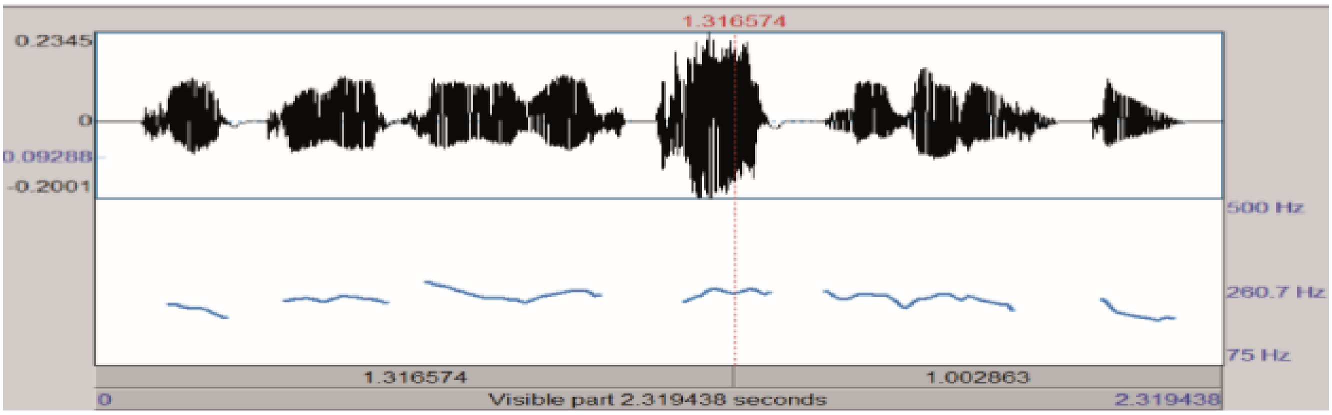

**fig.2** Pitch contour for Fear

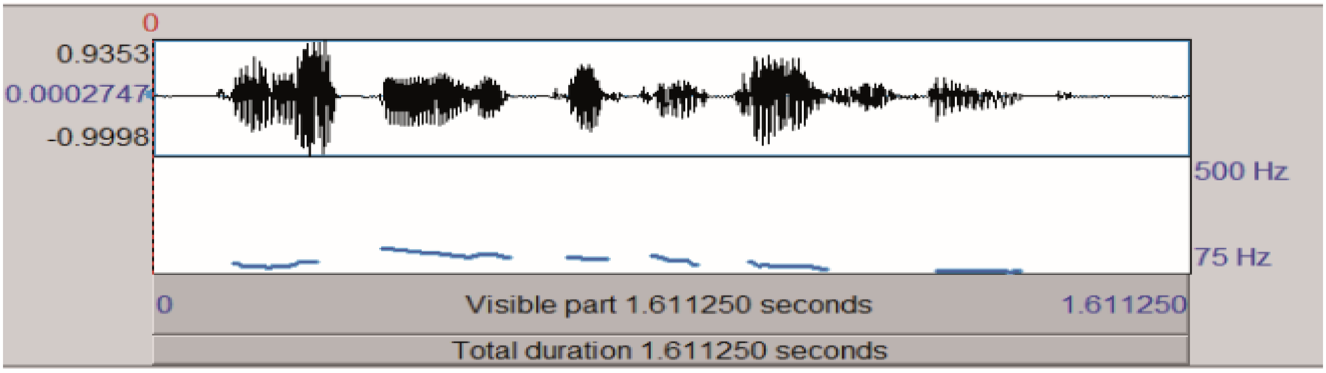

**Fig.3** Pitch contour for Neutral

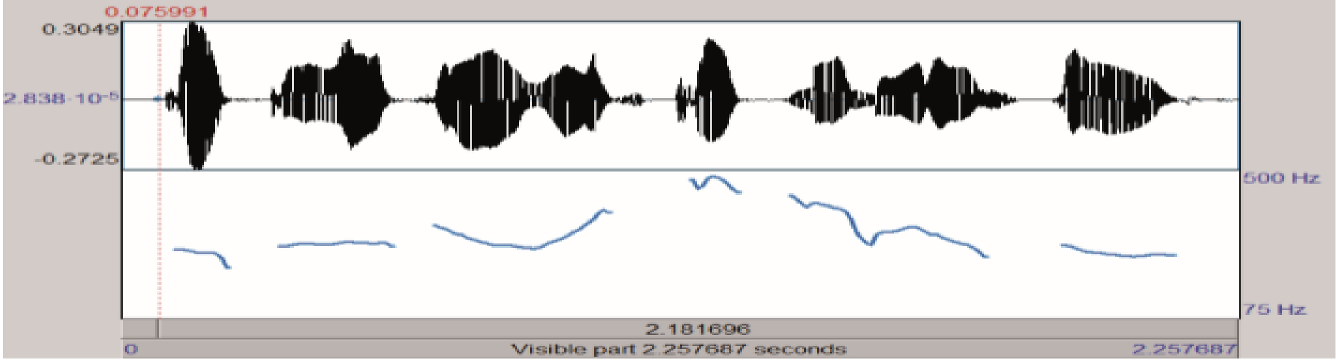

**Fig.4** Pitch contour for Happiness

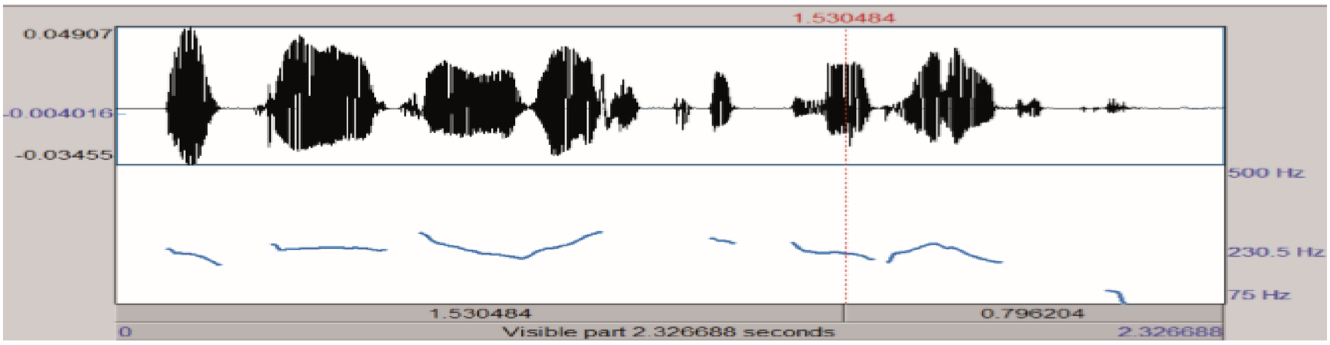

**Fig.5** Pitch contour for Boredom

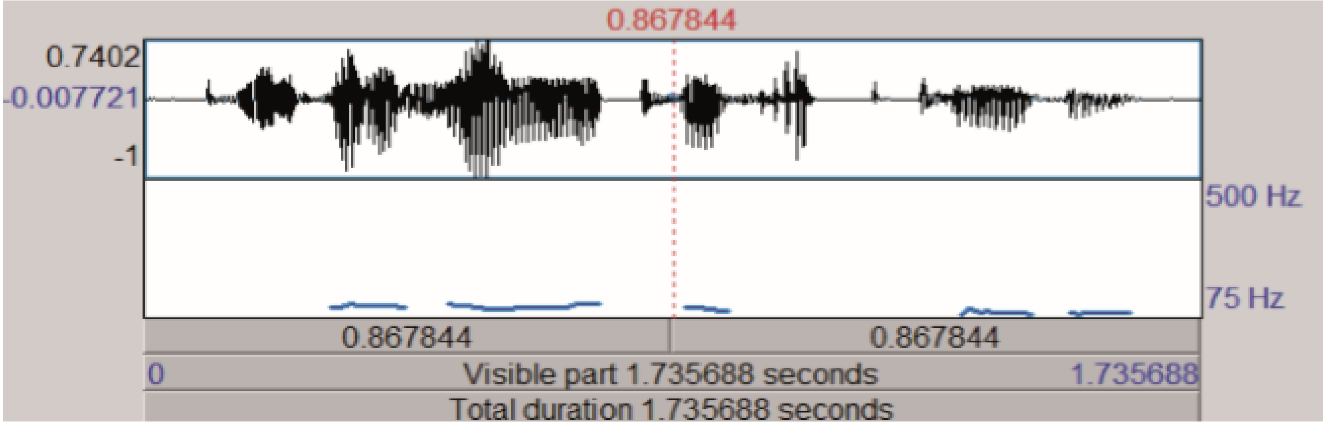

**Fig.6** Pitch contour for Sadness

| Emotional Class | Neutral | Anger | Boredom | Fear | Happiness | Sadness |
|---|---|---|---|---|---|---|
| Neutral | 80 | 0 | 20 | 0 | 0 | 0 |
| Anger | 0 | 100 | 0 | 0 | 0 | 0 |
| Boredom | 0 | 0 | 100 | 0 | 0 | 0 |
| Fear | 0 | 0 | 0 | 80 | 20 | 0 |
| Happiness | 0 | 0 | 0 | 0 | 100 | 0 |
| Sadness | 0 | 0 | 0 | 0 | 0 | 100 |

Table 4. Confusion Matrix for recognition using Fusion features

## 7 Results

The multi- algorithm approach is proposed for emotion recognition from speech signals. The system is designed with specifications as:  The Database created of one of Indian regional languages, the Marathi speech samples database created for the six emotions. The prosodic analysis on the Marathi speech samples shows that linguistic changes do not affect the prosody and emotion correlation. Feature extraction performed using the Acoustic features like Pitch and Formant frequency along with MFCC and Discrete Wavelet Transform based algorithms.  Evaluation carried out for recognition of emotions by using individual features first. Results show that recognition is better when pitch and MFCC algorithms are employed individually.

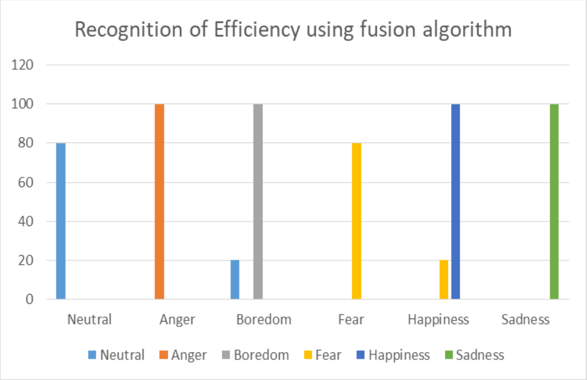

Fig 7. Recognition efficiency using fusion based features

## 8 Conclusion

When all of the characteristics are combined, the system's performance is deemed to be good. This suggests that the problem of speech emotion recognition is better handled when using a group of multiple descriptors. Moreover, our methodology significantly outperforms the ability of a human listener to classify the respective signals which is equal to 100 percent. The Classifier shows better results when fusion of all the features are used as shown in table 5 and 6. In future work, we aim to test system performance on spontaneously expressed (Non acted) emotions of Marathi language. To enhance system performance rate. further introduction of one more classifier can be proposed which can reduce confusion rate. classifier can be proposed which can reduce confusion

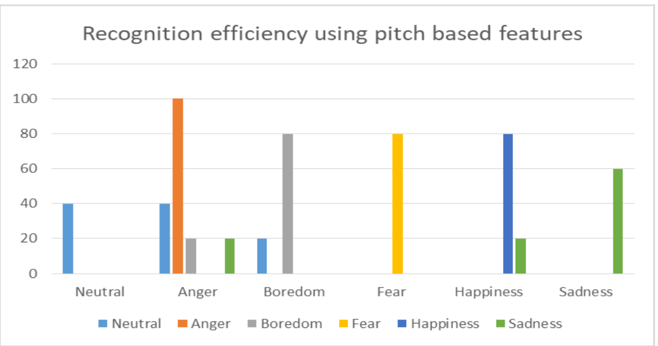

Fig. 8  Recognition efficiency using pitch based features

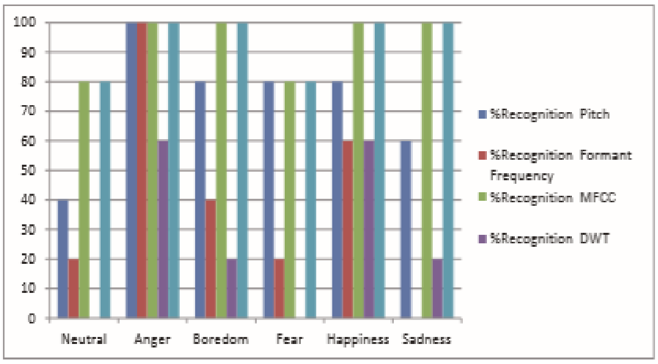

Fig 9. Comparison of % recognition for feature extraction algorithm

## References

[1] Zhang, Sheng, Min Chen, Jincai Chen, Yuan-Fang Li, Yiling Wu, Minglei Li, and Chuanbo Zhu, "Combining cross-modal knowledge transfer and semi-supervised learning for speech emotion recognition," in Knowledge-Based Systems, Vol.229, pp.107340, 2021.

[2] Zehra, Wisha, Abdul Rehman Javed, Zunera Jalil, Habib Ullah Khan, and Thippa Reddy Gadekallu, "Cross corpus multi-lingual speech emotion recognition using ensemble learning," in Complex & Intelligent Systems, Vol.7, no.4, pp.1845-1854, 2021.

[3] Guanghui, Chen, and Zeng Xiaoping, "Multi-modal emotion recognition by fusing correlation features of speech-visual," in IEEE Signal Processing Letters, Vol.28, pp.533-537, 2021.

[4] C. Zhang and L. Xue, "Autoencoder With Emotion Embedding for Speech Emotion Recognition," in IEEE Access, vol.9, pp.51231-51241, 2021.

[5] Hsu, Jia-Hao, Ming-Hsiang Su, Chung-Hsien Wu, and Yi-Hsuan Chen, "Speech emotion recognition considering nonverbal vocalization in affective conversations," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 29, pp.1675-1686, 2021.

[6] N. Liu, Baofeng Zhang, Bin Liu, Jingang Shi, Lei Yang, Zhiwei Li and Junchao Zhu, "Transfer Subspace Learning for Unsupervised Cross-Corpus Speech Emotion Recognition," in IEEE Access, vol. 9, pp. 95925-95937, 2021.

[7] M. B. Er, "A Novel Approach for Classification of Speech Emotions Based on Deep and Acoustic Features," in IEEE Access, vol. 8, pp. 221640-221653, 2020.

[8] S. Kanwal and S. Asghar, "Speech Emotion Recognition Using Clustering Based GA-Optimized Feature Set," in IEEE Access, vol.9, pp.125830-125842, 2021.

[9] Schlegel, Patrick, Stefan Kniesburges, Stephan Dürr, Anne Schützenberger, and Michael Döllinger, "Machine learning based identification of relevant parameters for functional voice disorders derived from endoscopic high-speed recordings," in scientific reports, Vol.10, no.1, pp.1-14, 2020.

[10] E. Cambria, S. Poria, A. Hussain and B. Liu, "Computational Intelligence for Affective Computing and Sentiment Analysis [Guest Editorial]," in IEEE Computational Intelligence Magazine, vol. 14, no. 2, pp. 16-17, 2019.

[11] Chen, Min, and Yixue Hao, "Label-less learning for emotion cognition," in IEEE transactions on neural networks and learning systems, Vol.31, no.7, pp.2430-2440, 2019.

[12] El Ayadi, Moataz, Mohamed S. Kamel, and Fakhri Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," in Pattern recognition, Vol.44, no. 3, pp.572-587, 2011.

[13] Zvarevashe, Kudakwashe, and Oludayo Olugbara, "Ensemble learning of hybrid acoustic features for speech emotion recognition," in Algorithms, Vol.13, no.3, pp.70, 2020.

[14] Liu, Zhen-Tao, Qiao Xie, Min Wu, Wei-Hua Cao, Ying Mei, and Jun-Wei Mao, "Speech emotion recognition based on an improved brain emotion learning model," in Neurocomputing, Vol.309, pp.145-156, 2018.

[15] Gideon, John, Melvin G. McInnis, and Emily Mower Provost, "Improving cross-corpus speech emotion recognition with adversarial discriminative domain generalization (ADDoG)," in IEEE Transactions on Affective Computing, Vol.12, no.4, pp.1055-1068, 2019.

[16] Lu, Guanming, Liang Yuan, Wenjuan Yang, Jingjie Yan, and Haibo Li, "Speech emotion recognition based on long short-term memory and convolutional neural networks," in Journal of Nanjing University of Posts and Telecommunications, Vol.38, no.5, pp.63-69, 2018.

[17] Liu, Zhen-Tao, Qiao Xie, Min Wu, Wei-Hua Cao, Ying Mei, and Jun-Wei Mao, "Speech emotion recognition based on an improved brain emotion learning model," in Neurocomputing, Vol.309, pp.145-156, 2018.

[18] Bhavan, Anjali, Pankaj Chauhan, and Rajiv Ratn Shah, "Bagged support vector machines for emotion recognition from speech," in Knowledge-Based Systems, Vol.184, pp.104886, 2019.

[19] Issa, Dias, M. Fatih Demirci, and Adnan Yazici, "Speech emotion recognition with deep convolutional neural networks," in Biomedical Signal Processing and Control, Vol.59, pp.101894, 2020.

[20] Schuller, Björn, Zixing Zhang, Felix Weninger, and Gerhard Rigoll, "Using multiple databases for training in emotion recognition: To unite or to vote?," in Twelfth Annual Conference of the International Speech Communication Association, 2011.

[21] Atmaja, Bagus Tris, and Masato Akagi, "Speech emotion recognition based on speech segment using LSTM with attention model," in 2019 IEEE International Conference on Signals and Systems (ICSigSys), pp.40-44, 2019.

[22] Schirmer, Annett, and Thomas C. Gunter, "Temporal signatures of processing voiceness and emotion in sound," in Social cognitive and affective neuroscience, Vol.12, no.6, pp.902-909, 2017.

[23] Nardelli, Mimma, Gaetano Valenza, Alberto Greco, Antonio Lanata, and Enzo Pasquale Scilingo, "Recognizing emotions induced by affective sounds through heart rate variability," in IEEE Transactions on Affective Computing, Vol.6, no.4, pp.385-394, 2015.

[24] F. Dellaert, T. Polzin and A. Waibel, "Recognizing emotion in speech," Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP, vol.3, pp. 1970-1973 1996.

[25] https://ieeexplore.ieee.org/document/1623803?reload=true

[26] Askarzadeh, Alireza, "A novel metaheuristic method for solving constrained engineering optimization problems: crow search algorithm," in Computers & Structures, Vol.169, pp.1-12, 2016.

[27] Rao, R. Venkata, "Teaching-learning-based optimization algorithm," in Teaching learning based optimization algorithm, pp. 9-39, 2016.
[28] Agrawal, Shyam Sunder. “Emotions in Hindi speech- analysis, perception and recognition.” *2011 International Conference on Speech Database and Assessments (Oriental COCOSDA)* (2011): 7-13.