Speech Emotion Recognition using GTCC, NN and GA

[1]Khushboo Mittal, [2]Parvinder Kaur
[1]Student, [2]Asst.Proffesor
[1]Computer Science and Engineering
[1]Shaheed Udham Singh College of Engineering and Technology, Tangori, Punjab

***Abstract* - Emotion identification from speech of a human being is a very important area in research, which represents the inhuman-computer interaction. Recently, increasing attention has been directed to the study of the emotional content of speech signals, and hence, many systems have been proposed to identify the emotional content of a spoken utterance. The recent literature on speech emotion recognition has been presented considering the issues related to emotional speech corpora, different types of speech features and models used for recognition of emotions from speech. In this work, three emotions JOY, SAD, and AGGRESSIVE, are used based on automatic speech emotion recognition system. A new system is proposed for the detection of emotion form speech of an individual, GTCC and GA are used for feature extraction and feature reduction purpose. And for classification neural network and support vector machine algorithm are used. Then after obtaining results from NN and SVM, they are compared using graph. This proves neural network does better classification.**

***Keywords* - Emotion Recognition, Happy, Sad, Anger, Feature Extraction, Classification.**

I. INTRODUCTION

The speech signal is the fastest and the most natural method of communication between several peoples [1]. This statement has motivated many researchers to consider speech as a fast and efficient method of interaction between an individual as well as machine. On the other hand, this necessitates that machine should have the sufficient intelligence to recognize human voices. In the late fifties, there has been remarkable research on speech recognition, which refers to the process of transforming the human-speech into an arrangement of words. However, despite the great progress made in speech recognition, the researchers are quite far from devising a natural communication amongst man and machine because the machine does not understand emotional-state of the particular speaker. This has presented a relatively recent research field, namely speech emotion-detection that is usually well-defined as extricating emotional-state of a speaker from his or her speech. It is believed that speech-emotion-detection could be utilized to extricate valuable semantics from speech, and hence, improves the performance of speech recognition systems [2].

Emotional speech recognition aims at involuntarily identifying the emotional or physical-condition of an individual through her/his voice [3]. A speaker has dissimilar stages throughout speech that are recognized as emotional characteristics of speech as well as are assimilated in the paralinguistic aspects. The linguistic content cannot modify by emotional state; in communication of individual this is a significant factor, since feedback information is delivered in plentiful applications [4].

Speech is perhaps the generally proficient way to correspond with each other. This also means that speech could possibly be a quite helpful periphery to cooperate with machines [5]. A few victorious examples based on it throughout the past years, while we have awareness about electromagnetism; includes the development of the megaphone, telephone. Even in the previous centuries people were researching on speech fusion. On Kempe Len developed an engine talented of 'speaking' words and phrases [6].

At the present time, it has happened to achievable not only to expand examination and execute speech recognition systems, but also to have systems competent to real-time alteration of text into speech [7]. Below we have given speech recognition system:

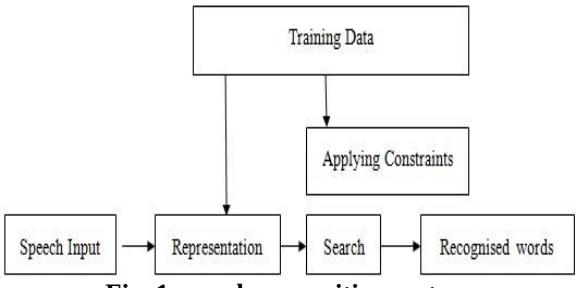


Fig. 1 speech recognition system

Speech emotion recognition is mainly beneficial for applications that usually necessitate natural man–machine communication for instance web, movies, along with computer-tutorial applications where the proper response of those frameworks towards the user be contingent on the perceived emotions [2].

Similarly, it is valuable for in-car-board system in which information of the mental-state of the driver might be provided towards the system to initiate his/her safety [2, 8]. It can be also employed as a diagnostic-tool for psychiatrists/counselors [3].

It could be beneficial in automatic-translation-systems, in which the emotional-state of the speaker plays an important-role in communication between multiple parties [9].

Criteria for Emotion Recognition

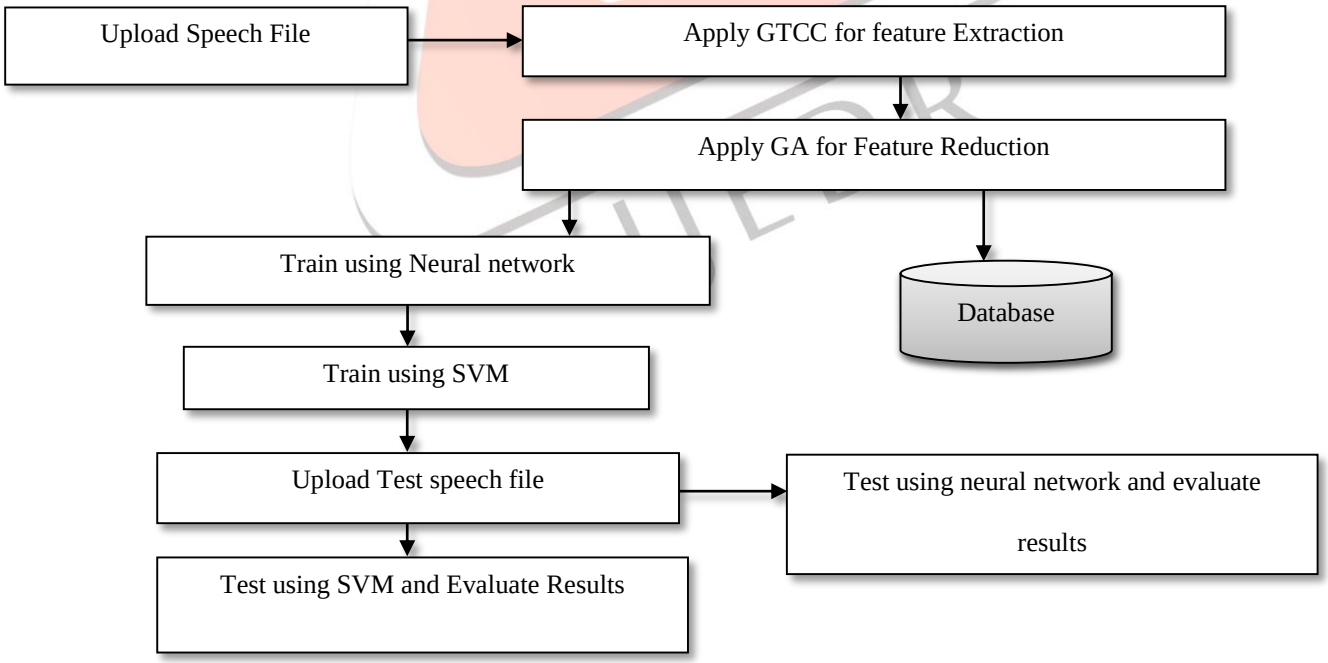
- INPUT: Receiving of input signals (i.e. raw semantic modulation data) [10].
- PATTERN RECOGINITION: Feature extraction and classification (i.e. relevant emotional features and structures for input signals).
- REASONING: Prediction of emotion based on knowledge about emotion generation and expression, i.e. reasoning about situations, goals, preferences, social rules, and other perceived context [11].
- LEARNING: Learning of person dependent factors and updating of rules used in future reasoning based on new information and reasoning [12].
- BIAS: Optional bias in recognition to account for internal emotional states, if emotionally induced behavior is defined.
- OUTPUT: Descriptions of recognized emotions and expressions (e.g. probabilities for current and predicted emotions).

II. SIMULATION MODEL

There are number of milestones that need to be achieved in order to reach the goal, so following are the sequential steps that will led to attain the signal.

Step 1 : Ready database.
Step 2 : Apply GTCC for feature extraction form uploaded speech file.
Step 3 : Apply Genetic algorithm for feature reduction.
Step 4 : Repeat step 1-3 for all categories.
Step 5 : Save data to db.
Step 6 : Then train the data by using Support Vector Machine
Step 7 : Train data by using Neural Network.
Step 8 : Upload a file to test.
Step 9 : Extract same features of the uploaded file.
Step 10 : Set Target values for the evaluation.
Step 11 : Apply SVM algorithm for classification
Step 12 : Apply BPNN algorithm for classification.
Step 13 : Evaluate the results using accuracy parameter.

Fig.2 proposed work flowchart



Technique Used

GTCC (Gammatone Cepstral Coefficients)

Gamma-tone-function models the human-auditory-filter-response. The association between the impulse response of the gamma-tone-filter and the one which is acquired from the beings was established in [12].

The computation procedure of the projected Gamma-tone-cepstral-coefficients is equivalent to the Mel-frequency cepstral coefficients extraction scheme [13]. The audio-signal is initially windowed into short-frames, generally of 10 to 50 ms. This method has a two-fold-purpose:

1) The non-stationary audio-signal could be assumed to be stationary for a short interval, thus enabling the spectro-temporal signal-analysis; and then
2) The proficiency of the feature-extraction procedure is augmented

Neural Network

Machine learning algorithms facilitate a lot in decision making and neural network has performed well in categorization purpose in medical field. Most popular techniques among them are neural network. Neural networks are those networks that are the collection of simple elements which function parallel [7]. A neural network can be trained to perform a particular function by adjusting the values of the weights between elements. Network function is determined by the connections between elements [9]. There are several activation functions that are used to produce relevant output.

Genetic Algorithm

Genetic Algorithms are adaptive heuristic search algorithm based on the evolutionary ideas of normal range and inheritance. As such they signify an intelligent operation of an arbitrary search used to solve optimization problems. Even if randomized, GAs are by no means random, as a substitute they develop past in sequence to direct the search into the region of better act within the search space. The basic techniques of the GAs are calculated to suggest processes in natural systems required for growth; especially those follow the values first laid down by Charles Darwin of "survival of the fitting."

III. EXPERIMENT RESULTS

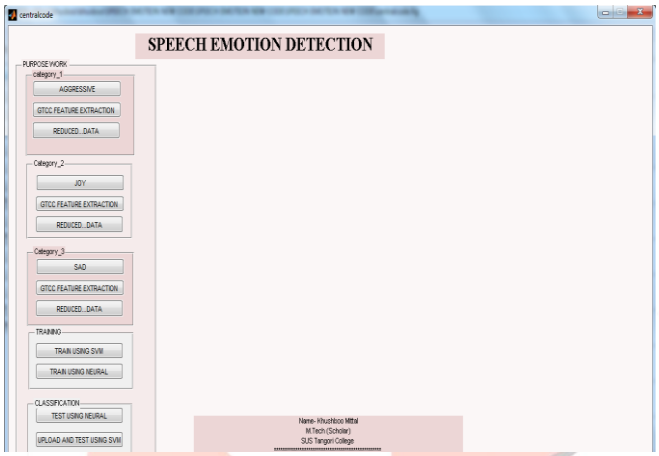


Fig.3 main gui

In above figure, main graphical user interface is shown of proposed system. In this there are three categories panel, training panel and testing panel. In category panel such as aggressive, joy and sad are in separate panels. In training panel two buttons are shown of training using neural network and training using support vector machine. In testing panel there are two testing buttons one is of testing using neural network, and other one is of testing using support vector machine.

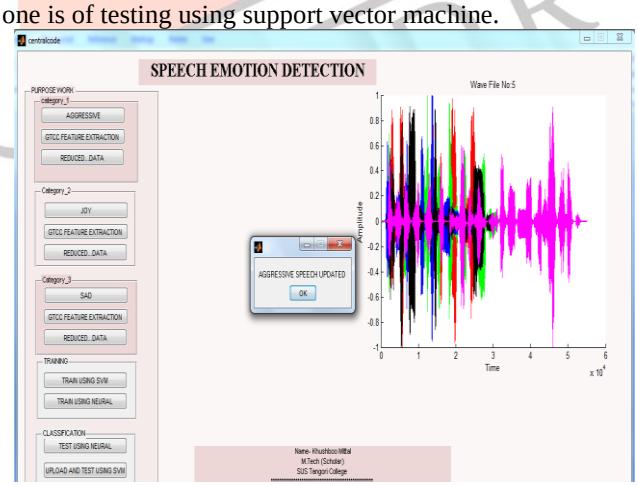


Fig.4 upload aggressive audio file

In above figure, it is shown that initially a speech file of .wav format 0is uploaded from the database. As the speech file is uploaded completely a graph will be plotted showing amplitude and time of the file as shown above. And also a dialog box will appear stating AGGRESSIVE SPEECH UPDATED.

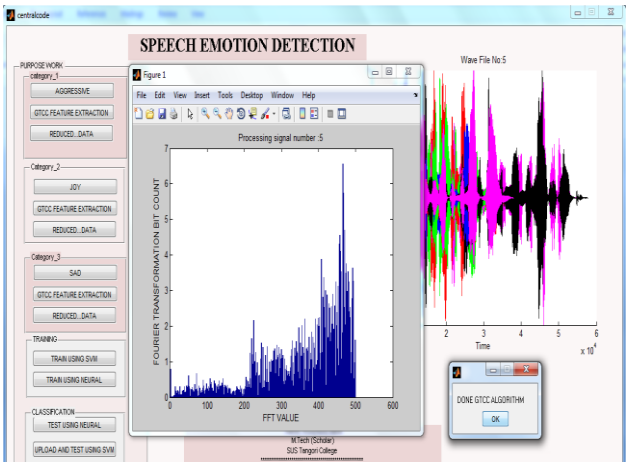


Figure.5 apply gtcc feature extraction

Once, speech file is uploaded, then apply gammatone cepstral coefficient for feature extraction form speech file. In above figure, when GTCC is applied a graph will appear which is plotted between FFT Value and Fourier transformation bit count as shown above. A dialog box will also appear stating Done GTCC algorithm that means execution of GTCC is completed.

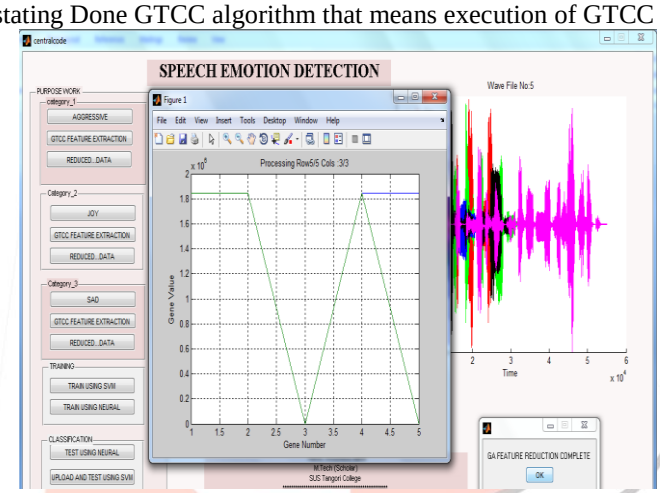


Fig.6 apply ga for optimization

In this figure genetic algorithm is applied on the previous GTCC results for optimization of extracted features from speech file. A graph is plotted between gene number and gene value showing GA processing in rows and columns. A dialog box will appear when execution of Genetic Algorithm is completed stating GA FEATURE REDUCTION COMPLETE as shown above.

Similarly, this process is repeated in other two category, such as speech file is uploaded of different emotion like joy and sad, then apply GTCC for feature extraction form Speech file and in the end apply genetic algorithm for feature reduction purpose.

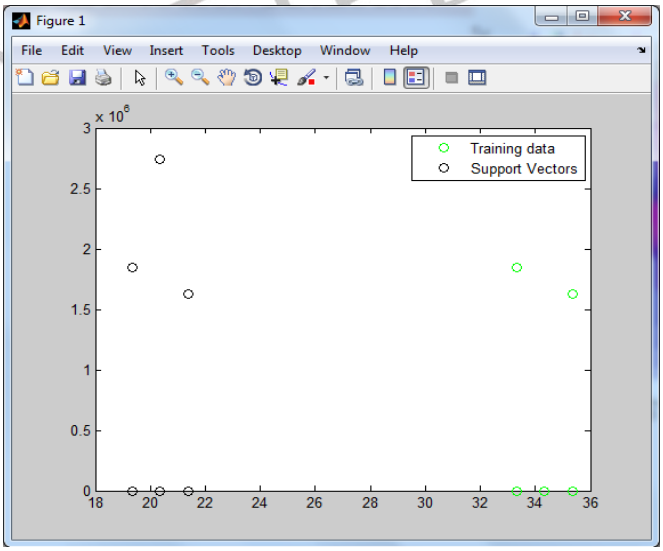


Fig.7 training using support vector machine algorithm

In above figure, once cat 1, cat 2, and cat 3 is uploaded then for training purpose Support Vector Machine will be applied. A graph will be generated showing training data with green color circle and SVM vectors with black circle.

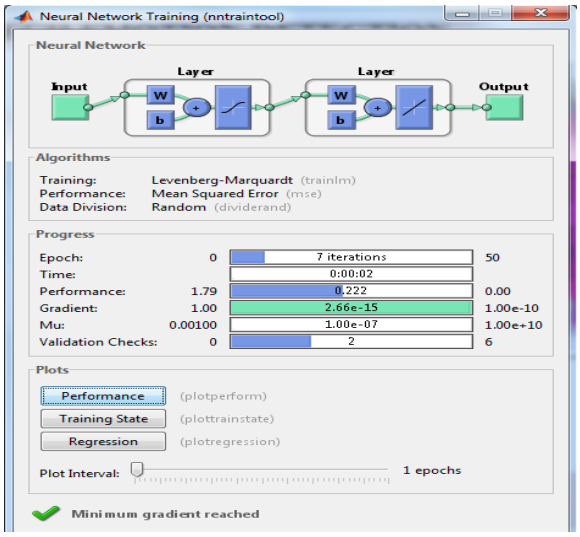


Fig.8 training by using neural network

In this figure, neural network is utilized for training purpose. Neural networks are those networks that are the collection of simple elements which function parallel. A neural network can be trained to perform a particular function by adjusting the values of the weights between elements. Network function is determined by the connections between elements.

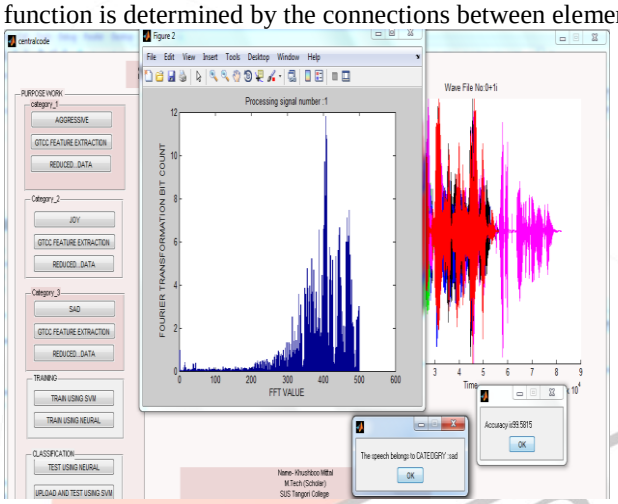


Fig.9 testing using neural network

In above figure, testing using neural network is done and above shown results are attained from it. In this first upload test speech file from the database, then neural network will be applied when speech file is uploaded which classify it that in which category it belongs to such that above figure shows test speech file belongs to category five and it also show the accuracy of the proposed system results which is 99.8639.

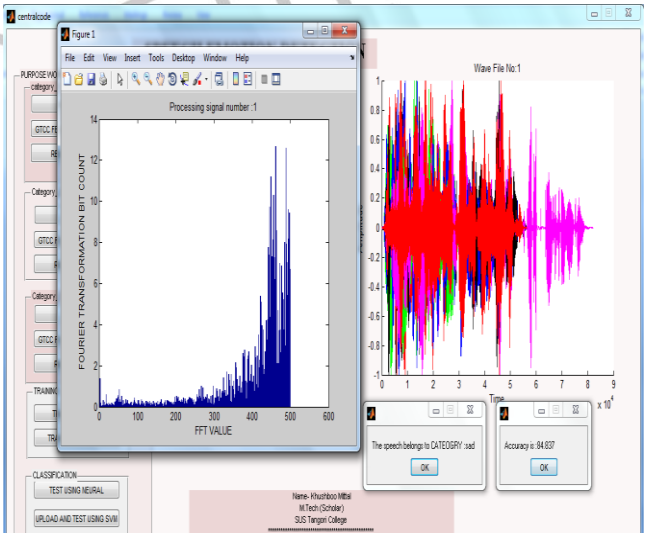


Fig.10 testing using svm

In above figure, testing using support vector machine is done and above shown results are attained from it. In this first upload test speech file from the database, then SVM will be applied when speech file is uploaded which classify it that in which category it belongs to such that above figure shows test speech file belongs to category five and it also show the accuracy of the proposed system results which is 84.911.

Table .1 result comparison table

Parameter	Result Value
Accuracy using SVM	84.911
Accuracy using NN	99.8639

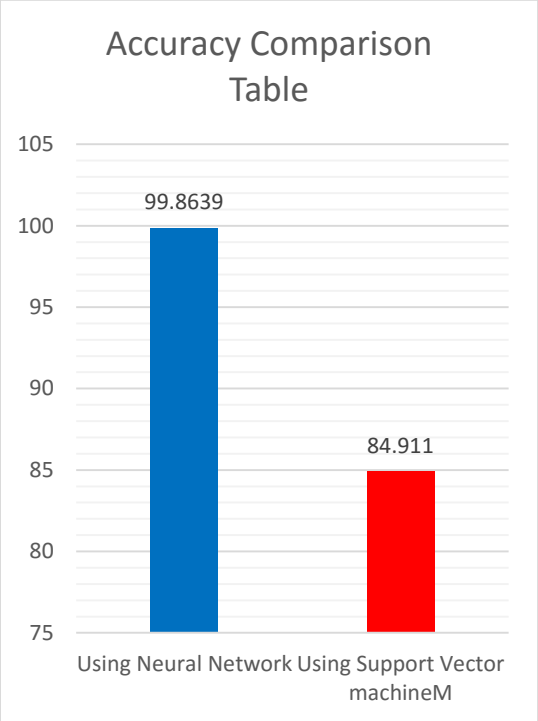


Fig.11 accuracy comparison graph

In above graph, it is shown that neural network works better for classification purpose as compared to Support Vector Machine.

IV. CONCLUSION

In this paper, the five emotions JOY, SAD, and AGGRESSIVE, based on automatic speech emotion recognition systems are overviewed. The Gamma Tone Cepstral Coefficient algorithm (GTCC) and Genetic algorithm approach is used for feature extraction and feature reduction. The methodology of emotion speech identification technique involves the optimization technique, Neural Network and Support Vector Machine method. The methodology does make available a real-world clarification to the problem of emotion recognition through speech and it can work well in constrained environment. In the end, accuracy results obtained from NN and SVM are compared using graph it shows that neural network works much better as compared to Support Vector machine as accuracy value of NN is 99.8639 and SVM is 84.911.
This work can be extended by integrating with Fuzzy C-means clustering algorithm for better efficiency.

V. REFERENCES

[1] J. Nicholson, K. Takahashi, R. Nakatsu, Emotion recognition in speech using neural networks,NeuralComput.Appl.9(2000)290–296.

[2] Schuller, B., Rigoll, G., & Lang, M. (2004, May). Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture. In Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on (Vol. 1, pp. I-577). IEEE.

[3] France, D. J., Shiavi, R. G., Silverman, S., Silverman, M., & Wilkes, D. M. (2000). Acoustical properties of speech as indicators of depression and suicidal risk. Biomedical Engineering, IEEE Transactions on, 47(7), 829-837.

[4] Fernandez, R. (2004). A computational model for the automatic recognition of affect in speech (Doctoral dissertation, Massachusetts Institute of Technology).

[5] Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. The Journal of the Acoustical Society of America, 52(4B), 1238-1250.

[6] R. Schluter, I. Bezrukov, H.Wagner, and H.Ney, "Gammatone features and feature combination for large vocabulary," in Proc. ICASSP, 2007.

[7] Md. Ali Hossain, Md. Mijanur Rahman, Uzzal Kumar Prodhan, Md.Farukuzzaman Khan "Implementation Of Back-Propagation NeuralNetwork For Isolated Bangla Speech Recognition" ,IJSCIT , 2013.

[8] DimitriosVerveridis and Constantine Kotropoulos, "Emotional speech recognition: Resources, features, and methods", Artificial Intelligence and Information Analysis Laboratory, Department of Informatics, Aristotle University ofThessaloniki, Box 451, Thessaloniki 541 24, Greece.

[9] Han, K., Yu, D., & Tashev, I. (2014). Speech emotion recognition using deep neural network and extreme learning machine. Proceedings of INTERSPEECH, ISCA, Singapore, 223-227.

[10] Rao, K. S., Kumar, T. P., Anusha, K., Leela, B., Bhavana, I., & Gowtham, S. V. S. K. (2012). Emotion recognition from speech. International Journal of Computer Science and Information Technologies, 3, 3603-3607.
[11] Vogt, T., André, E., & Wagner, J. (2008). Automatic recognition of emotions from speech: a review of the literature and recommendations for practical realisation. In Affect and emotion in human-computer interaction (pp. 75-91). Springer Berlin Heidelberg.
[12] R. Schluter, I. Bezrukov, H.Wagner, and H.Ney, "Gammatone features and feature combination for large vocabulary," in Proc. ICASSP, 2007.
[13] Y. Shao and D. Wang, "Robust identification using auditory features and computational auditory scene analysis," in Proc. ICASSP, 2008.