Speech Emotion Recognition using GFCC and BPNN

Shaveta Sharma[#1], Parminder Singh[#2]

[#1] Research scholar & Electronics and Communication Engineering & DIET Kharar-Punjab Technical University Jalandhar- India

[#2] Assistant Professor& Electronics and Communication Engineering & DIET Kharar-Punjab Technical University Jalandhar-India

***Abstract*— From the past years, researchers have showed a very interest in the speech recognition systems based on the emotions. Mainly the research has been done with the aim to bring closer both human and computer with each other by recognizing mood swings. In the recognizing process we must know how to represent the emotions on the basis of some features like sad, happy etc. When the verbal content is well recognized on the speaker's emotion, a promising enhancement of such systems would come. So recognition in speech is a crucial step. In the proposed work speech emotions will be recognized using hybridization of GFCC (Gammatone Frequency Cepstral Coefficients) and BPNN (Back Propagation Neural Network). The whole simulation will be done in the MATLAB environment to check the comparison result.**

***Keywords*— Back Propagation Neural Network, GFCC, Speech Recognition, Emotions.**

I. Introduction

Speech emotion recognition is one of the latest challenges in speech processing. Besides human facial expressions, speech has proven as one of the most promising modalities for the automatic recognition of human emotions. Especially in the field of security systems a growing interest can be observed throughout the last year. Besides, the detection of files, video games and psychiatric aid are often claimed as further scenarios for emotion recognition. Addressing classification in a practical [1] view it has to be considered that a technical approach can only rely on pragmatic decisions about kind, extent and number of emotions suiting the situation. It seems reasonable to adapt and limit this number and kind of recognizable emotions to the requirements given within the application to ensure a robust classification. Yet no standard exists for the classification of emotions in technical recognition. An often favoured way is to distinguish between a defined set of discrete emotions. However, as mentioned, no common opinion exists about their number and naming. A recent approach can be found in the MPEG4 standard, which names the six emotions anger, disgust, fear, joy, sadness and surprise. The addition of a neutral state seems reasonable to realize the absence of any of these emotions. This classification is used as a basis for the comparison throughout this work also expecting further comparisons. Most approaches in nowadays speech emotion recognition use global statistics of a phrase as basis [2].

We present two working engines using both alluded alternatives by use of continuous BPNN and GFCC method. Our system has been fully implemented and tested for audio wave files. Result analysis is done using neural and GFCC tool. The emotional speech input to the system is the collection of speech data. After collection of database which is considered as the training samples, necessary features were extracted from the speech signal to train the system using GFCC and BPNN algorithm. A feature set of various potentially features is extracted, analysed and database is prepared in excel spreadsheet. Then the recorded test samples is presented to the classifier which classifies the test sample into one of the emotion considered in our study and gives the recognized emotion as output. The remaining paper is organized as Section I will give the context of the speech recognition system, Section II explains the proposed emotion recognition system, Section III describes the step of feature extraction; Section IV discusses the results details. Finally, Section V discusses the Conclusion and future scope of the work.

II. Speech Emotion Recognition System

The proposed architecture of our SER system has following steps:

- Our speech processing system extracts some appropriate features from signal.
- Database is prepared for different emotions in excel spreadsheet
- Using BPNN and GFCC algorithm our system is trained in a supervised manner with example data how to associate the features to the different emotions. GFCC has been used to extract the features whereas the neural network aim is to train the network to achieve a balance between the ability to respond correctly to the input patterns that are used for training and the ability to provide good response to the input that are similar [3].
- Compare results.

III. Feature Extraction

Any emotion from the speaker's speech contains large number of parameters and the changes in these parameters will result in corresponding changes in emotions. In SER, feature extraction is one of the special forms of dimensionality reduction. Feature resources required to describe a large set of data accurately [4]. Basically feature extraction is based on

partitioning speech into small intervals known as frames. Speaker variability invariance degradation in the speech signal is due to the channel and noise. In proposed work feature extraction is done using GFCC method.

The goal is to find the set [5] of properties of an utterance that have an acoustic correlates in the speech signal, that is, the parameters that somehow are computed or estimated through processing of the signal waveform. Such parameters are termed as features.

IV. Implementation & Results

Implementation is taken place in MATLAB using BPNN classifier [6]. In Classification, training examples are used to learn a model that can classify the data samples into known classes. The Classification process involves following steps:

- Create training data set.
- Identify class attribute and classes.
- Identify useful attributes for classification (Relevance analysis).
- Learn a model using training examples in Training set.
- Use the model to classify the unknown data samples for angry emotion.

Table 1.

Modification Table for Angry emotion using Hybridization

Method	Accuracy
Hybridization (GFCC+BPNN)	98.5%

Above table shows that by using hybridization accuracy of angry emotion has been enhanced to 98.5%.

V. Conclusion & Future Scope

We believe that this contribution shows important results considering emotion recognition with BPNN and GFCC. The two introduced methods proved both capable of a rather reasonable model for the automatic recognition of human emotions in speech.

Furthermore some emotions seem to be recognized more easily. This may be due to the fact that the most test patterns were acted emotions and test-persons have difficulties with feigning certain emotions. Though the same training material and test sets were used, the two proclaimed solutions differ greatly in their behaviour. One reason for the better performance can be seen in the loss of information of durations of voiced sounds by eliminating these in the contours as described. The results of both engines reach the abilities of a human decider as described above.

In our future work we aim at a hybrid approach combining neural networks [7] and hidden Markov models for the automatic recognition. Also the integration of other modalities such as video based or manual interaction will be investigated further.

References

[1] J. C. Platt, "*Fast Training of Support Vector Machines using Sequential Minimal Optimization*", Advances in Kernel Methods: Support Vector Learning, pp 185–208, 2000.

[2] Md. Ali Hossain, Md. Mijanur Rahman, Uzzal Kumar Prodhan, Md. Farukuz zaman *Khan, "Implementation Of Back-Propagation Neural Network For Isolated Bangla Speech Recognition"*, International Journal of Information Sciences and Techniques , vol.3, pp 1-9,2013.

[3] Raul Rojas, "*Neural Networks: A Systematic Introduction. Springer*", 2005.

[4] Dimitrios Ververidis and Constantine Kotropoulos *"Emotional speech recognition:Resources, features, and methods"*,Speech Communication 48, pp. 1161-1182,2006.

[5] John C. Platt, "Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines", Advances in Neural Information Processing Systems 11, M. S. Kearns, S. A. Solla, D. A. Cohn, eds., MIT Press, 1999.

[6] Wouter Gevaert, Georgi Tsenov, Valeri Mladenov, "*Neural Networks used for Speech Recognition*", Journal of Automatic Control, University of Belgrade, vol. 20:1-7, 2010.

[7] Mirza Cilimkovic, "*Neural Networks and Back Propagation Algorithm*", Institute of Technology Blanchardstown, Blanchardstown Road North Dublin 15, Ireland.