

Speech Based Emotion Recognition Using Feature Normalization and Neural Network

Roshni Pandey, Y.A. Nafde, J.C. Kolte

PG Student, Dept. of Electronics and Telecommunication, Priyadarshini College of Engineering and Technology,
Nagpur, Maharashtra, India

Assistant Professor, Dept. of Electronics and Telecommunication, Priyadarshini College of Engineering and
Technology, Nagpur, Maharashtra, India

Assistant Professor, Dept. of Electronics and Telecommunication, Priyadarshini College of Engineering and
Technology, Nagpur, Maharashtra, India

ABSTRACT: In this paper, we have determined the emotional states of a person by analysing the speech features of the sound signals by employing feature normalization and neural network. Human emotion justifies the mental state of a person for the proper interaction with a machine and also helpful in many other areas. Here we will be extracting the various speech features by normalization and analysing the speech signal of a person in different mental states. Then these features are classified where majority voting technique applied over Support Vector Machine (SVM), Neural Network (NN) and the combination of SVM and neural networks interpret the exact class of the emotion. Nowadays, emotion recognition plays important role for the understanding of human reaction in various situations and towards one another.

KEYWORDS: Emotional states, Normalization, sound features, Neural networks (NN), and Support vector machine (SVM).

1. INTRODUCTION

Emotional states of a person interpret the mental state whether the person is happy, sad etc. Hence this identifies the importance of emotion recognition. The requirement of human emotion recognition is increasing rapidly so that we can introduce a revolution in the field of machine intelligence and make it possible for the machines to take decisions as human do. These machines will be able to interact with the other machines and the environment. It is also known to us that emotion recognition is very important for decision making hence it is very important for machines to determine emotion of a person.

Humans interact with each other mainly through speech, but also through body gestures, to emphasize a certain part of the speech and display of emotions. Normally humans exchange their information via the natural sensory modes of sight, sound, and touch. In face-to-face information exchange, humans employ these communication paths simultaneously and in combination. Machines may never need all of the emotional skills that humans need but they will inevitably require some of these skills to appear intelligent when interacting with people.

Human speech emotions are generally classified into six different categories as sadness, fear, happiness, surprise, anger and disgust but here we will recognize only four emotional states that are anger, sadness, happiness and normal. Firstly, the speech features are analysed. Then, these signals are normalised and classified by majority voting technique applied over Neural Network (NN), SVM and their combination. Some of the domains where it is highly usable to enhance working capabilities include tutoring systems, health care, games, call centres and ambient intelligent environments.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2015

Normalization scheme is used to make a database of speech signals and estimate their linear scaling parameters for each speaker depending upon his/her speech signal. Then, the normalization parameters are applied to the speech.

II. PROPOSED METHODOLOGY

a) . NORMALIZATION PROCESS

Feature Normalization –

Normalization can be done by separately estimating the normalization parameters for neutral speech signals and those estimated normalization parameters are applied to the speech signal. These normalizations are performed such that the properties of their neutral speech become similar across speakers (e.g., first and second order statistics). This can be done by applying feature normalization over training data and then test the data using mean and variance.

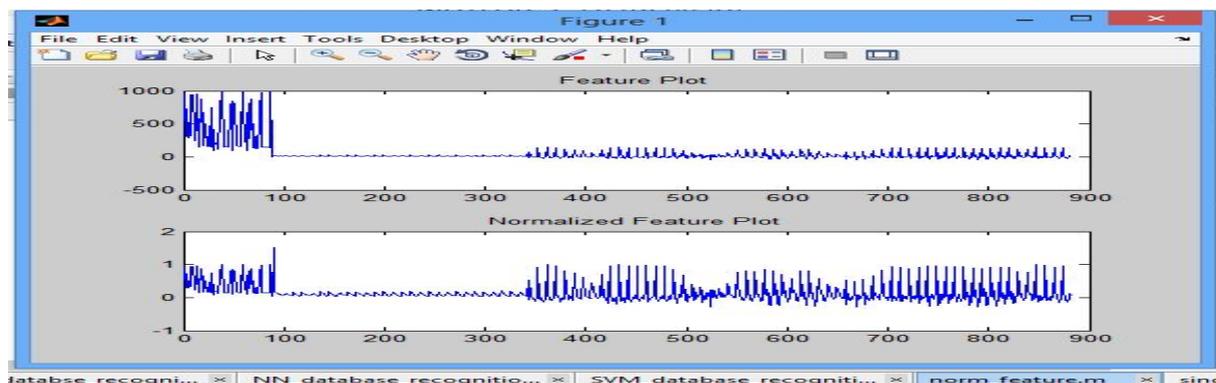


Fig [a]

A). Neural Network –

Neural networks are machine learning or computational model inspired by animals central nervous system (brain). It is used to find exact output from the large number of inputs. It uses large number of neurons that sends messages to one another. Neural network is used for training and classification of the normalized speech features.

C). Support Vector Machine (SVM) – SVM is a type of classifier which is very powerful and having good accuracy. It is also used as one of the tool in speech based emotion recognition for pattern recognition. Whenever we use SVM, the variable length data is first converted into the fixed length data as it can only be applied to linear data stream. It uses a fitting function.

III. RELATED WORK

In [7] authors tell about normalization related work. How this process occurs and why this process is employed in speech features. The authors in [8] explain the various speech features and their behaviour. Also it explains the process of feature extraction. V. Petrushin [4] proposed a system implemented an artificial neural network for emotion recognition for call centres from where i get the knowledge about neural networks as it is the main part of the system used for the classification of emotional states. Similarly Yixiong Pan, Peipei Shen and Liping Shen [2] proposed Speech Emotion Recognition Using Support Vector Machine that describes about SVM machine and its processing that helps me out to properly understand the SVM machine working.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2015

IV. STRUCTURE OF SPEECH EMOTION RECOGNITION SYSTEM

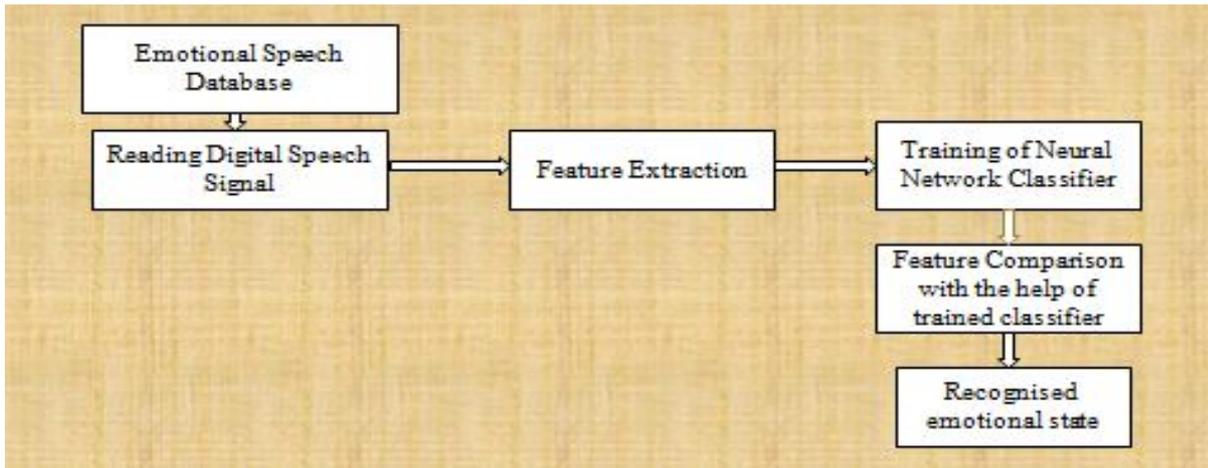
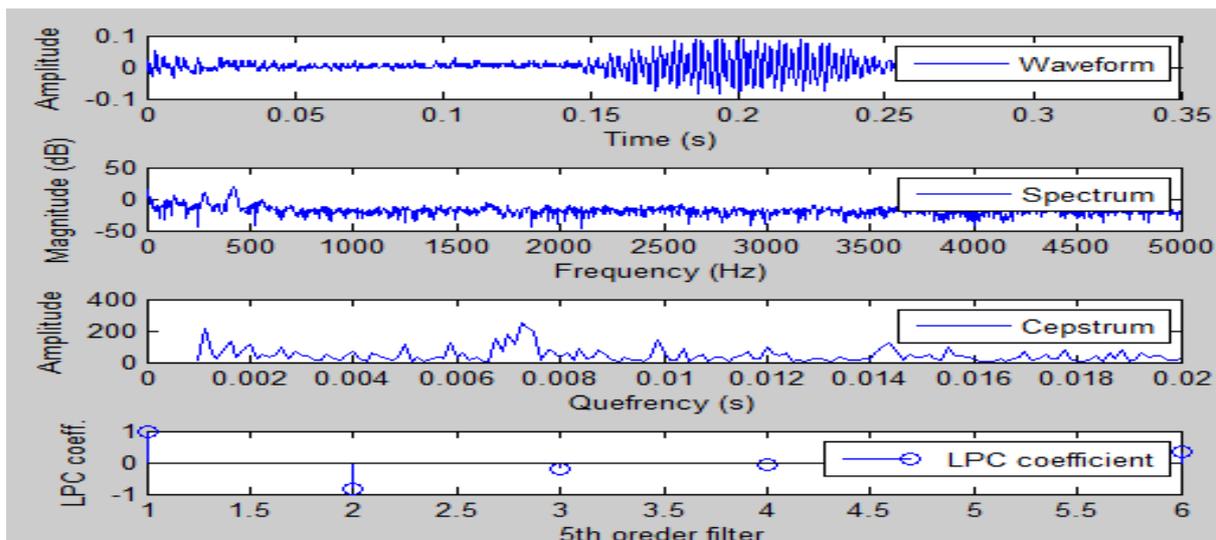


Fig [1]

V. EXPERIMENTS AND RESULTS

The experiment is performed in MATLAB 7.9 version. Fig [a] above shows the result of normalization process. Here in fig 2, Speech feature plot is shown.



Fig[2]

Here fig 3 shows the database selection process for the selection of input speech signals that are estimated and classified by SVM classifier to get the four emotional states (sadness, anger, happiness and normal) of a person shown in fig 5. SVM gives the output in binary state means considering only two signals at a time.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2015

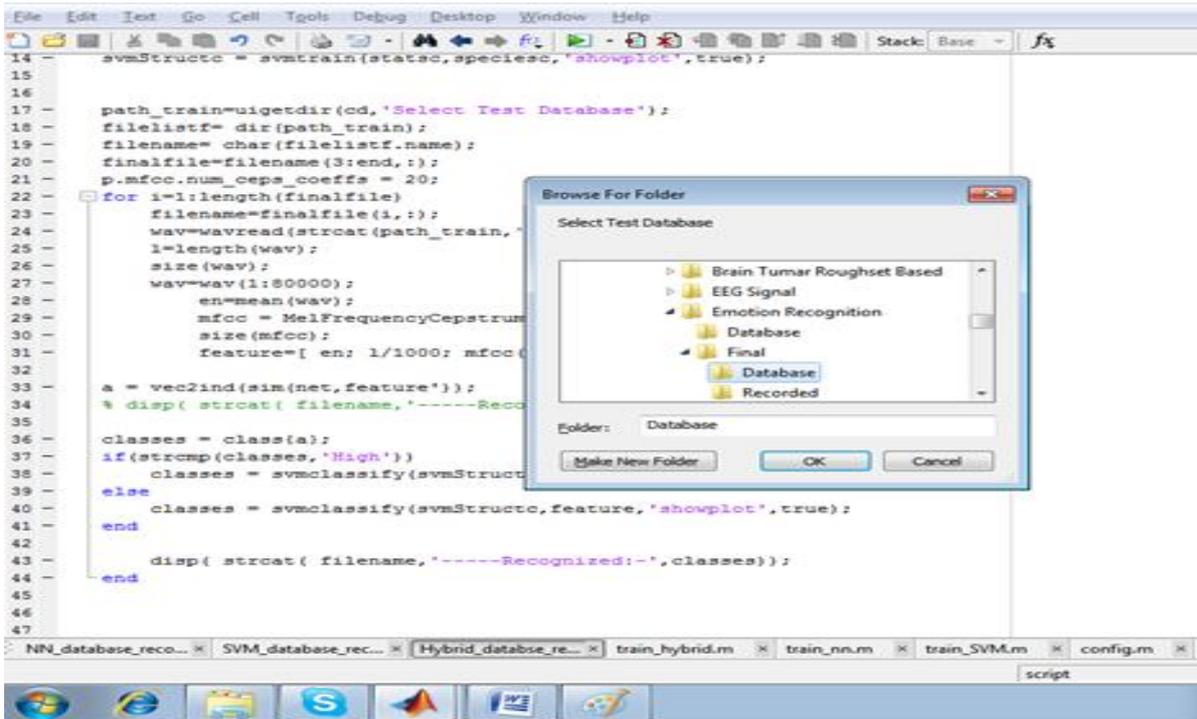


Fig [3]

Command Window	Command Window
'sa01.wav-----Recognized:-Sad'	'a01.wav-----Recognized:-Anger'
'sa02.wav-----Recognized:-Sad'	'a02.wav-----Recognized:-Anger'
'sa03.wav-----Recognized:-Sad'	'a03.wav-----Recognized:-Anger'
'sa04.wav-----Recognized:-Sad'	'a04.wav-----Recognized:-Anger'
'sa05.wav-----Recognized:-Sad'	'a05.wav-----Recognized:-Anger'
'sa06.wav-----Recognized:-Sad'	'a06.wav-----Recognized:-Sad'
'sa07.wav-----Recognized:-Sad'	'a07.wav-----Recognized:-Happy'
'sa08.wav-----Recognized:-Normal'	'a08.wav-----Recognized:-Anger'
'sa09.wav-----Recognized:-Normal'	'a09.wav-----Recognized:-Sad'
'sa10.wav-----Recognized:-Normal'	'a10.wav-----Recognized:-Sad'
'sa11.wav-----Recognized:-Normal'	'a11.wav-----Recognized:-Normal'
'sa12.wav-----Recognized:-Sad'	'a12.wav-----Recognized:-Happy'
'sa13.wav-----Recognized:-Sad'	'a13.wav-----Recognized:-Happy'

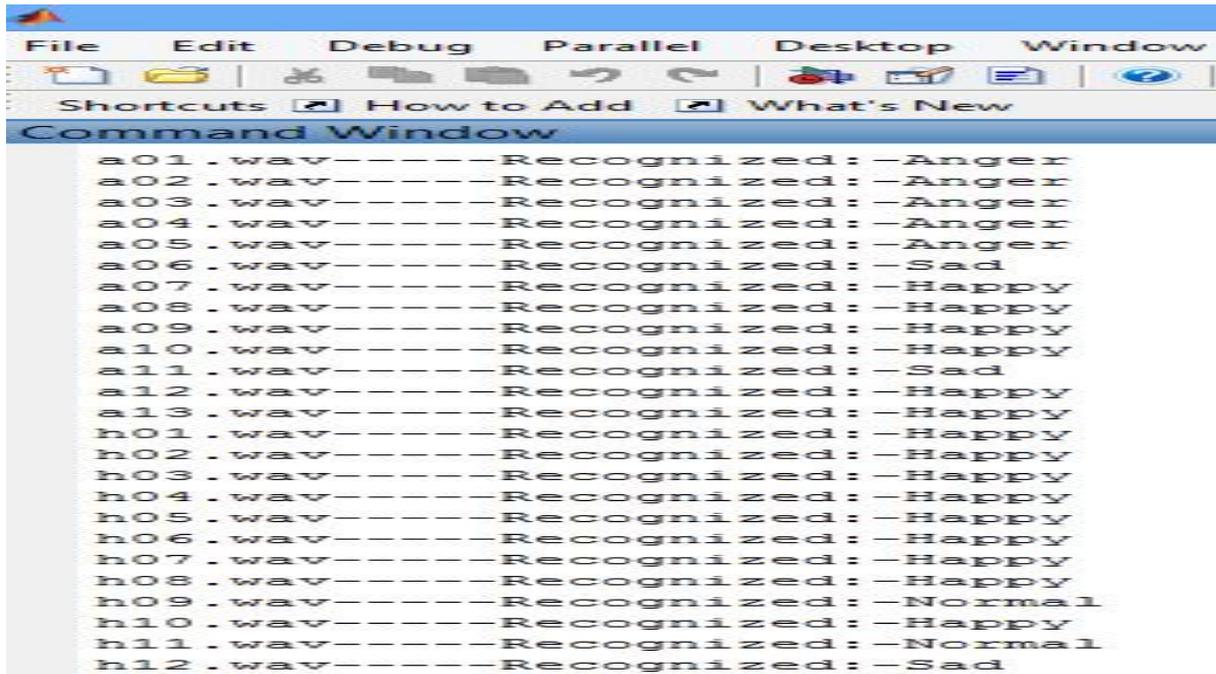
Fig [4]

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2015

The output of neural network is shown in fig 5. It classifies the various emotional states of a person.

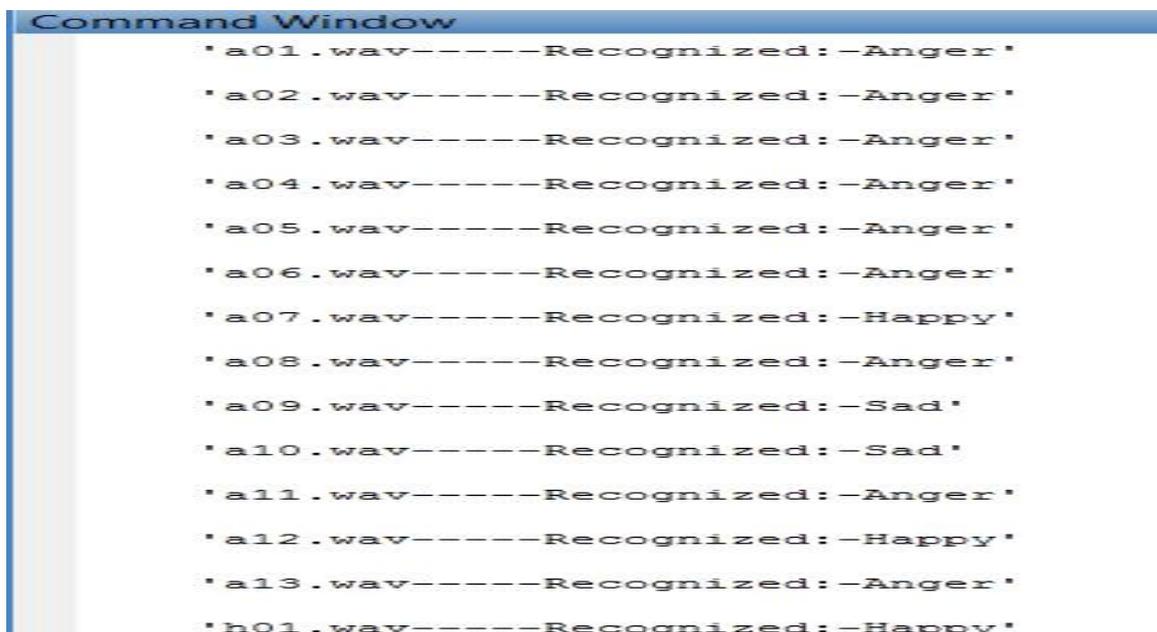


```

File Edit Debug Parallel Desktop Window
Shortcuts [x] How to Add [x] What's New
Command Window
a01.wav-----Recognized:-Anger
a02.wav-----Recognized:-Anger
a03.wav-----Recognized:-Anger
a04.wav-----Recognized:-Anger
a05.wav-----Recognized:-Anger
a06.wav-----Recognized:-Sad
a07.wav-----Recognized:-Happy
a08.wav-----Recognized:-Happy
a09.wav-----Recognized:-Happy
a10.wav-----Recognized:-Happy
a11.wav-----Recognized:-Sad
a12.wav-----Recognized:-Happy
a13.wav-----Recognized:-Happy
h01.wav-----Recognized:-Happy
h02.wav-----Recognized:-Happy
h03.wav-----Recognized:-Happy
h04.wav-----Recognized:-Happy
h05.wav-----Recognized:-Happy
h06.wav-----Recognized:-Happy
h07.wav-----Recognized:-Happy
h08.wav-----Recognized:-Happy
h09.wav-----Recognized:-Normal
h10.wav-----Recognized:-Happy
h11.wav-----Recognized:-Normal
h12.wav-----Recognized:-Sad
  
```

Fig [5]

The output of Hybrid network consists of neural network and SVM where SVM is used for training and neural network is used as classifier shown in fig 6.



```

Command Window
'a01.wav-----Recognized:-Anger'
'a02.wav-----Recognized:-Anger'
'a03.wav-----Recognized:-Anger'
'a04.wav-----Recognized:-Anger'
'a05.wav-----Recognized:-Anger'
'a06.wav-----Recognized:-Anger'
'a07.wav-----Recognized:-Happy'
'a08.wav-----Recognized:-Anger'
'a09.wav-----Recognized:-Sad'
'a10.wav-----Recognized:-Sad'
'a11.wav-----Recognized:-Anger'
'a12.wav-----Recognized:-Happy'
'a13.wav-----Recognized:-Anger'
'h01.wav-----Recognized:-Happy'
  
```

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2015

VI. CONCLUSION

Here we have shown the emotion recognition of a speech signal by using normalization technique and also shown its classification with the help of SVM, Neural network and the combination of both i.e., hybrid network. Speech feature extraction is also shown above. Our system successfully detects the four emotional states that is anger, sadness, happiness and normal of the speech signal.

REFERENCES

- [1]. Yixiong Pan, Peipei Shen and Liping Shen, "Speech Emotion Recognition Using Support Vector Machine", *International Journal of Smart Home*, vol. 6, No. 2, April, 2012.
- [2]. Shruti Aggarwal, Naveen Aggarwal, "Classification of Audio Data using Support Vector Machine", *IJCST*, vol. 2, Issue 3, September 2011.
- [3]. Carlos Busso, Soroosh Mariooryad, Student Member, Angeliki Metallinou and Shrikanth Narayanan, "Iterative Feature Normalization Scheme for Automatic Emotion Detection from Speech", *IEEE transactions on affective computing*, vol. 4, no. 4, october-december 2013
- [4]. F. Dellaert, T. Polzin, and A. Waibel, "Recognizing emotion in speech," in Proc. ICSLP, Philadelphia, PA, 1996, pp. 1970 – 1973
- [5]. V. Petrushin, "Emotion in speech: Recognition and application to call centers," *Artif. Neu. Net. Engr. (ANNIE)*, 1999
- [7]. V. Sethu, E. Ambikairajah, and J. Epps, "Speaker Normalisation for Speech Based Emotion Detection," Proc. 15th 15th International Conference on Digital Signal Processing Processing - , pp. 611-614, July 2007.
- [8]. M.N.Hasrul, M.Hariharan, Sazali Yaacob, "Human Affective (Emotion) Behaviour Analysis using Speech Signals: A Review," 2012 International Conference on Biomedical Engineering (ICoBE), 27-28 February 2012, Penang.