International Journal of Interdisciplinary Innovative Research &Development (IJIIRD)
ISSN: 2456-236X
Vol. **05** Special Issue **01** | **2020**

ANN for Machine Intelligence

Dr. Moiz A. Hussain[1], Prof. M. U Karande[2], Prof. Y. P. Sushir[3]
[1]Associate Professor, Department of Electrical Engineering, Padm. Dr. V. B. Kolte College of Engineering, Malkapur
[2,3] Assistant Professor, Department of Computer Science & Engineering, Padm. Dr. V. B. Kolte College of Engineering, Malkapur

Abstract: This paper presents a sex-independent emotion recognition system for the identification of human affective state in speech signal. A Berlin Database of emotional speech [4] consisting of ten (10) sentences spoken by five (5) male & five (5) female is used for testing the feasibility of the system. The potential prosodic features are first identified and extracted from speech data then we introduce the Generalized Feed Forward network to realize the classification of human emotions. The recognition rates by the Generalized Feed Forward Classifier were 100% for neutral, anger, disgust, fear, happiness, sadness & 95% for boredom.
Keywords: Generalized Feed Forward network; Emotion recognition; Prosodic features; Berlin Database of emotional speech.

1. INTRODUCTION

Research has long been done on emotion in the fields of psychology and physiology. More recently it is the subject of attention by engineers. Its most important application is in intelligent human-machine interaction. In today's human-machine interaction systems, machines can recognize "what is said" and "who said it" using speech recognition and speaker identification techniques. If equipped with emotion recognition techniques, machines can also know "how it is said" to react more appropriately, and make the interaction more natural. In the field of Human Computer Interaction (HCI), speech is primary to the objectives of an emotion recognition system, as are facial expressions and gestures. It is considered a powerful mode to communicate intentions and emotions.

This paper explores methods by which a computer can recognize human emotion in speech signal; such methods can contribute to applications such as learning environments, consumer relations, entertainment etc. Speech materials, which represent the emotion of neutral, anger, boredom, disgust, fear, happiness, and sadness were analyzed by studying the features of speech. Algorithms were developed to extract the features of speech. Then analyzed emotional speech parameters are learned by Neural Networks.

2. EMOTION RECOGNITION SYSTEM

The structural components of the emotion recognition system are depicted in Figure 1. It consists of five modules: speech input, feature extraction, Generalized Feed Forward neural network for classification, and the recognized emotion output. Here input is speech signals from Berlin Database [4] and after extracting speech features Neural Network is trained using GFF classifier, to recognize emotions as an output.

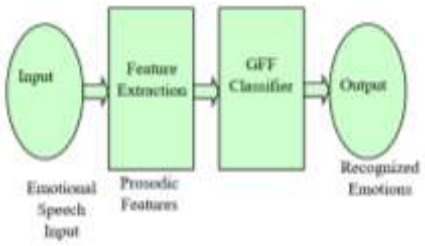

Figure1: Structure of Emotion Recognition System

2.1 Emotional Speech Database

In order to build an effective sex independent emotion recognition system and test its feasibility, a Berlin Database of emotional speech [4] was used consisting of ten (10) sentences spoken by five (5) male & five (5) female for seven classes: neutral, anger, boredom, disgust, fear, happiness, and sadness. Since our aim is to develop a sex independent system, subjects from German language backgrounds & both male & female are used in this paper. In this approach, speech signal was re-sampled to 10 kHz, and the silence segments at the beginning and the end of speech were cut out artificially. Then the whole database was divided into two parts for the purpose training & cross-validation.

2.2 Feature Extraction

It is believed that prosodic features are the primary indicator of speaker's emotional states. The rationale for

feature selection is that new or reduced features might perform better than the base features because we can eliminate irrelevant features from the base feature set. This can also reduce the dimensionality, which can otherwise hurt the performance of the pattern classifiers. Research to analyze emotional speech indicates that fundamental frequency, energy and formant frequencies are potentially effective parameters to distinguish certain emotional states. Five fundamental formant frequencies (f0 to f4), Minima (It describes the number of times minimum occurring in the speech signal), Five entropy (Shannon, Log, Threshold, Sure, Norm), median, variance, LPC were calculated from speech.

The fundamental formant frequencies were extracted by the method discussed in [8], entropies were estimated using speech toolbox in MATLAB; variance was calculated using the covariance matrix. A total number of 14 features are extracted by analyzing speech spectrogram. These 14 possible candidates are listed in Table 1 we consider all possible candidate features in feature extraction. However, some of the features may be redundant or even cause negative effects. But we have trained the neural network using all the feature vectors & recognized the human emotions.

Table 1: List of 14 Feature Vectors

Sr. No.	Feature	Sr. No.	Feature
1	Formant0	8	Log Entropy
2	Formant1	9	Threshold Entropy
3	Formant2	10	Sure Entropy
4	Formant3	11	Norm Entropy
5	Formant4	12	Median
6	Minima	13	Variance
7	Shannon Entropy	14	LPC

O/P Desired	On	Oa	Ob	Od	Of	Oh	Os
On	14	0	0	0	0	0	0
Oa	0	23	0	0	0	0	0
Ob	0	0	19	0	0	0	0
Od	0	0	1	4	0	0	0
Of	0	0	0	0	14	0	0
Oh	0	0	0	0	0	20	0
Os	0	0	0	0	0	0	10

3. GENRALIZED FEEDFORWRAD NEURAL NETWORK

Generalized feed forward networks are a generalization of the Multilayer Perceptron (MLP) such that connections can jump over one or more layers. In theory, a MLP can solve any problem that a generalized feed forward network can solve. In practice, however, generalized feed forward networks often solve the problem much more efficiently. A classic example of this is the two spiral problem. Without describing the problem, it suffices to say that a standard MLP requires hundreds of times more training epochs than the generalized feed forward network containing the same number of processing elements.The structure of feed forward neural network is shown in figure 2. In this type of neural netwrok information is proceeded for the input layer to hidden layer & through the hidden layer to output layer

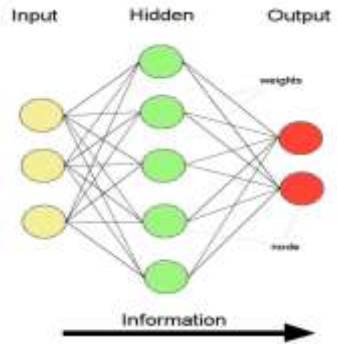

Figure 2: Feed forward network with one hidden layer & one output layer.

3.1 Experimental Results Using Neural Network

The GFF classifier was used to test the proposed feature vector of speech. The Leave N Out training method was used in the GFF classifier to train the neural network, and experimental results were obtained by using 20% cross-validation. The best recognition rate of 100% was obtained for neutral, anger, disgust, fear, happiness, sadness & 95% for boredom.

The experimental results for cross-validation data & for training data are shown in table 2 & 3.

Table 2: GFF recognition results of cross- validation dataset

O/P Desired	On	Oa	Ob	Od	Of	Oh	Os
On	62	0	0	0	0	0	0
Oa	1	103	0	0	1	0	0
Ob	0	0	57	0	0	0	0
Od	0	0	0	38	0	0	0
Of	0	1	0	1	55	0	0
Oh	0	0	0	1	0	49	0
Os	0	0	0	1	1	0	50

In table, **On**: Output Neutral, **Oa**: Output Anger, **Ob**: Output Boredom, **Od**: Output Disgust, **Of**: Output Fear, **Oh**: Output Happiness, **Os**: Output Sadness

Table 3: GFF recognition based on training dataset

Performance	On	Oa	Ob	Od	Of	Oh	Os
MSE	0.005	0.007	0.014	0.004	0.008	0.006	0.002
NMSE	0.039	0.039	0.094	0.100	0.073	0.039	0.028
MAE	0.053	0.057	0.060	0.049	0.059	0.056	0.046
Min Abs Error	0.001	0	0.001	0.004	0.001	0.001	0.003
Max Abs Error	0.309	0.484	1.055	0.223	0.511	0.316	0.103
r	0.984	0.981	0.955	0.954	0.970	0.984	0.994
%Correct	**100**	**100**	**95**	**100**	**100**	**100**	**100**

The table 4 & 5 shows the result of cross-validation data & training data for GFF neural network based on the performance

4. DISCUSSION & CONCLUSION

This paper proposes a sex-independent system for emotion recognition in human speech using neural networks.

Author have designed & implemented Generalized Feed Forward neural network for emotion recognition. Features based on the formant frequency, entropy, minima variance, LPC were extracted as the candidate input to the GFF classifier. The leave n out training method was used to train the GFF network. System obtained relatively high accuracy of 100% for neutral, anger, disgust, fear, happiness & sadness, 95% accuracy was obtained for boredom. The overall accuracy of the system comes out to be 100%.

5. REFERENCES

(1) Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., and Taylor, J. G..: "Emotion recognition in human-computer interaction", IEEE Signal Processing magazine, Vol. 18, No. 1, pp. 32-80, Jan. 2001.
(2) Muhammad Waqas Bhatti[1], Yongjin Wang[2] and Ling Guan[3].: "A Neural Network approach for Human Emotion Recognition in Speech", 0-7803-8251-X/04/2004 IEEE.
(3) Lili Cai, Chunhui Jiang, Zhiping Wang, Li Zhao, Cairong Zou, "A Method Combining The Global And Time Series Structure Features For Emotion Recognition In Speech", IEEE Int. Conf. Neural Networks & Signal Processing Nanjing, China, December 14-17, 2003 0-7803-7702-8/03/ 2003 IEEE.
(4) Felix Burkhardt, Miriam Kienast, Astrid Paeschke and Benjamin Weiss.:"Berlin Database of Emotional Speech" available at http://pascal.kgw.tu-berlin.de/emodb/.
(5) Yi-Lin, Gang Wei.: "Speech Emotion Recognition Based on HMM and SVM", Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, 18-21 August 2005 0-7803-9091-1/05/2005 IEEE.
(6) J. Nicholson, K. Takabashi and R. Nakatsu.: "Emotion Recognition in Speech Using Neural Network", Neural Information Processing, 1999.
(7) Li Zhuo, Xiungmin Qiun, Cuirong Zou, Zhenyung Wu.: "A Study on Emotional Feature Analysis and Recognition in Speech Signal" Journal of China Institute of Communications, Vo1.21, No.10, pp18-25, 2000.
(8) François Thibault," Formant Trajectory Detection Using Hidden Markov Models", Special Project Course Report MUMT 609, December 14 2003.