International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-6, Issue-4 E-ISSN: 2347-2693

An Application Using Radial Basis Function Classification in Stress Speech Identification

N.P. Dhole[1*], S.N. Kale[2]

[1]Department of ETE, PRMITR Badnera, Amravati, India
[2]Department of Applied Electronics, Sant Gadge Baba Amravati University, Amravati, India

**Corresponding Author: npdhole34@gmail.com*

Available online at: www.ijcseonline.org

Received: 20/Mar/2018, Revised: 28/Mar/2018, Accepted: 19/Apr/2018, Published: 30/Apr/2018

Abstract- Speech of human beings is the reflection of the state of mind. Proper evaluation of these speech signals into stress types is necessary in order to ensure that the person is in a healthy state of mind. In this work we propose a RBF classifier for speech stress classification algorithm, with sophisticated feature extraction techniques as Mel Frequency Cepstral Coefficients (MFCC). The RBF algorithm assists the system to learn the speech patterns in real time and self-train itself in order to improve the classification accuracy of the overall system. The proposed system is suitable for real time speech and is language and word independent. The human behaviour considers six basic emotions which are happiness, sadness, anger, fear, surprise & disgust. It becomes important to detect emotional state of a person which will be induced by workload, background noise, physical environmental factors (e.g. G-force) & fatigue. Broadly, stress identification becomes a scientific challenge to analyze a human being interaction with environment

Index Terms: RBF, MFCC, Stress Classification, Feature Selection.

I. INTRODUCTION

Stress Identification is remarkably gained high attention in various fields from two decades. The fields are Medical, Forensics, Smart Environments, Teaching Learning Education, Human computer interactions, Emergency services and of course Real Time situations which is utmost crucial. From many years different speech recognition software's [1] has been developed to speed up the accuracy using various classifiers on several databases [2]. We have also revised the literature review of numerous researchers for the same work [3,4,5,6,7,8,9]. We have used for this work the Berlin database [8,9] and Humane database [10,11,12,13] as Benchmark Datasets. Again we have recorded our speech samples using Audacity software with different frequencies. Speech signal recorded was of people having male, female voices including children above eight years and elder's up to age of 58. Stress is the output to physical or mental challenges. The emotional state can affect the speech characteristics. The speech signal is an interesting source for stress analysis, since it can be seen from literature review that various researchers have worked on stress analysis in different ways. A speech monitoring system is that system which is able to quantify a speaker's degree of stress. Recent studies demonstrate the potential for reliable stress classification via nonlinear, articulatory and speech production features .Once a period of speech under stress has been identified, an identification system incorporating a compensation procedures specific to that form of stress could be used.

This paper proposes RBF algorithm to detect and classify the human speech into different stress classes, and thereby provide a preliminary analysis of the type of stress which the person might be undergoing. Doing this can help the person to analyze the stress and obtain remedies for the same. The whole Algorithm is developed in MATLAB Software.

II. BERLIN DATABASE

The article describes a database of emotional speech. Ten actors (5 Female and 5 Male) simulated the emotions, producing 10 German utterances (5 short and 5 longer sentences) which could be used in everyday communication and are interpretable in all applied emotions [8]. The recordings were taken in an anechoic chamber with high-quality recording equipment. In addition to the sound electro-glottograms were recorded. The speech material comprises about 800 sentences (seven emotions * ten actors * ten sentences + some second versions). The complete database was evaluated in a perception test regarding the recognisability of emotions and their naturalness [9]. Utterances recognised better than 80% and judged as natural by more than 60% of the listeners were phonetically labelled

in a narrow transcription with special markers for voice-quality, phonatory and articulatory settings and articulatory features.

III. HUMAINE DATABASE

The database proper is a selected subset of the data with systematic labelling, mounted on the ANVIL platform [10,11,12,13,14]. It is designed to provide a concrete illustration of key principles rather than to be used as it stands in machine learning. Stage 1 (available via the HUMAINE portal at www.emotion-research.net) contains 50 'clips' from naturalistic and induced data, showing a range of modalities and emotions, and covering a balanced sample of emotional behaviour in a range of contexts. Emotional content is described by a structured set of labels attached to the clips both at a global level, and frame-by-frame, showing change over time. Labels for a range of signs of emotion have also been developed and applied to a subset of the clips: these include core signs in speech and language, and descriptors for gestures and facial features that draw on standard descriptive schemes.

IV. AUDACITY SOFTWARE

Audacity is a free and Open Source Software, it's an easy-to-use audio editor and recorder for Windows, Mac OS X, GNU/Linux, and other operating systems. Audacity is free software, developed by a group of volunteers and distributed under the GNU General Public License (GPL) [15]. We can use Audacity to Record live audio, Convert tapes and records into digital recordings or CDs Edit Ogg Vorbis, MP3, and WAV sound files to Cut, copy, splice, and mix sounds together to Change the speed or pitch of a recording. Audacity can record live audio through a microphone or mixer, or digitize recordings from cassette tapes, vinyl records, or minidiscs. In this research work we have recorded the speech using audacity with different frequencies 8 kHz, 16 kHz and 44.1 kHz.

Table no.1: Elements of Database

Databases	Marathi, Hindi, Berlin, Humaine
Features	MFCC
Classifier	RBF
Output	.mat files
Results	Images of MATLAB Software

Table no. 1 shows the elements of databases. Output of RBF Algorithm is saved in .mat files so that we can separately process various processes easily. Results of these are taken when we run the codes and get the images in MATLAB windows.

V. FLOWCHART OF RBF ALGORITHM

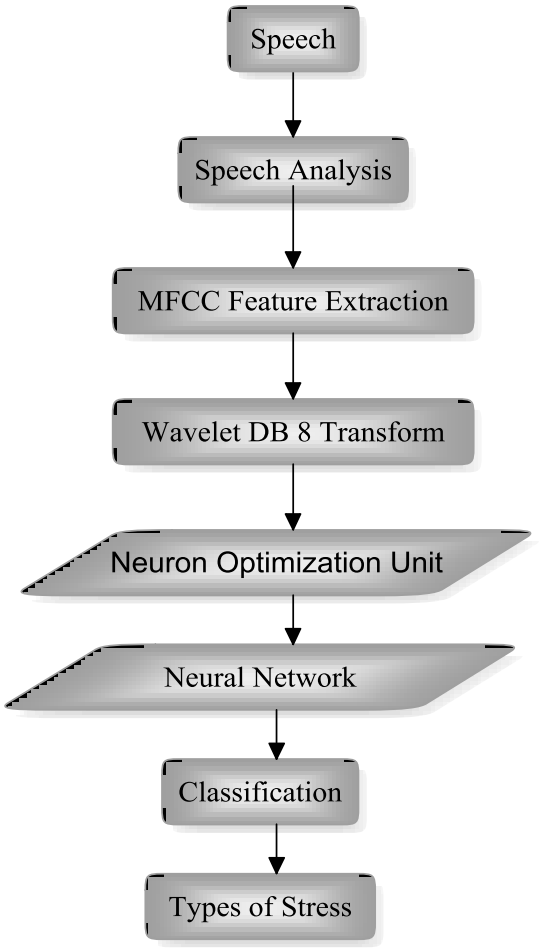


Figure 1 Flowchart of RBF Algorithm

A. Radial Basis Function-

Figure 2 shows an input vector x is used as input to all radial basis functions, each with different parameters. The output network is a linear combination of the outputs from radial basis functions. n the field of mathematical modelling, a radial basis function network is an artificial neural network that uses radial basis functions as activation functions. The output of the network is a linear combination of radial basis functions of the inputs and neuron parameters. Radial basis function networks have many uses, including function approximation, time series prediction, classification, and system control[16]. They were first formulated in a 1988 paper by Broom-head and Lowe, both researchers at the Royal Signals and Radar Establishment [17][18][19].

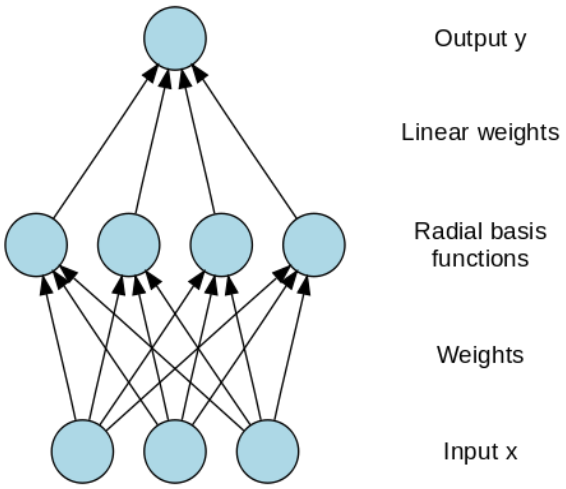


Fig. 2 Generalized Structure of RBF

B. Features (MFCC)-

Mel Frequency Cepstral Co-efficients is mostly used features for any speech recognition system. We are using MFCC for stress speech feature extraction. [20] Feature extraction undergoes raw speech transformation into useful parameters without changing speech information. It consists of Pre-emphasis, Framing, windowing, spectral estimation, Mel Filtering DCT etc. as procedures for this features extraction. In stress speech extraction we convert into useful data to classify and train the neural network.

C. Classifier training and testing-

RBF classifier is trained using neural network for stress speech identification using MFCC. These feature vectors are provided to test the stress types and classify using delay needed.

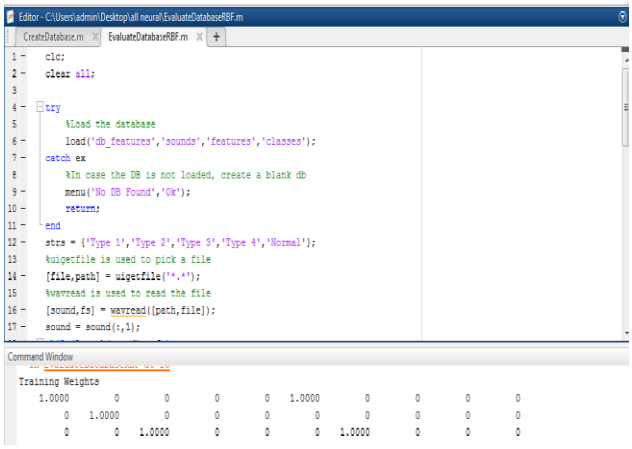
VI. RESULTS

We tested our stress detection systems under 5 different categories, namely,

- Stress Type 1
- Stress Type 2
- Stress Type 3
- Stress Type 4
- No Stress

Stress type 1 arises from problems like workload and anxiety. Stress Type 2 induces from noise and speech quality. Stress type 3 corresponds to effects causing due to medicines, illness and narcotics. Stress Type 4 refers to problem arises from vibration and acceleration. Finally No stress means persons is in normal condition.

Figure 3 is a Matlab screenshot for the code of RBF. This is the code of extracting the features which are mfcc and identifying using classifier as different stress type speech or a normal speech. This code is implementable on any Language. RBF suggests that it has classified a speech wave file is identified as Normal Speech. The same Procedure is repeated with Berlin Database and Humaine Database. The Speech file extension is .wav. The Speech undertaken is from Real Datasets which we have created.



```
clc;
clear all;

try
    %Load the database
    load('db_features','sounds','features','classes');
catch ex
    %In case the DB is not loaded, create a blank db
    menu('No DB Found','Ok');
    return;
end
strs = {'Type 1','Type 2','Type 3','Type 4','Normal'};
%uigetfile is used to pick a file
[file,path] = uigetfile('*.*');
%wavread is used to read the file
[sound,fs] = wavread([path,file]);
sound = sound(:,1);
```

```
Training Weights
    1.0000         0         0         0         0    1.0000         0         0         0         0
         0    1.0000         0         0         0         0         0         0         0         0
         0         0    1.0000         0         0         0    1.0000         0         0         0
```

Figure no. 3 MATLAB Code Screenshot for RBF classifier

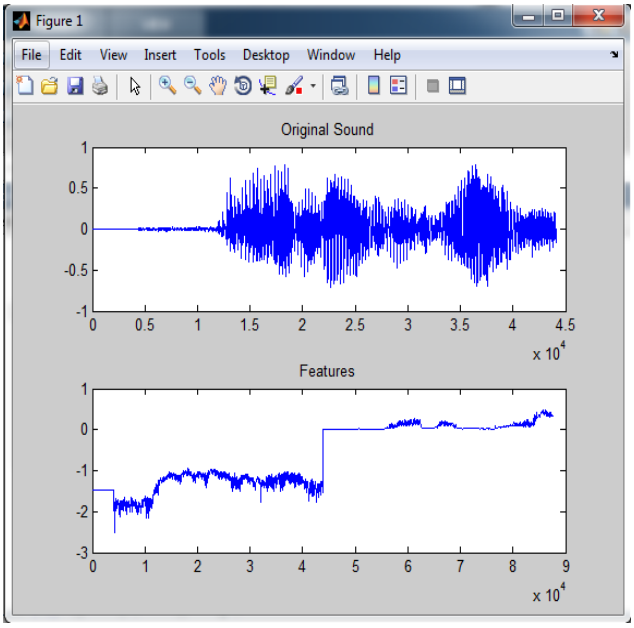


Figure no. 4 The Original Speech with its MFCC Features.

Above figure notifies the Matlab window original speech and its MFCC feature where on x-axis is Frequency and Y-axis is its Amplitude.

Figure no. 5 describes the Delay needed for RBF classifier for a particular Wave file and classified as Normal Speech. Delay arises due to echo and reverberations in speech.

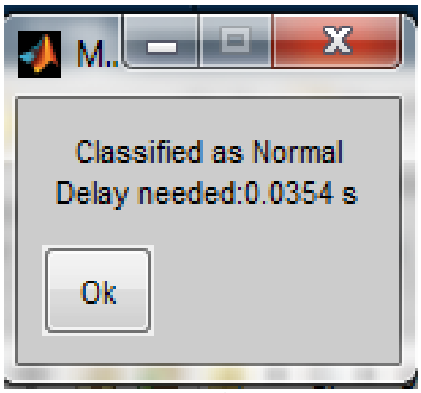


Figure no. 5. Delay and Classifier screenshot in MATLAB

VII. CONCLUSION

From the above results we are having the screenshots for Realtime Database for this research work. The similar procedures are operated onto the two standard databases which are BERLIN and HUMAINE Datasets. RBF is chosen to recognize the speech into stress types. RBF is approach based on Neural Network using MFCC.

VIII. FUTURE SCOPE

In the future we are going to find the percentage of efficiency of RBF Neural Network and compare it again different neural networks to get best classifier used for Stress Speech Identification.

REFERENCES

[1] Schuller, Bjorn, et al., *"Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first Challenge"*, Speech Communication 53.9, pp. 1062-1087, 2011.

[2] Anagnostopoulos, Christos-Nikolaos, Theodoros Iliou, and Ioannis Giannoukos, *"Features and Classifiers for Emotion Recognition from Speech: A survey from 2000 to 2011"*, Artificial Intelligence Review 43.2, pp.155-177, 2015.

[3] Dipti D. Joshi, M. B. Zalte, *"Speech Emotion Recognition: A Review"*, Journal of Electronics and Communication Engineering (IOSR-JECE) 4.4, pp.34-37, 2013.

[4] Ververidis, Dimitrios, and Constantine Kotropoulos, *"Emotional Speech Recognition: Resources, Features, and Methods"*, Speech Communication 48.9, pp.1162-1181, 2006.

[5] El Ayadi, Moataz, Mohamed S. Kamel, and Fakhri Karray, *"Survey on Speech Emotion Recognition"*, Features, classification schemes, and databases, Pattern Recognition 44.3 pp. 572-587,2011.

[6] Scherer, Klaus R., *"Vocal Communication of Emotion: A review of research paradigms"*, Speech communication 40.1, pp.227-256, 2003.

[7] Vogt, Thurid, Elisabeth Andre, and Johannes Wagner, *"Automatic recognition of emotions from speech: a review of the literature and recommendations for practical realization, Affect and emotion in human-computer interaction*", Springer Berlin Heidelberg, pp. 75-91, 2008.

[8] Burkhardt, Felix, et al., *"A Database of German Emotional Speech"*, INTER-SPEECH, Lisbon, Portugal, vol. 5, pp.1-4, 2005.

[9] Kwon, Oh-Wook, et al, "Emotion Recognition by Speech Signals, INTER-SPEECH, pp.1-4, 2003.

[10] Campbell, N. *"Recording and Storing of Speech Data"*. In: Proceedings LREC, pp. 12-25, 2002.

[11] Cowie, R., Douglas-Cowie, E., Savvidou, S., McMahon, E., Sawey, M., Schroder, M. *Feeltrace, "An Instrument for Recording Perceived Emotion in Real Time"*, In: Proceedings of the ISCA Workshop on Speech and Emotion, pp.19-24, 2000.

[12] Devillers, L., Cowie, R., Martin, J.-C., Douglas-Cowie, E., Abrilian, S., McRorie, M.: *"Real life emotions in French and English TV video clips: an integrated annotation protocol combining continuous and discrete approaches"*, 5th International Conference on Language Resources and Evaluation LREC, Genoa, Italy.2006.

[13] Douglas-Cowie, E., Campbell, N., Cowie, R.P. "*Emotional speech: Towards a new generation of databases"*, Speech Communication 40(1–2), pp.33-60, 2003.

[14] Douglas-Cowie, E., et al.: *"The description of naturally occurring emotional speech"*. In: Proceedings of 15th International Congress of Phonetic Sciences, Barcelona, 2003.

[15] http://audacity.sourceforge.net/download.

[16] Broomhead, D. S. Lowe, *David "*Radial basis functions, multi-variable functional interpolation and adaptive networks", Technical Report- RSRE 4148, 1988.

[17] Broomhead, D. S. Lowe, David *"Multivariable functional interpolation and adaptive networks"*, Complex Systems 2, pp. 321–355, 1988.

[18] Schwenker, Friedhelm; Kestler, Hans A.; Palm, Gunther, *"Three learning phases for radial-basis-function networks"*, Neural Networks 14, pp. 439–458, 2001.

[19] Park, J., I. W. Sandberg, *"Universal Approximation Using Radial-Basis-Function Networks"*, Neural Computation. 3 (2): pp. 246–257, Retrieved 26 March 2013.

[20] BOGERT, B. P.; HEALY, M. J. R.; TURKEY, J. W.: "*The Quefrency Alanysis of Time Series for Echoes: Cepstrum, Pseudo Autocovariance, Cross-Cepstrum and Saphe Cracking"*, Proceedings of the Symposium on Time Series Analysis, (M. Rosenblatt, Ed) Chapter 15, New York: Wiley, pp.209-243, 1963.

Authors Profile

Mrs. N.P. Dhole is born in Amravati, Maharashtra State, India. She is received the B.E. degree in Electronics & Telecommunication Engineering and the M.E. degree in Digital Electronics from Sant Gadge Baba Amravati University, Amravati. She is Assistant Professor in Electronics & Telecommunication Engineering department at PRMIT&R Badnera. She has 04 years of teaching experience also pursuing PhD in "Identification of Stress from Speech Signal using SOFT Computing Techniques".

Dr. Sujata N. Kale is born in Amravati, Maharashtra State, India. Currently she is Associate Professor in PG Department of Applied Electronics, Sant Gadge Baba Amravati University. She has 20 years of teaching experience. She received her doctoral degree in Machine Intelligence from Sant Gadge Baba Amravati University, Amravati. Her current research interests focus on the areas of image processing, signal processing, neural networks.

