JCSE International Journal of Computer Sciences and Engineering Open Access

Research Paper Volume-6, Issue-4 E-ISSN: 2347-2693

Study of Recurrent Neural Network Classification of Stress Types in Speech Identification

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

[1]Dept. of Electronics and Telecommunication Engineering, PRMIT&R Badnera, Amravati, India,
[2]Dept. 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. More than a decade has passed since research on stress types in speech identification has become a new field of research in line with its 'big brothers' speech and speaker recognition. This article attempts to provide a short overview on where we are today, how we got there and what this can reveal us on where to go next and how we could arrive there. In this work we propose a Recurrent Neural Network classifier for speech stress classification algorithm, with sophisticated feature extraction techniques as Mel Frequency Cepstral Coefficients (MFCC). The 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.

Index Terms: RNN, 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. 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 Recurrent Neural Network 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	Recurrent Neural Network
Output	.mat files
Results	Images of MATLAB Software

Table no. 1 shows the elements of databases. Output of Recurrent Neural Network 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 RECURRENT NEURAL NETWORK ALGORITHM

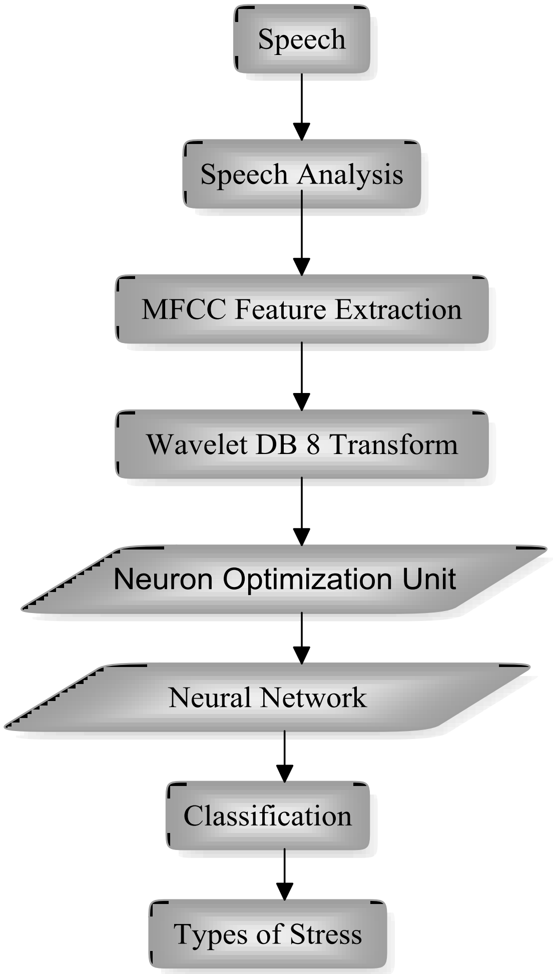


Figure no. 1: Flowchart of Recurrent Neural Network Algorithm

A. Recurrent Neural Network-

Recurrent FNNs have an important drawback: only a fixed number of previous words can be taken into account to predict the next word [16]. This limitation is inherent to the structure of FNNs, since they lack any form of 'memory': only the words that are presented via the fixed number of input neurons can be used to predict the next word and all words that were presented during earlier iterations are 'forgotten', although these words can be essential to determine the context and thus to determine a suitable next word. Ingenious solutions were proposed to introduce some memory in the FNN architecture to overcome the limited context length. Noteworthy variants of the original FNN are the Jordan (1986) and Elman (1990) networks where extra neurons are incorporated that are connected to the hidden layer like the other input neurons. These extra neurons are called context neurons and hold the contents of one of the layers as it existed when the previous pattern was trained. This model allows a sort of short term memory. Training can still be done in the same way as for FNNs. Fig. 2 shows the network output fed back into the context unit, which in turn sends it to the hidden layer in the next iteration [17]. In an Elman network a context neuron is fed by a hidden layer.

The context length was extended to indefinite; one could even say infinite, size by using a recurrent version of neural networks, conveniently called recurrent neural networks (RNN), which can handle arbitrary context lengths. Initial enthusiasm about RNN as statistical language modellers, mainly driven by their abilities to learn vector representations for words(assign FNN) and to handle arbitrarily long contexts, in addition to the fact that they are universal approximates (Schafer and Zimmerman, 2007), was quickly tempered by their extremely slow learning and by the observation that W. De Mulder *et al*. Computer Speech and Language 30 (2015) 61–98 65 the theoretical incorporation of arbitrarily long context lengths does not manifest itself in practice. Consequently, many variations on the original RNN have been developed to cope with these limitations and, at the same time, to specialize their structure towards SLM (Kombrink *et al*., 2011). This survey presents the main RNN architectures that resulted from the attempts to turn the original, theoretically well founded RNN with its limited practical usefulness into more practically oriented structures, where researchers were willing to exchange the theoretical soundness with heuristic reasoning (at least to some degree), but without giving up the aforementioned strengths, each R.N.N. architecture is compared to the original structure, and empirical evaluations, as done by researchers in the field, are presented [19].

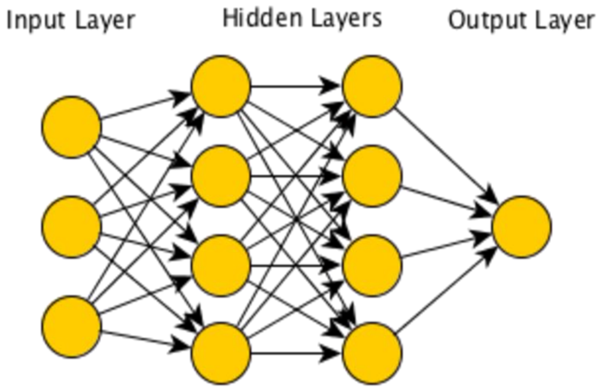


Figure 2 A Generalized Recurrent Neural Networks

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 [19]. 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-

The performance of most speech-recognition systems, whose designs are predicated on assumptions about the ambient conditions, such as low noise background noise background, degrades rapidly in the presence of noise and distortion. Recurrent neural network 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

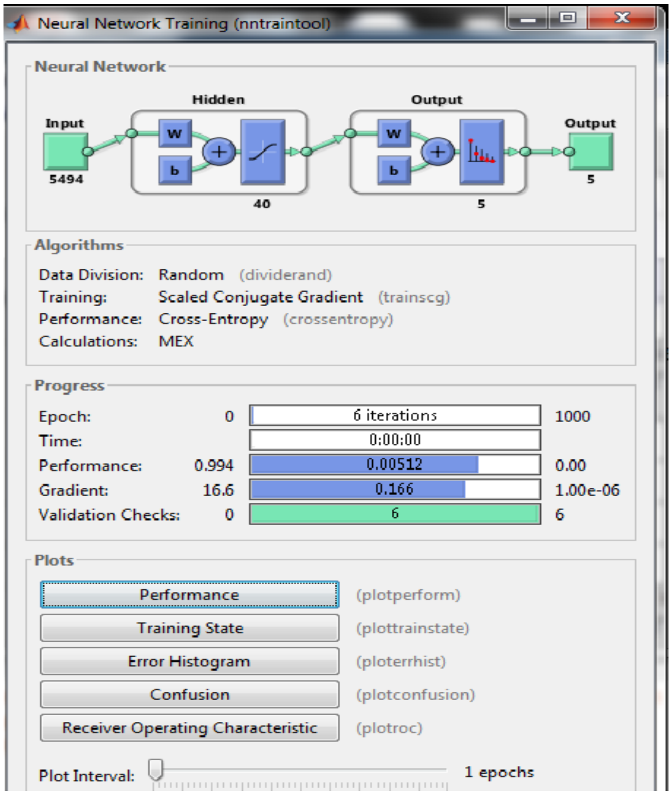
We tested our stress detection systems under 5 different categories, namely,

- 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 training recurrent neural Network. The algorithm is trained through Scaled Conjugate Gradient. It can train any network as long as its weight, net input, and transfer functions have derivative functions. Back propagation is used to calculate derivatives of performance with respect to the weight and bias variables .The scaled conjugate gradient algorithm is based on conjugate directions, as in traincgp, traincgf, and traincgb, but this algorithm does not perform a line search at each iteration.

Recurrent neural network suggests that it has classified a speech wave file as Stress Type 4. The same Procedure is repeated with Berlin Database and Humaine Database. The Speech file extension is .wav file. The Speech undertaken is from Real Datasets which we have created. From Fig 3 we can see the performance of Recurrent Neural Network taking six iteration to complete the training consisting of 40 hidden neurons.



Neural Network Training (nntraintool)

Neural Network

Algorithms

Data Division: Random (dividerand)
Training: Scaled Conjugate Gradient (trainscg)
Performance: Cross-Entropy (crossentropy)
Calculations: MEX

Progress

Epoch:	0	6 iterations	1000
Time:		0:00:00	
Performance:	0.994	0.00512	0.00
Gradient:	16.6	0.166	1.00e-06
Validation Checks:	0	6	6

Plots

Performance (plotperform)
Training State (plottrainstate)
Error Histogram (ploterrhist)
Confusion (plotconfusion)
Receiver Operating Characteristic (plotroc)

Plot Interval: 1 epochs

Figure no. 3 MATLAB training for Recurrent Neural Network Classifier

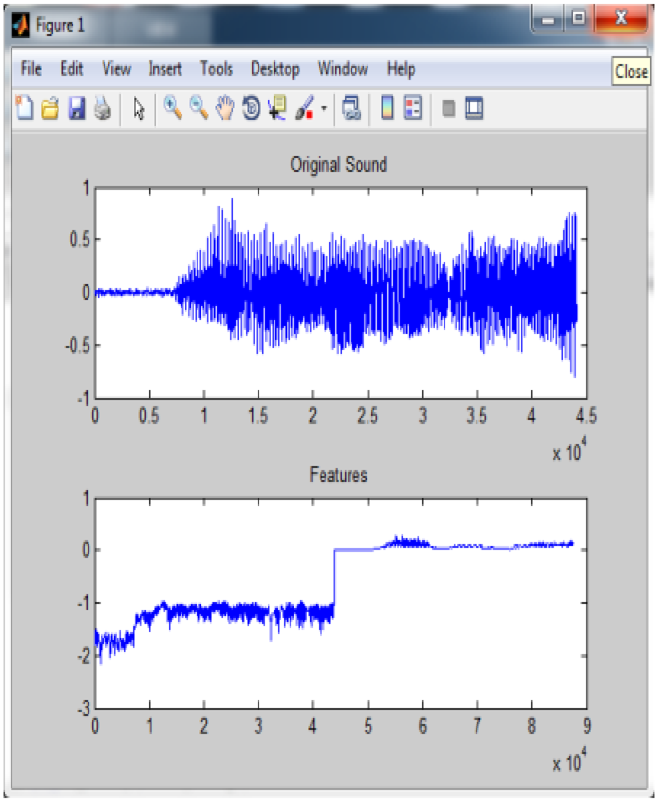
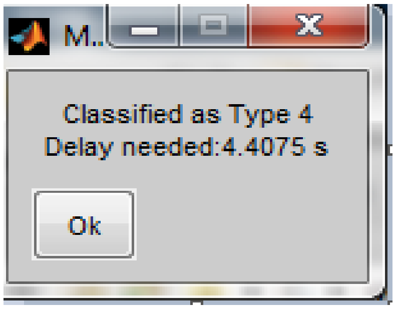


Figure no. 4 The Original Speech with its MFCC Features.

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 Recurrent Neural Network classifier for a particular Wave file and classified as Stress Type 4. Delay arises due to echo and reverberations in speech.



M..

Classified as Type 4
Delay needed:4.4075 s

Ok

Figure no. 5. Delay and Classifier screenshot in MATLAB

VII. Conclusion

We have contributed the stress classification using RNN. It almost gives best classification of speech. From the above results we are having the screenshots for Real-time Database for this research work. Recurrent Network works as good Identifier in speech Identification. The similar procedures are operated onto the two standard databases which are BERLIN and HUMAINE Datasets. Recurrent Neural Network is chosen to recognize the speech into stress types. Recurrent Neural Network is approach based on Neural Network using MFCC.

VIII. Future Scope

In the future we are going to find the percentage of efficiency for Recurrent 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] A. J. Robinson, *"An Application of Recurrent Nets to Phone Probability Estimation,"* IEEE Transactions on Neural Networks, vol. 5, no. 2, pp. 298-305, 1994.

[17] Oriol Vinyals, Suman Ravuri, and Daniel Povey, *"Revisiting Recurrent Neural Networks for Robust ASR,"* in ICASSP, 2012.

[18] A. Maas, Q. Le, T. O Neil, O. Vinyals, P. Nguyen, and A. Ng, "Recurrent neural networks for noise reduction in robust asr," in INTERSPEECH, 2012.

[19] 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.