International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org ||Volume 7 Issue 5 Ver. IV|| May 2018 || PP 85-89

Stress Speech Identification Using Various Neural Networks

Mrs. N.P. Dhole[1], Dr. S.N. Kale[2]
[1]Assistant Professor, Department of Electronics and Telecommunication Engineering, PRMIT&R Badnera, Amravati (Maharashtra), India, [2]Associate Professor, Department of Applied Electronics, Sant Gadge Baba Amravati University, Amravati, (Maharashtra), India
Corresponding Author: Mrs. N.P. Dhole1

Abstract: *There has been a Speech Identification technology already available in the market based on different assumptions. The work presented in this paper investigates the feasibility of alternative approaches for solving the problem more efficiently. A Speech Identification particularly Stress component in the research comprised of three distinct blocks, a Feature Extraction Neural Networks and Identifier. The Feature Extractor block uses MFCC which is most prominent and efficient for Stress Speech Identifier. Designs of the Neural Network blocks based on four different approaches are compared. The performance of Support Vector Machine, Multi-Layer Perceptrons, Radial Basis Functions and Recurrent Neural Networks based Identifier is tested on Speech which consists of Stress components Identification problem. Experimental results indicate that applying these neural networks reduces the training complexity and the operation of the Identifier. During the implementation of this all algorithms some results have been obtained in terms of the accuracy and the quality of the Identifier. The comparison between the different approaches to the design of the Identifier conducted here gives a better understanding of the problem and its possible solutions for well being of the society.*

Keywords: *RBF, SVM, MLP, RNN, MFCC, Stress Classification, Feature Extraction.*

Date of Submission: 07-05-2018 Date of acceptance: 22-05-2018

I. Introduction

Through spoken Languages we are able to communicate from person to person. From country to country and even states to states spoken languages are different. So here it becomes compulsory to design a system independent of language barrier and to ensure efficient stress identification. If an efficient speech recognition machine is enhanced by natural language systems and speech producing techniques, it would be possible to produce computational applications that do not require a keyboard and a screen [1]. This would allow incredible miniaturization of known systems facilitating the creation of small intelligent devices that can interact with a user through the use of speech [2].

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 [3,4] has been developed to speed up the accuracy using various classifiers on several databases [5]. We have also revised the literature review of numerous researchers for the same work [6,7,8,9,10]. We have used for this work the Berlin database and Humaine database 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.

It is important to understand that it is not the purpose of this work to develop a full-scale Stress Speech Identifier but only to test new techniques and explore their usefulness in providing more efficient solutions. Doing this can help the person to analyze the stress and obtain remedies for the same. The whole Algorithm is developed in MATLAB Software.

1.1 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 [11][12]. 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.

1.2 Humaine Database

The database proper is a selected subset of the data with systematic labelling, mounted on the ANVIL platform [13,14,15,16,17]. 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.

II. 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) [18]. 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.

III. Biological Neural Networks

The central idea behind most of the developments in neural sciences is that a great deal of all intelligent behaviour is a consequence of the brain activity. Even more, the different brain functions emerge as a result from the dynamic operation of millions and millions of interconnected cells called neurons. Roughly speaking, a typical neuron receives its inputs through thousands of dendrites, processes them in its soma, transmit the obtained output through its axon, and uses synaptic connections to carry the signal of the axon to thousands of other dendrites that belong to other neurons. It is the manner in which this massively interconnected network is structured and the mechanisms by which neurons produce their individual outputs that determine most of the intelligent behaviour. These two aspects are determined in living beings by evolutionary mechanisms and learning processes.

4.1 Artificial Neural Networks

During the last 30 years, a new computational paradigm based on the ideas from neural sciences has risen [17]. Its name, neural networks, comes from the fact that these techniques are based upon analogies drawn from the inner workings of the human nervous system. The main idea behind the neural network paradigm is to simulate the behaviour of the brain using interconnected abstractions of the real neurons in fig.1, to obtain systems that exhibit complex behaviours, and hopefully intelligent ones. In an artificial neuron, numerical values are used as inputs to the "dendrites." Each input is multiplied by a value called weight, which simulates the response of a real dendrite. All the results from the "dendrites" are added and threshold in the "soma." Finally, the threshold result is sent to the "dendrites" of other neurons through an "axon." Although the model suggested by Figure 1 is a simplistic reduction of a real neuron, its usefulness is not based on the individual capacities that each individual neuron exhibits but on the emergent behavior that arises when these simple neurons are interconnected to form a neural network. As in biological neurons, the interconnection architecture of the neurons and the manner in which neurons process their inputs determines the behavior exhibited by the neural network. In contrast with biological neural networks, the architecture of a neural network is set by a design process, and the parameters that define the way its neurons process their respective inputs are obtained through a learning process. The learning process involves a search in a parameter space for determining the best parameter values according to some optimality criteria.

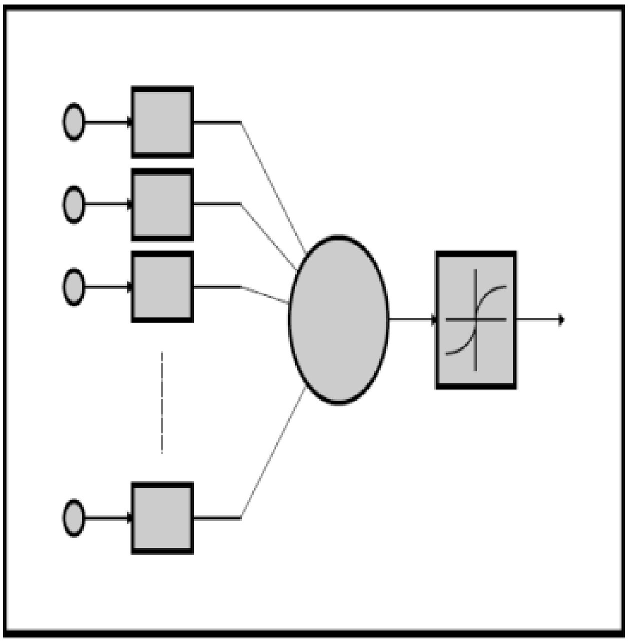

Fig.1 simplistic reduction of a real neuron

IV. Flowchart Of Neural Network Algorithm

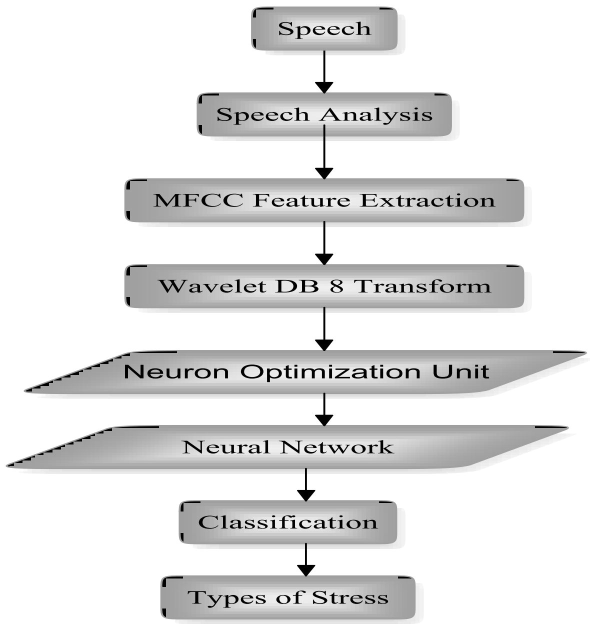

Fig. 2 Flowchart of NN Algorithm

5.1 Speech Analysis-

It consists of speech as input to the Identifier system. Filtering of the speech is done by using LMS Algorithm. The LMS algorithm is best suited for suppressing noise from the speech signal by simulations in MATLAB software.

5.2 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. 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.

5.3 Classifier training and testing-

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. We may term the classifiers as Identifiers. In this Work we are comparing all Classifiers which are SVM, MLP, RNN and RBF [20, 21, 22]. Here we have used *Levenberg-Marquardt (LM) Algorithm.*

V. 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.

In the last work we have implemented all four classifiers and seen its original signal, MFCC features and delay needed. From them we got the idea that a person is under Stress Type1, Stress Type2, Stress Type 3, and Stress Type 4 or in Normal condition. 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.

Table I. Stress Detection Results

No. Of Entries in dB	No. Of Entries Tested	Accuracy SVM (%)	Accuracy RBF (%)	Accuracy RNN (%)	Accuracy MLP (%)
10	12	50.75	58	65	72.5
20	25	54.6	62.4	69.33333	78
30	39	56.35	64.4	71.66667	80.5
40	55	57.4	65.6	73.33333	82
50	65	58.625	67	74.83333	83.75
75	85	60.305	68.92	77.43333	86.15
90	100	61.145	69.88	79.03333	87.35
100	120	61.565	70.36	80.36666	87.95

VI. Conclusion

From the above results of Accuracy all classifiers are trained and tested for Realtime Database for this research work. The inputs are in decibels to find accuracy of all classifiers. The number of entries is increasing and accordingly the four classifiers are tested. SVM is falling in the lowest percentage of accuracy as compared to RBF. Moreover RBF is substantially lower as in comparison to RNN. But when RNN is compared to MLP it is approximately less than MLP. So MLP can be termed as best Identifier of stress in speech. But this study is limited to our Real Time Database. The similar procedures are operated onto the two standard databases which are BERLIN and HUMAINE Datasets. When we will apply the same classifiers to Berlin and Humaine Databases RNN and MLP will work more or less the same. It all depends on the datasets on which we apply this algorithm .Many experiments were conducted for training/testing each model separately with different: number of hidden layer neurons, learning rate, and training algorithm to judge the classifier or Identifier with best results of Stress Speech signal or Normal Signal. So from the above results we founded that MLP is best Identifier for the real datasets for Stress in Speech Signal.

References

[1]. L. Rabiner and G. Juang, *Fundamentals of Speech Recognition,* Prentice-Hall 1993.

[2]. N. Negroponte, *Being Digital* Vintage books 1995.

[3]. Farhang – Borounjeny, B. Adaptive Filters, Theory and applications, John Wiley & Sons, Chichester 2005, ISBN 0–471–98337–3, pp. 139-168.

[4]. Schuller, Bjorn, et al., Recognizing 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.

[5]. 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.

[6]. 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.

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

[8]. 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.
[9]. Scherer, Klaus R., "Vocal Communication of Emotion: A review of research paradigms", Speech communication 40.1, pp.227-256, 2003.
[10]. 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.
[11]. Burkhardt, Felix, et al., "A Database of German Emotional Speech", *INTER-SPEECH, Lisbon, Portugal, vol. 5,* pp.1-4, 2005.
[12]. Kwon, Oh-Wook, et al, "Emotion Recognition by Speech Signals, *INTER-SPEECH,* pp.1-4, 2003.
[13]. Campbell, N. *"Recording and Storing of Speech Data".* In: Proceedings LREC, pp. 12-25, 2002.
[14]. 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.
[15]. 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.
[16]. 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.
[17]. Douglas-Cowie, E., et al.: "The description of naturally occurring emotional speech". *In: Proceedings of 15th International Congress of Phonetic Sciences,* Barcelona, 2003.
[18]. http://audacity.sourceforge.net/download.
[19]. S. Haykin, "Neural Networks: A Comprehensive Foundation*," Macmillan College Publishing Company*, 1994.
[20]. N. P. Dhole, Dr.S. N. Kale, "An Application Using Radial Basis Function Classification in Stress Speech Identification,"*International Journal of Computer Sciences and Engineering (ISSN: 2347-2693), Vol.6, Issue.4,* 2018
[21]. N.P. Dhole, Dr.S.N. Kale, "Study of Recurrent Neural Network Classification of Stress Types in Speech Identification," *International Journal of Computer Sciences and Engineering (ISSN: 2347-2693), Vol.6, Issue.4,* 2018
[22]. N. P. Dhole, Dr.S.N. Kale, "Multilayer Perceptron Classification in Stress Speech Identification," *International Journal of Computer Sciences and Engineering (ISSN: 2347-2693), Vol.6, Issue.4*, 2018

Mrs. N.P. Dhole" Stress Speech Identification Using Various Neural Networks "International Journal of Engineering Science Invention (IJESI), vol. 07, no. 05, 2018, pp 85-89