IOSR Journal of VLSI and Signal Processing (IOSR-JVSP)
Volume 8, Issue 2, Ver. I (Mar.-Apr. 2018), PP 44-47
e-ISSN: 2319 – 4200, p-ISSN No. : 2319 – 4197
www.iosrjournals.org

Support Vector Machine Classification of Stress Types in Speech

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

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 SVM classifier for speech stress classification algorithm, with sophisticated feature extraction techniques as Mel Frequency Cepstral Coefficients (MFCC). The SVM 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: *Support Vector Machines, MFCC, Stress Classification, Feature Selection.*

Date of Submission: 17-04-2018 Date of acceptance: 04-05-2018

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.

This paper proposes SVM 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	Support Vector Machine
Output	.mat files
Results	Images of MATLAB Software

Table no. 1 shows the elements of databases.Output of SVM Algorithm are 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 SVM Algorithm

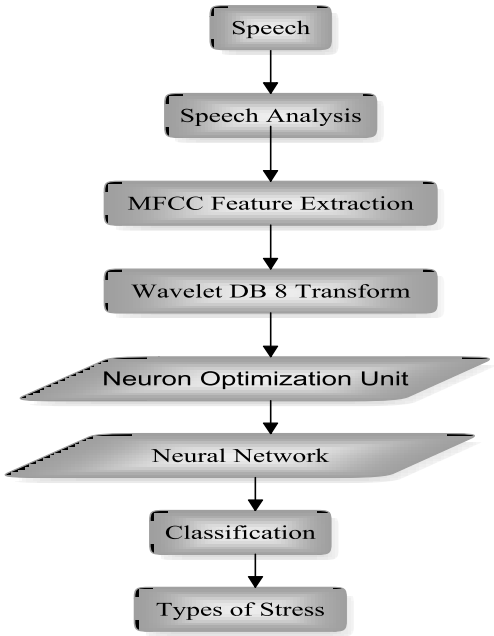

Figure no. 1: Flowchart of SVM Algorithm

SVM: SVM is a machine learning technique in pattern recognition specially speech identification. SVM is Kernel based margin maximization [16]. The Maximal-Margin Classifier is a hypothetical classifier that best explains how SVM works in practice. The numeric input variables (x) in your data (the columns) form an n-dimensional space. For example, if you had two input variables, this would form a two-dimensional space. A hyper plane is a line that splits the input variable space [16, 17]. In SVM, a hyper plane is selected to best separate the points in the input variable space by their class, either class 0 or class

Features (MFCC): Mel frequency cepstral co-efficients are mostly used features for any speech recognition system. We are using MFCC for stress speech feature extraction. [17] 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.

Classifier training and testing*:* SVM 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 2 is a Matlab screenshot for the code of SVM. SVM suggests that it has classified a speech wavefile as in the Class 2 and Class 4 with number 4 counts in Class 1 and Class 2 as number of Counts 2 which means that Class 1 is Stress Type 1 with recognition with 4 counts. 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.

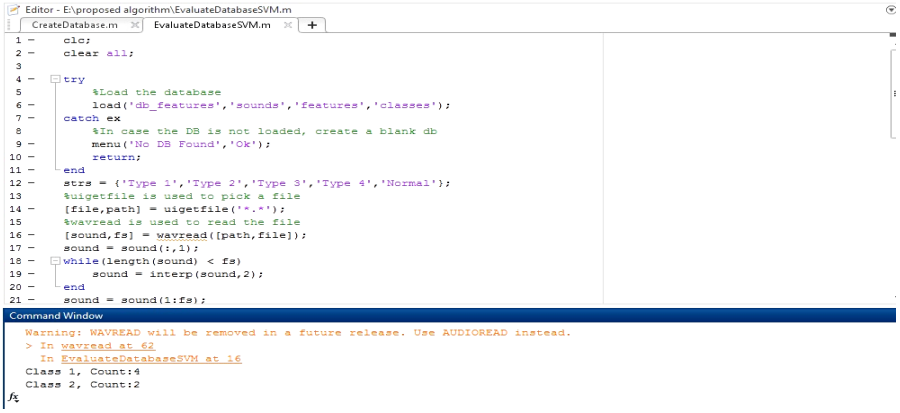

Figure no. 2 MATLAB Code Screenshot for SVM classifier

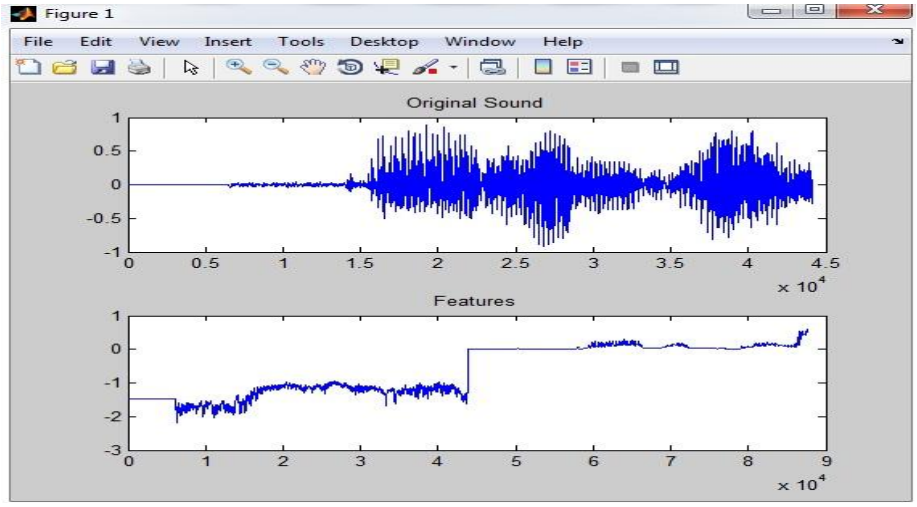

Figure no. 3 The Original Speech with its MFCC Features.

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. 4 describes the Delay needed for SVM classifier for a particular Wave file and classified as Stresss Type 1. Delay arises due to echo and reverberations in speech.

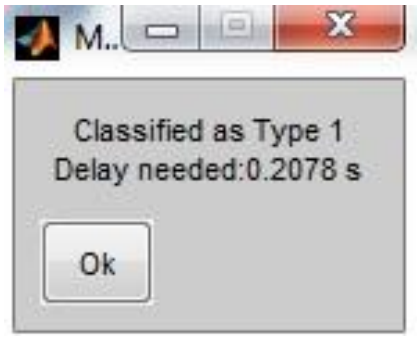

Figure no. 4. Delay and Classifier screenshot in MATLAB

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. SVM is chosen to recognize the speech into stress types. SVM is approach based on Neural Network using MFCC.

Future Scope

In the future we afe going to find the percentage of efficiency 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 (2011): 1062-1087.
[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 (2015): 155-177.
[3]. Joshi, Dipti D., and M. B. Zalte, Speech Emotion Recognition: A Review, Journal of Electronics and Communication Engineering (IOSR-JECE) 4.4 (2013).
[4]. Ververidis, Dimitrios, and Constantine Kotropoulos, Emotional speech recognition: Resources, features, and methods, Speech communication 48.9 (2006): 1162-1181.
[5]. El Ayadi, Moataz, Mohamed S. Kamel, and Fakhri Karray, Survey on speech emotion recognition: Features, classification schemes, and databases, Pattern Recognition 44.3 (2011): 572-587.
[6]. Scherer, Klaus R., Vocal communication of emotion: A review of research paradigms, Speech communication 40.1 (2003): 227-256.
[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, 2008. 75-91.
[8]. Burkhardt, Felix, et al., A database of German emotional speech, Interspeech. Vol. 5. 2005.
[9]. Kwon, Oh-Wook, et al.,Emotion recognition by speech signals, INTERSPEECH. 2003.
[10]. Campbell, N.: Recording and storing of speech data. In: Proceedings LREC (2002) 7. Cowie, R., Douglas-Cowie, E., Apolloni, B., Taylor, J., Romano, A., Fellenz, W.: What a neural net needs to know about emotion words. In: Mastorakis, N. (ed.) Computational Intelligence and Applications. World Scientific Engineering Society,(1999):109–114.
[11]. Cowie, R., Douglas-Cowie, E., Savvidou, S., McMahon, E., Sawey, M., Schröder, M.: Feeltrace: an instrument for recording perceived emotion in real time. In: Proceedings of the ISCA Workshop on Speech and Emotion, (2000):19–24.
[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 2006), 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), (2003):33–60.
[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]. HANSEN, John H. L.; HASAN, Tauq: Speaker Recognition by Machines and Humans, IEEE Signal Processing Magazine,(2015): 74-99.
[17]. 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, (1963): 209-243.

Mitali Sharma. “Carbon Nanotube FET based high performance Universal logic using Cascade Voltage Switch Logic.” IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) , vol. 8, no. 2, 2018, pp. 44-47