ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 10, April 2013

Detection of Speech under Stress: A Review

Neha P. Dhole, Dr. Ajay A.Gurjar

ABSTRACT- This paper deals with speech signal as significant indicator of psychological stress when the speaker is involved in a stressful activity. The investigation of speaker's stress is based on specific changes in short-time spectrum of vowel phonemes. For each selected signal segment, the spectrum computed by means of different methods by different researchers. By comparing the results of different methods by different researchers we detect the speaker's stress. In case of speech under stress, the obtained spectra differ towards the higher frequencies due to enhanced pitch modulation. Different experiments b considered by different researchers, a database of speech under stress consisting of data collected during different researchers and speech was analyzed as normal or stressed. Stress is a constant factor in most peoples' lives, often at times it is undesirable but other times it is beneficiary. Detecting stress and the level can be crucial in applications that are stress sensitive such as voice activated military equipment, psychological testing and deception detection.

Index Terms- **DWT, Speech, Stress, Spectral Analysis.**

I. INTRODUCTION

Emotions have long been recognized to be an important aspect of human beings. More recently, psychologists have begun to explore the role of emotions as a positive component in human cognition and intelligence. Spoken language comes from our inside. Factors such as mood, emotion, physical characteristics and further pragmatic information are contained in speech signals. Many of these characteristics are also audible. An emotional speech with high content differs in some parameters from a neutral speech. In recent years, the interest for automatically detection and interpretation of emotions in speech has grown and vocal emotions have also tended to be studied in isolation. About 25% of information contained in a clean speech signal refers to the speaker. These linguistically irrelevant speaker characteristics make speech recognition less effective but can be used for speaker recognition (ca. 15% of information) and analysis of the speaker's emotional and health state (ca. 10% of information). With increasing demand for speech technology systems, there is an increasing need for processing of emotion and other pragmatic effects (simulation in synthetic speech, elimination in robust speech recognition). In some cases, it is very important to detect the emotional state of a person (e.g., stress, fatigue or use of alcohol) from his/her voice. Because stress imparts negative public health consequences, it is advantageous to consider automatic and ubiquitous methods for stress detection. Ubiquitous stress detection can help individuals become aware of and manage their stress levels. Meanwhile, distributed stress monitoring may allow health professionals the ability to examine the extent and severity of stress across populations. Many physiological symptoms of stress may be measured with sensors, e.g., by chemical analysis, skin conductance readings, electrocardiograms, etc. However, such methods are inherently intrusive upon daily life, as they require direct interaction between users and sensors. We therefore seek less intrusive methods to monitor stress.

A. Speech under Stress-

Stress is a psycho-physiological state characterized by subjective strain, dysfunctional physiological activity and deterioration of performance. The accepted term for speech signal carrying information on the speaker's physiological stress is "stressed speech".

B. Speech Production under Stress-

Speech production begins with abstract mental processes, the desire to communicate and the idea which is to be communicated. Suitable linguistic units have to be chosen from memory and formed into a sentence, subject to grammatical constraints. From the abstract sequence of words, a corresponding sequence of articulately targets must be generated. Then appropriate motor programs for the targets must be activated, with modifications to take account of context and paralinguistic information. This results in patterns of nerve impulses being transmitted to the muscles which control the respiratory system and vocal tract. The final stages are purely physical, the generation of acoustic energy, the shaping of its spectral and temporal characteristics, and its radiation from the mouth.

II. DATABASES

The presented work investigates speech data from three different databases, which are briefly presented in the following paragraphs.

A. SUSAS-

The SUSAS database comprises speech under simulated and actual stress in five different domains and was especially created for investigating speech under stress. The vocabulary covers 35 single-word utterances from aircraft communication. Three of these five domains are selected for analysis, namely Talking Styles (neutral, slow, fast, soft, loud, clear, angry and question), Single Tracking and Dual Tracking computer response tasks. SUSAS is employed in the majority of literature on speech under stress research and thus taken as a reference for the current work.

B.EMO-DB:-

The EMO-DB database consists of 10 different German utterances produced in seven (acted) emotional states, each produced by 5 female and 5 male speakers. The simulated emotions include angry, anxious, bored, disgusted, joyful, sad, and neutral. This database has been chosen in order to be tested against SUSAS talking styles domain, which

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 10, April 2013

contains two equivalent emotional classes, neutral and angry.

C. ATCOSIM-

The ATCOSIM speech corpus contains about 10 hours of non-prompted, clean air traffic control simulation speech recorded during real-time simulations. In contrast to the remaining two databases, the data is not categorized into emotional classes, and no metadata in terms of label files with phoneme information is provided, but since a list of appearing utterances exists, phoneme labels and boundaries can be nevertheless determined with the HTK toolkit using a standardized lexicon file. With a check against the SUSAS computer response domains in mind, emotional classes are assigned approximately.

III. LITERATURE REVIEW

Dr. MILAN SIGMUD conducted the research at university; the most suitable situation with realistic stress took place during the final state examinations held in oral form in front of a board of examiners. The test speakers in his experiment were 31. He created a database of them and analyzed the speech as normal or stressed. Spectral analysis was done with the help of chirp transform. Changes in spectrum of speech signal have shown to be a reliable indicator of the internal emotional state of a person. The speaker's stress can be detected from short segments of vowels comparing their Fourier and chirp spectra. Long-term goal is to automatically detect and quantify the actual stress influencing a person, on the basis of acoustic and prosodic information extractable from utterances. Researcher *Hong Lu* stated in [1], [2] that Stress is a subjective phenomenon, and stressors are the observable or imagined events and stimuli that cause stress. He studied stress associated with the cognitive load experienced by a participant during a job interview as an interviewee and conducting a marketing task as an employee. The most widely investigated acoustic feature for stress is pitching. Pitch reflects the fundamental frequency of vocal cord vibration during speech production. Normally, the mean, standard deviation and range of pitch increase when somebody is stressed while the pitch jitter in voice usually decreases. The degree of change varies from person to person depending on the person's experience and arousal level. He presented an adaptive method for detecting stress in diverse scenarios using the microphone on smart phones. Unlike physiological sensors, microphones do not require contact with the human body and are ubiquitous to all mobile phones. Using a non-iterative MAP adaptation scheme for Gaussian mixture models, we demonstrated that it is feasible to customize a universal stress model to different users and different scenarios using only few new data observations and at a low computational overhead. Our proof-of-concept software demonstrates that Stress Sense can run on off-the-shelf smart phone in real time. Researcher *Huayang Xie* detected Stress in Spoken New Zealand English using decision trees and support vector. After identifying the vowel segments of the speech signal, the approach extracts two different sets of features prosodic features and vowel quality features from the vowel segments stated in [3]. These features are then normalized and scaled to obtain speaker independent feature values that can be used to classify each vowel segment as stressed or unstressed. He used Decision Trees (C4.5) and Support Vector Machines (LIBSVM) to learn stress-detecting classifiers with various combinations of the features. The approach was evaluated on 60 adult female utterances with 703 vowels and a maximum accuracy of 84.72% was achieved. The results shows that a combination of features derived from duration and amplitude achieved the best performance but the vowel quality features also achieved quite reasonable results. To determine the actual vowel quality of the vowels, he applied a very constrained form of full recognition to each of the vowel segments identified by forced alignment, and use the probability scores of the individual HMM phoneme models to compute several features that indicate whether the vowel is reduced or not. *Firoz Shah, Raji Sukumar A, Babu,* works deals with automatic recognition of stress from spoken words in Malayalam language [4]. Automatic Stress Recognition from speech is one of the most interesting areas in speech and emotion related studies and applications. Automatic recognition of stress from speech finds applications mostly in affective computing. Stress detection from speech means to understand the exact stress level from human speech by using a machine with the help of some machine learning algorithms. They have created and evaluated an elicited mode database consisting of a total number of four hundred isolated spoken words. Discrete Wavelet Transform (DWT) is used for the feature extraction and Artificial Neural Network (ANN) is used for the training and testing phase of the machine learning. They have obtained an overall recognition accuracy of 85% from this experiment. *Min Lai Yining Chen Min Chu Yong Zhao Fangyu Hu* Their work consist a hierarchical framework, which consists of three layers of classifiers, for automatic stress detection in English speech utterances [4], [5]. The top two layers are a linguistic classifier, which assigns stressed labels to all content words and unstressed labels to all functions words, and an acoustic classifier, which assigns stressed and unstressed labels with HMM based models and using only acoustic features such as MFCC, energy and fundamental frequency. When there is no manual stressed label available, only the top two layers are activated. The best performance we achieved is 92.9%. The third layer in the framework is an Ada Boost classifier that can improve the accuracy by using more features and manual labels. The best result they obtained is 94.1%, which is approaching to the self-agreement ratio (97.4%) of the same annotator, or the upper bound of the performance.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 10, April 2013

IV. PROPOSED WORK

In proposed work, samples of the persons those who are in the age of 22 to 24 year are considered and after that collecting all the samples of them at the time of their viva examination before and after the examination for analysis of stress in speech. All the samples collected used as a database for the spectrum analysis of the speech under stress. In order to check the whether the speech have stress or not. For this checking we are using the neural network, anfis & the Matlab Tool box. Discrete Wavelet Transforms (DWT) is the process of transformation of a signal to the high frequency and low frequency components by using digital filtering techniques. In DWT we are taking into account only the low frequency components of the signal under consideration because low frequency components characterize a signal more than its high frequency components and finally features of speech are considered.

V. CONCLUSION

The presented approach by the authors mentioned above proposes a stress detection system able to cope with this lack of security. DWT is the process of transformation of a signal to the high frequency and low frequency components by using digital filtering techniques. In DWT we are taking into account only the low frequency components of the signal under consideration because low frequency components characterize a signal more than its high frequency components, recognition stress from speech. From the obtained results we could understand that DWT is a good feature extraction method for detecting stress from speech. Also artificial neural network is used for the training and testing of the network. The efficiency of the algorithm can be evaluated by using different machine learning techniques.

REFERENCES

[1] S. Cohen, R.C. Kessler, and L.U. Gordon. Measuring stress: A guide for health and social scientists. Oxford University Press, USA, 1997.

[2] A. Perkins. "Saving money by reducing stress," Harvard Business Review, 72(6):12, 19942. S. Cohen, R.C. Kessler, and L.U. Gordon. "Measuring stress: A guide for health and social scientists," Oxford University Press, USA, 1997.

[3] Xie H, Andreae P, Zhang M, & Warren, P. (2004), "Learning models for English speech recognition", Proceedings of the 27th Australasian Computer Science Conference, Dunedin, New Zealand.

[4] Aull, A. M. & Zue, V. W. (1985) "Lexical stress determination and its application to speech," International Conference on Acoustics, Speech & Signal Processing, pp. 1549-1552.

[5] D. A. Cairns, J. H. L. Hansen, "Nonlinear Analysis and Detection of Speech under Stressed Conditions," Journal of the Acoustical Society of America, vol. 96, no. 6, pp. 3392-3400, 1994.

AUTHOR'S PROFILE

Neha P. Dhole received a Bachelor degree in 2007 in Electronics & Telecommunication from Sant Gadge Baba Amravati university and pursuing M.E. (Digital Electronics) from Sant Gadge Baba Amravati university, her papers was published in namely following journals [1] "Automatic Induction Motor Starter with programmable Timmer", International journal of advance management technology & engineering, science 2011, No. 2249-7455. [2] "Reconstruction of Watermark from digital images", "Research link, 2011 ISSN-0973-1628." Currently, she is a Lecturer in Government college of Engineering Amravati. His research interests include signal processing, control system and CMOS VLSI.