RESEARCH ARTICLE OPEN ACCESS

An Overview on Stress Identification in Speech

N.P. Dhole[1], S.N. Kale[2]
[1]*Assistant Professor, Department of Electronics and Telecommunication Engineering, PRMIT&R Badnera, Amravati (Maharashtra), India,*
Email Id-npdhole34@gmail.com
[2]*Associate Professor, Department of Applied Electronics, Sant Gadge Baba Amravati University, Amravati, (Maharashtra), India*
Email Id-sujatakale@sgbau.ac.in

Abstract:

Twenty papers are reviewed. Each paper consists of human speech under different stressful or emotional conditions. A basic description of each paper and its applications is provided. The conclusion of this study is that we find different algorithms with various databases and technologies are envisaged to detect stress under speech.

Keywords: Speech Recognition, Speech Databases, Stress Detection

I. INTRODUCTION

Speech signal is that signal in which a person communicates with one another. Speech production consists of sequence of articulatory movements, airflow from respiratory system & timing of vocal system. Again human emotion plays an important role in speech production which can analyze an individual state of mind in which he/she acts or reacts relating to surrounding. It may be termed as human behavior which 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 .Therefore stress can be defined as a psycho-physiological characterized by emotions, strain & deterioration of performance[1] . Therefore, it has become increasingly important to study speech under stress in order to improve the performance of speech recognition systems, to recognize when people are in a stressed state and to understand contexts in which speakers are communicating.

Following are the areas where identification of stress from speech includes:

A. Forensics

Deception detection systems, analysis of 911 phone calls that can include threats [2, 3].

B. Safety and Security

Air traffic controllers and pilots in noisy high stress environments, deep sea divers, NASA-space explorations, power system operators [2], military persons facing examination panel [4, 5], law enforcement training.

C. Psychology

Emotional state of patients- The levels where speech communication/production occurs and their corresponding Stress order [6, 7].

II. DESCRIPTION OF STRESS SPEECH WORK

The problem of stress identification has been receiving an increasing attention in related research communities due to wider recognition of potential problems caused by several reasons & due to the recent developments of technologies providing non- intrusive ways of collecting continuously objective measurements to monitor stress level.

Following literature review enlights work done by researchers on analyzing stress by different ways using speech.

Hindra Kurniawan, *et.al.* [11] have analyzed that stress level can be judged based on Galvanic Skin Response(GSR) & speech signal. But GSR & speech signal under stress may not be available at same time. The speech was sampled at a sampling rate of 44,100 Hz by using two channels. Facial expression was recorded using Handy cam Camcorders with High Definition (HD) resolution at 1, 440 × 1, 080 pixels .To make a GSR sensor measuring the changes in skin conductance they used the LEGO Mind storms NXT1 and an RCX wire connector sensor, which converts the analog reading to digital raw values in the range of 0 to 1,023.

Xiao Yao, *et.al* [12] have proposed a method for the classification of speech under stress that is based on a physical model for classification of neutral & stressed speech.

Bahador Makkiabadi & Saeid Sanei [13] have proposed a novel tensor factorization method which is developed to solve the under-determined blind source separation (UBSS) and especially under-determined blind identification (UBI) problems in mixed speech signals.

Gopala Krishna Anumanchipalliyz, *et.al.* [14] have proposed a new approach to pitch transformation, that can capture aspects of speaking style to convert speech from one speaker and make it sound like another speaker. They choose the CMU ARCTIC databases which has multiple male/female speakers delivering the same set of sentences, with roughly 1 hour of speech each.

Tom Giraud, *et.al*. [15] have presented a protocol for collecting multimodal non-acted emotional expressions in a stressful situations of speech. For the voice, we used a wireless microphone system attached on the participant's clothes. They collected audio with lapel-microphone (AKG PT40 FLEXX with Signal/noise ratio: 110 dB) at 16 kHz. The gain was adapted manually. Whole body behaviors were recorded by one Sony HDR-CX550 camera at 25fps in full HD and one Kinect in front of the participant.

S. Kirbız [16] has proposed an adaptive time-frequency resolution based on single channel sound source separation method using Non-negative Tensor Factorization (NTF).

Cecilia Damon, *et.al*. [17] have introduced a new method for artifact removal for single-channel Electro Encephalogram (EEG) recordings using nonnegative matrix factorization (NMF) in a Gaussian source separation framework.

James Z. Zhang, *et.al*.[18] have proved a new method called Adaptive Empirical Mode Decomposition(AEMD) applied to voice stress detection where voice stress is found by finding microtremour frequency. Experimental data for this research came from two sources, one is the Speech under Simulated and Actual Stress (SUSAS) database, and the other is "Deception Interviews of Volunteers" developed for this research.

Ling He, *et.al.*[19] have presented an automatic stress recognition methods based on acoustic speech analysis. The Speech under Simulated and Actual Stress (SUSAS) database comprises a wide variety of acted and actual stresses and emotions.

Ya Li et.al [20] have considered a HMM-based expressive speech synthesis system which supports Mandarin stress synthesis. The audio corpus used in this work contains 6000 sentences (about 73000 syllables), which are read by a professional female speaker.

Ling He, *et.al.* [21] have presented a new system for automatic stress detection in speech. Results indicate that the proposed method can be applied to voiced speech in speech independent conditions. The Speech under Simulated and Actual Stress (SUSAS) database [7] was used to select the training and testing sets. The SUSAS data base contains speech recordings under stress made by actors as well as speech recordings made in actual stressful work situations. Only the speech signals recorded under actual stressful work situations were used to perform the classification into three classes of speech: high-level stress, low-level stress and neutral.

Sumitra Shukla, *et.al*.[22] have studied the effect of stress in human and automatic stressed speech processing tasks for speech collected from non-professional speakers. A simulated stressed speech database is collected in Hindi, Indian language. Recording is done using fifteen nonprofessional adult speakers under five most frequently used stress conditions, namely, neutral, angry, happy, and sad and Lombard.

Sumitra Shukla, *et.al* [23] have presented a subspace projection based approach which is tested for separation of speech and stress information present in stressed speech.

Ling He, *et.al.* [24] have presented a new method that extract characteristic features from speech magnitude spectrograms. It proves that the Gaussian mixture model is the best.

Xiao Ya *et.al* .[25] have proposed a two-mass vocal fold model is fitted to estimate the stiffness parameters of vocal folds during speech, and the stiffness parameters are then analyzed in order to classify recorded samples into neutral and stressed speech. A soundtrack of video recordings from the Oregon Research Institute (ORI) was used to select speech samples for processing. The data included 71 parents (27 mothers and 44 fathers) video recorded while being engaged in a family discussion with their children.

Nurul Aida *et.al*.[26] have proposed a feature extraction method using two different wavelet packet filter bank structures which are based on bark scale and equivalent rectangular bandwidth (ERB) scale for identifying the emotional/stressed states of a person. In the experiment, we used a database collected by the Fujitsu Corporation [10]. This database contains speech samples from eleven subjects, four for male, and seven female. To simulate mental pressure resulting in psychological stress, three different tasks were introduced. These tasks were performed by the speaker while having a telephone conversation with an operator, in order to simulate a situation involving pressure during a telephone call.

Nurul Aida, *et.al*. [27], have presented a feature extraction method based on wavelet packet decomposition for detecting the emotional or stressed states of the person.

Clara Vania, *et.al*. [28], have analyzed the effect of syllable and word stress on the quality of synthesized speech. Subjective assessment of Indonesian speech synthesis system in terms of both quality and intelligibility result then conducted to evaluate the results of our systems.

Pavel Sala & Milan Sigmud [29] have worked on methods for the estimation of glottal pulses from speech signal .This method was applied to the speech under stress aimed to investigate the influence of stress by speaker on the generating glottal flow. Presented results were obtained using speech data from the Exam Stress database (recordings of 9 speakers only) and applying the IAIF method for the glottal flow estimation.

Oliver Jokisch *et.al*.[30] summarize stress-related analysis results for pitch, duration and intensity and their interrelations.

III. CONCLUSIONS

From the study of these twenty papers, we find that there are numerous techniques to detect stress into the speech. Again different researchers have worked on different databases as well their own datasets to formulate the stress components present into speech. We conclude that there is a contribution of novel approaches by different Researchers in this vast Stress detection field through speech signal.

REFERENCES

1. *HJM Steeneken, JHL Hansen, "Speech under stress conditions: overview of the effect on speech production and on the system performance", in Proc. ICASS (ATLANTA, GEORGIA, 1996).*
2. *Alm C.O., Roth D., Sproat R, " Emotions from Text: Machine Learning for Text based Emotion Prediction", Proceedings of HLT/EMNLP 05,Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP), pp. 579–586, Vancouver, October 2005.*
3. *Hollien H., "Forensic Voice Identification", Academic Press, London 2002.*
4. *Prahallad K., Black A., Mosur R, "Sub-Phonetic Modeling for Capturing Pronunciation Variation in Conversational Speech Synthesis", In Proceedings of the 31st IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2006.*
5. *Ruzanski E., Hansen J.H.L., Meyerhoff J., Saviolakis G., Koenig M., " Effect of phoneme characteristics on TEO Feature-based Automatic Stress Detection in Speech", Proceedings of the 30th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05), Philadelphia, vol. 1, pp. 357–360,2005.*
6. *Hansen J.H.L., Swail C., South A.J., Moore R.K., Steeneken H., Cupples E.J., Anderson T., Vloeberghs, C.R.A., Trancoso I., Verlinde P, "The Impact of Speech under Stress on Military Speech Technology in NATO RTO-TR-10, AC/323(IST) TP/5 IST/TG-01", 2000.*
7. *Murray I.R., Baber C., South A., "Towards a Definition and Working Model of Stress and its Effects on Speech. Speech Communication", Speech Communication Journal, vol. 20, Issue 1-2, Nov.1996, pp.3-12.*
8. *Schreuder M.J., "Prosodic Processes in Language and Music", PhD thesis, Research School of Behavioral and Cognitive Neurosciences (BCN), Groningen University of Groningen 2006.*
9. *Hansen J.H.L., Cairns D.A, "Source Generator based Real-Time Recognition of Speech in Noisy Stressful and Lombard Effect Environments" Speech Communications Journal vol. 16 (4), pp. 391–422, 1995.*
10. *Yapanel U.H., Hansen J.H.L, "A New Perspective on Feature Extraction for Robust In-Vehicle Speech Recognition". Proceedings of the 8th European Conference on Speech Communication and Technology (Eurospeech '03), Geneva, Switzerland, pp. 1281–1284, 2003.*
11. *Hindra Kurniawan, Alexandr V. Maslov, Mykola Pechenizkiy, "Stress Detection from Speech and Galvanic Skin Response Signals", IEEE Conference Board of Mathematical Sciences ,Press 2013, pp. 209-214.*
12. *Xiao Yao, Takatoshi Jitsuhiro, Chiyomi Miyajima, Norihide Kitaoka1, Kazuya Takeda "Estimation of vocal tract parameters for the classification of speech under stress", 'IEEE International Conference on Acoustics, speech, & Signal Processing 2013, pp.7532-7536.*
13. *Bahador Makkiabadi, Saeid Sanei, "Orthogonal Segmented Model for Underdetermined Blind Identification and Separation of Sources with Sparse Events", 2013 IEEE International Workshop On Machine Learning For Signal Processing, Sept. 22-25, Southhampten, UK 2013, pp.1-6*
14. *Gopala Krishna Anumanchipalli, Luıs C. Oliveira, Alan W Black,"A Style Capturing Approach To F0 Transformation In Voice Conversion", 'IEEE International Conference on Acoustics, speech,& Signal Processing ' 2013, pp.6915-6919.*
15. *Tom Giraud, Mariette Soury, Jiewen Hua, Agnes Delaborde, Marie, "Multimodal Expressions of Stress during a Public Speaking Task", 'IEEE 2013 Humane Association Conference on Affective Computing and Intelligent Interaction', pp. 417-422.*
16. *S. Kirbız, B. Gunsel, "An Adaptive Time Frequency Resolution Framework for Single Channel Source Separation Based on Non-Negative Tensor Factorization", 'IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)' 2013, pp. 905-909.*
17. *Cecilia Damon, Antoine Liutkus, Alexandre Gramfort, Slim Essid, "Non-Negative Matrix Factorization for Single Channel EEG Artifact Rejection", 'IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)' 2013, pp.1177-1181.*
18. *James Z. Zhang, Nyaga Mbitiru, Peter C. Tay, Robert D Adams, "Analysis of Stress in Speech Using Adaptive Empirical Mode Decomposition", IEEE 2009, 43rd ASILOMAR conference on signals, systems & computers, pp. 361-365.*
19. *Ling He, Margaret Lech, Namunu C. Maddage and Nicholas Allen, "Neural Networks and TEO features for an Automatic Recognition of Stress in Spontaneous Speech", IEEE Fifth International Conference on Natural Computation, Computer Society',2009, pp. 227-231.*
20. *Ya Li, Shifeng Pan, Jianhua Tao, "HMM-based Speech Synthesis with a Flexible Mandarin Stress Adaptation Model", 'International Conference on Signal Processing (ICSP)' 2010, pp.625-628.*

21. *Ling He, Margaret Lech, Namunu C. Maddage, Nicholas Allen, "Stress Detection Using Speech Spectrograms and Sigma-pi Neuron Units", IEEE Computer Society 2009, pp. 260-264.*
22. *Sumitra Shukla, S R M Prasanna, S. Dandapat, "Stressed Speech Processing: Human Vs Automatic in Non-professional Speakers Scenario", IEEE National Conference Communications (NCC), 2011, pp. 978-982.*
23. *Sumitra Shukla, S.Dandapat and S. R. Mahadeva Prasanna, "Subspace Projection Based Analysis of Speech under Stressed Condition", 'IEEE Information & Communication Technologies', 2012, pp. 831-834.*
24. *Ling He, Margaret Lech, Namunu Maddage Nicholas Allen, "Stress and Emotion Recognition Using Log-Gabor Filter Analysis of Speech Spectrograms", 'IEEE 3rd International Conference Affective Computing and Intelligent Interaction and Workshops (ACII)', 2009, pp. 4244-4250.*
25. *Xiao Yao, Takatoshi Jitsuhiro, Chiyomi Miyajima, Norihide Kitaoka Kazuya Takeda, "Physical Characteristics of Vocal Folds during Speech Under Stress", 'IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)' 2012, pp.4609-4612.*
26. *Nurul Aida Amira Bt Johari, M.Hariharan, A.Saidatul, Sazali Yaacob, "Multistyle Classification of Speech Under Stress using Wavelet Packet Energy and Entropy Features", 'IEEE Conference on Sustainable Utilization and Development in Engineering and Technology (STUDENT)'.The University of Nottingham, Semenyih, Selangor, Malaysia. 20-21 October 2011, pp.74-78.*
27. *Nurul Aida Amira Bt Johari, M.Hariharan, A.Saidatul, Sazali Yaacob, "Assimilate the Auditory Scale with Wavelet Packet Filters for Multistyle Classification of Speech Under Stress", 'IEEE 2012 International Conference on Biomedical Engineering (ICoBE)',27-28 February 2012,Penang 2012, pp.537-542.*
28. *Clara Vania, Mirna Adriani, "The Effect of Syllable and Word Stress on the Quality of Indonesian HMM-based Speech Synthesis System", 'International Conference on Advanced Computer Science and Information Systems', 2011, pp.413-417.*
29. *Pavel Sala, Milan Sigmud, "Diagnostical analysis of voice based on glottal pulses", 'IEEE 2009 19th International Conference Radioelektronika', 22-23 April 2009, pp. 3538-3542.*
30. *Oliver Jokisch, Yitagessu Birhanu, Rudiger Hoffmann, "Syllable-Based Prosodic Analysis of Amharic Read Speech", 'IEEE Spoken Language Technology Workshop (SLT)', 2012, pp. 258-262.*