International Journal of Electrical, Electronics and Computer Engineering **2**(2): 148-154(2013)

ISSN No. (Online): 2277-2626

Voice Stress Detection

Vijay P. Patil*, Krishna Kant Nayak** and Manish Saxena**

** PG Scholar, Department of Electronics and Communication Engineering, BIST, Bhopal, (MP)*
***Assistant Professor, BIST, Department of Electronics and Communication Engineering, BIST, Bhopal, (MP)*

(Received 05 November, 2013 Accepted 07 December, 2013)

ABSTRACT: Studies have shown that distortion introduced by stress or emotion can severely reduce speech recognition accuracy. Techniques for detecting or assessing the presence of stress could help neutralize stressed speech and improve robustness of speech recognition systems. The performance of a speaker recognition system decreases when the speaker is under stress or emotion. In this dissertation we explore and identify a mechanism that enables use of inherent stress-in-speech or speaking style information present in speech of a person as additional cues for speaker recognition. Stress in human speech can be detected by various methods know as Voice Stress Analysis (VSA).Stress detection in voice gives a great alternative for obtaining a noninvasive way to extract information about a possible deception from a person declaration. This article contains information and results of a primary work done to show how changes the stress For Mel Frequency Cepstral Coefficient features can be detected through FFT signal processing when a person is under psychological pressure. The principal purpose is to obtain a tool that could help innocent people to prove their guiltlessness of having committed an offense or a crime

Keywords: Speech recognition, FFT, voice stress analyzer

I. INTRODUCTION

Voice stress detection has found its applications in many areas including voice activated military equipment, psychological testing and deception detection. Voice stress analysis (VSA) is accomplished by measuring fluctuations in the physiological micro tremor present in speech. A micro tremor is a low amplitude oscillation of the reflex mechanism controlling the length and tension of a stretched muscle caused by the finite transmission delay between neurons to and from the target muscle. Micro tremors are present in every muscle in the body including the vocal chords and have a frequency of around 8 – 12 Hz. During times of increased stress this microtremor shifts in frequency. This change in frequency transfers from the muscles in the vocal tract to the voice produced. Stress can thus be detected by analyzing the change in micro tremor frequency of an individual's voice. Conventional algorithms include Fast Fourier Transform (FFT) as well as McQuiston- Ford algorithm. However, the accuracy of voice stress detection depends on the algorithms in use as well as the effectiveness of the examiners. Speech is the most important means of communication among humans. Speech, however, is not limited only to the process of communication, but is also very important for transferring emotions, expressing our personality, and reflecting situations of speaking. Modern lifestyles have increased the risk of experiencing some kind of voice alteration. Speech system performance is greatly impacted by speech produced under stress. Speakers encounter various forms of stress, namely, physical (e.g. running), cognitive (*e.g.* driving), chemical (*e.g.* medication), fatigue (*e.g.* sleep deprivation), and Lombard-effect (*e.g.* speech in noisy environment) [1]. Here, stress is known to induce changes in the spectral and temporal patterns of speech.

As a result, it is important to establish the nature of these changes in order to compensate for the effects and improve speech system performance. This work focuses on studying the impact of physical stress on temporal patterns of speech.

II. LITERATURE REVIEW

Stress and its manifestation in the acoustic signal have been the subject matter of many studies in literature. Researchers have attempted to determine reliable indicators of stress by analyzing certain variable parameters of speech such as fundamental frequency (pitch), amplitude, concentration of spectral energy, duration and several others. In literature, analysis of stress is performed through analysis of some parameters of stress like fundamental frequency (F0), pitch, vowel duration and formants in recorded emotional speech, namely, analyzing a speaker's speech when they are under stress, fatigue, heavy workload, environmental noise, sleep loss or expressing some emotion like happiness, anger or sorrow.

III. SPEECH UNDER STRESS

Stress is more or less present in all professions in today's hectic and fast-moving society. Stress is generally defined as a strain upon a bodily organ or mental power Depending on its duration and intensity, stress can have short- or long-lasting effects. Negative influence of stress on health, professional performance as well as interpersonal communication is well known. A comprehensive reference source on stressors, effects of activating the stress response mechanisms, and the disorders that may arise as a consequence of acute or chronic stress is provided, for example, in the Encyclopedia of Stress.

A. Theory behind stress detection

Corpus: The audio corpus used in this work contains 6000 sentences (about 73000 syllables), which are read by a professional female speaker. All the utterances are segmentally labeled according to the audio data by research assistants. During the stress labeling, three assistants were trained with a subset corpus several times first. A small percentage of disagreement is acceptable due to the frequent perception confliction among tone, intonation and stress. The aim of training is to keep consistency of each annotator with their own during the whole annotating process and among annotators as much as possible. Three levels of stress are adopted here, namely stressed, regular and unstressed syllable according to their prominence degrees within a prosodic word. To reduce the impact of the surrounding syllables on the perception of the current syllable, we segmented the utterance into prosodic words and stored them according to their tone patterns separately.

B. Voice stress Analyzer

A VSA can be defined as a tool that shows measurements, without body contact, of the involuntary psychology answer from a person voice that is under stress. The measurement can be either in a graphical or non-graphical form. The theory that supports VSA is the one whose subject matter is the assumption that it had been discovered that human body has involuntary responses that are associated with deceptive answers and/or stressful situations. Some of these responses are the lack of Lippold micro tremor or infrasonic frequency modulation in voice. An analyzer of stress measures the inexistence of micro oscillations that modulate human voice when the person under test is under stress. The phenomenon can be explained with the following: the muscles in the throat, which mainly modulate voice, get rigid suppressing the frequency modulation on the voice, especially infrasonic modulation, which means that fundamental frequency of the voice does not show infrasonic frequency modulation.

C. Polygraph

A **polygraph** (popularly referred to as a **lie detector**) measures and records several physiological indices such as blood pressure, pulse, respiration, and skin conductivity while the subject is asked and answers a series of questions. The belief underpinning the use of the polygraph is that deceptive answers will produce physiological responses that can be differentiated from those associated with non-deceptive answers.

IV. FEATURES ANALYZED

DFs in this thesis are analyzed with respect to "features," where features are defined as observable characteristics in the data.

A distinction is drawn between "features" and "variables." Variables are defined as underlying explanatory factors to which an observable trend can be attributed. Features are the surface manifestations of these variables. While the data-driven approach adopted dictates that the analyses herein are described in terms of features, the long-term goal is of course to discover the variables underlying these features, since it is in terms of variables rather than features that a unified theory must be phrased. We expect that the study of DFs should cut across areas relevant to the study of fluent speech. Therefore, a variety of different features are examined. As an arbitrary convenience for organizing the large number of features, as well as for relating the present studies to previous work, features are organized into a small renumber of logically orthogonal "dimensions." Note that while this organization of features is useful for discussion, no *theoretical* significance is attached to these groupings. Features are epiphenomena of underlying explanatory variables, and the relationships among these explanatory variables are inherently unknown. Therefore, feature dimensions have no reality at the level of interpretation.

A. Domain features

Domain features include observable or manipulated aspects of the speech setting in which a DF is produced, such as the purpose of the discourse, or the communication mode. In this work, a single domain feature is examined, corresponding to the corpus of speech data from which the DF was drawn.

B. Speaker features

Speaker features are aspects associated with the individual who produces a DF. Speaker features examined include "speaker identity" in analyses of individual differences, speech rate, and speaker gender.

C. Sentence features

Sentence features make reference to the sentence in which a DF occurs, without making reference to the internal arrangement of words in the sentence. This is a somewhat contrived dimension, but useful for distinguishing the types of features examined in this paper from syntactic features, which do not play a direct role in this work. Sentence features examined include the length of the sentence in which a DF occurs, the sentence position of a DF (sentence-initial or sentence-medial), and the presence of other DFs in the same sentence.

D. Acoustic features

Acoustic features are concrete measurements of properties of the speech signal, including fundamental frequency and duration. These features are studied for only the two simplest classes of DF phenomena, filled pauses and repetitions.

V. PROPOSED MODEL

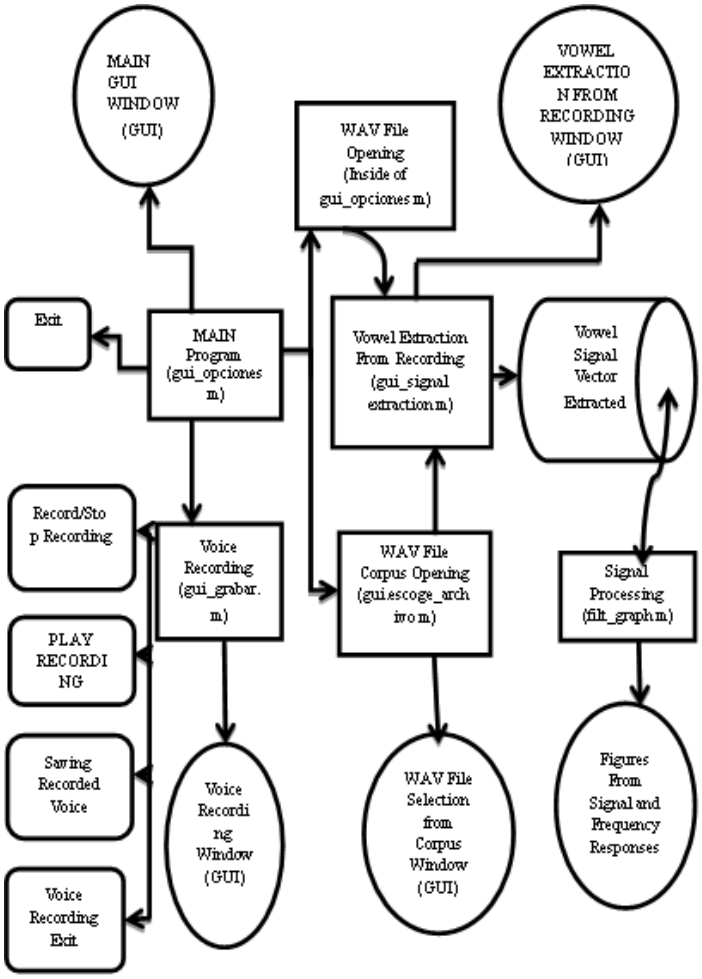

Fig. 1. GUI's Communication & Processing Mode.

VI. ALGORITHM USED

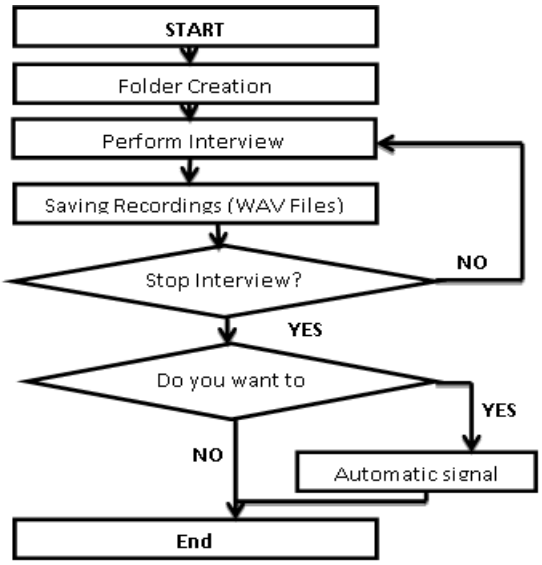

Fig. 2. Questioning Signal Process.

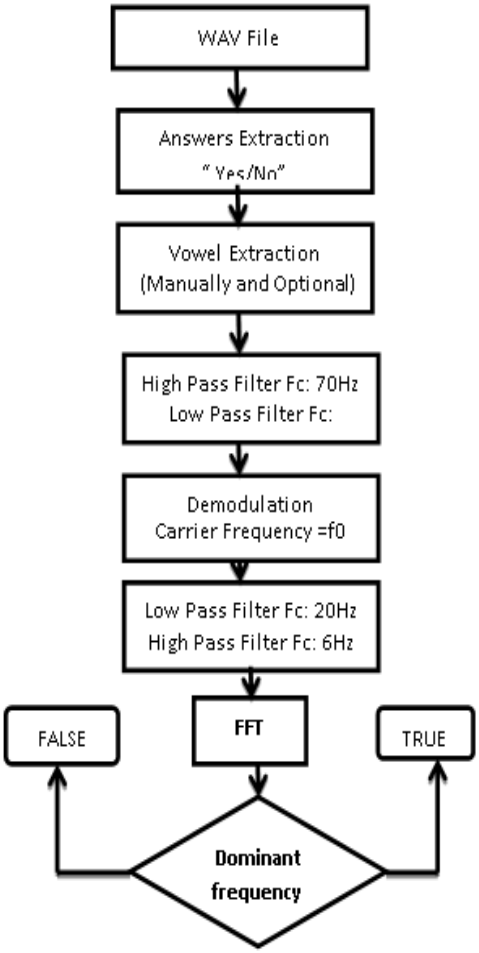

Fig. 3. Voice Signal Processing.

VII. RESULT AND DISCUSSION

A. Training dataset

In training dataset we generate the 10 question for the training purpose. Using this database we create the different interview with different five users.

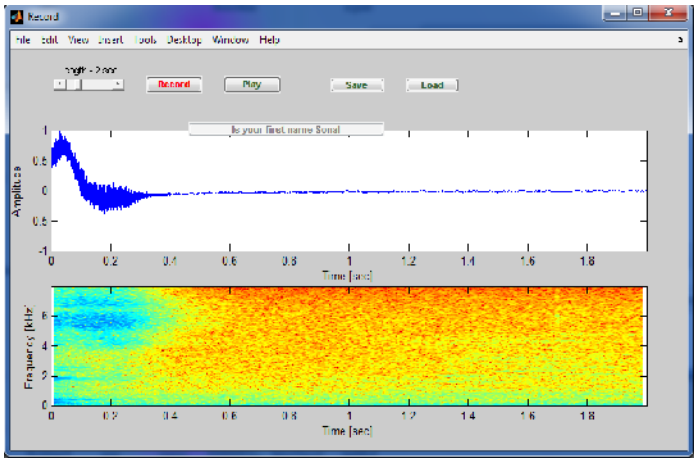

Fig. 4. Output window of training dataset sample.

B. Testing Dataset

In Testing dataset we test the 10 question using multiple users Using this database we create the different output with different multiple users.

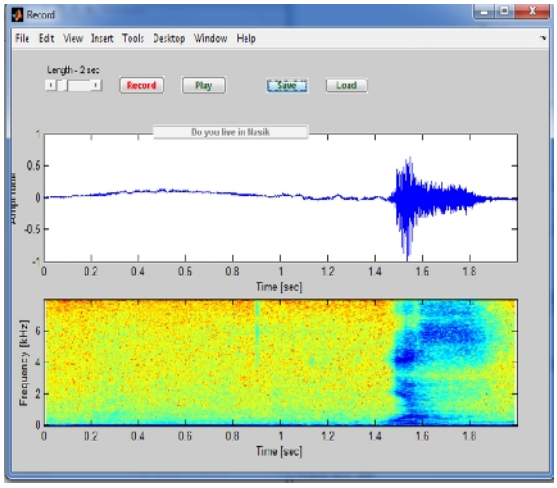

Fig. 5. Output window of testing dataset sample.

C. Interview results

Table 1: Result from the test applied over the first interview.

Sr. No	**Questions**	**Answer**	**Result**
1	Is your first name Vijay?	YES	TRUE
2	Do you live in Nasik?	YES	TRUE
3	Do you know the address of Taj hotel?	NO	FALSE
4	Are you married?	YES	TRUE
5	Did you have a smartphone?	NO	TRUE
6	Are you working in college?	NO	FALSE
7	Did you take the laptop?	YES	TRUE
8	Whether the selected area is optimal	NO	TRUE
9	Whether the production process chosen is suitable?	YES	TRUE
10	Do I have the right strategy?	YES	TRUE

Table 2: Result from the test applied over the second interview.

Sr. No	Questions	Answer	Result
1	Are you a very good decision maker?	YES	TRUE
2	Would you like to play cricket right now?	NO	TRUE
3	Do you read Times Of India daily?	NO	TRUE
4	Would you like to sing a song right now?	YES	TRUE
5	Do I have the right strategy?	YES	FALSE
6	Are you working in industry?	YES	TRUE
7	Is the capital of Madhya Pradesh is Bhopal?	NO	TRUE
8	Do you know who invented the digital camera?	NO	FALSE
9	Do you know the different kinds of Stress?	YES	FALSE
10	Are you over 28 years of age?	YES	TRUE

Table 3: Result from the test applied over the third interview.

Sr. No.	Questions	Answer	Result
1	If today is Friday, then coming day after tomorrow is Sunday?	YES	TRUE
2	Is February has 29 days in 2014?	NO	TRUE
3	Is the number 17 is a prime number?	NO	FALSE
4	Is Shahrukh Khan is the best actor as compared to Salman Khan?	YES	FALSE
5	Did you get it?	YES	TRUE
6	Have you ever done sex with anyone apart from your life partner?	NO	FALSE
7	Are you a Gay?	NO	TRUE
8	Are you sure?	NO	TRUE
9	Is December has 31 days?	YES	TRUE
10	Is the number 72 is divisible by 9?	YES	TRUE

VIII. CONCLUSION

FFT is very used method for speech recognition and it was applied to show that in effect there occur changes in the frequency component of demodulated voice signal. The majority of people, who were interviewed to give their answers knowing that no critical consequences could derive from their answers. In some cases, no stress is detected. In order to obtain more clearly results, it is proposed to perform recordings to interviews sessions over people that are in crime. These people naturally will be under real pressure and then when the answers are analyzed we would obtain better results of stress detection. Voice stress analysis (VSA) technology is evaluating its effectiveness for both military and law enforcement applications. Finally we conclude that VSA technology can identify stress better than polygraph systems that experience and training improves the accuracy of result.

REFERENCES

[1]. S.T. Jovicic, Z. Kasic, and M. Dordevic, "Corpus creating of speech expression of emotions and attitudes in serbian language," in GEES. Conference TELFOR, 2003.

[2]. B.D. Womack and J. H. L. Hansen, "Classification of speech under stress using target driven features," *Speech Commun.*, vol. **20**, pp. 131–150, June 1996..

[3]. F. Burkhardt, A. Paeschke, M. Rolfes, W. S. ndlmeier, and B. Weiss., "A database of german emotional speech," in Interspeech 2005 *Eurospeech, 9th European Conference on Speech Communication and Technology*, 2005.

[4]. B. Schuller, S. Steidl, and A. Batliner, "The interspeech 2009 emotion challenge," 2009.

[5]. S. Steidl, "Automatic classification of emotion-related user states in spontaneous childrens speech," Ph.D. dissertation, Erlangen, 2009.

[6]. D. Ververidis and C. Kotropoulos, "A review of emotional speech databases," in In: PCI 2003, 9th Panhellenic Conference on Informatics, November 1-23, 2003.

[7]. J. H. L. Hansen, "Morphological constrained feature enhancement with adaptive cepstral compensation (MCE-ACC) for speech recognition in noise and lombard effect," *IEEE Trans. Speech Audio Process*., vol. **2**, pp. 598–614, Oct. 1994.

[8]. S. I, S. I, Navas, E. Hernez, I. Sanchez, J. Luengo, and I. O. I., "Subjective evaluation of an emotional speech database for basque," in Sixth International Language Resources and Evaluation (LREC'08), 2008.

[9]. S. B. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabicword recognition in continuously spoken sentences," *IEEE Transactions on Acous. Speech and Signal Process*. vol. **28**, pp. 357–366, August 1980.

[10]. Y. linde, A. BUzo, and R. M. . Gray, "An algorithm for vector quantizer design," *IEEE Trans. Communication*, vol. **28**, pp. 84–96, Jan 1980.

[11]. L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proc. IEEE*, vol. **77**, pp. 257–286, 1989

[12]. B. D. Womack and J. H. L. Hansen, "Classification of speech under stress using target driven features," *Speech Commun.*, vol.**20**, pp. 131–150, June 1996.

[13]. S. Steidl, "Automatic classification of emotion-related user states in spontaneous childrens speech," Ph.D. dissertation, Erlangen, 2009.