Vocal Indicators of Emotional Stress

Savita Sondhi
Electrical, Electronics & Communication Engg
ITM University
Gurgaon-122017, India

Munna Khan
Electrical Engineering
Jamia Millia Islamia
(Central University)
New Delhi-110025, India

Ritu Vijay
Department of Electronics
Banasthali, Banasthali University
Rajasthan-304022; India

Ashok K. Salhan
Defence Institute of Physiology and Allied Sciences (DIPAS),
DRDO,
New Delhi-110054, India

ABSTRACT
Background: Voice, apart from its semantic content also carries information about the speaker's psychological and physical state. Emotional stress or physical fatigue, are the pathological elements of this condition. The possible relationship between emotional stress and the measurable changes to the voice signal was the subject of this study. **Method:** Eleven subjects were interviewed with questions from two domains and their responses were recorded. In the first domain, two men, two women and three teenagers were asked to remember an incident from their past where they felt embarrassed or ashamed of their own act. In the second domain, three women and one man from the house keeping staff were interviewed for the stolen mobile phone. These subjects were different from the subjects who participated in domain 1. Stress in voice was detected as a measure of shift in the acoustic parameters with respect to the baseline. All recordings were analyzed using PRAAT software. Spectrograms were also plotted for qualitative comparison between normal speech and stressed speech. **Result:** Significant increase in mean pitch and substantial decrease in the first two formants (F1 and F2) were observed under stress. Other acoustic measures did undergo change under stress but failed to reveal any significance. Spectrograms were distinct for the two conditions. **Conclusion:** Obtained results indicate that, when a person is emotionally charged, stress could be discerned in his voice. Mean pitch and Formants F1 and F2 have been obtained as reliable vocal indicators of emotional stress. This study proposes a simple non-invasive approach which can act as an alibi for innocent people.

General Terms
Voice stress analysis, speech processing, human computer interface.

Keywords
Deception, anxiety, stress, spectrogram, mean pitch, formants.

1. INTRODUCTION
Intentional deception is not only pervasive but also widely accepted form of communication. Knapp and Comadena (1979) reported act of deception as an ancient technique used by human beings to adapt with their surroundings (p.275) [12]. Knapp, Hart and Dennis (1974) described deception as a communication strategy to avoid conflicts and hide veracity [11]. Today deceptive discourses are common during political campaigns, diplomatic negotiations and also in family relationships. Interrogators investigating criminal cases as well as human resource professionals during the employment screening are often faced with the challenge to differentiate between truth and deception. Fielder and Walka, (1993) suggested that, due to lack of knowledge of cues, these professionals in most cases use heuristics i.e. simple rules of thumb [8]. Drawback of using heuristics is that, they prevent critical thinking and detailed in-depth analysis of the clues. Therefore, instead of relying on heuristics for detection of deception, it is important to carefully analyze the oral discourse to identify vocal cues of stress. Voice is a part of physical body. Voice is unique like fingerprints. Although the main role of voice is to communicate, it can also express emotions, intention, stress as well as the context in which the speaker is talking. Warren and Riedel (2004) explains, that the moment an emotion is distorted, it instantly distorts the voice [32]. During a conversation no matter how efficiently a person tries to hide his emotions, voice reveals the truth. Previous studies by (Williams and Stevens, 1972; Cummings and Clements, 1995; Scherer, 1986; Scherer et al., 2001) have reported that facial and vocal features are the potential indicator of human emotions like happiness, anger, fear, sorrow, disgust, and stress [3, 27, 29, 31]. Emotions are strong innate components. As per Scherer, (2000) emotions affect the psychophysical state of an individual in the form of physiological arousal, motor expression, and subjective feeling [28]. While subjective feeling refers to the internal condition of an individual under stress, physiological arousal and motor expression can be observed from facial expression, body language, and speech. In yet another study, Zhou, Hansen et al., (2001) reported that emotions also affect the muscles of the vocal fold and the excitation source of speech, which in turn depends on the vibration of the vocal folds [34]. Therefore, in situations of increased emotional stress, the striated muscles surrounding the vocal cords contract in response to the stimulus. Thus stress possess an inherently communicative role which can be discerned in the voice. While classifying neutral and stressed speech based on stiffness parameter of vocal folds, Xiao, Jitsuhiro et al; (2012) observed that, "when a speaker experiences stress, the contraction of cricothyroid muscle causes higher tension, while the activity of thyroarytenoid muscle becomes relatively low, which causes lack of articulation in the speech produced" [33]. Thus stress changes the spectral and temporal characteristics of speech and degrades the voice quality. Numerous studies since 1960's (Williams and Stevens, 1972; Cosetl et al. 2011; Smith, 1977; Ruiz, 1990, Ruiz, 1996; Sigmund, 2007; Bageshree V et al. 2012; Costantini et al. 2014) have attempted to identify potential indicators of stress by analyzing the acoustic parameters like fundamental frequency (F_0), voice perturbations and formants in recorded speech as well as real time voicing [2, 4, 5, 18, 19, 20, 21, 31]. It has been well documented that, fundamental frequency (F_0) of voice is a potential indicator of stress [10, 20 25, 25]. Literature describes that when a person experiences stress, F_0 shifts from its baseline. However, much of the research draws from simulated laboratory studies, therefore they lack the element of deception. It is felt that vocal measures under natural and real life, stressful conditions might be different.

Therefore this study was aimed at analyzing voice samples of subjects under real life stressful event. Following this introduction, section 2 describes the method and materials used for the study. Results of emotional stress are given in section 3. Discussion and concluding remarks are presented in section 4 and 5 respectively.

2. MATERIALS AND METHODS

Eleven subjects participated in this study. They were interviewed with questions from two different domains and their answers were recorded. In the first domain, two men, two women and three teenagers were asked to remember an incident from their past where they felt embarrassed or ashamed of their own act. In the second domain, three women and one man from the house keeping staff were interviewed for the stolen mobile phone. These subjects were different from the subjects who participated in domain 1. The interview consisted of a questionnaire which started with simple questions but gradually became more stressful to answer due to the nature of the question. This paper present results from both the domains. Voice was recorded using a mobile phone of Samsung S-II model in .MP3 format with 44.1 KHz sampling rate. As quality of recording was very important in this study therefore, voice was recorded in a closed noise free room with fans and air conditioner turned off. Teenager had no knowledge that the discussion was being recorded. This was intended to eliminate any possibility of additional stress. Henceforth in this paper this teenager will be referred as subject-1. The discussion started in a normal relaxed tone, where subject-1 shared a painting, signed and gifted to all his classmates by a renowned painter at school. The discussion continued further and the moment a specific question referring to an incident at his school was asked, he gets nervous. Subject-1 had bluntly refused to acknowledge an elderly family friend who met him at school. In another domain, four housekeeping staff were interviewed, who had visited the room where the mobile was kept before being stolen. Interview questions for this domain were carefully framed based on MZOC-GC protocol. Answers were recorded using an HCL Desktop placed on a table along with the Philips SHM1500 PC VOIP microphone connected to the sound port of the desktop. Past studies have confirmed pitch to be a reliable indicator of stress [10]. Therefore this study was aimed to evaluate the influence of emotional stress on other acoustic features like jitter, shimmer and first four formants F1-F4 as well. In order to observe the influence of nervousness (anxiety) and deception on acoustic characteristics of voice, the recorded audio files were analyzed using PRAAT software version 5.3.56 Boersma and Weenink, (2010). PRAAT is a freeware program for the analysis of speech in phonetics [1]. PRAAT applies autocorrelation method for pitch analysis. Duffy, (2003) explains that the algorithm performs acoustic periodicity detection based on accurate autocorrelation method [6].

RESULT

2.1 Analysis for anxiety: interview results of subject-1

The recorded audio file of subject-1 was read using PRAAT software. The discussion was in Indian language. Therefore the word ‘haan’ which mean ‘yes’ in Hindi spoken by subject-1, during normal (relaxed) discussion and later under nervous state were selected for detailed analysis and comparison. ‘Stressed speech’ in this paper will refer to speech under nervousness. This study explores the effect of stress on the acoustic measures like fundamental frequency, jitter, shimmer and first four formants (F1-F4) of voice. Fundamental frequency is the frequency of vibration of vocal cord inside the larynx, whereas, formants are formed due to the vibration of air inside the vocal tract. As jitter and shimmer reveal source of variability in the frequency and amplitude of vocal fold vibrations, therefore, both these basic measures were also considered for acoustic measurements. Jitter (local) and shimmer (local) measurements were extracted in this study. Jitter (local) is the average absolute difference between consecutive intervals, divided by the average interval. Similarly, shimmer (local) is the average absolute difference between amplitudes of consecutive periods, divided by the average amplitude. Threshold measures for pathology, indicate that jitter (local) is 1.040% and shimmer (local) is 3.810%. Table 1 shows the acoustic measures (averaged across all subjects of domain 1) for normal speech and stressed speech. These acoustic measures are obtained from the built in functions of PRAAT software. Spectrograms obtained from PRAAT are distinct for both normal and stressed speech as shown in Fig 1. The blue dots represent pitch contour. It can be clearly seen, that pitch contour shifts upward for the stressed speech indicating significant increase in fundamental frequency (also referred as pitch) under stress. Significant increase in mean pitch and decrease in F1 and F2 was observed across all subjects under stress.

Table 1: Acoustic measures for normal as well as stressed speech (average across subjects).

	Normal speech	**Stressed speech**
Mean Pitch (Hz)	185.38	216.90
Formant F1 (Hz)	335.93	269.03
Formant F2 (Hz)	2256.22	1941.67
Formant F3 (Hz)	2860.49	3142.49
Formant F4 (Hz)	4267.45	3987.12
Jitter (%)	1.36	1.18
Shimmer (%)	8.60	6.82

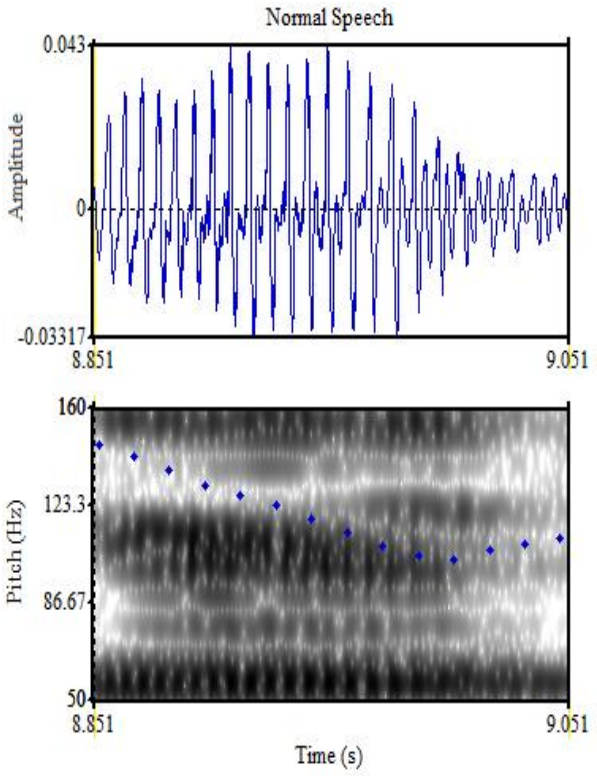

Fig. 1 (a): Spectrogram of normal speech.

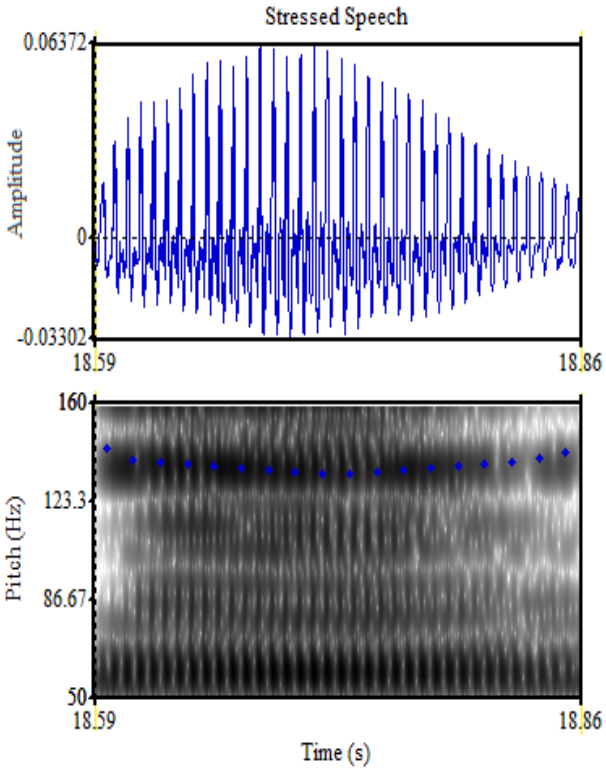

Fig. 1 (b): Spectrogram of stressed speech.

2.2 Analysis for deception: interview results of the suspicious female subject for the stolen mobile case

Interview for the stolen mobile was prepared as per the formal protocol of MZOC-GC of Diogene Company. Subjects were briefed about the procedure and were instructed to carefully listen and understand the questions. They were asked to answer either in yes/no only. Answers were recorded using Gold wave software v5.58 (2010) (freeware) on an HCL Desktop placed on a table along with the Philips SHM1500 PC VOIP microphone connected to the sound port of the desktop. Only the subject and the interviewer were present in the room. The subject was seated on a chair in a comfortable position, heard the questions and answered into the microphone. The interviewer asking the questions was at a distance, so that mostly the voice of the subject could be acquired by the mike. The interview was conducted in a closed noise free room with fans and air conditioners turned off to prevent any low frequency from being acquired by the mike.

In order to observe the shift in the acoustic measures, the questionnaire consisted of general, related and direct questions using a specific order in the interview. General questions were asked to obtain the baseline of the subject. Related and direct questions were intended to induce the stress. Each yes/no answer of the subjects were analyzed using PRAAT software. Table 2, gives the result of the interview of the female subject suspected of have stolen the mobile phone. The table labels the questions as general, related or direct and also lists the value of mean pitch, jitter, shimmer and formants (F1-F4) obtained using PRAAT software for the respective answers.

Obtained results indicate increase in the mean pitch for related and direct questions. However it was also observed that the percentage rise in mean pitch was maximum for direct question indicating significant stress. Formants F1-F4 were also obtained for all answers. Table 2 shows that F1 and F2 were remarkably low under stress (i.e. for answers to direct questions). Fig 2 shows the spectrogram of answers of suspect under normal condition and as well as under stress. Both the spectrograms are distinct and also indicate significant upshift in the pitch contour (shown by blue dots) under stress. Shift in acoustic measures of other subjects of second domain were also in the same direction as shown in Table 2, but it was noted that the percentage change in the acoustic measures of the female suspect was much higher than others. Due to the higher percentage change in acoustic measures under stress, this female was labeled as suspect. For the sake of comparison, spectrogram of another female subject of second domain was also shown (Fig 3). Fig 3(a) refers to her answer to the general question and Fig 3(b) refers to her answer to the direct question. Marginal increase in the mean pitch of this another female subject was observed while answering to direct question (Fig 3b). The probable reason for this marginal rise may be mild stress due to interrogation. However, shift in mean pitch of this female subject (Fig 3) was observed to be less than that of female suspect (Fig 2).

Table 2: Acoustic measures of the female suspect

Questions	Type of question	Answers	Acoustic measures						
			Mean Pitch (Hz)	Jitter (%)	Shimmer (%)	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
1. Is your name XYZ?	General	Yes	182.16	1.765	5.181	450.92	1532.53	2123.18	3288.08
2. Do your work at ABC?	General	Yes	181.548	1.162	4.056	396.12	1472.55	1995.14	3402.60
3. Are the lights on in the room?	General	Yes	187.951	1.273	5.226	393.12	1418.64	2057.06	3376.13
4. Did you see the mobile phone on the table?	Related	Yes	198.748	1.373	6.025	331.81	1574.33	2097.84	3383.85
5. Today is Monday?	General	Yes	183.020	1.834	7.114	401.75	1511.68	2058.19	3355.73
6. Do you speak lies?	Related	No	196.576	1.411	6.182	301.26	1394.88	2423.73	3301.25
7. Is the glass on the table filled with water?	General	Yes	189.022	1.584	7.506	470.87	1419.15	2137.82	3412.84
8. Have u stolen the mobile phone?	Direct	No	213.569	1.142	7.312	302.39	1287.28	2441.50	3310.76
9. Have you stolen things like this before?	Related	No	194.935	1.538	8.653	271.45	1484.76	2359.26	3276.02
10. Do you know where the mobile is right now?	Direct	No	199.454	1.620	7.981	309.60	1314.26	2367.57	3215.59

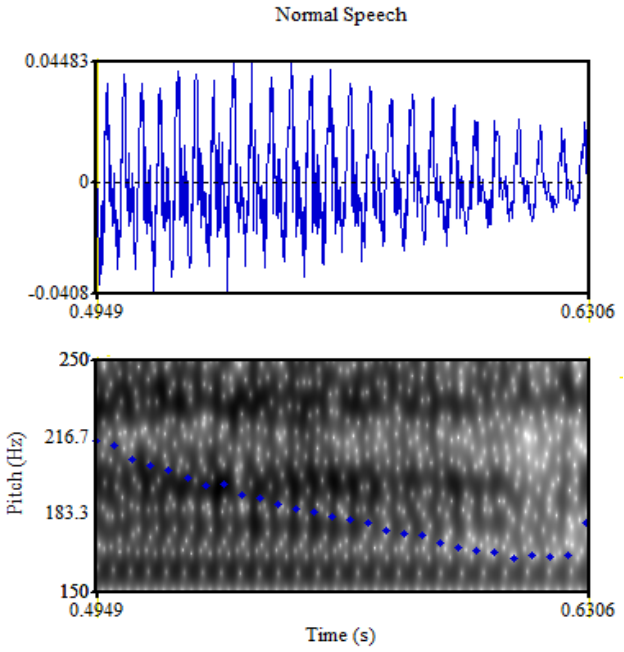

Fig. 2(a): Spectrogram of female suspect under normal condition

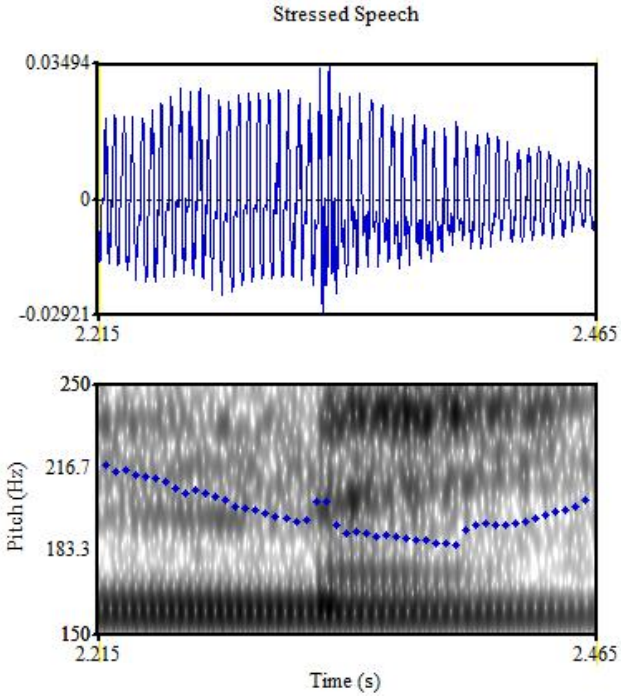

Fig. 2(b): Spectrogram of female suspect under stressed condition

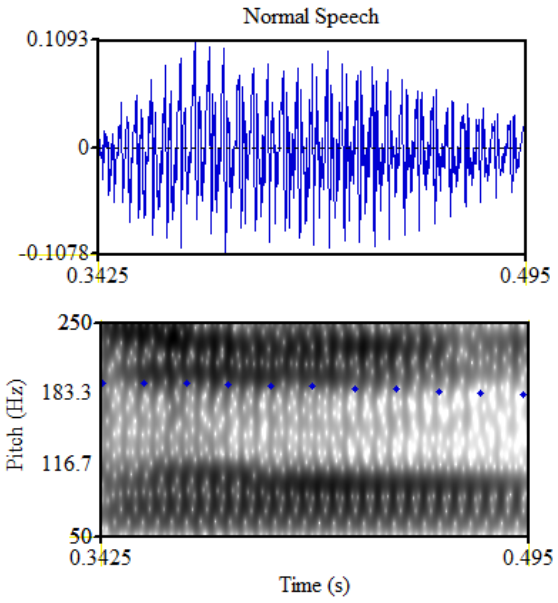

Fig. 3(a): Spectrogram of other female subject (domain-2) while answering to a general question.

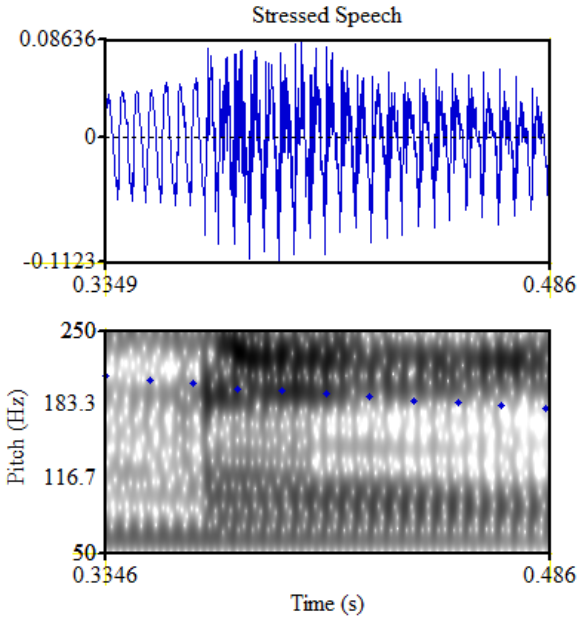

Fig. 3(b): Spectrogram of other female subject (domain-2) while answering to a direct question.

3. DISCUSSION

Stress is a person's response to an unexpected and unfavorable environmental condition or a stimulus. Ever since, Lippold, (1971) announced the presence of micro tremor in the muscles of human body, which attenuates in states of arousal, influence of psychological stress on the laryngeal muscles have been studied intensively [13]. Lot of commercial devices like voice stress analyzers, CVSA, Lantern instruments etc. also claim to detect deception by analyzing the shift in the micro tremors in the laryngeal muscles during stress. Muscle control during voice production could be influenced, if the speaker is under stress, however, it is still not certain, how and to what degree, this change could manifest itself into micro tremors. Previous studies have confirmed F_0 as a reliable indicator of human emotions like fear, happiness, anger, sorrow, fatigue etc. This paper attempts to detect the influence of emotions like anxiety and deception on more number of acoustic parameters. Therefore mean pitch (mean fundamental frequency), jitter, shimmer and first four formants of voice were compared under stress and normal (relaxed) condition using autocorrelation method. Darren, et al, (2002) suggested that autocorrelation method is more accurate and noise-resistant than CEPSTRUM method [7]. Result of domain-1 of this study indicated significant increase in mean pitch and decrease in formants F1 and F2 under anxiety/ embarrassment (domain-1). Answers of female suspect of stolen mobile case (domain-2), indicated substantial increase in mean pitch and significant decrease in the value of first two formants (F1 and F2) for all the related and direct questions. Other acoustic measures also indicated change under stress, however they failed to reveal any significance. Percentage change in the acoustic measures of female suspect were found to be higher than other subjects of second domain, indicating presence of significant stress.

Obtained results agree with the findings of (Streeter et al, 1977; Protopapas and Lieberman, 1997; Sigmund M, 2008, 2013; Mohanty M and Jena B, 2011) and confirm that mean pitch (F_0) and formants F1 and F2 are potential indicators of vocal stress [15, 17, 22, 23, 30]. (Williams and Stevens, 1972; Ling He, 2009) emphasized visual observation of spectrograms and qualitative reasoning in their work [14, 31]. They suggested that a person under stress may devoice and may not be precise in his articulation. It is also possible that he may slur syllables or omit certain speech sounds. Since these effects are difficult to quantify, they can be easily demonstrated with the help of spectrograms. Therefore last and most extensive analysis consisted of qualitative comparison of spectrograms under relaxed and stressed condition. Comparison yielded distinct spectral representation for normal as well as stressed speech. Stress classification and detection of deception has potential applications in law enforcement, employment screening and in the military field.

4. CONCLUSION

Mean pitch and formants F1 and F2 of human voice were obtained as reliable and non-invasive indicators of emotional stress. This study proposes a computer software based acoustic analysis of already recorded speech. Emotion analysis from real time voicing with simultaneous display of result can be carried out for further work.

5. REFERENCES

[1] Boersma P, Weenink D, PRAAT: doing phonetics by computer. (v5.3.56) 2010. Available from http://www.praat.org/ [Computer program]

[2] Bageshree, V., Pathak, S. and Panat, A.R. 2012. Extraction of Pitch and Formants and its Analysis to identify 3 different emotional states of a person. International Journal of Computer Science. Vol. 9, No. 4, pp. 296-299.

[3] Cummings, K.E., and Clements, M.A. 1995. Analysis of the glottal excitation of emotionally styled and stressed speech. Journal of Acoustical Society of America, Vol. 98, pp. 88–98

[4] Costantini, G., Iaderola, I., Paoloni, A. and Todisco, M. 2014. EMOVO corpus: an Italian emotional speech database, Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14). European Language Resources Association (ELRA), pp. 3501-3504.

[5] Cosetl, R.C., and Lopez, D.B. 2011. Voice Stress Detection: A method for stress analysis. Detecting fluctuations on Lippold Microtremor spectrum using FFT. 21st International Conference on Electrical Communications and Computers (CONIELECOMP) IEEE Xplore 2011; 184-189.

[6] Duffy, D. G. 2003. Advanced engineering mathematics with MATLAB, Boca Raton, Fla. Chapman & Hall/CRC, 2nd ed.

[7] Darren, H., Sharon, W., Roy, R., Megan, S. 2002. Investigation and Evaluation of Voice Stress Analysis Technology. The U.S. Department of Justice report (98-LB-VX-A013)

[8] Fielder, K., and Walka, I. 1993. Training lie detectors to use nonverbal cues instead of global heuristics. Human Communication Research, Vol. 20, No. 2, pp. 199-223.

[9] Hopkins, C., and Daniel, B. 2005. Evaluation of Voice Stress Analysis Technology. Proceedings of 38th Hawaii International Conference on System Science. IEEE. 2005.

[10] Heisse, J. W. 1976. Audio Stress Analysis—A Validation and Reliability Study of the Psychological Stress Evaluator (PSE). Proceedings of Carnahan Conference on Crime Countermeasures, Lexington, KY, pp. 5-18.

[11] Knapp, M. L., Hart, R. P., and Dennis, H. S. 1974. An exploration of deception as a communication construct. Human Communication Research. Vol. 1, pp.15-29.

[12] Knapp, M. L. and Comadena, M, E. 1979. Telling it like it isn't: A review of theory and research on deceptive communication. Human Communication Research. Vol. 5, pp. 270-285.

[13] Lippold, O. 1971. Physiological Tremor. Scientific American. Vol. 224, No.3, pp. 65-73.

[14] Ling He; Lech, M.; Maddage, M.C.; Allen, N., 2009. Stress Detection Using Speech Spectrograms and Sigma-pi Neuron Units, Natural Computation, 2009. ICNC '09. Fifth International Conference on, vol.2, no., pp.260-264.

[15] Mohanty, M.N. and Jena, B. 2011. Analysis of stressed human speech. Int. J. Computational Vision and Robotics. Vol. 2, No. 2, pp. 180–187.

[16] Mencattini, A., Martinelli, E., Costantini, G., Todisco, M., Basile, B., Bozzali, M., and Di Natale, C. 2014 'Speech emotion recognition using amplitude modulation parameters and a combined feature selection procedure', Elsevier, Knowledge-Based Systems, Vol. 63, pp. 68–81.

[17] Protopapas, A., and Liberman, P. 1997. Fundamental frequency of phonation and perceived emotional stress. Journal of Acoustical Society of America. Vol. 101, No. 4, pp. 2267– 2277.

[18] Ruiz, R., Legros, C., and Guell A. 1990. Voice Analysis to Predict the Psychological or Physical State of a Speaker. Aviation Space and Environmental Medicine. Vol. 61, No.3, pp. 266-71

[19] Ruiz, R., Absil, E., Harmegnies, B., and Legros, C. 1996. Time and spectrum-related variability's in stressed speech under laboratory and real conditions. Speech Communication. Vol. 20, pp. 111 - 129

[20] Smith, G. A. 1977. Voice analysis for the measurement of anxiety. British Journal of Medical Psychology. Vol. 50, pp. 367-73

[21] Sigmund, M. 2007. Spectral Analysis of Speech under Stress. ICSNS International Journal of Computer Science and Network Security. Vol. 7, No.4, pp. 170-72

[22] Sigmund, M., Prokes, A. and Brabec, Z. 2008. Statistical analysis of glottal pulses in speech under psychological stress. Proceedings of the 16th European Signal Processing Conference (EUSIPCO 2008), August 25-29.

[23] Sigmund, M. 2013. Statistical Analysis of Fundamental Frequency Based Features in Speech under Stress. Information Technology and Control. Vol. 42, No. 3, pp. 286-291.

[24] Salhan, A., Khan, M., Sondhi, S., and Vijay, R. 2012. Online offline voice stress analyzer. Aviation Space and Environmental Medicine. Vol. 83, No.3, pp. 309.

[25] Sondhi, S., Khan, M., Vijay, R., Salhan, A., and Vashisth, S. 2012. Real time speech analysis for detection of stress using Autocorrelation function. Proceeding of 11th International Conference on Information Technology and Telecommunication March 29- 30, 2012: 38 – 44 at Cork Institute of Technology, Cork, Ireland.

[26] Scherer, K.R. 2003. Vocal communication of emotion: a review of research paradigms. Speech Comm. Vol. 40, pp. 227–256.

[27] Scherer, K.R. 1986. Voice, Stress and Emotion: In: H. Appley and R. Trumbull, Editors. Dynamics of Stress: Physiological and Psychological Social Perspective. New York: Plenum Press. pp. 157-179.

[28] Scherer, K.R. 2000. The neuropsychology of emotion, chapter Psychological models of emotion. Oxford University Press, Oxford 2000. pp. 137–162.

[29] Scherer, K.R., Banse, R., and Wallbott, H.G. 2001. Emotion inferences from vocal expression correlate across languages and cultures. J. Crosscult. Psychol. Vol. 32, pp. 76–92

[30] Streeter, L.A., Krauss, R.M., Geller, V., Olson, C., and Apple, W. 1977 Pitch changes during attempted deception. Journal of Personality and Social Psychology. Vol. 35, No. 5, pp. 345–350.

[31] Williams, C.E., and Stevens, K.N. 1972. Emotions and Speech: Some Acoustical Correlates. J. Acoust. Soc. Amer. Vol. 52, pp. 1238-1250.

[32] Warren, J. and Riedel, R. 2004. Emotional Power: tapping the inexhaustible energy of your spirit. Malaysia, Axiom.

[33] Xiao, Y., Jitsuhiro, T., Miyajima, C., Kitaoka, N., and Takeda, K. 2012. Physical characteristics of vocal folds during speech under stress. Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference, 25-30 March 2012. pp. 4609-4612.

[34] Zhou, G., Hansen, J. H. L., and Kaiser, J. F. 2001. Nonlinear Feature based Classification of Speech under Stress. IEEE Trans. On Speech and Audio Processing, Vol. 3, pp. 201-206.