Voice Analysis for Detection of Deception

Savita Sondhi
Bina Nusantara Group, Jakarta
Indonesia
ssondhi@binus.edu

Ritu Vijay
Department of Electronics
Banasthali, Banasthali University
Rajasthan, India

Munna Khan
Electrical Engineering Department
Jamia Millia Islamia
New Delhi, India

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

***Abstract*—Human voice can signal the psychological state of an individual. Be it emotional, physical, intentional or unconscious, voice reveals the internal state of the speaker. Therefore, the main aim of the present study is to identify which voice parameter could be used as a reliable indicator of stress due to deception. For analyzing real life natural stress, the voice of crime suspects was recorded during ongoing interrogations at the police station. PRAAT software was used for analysis. Results indicated a significant increase in the mean fundamental frequency (F_0) and formant frequencies F1, F2. Marginal decrease in jitter was observed, however, shimmer did not at all correlate to stress due to deception. Pitch contours indicated considerable rise in the pitch of crime suspects while answering specific direct questions. Thus, it was concluded that mean F_0 and formant frequencies F1, F2 correlate strongly to stress.**

Keywords—Voice stress analysis; detection of deception; fundamental frequency; jitter; shimmer; spectrogram; pitch contour

I. INTRODUCTION

Emotions are strong innate components. They affect the psycho-physiological state of an individual in the form of physiological arousal, motor expression, and subjective feelings [1]. While subjective feelings refer to the internal condition of an individual under stress, physiological arousal and motor expression can be observed from facial expressions, body language, and voice [2]. Past studies have shown that physiological arousal accompanied with fear affects the respiration, phonation, and articulation of the speaker in a way that it produces stress specific acoustic patterns. The main aim of this study is to detect such stress specific patterns in the spectral and vocal characteristics of the speaker trying to deceive during a police interrogation. Deception can be in the form of exaggeration, equivocation, concealment, or an outright lie. Deception is defined as "a deliberate attempt to mislead others" [3], alternatively, as, "the deliberate attempt, whether successful or not, to conceal, fabricate, and/or manipulate in any other way factual and/or emotional information, by verbal and/or nonverbal means, in order to create or maintain in another or in others a belief that the communicator himself or herself considers false" [4] (p. 148). Ever since, [5] announced the presence of micro tremor in the muscles of human body, which attenuates in states of arousal, voice analysis has gained sufficient popularity to detect deception and stress. According to the Merck Manual, (1999) "enhanced physiologic tremor may be produced by anxiety, stress, fatigue, metabolic derangements or by certain drugs", [6]. Few studies by [7, 8] have also analyzed changes in the micro tremor frequency to detect stress in voice. Although muscle control during voice production could be influenced, if the speaker is under stress, it is still not certain, how and to what degree; this change could manifest itself into micro tremors.

Lot of commercial devices like voice stress analyzers, CVSA, Lantern instruments etc. have claimed to detect deception by analyzing shift in the micro tremors in the laryngeal muscles under the influence of stress. The best known and most commonly used deception detection technology today is the polygraph test. It is a scientific instrument that uses electrodes, blood pressure cuffs and pneumatic gauges to record cardiovascular activities, respiration rate and skin conductance. It is assumed that the physiological responses of the liars will be different to those displayed by truth-tellers. Although polygraph is accepted as a deception detection tool, there is considerable controversy as to the effectiveness of the polygraph which is amply documented in National Academies study, 2003 [9].

Main challenge in detection of deception is the design of an ethically approved and scientifically sound experiment. Due to ethical limitations, most of past research has been conducted in laboratories. However, all these laboratory studies lacked the real element of deception and fear. Recognizing the constraints of laboratory studies, researchers focused on collecting data from real life natural settings. Reference [10] presented a study wherein the participants were involved in a simulated crime and were then subjected to a police interview concerning the criminal offence.

Furthermore, due to ethical and practical constraints, few studies selected university students as their subjects as in [11, 12, 34]. Despite all these efforts, deception employed by lay persons and university students in mock or simulated conditions failed to resemble real-life deception of skilled and trained liars.

Literature documents few studies that have examined real-life police interviews with criminal suspects using behavior cues [13, 14] and of a murderer during his police interview using videotapes [15]. Although few, these real-life investigations have provided an understanding of the behavior

978-1-5090-5130-4/16/$31.00 ©2016 IEEE

of liars, who use deception by their own wish in high-stake real-life settings.

While most of the past studies on deception, conducted by psychologists, have focused on standard biometric indicators which are commonly measured in polygraph (cardiovascular, electrodermal, and respiratory) [28] and non-verbal clues like gestures, body movement, facial expressions, brain imaging, body odor, vocal behavior [3, 23, 35], it appears that literature on the acoustic correlates of deception by real criminals is limited. Therefore further research into this area would provide a valuable contribution.

In order to obtain a reliable, non-invasive and non-contact system which could differentiate between truth and deception, it is important to determine whether a consistent and reliable relationship exists between stress (while being deceptive) and the basic parameters of voice like fundamental frequency, formant characteristics and voice perturbation measures like jitter and shimmer.

II. MATERIALS AND METHOD

A. *Subject involvement and data collection*

To investigate whether a reliable relationship exists between the above mentioned acoustic parameters and deception, the inspector in-charge of the police station was requested permission to visit the police station and record the voice of crime suspects. Permission was granted to visit the police station for five days. During this span, the voice of twelve subjects (10 male, 2 female) was recorded. Out of these, one subject was a shop keeper who was being interrogated for running a racket of stolen mobile phones. Other subjects had come to report matters like snatch and run, theft at home, domestic violence and threatening calls being received. Normal conversation between an informer of police and the inspector in charge was also recorded. Full conversation between the policemen and the subjects was recorded using Samsung (S-II model) mobile phone in .AMR file format. The sampling rate was selected to be 44.1 kHz (16 bit). Subjects introduced themselves with basic details like name and address which served as the baseline for the study. In order to avoid any possibility of additional stress, voice was recorded without the knowledge of the subjects. While listening to their complaint, the police inspector questioned and also grilled them to gather more clues. Response of the subjects to such deliberate questions served as stressed voice samples. All audio clips were archived in .WAV format for further processing. In addition to these twelve subjects, pre-recorded audio clips of criminals involved in abduction, murder and robbery were also provided by the police station for acoustic analysis.

B. Data Analysis

As the recordings were done in real field conditions, audio files contained background noise. All background disturbances as well as voice of the police inspector was removed using Audacity 2.0.3 software [20]. Further analysis was performed using PRAAT software (version 5.3.56) [21]. Following the PRAAT software instructions, for the sake of analysis, the pitch range for female subjects was taken as 75–500 Hz, and for male subjects was taken as 50–300 Hz.

III. RESULTS

Result of the analysis is presented in the form of numerical data, graphs, spectrograms and pitch contours. The parameters selected for measurement were fundamental frequency of voice (F_0), variation in F_0 (i.e. F_0 SD), jitter, shimmer and formants F1 and F2. Table I shows the voice parameters (averaged across ten male subjects) under baseline and stressed state. Results of female subjects are shown in graphs below.

A. *Mean Fundamental Frequency F_0*

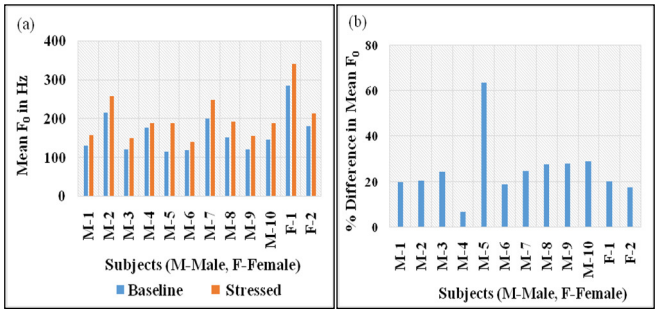

Fig. 1. (a) Mean F_0 in baseline and under stress and (b) Percentage difference in mean F_0

Fig. 1 (a) shows the graph between mean F_0 and the psychological state (i.e. baseline or under stress) of each subject while Fig. 1 (b) gives the % difference in mean F_0. To obtain the percentage difference the following calculation has been used:

$$\%\, difference = \left(\frac{F_0 \text{ under stress} - F_0 \text{ baseline}}{F_0 \text{ baseline}}\right) \times 100 \quad (1)$$

Obtained results indicate that deception resulted in increased levels of stress. These emotional and cognitive states experienced by the subjects during interrogation led to an increase in mean F_0. The increase ranged from 6.5% (11.5 Hz) to 63.3% (73.56 Hz) as shown in Fig. 1(a). For majority of the subject, the percentage change was around (20-40) %. In contrast, one subject demonstrated change less than 10%, and another showed change higher than 60%. Positive values in the '% difference' as shown in Fig. 1(b) indicate that mean F_0 measured during the stressed states are higher compared to the baseline condition.

TABLE I. VOICE PARAMETERS (AVERAGED ACROSS TEN MALE SUBJECTS) UNDER RELAXED AND STRESSED STATE. VALUES ARE MEAN ± SD

Voice parameters	Relaxed state	Stressed state
Mean F_0 (Hz)	154.05 ± 36.02	191.16 ± 38.48
Formant F1 (Hz)	492.07 ± 60.65	499.57 ± 102.21
Formant F2 (Hz)	1585.10 ± 277.82	1748.21 ± 277.72
Jitter (%)	1.46 ± 0.37	1.07 ± 0.33
Shimmer (%)	7.01 ± 1.91	6.86 ± 1.46

B. F_0 variability (F_0 SD)

In addition to measuring mean F_0, variation in F_0 was also measured. F_0 variability in the present study was expressed by the standard deviation of mean F_0 values.

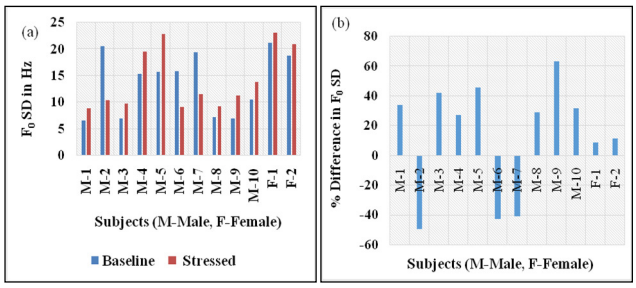

Fig. 2. (a) F_0 SD in baseline and under stress and (b) Percentage difference in F_0 SD

Nine out of twelve subjects indicated increase in F_0 SD as shown in Fig. 2(a). The magnitude of increase ranged from 8% (1.8 Hz) to 63 % (4.36 Hz) as shown in Fig. 2(b).

C. Formant Frequency F1 and F2

To obtain the values of formant frequencies F1 and F2, vowel /ã/ was analyzed. For the analysis, vowel /ã/ was extracted from the audio clips of each subject. The sound of vowel /ã/ is the same as in the word 'pass'. The vowel /ã/ was chosen as it is the basic sound with least modification along the path it takes from the vocal cords to the lips. Results indicate that formant F1 and F2 also increased under stress for majority of subjects as shown in Fig. 3 and 4. respectively. However, it was also observed that, three subjects indicated decrease in F1 and four subjects showed decrease in F2 values.

D. Voice perturbations (Jitter and shimmer)

Jitter and shimmer are perturbation measures of voice signals. For detailed acoustic measurements, jitter and shimmer were also considered because they describe the disturbances, or changes in the output of the voice. Jitter reveals the source of variation in the frequency whereas shimmer reveals the source of variation in amplitude of vocal fold vibrations. Jitter (local) and shimmer (local) measurements were obtained using built in function of PRAAT software.

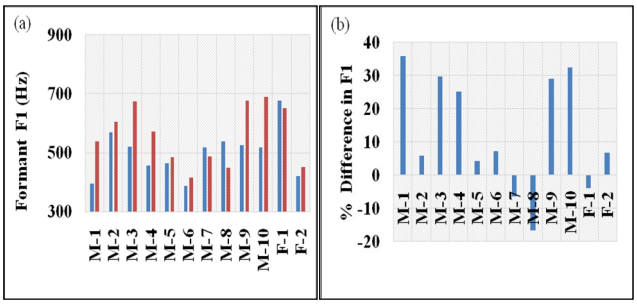

Fig. 3. (a) Formant F1 in baseline and under stress and (b) Percentage difference in F1

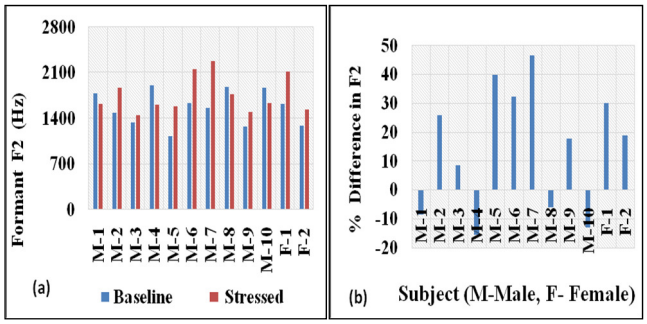

Fig. 4. (a) Formant F2 in baseline and under stress and (b) Percentage difference in F2

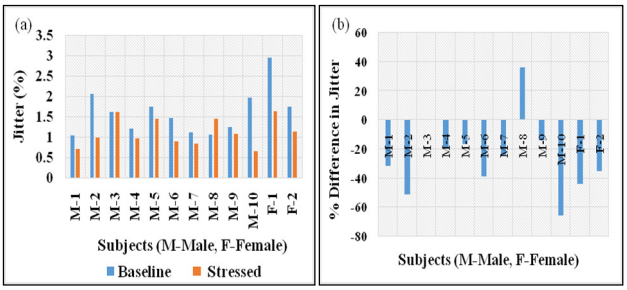

Fig. 5. (a) Jitter in baseline and under stress and (b) Percentage difference in Jitter

It had been reported that pathological threshold measures indicate jitter (local) to be 1.040% and shimmer (local) to be 3.810%. In the present study, out of the twelve subjects, ten subjects indicated decrease in the value of jitter under stress as shown in Fig. 5. Only one subject indicated an increase in jitter and one subject showed no change.

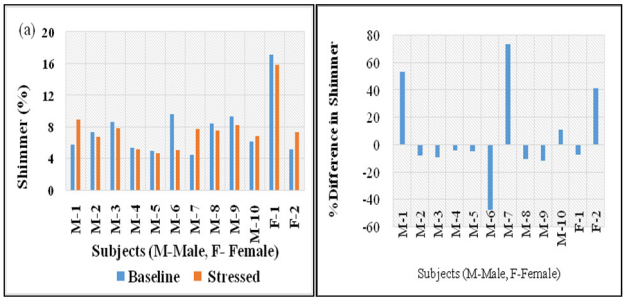

Fig. 6. (a) Shimmer in baseline and under stress and (b) Percentage difference in shimmer

Shimmer on the other hand did not show any significant pattern for psychological stress (Fig. 6a). The general trend was that of decrease in the shimmer values. While three subjects displayed increase in shimmer value (Fig. 6b). For sake of comparison, research into jitter and shimmer in connection with deceptive speech could not be located. Thus, obtained results indicate that among all the acoustic parameters analysed, mean F_0 correlates strongly with stress.

The following section presents the analysis of the interrogation conducted on criminals who were involved in abduction, murder, robbery. Since the audio was recorded

during the interrogation, the baseline of these accused was not available. Therefore for analysis only the voice of accused was extracted from the audio clips. The result is presented in the form of pitch contours and spectrograms.

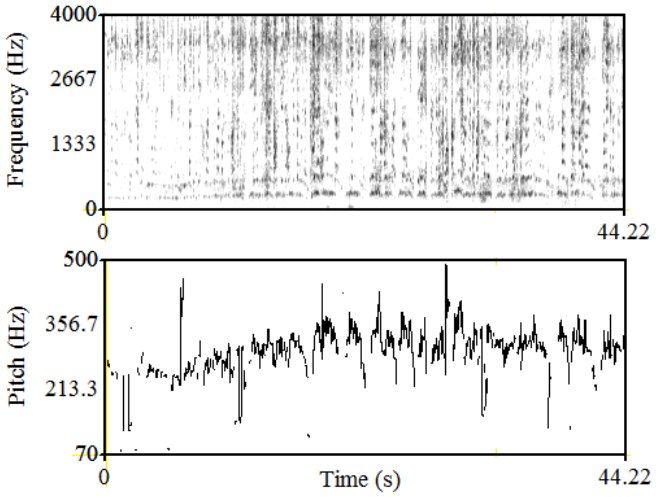

Fig. 7. Spectrogram and pitch contour of a lady showing significant stress during the interrogation.

Fig. 7 represents the spectrogram and pitch contour of a lady suspected of being involved in murder. She was initially relaxed but got stressed as the police inspector grilled her further during the interrogation. Increase in stress is shown by the rise in the pitch contour.

Fig. 8 shows the spectrogram and pitch contour of a man who was initially relaxed but the moment a specific question was asked, he got stressed which resulted in noteworthy rise in F_0. As he continued talking further, his F_0 returned back to the original range. Further investigation revealed that this subject was telling lies during the interrogation. He also accepted that he was deceitful at that moment which showed sudden rise in F_0. This confirmed that mean F_0 increases with deception.

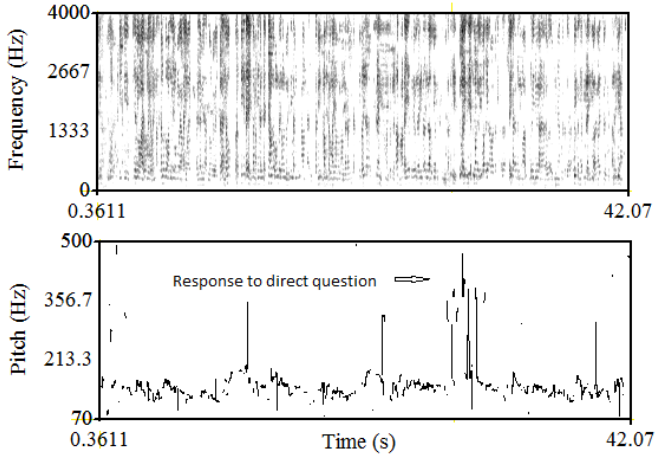

Fig. 8. Spectrogram and pitch contour of a man showing increased stress while a specific question was asked during interrogation.

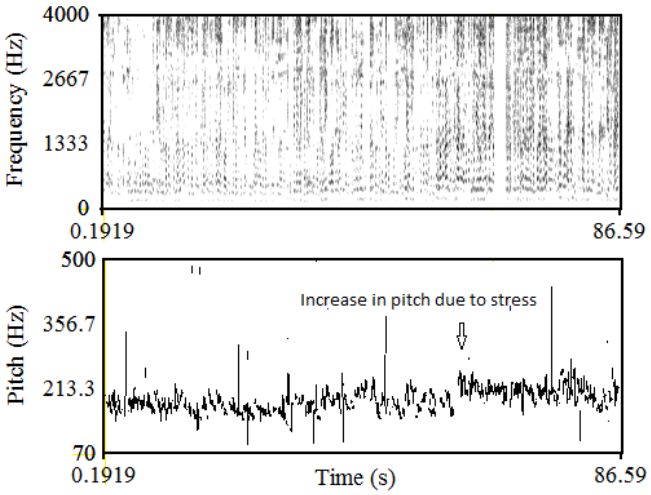

Fig. 9. Spectrogram and pitch contour of an accused showing a rise in pitch contour towards the end as he feels trapped for his transgression.

Fig. 9 shows the spectrogram and pitch contour of an accused involved in robbery and illegal activities. He was initially making stories to save himself. However, the moment policemen showed the proofs collected against him he got stressed. This part of the conversation where the policemen revealed the proof and he felt trapped can be seen as an upward shift in the pitch contour (marked). He finally accepted all his transgressions.

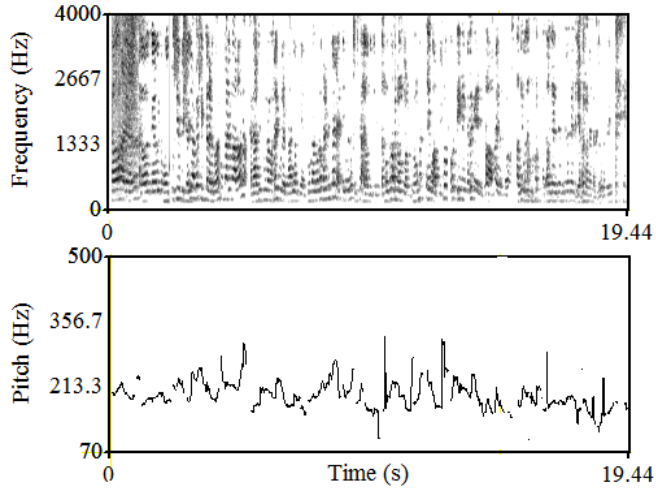

Fig. 10. Spectrogram and pitch contour of the accused under constant stress.

Fig. 10 shows the spectrogram and pitch contour of an accused involved in abduction. His pitch contour indicates constant stress during the interrogation.

Fig. 11 shows the spectrogram and pitch contour of an informer of police who was in neutral state as he conversed with the police inspector. As this subject was well acquainted with the inspector, his pitch contour revealed his relaxed state of mind.

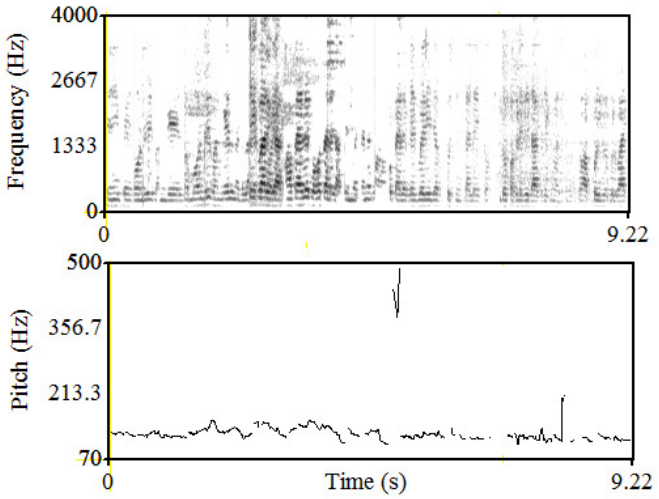

Fig. 11. Spectrogram and pitch contour of an informer of police who was in neutral state as he conversed with the police Inspector.

IV. DISCUSSION

The utility of voice analysis to identify the psycho-physiological stress during deception is investigated in this study. The primary objective was to identify which voice parameter/s is a reliable indicator of stress during deception. Obtained results indicated that deception provoked an increment in mean F_0. In a similar study, the author conducted a laboratory based mock-theft investigation in combination with 'security interview' to incite deception from among a total of ten male native British English speakers [22]. It was reported that mean F_0 increased from 0.9% to 9.3% in response to the direct questions intended to induce stress. In contrast, subjects in the present work were Hindi speakers and were being interrogated by Indian Police. They were under the influence of real stress which led to higher percentage increase in mean F_0 as compared to that reported by [22]. Increase in pitch was also reported to be a reliable clue associated with deception in [23-27]. Fundamental frequency (F_0) of subject responses was reported to be higher for deceptive than non-deceptive responses in [28].

Reference [29] had analyzed voice of pilots in an emergency situation and reported an increase in mean F_0 from 208 Hz to 432 Hz. Similarly, an increase of over 60Hz in mean F_0 under stressed state when compared with neutral state was also reported by [30]. However, in yet another study an increase in mean F_0 from 149 Hz to 264 Hz for one of the pilot subjects under stress was noted [31]. Percentage increase in mean F_0 reported by these studies is much higher than that obtained in the present work. The probable reason for this may be that these studies were based on catastrophic events such as aero plane crash or flight emergencies, which involved risk of life. While the subjects in the present case were general people who had come to file an FIR (first information report), deliberate questions asked by the police inspector to assess authenticity of their complaint would have added to their psychological stress. Still the common link between these studies and the present research work was the analysis of natural data. This confirms that percentage increase in mean F_0 for real natural stress is much higher than the laboratory induced stress. In addition to increase in the mean value of F_0, greater F_0 variability, expressed in terms of F_0 SD, was also reported by [22, 32].

Results of the present study also indicate increase in the formants frequencies F1 and F2 during deception. These results are also in confirmation with the results of [33, 34], however, the authors did not specify which vowel phonemes were analyzed. Literature describes that formant values depend a lot on the method of formant extraction. In the present study, formants had been extracted using built in function of PRAAT software. Obtained results indicate that the increase in F1 and F2 was not consistent across all subjects. The probable reason for this may be that, the recordings have been performed in real field conditions with background noise which may have affected the results. Jitter was found to decrease during deception however, shimmer did not show any significant pattern for psychological stress. As, literature into jitter and shimmer in connection with deceptive speech could not be located, further research on the parameters should be conducted to examine the reliability of these parameters.

V. CONCLUSION

Results indicate that mean F_0 can be used as a reliable and non-invasive indicator of stress due to deception. Formants frequencies F1 and F2 were found to increase under stress, however, the increase was not consistent across all subjects. Spectrograms were qualitatively distinct for relaxed and stressed states. Jitter was found to decrease with stress; however, shimmer did not reveal any significant impact of stress. This study proposes a non-invasive non-contact technique to detect psychological stress due to deception. No electrodes or wires are to be connected to the body of the subject, because such gadgets would just add to the stress. All that would be needed is a laptop with the software installed. This way the setup would be simplified and will also be financially economical.

REFERENCES

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

[2] C. Darwin, The expression of the emotions in man and animals, 3rd ed., Oxford University Press, Oxford 1872, reprint 1998.

[3] B.M. DePaulo, J.J. Lindsay, B.E. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper, "Cues to deception," Psychological Bulletin, vol. 129(1), pp. 74-112, 2003.

[4] J. Masip, E. Garrido, and C. Herrero, Defining deception. Anales de psicologia, vol. 20(1), pp. 147-171, 2004.

[5] O. Lippold, "Physiological Tremor," Scientific American, vol. 224(3), pp. 65-73, 1971.

[6] M.H. Beers and R. Berkow, The Merck Manual of Diagnosis and Therapy, 17th ed., John Wiley & Sons, New York, 1999.

[7] G.A. Smith, "Voice analysis for the measurement of anxiety," British Journal of Medical Psychology, vol. 50(4), pp. 367-373, December 1977.

[8] R.C. Cosetl and J.M.D.B. Lopez, "Voice Stress Detection: A method for stress analysis detecting fluctuations on Lippold microtremor spectrum using FFT," in Proc. of 21st International Conference on Electrical

Communications and Computers (CONIELECOMP), San Andres Cholula, 28th Feb-2nd March 2011, pp.184-189.

[9] National Research Council, The polygraph and Lie detection. Committee to review the scientific evidence on the Polygraph, Washington, DC: The National Academic Press, 2003.

[10] L.A. Stromwall, M. Hartwig, and P.A. Granhag, "To act truthfully: Nonverbal behavior and strategies during a police interrogation," Psychology, Crime & Law, vol. 12(2), pp. 207-219, 2006.

[11] R. Edelmann, "Non-verbal behaviour and deception," in D. Canter and L. Alison Eds. Interviewing and deception. Aldershot: Ashgate Dartmouth, pp. 157-182, 1998.

[12] L. Anolli and R. Ciceri, "The voice of deception: Vocal strategies of naive and able liars," Journal of Nonverbal Behavior, vol. 21(4), pp. 259-284, 1997.

[13] S. Mann, Suspects, lies and videotape: An investigation into telling and detecting lies in police/suspect interviews. Ph.D. Thesis, University of Portsmouth, 2001.

[14] M. Davis, K.A. Markus, S.B. Walters, N. Vorus, and B. Connors, "Behavioral cues to deception vs. topic incriminating potential in criminal confessions," Law and Human Behavior, vol. 29, pp. 683-704, December, 2005.

[15] A. Vrij and S. Mann, "Telling and detecting lies in a high-stake situation: the case of a convicted murderer," Applied Cognitive Psychology, vol. 15(2), pp. 187-203, Mar-Apr 2001.

[16] D. Haddad and R. Ratley, Investigation and evaluation of voice stress analysis technology, March 2002. http://www.ncjrs.org/pdffiles1/nij/193832.pdf

[17] H. Hollien and J. Harnsberger, "Voice stress analyser instrumentation evaluation," Technical report, Counterintelligence Field Activity, CIFA CONTRACT-FA 4814-04-0011, 2006.

[18] S. Sondhi, M. Khan, R. Vijay, A.K. Salhan, and S. Chouhan, "Acoustic analysis of speech under stress," Int. J. Bioinformatics Research and Applications, vol. 11 (5), pp. 417–432, 2015.

[19] S. Sondhi, M. Khan, R. Vijay, A.K. Salhan, "Vocal indicators of emotional stress," International Journal of Computer Applications, vol. 122(15), pp. 38-43, 2015.

[20] D. Mazzoni, Audacity(R): A Free, Cross-Platform Digital Audio Editor, (version 2.0.3), 2013. Available online at: http:// www: audacity.sourceforge.net.

[21] P. Boersma and D. Weenink, Praat: doing phonetics by computer (version 5.3.56), 2010. Available online at: http://www.praat.org/ [computer program].

[22] C. Kirchhubel, The acoustic and temporal characteristics of deceptive speech. Ph.D. Thesis, Department of Electronics, The University of York, 2013.

[23] M. Zuckerman, R. Koestner, and R. Driver, "Beliefs about cues associated with deception," Journal of Nonverbal Behavior, vol. 6(2), pp. 105-114, 1981.

[24] A. Vrij and G.R. Semin, "Lie experts' beliefs about nonverbal indicators of deception," Journal of Nonverbal Behavior, vol. 20(1), pp. 65-80, March 1996.

[25] D.E. Anderson, B.M. DePaulo, M.E. Ansfield, J.J. Tickle, and E. Green, "Beliefs about cues to deception: mindless stereotypes or untapped wisdom," Journal of Nonverbal Behavior, vol. 23(1), pp. 67-89, March 1999.

[26] M. Lakhani, and R. Taylor, "Beliefs about the cues to deception in high- and low-stake situations," Psychology, Crime & Law, vol. 9(4), pp. 357-368, December 2003.

[27] L.H. Colwell, H.A. Miller, R.S. Miller, and P.M. Lyons, "US Police officers' knowledge regarding behaviors indicative of deception: Implications for eradicating erroneous beliefs through training," Psychology, Crime & Law, vol. 12(5), pp. 489-503, 2006.

[28] L.A. Streeter, R.M. Krauss, V. Geller, C. Olson, and W. Apple, "Pitch changes during attempted deception," Journal of Personality and Social Psychology, vol. 35(5), pp. 345-350, 1977.

[29] C.E. Williams, and K.N. Stevens, "On determining the emotional state of pilots during flight: an exploratory study," Aerospace Medicine, vol. 40, pp. 1369-1372, 1969.

[30] R. Ruiz, E. Absil, B. Harmegnies, and C. Legros, "Time and spectrum-related variability's in stressed speech under laboratory and real conditions," Speech Communication, vol. 20(1-2), pp. 111-129, 1996. doi.org/10.1016/S0167- 6393(96)00048-9.

[31] M. Brenner, T. Shipp, E.T. Doherty, and P. Morrissey, "Voice measures of psychological stress-laboratory and field data," in I.R. Titze and R.C. Scherer Eds. Vocal fold physiology: biomechanics, acoustics and phonatory control. Denver: Denver Centre for Performing Arts, Dencer, Colorado, USA. 1983, pp. 239-248.

[32] J. Krause and L.D. Braida, "Acoustic properties of naturally produced clear speech at normal speaking rates," Journal of the Acoustical Society of America, vol. 115(1), pp. 362-378, January, 2004.

[33] P. Benson, "Analysis of the Acoustic Correlates of Stress from an Operational Aviation Emergency," in I. Trancoso and R. Moore Eds. ESCA-NATO Tutorial and Research Workshop on Speech under Stress, Lisbon, Sept 14-15, 1995, pp. 61-64.

[34] M. Sigmund, "Introducing the Database Exam Stress for Speech under Stress," in Proc. of 7th IEEE Nordic Signal Processing Symposium, (NORSIG 2006), Reykjavik, pp. 290-293, (7-9) June, 2006. doi.org/10.1109/NORSIG.2006.275258.

[35] S.L. Sporer, and B. Schwandt, "Para-verbal indicators of deception: A meta-analytic synthesis," Applied Cognitive Psychology, vol. 20(4), pp. 421-446,2006.