■ REVIEW ARTICLE

Voice Stress Analysis for Deception Detection

Shyam Narayan Singh[1], Priyanka Raj[2], Aditya Saini[3], Md Azhan Tariq[4], Vinny Sharma[5]

ABSTRACT

The Air Force Research Laboratory (AFRL) has been asked by the National Institute of Justice to investigate voice stress analysis (VSA) technology and evaluate its effectiveness for both military and law enforcement applications. This technology has been marked as commercially available in computer based form, and marketed as being capable of measuring stress and, in some systems, deception. This technology is reported to be easier to use, less invasive and less constrained in its operation than standard polygraph technology. This study has found the VSA technology can identify stress better than chance with performance approaching that of current polygraph systems. However, it is not a technology that is mature enough to be used in a court of law. We also found that experience and training improve the accuracy than less trained individuals. Lastly, we explored how this technology may become an effective interrogation tool, when combined with polygraph technology. This article contains information and results of a primary work done to show how the stress changes 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 the accused to prove their innocence in an offense or a crime.

Authors' Affiliations:
[1-3] Student
[5] Assistant Professor, School of Basic and Applied Sciences, Galgotias University, Greater Noida, Uttar Pradesh
[4] Student, School of Computing Science and Engineering, Galgotias University, Greater Noida, Uttar Pradesh

Corresponding Author:
Vinny Sharma
Assistant Professor, School of Basic and Applied Sciences, Galgotias University, Greater Noida, Uttar Pradesh

Email: vinnysharmashah @gmail.com

Keywords | VSA, Voice Stress Analysis, Polygraph, Deception Detection

INTRODUCTION

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. Deception Detection refers to the investigative practices used to determine a person's truthfulness and credibility. This is largely determined through the consideration of certain behavioural and physiological cues as well as larger contextual and situational information.

The use of voice stress analysis for deception detection is arguable, since it's easy for an individual to change his/her feelings. On a voice decision, however, the recent advancements in computing have made it possible for humans to watch and analyse their inner feelings simply by analysing them. VSA is accomplished by activity fluctuations within the physiological small tremor gift in speech. A small tremor could be a low amplitude oscillation of the reflex mechanism dominant the length and tension of a stretched muscle caused by the finite transmission delay between neurons to and from the target muscle. Small tremors are felt in each muscle within the body as well as the vocal cords

Objective

- The main idea behind the research was to understand the voice stress analysis for deception detection among criminals.
- To understand how stress in voice reflects the emotions of a criminal.
- Graphical representation of the emotions examined through the voice.
- To create a dataset of criminals so that it can be used in future to get the criminals behind the bars.

Review of Literature

Stress and its manifestation within the acoustic signal are the topic matter of the many studies in literature. Researchers have tried to see reliable indicators of stress by analysing sure variable parameters of speech like first harmonic (pitch), amplitude, concentration of spectral energy, period and a number of other others. In literature, analysis of stress is performed through analysis of some parameters of stress like first harmonic, pitch, vowel period and formants in recorded emotional speech, namely, analysing a speaker's speech once they ar below stress, fatigue, significant work load, environmental noise, sleep loss or expressing some feeling like happiness, anger or sorrow.

Voice stress analysis originated from the idea that once an individual is below stress, micro-muscle tremors (MMT) occur within the muscles that compose the vocal tract that are transmitted through the speech. VSA literature points to a descriptor because the physiological basis for the MMT. This paper describes "a slight oscillation at or so ten cycles per second" (i.e., physiological tremors) throughout the conventional contraction of the skeletal muscle. All muscles within the body, as well as the vocal cords, vibrate within the eight to twelve cps vary. It's these MMTs that the VSA vendors claim to be the only real supply of detective work, if a person is lying.

In moments of stress, particularly if an individual is exposed to hazard, the body prepares for fight or flight by increasing the readiness of its muscles to spring into action. This successively causes the muscle vibrations to extend. In step with the Merck Manual, "enhanced physical tremors is also made by anxiety, stress, fatigue, or metabolic derangements or by sure medicine. VSA systems claim to live these tremors transmitted through the speech.

METHOD & MATERIALS
---

200 voice samples were collected and analysed for our research and a database was created (DB00002). The voice samples were then analysed on the basis of frequency. Change in frequency transfers from the muscles in the vocal tract to the voice produced.

To better perceive the aspects of stress speech during a human, the Air Force Research Lab (AFRL) worked with Dr. John Hansen of the University of Colorado, to see if it's possible to acknowledge and classify stress in a person's voice. Dr. Hansen is a world renowned professional within the space of voice stress. The report was enclosed during this report, associate degree is connected as an Appendix C. He states "it's not impossible that below extreme levels of stress, that muscle management throughout the speaker are going to be affected, as well as muscles related to speech production". During this study, he used the Speech Under Simulated and Actual Stress (SUSAS) info. This info includes stress speech like angry, loud, European (speaking below shouting conditions), and worry stress. In his report, he reviewed literature that mentioned past speech below stress studies. He analyzed stress in speech, within which he complete that voice stress is caused by factors that introduce variability into the vocalization method. These variabilities or options embody period, vocal organ supply factors, pitch distribution, spectral structure and intensity.

The workflow includes four areas:
1. Voice i.e., Overall word period,
2. Analysis: Individual speech category (vowel, consonant, semivowel, and diphthong) period,
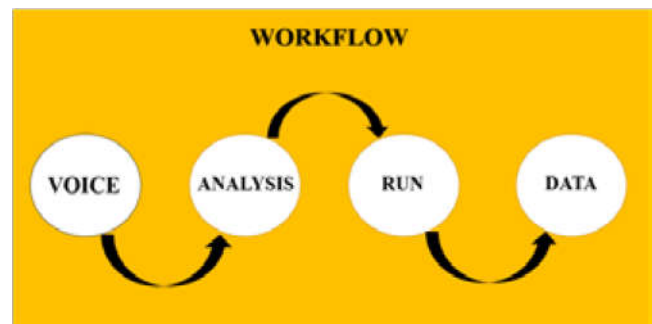
3. Data running and Period shifts between categories
4. Data acquisition and Speech category period ratios

Glottal supply factors measured the spectral slope of these vowels, that were longer than five frames or ninety-six-time unit. the primary and second formant's locations ar measured to see the spectral structure. Intensity could be a calculation of energy during a voice signal. These variabilities might even be speaker-dependent. By victimization these numerous linear and nonlinear options, and testing with the Bayesian hypothesis technique, it had been complete that differing kinds of emotional stress may well be classified. The Bayesian hypothesis technique could be a stress detection technique to see if a given piece of audio information is either neutral speech or an explicit classification of stress speech. From the results, it suggests that it's unlikely that one feature may well be wont to accurately sight deceptive stressed speech. There are lot of options that are united along, the strain kind recognition improves. It conjointly shows that some options, single two-handed, will sight a particular style of stress higher than alternative options. For associate degree example, the pitch feature might sight loud stress higher than angry and European stress.

Whereas, the spectral structure feature may find angry stress higher. Classification of deceptive stress wasn't tested because of the inaccessibility of a deceptive information. the gathering of a deceptive information may be a recommendation of future work. The tests were performed, the information was documented, and therefore the results were compared. The Vericator and Diogenes lamp Systems were utilised during this analysis and their technology was tested. the first goal of this part of the VSA analysis was to work out if the microtremor claim is that the VSA's true theory of operation. For the needs of this check the character of the results, stress or no stress indicated, weren't taken into consideration. The results were found to be consistent across the board with very little variation within the leads to response to the adjustments/changes within the modulation or depth of modulation rates. for instance, the analysis of the 80Hz FM check wave, with a depth of modulation rate of one Hz and a modulation rate of one Hz, differed little or no from associate degree eighty Hz FM check wave with a depth of modulation of 4Hz and a modulation rate of twenty-five Hz.

```
['JE_h09.wav', 'KL_f12.wav', 'DC_h03.wav', 'DC_d04.wav', 'KL_a14.wav']
```

```
male_neutral      120
male_sad           60
male_happy         60
male_disgust       60
male_surprise      60
male_fear          60
male_angry         60
Name: labels, dtype: int64
```

```
male_neutral      144
female_neutral    144
male_happy         96
female_fear        96
female_sad         96
male_disgust       96
female_disgust     96
male_fear          96
female_angry       96
female_happy       96
male_angry         96
male_sad           96
female_surprise    96
male_surprise      96
Name: labels, dtype: int64
```
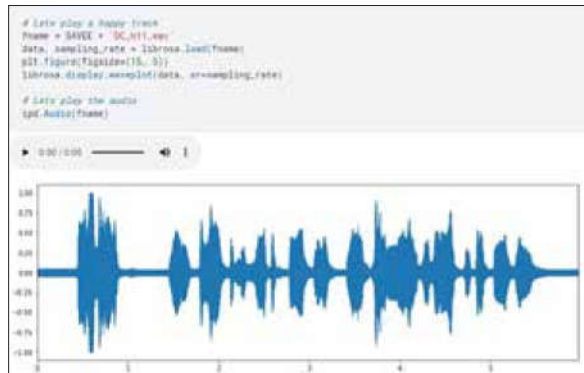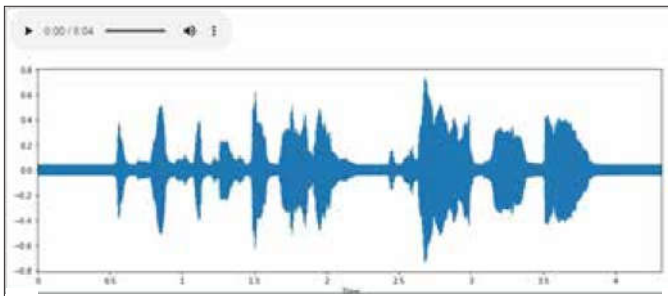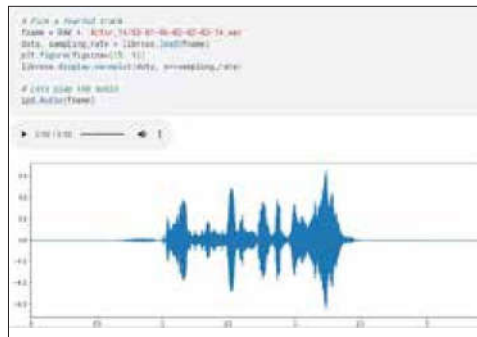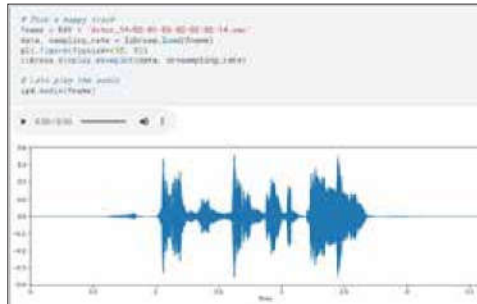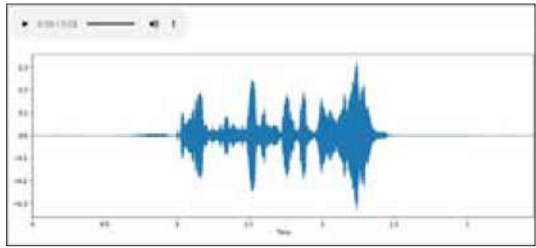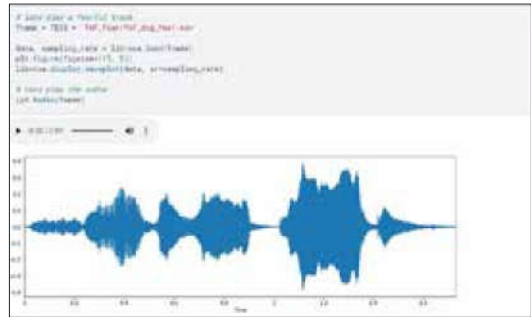
```
dir_list = os.listdir(TESS)
dir_list.sort()
dir_list
```

```
['OAF_Fear',
 'OAF_Pleasant_surprise',
 'OAF_Sad',
 'OAF_angry',
 'OAF_disgust',
 'OAF_happy',
 'OAF_neutral',
 'YAF_angry',
 'YAF_disgust',
 'YAF_fear',
 'YAF_happy',
 'YAF_neutral',
 'YAF_pleasant_surprised',
 'YAF_sad']
```

```
female_sad        400
female_disgust    400
female_happy      400
female_fear       400
female_neutral    400
female_angry      400
female_surprise   400
Name: labels, dtype: int64
```

Since there was no variation of indicated stress from totally different input signals, it may be assumed that the systems tested don't use microtremors as indicated in their claims. It was absolutely determined late within the testing part of this project, that the Diogenes lamp System live the energy amendment of the spectrum envelope between 20Hz and 40Hz. This can be what the Diogenes lamp System claims to be microtremors. It's the amendment of energy within the speech envelope. If a private is below stress, their vocal tract muscles square measure possible to constrain. Once the vocal tract muscles constrain the energy of the voice, signal becomes abrupt once the individual starts and finishes talking. Throughout the time a private talk, there's less variation of energy at intervals this the 20Hz band pass. Once a private isn't stressed, their voice energy slowly results in a peak after they begin to talk, then the energy varies till the individual stops speaking wherever the energy slowly trails off. This formula was coded within the laboratory with a similar audio signal input.
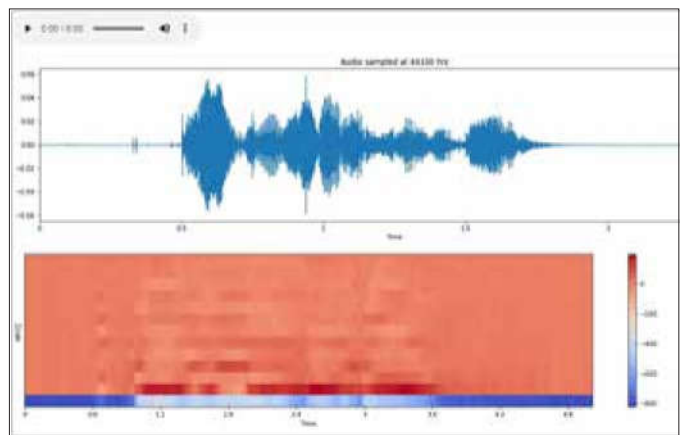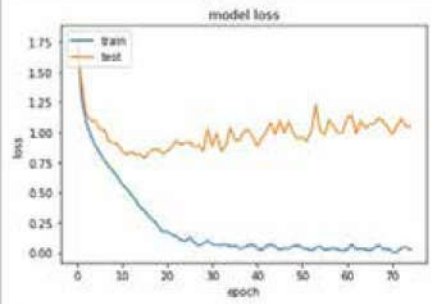

Source: All pictures are author self.

```
#concatinating the feature column into the complete dataframe
df = pd.concat([ref,pd.DataFrame(df['feature'].values.tolist())],axis=1)
df[:5]
```

	labels	source	path	0	1	2	3	4	5	6	...	206	207	208	209	210	211	212	213	214
0	male_happy	SAVEE	/kaggle/input/surrey-audiovisual-expressed-emo...	-25.742002	-26.182264	-25.468557	-25.198936	-25.429790	-23.347939	-16.800349	...	-3.119431	-4.260779	-5.274271	-5.775263	-7.272358	-7.340224	-7.019254	-8.643790	-15.420984
1	male_fear	SAVEE	/kaggle/input/surrey-audiovisual-expressed-emo...	-41.184326	-38.827896	-39.008781	-41.509396	-39.770164	-34.743958	-30.551401	...	-27.156342	-25.647068	-25.140005	-26.746456	-27.039560	-27.250130	-26.599070	-25.778673	-24.685337
2	male_happy	SAVEE	/kaggle/input/surrey-audiovisual-expressed-emo...	-25.528027	-23.866304	-22.509321	-22.320293	-17.513348	-12.073632	-9.537952	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	male_disgust	SAVEE	/kaggle/input/surrey-audiovisual-expressed-emo...	-24.961075	-23.702238	-24.368538	-23.824282	-20.085262	-20.491526	-21.797201	...	-11.261946	-12.151450	-13.157601	-13.586168	-14.427001	-15.093400	-13.683898	-13.934835	-9.288823
4	male_angry	SAVEE	/kaggle/input/surrey-audiovisual-expressed-emo...	-16.746674	-18.525646	-22.264179	-22.016902	-23.868864	-23.066270	-23.807617	...	-27.922167	-27.525471	-27.459106	-26.916111	-27.472073	-29.056606	-28.666271	-28.376913	-24.786333

```
plt.plot(model_history.history['loss'])
plt.plot(model_history.history['val_loss'])
plt.title('model loss')
plt.ylabel('loss')
plt.xlabel('epoch')
plt.legend(['train', 'test'], loc='upper left')
plt.show()
```
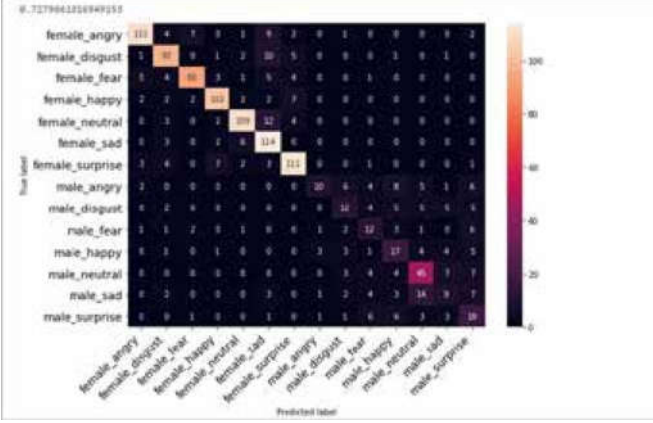


```
finaldf.to_csv('Predictions.csv', index=False)
finaldf.groupby('predictedvalues').count()
```

	actualvalues
predictedvalues	
female_angry	124
female_disgust	123
female_fear	105
female_happy	118
female_neutral	124
female_sad	159
female_surprise	133
male_angry	16
male_disgust	30
male_fear	37
male_happy	47
male_neutral	77
male_sad	30
male_surprise	57

```
classes = finaldf.actualvalues.unique()
classes.sort()
print(classification_report(finaldf.actualvalues,
```

```
                 precision    recall  f1-score   support

   female_angry       0.89      0.81      0.85       136
 female_disgust       0.79      0.82      0.80       118
    female_fear       0.89      0.80      0.84       116
   female_happy       0.86      0.86      0.86       119
 female_neutral       0.88      0.84      0.86       130
     female_sad       0.72      0.91      0.80       125
female_surprise       0.83      0.84      0.84       132
     male_angry       0.62      0.24      0.34        42
   male_disgust       0.40      0.32      0.35        38
      male_fear       0.32      0.40      0.36        30
     male_happy       0.36      0.44      0.40        39
   male_neutral       0.58      0.64      0.61        70
       male_sad       0.30      0.20      0.24        45
  male_surprise       0.32      0.45      0.37        40

       accuracy                           0.73      1180
      macro avg       0.63      0.61      0.61      1180
   weighted avg       0.73      0.73      0.72      1180
```


RESULT & DISCUSSION

The stress ground truth was obtained through the medical instrument examiner and court proceedings via the outcomes of every of the interviews each suspect confessed and were afterward guilty of murder. All of the relevant

stress sentences were verified. Every of the forty-eight utterances was analysed and compared to the bottom truth. Every system gave indications of high levels of stress wherever stress indicators were verified. The Vericator system scored 100 percent in its indication of some kind of stress, wherever because it displayed deceitful, high stress, or in all probability lying. The lamp system additionally scored 100 percent in its indication of stress through the wave shape analysis each system gave the examiner a conclusive indication of relevant stress.

Voiced analysis report table shows consistent results utilizing DAT and live voice. Every auditory communication was examined and located that everyone the waveforms and analysis was systematically identical. Once employing a recorder similar results were obtained as within the live information. Once recording with a electronic equipment, care has to be taken once adjusting the automated gain management (AGC). If the recording volume isn't set accurately, the input voice signal gets clipped. Thus once the output wave shape is processed, it gets distorted too. This might end in associate degree analysis that is totally different from the reality, so providing and incorrect result by the examiner. These discrepancies may be seen in figures, once exploitation the Diogenes lamp system. Once exploitation of the Vericator these discrepancies are evident. The Vericator results reacted otherwise everytime a similar clipped information was inputted into the system. For instance, if a clipped audio phase was processed within the Vericator, the system might show truth, whereas once more that very same clipped information would cause the system to show untruth.

CONCLUSION

The research was a success with an accuracy of 99.01%, we can say VSA can be done on an individual. For Deception Detection 200 samples were collected for this research and achieved successfully. Database can be created from Voice samples of Criminals which can be used in future for the purpose of identification of criminals.

It was not the target of this study to suggest one Psychophysiological Deception Detection technology was better than the other. Rather, it was our intent to produce unbiased analysis of VSA technology in conjunction with enough data to help them in creating selections on what kind of system to use. These instruments aren't "lie detectors". The choice on whether or not an issue is being truthful or lying ought to solely be created by a trained examiner. This call ought to be primarily based upon reviewing the information conferred by the instrument, the behavior of the topic, and different proof from the case. VSA systems square measure capable of providing associate degree examiner with a wave shape or different response which will be an affordable reflection of the strain level being toughened by the topic, in an exceedingly majority of the cases. The proper interpretation of this indicator is that the responsibility of the examiner. The goal in employing a VSA system or medical instrument ought to be to win over the topic that they cannot deceive the operator, which the instrument can find their deception and their best avenue is to confess to the crime. This study has shown that VSA systems can manufacture results that trained operators will use confidently to get confessions.

Finally, we tend to conclude that VSA technology will determine stress higher than medical instrument systems that have and coaching improves the accuracy of result. **IJFMP**

REFERENCES

1. Chen Y. Cepstral domain talker stress compensation for robust speech recognition. IEEE Trans Acoust 1988;36:433–9.
2. Cummings K, Clements M. Analysis of glottal waveforms across stress styles. Proceedings of the IEEE, International Conference on Acoustics, Speech, and Signal Processing. New York, NY: Institute of Electrical and Electronics Engineers, 1990;1:369–72.
3. Cummings K, Clements M. Analysis of glottal excitation of emotionally styled and stressed speech. J Acoust Soc Am 1995;98:88–98.
4. Frick RW. The prosodic expression of anger: differentiating threat and frustration. Aggress Behav 1986;12:121–8.
5. Hicks JW, Hollien H. The reflection of stress in voice-1: understanding the basic correlates. Proceedings of the 1981 Carnahan Conference on Crime Countermeasures. Lexington, KY: ORES Publications, 1981;189– 94.
6. Hollien H. Vocal indicators of psychological stress. In: Wright F, Bahn C, Rieber RW, editors. Forensic psychology and psychiatry. New York: NY: Academy of Sciences, 1980;47–72.
7. Hollien H. Acoustics of crime. New York, NY: Plenum Press, 1990. 8. Hollien H, Saletto JA, Miller SK. Psychological stress in voice: new approach. Studia Phonet Posnan 1993;4:5–17.
8. Lazarus RS. From psychological stress to the emotions—a history of changing outlooks. Ann Rev Psychol 1993;44:1–21.
9. Scherer KR. Vocal indicators of stress. In: Darby J, editor. Speech evaluation in psychiatry. New York, NY: Grune and Stratton, 1981;171– 87.
10. Scherer KR. Voice, stress and emotion. In: Appley H, Trumbull R, editors. Dynamics of stress: physiological, psychological, and social perspectives. New York, NY: Plenum Press, 1986;157–79.
11. Siegman AW, Boyle S. Voices of fear and anxiety and sadness and depression: the effects of speech rate and loudness on fear and anxiety and sadness and depression. J Abnorm Psychol 1993;102:430–7.
12. Williams CE, Stevens KN. Emotions and speech: some acoustical correlates. J Acoust Soc Am 1972;2:1238–50.
13. Williams CE, Stevens KN. Vocal correlates of emotional states. In: Darby J, editor. Speech evaluation in psychiatry. New York, NY: Grune and Stratton, 1981;221–40.
14. Anolli L, Ciceri R. The voice of deception: vocal strategies of nave and able liars. J Nonverbal Behav 1997;21:259–84.
15. Brenner M, Shipp T, Doherty ET, Morrisey P. Voice measures of psychological stress—laboratory field data. In: Titze IR, Scherer RC, editor. Vocal fold physiology. Denver, CO: Denver Center for the Performing Arts, 1983;239–48.
16. DePaulo BM, Lindsay JJ, Malone BE, Muhlenbruck L, Charlton K, Cooper H. Cues to deception. Psychol Bull 2003;29:74–118.
17. Doherty ET. Speech analysis techniques for detecting stress. Proceedings of the Human Factors Society. Santa Monica, CA: Human Factors & Ergonomics Society, 1991;1:689–93.
18. Hicks JW. An acoustical/temporal analysis of emotional stress on speech. Doctoral dissertation. Gainesville, FL: University of Florida, 1979.
19. Hollien H. Acoustical analysis of psychological stress. In: Lawrence V, editor. Tenth symposium care of the professional voice. New York, NY: The Voice Foundation, 1981;145–58.
20. Kuroda I, Fujiware O, Okamur N, Utsukli N. Method for determining pilot stress through analysis of voice communication. Aviat Space Environ Med 1976;47:528–33.
21. McGlone RE, Hollien H. Partial analysis of the acoustic signal of stressed and unstressed speech. Proceedings of the 1976 Carnahan Conference on Crime Countermeasures. Lexington, KY: ORES Publications, 1976;83–6.
22. McGlone RE, Petrie C, Frye J. Acoustic analysis of low-risk lies. J Acoust Soc Am 1974;55:S20(A).
23. Nilsonne A, Sundberg J. Differences in ability of musicians and nonmusicians to judge emotional state from the fundamental frequency of voice samples. Music Percept 1985;2:507–16.
24. Rosenfield JP, Soskins M, Bosh G, Ryan A. Simple, effective countermeasures to P300-based tests of the detection of concealed information. Psychophysiology 2004;41:205–19.
25. Scherer KR, Banse R, Walbott HG, Goldbeck T. Vocal cues in emotion encoding and decoding. Motiv Emot 1991;15:123–48.
26. Streeter LA, Krauss FM, Geller V, Olson C, Apple W. Pitch changes during attempted deception. J Pers Soc Psychol 1977;35:345–50.
27. Tolkmitt FJ, Scherer KR. Effects of experimentally induced stress on vocal patterns. J Exp Psychol 1986;12:302–13.
28. Vrij A, Edward K, Roberts KP, Bull R. Detecting deceit via analysis of verbal and nonverbal behavior. J Nonverbal Behav 2000;24:239– 63.
29. Vrij A, Edward K, Roberts KP, Bull R. Stereotypical verbal and nonverbal responses while deceiving others. Pers Soc Psychol Bull 2001; 27:899–909.
30. Cestaro VL. A comparison of accuracy rates between detection of deception examinations using the polygraph and the computer voice stress analyzer in a mock crime scenario. Ft. McClellan, AL: US Department of Defense Polygraph Institute, 1996. Report no.: DoDPI95-R-0004.
31. Haddad D, Walter S, Ratley R, Smith M. Investigation and evaluation of voice stress analysis technology. Rome AFB: US Dept Justice Report, 2002. Grant 98-LB-VX-A103.
32. Heisse JW. Audio stress analysis—a validation and reliability study of the psychological stress evaluator (PSE). Proceedings of the 1976 Carnahan Conference on Crime Countermeasures. Lexington, KY: ORES Publications, 1976;5–18.
33. Hollien H, Geison LL, Hicks JW Jr. Voice stress evaluators and lie detection. J Forensic Sci 1987;32:405–18.
34. Horvath F. The effects of differential motivation on detection of deception with the psychological stress evaluator and the galvanic skin test response. Appl Psychol 1979;64:323–30.
35. Kubis J. Comparison of voice analysis and polygraph as lie detection procedures. Aberdeen Proving Ground, MD: US Army Land Warfare Laboratory, 1973. Technical Report LWL-CR-U3B70.
36. Meyerhoff JL, Saviolakis GA, Koenig ML, Yourick DL. Physiological and biochemical measures of stress compared to voice stress analysis using the Computer Voice Stress Analyzer (CVSA). Fort McClellan, AL: Department of Defense Polygraph Institute, 2001. Report No. DoDPI01-R-0001.
37. Brenner M, Branscomb HH. The psychological stress evaluator, technical limitations affecting lie detection. Polygraph 1979;8:127–32.
38. Brockway BF, Plummer OB, Lowe BM. The effects of two types of nursing reassurance upon patient vocal stress levels as measured by a new tool, the PSE. Nurs Res 1976;25:440–6.
39. McGlone RE. Tests of the psychological stress evaluator (PSE) as a lie and stress detector. Proceedings of the 1975 Carnahan Conference on Crime Countermeasures. Lexington, KY: ORES Publications, 1975;83– 6.
40. J. H. I. Hansen, et. al., Methods for Voice Stress Analysis and Classification, as appendix to Investigation and Evaluation of Voice Stress Analysis Technology, Final Report for National Institute of Justice, Interagency Agreement 98-LB-R-013. Washington, DC, 2002. NCJRS, NCJ 193832.