Acoustic Correlates of Deceptive Speech – An Exploratory Study

David M. Howard and Christin Kirchhübel

Audio Laboratory, Department of Electronics, University of York, UK
{dh,ck531}@ohm.york.ac.uk

Abstract. The current work sets out to enhance our knowledge of changes or lack of changes in the speech signal when people are being deceptive. In particular, the study attempted to investigate the appropriateness of using speech cues in detecting deception. Truthful, deceptive and control speech was elicited from five speakers during an interview setting. The data was subjected to acoustic analysis and results are presented on a range of speech parameters including fundamental frequency (f0), overall intensity and mean vowel formants F1, F2 and F3. A significant correlation could not be established for any of the acoustic features examined. Directions for future work are highlighted.

Keywords: Deception, speech, acoustic, Voice Stress Analyzer.

1 Introduction

It is acknowledged that information can be gained about a human speaker from the speech signal alone. Possible knowledge that can be derived include a speaker's age, regional and social background, the presence of speech or voice based pathology, voice/language disguise, speaking style, and influence of alcohol intoxication.

The voice can also give information about a speaker's affective state. Listening to a third party conversation, lay listeners are usually able to tell whether the speakers are happy, sad, angry or bored. Whilst at an interpersonal level it is possible to perceive accurately emotional states, empirical research has not been successful in identifying the speech characteristics that distinguish the different emotions. Compared to the investigation of speech and emotion, research into psychological stress has been somewhat more successful in establishing the acoustic and phonetic changes involved [1]. However, facing similar methodological and conceptual obstacles, it is more appropriate to refer to the correlations that have been discovered so far as 'acoustic tendencies' rather than 'reliable acoustic indicators'.

If it is possible to deduce a speaker's emotional condition from listening to their voice, could it also be viable to make judgements about their sincerity from speech as well? For centuries people have been interested and fascinated by the phenomenon of deception and its detection. Indeed, a device that would reliably and consistently differentiate between truths and lies would be of great significance to law enforcement- and security agencies [2]. In recent times, claims have been brought forward involving voice stress analysers (VSA) which are said to measure speaker

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2011, LNAI 6781, pp. 28–37, 2011.
© Springer-Verlag Berlin Heidelberg 2011

veracity based on the speech signal. Scientific reliability testing of these products has mainly resulted in negative evaluations [3,4,5]. While testing of these products is a necessary part of their evaluation, it is believed that a more fundamental step has been overlooked. Prior to examining the reliability of a test it should be ascertained whether the assumptions on which the test is based are valid [4]. In other words, it needs to be established whether a relationship exists between deception, truth and speech, and if so, what the nature of this relationship is.

Surprisingly, very little research has been carried out on the acoustic and phonetic characteristics of deceptive speech. There are a number of studies [6,7,8] that have analysed temporal features such as speaking rate, pauses, hesitations and speech errors but only a few studies have investigated frequency based parameters as, for instance, mean f_0 and f_0 variability [9,10]. Evidence for the analysis of vowel formants and voice quality in connection with deceptive speech is rare. Recently completed work by Enos [11] is one of the first attempts to analyse deceptive speech using spoken language processing techniques. This work provides a basis in this complex area but more research is needed within this subject matter in order to improve our understanding of deceptive speech and consequently, to assess whether differentiating truthfulness and deception from speech is a realistic and reasonable aspiration.

2 Method

2.1 Participants

The data consists of an opportunistic sample of five male native British English speakers between the ages of 20 to 30 years (mean age = 24.7 years; SD = 3.65 years). The majority were drawn from the student population at the University of York and were from the northern part of England. None of the speakers had any voice, speech or hearing disorders.

2.2 Experiment Procedure

The procedure was modified from Porter [12] and is based on a mock-theft paradigm. The experiment was advertised as being part of a security research project and participants received £5 for participating with the chance of earning more money through the trial. Having arrived at the experimental setting participants were told that the University was looking to implement a new security campaign in order to reduce small scale criminal activity (e.g. theft on campus). The security scheme would involve employing non-uniformed security wardens who:

a) Patrol in selected buildings
b) Perform spot checks on people
c) Interview students suspected of having been involved in a transgression

The participants were then led to believe that the researchers were testing the effectiveness of this security system and, in particular, the extent that wardens would

be able to differentiate between guilty and non-guilty suspects. Further to this the volunteers were informed that the experiment was also part of a communication investigation and therefore audio data would be collected. Having given written consent that they are prepared to continue with the study participants had to complete three tasks:

Task 1: sitting in a quiet office room ('preparation room'), participants were asked to complete demographic details. On completion of the forms they were taken to an interview room where they were involved in a brief conversation which formed the baseline/control data.

Task 2: participants were provided with a key and directions. They were asked to go into an office and take a £10 note out of a wallet located in a desk drawer and hide it on their body. They were advised to be careful in order not to raise suspicion or to draw attention of the security warden who was said to be in the building and who might perform a spot check.

Task 3: a security interview was conducted in which the mock security warden questioned the participant about two thefts that had allegedly occurred in the previous hour. The participant committed one of the thefts (theft of £10 note) but not the other (theft of digital camera from the 'preparation room'). Participants were required to convince the interviewer that they were not guilty of either theft. With respect to the camera theft, participants could tell the truth but when the interviewer asked about the £10 note the participant had to fabricate a false alibi. Each participant had 10 minutes prior to the interview to formulate a convincing story.

If the participants were successful in convincing the interviewer that they did not take either the camera or the £10 they could earn an extra £5 in addition to their basic £5 participation payment. If they failed on either however, they would lose the extra payment and be asked to write a report about what had happened.

2.3 Recording Setting/Equipment

The experiment was conducted in the Linguistics department at the University of York. A vacant office room was used as the 'preparation room' in which participants completed task 1 and prepared for task 3. The 'target room', an unoccupied office room, from which the £10 was taken, was situated at the other end of the corridor approximately 200m away from the 'preparation room'. The baseline/control data and the security interview were recorded in a small recording studio. The interviewer and participants were sat down, oriented at approximately 90 degrees to each other. To ensure that the distance between microphone and speaker was kept constant an omnidirectional head-worn microphone of the type DPA 4066 was used. The microphone was coupled to a Zoom H4 recorder.

2.4 Parameters Analysed

The experiment took the form of a within-subjects design. Truthful and deceptive speech was elicited from participants during the interviews. In addition baseline

(control) data was recorded prior to the interviews. The three speaking conditions will be referred to as Baseline (B), Truth (T) and Deception (D) in this article.

A number of speech parameters was analysed including mean fundamental frequency (f_0 mean) and standard deviation (f_0 SD). The changes in mean energy across speaking conditions were computed as well as mean vowel formants F1, F2 and F3. Every speaker provided one file for each of the three speaking conditions, resulting in 60 files for analysis. The duration of the files was between about 3-5 minutes.

2.5 Measurement Equipment/Technique

'Sony Sound Forge' software was used for initial editing of the speech files. The acoustic analysis was performed using 'Praat' speech analysis software [13].

The f_0 based parameters were measured on the previously edited files using a Praat script developed by Philip Harrison.

In order to measure intensity, the files were edited in Praat so as to only contain speech (i.e. all silences were removed). The mean energy was then determined using a function of the Praat software. Rather than expressing the intensity values in absolute form, the differences between Baseline, Truth and Deception are reported in this paper.

Vowel formant measurements were extracted from Linear Predictive Coding (LPC) spectra using Praat's inbuilt formant tracker. The mean F1, F2 and F3 values were taken from an average of 10-20 milliseconds near the centre of each vowel portion. Any errors resulting from the in-built formant tracker were corrected by hand. To be counted in the analysis the vowels had to show relatively steady formants and be in stressed positions. For all speakers, 8 vowel categories[1]-FLEECE, KIT, DRESS, TRAP, NURSE, STRUT, LOT, and NORTH- were measured with one to 15 tokens (average 10 tokens) per category for each condition yielding a total of around 1200 measurements.

3 Results

The following section presents result for all 5 speakers. Statistical tests (T-tests) were employed where possible to assess the significance of the difference between the three conditions.

3.1 Fundamental Frequency (f_0)

Based on the f_0 mean and f_0 SD values obtained for each speaker and each condition, bar graphs were generated (Figure 1).

[1] Standard Lexical Sets for English developed by John C. Wells [14]. There are 24 lexical sets which represent words of the English language based on their pronunciation in Received Pronunciation (RP).

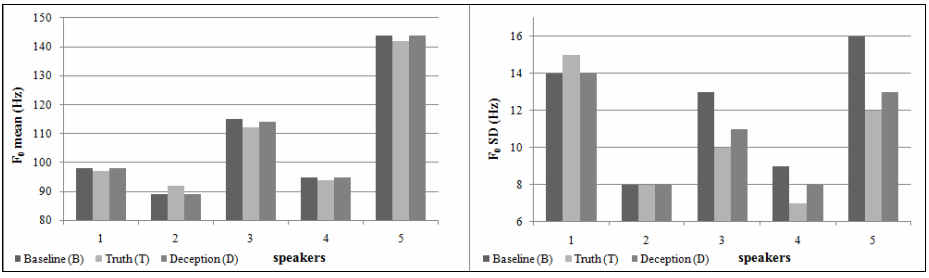

Fig. 1. f_0 mean (left) and f_0 SD (right) in Hz for all three speaking conditions for every speaker

Looking at Figure 1 (left) we can immediately see that there is not a great amount of difference in f_0 mean across conditions. There is a tendency that the f_0 mean of T is slightly lower than the values for B and D, which are close, but overall the mean f_0 values of all conditions are essentially similar. Examining the f_0 SD measures illustrated in the right hand side of Figure 1, we can perceive an overall trend in that there is less variation in f_0 in the Truth and Deception condition compared to the Baseline. The values of T and D for each speaker, in contrast, are rather comparable.

3.2 Intensity

Mean energy for each speaker is represented in Figure 2. No specific patterns can be generalized from the results. There is variability in direction and extent of change across speakers for both Truth and Deception. Apart from speakers 1, 3 and 5 the changes in mean energy are very small and interestingly, these three speakers show a uniform change in direction for both Truth and Deception.

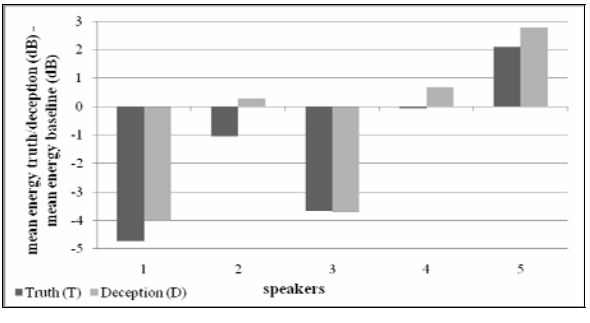

Fig. 2. Overall mean energy changes between Baseline and Truth/Deception for all speakers

3.3 Vowel Formants

F1

Figure 3 illustrates the mean F1 values for each measured vowel category from each individual speaker. The x-axis represents the F1 taken from the Baseline condition

and the difference between Baseline and Truth/Deception is shown along the y-axis. The majority of tokens lay on or slightly above the origin on the y-axis, which indicated that the F1 values from Truth/Deception were similar to those from the Baseline speech. There was some variability across tokens, demonstrated by the moderate spread of data points. As suggested by the almost horizontal trend lines, no significant correlations existed between F1 in the Baseline condition and the change in F1 in the Truth ($r = 0.03926$, $df = 38$, $p = 0.8099$) or Deception conditions ($r = 0.00948$, $df = 38$, $p = 0.9537$).

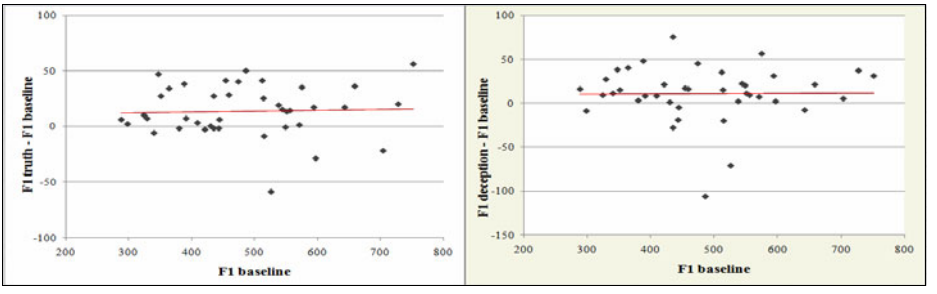

Fig. 3. Scatter-plot of F1 measures, showing value in the Baseline condition (x-axis) against shift observed in the Truth/Deception condition (y-axis)

F2

Figure 4 reflects the observed directional inconsistencies for F2. Some values were slightly increasing, some were decreasing and others were not changing. Overall the change was not considerable for any of the vowel categories and T-tests did not illustrate any significant differences at the 5% level.

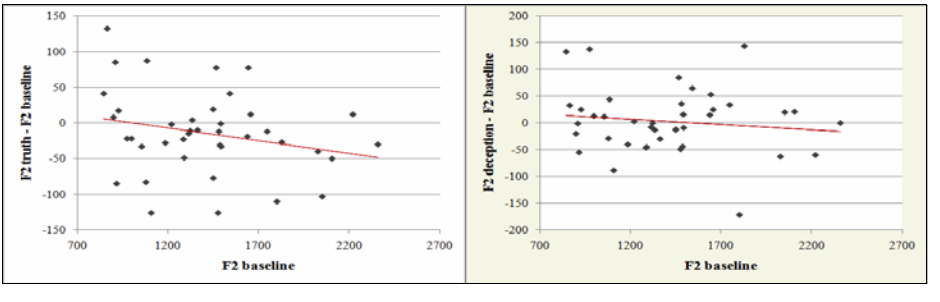

Fig. 4. Scatter-plot of F2 measures, showing value in the Baseline condition (x-axis) against shift observed in the Truth/Deception condition (y-axis)

In particular for the Truth condition, there was a substantial amount of variation between tokens in the size of the difference as well as the direction. The same vowel category might be increasing, decreasing or not changing between different speakers. There was less variation in the Deception condition and most of the tokens were

grouped around the origin. The nearly horizontal trend-line reinforces the observation that no real pattern can be detected. The correlation between Baseline vowel measurements and the effect of Truth (-0.24396, df = 38, p = 0.1292) and Deception (r = -0.12698, df = 38, p = 0.4349) was weak and not statistically significant at any conventional significance level.

F3

The F3 values of Baseline and Truth/Deception seemed to correspond very closely to each other and there did not appear to be a difference for any of the vowel categories. If changes did occur then it tended to be a decrease in Truth and Deception compared to Baseline.

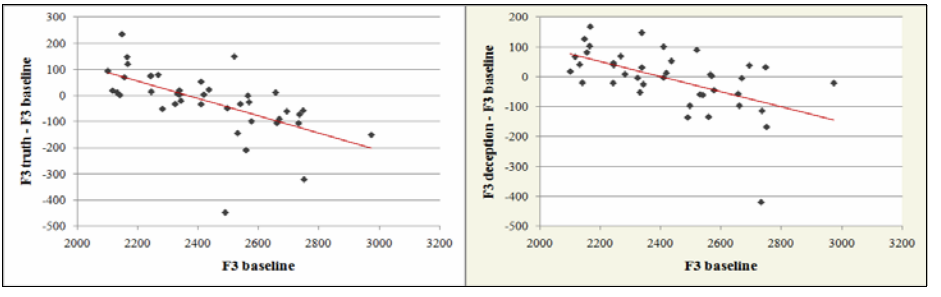

Fig. 5. Scatter-plot of F3 measures, showing value in the Baseline condition (x-axis) against shift observed in the Truth/Deception condition (y-axis)

There was a tendency that high F3 values in the Baseline condition were more likely to be subject to a decrease in Truth/Deception than low F3 values. The overall correlation between F3 in Baseline and F3 decrease in Truth (r = -0.59525, df = 38, p < .001) and Deception (r = -0.54605, df = 38, p < .001) was statistically significant.

4 Discussion

The results suggest that truth-tellers and liars cannot be differentiated based on the speech signal measures analysed in this study. Not only was there a lack of significant changes for the majority of parameters investigated but also, if change was present it failed to reveal consistencies within and between the speakers.

F_0 mean varied little across conditions. The reduced F_0 SD values in T and D suggest that speakers were less variable and perhaps spoke with a more monotone voice in these conditions. This could be further investigated by auditory analysis.

With regard to overall intensity changes, the findings also did not offer grounds for a reliable distinction between those telling the truth or lying. If speakers showed a change in mean energy then it was uniform in terms of direction and size across Truth and Deception.

The majority of F1 and F2 differences between conditions were not statistically significant. For F2 in particular, there was a considerable amount of variation across

conditions with values increasing, decreasing or not changing. The F3 results point towards a negative correlation between Truth/Deception and Baseline speech. Of note again is the parallelism between Truth and Deception in that both show a significant negative correlation compared to the Baseline. F3 is linked to voice quality and vocal profile analysis of the speakers would cast more insights.

The remarkable amount of inter- and intra-speaker variability underlines the fact that deceptive behaviour is individualised, very multifaceted and far from being clear cut and straightforward. Despite their non-significance the findings are of interest since they point to some potential limitations when trying speech analysis for deception detection purposes only.

It may be argued that the lack of significant findings is a product of the experimental arrangement as a laboratory induced deception which does not adequately represent deception as it might occur in real life. This is a methodological limitation which, due to ethical considerations, cannot be overcome in the majority of studies on deception. In order to maintain the impact of the scientific validity of this study, it can be said that post-interview rating scales confirmed that all the participants were highly motivated to succeed in the deceptive act (score of 6 or higher on a 7- point Likert scale). It should also be stated that research into a relatively unexplored area, such as speech and deception, needs to start off with fully controlled experiments where variables can be controlled more strictly. Clean, high quality recordings must provide the starting point for the acoustic and phonetic analysis. If differences between truth and deception are found in these ideal conditions, research can then move on to investigating less controlled data in the field.

One of the assets of the present research design concerns the separation of stress and deception in that the latter was not inferred from the former. The polygraph and most of the VSA technologies are based on the assumption that liars will show more emotional arousal i.e. will experience more stress than truth-tellers [2]. However, such a direct linkage cannot be presupposed. Certainly, there will be liars who do manifest the stereotypical image of nervousness and stress. At the same time, however, truth-tellers may also exhibit anxiety and tension, especially if in fear of not being trusted. And on the contrary, liars might not conform to the stereotypical image described above but rather display a composed and calm countenance. As the following quote illustrates:

> *'Anyone driven by the necessity of adjudging credibility who has listened over a number of years to sworn testimony, knows that as much truth has been uttered by shifty-eyed, perspiring, lip-licking, nail-biting, guilty- looking, ill-at-ease fidgety witnesses as have lies issued from calm, collected, imperturbable, urbane, straight-in-the-eye perjurers.'(Jones, E.A. in [2, p.102])*

Harnsberger et al. [5] for example only included participants in their analysis who showed a significant increase in stress levels during deception. Given that the goal of their research was to test the validity of VSA technology this may be a justified methodological choice. However, as the aim of the present study was to attain a more comprehensive knowledge of the fundamental relationship between deception and speech it was essential to disassociate deception and stress.

Further acoustic and phonetic analysis is under way to expand the analysis beyond measures of f_0, intensity, and formant measurements to include measurement of

diphthong trajectories, consonant articulation, jitter, shimmer and spectral tilting. In addition, laryngograph recordings have been made with 10 speakers which will provide opportunity to analyse the glottal wave signal and this in turn will contribute further to our knowledge of truth/deception specific speech characteristics. Furthermore, the hypothesis that increasing cognitive load during interview situations has the potential of magnifying the differences between truth-tellers and liars in the speech domain will also be evaluated.

5 Conclusion

This paper summarised an exploratory investigation into the relationship between acoustic parameters of speech and truth/deception. So far the analysed data does not suggest that a reliable and consistent correlation exists. As well as providing a basis for future research programs the present study should encourage researchers and practitioners to evaluate critically what is and what is not possible, using auditory and machine based analyses, with respect to detecting deception from speech.

Acknowledgement. This work was made possible through EPSRC Grant number: EP/H02302X/1. Thanks to Philip Harrison for providing the Praat script and Francis Newton for his time and assistance during the data collection process.

References

1. Jessen, M.: Einfluss von Stress auf Sprache und Stimme. Unter besonderer Beruecksichtigung polizeidienstlicher Anforderungen. Schulz- Kirchner Verlag GmbH, Idstein (2006)
2. Lykken, D.: A Tremor in the Blood: Uses and Abuses of the Lie Detector. Perseus Publishing, Reading (1998)
3. Damphousse, K.R., Pointon, L., Upchurch, D., Moore, R.K.: Assessing the Validity of Voice Stress Analysis Tools in a Jail Setting. Report submitted to the U.S. Department of Justice (2007)
4. Eriksson, A., Lacerda, F.: Charlantry in forensic speech science: A problem to be taken seriously. International Journal of Speech, Language and the Law 14(2), 169–193 (2007)
5. Harnsberger, J.D., Hollien, H., Martin, C.A., Hollien, K.A.: Stress and Deception in Speech: Evaluating Layered Voice Analysis. Journal of Forensic Science 54(3), 642–650 (2009)
6. Benus, S., Enos, F., Hirschberg, J., Shriberg, E.: Pauses in deceptive Speech. In: Proceedings ISCA 3rd International Conference on Speech Prosody, Dresden, Germany (2006)
7. Feeley, T.H., deTurck, M.A.: The behavioural correlates of sanctioned and unsanctioned deceptive communication. Journal of Nonverbal Behavior 22(3), 189–204 (1998)
8. Stroemwall, L.A., Hartwig, M., Granhag, P.A.: To act truthfully: Nonverbal behaviour and strategies during a police interrogation. Psychology, Crime and Law 12(2), 207–219 (2006)
9. Anolli, L., Ciceri, R.: The Voice of deception: Vocal strategies of naïve and able liars. Journal of Nonverbal Behavior 21(4), 259–284 (1997)

10. Rockwell, P., Buller, D.B., Burgoon, J.K.: The voice of deceit: Refining and expanding vocal cues to deception. Communication Research Reports 14(4), 451–459 (1997)
11. Enos, F.: Detecting Deception in Speech. PhD Thesis submitted to Columbia University (2009)
12. Porter, S.B.: The Language of Deceit: Are there reliable verbal cues to deception in the interrogation context? Master's thesis submitted to The University of British Columbia (1994)
13. Boersma, P., Weenink, D.: Praat: doing phonetics by computer (Computer program). Version 5.2.12, http://www.praat.org/ (retrieved January 28, 2011)
14. Wells, J.C.: Accents of English I: An Introduction. Cambridge University Press, Cambridge (1982)