INVESTIGATING THE ACOUSTIC CHARACTERISTICS OF DECEPTIVE SPEECH

Christin Kirchhübel & David M Howard

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

ABSTRACT

This study explored the changes in the speech signal when people were being deceptive. Truthful, deceptive and control speech was elicited from ten speakers during an interview setting. Results are presented on parameters including f_0, intensity and vowel formant frequencies. A significant correlation could not be established for any of the acoustic features. As well as providing a basis for future research programs, the present study should encourage researchers and practitioners to evaluate critically what is (im)possible using auditory and machine based analyses with respect to detecting deception from speech.

Keywords: deceptive speech, Voice Stress Analysis (VSA), acoustic and phonetic features

1. INTRODUCTION

The human voice can provide information about a number of characteristics including, but not limited to, the speaker's age, ethnic and social background, and presence of voice pathology or alcohol intoxication. Hearing the voice alone enables listeners to make inferences about a speaker's affective state including the emotions experienced and the presence of psychological stress. Considerable interest has been devoted to the identification and decoding of the acoustical characteristics of affective speech [8]. If it is possible to deduce speakers' emotional condition from listening to their voice, could it be viable to make judgments about their sincerity from speech as well? A method that reliably detects deception would indeed be of considerable practical relevance and of benefit to police, intelligence agencies, military and security personnel.

The majority of previous research on deception has been conducted by psychologists who were particularly interested in the non-verbal aspects of deceptive behaviour such as gestures, movement and facial expressions [3]. More recently, interest into the verbal aspects of deception has grown and techniques employed in the linguistic analysis of language such as Reality Monitoring [10] and Statement Validity Assessment (SVA) [11] are tested with deceptive speech data.

In comparison to the behavioural, physiological and linguistic aspects of deception, surprisingly, very little research has been carried out on the acoustic and phonetic characteristics of deceptive speech. There are a number of studies that have analysed temporal features such as speaking rate, pauses, hesitations and speech errors, but only a few studies have investigated frequency-based parameters such as mean f_0 and f_0 variability (for an overview see [13]). Evidence for the analysis of other phonetic and acoustic features such as vowel and consonant articulation or voice quality in connection with deceptive speech is rare in the research published to date. Recently completed work by Enos [4] is one of the first attempts to analyse deceptive speech using spoken language processing techniques and Torres, et al. [12] investigated the relationship between features of the glottal waveform and deception.

In recent times, technologies have been promoted, claiming to measure peoples' veracity based on the speech signal. Voice Stress Analysis (VSA) relies on the theory of 'microtremors', while Layered Voice Analysis (LVA) is said to be based on the analysis of multiple layers of the voice. Both VSA and LVA have been subjected to reliability testing, which in general resulted in negative outcomes [2, 7]. While reliability 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 [5]. In other words, whether a relationship exists between deception, truth and speech, and if so, what the nature of this relationship is.

This paper describes an initial investigation into the acoustic and phonetic correlates of deceptive speech using auditory and acoustic analysis.

2. METHODOLOGY

2.1. Data and experimental design

The data consisted of baseline, truthful and deceptive speech from ten male native British English speakers. The experimental procedure was based on a mock-theft paradigm and a subsequent 'security interview' in which the participant was questioned about two thefts. Participants committed one of the thefts but not the other; however, it was their task to convince the interviewer that they were not guilty of either. The interviewees were motivated to succeed by financial as well as self-presentational incentives. In a similar fashion to [4], the data was divided into global and local Truth/Lie. Global Truth/Lie reflects the overall intention to be truthful or to deceive, whereas local Truth/Lie represents actual truthful or deceitful statements.

2.2. Recordings

The interviews were conducted in a recording studio in the Linguistics department at the University of York. Subjects were seated, with an omnidirectional head-worn microphone which was coupled to a Zoom H4 recorder. Recordings were made using a sampling rate of 44.1 kHz and a bit depth of 16 bit. Every speaker provided one file, of 3-5 minutes duration, for each of the three speaking conditions.

2.3. Measurements

The acoustic analysis was performed using the Speech Analysis Software *Praat 5.1.44* [1]. F_0 mean and f_0 Standard Deviation (SD) values were measured by means of a *Praat* script developed by Philip Harrison using the autocorrelation method.

Intensity was measured in terms of dB using Praat's in-built function. Rather than expressing the values in absolute form, the relative differences between the three speaking conditions will be reported.

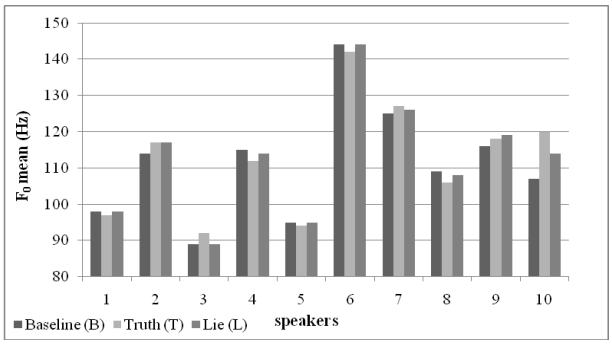
Vowel formant measurements were extracted from Linear Predictive Coding (LPC) spectra using *Praat's* inbuilt formant tracker. The mean F_1, F_2 and F_3 values were taken from an average of 10-20 ms near the centre of each vowel portion. Any errors resulting from the inbuilt formant tracker were corrected by hand. For all speakers, 8 vowel categories - FLEECE, KIT, DRESS, TRAP, NURSE, STRUT, LOT, and NORTH [14] - were measured with one to 15 tokens (average 10 tokens) per category for each condition.

3. RESULTS

3.1. F_0

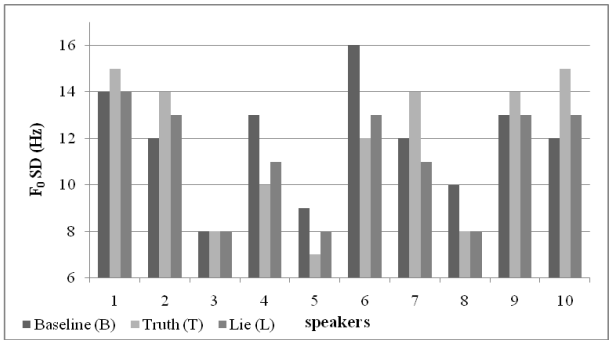
Apart from one speaker (speaker 10) the differences in f_0 mean across conditions were not notable. The values were essentially similar for Baseline, Truth and Lie. With a rise in f_0 mean of 13 Hz and 7 Hz for Truth and Lie respectively, speaker 10 was the only speaker who showed a dramatic change.

Figure 1: F_0 mean for all three speaking conditions for every speaker.



With regards to f_0 SD no general trend emerged. While four speakers (4, 5, 6, 8) demonstrated less variation in f_0 in the Truth and Lie conditions compared to the Baseline, two (2, 10) exhibited a rise in f_0 variability and others did not change substantially. Of interest yet again was the homogeny in direction of change for Truth and Lie. As can be seen in Figure 2, with the exception of speaker 7, the values of Truth and Lie for each speaker were rather comparable. The values of the local and global measurements for f_0 mean and f_0 SD showed large correspondence. A Friedman's ANOVA returned the differences in mean f_0 and f_0 SD across conditions to be non-significant.

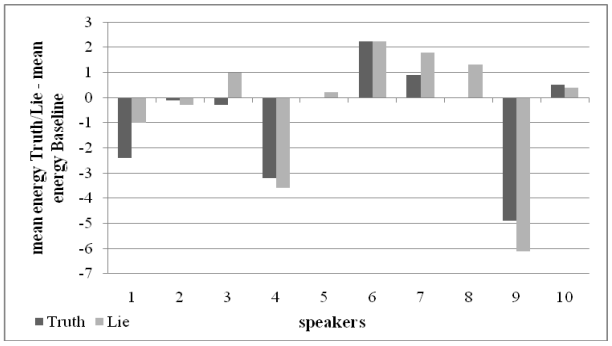
Figure 2: F_0 SD for all three speaking conditions for every speaker.



3.2. Intensity

A pattern could not be generalized from the intensity measurements. There was variability in direction and extent of change across speakers for both Truth and Lie. For some speakers (2, 3, 5, and 10) the changes were relatively small whereas for others they were somewhat larger (4, 6, and 9). Of interest is that almost all of the speakers showed a uniform change in direction for both Truth and Lie. When contrasted, the values of the local and the global measurements were in general analogous. The differences in intensity across conditions were non-significant. The investigation of the interdependence of f_0 mean and intensity changes did not result in strong correlations. Speakers who did not have a remarkable change in intensity between Baseline and Truth/Lie nevertheless showed f_0 mean differences.

Figure 3: Overall intensity changes between Baseline and Truth/Lie for all speakers.



3.3. Vowel articulation

The majority of F_1, F_2 and F_3 differences between conditions were statistically non-significant (Friedman's ANOVA). For F_2 in particular, there was a considerable amount of variation across conditions with values increasing, decreasing or not changing. Some of the F_1 values exhibited a slight increase and F_3, if changing, tended to reduce in the Truth and Lie conditions as compared to the Baseline. By breaking the results down according to individual speakers more variability emerged. For F_1, there appeared to be variation in magnitude as well as direction of change in Truth and Lie. F_3 was relatively stable within speakers across conditions and for F_2 it was the back vowels which seemed to be affected more than the front and central vowels. There was motivation to analyse the possibility of a relationship between intensity changes and formant behavior. However, when correlations were run between the two variables, no significant correlation existed.

4. DISCUSSION

Preliminary analysis suggests that truth-tellers and liars cannot be differentiated based on the speech parameters measured in this study. Overall, there was a lack of sizeable changes and, if change was present, it tended to be uniform for both Truth and Lie. 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.

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 90% of the participants were highly motivated to succeed in the deceptive act (score of 5 or higher on a 7-point Likert scale). Harnsberger, et al. [6] have already stated the necessity of fully controlled experiments but to reiterate again, research into a relatively unexplored area, such as speech and deception, needs to start off with fully controlled experiments. 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.

One of the contributions 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 the majority of voice stress analysis technologies are based on the assumption that liars will show more emotional arousal, i.e. will experience more stress than truth-tellers [9]. However, such a direct relationship 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. Furthermore, 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-bitting, 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 Lykken [9] p. 102])*

Harnsberger, et al. [6] only included those subjects into analysis who showed a significant increase in stress levels during deception. Given that the aim 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.

A further advantage of the current study was the fact that it focussed on investigating the acoustic correlates of deception as well as truth. The relationship between speech and veracity is just as important, but has mostly been overlooked in previous research designs.

5. CONCLUSION

This paper summarised an initial investigation into the relationship between some acoustic parameters of speech and truth/lying. So far the analysed data does not suggest that a reliable and consistent correlation exists. Further work involving laryngograph recordings and additional acoustic and phonetic features, e.g. measurement of diphthong trajectories, jitter, shimmer, consonant articulation, spectral tilting, and parameters of speaking flow, is currently in progress. This will provide the basis for a comprehensive evaluation of the nature of the manifestation of truth and deception in the speech domain.

6. REFERENCES

[1] Boersma, P., Weenink, D. Praat: Doing phonetics by computer [Computer program]. Version 5.2.12, retrieved 28 January 2011 from *http://www.praat.org/*.

[2] Damphousse, K.R., Pointon, L., Upchurch, D., Moore, R.K. 2007. *Assessing the Validity of Voice Stress Analysis Tools in a Jail Setting.* Report submitted to the U.S. Department of Justice. *http://www.ncjrs.gov/pdffiles1/nij/grants/219031.pdf*

[3] DePaulo, B.M., Lindsay, J.J., Malone, B.E., Muhlenbruck, L., Charlton, K., Cooper, H. 2003. Cues to deception. *Psychological Bulletin* 129(1), 74-118.

[4] Enos, F. 2009. *Detecting Deception in Speech*. Ph.D. Thesis submitted to Columbia University.

[5] Eriksson, A., Lacerda, F. 2007. Charlatanry in forensic speech science: A problem to be taken seriously. *International Journal of Speech, Language and the Law* 14(2), 169-193.

[6] Harnsberger, J.D., Hollien, H., Martin, C.A., Hollien, K.A. 2009. Stress and deception in speech: Evaluating layered voice analysis. *Journal of Forensic Science* 54(3), 642-650.

[7] Hollien, H., Harnsberger, J.D. 2006. Voice stress analyzer instrumentation evaluation. *Final report, CIFA Contract*–FA4814-04-0011. *http://www.clas.ufl.edu/users/jharns/Research%20Projects/UF_Report_03_17_2006.pdf*

[8] Kappas A., Hess, U., Scherer, K.S. 1991. Voice and emotion. In Rime, B., Feldman, R.S. (eds.), *Fundamentals of Nonverbal Behavior*. Cambridge and New York: Cambridge University Press, 200-238.

[9] Lykken, D. 1998. *A Tremor in the Blood: Uses and Abuses of the Lie Detector*. *Reading*, MA: Perseus Publishing.

[10] Sporer, S.L. 2004. Reality monitoring and detection of deception. In Granhag, P.A., Stromwall, L.A. (eds.), *Deception Detection in Forensic Contexts*. Cambridge, UK: Cambridge University Press, 64-102.

[11] Steller, M., Köhnken, G. 1989. Criteria-based content analysis. In Raskin, D.C. (ed.), *Psychological Methods in Criminal Investigation and Evidence*. NewYork, NJ: Springer-Verlag, 217-245.

[12] Torres, J.F., Moore, E. II, Bryant, E. 2008. A study of glottal waveform features for deceptive speech classification. *Proceedings ICASSP* Las Vegas, 4489-4492.

[13] Vrij, A. 2008. *Detecting Lies and Deceit: Pitfalls and Opportunities*. Chichester: Wiley

[14] Wells, J.C. 1982. *Accents of English I: An Introduction*. Cambridge: Cambridge University Press