Analyzing Deceptive Speech

Christin Kirchhübel[1], Alex W. Stedmon[2], and David M. Howard[1]

[1] Audio Laboratory, Department of Electronics, University of York, UK
{ck531,david.howard}@york.ac.uk
[2] Cultural Communications and Computing Research Institute (C3RI),
Sheffield Hallam University, UK
a.stedmon@shu.ac.uk

Abstract. This current work explored the speech-based attributes of participants who were being deceptive in an experimental interrogation setting. In particular, the study attempted to investigate the appropriateness of using temporal speech cues in detecting deception. Deceptive and control speech was elicited from nineteen speakers and the data was analyzed on a range of speech parameters including Speaking Rate (SR), Response Onset Time (ROT) and frequency and duration of Hesitation markers. The findings point to a significant increase in SR, a significant decrease in ROT and a reduction in hesitation phenomena in the deceptive condition suggesting an acceleration of overall speaking tempo. The potential significance of temporal parameters for detecting deception in speech is recognized. However, the complex and multifaceted nature of deceptive behaviour is highlighted and caution is advised when attempting veracity judgments based on speech.

Keywords: Deception, Speaking Rate, Response Onset Time, Hesitations.

1 Introduction

Early research into characteristics of deception tended to focus on specific cues or behaviours that would reliably indicate that a deception was taking place. This has proved particularly problematic and more recently researchers have begun to investigate the emotional, cognitive and communicative processes that tend to accompany deception. Following this a number of theoretical frameworks have been developed to predict and account for the behaviour liars may display (De Paulo et al. 2003, Ekman, 1985, Miller & Stiff, 1993, Vrij, 2008). One such approach is the Cognitive Theory of deception that considers lying to be cognitively more demanding than truth-telling and empirical evidence supporting this line of thought can be found in the work of Walczyk et al. (2003, 2005).

From a speech analysis perspective, it has also been established that filled and unfilled pauses in speech are reactions associated with cognitive processes (Goldman-Eisler 1968). Based on this it may be hypothesised that the increase in cognitive load required for deception may lead to specific speech dependent measures of deception that would manifest themselves in the temporal domain, specifically, in an overall slowing down of speech.

D. Harris (Ed.): EPCE/HCII 2013, Part I, LNAI 8019, pp. 134–141, 2013.
© Springer-Verlag Berlin Heidelberg 2013

This current study explored the speech-based attributes of participants who were being deceptive in an experimental interrogation setting. Specific data were taken from audio recordings of the interrogation sessions that were part of a broader research study investigating a number of human deception responses across biological, physiological, psychological and behavioural dimensions. This paper presents an overview of the methodology that was relevant to the speech analysis and focuses specifically on the deceptive participants. More detailed descriptions of this research are contained in (Eachus et al. 2012).

2 Method

A total of 19 male participants were drawn from the staff and student population at the University of Nottingham where the experiment was conducted. All participants were native British English speakers and none had any self-reported voice, speech or hearing disorders.

A scalable interrogation paradigm was developed specifically for this study in which participants progressed from a baseline interview through two levels of interrogation (e.g. Baseline, Interview 1 and Interview 2). Participants were given a 'token' containing pictorial and verbal information that they had to conceal from the interviewers during the scaled interrogations. The Baseline interview contained neutral and relaxation based questions, designed to elicit control data and non-deceptive speech data. Both the interviews aimed to increase participant arousal by asking more probing and penetrating questions. Interview 1 provoked a low level of emotional involvement by posing general questions about social desirability and information concealment, whereas Interview 2 was more provocative by directly challenging participants about their truthfulness.

The questions for the three conditions were pre-recorded as audio files and presented via loudspeaker from a standard laptop computer to ensure that they remained constant across participants. For the most part, the questions were of a yes/no format resulting in the generation of short answer/monosyllabic responses. The three conditions contained 20 questions but in order to avoid participants anticipating the end of the interviews, the questions were not numbered in a serial fashion. The order of the conditions (e.g. Baseline, Interview 1 and Interview 2) was kept the same for all participants. Overall, the experiment took approximately 75 minutes to complete after which participants were debriefed and received a £30 participation reward.

3 Parameters Analysed

Every speaker provided 1 file for each of the three speaking conditions, resulting in 54 files for analysis. Given the nature of the data only a selected number of temporal parameters could be investigated. Amongst these were Speaking Rate (SR), Response Onset Time (ROT) and frequency as well as duration of filled pauses.

4 Apparatus

'Sound Forge TM Pro 10' software (Sony Creative Software) was used for initial editing of the speech files. The temporal analysis was performed using Praat 5.1.44 speech analysis software (Boersma & Weenink 2005). SR measurements were based on the number of phonetic syllables in each participant's speaking turns. ROT was measured as the time in between the end of a question and the beginning of the participant's response. With regards to hesitation markers, the frequency and duration of the 'vocalic' [ɛ] as well as 'vocalic + nasal' [ɛm] variants were taken into account. In order to control for differences in length of speaking time across conditions and speakers, the frequency aspect was conveyed in the form of a Hesitation Rate (HR) measurement calculated as number of hesitations per minute. Textgrids generated by Praat facilitated easy access to durational calculations which were then transferred into Microsoft Excel for a more accessible examination.

5 Results

The following section presents the results for all 19 speakers. Repeated Measures ANOVAs were employed to assess the significance of the inter- and intra-speaker comparisons for SR, ROT and frequency of hesitations. In cases where Mauchly's test indicated that the assumption of sphericity had been violated degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity. Post-hoc tests, if applicable, are reported using Bonferroni correction. Not every speaker employed hesitation markers which resulted in a lack of durational measures for some and, consequently, a reduction in sample size. Therefore, in order to account for the relatively small sample of durational measures, the non-parametric Friedman's ANOVA test was chosen.

5.1 Speaking Rate (SR)

Mean SR appeared to be affected by the different interview conditions. About half of the speakers tended to decrease their mean SR in Interview 1 while the other half showed an increase. Irrespective of the direction of change, the extent of change was similar ranging from 4.1%/-2.9% at the lower end to 29.5%/-31.2% at the upper spectrum. The results of Interview 2 were more coherent with 15 out of 19 speakers showing a higher mean SR compared to the Baseline. This increase spanned from 4.3% to 48.2%. A decrease was only noteworthy for 3 out of the 4 speakers covering a smaller range from -9.8% to -18.4%. Not only was there a traceable increase in mean SR in Interview 2 when compared to the Baseline, the effect was also evident, perhaps even more so, when contrasting the two interviews against each other. Almost all speakers exhibited a faster mean SR in Interview 2 than Interview 1 with values ranging from a 5.7% increase to a substantial 56.1% increase. At the inter-speaker level, mean SR changed significantly across the three conditions ($F(2, 36) = 7.271$, $p \leq .01$). Pairwise comparisons revealed that mean SR did not change significantly between Baseline (mean = 3.8 syll/sec) and Interview 1 (mean = 3.7 syll/sec) but that there was a significant increase in mean SR between Baseline and

Interview 2 (mean = 4.2 syll/sec) and between the two interviews. The observed trend of an increase in mean SR in Interview 2 also held true at the intra-speaker level. Post-hoc comparisons revealed a significant difference in mean SR between Baseline and Interviews for 7 out of the 19 participants. For the remaining 12 participants the change in mean SR between the three conditions was not significant ($p \geq .05$).

5.2 Response Onset Time (ROT)

Results from the ANOVA illustrated that mean ROT differed significantly between the three experimental tasks ($F (1.316, 23.689) = 5.802$, $p \leq .01$). Post hoc tests revealed that there was a significant decrease in ROT in Interview 1 (mean = 171ms) as compared to Baseline (mean = 228ms) and Interview 2 (mean = 196ms) ($p \leq .05$). Although no statistically significant change was observed between Baseline and Interview 2, a similar pattern of a reduction in ROT in the latter as compared to the Baseline was apparent. The majority of speakers showed a decrease in ROT for interview 1 which varied from 0.2 % to 63.7%. The magnitude of change for the four participants who increased their ROT in interview 1 only spanned from 1.7 % to 32.4% so the amount of maximum decrease in ROT was considerable larger than the amount of maximum increase. More participants increased their ROT in Interview 2 than Interview 1. The range of the increase in ROT displayed by seven participants was from 3.4% to 43.9%. The range of the reduction in ROT for the remaining 12 participants extended from 0.9% to 60.5%. Once again, the magnitude in decrease (227.6 ms) is greater than the magnitude in increase (146.9 ms). In terms of the intra-speaker analysis, seven of the participants showed a significantly lower ROT in one or both of the interview conditions when compared to the Baseline. Only one participant showed a significant difference in ROT between the two interview conditions.

5.3 Hesitations

Hesitations were analysed according to frequency and duration. As the calculation of HR resulted in only one numerical result per speaker per condition, statistical testing could only be performed on the inter-speaker level. The change in frequency of hesitations between the Baseline and two interview conditions was characterized by a decrease. A repeated-measures ANOVA illustrated that this change was statistically significant ($F (1.185, 21.326) = 4.598$, $p \leq .05$) and post-hoc comparisons identified that there were significantly less hesitations in Interview 2 (mean = 6.831 Hes/min) as compared to the Baseline (mean = 11.861 Hes/min). Although not statistically significant, the trend of a decrease in hesitations could also be observed for Interview 1 (mean = 7.750 Hes/min) as opposed to the Baseline. While a hesitation measure was obtained for all participants in the Baseline condition, the interview conditions often featured no hesitations at all. This was the case for 9 participants in Interview 1. 8 of the remaining 10 speakers decreased their HR by an average of 55.1%. Merely two participants employed more hesitations in Interview 1 as compared to the Baseline but for one of these the increase was striking reaching a doubling in HR from 30.18 Hes/min to 61.34 Hes/min. Similarly to interview 1, 7 participants did not produce

any hesitations in interview 2 and 9 showed a decrease in HR averaging 55.4%. Three participants illustrated an increase in HR from Baseline to Interview 2 but the magnitude of increase was only noteworthy for one of the speakers reaching 57.6% as compared to 6% and 12% for the other two. 6 participants had no hesitations in either of the interviews. Thirteen participants demonstrated changes in hesitation frequency between the interviews but the direction of change varied amongst them with an equal number increasing and decreasing. While the range of decrease spanned from -0.7% to -33.4%, the increase was more remarkable stretching from 21% to 106.4%. Statistical examination confirmed no difference between Interview 1 and Interview 2 ($p \geq .05$). Only 9 out of the 19 speakers offered hesitation markers in all three conditions and even then the number of occurrences tended to be very low sometimes merely reaching 1 per condition. Therefore, it was decided to limit statistical testing to the inter-speaker level once again only using this subset of 9 participants. Duration of hesitations was significantly affected by the interviews ($x2\ (2) = 8$, $p \leq .05$). Wilcoxon Sign Ranked tests were used to follow up on this finding and it appeared that duration was no different between Baseline (mean = 43.9ms) and Interview 1 (mean = 41ms) or between Interview 1 and Interview 2 (mean = 36ms). However, for Interview 2, duration of hesitations was significantly lower compared to the Baseline ($T = 0$, $p \leq .01$, $r = -.63$). None of the participants showed longer durations in Interview 2 when compared to the Baseline and for 5 out of the 9 participants this durational drop was particularly noticeable averaging around 30ms. For the remaining 4 participants the decrease was less striking ranging from 0.2ms to 3.2ms. While 6 participants also showed a reduction in hesitation length in Interview 1 ranging from 3.5ms to 17.1ms, we find that 3 participants produced longer hesitation markers as compared to the Baseline.

6 Summary and Discussion

A summary of the results is presented in Table 1. A significant increase in mean SR was observed for Interview 2 when contrasted with the Baseline and Interview 1. Mean ROT appeared to be decreasing across the two interview conditions compared to the Baseline but this reduction was only significant for Interview 1. Number of hesitations significantly declined in both interviews compared with the Baseline, but duration was only affected in Interview 2 which was characterized by a significant shortening in length of hesitation markers when compared against the Baseline values.

For the majority of parameters examined the experimental effect manifested itself between the Baseline and either or both of the interview conditions. There tended to be little difference between the two interviews themselves despite the heightened interrogative pressure.The increase in mean SR in the interviews corresponds with the observed reduction in the number of hesitation pauses. In addition to this, the shorter ROTs would further indicate a general acceleration of speaking tempo when being deceptive.

Table 1. Summary of the results for all parameters investigated

Parameter	*Interview 1*	*Interview 2*
Speaking Rate (SR)	• Mixed results	• Significant increase compared to Baseline and Interview 1
Response Onset Time (ROT)	• Significant decrease compared to Baseline and interview 2	• Tendency for a decrease compared to Baseline
Frequency of Hesitations (HR)	• Significant decrease compared to Baseline • No change to Interview 2	• Significant decrease compared to Baseline • No change to Interview 1
Duration of Hesitations	• No significant change to Baseline or interview 2	• Significant decrease compared to Baseline

When probing the available literature, it appears that research into the temporal aspects of deceptive speech has resulted in conflicting observations. Indeed, the majority of studies tended to observe an overall slowing down of speaking tempo and, in particular, an increase in hesitation phenomena supporting the cognitive theory of deception as briefly outlined above. However, a number of studies exist which suggest the opposite, namely, a decrease in ROT and speech disturbances which complements the results of the current experiment (Benus et al. 2006, Vrij & Heaven 1999). A solution to the apparent disparity is offered by Vrij & Heaven (1999) who concluded that lie complexity affects pausing behavior. In their research the authors illustrated that liars made fewer speech disturbances when the lie was easy to fabricate as opposed to a more cognitively complex fabrication. Furthermore, research into prepared and spontaneous lying has shown that anticipated lies carried shorter ROTs compared to truthful utterances while spontaneous lies did not conform to this pattern (O'Hair et al. 1981).

In this study, participants were able to prepare for their deceptive act as they were informed, prior to the interrogation process, that they would be required to conceal knowledge. In addition, 'yes' or 'no' were legitimate answers to the majority of the questions and, therefore, it could be envisaged that the amount of cognitive energy necessary to carry out the deception in the present experiment was minimal. Participants may have resorted to some form of automated responding, akin to a suspect's 'no comment' strategy in PACE (Police and Criminal Evidence Act) interviews as the interrogations progressed. Goldman-Eisler (1968:58-59) has shown that routine and automated speech results in a decrease of pausing behavior and this could provide a viable explanation for the present findings.

Vrij & Heaven (1999) make reference to the attempted control theory in order to account for the decrease in speech disturbances apparent in their study. They argue that in order to present a truthful demeanour people may try to suppress or control behaviours which they associate with lying and consequently expect their target to be associating with lying. This type of behavioural management could lead to an

'overcontrol' of speech resulting in the decrease of hesitation phenomena. Both rationalizations are plausible and may work in tandem; however more empirical research is needed to assess the relationship between behavioural control and speech (Kirchhübel & Howard in press).

Having said all this and when taking a closer look at the experimental design one has to be careful not to come to premature conclusions. The nature of the data tended to reflect short answer responses and speaking turns tended to contain little speech and at times only 'yes' or 'no' responses. There are limitations of measuring tempo related speech characteristics using these types of utterances. Sound prolongation is one of the more obvious factors but at the other end of the scale we might also find syllable shortening both of which would unjustly influence mean SR measurements. The nature of yes/no responses also limit the opportunity for the realisation of hesitation markers and may affect ROT measures due to participants' engaging in automated/routine responding.

Amongst the many factors affecting speech, speaking context and setting have been shown to be influential (Giles et al. 1991). Compared to the Baseline which aimed to create a positive and warm atmosphere, the two interviews posed a more threatening communicative environment. Setting aside the participants' intent to deceive, the interrogation process itself could have contributed to participants feeling intimidated and affronted and may therefore have resorted to speech divergence and limited information sharing. The presence of variability within as well as between participants underlines the fact that deceptive behaviour is individualised, very multifaceted and far from being clear cut, a finding which repeatedly emerges in research on deception (Kirchhübel & Howard 2012).

7 Conclusion and Limitations

The present research highlights the complex nature of deception and underlines the difficulties in locating consistent cues of deceptive behavior. There are a range of factors that affect deceptive behaviour with 'lie complexity' and 'time to prepare for deception' being of most relevance to the present experiment. Differences in research design and analysis methods as well as failure to control for the various confounding variable have led to discrepancies in results. Temporal analysis may be useful for deceptive speech detection but may only apply in specific settings and certainly needs to take into account situational as well as individual factors. Applying the research to a practical setting, caution needs to be employed when attempting veracity judgments based on the temporal characteristics of speech regardless of whether the speech stemmed from formal interviews or polygraph examinations using the guilty knowledge paradigm (Lykken 1998). It is worth emphasizing that more data is in existence which was taken from of a group of truth-tellers undergoing the same experimental procedure. Unfortunately, the deceptive and truthful speech samples are based on a between-subjects design and, therefore, the scope for direct comparison of features is limited; however, temporal speech analysis of the truth-tellers would nevertheless offer additional insight and is planned in future work.

References

1. Benus, S., Enos, F., Hirschberg, J., Shriberg, E.: Pauses in deceptive Speech. In: Proceedings ISCA 3rd International Conference on Speech Prosody, Dresden, Germany (2006)
2. DePaulo, B.M., Lindsay, J.J., Malone, B.E., Muhlenbruck, L., Charlton, K., Cooper, H.: Cues to deception. Psychological Bulletin 129(1), 74–118 (2003)
3. Eachus, P., Stedmon, A.W., Baillie, L.: Hostile Intent in Crowded Places: A Field Study. Applied Ergonomics (2012), http://dx.doi.org/10.1016/j.apergo.2012.05.009
4. Ekman, P.: Telling lies: Clues to deceit in the marketplace, politics and marriage. W. W. Norton, New York (1985) (Reprinted in 1992 and 2001)
5. Giles, H., Coupland, N., Coupland, J.: Accommodation Theory: Communication, Context and Consequence. In: Giles, H., Coupland, J., Coupland, N. (eds.) Contexts of Accommodation: Developments in Applied Sociolinguistics. CUP, Cambridge (1991)
6. Goldman-Eisler, F.: Psycholinguistics. Experiments in Spontaneous Speech. Academic Press, London (1968)
7. Kirchhübel, C., Howard, D.M.: Detecting suspicious behavior using speech: Acoustic correlates of deceptive speech – an exploratory investigation. Applied Ergonomics (2012), http://dx.doi.org/10.1016/j.apergo.2012.04.016
8. Kirchhübel, C & Howard, D. M.: Deception and Speech – a theoretical overview (in press)
9. Lykken, D.: A Tremor in the Blood: Uses and Abuses of the Lie Detector. Perseus Publishing, Reading (1998)
10. Miller, G.R., Stiff, J.B.: Deceptive Communication. Sage, Newbury Park (1993)
11. O'Hair, H.D., Cody, M., McLaughlin, M.L.: Prepared lies, spontaneous lies, Machiavellianism and nonverbal communication. Human Communication Research 7, 325–339 (1981)
12. Vrij, A.: Detecting lies and deceit: pitfalls and opportunities, 2nd edn. Wiley, West Sussex (2008)
13. Vrij, A., Heaven, S.: Vocal and verbal indicators of deception as a function of lie complexity. Psychology, Crime & Law 5, 203–315 (1999)
14. Walczyk, J.J., Roper, K.S., Seemann, E., Humphrey, A.M.: Cognitive mechanisms underlying lying to questions: Response time as a cue to deception. Applied Cognitive Psychology 17, 744–755 (2003)
15. Walcyyk, J.J., Schwartz, J.P., Clifton, R., Adams, B., Wei, M., Zha, P.: Lying person-to-person about live events: A cognitive framework for lie detection. Personnel Psychology 58, 141–170 (2005)