The acoustic and temporal characteristics of deceptive speech

Christin Kirchhübel

Department of Electronics
The University of York

Submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

Submitted October 2013

Abstract

Voice stress analysis based technologies, commonly known as Voice Stress Analysers (VSA), which are said to measure peoples' veracity based on the speech signal, have come under a large amount of scientific scrutiny in recent years. Scientific reliability testing of these products has exclusively resulted in negative evaluations. 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. In other words, whether a relationship exists between deception, truth and speech, and if so, what the nature of this relationship is.

Using two empirical studies, the present research explores the viability of using speech analysis as a means of differentiating between deceptive and non-deceptive speech. Experiment 1 takes the form of a laboratory-based study which employs a mock-theft paradigm in conjunction with a 'security interview' to elicit baseline/control and deceptive speech from a total of ten male native British English speakers. The data for experiment 2 is 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. The recordings provided control, truthful and deceptive speech from a total of 37 speakers. The speech samples were analysed on a range of acoustic and temporal parameters.

For the majority of acoustic parameters no significant differences or trends can be discerned that would allow for a reliable differentiation between deceptive and non-deceptive speech. The changes that do occur are very minute and not consistent across speakers. The results of the temporal measurements seem more promising with respect to highlighting differences between deceptive and non-deceptive samples. However, it transpired that contextual factors such as type of interview and nature of questions asked have conflicting effects on the temporal aspects of liars' speech. Moreover, context appears to affect the speech of truth-tellers in similar ways to that of liars.

Deceptive behaviour is individualised, very multifaceted and far from being clear cut. The successful development of a method that would reliably and consistently differentiate between truths and lies, based on speech analysis, is highly questionable.

Contents

List of Figures

List of Tables

List of Abbreviations

ADCM	Activation-decision-construction model
AR	Articulation rate
ANS	Autonomic nervous system
BAI	Behavioural analysis interview
CBCA	Criteria based content analysis
CNS	Central nervous system
CKT	Concealed knowledge test
CSC	Columbia/SRI/Colorado corpus
CV	Coefficient of variation
DyViS	Dynamic variability in speech
EEG	Electroencephalography
ERP	Event related brain potentials
F0	Fundamental frequency
F1-4	Formants (i.e. F1 = first formant)
fMRI	Functional magnetic resonance imaging
GKT	Guilty knowledge test
HR	Hesitation rate
IP	Inter-pause stretch
LPC	Linear predictive coding
LTFD	Long term formant distribution
LTF1-4	Long term formants (i.e. LTF1 = first long term formant)
LVA	Layered voice analysis
MCQ	memory characteristic questionnaire
PACE	Police and criminal evidence act
PET	Positron emission tomography
POMP	Percent of the maximum possible score
PSE	Psychological stress evaluator
PR	Pause rate
RM	Reality monitoring
ROT	Response onset time
SCAN	Scientific content analysis
SD	Standard deviation
STAI	State trait anxiety questionnaire
SR	Speaking rate
SVA	Statement validity analysis
VOT	Voice onset time
VPA	Vocal profile analysis
VSA	Voice stress analysers
VSAE	Vowel space area estimation

Acknowledgements

Thank you to my supervisor Professor David Howard for enabling me to embark on this research and for having the confidence in letting me steer it in the chosen direction. His enthusiasm and drive are inspiring and I must thank him for creating broader research opportunities along the way.

I am immensely grateful to Professor Peter French for his continuous support both professional and personal. His encouragement and understanding have been invaluable.

Thank you to my family, friends and colleagues. I will thank each and every one of you in person. I do want to give special appreciation to my mother, Martina Kirchhübel, however. I want to thank her for giving me, amongst many other things, the inner strength to face and master any challenge. Du bist mein Vorbild!

Declaration

This is to certify that this thesis comprises original work and that all contributions from external sources have been explicitly stated and reference appropriately.

I also declare that aspects of the research have been previously presented at conferences and in journals. These publications are listed as follows:

Publications

Kirchhübel, C., Howard, D. M., and Stedmon, A. (2013). Analysing deceptive speech. In D. Harris, (Ed.). *EPCE'13 Proceedings of the 10th International Conference on Engineering Psychology and Cognitive Ergonomics*. Springer-Verlag Berlin, Heidelberg, 134—141.

Kirchhübel, C. and Howard, D. M. (2013). Detecting suspicious behaviour using speech: Acoustic correlates of deceptive speech - an exploratory investigation. *Applied Ergonomics, 44(5), 694—702.*

Kirchhübel, C. and Howard, D. M. (2011) Acoustic correlates of deceptive speech: an exploratory study. In D. Harris, (Ed.). *EPCE'11 Proceedings of the 9th International Conference* *on Engineering Psychology and Cognitive Ergonomics*. Springer-Verlag Berlin, Heidelberg, 28—37.

Kirchhübel, C. and Howard, D. M. (2011) Investigating the acoustic characteristics of deceptive speech. *Proceedings of the 17th International Congress of Phonetic Sciences, Hong Kong*, 1094—1097.

Kirchhübel, C., Howard, D. M., and Stedmon, A. (2011). Acoustic correlates of speech when under stress: research, methods and future directions. *International Journal of Speech, Language and the Law,* 18, 75—98.

Presentations

Kirchhübel, C., Howard, D. M., and Stedmon, A. (2013). Analysing deceptive speech. The *10th International Conference on Engineering Psychology and Cognitive Ergonomics*, Las Vegas, USA.

Kirchhübel, C. (2012). Investigating temporal correlates of deceptive speech. The 21st Annual Conference of the International Association of Forensic Phonetics and Acoustics (IAFPA), Santander, Spain.

Kirchhübel, C. and Howard, D. M. (2011). Investigating the acoustic characteristics of deceptive speech. The 17th International Congress of Phonetic Sciences, Hong Kong (poster presentation).

Kirchhübel, C. (2011). Investigating the acoustic correlates of deceptive speech. The 20th International Conference on Forensic Phonetics and Acoustics (IAFPA), Vienna, Austria.

Kirchhübel, C. and Howard, D. M. (2011). Detecting suspicious behaviour using speech: Phonetic and acoustic correlates of deceptive speech. The *9th International Conference on Engineering Psychology and Cognitive Ergonomics*, Orlando, USA.

Kirchhübel, C. (2010). The acoustic correlates of speech under stress- a review. *AVFA* workshop, York, UK.

1 Introduction

1.1 Background and aim of research

It is acknowledged that information can be gained about a speaker from the speech signal alone. Possible knowledge that can be derived includes, but is not limited to, a speaker's age, sex/gender, regional and social background, the presence of a speech or voice based pathology, voice/language disguise, speaking style (reading vs. spontaneous speech) and influence of alcohol intoxication (French and Harrison, 2006). The voice can also give information about a speaker's mental or affective state. Listening to a third party conversation, lay listeners are usually able to tell whether the speakers are stressed, happy, sad, angry or experiencing cognitive workload, for example. If it is possible to deduce a speaker's emotional condition from listening to his/her voice, could it also be viable to make judgements about speakers' sincerity based on their speech?

The present research explores the viability of using speech analysis as a means of detecting deception. A number of disciplines have been interested in studying the phenomenon of deception and deception detection including psychologists, communication researchers and people involved in security research. Attention has been devoted to the behavioural, verbal and physiological correlates of deception, whilst relatively little research has been carried out on the vocal characteristics of deceptive speech. Nevertheless, more recently, claims have been brought forward concerning deception/stress detection based on voice analysis. Psychological Stress Evaluators (PSE) and Voice Stress Analysers (VSA) are said to measure so-called 'micro-tremors' in the voice which are believed to be physiological indicators of stress and, by extension, of deception. Layered Voice Analysis (LVA) based products are promoted as relying on a link between certain types of brain activity and deception related tremors in the voice (Eriksson and Lacerda, 2007). Scientific reliability testing of these products has resulted in exclusively negative evaluations (Bhatt and Brandon, 2008; Damphousse et al., 2007; Eriksson and Lacerda, 2007; Harnsberger et al., 2009). Despite the lack of scientific support, parts of the non-scientific community, e.g. some local governments, insurance companies, law enforcement agencies, the military and the popular media still believe in the merit of these tools and large amounts of money have been spent buying into these systems (Lykken, 1998, p. 176).

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. As the British Psychological Society (BPS) working group (2004) state: “a reliable procedure is not necessarily a valid procedure, for it might produce consistent but inaccurate results” (p. 9). In order to assess validity, it needs to be established whether a relationship exists between deception and speech, and if so, what the nature of any such relationship is.

Given the rise in VSA based lie detectors as outlined above, one would have thought that the scientific evidence for these was overwhelming. As will become clear in Chapter 5, the reality is to the contrary; a number of studies have analysed temporal aspects of speech, such as speaking rate, pauses, hesitations and speech errors, but only a few studies have investigated frequency based parameters such as fundamental frequency (f_0) and amplitude. Evidence for the analysis of vowel and consonant articulation or voice quality in connection with deceptive speech is almost non-existent.

The main aim of the current research is to investigate various frequency and temporal parameters of deceptive speech in order to assess whether a consistent and reliable relationship exists. This aspect of the research is formalised in research question 1 below:

1. **Can deception be identified based on frequency and temporal characteristics of speech?**

A number of theoretical frameworks have been developed to predict and account for the behaviour that liars may display. Detailed descriptions of each of these can be found in Chapter 2. One of these, the Cognitive Theory of deception, which considers lying to be cognitively more demanding than truth-telling, has been supported by empirical evidence (Walczyk et al., 2003; 2005). Psycholinguistic research on pausing has established that filled and unfilled pauses are behavioural reactions associated with cognitive processes. In the words of Goldman-Eisler (1968):

> "Pausing during the act of generating spontaneous speech is a highly variable phenomenon which is symptomatic of individual differences, sensitive to the pressure of social interaction and to the requirements of verbal tasks and diminishing with learning, i.e. with the reduction in the spontaneity of the process." (p. 15)

Goldman-Eisler's early findings have been corroborated by more recent investigation and Kowal & O'Connell (1980, p. 61) confirm that there is "a lawful relationship between temporal phenomena in human speech and concurrent cognitive processes". Based on this it may be hypothesised that the increase in cognitive load required for deception may lead to a heightened occurrence of speech-planning problems, which, in turn, would manifest themselves in the temporal domain. As a second goal, this study explores the effectiveness of using temporal parameters of speech to differentiate between truth-tellers and liars, and this is specified in research question 2:

2. **Do temporal parameters of speech provide a better means of detecting deception than frequency-based parameters?**

Deception can occur in many different contexts including advertising, sales, workplace, history, interpersonal relations and the legal/forensic setting. This thesis investigates deception within the latter environment (i.e. the criminal arena), and in that sense is concerned with 'serious lies': lies about criminal transgressions including terrorist related plotting and intentions. A critical component of criminal investigations is the gathering of accurate and truthful information by interviewing suspects. Different interviewing techniques have been developed over the years and social psychological research into establishing the most appropriate methods of obtaining information to aid effective law enforcement is thriving (Milne and Bull, 1999; Hartwig et al., 2005). Moston and Engelberg (1993) identified that, generally, two types of interviewing styles were used when police questioned suspects. These are 'information gathering interviews' on the one hand and 'accusatory interviews' on the other. While the former attempts to elicit information using a non-accusatory style of questioning, the latter commonly employs coercive methods to obtain factual details. Chapter 3 will provide a more detailed overview of these two interviewing strategies. The third point of interest of the current research focuses on comparing the latter two interview types with regards to their effectiveness in eliciting temporal speech cues which could be indicative of deception. Research question 3 reads as follows:

3. **Is the use of information gathering interview techniques more effective in eliciting deceptive speech cues than interrogation type interviews?**

A lot of thought has been devoted to identifying behavioural differences between liars and truth-tellers. In many experimental designs deceptive behaviour would be compared against truthful behaviour. These behavioural differences tended to be specified in a unilateral fashion however, with deception being regarded as a deviation from 'normal' behaviour, i.e. the behaviour of truth-tellers. If one seeks more knowledge about the characteristics of speech when people are being deceptive, the main focus will be on the factor 'deception' and a logical starting point would be to design an experiment in which people are enticed to lie and record them while they are committing the deceptive act. Indeed, the majority of previous speech research into deception concentrated on utilising such a paradigm. However, just as it is important to investigate whether there is a relationship between speech and deception, an approach that looked at the behaviour of truth-tellers when placed in similar experimental settings may offer insightful findings, too. It may be that when speaking the truth, people also show behavioural patterns which differ from their usual characteristics. This is particularly conceivable in situations in which telling the truth is of paramount importance such as police interrogation settings. To refer back to research question 2, it was suggested that liars experience heightened cognitive load which could manifest itself in the temporal domain of speech. Truth-tellers may also experience increased cognitive load. Having to retell events that happened in the past could be a challenging exercise. If the stakes are high and people want to make sure they are being believed they may show increasing efforts to portray the truth which, in consequence, could lead to more planning and monitoring in behaviour resulting in more cognitive involvement and by extension changes in temporal parameters of speech. The fourth and final research question addressed in this current study reads as follows:

4. **To what extent does the speech of truth-tellers differ to that of liars when placed in a similar experimental setting?**

1.2 Application of research

A greater understanding of the acoustic and temporal characteristics of speech elicited during situations when one is trying to hide something or lie has potential benefit to a number of professions. Law enforcement agencies, such as the police, would be aided if they could tell if their suspect is lying or telling the truth during an interview. Similarly, intelligence agencies, the military and security officers would benefit from being able to use 'reliable' means of detecting deception (Hirschberg et al., 2005).

Although the current research focuses on deception as it occurs in the criminal domain, the scope of the present findings may very well inform other areas where deception could take place. Mental health practitioners assessing the risk of suicide in their patients would profit from the ability to distinguish if patients are giving a true account of their inner state or are pretending to feel well in order to be released from the clinic. Being able to detect deception would be advantageous to the realms of politics and business as well. Investigations into social lies, or so called 'white lies', could be of benefit to communication researchers aiming to enhance our present knowledge of the workings of social interaction and social relationships.

The academic research community in phonetics and linguistics and, in particular, the forensic speech science community would also profit from further theoretical and practical advancement on the topic of speech and deception. Forensic Speech Science is a relatively young but growing field of research, applying knowledge gained from linguistics, phonetics, dialectology and sociolinguistics, in combination with expertise from speech and sound analysis for use in legal investigations and proceedings. Forensic speech scientists are sometimes asked by instructing parties (e.g. police officers, lawyers) to make assessments of the sincerity of suspects in recorded police interviews. Similarly, they may be asked to give an opinion on the genuineness of recordings containing bomb threats or recordings in which victims/witnesses report violent attacks. As very little is known about how deception is manifested in voice and speech such evaluations are currently prohibited by the International Association of Forensic Phonetics and Acoustics (IAFPA code of practice, clause 9).

1.3 Research approach

The four main research questions under investigation are re-iterated below:

1. **Can deception be identified based on frequency and temporal characteristics of speech?**

2. **Do temporal parameters of speech provide a better means of detecting deception than frequency-based parameters?**

3. **Is the use of information gathering interview techniques more effective in eliciting deceptive speech cues than interrogation type interviews?**

4. **To what extent does the speech of truth-tellers differ to that of liars when placed in a similar experimental setting?**

An eclectic approach was employed in order to answer these research questions. In particular, methodologies from the fields of phonetics, speech research and psychology were combined during the data collection and analysis stages. Research into speech and deception is a relatively unexplored area and in the design of an empirical study the following criteria need to be considered:

(a) Fully controlled experiments, where variables can be strictly manipulated
(b) Clean, high quality recordings with a sufficient amount of analysable tokens
(c) A within-speaker paradigm - control and deceptive speech from the same speaker
(d) Knowledge of ground truth.

When embarking on such research, there are merits in collecting one's own database. Not only does this enable one to tailor the experimental design according to one's wishes, it may also alert the researcher to aspects of the research which were not considered in the planning stages. Despite the advantages connected to the self-collection of corpora, there are practical benefits in using existing speech databases. A small number of databases were identified which could potentially provide a basis for deception related speech research. The following provides a brief overview of these along with the reasons for their inclusion or exclusion.

1.3.1 Columbia/SRI/Colorado corpus

The Columbia/SRI/Colorado (CSC) deception corpus is one of the few corpora containing high quality deceptive and non-deceptive speech. It was recorded by Julia Hirschberg and her colleagues for the purpose of examining the feasibility of detecting deception from speech using automatic means. Deceptive and non-deceptive speech was collected from 32 speakers of Standard American English. Participants had to perform a number of activity- and question-based tasks after which they were interviewed about their level of performance in each. The design was arranged in such a way that speakers were required to tell the truth about their achievement for some of the activities but lie about their score in others. The interviewees were expected to indicate whether their statements were true or false by pressing one of two pedals that were hidden underneath the interview table. Using this rather innovative strategy, the authors managed to distinguish between what they call 'global truth/lie' and 'local truth/lie'. 'Global' refers to the overall intention to deceive or not deceive while 'local' refers to the individual statements used to achieve global truth/lie. Motivation to deceive successfully was induced by financial as well as social incentives. A thorough description of the corpus can be found in Enos (2009). This corpus would have provided for an excellent source of deceptive speech but, unfortunately, it has not yet been made available to the research community at large.

1.3.2 Idiap Wolf corpus

The Idiap Wolf recordings are part of an audio-visual corpus which was collected at the IDIAP research institute, Switzerland, by Gokul Chittaranjan and Hayley Hung. It represents the first database that captures deceptive speech in a multi-speaker group scenario involving a conversational role playing game called 'Are you a werewolf?' (Plotkin, 1997). It is a game of suspicion, accusation, lying and second-guessing in which players are randomly divided into villagers and werewolves as well as one narrator. During the 'night phase' of the game, in which the villagers are asleep, the werewolves 'kill' one of the villagers. During the 'day phase' the narrator will reveal which villager has been murdered and following this all players, including the werewolves, debate and discuss the identities of the werewolves and jointly vote which player to eliminate. The game ends when all werewolves have been detected or when the werewolves outnumber the villagers. Some additional novelty of the database rests in the fact that it does not only allow for the investigation of deceptive/non-deceptive behaviours but also which behavioural actions, including speech, evoke suspicion. Those playing the roles of werewolves would be classed as liars and those acting the part of the villagers portray non-liars. Players who are 'killed' in the course of the game may be regarded as behaving suspiciously while those who survive may be seen as appearing trustworthy. Using head-mounted microphones, high-quality audio data was collected from four groups playing fifteen games in total. Groups had an average number of ten players per game of whom two were always werewolves. In addition to the game recordings the players were also recorded in non-game conditions which provides for baseline data. The reader is directed to Hung and Chittaranjan (2010) for a more thorough description of the corpus which has been made freely available to researchers.

A corpus that is freely available is rare and, indeed, it appears as though the speech collected would present for an interesting study. However, it was believed that the multi-speaker environment was too complex for the purposes of the present study. As indicated, the contextual focus of this work revolves around the criminal setting and, specifically, police or security interview environments. These tend to be dyadic or triadic interactions rather than group based and for this reason it was decided not to work with this corpus at the present moment in time.

1.3.3 DyViS

The DyViS database (Dynamic variability in speech) was produced by a UK funded research project led by Francis Nolan and his team at the University of Cambridge (Nolan et al., 2009). The project arose from the objective of exploring dynamic variability in speech from two angles. One of the focal points was to assess the effectiveness of dynamic features of the speech signal such as sound- and, in particular, formant transitions in characterising speakers. The second core interest concerned the wider aspect of sound or linguistic change and how speakers may be differentiated based on the nature of their participation in such change. Specific attention was placed on those phonetic variables that were thought to be undergoing change in the speech community under question. A full overview of the DyViS project can be found at http://www.ling.cam.ac.uk/dyvis/index.html.

Despite the fact that the database was not inspired by the desire to study deceptive speech as such, it does have the potential for a variety of research purposes. It contains high-quality audio recordings from 100 male Standard Southern British English speakers communicating in a number of different contexts, conditions and styles. Of particular interest for the goal of examining deception in speech would be the mock police interview in which participants were questioned about a drug-trafficking incident they were involved in. The interview employs an information gathering style and it presents a situation of 'cognitive load' on the part of the interviewees who are required to cover up certain facts (Nolan et al., 2009). The interview closely replicated a real-life scenario and it is safe to say that participants were sufficiently involved. One aspect of the design, however, complicated its use for research into deceptive speech. Participants were asked to act and pretend that they were involved in a crime. It could be argued that the task of acting itself could be seen as deceptive. In addition to acting out their given role, participants were expected to lie during the police interview. So, to a certain extent, they were being doubly deceptive (Gea de Jong, personal communication). Stated another way, when participants were telling the truth they were actually lying, and vice versa. For example, when the participants deny knowing one of the other suspects, their answer is in conflict with the scripted scenario guiding their responses - but of course their denial is true outside of the experiment (Francis Nolan, personal communication). After some contemplation the decision was made not to proceed with analysing this corpus in the context of the present work.

1.3.4 'Smell of Fear'

The 'smell of fear' database grew out of a larger research project looking into the biological, physiological, psychological and behavioural factors associated with the act of concealing knowledge (Eachus et al., 2013). In a controlled laboratory study involving multiple interviews, participants were required to conceal the possession of an object in return for a financial reward. The interviews followed a scalable interrogation protocol with twenty yes/no questions for each session. Initially, participants completed a baseline interview which included demographic- and hobby- related questions. This was designed to collect control data from each participant. The two subsequent interviews employed an accusatory style and were structured in such a way as to increase participants' levels of arousal. The questions comprising the first interview focused on social desirability and trustworthiness while the second interview contained direct probing about possession of the object in question. In order to achieve consistency in representation of the questions they were pre-recorded and then presented over a tape recorder. As part of the behavioural measures collected, 38 male native English speakers were audio and video recorded during the three interview sessions. A more detailed description of the database can be found in Eachus et al. (2013) and Chapter 6 of the present work.

Several aspects of the design of the corpus presented shortcomings for research into speech related factors. Firstly, concealment of the object was treated as a between-subjects variable. Half the number of participants (liars) received the object and therefore had to deceive the interviewer whereas the other half (truth-tellers) was not in the possession of any object which cancelled the need for deception for that group. Given the speaker dependent nature of speech and voice, the two groups could not be compared against each other. Secondly, it transpired that there were limitations with the quality of the recordings. Participants did not wear headset microphones and basic recording equipment (i.e. standard tape recorder) was used to make the speech recordings.

The between-subjects paradigm cast doubts on whether this database should be considered for analysis. However, the presence of the baseline interview enabled 'within group - within speaker' comparisons and therefore, what initially seemed a major limitation, became less of a problem. In fact, the between subjects paradigm actually presented an advantage in that it enabled research into the behaviour of truth-tellers when placed in the same environment as liars which is the core of research question 4.

Out of the four corpora that could be located, only three were accessible for research purposes. The Idiap and DyViS corpora met all the criteria set out in (a) – (d) above. However, they presented aspects (namely, multi-group setting and acting), which were thought to be too complex for the purposes of the present work. The '*smell of fear*' dataset did not meet all the desirable criteria but it was nevertheless deemed to be appropriate for the purposes of the present work and was therefore selected[1]. In particular, the accusatory style of interviewing and the inclusion of a group of truth-tellers will be valuable in relation to addressing research questions 3 and 4. Despite the selection of the '*smell of fear*' corpus, another set of deceptive speech data was vital in order to satisfactorily address the four research questions. It was decided, as part of the present work, to record a new corpus of deceptive speech which would meet all the criteria set out in (a)-(d) and which employs an information gathering interview style. A detailed description of this corpus is presented in Chapter 6.

[1] Acknowledgements to Alex Stedmon for making the recordings available for the purposes of the present research.

1.3.5 Real-life data

In addition to looking into using already existing corpora, consideration was also given to analysing real-life data. Undoubtedly, authentic deceptive speech has more ecological validity than deceptive speech collected in the 'fake' setting which is the laboratory. Under the permission of J P French Associates[2], relevant archived data from authentic forensic casework was identified and evaluated. Unfortunately, it transpired that ground truth could not be reliably established for the recordings in question and it was decided not to pursue this line of analysis.

It is acknowledged that laboratory research raises shortcomings in terms of practical applicability. However, while it is important to emphasise the limitations of laboratory based research, its value needs to be highlighted as well. Not only does a laboratory setting provide the opportunity to study a range of parameters, it also allows for tight control over variables. The research on the acoustic and temporal aspects of deceptive speech is sparse. Any findings of laboratory based investigations should be seen as giving background knowledge and a starting point for research into less controlled environments. If robust differences between control and deceptive speech are found in these ideal conditions, research can then move on to investigating more natural situations in the field.

1.4 Structure of thesis

This chapter introduced the reader to the aim of the study, the overall approach taken and its practical as well as theoretical value. Chapters 2, 3, 4 and 5 cover the theoretical background to the study. Chapter 2 introduces the reader to the concept of deception and highlights various factors that may affect research into deceptive behaviour. Chapter 3 covers the area of deception detection introducing traditional as well as emerging research techniques. Chapter 4 attempts to combine the theoretical knowledge of deception presented in the previous chapter with knowledge that has been gained from speech research. In particular, an overview is given of the literature on speech under stress, cognitive load and attempted control and attempts are made to highlight how these findings are of relevance to deceptive speech research. Chapter 5 then narrows the focus down to summarising previous research studies that have investigated acoustic and/or temporal parameters of deceptive speech. Chapters 6,

[2] The UK's longest established independent Forensic Speech and Acoustics laboratory.

7 and 8 present the empirical research which underlies the current thesis. Deceptive speech was analysed in two different communication environments by way of two laboratory based experiments. The findings were subject to various quantitative statistical testing and the speech variables under investigation were theoretically motivated and their genesis was explained in Chapter 4. Chapter 6 provides a thorough description of the methodological design of each of the two experiments. Chapter 7 cover the acoustic and temporal analysis of experiment 1 and Chapter 8 presents the findings of the temporal analysis of the liars and truth-tellers from experiment 2. Chapter 9 discusses the results in light of the four research questions posed and closes the study by highlighting future research ideas.

2 Deception

People have been interested in lies and how they could be detected since ancient times. A comprehensive and illustrative account of the history of deception detection is given by Trovillo (1939a; 1939b). Through 50 pages the reader is introduced to the methods of the Ordeal and Torture as well as the early attempts to detect deception using physiological cues. While the historical account of the psychological and philosophical nature of human deception is a fascinating topic in its own right it is not the focus of this thesis. Rather, the bulk of attention of the present work concerns the comparatively younger scientific study of human deception.

The present chapter will present preliminary background information on the general nature of deception research including definitions of the concept of deception as pertinent to the present study, types of and motives for deception, and factors that have been found to affect the behaviour of liars, i.e. moderator variables. Furthermore, the theoretical foundation for research on deception generally and as pertaining to the current work is highlighted. Once the broader issues surrounding deception research are outlined the focus is then narrowed down to specifically describe the nature of empirical investigations into deceptive behaviour. To this end, a brief overview is given of the methodological designs employed and the findings of studies looking into the behavioural cues to deception. The chapter closes by shifting the reader's attention to the central aspects of the work: the speech related cues to deception.

2.1 Definition of deception

In the words of the philosopher Montaigne deception 'has a hundred thousand forms, and a field indefinite, without bound or limit' (Montaigne Essays, 1877). Deception and lying have presented philosophers with moral and conceptual questions for centuries and Carson (2010) provides a comprehensive overview of the nature of human deception. Lying and deception are often used interchangeably and due to practical considerations both terms will be employed reciprocally in this thesis, too. It is important to understand the difference between the two concepts, however. For lying to take place, it is necessary to make a false statement. Deception, on the other hand, may involve no false statements at all. Indeed, deception may be achieved without using any form of oral message. Secondly, deception implies success; lies, however, may not be successful in misleading the target. Investigating deception and lying from a scientific point of view is a relatively recent phenomenon and it is essential to establish a clear conceptualisation of what exactly is investigated when embarking on this type of research. There are broad and narrow concepts of deception and rather than talking about *the* definition of deception, it may be more appropriate to view the concept of deception as revolving around a space of plausible definitions (Carson, 2010, p. 39).

Masip et al., (2004) provide a thorough theoretical dismantling of the different aspects contributing to the definitions that have been formulated so far. Prototypical lies, as reported by Coleman and Kay (1981), contain three elements and these are, in order of importance, the objective falsity, the sender's belief in this falsity and the sender's intention to deceive. In line with this, contemporary social-scientific definitions of deception tend to entail all or parts of these semantic components. As the reader may be able to imagine, definitions of deception are plentiful and rather than merely listing each and every one it was decided to highlight those which have proven to be most influential within social science research.

The definitions by Krauss (1981) and Miller (1983), stated below, highlight the importance of the concept of intention and belief[3] in order to exclude cases where deception is the result of pathological behaviour or where it occurred due to ignorance, error or failing memory. They also underline the communicative aspect of deception thereby eliminating self-deception from the working definition.

> "message distortion resulting from deliberate falsification or omission of information by a communicator with the intent of stimulating in another, or others, a belief that the communicator himself or herself does not believe" (Miller, 1983 as cited in Miller, 1993, p. 20).

> "an act that is intended to foster in another person a belief or understanding which the deceiver considers to be false" (Krauss, 1981 as cited in Vrij, 2008, p. 13).

There are aspects which the definitions fail to exclude however, namely cases where people expect to be fooled such as magicians' shows and poker games. Ekman (1985) accounts for this by defining lying:

[3] This feature of deception is often disregarded by dictionary definitions.

"as a deliberate choice to mislead a target without giving any notification of the intent to do so." (p. 41).

Vrij (2008) refines this by acknowledging that lies may not be successful:

[Deception is] "a successful or unsuccessful deliberate attempt, without forewarning, to create in another a belief which the communicator considers to be untrue." (p. 15).

Masip et al. (2004) have arrived at an integrated yet rigorous definition of deception which takes into account the views of the most prominent and influential scholars within the field. Employed in the current work, their definition of deception reads as follows:

"the deliberate attempt, whether successful or not, to conceal, fabricate, and/or manipulate in any other way factual and/or emotional information, by verbal and/or nonverbal means, in order to create or maintain in another or in others a belief that the communicator himself or herself considers false." (p. 148).

2.2 Types of deception

There are several ways deception can be achieved and various typologies have been suggested to describe the different strategic means (Ekman, 1985; Hopper and Bell, 1984; O'Hair and Cody, 1994; Turner et al., 1975). People may conceal information or emotions. For example, a researcher might not tell a participant that he/she is being recorded or a criminal might try to conceal his knowledge of crime-relevant information in order not to be found out. A more active form of deception is falsification which requires creating a false account of an event or of personal feelings, for example. A suspect might present a false alibi or an interviewee for a competitive job may falsify his personal traits and experience. Other types of lying include half-concealments, i.e. leaving out crucial parts of a story for example, or telling the truth falsely by exaggerating it. A police officer working undercover in a drug circle might blatantly confirm the suspicion that he is an informant and invite members of the gang to search him. He thereby ridicules the distrust of the gang members. Yet other techniques involve misdirection, i.e. acknowledging an event or feeling but lying about how the event came about or what caused the feeling, and what Ekman (1985) termed the "incorrect-inference dodge" (p. 39). Ekman

(1985, Chapter 2) provides a very accessible, illustrative and detailed description of the different types of lies.

Masip et al. (2004) make an interesting point in that the distinction between emotional and factual lies might not be as straightforward as initially assumed. Although from a theoretical perspective it is useful to regard both as separate categories, in practice one should think of an emotional-factual continuum. Real-life deception is not simply characterised by a single deceptive utterance but should be viewed more as a network comprising multiple, simultaneous lies (and truths) with supplementary deceptive messages (emotional and factual) accompanying. Metts (1989) places the different types of lies along a continuum ranging from complete omission of information at one end to absolute fabrication on the other. The intermediate points of this continuum correspond to deceptive messages of different degrees of distortion or misrepresentation.

A study by DePaulo et al. (1996) investigated lying in everyday life by asking participants to keep a 'lying diary' for one week. It was revealed that, in everyday situations, lying is a frequent activity which usually does not entail preparation or strong feelings of guilt. Furthermore, the results showed that the majority of lies people told were falsifications. It could be argued that in the context of police interrogations however, passive lying e.g. using concealments, would be preferable than active lying. One reason for this could be seen in the fact that concealing information tends to be easier to conduct as it requires less cognitive capacity compared to falsifications (Vrij and Heaven, 1999). Concealments are also more easily justified should they be detected by way of claiming forgetfulness or inattention (Metts, 1989). On the other hand, police interviews tend to be encounters involving a large amount of information seeking which in turn could lead to the necessity of using falsifications to a greater extent. Although laboratory research benefits from drawing a distinction between the different types of deceptive strategies, e.g. concealments vs. fabrications, in real-life deception concealments and fabrications will go hand in hand as will emotional and factual lies (Ekman and Frank, 1993).

2.3 Motives for deception

Usually lying is stigmatised; stereotypically, it is seen as a selfish act that is committed in order to gain personal advantage or to avoid negative consequences. While deception indeed is an objectionable and immoral action in the majority of situations, there are cases where the use of deception could be viewed as appropriate and justifiable. Everyday lies, or *altruistic lies* tend to serve the protection and maintenance of social relationships. For example, in their study of information control in everyday conversations Turner et al. (1975, p. 83) discovered that "honesty is not always the best policy" but that deception is an important and frequently used discursive element. Similarly, Camden et al. (1984) reported positive incentives in connection with the use of every-day lies including protecting self-esteem and managing relationships. Beyond its functional use in interpersonal communication, deception is pervasive in numerous other areas such as law enforcement or the medical and political environment, for example. As pointed out, the focus of the present investigation revolves around deception as it occurs in a more high-stakes criminal setting rather than the white lies we find in close relationships and, therefore, it may be expected that the motives for the use of deception are more self-serving and carry less of a relational/interactional function.

Motives for deception have been dichotomized in various ways and a comprehensive summary of the various taxonomies of motives that have been suggested can be found in O'Hair and Cody (1994). Goffman (1974) distinguishes between 'benign fabrications' which serve the benefit of others as opposed to 'exploitative fabrications' which are guided by self-interest. O'Hair and Cody (1994) developed a two-dimensional categorization taking into account 'valence', i.e. the goal of the deceptive act (negative or positive), and 'target', i.e. the recipient of the deceptive act. More recently, Vrij (2008) developed a classification of reasons for lying according to three dimensions: (a) self-centred vs. other-centred lies, (b) to gain advantage vs. to avoid loss, (c) for materialistic vs. psychological reasons. Self-oriented lies, as the name suggests, are told for one's own benefit. Examples include a car owner who does not disclose all the faults to the potential buyer, a job applicant who exaggerates his experience in order to secure the vacancy or a suspect who denies involvement in a crime. Lies may be told for several reasons and are, therefore, not restricted to only one category. However, deception as it occurs within the realms of politics and business may be fuelled by instrumental, i.e. financial reasons more so than crime or terrorism related deception which tends to be framed by psychological and ideological motivations (Fussey, 2013).

2.4 Factors affecting deception

Various factors, situational as well as individual, could influence deceptive behaviour and in turn the magnitude and transparency of deceptive cues. DePaulo et al. (2003) discovered that the salience of deception cues depends on the motive for lying. Cues are more prominent when attempts are made to cover up criminal transgressions rather than when lying about opinions or emotions. Motivation for lying is another factor which has an effect on the deceptive behaviour displayed. In their meta-analysis, Zuckerman et al. (1981) showed that highly motivated liars revealed more deception cues than liars who were not as involved in their deceit. DePaulo et al. (1988) and DePaulo & Kirkendol (1989) explain this in terms of 'the motivational impairment effect'. Motivation may lead to the increased planning and preparation for the deceptive act which in turn could result in behavioural over-controlling. On the other hand, a high level of motivation and involvement may lead to a failure to suppress possibly suspicious behavioural cues. Motivation is judged to be higher if it is identity-related or ideological instead of instrumental (DePaulo and Kirkendol, 1989).

Not just the motivation of the deceiver but also the eagerness or suspicion of the 'lie-catcher' may affect the deceiver's performance (Buller et al., 1989). Deceptive circumstances may differ with respect to context and complexity of the lie. Deceivers may have had the opportunity to plan and rehearse their deceptive account; other occasions might necessitate spontaneous lying. Indeed, there might be people who are trained and experienced in lying. Liars might face the requirement of having to fabricate a story; or they may 'merely' need to conceal information. Complicated lies, i.e. those that require spontaneity and falsification, tend to exhibit more deceptive behavioural patterns than easier lies such as concealments or well-rehearsed lies (Vrij and Heaven, 1999). Furthermore, the perceived chances and consequences of getting caught, commonly referred to as 'stakes', vary from situation to situation and influence the nature of the deceptive act. The majority of every day lies are trivial and do not bear serious consequences if discovered. In other words, the stakes are low. Lies occurring in the legal setting on the other hand are usually connected to severe punishment if unsuccessful, i.e. they are high-stake lies. It is conceivable that there is heightened emotional and cognitive involvement in high-stake deception as compared to low stake situations and this may result in more or differing deceptive cues in the former.

Sporer and Schwandt (2007) present a summary of the effect of various moderator variables on deceptive behaviour, including motivation to deceive, content of lie, time to prepare the lie, whether the lie was sanctioned or not and what they call operationalisation, i.e. how the data was analysed. All variables were found to have an impact on the behaviour of liars.

The personality of the deceiver has been shown to contribute to the manifestation and expression of deception cues. Extroverts and people possessing good acting ability have been shown to convey fewer deception cues than introverts or 'bad actors' (Riggio and Friedman, 1983; Siegman and Reynolds, 1983). A reasonable conjecture would be to assume that people scoring high in Machiavellianism or Public Self-Consciousness (PSC) would be more successful at hiding deceptive behaviour due to the absence of a feeling of guilt or due to a heightened sense of self-awareness. For example, people high in PSC may be more aware of the link between frequency of body movement and level of physiological arousal. If they believe that heightened level of physiological arousal is indicative of deception, they might try to suppress their body movements. Vrij et al. (1997) showed that participants with high PSC ratings showed decreased hand movement during deception compared to those scoring low on this trait. Vrij (2008, p. 73) points out however, that evidence for a consistent relationship between different types of personality and deception has so far been inconclusive. As pointed out by Ekman and Frank (1993) the success of attempted deception might also be related to the liar's level of intelligence.

2.5 Perspectives on deception

Initially research on deception was motivated by the wish to locate a cue or behaviour that would reliably indicate that a deception was taking place. It was idealised that analogous to Pinocchio's growing nose in fiction there would be a trait or sign indicating deception in reality. This has proved particularly problematic and researchers started to acknowledge that lying is complex and flexible (Bond et al., 1985; Kraut, 1980). Instead of searching for behavioural correlates of deception per se, more recently researchers have begun to investigate the emotional, cognitive and communicative processes that tend to accompany deception. A number of theoretical frameworks, which are described in turn below, have been developed to predict and account for the behaviour liars may display. Detailed descriptions can be found in DePaulo et al. (2003), Ekman (1985), Miller and Stiff (1993, pp. 52-55) and Vrij (2008).

2.5.1 Multifactor model (Zuckerman et al., 1981)

2.5.1.1 Emotional Approach

Arousal Theory assumes that liars are more stressed and aroused than truth-tellers. The cause for this arousal is thought to be the presence of heightened emotions, the most common being fear, guilt and 'duping delight' (Ekman, 1985; Ekman and Frank, 1993). Liars may feel guilty for a transgression which they are trying to hide or simply for the act of lying itself. They may experience fear of getting caught or of the consequences if their lies did get discovered. They may also experience excitement in the prospect of fooling someone. The strength of the emotional involvement may be reflected in liars' demeanour and the stronger the emotions experienced the more likely they will leak through (Ekman, 1985). This is referred to as leakage and was first introduced by Ekman and Friesen (1969). As well as leakage, they also introduced the term 'deception clues' which incorporates the various behavioural alterations, in the face, body or voice, by which deception could be perpetrated. There are weaknesses with this theory in that there is no direct link between being deceptive and being emotionally aroused or stressed (Lykken, 1998). 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 might rather display a composed and calm countenance. As the following quotation 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. cited in Lykken, 1998, p. 102)

Based on the play by Shakespeare, Ekman (1985) refers to this as the Othello error. Essentially, Othello Error occurs "when the lie catcher fails to consider that a truthful person who is under stress may appear to be lying" (Ekman, 1985, pp. 169–70). A second error Ekman (1985) presents is the idiosyncrasy error or 'Brokaw Hazard' which signifies instances where the lie catcher fails to take into account individual differences in behaviour shown when lying. This highlights the inevitable requirement for establishing baseline behaviour before attempting lie detection.

2.5.1.2 Content Complexity Approach

The foundation of the content complexity approach rests on the fact that lying can often be cognitively more demanding compared to truth-telling. Deceivers need to suppress the truth, which is automatically activated. In order to get away with their lies deceivers need to make their stories plausible and consistent and they constantly need to monitor their own and their target's behaviour. Yet, within the context of social relations lying might be cognitively less charging compared to telling the truth as, for example, illustrated by insincere compliments. As with emotional arousal, there may be situations which prompt hesitation and uncertainty in truth-tellers and therefore caution needs to be taken in how behavioural signs are interpreted. In addition to this, personality traits may determine a person's non verbal appearance. Schlenker and Leary (1982), for example, found that socially anxious people are naturally slower in their responses.

2.5.1.3 Attempted Control Theory

People have stereotypical views about how liars behave (Hocking and Leathers, 1980; Strömwall et al., 2004). 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 control may be conscious but can also be the result of subconscious mechanisms. Paradoxically, this type of control could result in a less natural, more rigid appearance. In order to make an honest impression deceivers need to be able to manage their behaviour. Ekman et al. (1991) predict that people can control different communication channels to differing degrees. While they tend to be skilful at controlling the face because the face is such an important means of communication they are likely to be less adept at controlling the body or voice however. Once more it has to be pointed out that signs of control are not necessarily unique to attempted deception. Truth-tellers may well adopt a controlled and careful demeanour in order to ensure that their message is judged to be sincere.

2.5.2 Self-presentational perspective (DePaulo et al., 2003)

In 1992, Bella DePaulo and her colleagues first described the self-presentational perspective to explain non-verbal communication. In subsequent years this framework was applied for the prediction of possible cues to deception. Self-presentation in DePaulo's work is defined as 'peoples' attempts to control the impressions that are formed of them' (DePaulo et al., 2003, p. 77). Crucially, within this perspective, both truth-tellers and liars are seen as working towards the same goal, which is, to portray honesty. It highlights the similarities between truthful and deceptive interactions and accounts for what the emotional approach failed to acknowledge, i.e. the possibility that truth-tellers may fail to appear honest. The important difference between truthful and deceptive messages however, is that liars do not have the right to warrant truth. Liars' awareness of their illegitimate claims of honesty may affect their deceptive presentations in a number of ways. Firstly, due to possible moral apprehensiveness they may be less emotionally involved with their account, or even experience negative feelings. As a consequence liars might appear tense, distant and unpleasant. As a result of a lack of familiarity with the events and also in order to avoid contradiction, they may also, at first glance, portray reduced cognitive investment; their answers may come across as elusive and vague. In fact, given the continuous self-management, liars are likely to experience a cognitively demanding situation. Secondly, deceptive self-presentations may bear a sense of deliberateness whereas truthful ones would be expected to develop more effortlessly. Behavioural and cognitive self-monitoring and regulating may put strain on the liar and further enhance the impression of tension and amplified control.

2.5.3 Intercultural deception theory (Buller and Burgoon, 1996)

Developed in 1996 by Buller and Burgoon, this theory emphasises the importance of the relationship between sender and receiver and how interpersonal communication factors such as expectations, goals and motivation may modify the deceptive interaction. As well as monitoring their own behaviour, liars will also observe their target's reactions for signs of suspiciousness or distrust and, if thought necessary, they will adjust their verbal and non-verbal display. At the same time, targets or receivers will assess the performance of their communication partner and, in turn, amend their credibility judgement. It is hypothesised that detection apprehension on the deceiver's part is increased if the target of deception is thought to be a skilled lie catcher. Deception detection accuracy on the receiver's part is enhanced by his/her ability to decode a message as well as the sender's deception skills. Liars who experience increased detection apprehension will demonstrate heightened *nonstrategic* behaviours, i.e. cues related to arousal, cognitive load and attempted control. In addition, they may also exhibit *strategic cues* which are related to advanced information-, behaviour- and image management.

While all of the theoretical frameworks mentioned have been substantiated by empirical evidence (Vrij and Heaven, 1999; Vrij et al., 1996; Walczyk et al., 2003; 2005) some have received a greater strength of support than others. Furthermore, it needs to be stressed that no single approach is sufficient in explaining deceptive behaviour. Although the different deceptive processes where summarised in a separate fashion above, it ought to be noted, that they should not be viewed as mutually exclusive. Rather, it would be more appropriate to expect that, if they occur, they occur simultaneously and to varying degrees. As a consequence, we might find contradictory behavioural effects during deception.

2.6 Empirical research on deception

2.6.1 Experimental designs

One of the most challenging issues in empirical deception research is the design of an ethically approvable yet scientifically sound and valid experiment. Not only would it be beneficial if the study were based on a relevant real-world high-stakes scenario, it is also necessary that ground truth is known for every statement made. Controlling for possible individual differences due to personality, i.e. high anxious vs. low anxious, could be seen as another crucial part of the methodology when investigating deception.

Due to ethical considerations the majority of research so far has been conducted in the laboratory. Various strategies have been employed in eliciting deceptive behaviour and for ease of reference it was decided to categorise them into three major groups:

- 'false opinion paradigms' = lying about feelings, opinions and attitudes
- 'false information paradigms' = lying about the possession of information or objects
- 'false event paradigms' = the completion of mock/simulated crimes

'Topics' participants were asked to lie about tended to revolve around their emotions (Ekman and Friesen, 1974; Ekman et al., 1976; Stiff et al., 1992), their opinions and attitudes about various issues and people (Arciuli et al., 2009; DePaulo and Rosenthal, 1979; DePaulo et al., 1982; Knapp et al., 1974; Mehribian, 1971) and personal information such as holiday preferences (Cody and O'Hair, 1983; Cody et al., 1984). One of the problems with designs that require people to lie about their emotions and feelings relates to the difficulty in reliably differentiating between behaviours due to deception and behaviours resulting from emotional involvement with the stimuli, per se. In addition to this, lying about feelings also entails difficulties in establishing the ground truth. It is impossible for the experimenter to control and measure the true emotions experienced and therefore he/she has to rely on participants' integrity. In order to overcome differential involvement researchers have asked participants to lie about more neutral stimuli such as pictures (Anolli and Ciceri, 1997; Riggio and Friedman, 1983), events in videos (Vrij and Heaven, 1999; Vrij et al., 2001) and the possession of objects/things (Höfer et al., 1993; Vrij, 1995; Vrij and Winkel, 1991). After completion of the assigned task deceivers and non-deceivers were usually interviewed. The interviews may take the form of short-answer based questions or longer sessions akin to a police interview setting. Participants' truthful and deceptive responses were video- and/or audio-taped enabling analysis of various behavioural cues.

As Miller & Stiff (1993, pp. 42-43) point out, one short-coming with these types of studies is the fact that deception has been sanctioned by the experimenter. In real life this is hardly the case. The experimenter's approval may augment participants' detection apprehension which, in turn, would affect the presence, or better, the absence of behavioural cues. To facilitate the opportunity for people to choose whether to lie or tell the truth, some studies adopt what Miller and Stiff (1993) termed the 'Exline procedure' (DeTurck and Miller, 1985; Exline et al., 1970; Feeley and DeTurck, 1998; Frank and Ekman, 1997; Mehribian, 1971; Stiff and Miller, 1986). In this approach, participants are asked to complete a cognitive task, such as estimating dots or unscrambling anagrams, with a partner who was usually a confederate of the experimenter. The confederate either truthfully worked on the task with the partner or he/she prompted the participant to cheat. In subsequent interviews participants were asked to describe the methods they had used to complete the tasks. Those who were part of the deceptive condition could make a choice whether to admit their cheating or whether to lie about their strategies used. Although the Exline procedure controls for 'intent to deceive', one of the drawbacks of it relates to the unethical nature of deceiving participants.

Recognising the constraints of the laboratory some researchers attempted to collect controlled data from more natural settings. Strömwall et al. (2006) asked their participants to take part in a simulated crime and a subsequent police interview concerning the criminal offence.

Researchers have employed various means in order to increase the stakes including rewards, usually financial in nature, for successful deceit and punishments (Frank and Ekman, 1997; Mehrabian, 1971) for failed deception. Others have motivated participants by heightening their sense of self-presentation/self-awareness. For example, Ekman and Friesen (1974) and Ekman et al. (1976) manipulated the motivation of their participants (who were all nurses) by reminding them that the ability to achieve successful deception is of significant relevance to their chosen career.

Despite these attempts, laboratory studies will never be able to recreate the stakes involved in real-life deception situations and the question of ecological validity is paramount. Largely due to practical constrains the majority of studies have resorted to the student population in order to collect their participants (Edelmann, 1998). The question arises whether deception as employed by lay persons and the student population resembles real-life deception of skilled and trained liars. Eager to address this, researchers have looked for ways to study real-life deception. To this extent, Davis and Hadiks (1995) analysed the interview between Saddam Hussein and a CNN journalist during the Gulf War (1991) and Hirsch and Wolf (2001) investigated parts of Bill Clinton's testimonial to the Grand jury concerning his alleged affair with Monica Lewinsky. While it can be assumed that these are examples of high-stakes deception, these types of studies yet bear a different weakness; ground truth could not be established objectively and, consequently, it was unknown whether and when in the testimony/interview deception took place.

Mann (2001) and Davis et al. (2005) had the opportunity to examine real-life police interviews with criminal suspects (arson, robbery) in which ground truths had been established by independent evidence such as forensic evidence and witness accounts. Similarly, Vrij and Mann (2001) analysed truthful and deceptive fragments from recordings of a murderer during his police interviews. Again, reliable truths could be established via extramural evidence. Perhaps surprisingly, a comparison of objective cues from laboratory induced deception with actual cues observed in the murderer during his deceptive police interview revealed a large amount of correspondence, albeit, the cues were more eminent in the real-life setting (Vrij, 2008, p. 87). Although few, these real-life studies have given us an insight into the behaviour of liars, who use deception by their own volition in high-stakes real-life settings.

As stated above, it may be assumed that the majority of everyday lies, so called 'white lies' are low stake lies whereas lies related to criminal transgression or terrorist intent are high-stake lies. This inference presupposes that criminals and terrorists perceive the stakes to be high. It presupposes that those with criminal intent experience detection apprehension and are concerned about the consequences of their criminal activity. The literature on terrorism has shown however, that terrorists are not afraid of punishment or death; they do not regard suicide attacks as a risk but rather view it as an opportunity to accomplish their ideological goals (Hoffman, 2006; Post et al., 2003; Moghadam, 2003; Pedahzur et al., 2003; Silke, 2003). Instead of drawing a fixed division between laboratory-induced low stakes deception by students and real-life high stake deception by criminals, the picture should be viewed in a more complex light.

2.6.2 Cues to deception

Communication can be achieved by various means encompassing verbal as well as non-verbal behaviour and, consequently, deceptive messages can arise from these many different channels. A great 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, vocal behavior, and facial expressions[4]. There are a handful of reviews evaluating the various investigations into locating reliable deceptive behavioural cues (DePaulo et al., 2003; Sporer and Schandt, 2006; Vrij, 2008; Zuckerman et al., 1981).

The meta-analysis by DePaulo et al. (2003) provides one of the most comprehensive works to date. Examining a total of 116 research reports that explored the relationship between 158 behavioural cues and the act of lying or telling the truth, the authors concluded that there are very few behavioural differences between lying and truth-telling. One finding that was relatively consistent across the analysed studies however relates to the observation that liars act and are also perceived as more tense. The tension manifested itself both bodily and vocally as proven by increased pupil dilation and pitch. Furthermore, there was a tendency of reduced body movement during lying but this was subject to increased individual variability. Gaze aversion and nervous, fidgety behavior, which is stereotypically associated with lying, was rarely found.

[4] The analysis of facial expressions or 'microexpression' as a method of detecting deception has been most famously advocated by Paul Ekman (1985). It has received media attention through the popular US TV series 'Lie to me'.

More recently, interest into the verbal aspects of deception has grown and a review of verbal cues to deception concludes that deceptive accounts are shorter, contain fewer details and are less plausible compared to truthful statements (Vrij, 2008). Examples of studies that have investigated verbal aspects can be found in Adams (2002), Porter (1994) and Porter and Yuille (1996).

The research of DePaulo et al. (2003) has shown that liars' speech was perceived as more tense. This reference to increased pitch was the single vocal behaviour that featured in the meta-analysis. A meta-analytic study conducted by Sporer and Schwandt (2006) specifically focused on the vocal aspects of deceptive behaviour. The nine parameters analyzed included pitch, speech rate, filled and unfilled pauses, speech errors and repetitions, response latency, overall message duration and number of words. Rather disappointingly, the majority of these were not significantly correlated with deception; only pitch and response latency showing a reliable positive correlation. Of note is the predominance of temporal parameters over frequency based parameters analysed, and, in fact, it is only pitch which would fall into the latter category.

It appears as though the literature on the acoustic correlates of deceptive speech is scarce and this indicates that further research into this area would provide a beneficial contribution. It may, at first sight, seem that plenty of research has already been conducted into the temporal speech cues associated with deception. However, as will become clear in Chapter 5, in addition to examining frequency based parameters, it is equally worth further investigating the temporal aspects of deceptive speech.

One might ask the question why vocal cues should be advantageous in deception detection. Indeed, why would the voice be a better carrier of deception cues than the face or the body? Ekman and Friesen (1969) introduced the leakage hierarchy hypothesis which assumes that some communication channels are more controllable than others. Specifically, the authors conjectured that the body, as compared to the face, is less controllable and, therefore, more likely to show signs of deceit, which would ultimately be of benefit to deception detection, and indeed, empirical support exists that confirms their hypothesis (Ekman and Friesen, 1969; 1974). This train of thought was extended to include the vocal channel and studies have documented that, similarly to the body, the voice tends to be less susceptible to conscious control (Ekman et al., 1991; Zuckerman et al., 1982). Scherer (1985) reported that the vocal channel does carry leakage information and voice quality based features in particular seem to play the most influential role.

2.7 Summary

This chapter provided the reader with a concise overview of the concept of deception. A definition of the notion of deception was offered along with a description of different types and motives. Factors affecting the process of lying were also briefly summarised. The theoretical foundation for the empirical study of deception was presented in form of three processes thought to possibly be involved during the act of deceiving. From describing the theory, the chapter then moved on to give the reader an idea of the methodological characteristics and findings of previous empirical research into behavioural correlates of deception. In addition to finding empirical evidence for the occurrence of behaviours believed to be underlying deception, researchers have been eager to discover how best to detect these behaviours. Chapter 3 will introduce the reader to this area of the research on deception.

3 Detecting deception

This chapter introduces the reader to the research on deception detection. To start off with, the focus is on people's ability to detect lies. In other words, how good people are, based on empirical research, in detecting deceit without using specialised technology. Attention will then shift to lie detection using specialised tools and technological aids, and the reader is presented with the traditional as well as the more pioneering methods that are emerging. The chapter concludes by rationalising the inclusion of the influence of social-interactional settings on deception as a crucial aspect of the present work.

3.1 Deception detection without specialised tools

People are not very accurate at spotting deceivers; numerous studies have analysed deception detection accuracy and the result tends to be just slightly above chance, i.e. 50% (Aamodt and Custer, 2006; Bond and DePaulo, 2006). Several reasons are advanced to explain this finding. Firstly, beliefs about how liars behave, also called subjective cues, do not correspond to actual deceptive behaviour, or so called objective cues (Fiedler and Walka, 1993; Masip, 2006). Secondly, people tend to rely on using heuristics, rules of thumbs or intuitive judgements to detect deception (Strömwall et al., 2004; Vrij, 2008). The majority of lay people as well as professional lie-catchers such as police/security officers associate deception with nervousness. For example, participants generally draw a connection between deception and gaze aversion, fidgety movements, eye blinking and self-manipulators such as playing with the hair or biting the nails (DePaulo and Pfeifer, 1986; Köhnken, 1987; Kraut and Poe, 1980; Mann et al., 2004). Indeed, this seems to be a universal belief as suggested by an investigation by the Global Deception Research team in 2006 and, moreover, it appears to have been prevalent since ancient times:

> A person who gives poison may be recognized. He does not answer questions, or they are evasive answers; he speaks nonsense, rubs the great toe along the ground, and shivers; his face discolored; he rubs the roots of the hair with his fingers; and he tries by every means to leave the house... (Wise 1860 as cited in Trovillo, 1939a, p. 849).

As was pointed out above, the literature on actual cues of deception revealed markers of nervousness to be rather infrequent. A truth bias, i.e. the general predisposition to believe the truth and lack of adequate feedback on the success of deception detection are also popular explanations for the low detection accuracy (Levine et al., 1999). More recently, an alternative reason has been proposed in order to account for the lack of sizeable deception detection ability amongst many. Rather than a result of using invalid subjective cues, the meta-analysis by Hartwig and Bond (2011) puts forward that it may be due to a lack of valid objective cues in the first place.

Despite the limited number of observable cues indicative of deception, training programmes have been designed aimed at improving people's deception detection accuracy. A meta-analytic review by Frank and Feeley (2003) has shown that training, on average, increased detection accuracy by 4%. Although a significant increase in the broader scheme it is still a relatively small success rate. Given the lack of reliable indicators of deception this increase appears paradoxical. Levine et al. (2005) argued that the observed improvement might not necessarily be related to the training content but due to the mere act of receiving training, as it will have stimulated peoples' alertness and vigilance to processing source messages and source behaviour.

3.2 Deception detection using specialised tools

It has been shown that people are not very good at detecting deception based on lay judgement only. Could deception detection be improved by using specialised equipment? VSA based technology and the limitations in using it for the purposes of deception detection has already been introduced in Chapter 1. Other methods have been proposed which employ specialised tools in order to detecting deceit and the following section will present an overview of these. The aim is for the reader to get an overall idea of what is available and, therefore, it only presents a synopsis of the literature. References to more comprehensive descriptions are suggested throughout.

3.2.1 Linguistic analysis

When one hears the terms 'tools' and 'technologies' initial associations often revolve around concrete machines and technological devices. However, 'tools' can also refer to specific techniques which may not require a physical device as such. These techniques could involve specific methods of analysis. As briefly mentioned in Chapter 2, the study of deceptive language, rather than deceptive speech or deceptive behaviour more general, is growing. A number of verbal veracity tools are in existence. One of these is Statement Validity Analysis (SVA) formalised by Steller and Köhnken (1989). SVA was first developed to assess child witness accounts of alleged sexual abuse but attempts to broaden its application to suspect accounts in criminal proceedings are growing (Vrij, 2008).The core of SVA comprises Criteria-Based Content Analysis (CBCA) which rests on the atheoretical assumption that truthful and fabricated utterances differ in terms of content and quality. This assumption has been termed the 'Undeutsch hypothesis' in reference to the German researcher Werner Undeutsch, who first developed it (Vrij, 2008). It has been translated into practice by evaluating written text with regards to nineteen criteria which are thought to be present in truthful accounts. In view of this, CBCA should more appropriately be viewed as a verbal truth detection tool rather than an instrument for lie detection (Vrij, 2008). These nineteen criteria are divided into four subcategories which relate to a statement's general characteristics, the presence of specific contents and crime specific references and finally, to aspects of its presentation. The reader is referred to Vrij (2008, p. 208) for a full list of these criteria. As is the case with nonverbal cues to deception, SVA or CBCA are affected by speaker internal (e.g. false memories) and speaker external factors (e.g. type of interview, personality traits). Furthermore, despite attempts to such an end, the test is not standardised and this can result in two experts reaching two different opinions when interpreting the findings. SVA has important legal standing as it is used in several courts across the world. Although numerous laboratory studies exist which validate the method, field research has not been as successful and questions over the validity and reliability of SVA as a tool to detect fabricated statements are still present (Vrij, 2008).

Research into alternative verbal veracity assessment tools or methods is ongoing. In particular, investigations that are more theory-driven are gaining attention. One of these is Reality Monitoring (RM) first developed by Johnson and Raye (1981). RM was originally used to differentiate between events that people truly experienced and those that were the product of imagination by examining how memory manifests itself in speech. It is predicted that verbal memory recollections that are based on real events contain sensory, contextual, and affective information whereas those that are derived from imagination are characterised by cognitive reasoning. The former are rooted in external sources whereas the latter stem from internal operations. As part of RM, the Memory Characteristic Questionnaire (MCQ) was devised, which enables people to perform self-assessments with regards to the quality of their memory (Johnson et al., 1988). Scholars have proposed that the MCQ has the potential of a lie detection tool and a number of studies exist which test this assumption. Similarly to previous research designs, tape recorded truthful and deceptive interviews provided the data which was then analysed using a set of RM criteria, some of which are very similar to those used in CBCA (Vrij, 2008). Vrij (2008, pp. 278-279) provides an overview of deception studies using RM. In general, RM performs above chance in differentiating between truthful and false accounts; however, there is no uniform set of criteria or guidelines on its application, which considerably limits the method's reliability and application in practice. Vrij (2008, p. 276) argues that RM could be used for investigative purposes and in combination with CBCA may offer a comprehensive tool that assesses veracity in verbalised statements by taking into account factors connected to cognition, memory and motivation.

Verbal or text based detection of deception has not been subject to as much publicity or popularity as other methods based on voice analysis, i.e. VSA, PSE. This is paradoxical given that far more scientific research is available in relation to the former. One possible exception to this is SCAN (Scientific Content Analysis), a verbal veracity assessment tool developed by Avioam Sapir. The marketing strategies used to promote SCAN parallel those used for VSA products. Despite its lack of scientific support and standardisation, SCAN appears to be used by several countries across the world for lie detection purposes. The reader is directed to Vrij (2008) who provides an honest and fair evaluation of the SCAN technique as a lie detection tool.

Exactly why text-based analysis of deception has not been in the foreground of popular culture as much as voice based analysis is unclear. It could be that it is believed to be of less utility in the practical setting or, perhaps, it is a product of polygraph based methods which place restrictions on the quantity of utterances produced by the examinee. People may have the misconception that accurate voice analysis can be performed on a minimal amount of net speech whereas text-based analyses require longer samples. One of the major advantages of verbal veracity tools is the fact that they take into account cognitive processes and, as has been and will be demonstrated in later chapters, the cognitive theory has been at the forefront of deception research in recent years. It will be interesting to observe whether the status of the linguistic analysis of deceptive language changes in future. In the words of the US committee to review the scientific evidence on the polygraph – "efforts to design measures for the detection of deception based on language use may have untapped potential" (National Research Council, 2003, p. 165).

3.2.2 Polygraph analysis

The following will only provide a brief overview and the reader is redirected to Vrij (2008), National Research Council (2003) and Lykken (1998) for a thorough understanding of the many aspects surrounding the polygraph. The polygraph is the most common physiological method of deception detection. It is often referred to as a 'lie detector' but this term is deceptive in itself. The polygraph is a scientific instrument that measures autonomic nervous system (ANS) activity. Relying on the use of electrodes, blood pressure cuffs and pneumatic gauges it records cardiovascular activity, respiration rate and skin conductance. The assumption is that liars will show physiological responses that are different to those displayed by truth-tellers. In addition to the 'physiological' component represented by the various measures of the ANS, polygraph tests also have a 'psychological' element which relates to the nature of the questioning of the examinee. Whereas the physiological component usually does not vary, the psychological one does as different methods/techniques of questioning have been developed (Vrij, 2008). The different tests can be separated into two main groups and proponents of each have come into conflict over the years. While both quarters agree on using the polygraph as a deception detection tool, they disagree on the theoretical foundations on which the testing is based.

The first approach assumes that liars will experience heightened physiological arousal because of increased stress, concern or fear of being detected. This increase in stress is interpreted as a sign of lying and, recalling Chapter 2, this assumption links back to the emotional theory of deception. Tests which would fall into this category are the 'relevant/irrelevant test', 'the direct lie test' and, most popularly, 'the control question test/comparison question test'. Rather than identifying whether an examinee is lying, the second approach aims to measure whether an examinee possesses certain concealed knowledge about a crime which would indicate that he or she was the perpetrator. It relies on the theoretical notion of the orienting response which is an individual's immediate reaction to new or significant stimuli. It is associated with physiological changes including dilation of the pupils, a rise in skin conductance and heart rate, and changes in brain wave activity (Nieuwenhuis et al., 2011; Williams et al., 2000). For an overview of the orienting reflex the reader is referred to the work of Pavlov (1927) and Sokolov (1963). The assumption then is that liars will show orienting responses when presented with crime-relevant information which will be exposed by the polygraph readings. By far the most popular test which would fall into this category of polygraph testing is the Guilty Knowledge Test (GKT), sometimes also referred to as the Concealed Knowledge Test (CKT) (Lykken, 1998).

Porges (2006) ingeniously referred to the debate amongst these two groups as “one machine, two world views” (p. 47). In the published report of the US National Research Council (2003) on the scientific evidence underlying the polygraph the conclusion was less than favourable. It stated that anxiety/arousal based lie detection, i.e. polygraph testing using the control question test or similar, is not based on sound and scientific principles. Amongst the identified weaknesses were poor standardisation, lack of theoretical base, susceptibility to counter-measures and subjective scoring. The conclusions with regard to the GKT were more positive and in agreement with some earlier laboratory studies which concluded the test to be capable of being an effective deception measure (Lykken, 1998). Although the theoretical foundation of the GKT is approved, weaknesses are nevertheless apparent including the difficulties in designing an unbiased GKT test and the restricted applicability of the GKT test to specific crime scenarios only.

From a recent survey it became clear that within the scientific community the GKT is approved and regarded as theoretically sound whereas the premise of the CQT was generally rejected (Iacono and Lykken, 1997). Whether this holds true outside of the scientific community is questionable as professional polygraphers tend to be removed from the academic and scientific environment. Although not of evidential value, the polygraph, despite its proven flaws, is still used as an investigative tool in numerous countries.

As mentioned above, as well as leading to changes in ANS measures, the orienting reflex manifests itself in event related potentials (ERP), which refers to brain wave activities in response to any type of stimuli. ERPs are measured using Electroencephalography (EEG). Specifically, evidence from neuroscience has shown that stimuli of strong personal or emotional significance elicit brain waves which peak after 250-500 ms after onset of the stimuli. These are called ‘P300 waves’ and, in light of the high demand for the development of objective lie detection tools, it is not surprising that interest into this area has grown amongst deception researchers. Reference is made to deception detection using ERPs at this point in the text, rather than under the following section on neuroimaging, as it is closely related to polygraph testing using the guilty knowledge paradigm. The reader is referred to the research of Rosenfeld and his colleagues for a full insight into this type of deception research (Rosenfeld, 2002; Rosenfeld, 2011).

Deception detection using the polygraph relies on physiological changes assumed to be indicative of lying or the possession of incriminating information. Putting aside the GKT, polygraph testing is based on assumptions that correspond to the emotional perspective on deception as outlined in Chapter 2. Also highlighted in Chapter 2 was the cognitive basis underlying deception. In view of this the question has been asked whether lie detection based on physiological measures taken from the central nervous system would be more accurate than lie detection based on peripheral nervous system measures. The following section will briefly describe the recent attempts of scientists to answer this question.

3.2.3 Functional neuroimaging

As explained in Chapter 2, one of the processes believed to be underlying deception relates to the increase in cognitive load required for successful lying. Successful liars need to engage in information and image management. In order to accomplish the deceptive act, liars need to be able to plan their responses, they need to pay attention to their surroundings, they need to control their behaviour and require mental flexibility. From a neurological point of view, planning, problem solving, decision making, attention, behaviour control and manipulation of information are all executive functions which are, in part, regulated by prefrontal regions of the brain, specifically the prefrontal cortex (Spence, 2004).

Lie detection based on functional neuroimaging using functional magnetic resonance imaging (fMRI) or positron emission tomography (PET) is still in its infant stages. fMRI is similar to voice research in that it is dependent on assessing the difference in brain response between two types of stimuli, states or behaviours, i.e. a control/baseline and an experimental manipulation such as deception. Amongst other things, fMRI scanners severely restrict mobility as well as face-to-face interaction and this affects the type of experimental designs that can be employed. Commonly, participants lie or tell the truth by pressing buttons. PET is more advantageous in terms of enabling direct communication between participants and experimenter resulting in the collection of verbal responses. Experimental tasks usually involve answers to autobiographical questions (Abe et al., 2007; Spence et al., 2004), answers to questions about a mock-crime committed (Kozel et al., 2005) or what has been termed the 'play card paradigm' (Langleben, 2008; Davatzikos et al., 2005). In the latter, participants are shown a play card prior to the scan and asked to keep it. During the scan they are presented with an array of cards including the one seen at the beginning of the test and are asked to deny possession of every one of them. All of the conducted fMRI studies agree on the fact that lying resulted in the activation of regions within the prefrontal cortex and to a lesser extent cognitive demand (Abe et al., 2007; Ganis et al., 2003; Langleben et al., 2002; Spence et al., 2004). These studies not only present empirical evidence for the cognitive theory of deception they also substantiate earlier claims that the suppression of salient information is part of, although not specific to, the deception process (Spence et al., 2001; Walcyk et al., 2003). Amongst these similarities in findings, differences are also apparent across the studies. Whilst there is agreement on the involvement of the prefrontal cortex during deception, inconsistency with regards to exactly which sub-regions are activated is apparent. Moreover, prefrontal-parietal activation also occurs during truth-telling (Langleben et al., 2005). Analogous to behavioural research on deception, fMRI research revealed strong inter and intra-individual differences. Situational and contextual factors have an effect on what part of the brain is activated and just as there is no single consistent behavioural cue to deception, there is no one specific area of the brain that is always associated with telling lies (Ganis et al., 2003; Kozel et al., 2004a; 2004b). As Vrij (2008) stresses, fMRI research into deception has, so far, not convincingly shown that it has practical advantages over traditional detection methods. Despite these practical limitations, this type of research is nevertheless valuable as contributing to our theoretical understanding of the process of deception. For example, an understanding of the neurological basis of deception may inform concepts in social cognition such as the 'theory of mind' as well as generate implications for legal definitions of responsibility, guilt and mitigation. The latter, by extension, would in fact be of relevance to criminal and mental law making.

3.2.4 Other emerging technologies

As indicated above, the overriding association when one thinks of 'lie detection tools' relates to machine-based technologies. In 2005 the National Science Foundation funded a series of workshops on the use of autonomic and somatic measures for security evaluations. Subsequent to this, proceedings were published compiling all the oral presentations in one volume (Webb and Kircher, 2006). One of the areas covered related to emerging and alternative methods for detecting deception including thermal imaging, gait, odour and occulomotor activity. Voice is also mentioned as a promising means in deception detection and security evaluation, in general. In their contribution to the workshop Hollien and Harnsberger (2006) expand on what was addressed in Chapter 1 of the present work, and state "basic research in the relationship between voice, stress and deception is required in order to develop deception detection models and their ultimate security applications." (p. 78). As the quotation indicates, their research program revolves around the assumption that the majority of instances of deception involve heightened levels of stress. While stress is acknowledged as being a possible concomitant to deception, the research effort of the present work focused more on the cognitive load aspect of deception as introduced in Chapters 1 and 2.

All these research activities are scientifically interesting and stimulating; they are innovative and have the potential of broadening our understanding of the concept of deception. However, one must wonder whether they are useful from a practical perspective. Ultimately, the aim of the majority of deception related research is to improve real-life deception detection and to better identify security threats. Often the emerging technologies and methods are impractical and difficult to apply in a real-life setting. They often require very expensive equipment, are time-consuming to carry out and rely on specialised knowledge in a given field e.g. neuroscience, physics, phonetics. Are there potential methods that would provide for easier application in real life, methods that do not require specialised, expensive and heavy equipment?

On a more theoretical level, the emerging research into developing deception detection technologies stands up against similar problems as faced by traditional methods using the polygraph or detection without using specialised equipment. It is true that we can identify whether a person experiences cognitive load or emotional involvement and we can also detect whether persons attempt to control their behaviour or emotions. Research has made striking progress in developing and improving the methods of detecting what are believed to be behavioural correlates of deception. However, numerous problems still remain including the neglect of the social dimension and context when studying deception, making reverse inference from physiological observations to deceptive behaviour, the presence of individual variability which is paramount in so many different aspects of life and the constraints of laboratory research, in general (Sip et al., 2007). Researchers are aware of these issues and continually acknowledge the problems faced but little has been done to tackle them. While some of the above issues could be addressed by appropriately modifying experimental paradigms used, others, it is believed, may be insurmountable e.g. inter-individual variation.

The present work attempted to take into account the role of social interaction when studying deception. This is reflected in research question 3. Deception commonly involves face to face interaction. For example, police are investigating a crime that has been committed. In their information gathering process they have identified a suspect and are subsequently interviewing him or her with regards to the crime in question. Terrorist attacks are usually preceded by pre-operation planning phases which involve various processes of intelligence, surveillance and reconnaissance (ISR) including, for example, gathering overt and covert information, completing dry-runs and testing security responses (O'Brien, 2008). At some point during their ISR activity potential terrorists are required to be physically present at their target location. For this period they need to conceal their intentions and emotions from the security personnel and the general public in order to proceed with the planned operation (Porter et al., 2010). Be it a suspect in a police interview room or a terrorist completing reconnaissance at a target location, the face-to-face interaction process may be viewed as an opportunity for identifying suspicious or deceptive behaviour using methods grounded in different techniques of interviewing.

3.3 Deception detection using interviewing techniques

As pointed out in Chapter 1 research into interviewing techniques is thriving. This section will briefly outline two different interviewing styles that are currently in operation. Research question 3 focuses on these two interviewing types and evaluates differences between them in terms of efficiency in eliciting cues to deception. It only is a brief summary as anything more extensive would go beyond the limits of this thesis. The reader is directed to the research of Maria Hartwig for a more comprehensive overview of the literature (Hartwig, 2005).

3.3.1 Accusatory interviews vs. investigative interviews

Accusatory interviews rely on an aggressive and coercive interrogation style aimed at eliciting a suspect's confession. Questions are confrontational and often interviewees are directly accused of having committed a crime or of lying. By far the most influential accusatory interrogation technique on an international level is the Reid technique. It is based on the guidelines by Inbau et al. (2001) featured in their published manual *Criminal Interrogations and Confessions*. Not only do the tactics proposed by the Reid style raise a number of ethical issues, there is virtually no empirical basis for the majority of their recommendations (Gudjonsson, 2003; Masip, 2005). Part of the interrogation includes what Inbau and colleagues called the Behavioural Analysis Interview (BAI) which consists of fifteen questions designed to elicit different verbal and non-verbal cues from liars as opposed to truth-tellers. The questions tend to be of the closed type and are formulated in such a way as to obtain short, at times even monosyllabic, responses. As discussed above, the majority of cues stereotypically relied upon when attempting to detect deception have been shown to be unreliable and non-predictive by previous research. Apparently disregarding the value of the contemporary research literature, the core of the BAI is based on these non-diagnostic cues. Recently, and as the result of direct empirical testing, doubts over the BAI's premises are increasing (Vrij et al., 2006). As a matter of interest, the manual does not provide recommendations on vocal cues to deception.

Following the introduction of the Police and Criminal Evidence Act (PACE) in the UK in 1986, accusatory interrogation techniques increasingly gave way to information-gathering approaches one of which being the investigative interview (Hartwig, 2005). Investigative interviewing is a more ethical interrogation style which seeks to gather information rather than confessions (Milne and Bull, 1999). Based on a conversation management approach it employs methods that evoke informative rapport by creating a positive atmosphere. Some notable attributes include allowing the interviewee most of the talking time, using open-ended questions that do not constrain, attentive listening and appropriate pausing as well as limited interruptions. Typically, the interviewer will initially set the scene by providing a road map explaining the motivations and aims of the interview and, following this, the interviewee is encouraged to disclose as much detail as possible (Shepherd, 2007).

Vrij et al. (2006b) compared information gathering and accusatory interviews with regards to how truth-tellers and liars experience these two events. In comparison to accusatory interviews, information gathering interviews were perceived as cognitively more demanding. This is not surprising given that information gathering interviews commonly require comprehensive fabrications whereas accusatory questions usually produce short answer type responses. Accusatory interviews were perceived as more discomforting. Again, this does not surprise as accusatory interviews, as the name suggests, are highly coercive. However, experienced pressure, as indicated by participant self reports, was equal across the two interviewing styles. Of particular note was the fact that veracity did not affect how the interviews were perceived. In other words, both truth-tellers and liars shared the same experiences. In light of research question 4, this is an important observation and will be further addressed in Chapters 8 and 9.

3.4 Summary

Chapter 3 has provided the reader with an overview of the current research into detecting deception. Lie detection was split into detection without using any specialised tools and detection using various methodological and technological aids. Whilst acknowledging the potential of the progressive and forward looking research that is being conducted, long standing problems revolving around empirical research into deception are still in existence. One of these is the inattention to the importance of social interaction and context when studying deception. So far, the reader has been presented with a broad overview of deception and its detection. The next chapter will narrow the focus down from looking at deceptive behaviour in general to deception as manifested in speech.

4 Deception and speech

The current chapter will give an overview of the possible vocal correlates of deception based on the central aspects of each of the three deceptive processes – Emotion/Stress, Cognitive Load and Hyper-control – explained in Chapter 2. Whilst at an interpersonal level it is possible to accurately perceive various psychological and emotional states, empirical research has only been moderately successful in establishing the associated acoustic and phonetic characteristics (Beckford Wassink et al., 2006; Hammerschmidt and Jürgens, 2007; Jessen, 2006; Köster, 2001; Lively et al., 1993; Meinerz, 2010; Scherer, 2003; Yap et al., 2011). This is largely attributable to methodological and conceptual obstacles. Hence, rather than referring to the correlations that have been discovered so far as 'reliable acoustic indicators', it is more appropriate to regard them as 'acoustic tendencies'. Further to this, the chapter discusses the possible impact on speech when people have a heightened sense of self-awareness and increased need for image management as set out in the self-presentational perspective of DePaulo et al. (2003). It also touches upon possible interlocutor effects on the speech of deceivers which was the central notion of Buller and Burgoon's interpersonal deception theory (1996). To this end the sociolinguistic research literature has been consulted and, in particular, research in relation to stylistic variation as a function of the speaker and the audience. Before presenting the literature and the theoretical argument a brief overview is given of the analysis of speech relevant to deception, i.e. forensic speech analysis.

4.1 Forensic speech analysis

Forensic analysis of speech can be conducted using auditory and acoustic analysis, as well as analysis employing automatic systems. A detailed overview of the various methods can be found in Jessen (2008, 2012). It is assumed that the reader is familiar with the scientific analysis of speech as it occurs in a forensic setting but for reasons of completeness a brief description is given. Specific focus is placed on the method of speech analysis adopted in the current work, i.e. auditory-acoustic analysis.

The acoustic analysis of speech involves using specialised computer software to quantify various physical parameters of the speech signal including fundamental frequency (f_0), speaking intensity and vowel formant frequencies, for example. f_0 is determined as the number of vocal fold vibrations per second and its perceptual correlate is pitch; the higher the number of vibrations per second the higher the f_0. Mean and variability are the most commonly reported aspects of this parameter within the field of forensic speaker identification (Nolan, 1983). f_0 variability is less influenced by the anatomy or physiology of the vocal folds. As Jessen et al. (2005) describe it, "[f_0 variability] is much more a behavioural variable that captures the choice a speaker makes on the scale between more monotonous and more melodic speaking." (p. 194).

Most Speaking intensity reflects, amongst other things, the level of sub-glottal pressure and its perceptual correlate is loudness. Accurate absolute amplitude measurements are difficult to obtain. Distance to the microphone and recording level will have an influence on absolute values and speakers' chosen level of speaking also presents a confounding factor. Whilst the technical aspects can be controlled with relative ease by using, for example, a headset microphone, the speaker-dependent factor presents a more challenging issue. Therefore, research tends to report on relative differences (Jessen, 2006).

The vocal tract, which could be simply viewed as a tube closed at one end by the vocal folds, acts as a resonating chamber and has natural frequencies of vibration, also called resonant frequencies. The input energy, i.e. the glottal pressure wave, is filtered by the vocal tract and those frequencies of the glottal wave that are closest to the resonant frequencies are enhanced and appear as peaks in the energy spectrum of vocalic sounds. These peaks are called 'formants' in acoustic phonetics and are numbered starting with the lowest in frequency (1^{st} formant or F_1).

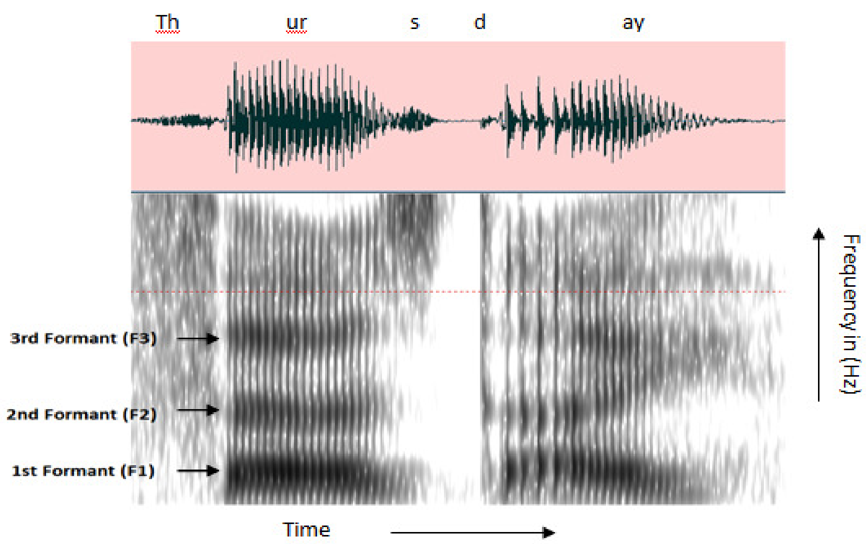

Figure 1 – Graphical illustration of formants F_1, F_2 and F_3 for the vowel /ɜ:/

Formants can be visualised on spectrograms as seen in Figure 1. Depending on the shape of the vocal tract it will resonate at different frequencies. Each vowel, having a different articulatory configuration, will be characterised by different formant frequencies, and the frequencies of F_1 and F_2 in particular determine the phonetic quality of vowels. While F_1 for vowels is correlated with the position of the tongue in the mouth in terms of its height on the 'open-close' dimension, F_2 is correlated with the position of the tongue in the mouth on the 'front-back' dimension (Denes and Pinson, 1993). Figure 2 below illustrates this relationship between tongue positions and formant frequencies. The three fMRI scans at the bottom show the author producing sustained /i/, /ɑ/ and /ʊ/, respectively. Immediately above each scan are the spectrograms of the respective vowels. Formant frequencies are marked by the red lines. As becomes obvious when comparing the marked formants across the three vowels, /i/ has a low F_1 and a high F_2, whereas /ʊ/ has low values for F_1 and F_2.

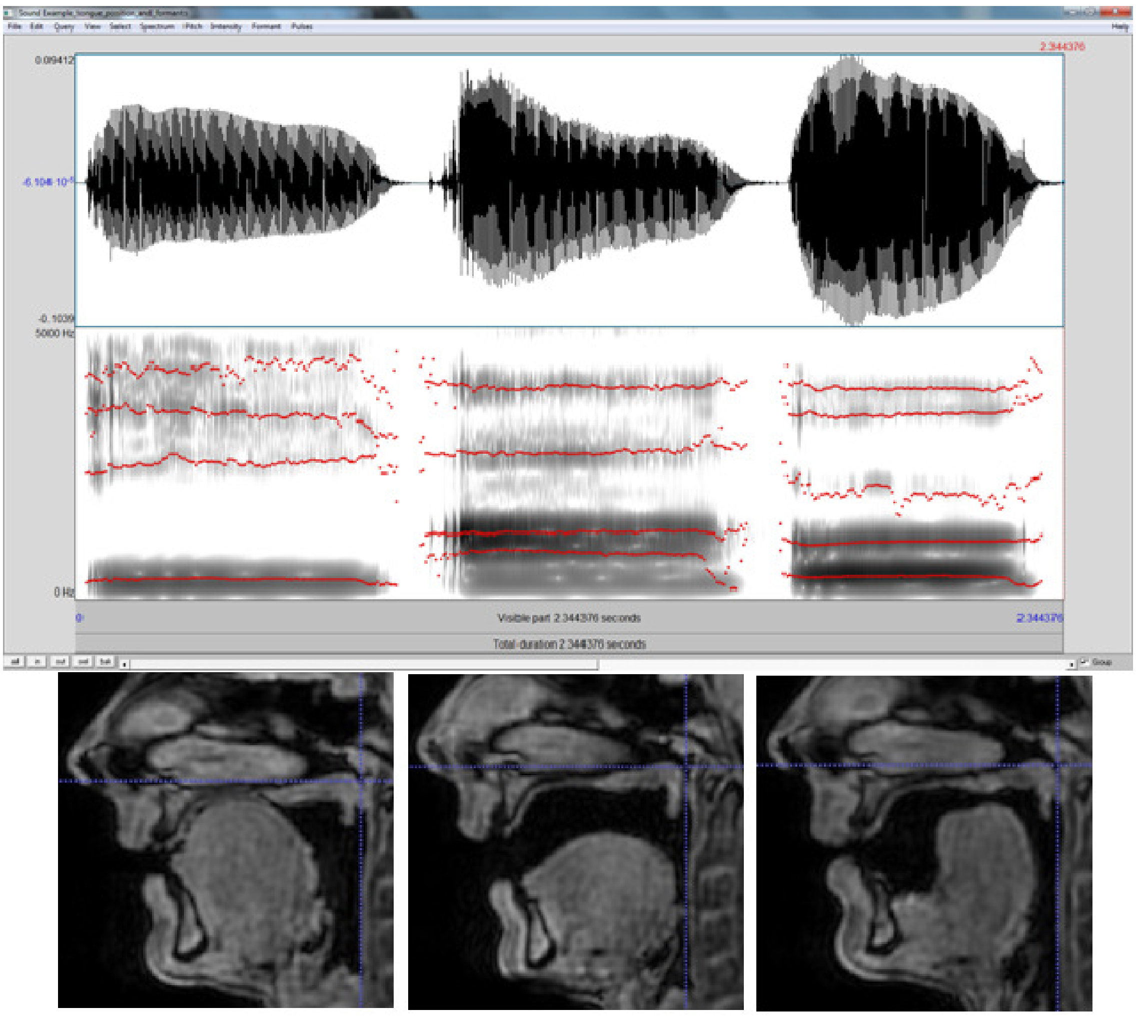

Figure 2 – Relationship between vowel formants F_1 and F_2 and tongue position

As well as determining the phonetic quality of vowels, formants also carry speaker-specific information in that they depend on a speaker's vocal tract anatomy and habits of articulation (French and Harrison, 2006, p. 254). Hence, the cause for formant differences may also be found in non-lingual, long term adjustments to the configuration of the vocal tract such as raising/lowering of the larynx or distension/constriction of the pharynx (Laver, 1980). In particular, the higher formants F_3 and F_4 have been linked to such long term settings (Fant, 1970; Ladefoged, 2006). One method of analysing formants which has the power to reflect speakers' long term settings and articulatory habits may be seen in long-term formant frequency distribution (LTFD) (Nolan and Grigoras, 2005). The method of LTFD involves averaging formant frequency values across all vocalic portions of a recording resulting in one LTF value per formant, i.e. LTF_1, LTF_2, and LTF_3. Vocalic extraction can be performed manually or using automated methods. Figure 3 shows an example of a sound file which only contains vocalic material.

Moos (2008) discovered that as little as six seconds of pure vocalic stream is sufficient for reliable LTF estimation. LTFD is a relatively recent development and has, so far, not been widely used in the analyses of speech or speakers.

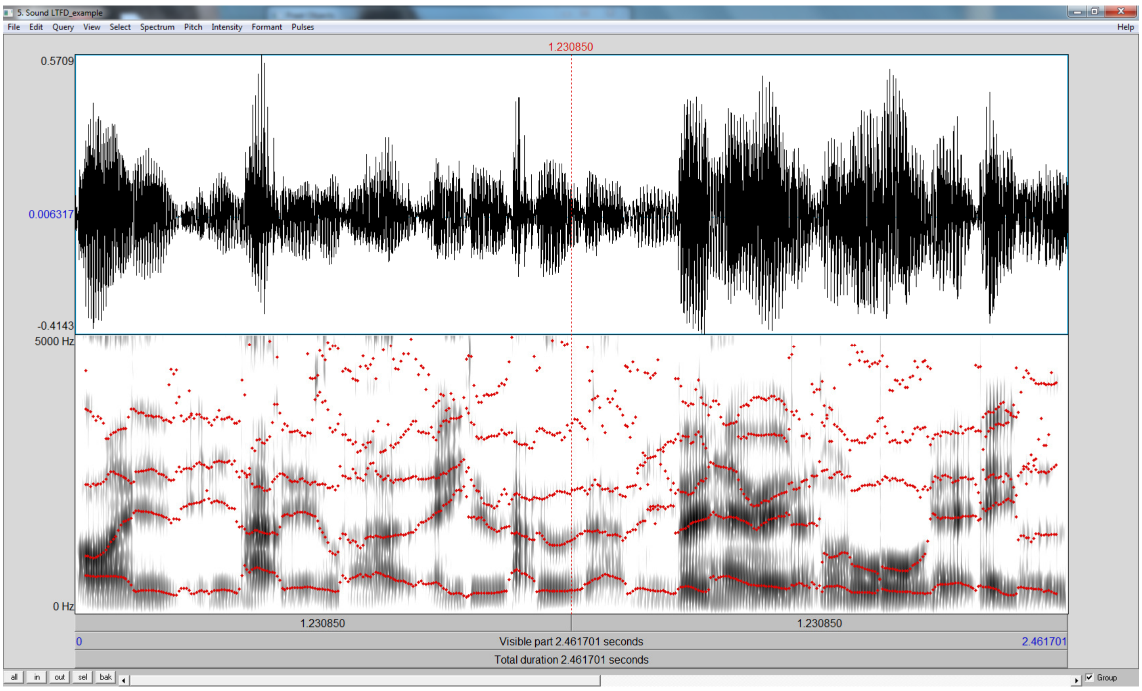

Figure 3 – Example of a sound file only containing vocalic material

Differences in F_1 and F_2 between different speaking conditions may suggest the occurrence of hyper-articulation or at the other extreme, target undershoot. Mean formant frequencies of the corner vowels, usually /i/, /a/, /ɒ/ and /u/, may be used to calculate a speaker's articulatory working space, or vowel space area. The assumption is that larger vowel space areas imply greater excursions of the articulators in terms of tongue height (F_1 dimension) or tongue advancement (F_2 dimension).

Temporal aspects of speech are commonly addressed using the measures of Speaking Rate (SR) and Articulation Rate (AR). Variably expressed in words, syllables, or sounds per seconds, rate parameters give an objective indication of how fast or slow people speak. SR refers to the rate of speech including all pauses and, therefore, incorporates aspects of peoples' speech planning processes. AR, on the other hand, excludes all silent intervals and, therefore, purely encompasses the net speech produced which, to a large extend, reflects the speed of peoples' articulatory movements (Goldman-Eisler, 1968). Pausing and hesitation phenomena are usually expressed in terms of frequency and duration of silent/unfilled pauses, as well as filled/hesitation pauses. Response Onset Time (ROT), i.e. the time it takes a speaker to respond to a question, and Voice Onset Time (VOT), which refers to time it takes for a speaker to produce glottal voicing after the closure period of a stop consonant, e.g. /p, t, k, b, d, g/, are additional parameters that tend to be quantified in research studies looking at aspects of speech tempo.

Most commonly, voice quality is assessed by auditory analysis. The Vocal Profile Analysis Scheme (VPA), which is based on the work of Laver (1980), allows for a detailed componential analysis of different vocal tract settings (e.g. nasality) and phonation features (e.g. creaky voice, breathy voice). Types of phonation and, in particular, breathy and creaky voice can also be characterised acoustically by taking measures of a) signal periodicity or harmonics-to-noise ratio (HNR), b) first harmonic amplitude and/or c) spectral tilt (see Hillenbrand et al., 1994). Different methods exist for calculating these spectral properties. For example, the most frequently used parameter for quantifying the amplitude of the first harmonic is H_1-H_2, i.e. the amplitude difference between the first and second harmonics. While breathiness has been shown to be characterised by enhanced H_1 amplitude, creaky voice commonly shows the opposite, i.e. H_2 prominence (Hillenbrand et al., 1994 for an overview of studies). A measure of spectral tilt may be seen in the amplitude difference between H_1 and the harmonic close to F_3, i.e. H_1-A_3 (Hanson, 1997). Research has shown conflicting findings in relation to the importance of each of the three spectral properties in distinguishing phonation types. Moreover, speakers may differ in how 'their' phonation type is implemented acoustically (Wayland and Jongman, 2003).

4.2 Research on speech under stress

A number of acoustic correlates of stress have been identified in previous studies. Before considering the actual acoustic correlates of stress however, some predictions about possible acoustic changes are listed. Based on the physiological aspects of the human stress response, in particular increased respiration rate, reduced salivation and increase in muscle tension, several predictions can be made about the possible vocal/acoustic changes involved, some of which have been discussed in Scherer (1979, 1981). A rise in respiration rate leads to an increase in sub-glottal pressure which, in turn, will increase the amplitude of vocal fold vibration. As a result, this could lead to higher amplitude of the speech signal. Increased sub-glottal pressure can also influence the frequency of vocal fold vibration. Jessen et al. (2005) discuss various anatomical, physiological and perceptual reasons for this correlation. One explanation may be seen in the fact that increased sub-glottal pressure produces increased trans-glottal airflow. This intensifies the Bernoulli effect and the vocal folds come together more rapidly. A raising in f_0 could also be caused by increased tension of the vocal folds caused by heightened muscle tension, in general. Turbulence and friction could be a further possible outcome of increased airflow which would manifest itself acoustically in the form of an increase in noise (e.g. affrication and aspiration). Increased muscle tension is also likely to affect the facial muscles of the lips, jaw, tongue and velum, and as a consequence, modifications in articulation of vowels and consonants may be observed. For example, it may be the case that tension in the articulatory muscles prevents the speaker from completely reaching consonantal and vocalic targets, also known as target undershoot. Acoustically, this could take the form of a contraction of the vowel space. On the other hand, a tense tongue might lead to vigorous and more rapid movement towards target sounds. If, at the same time, the temporal aspects of speech cannot be controlled this could lead to target overshoot. Scherer (1979) asserts that a possible lack of coordination between the laryngeal muscles involved in phonation on the one hand and irregular respiration patterns on the other could lead to voicing irregularity and voicing variability (jitter). In addition, a lack of coordination between phonation and articulation could lead to inappropriate voicing patterns in that, for example, voiced segments are devoiced in conditions that do not lend themselves to phonological or sociolinguistic explanations. In terms of voice quality, the acoustic consequences of an overall increase in tension within the vocal tract have been collectively referred to as ‘tense voice’ by Laver (1980, 1991). According to Laver, tense voice is characterised by harsh phonation which will sound louder and higher-pitched compared to modal voice (1991, p. 205).

The acoustic effects of the physiological response to stress on speech can, to some extent, be predicted for a number of acoustic parameters. A different situation exists with respect to the acoustic correlates of the psychological response to stress. Whilst there is active research on the acoustics of emotions, due to immense challenges in data collection methodology the results still have to be viewed with caution (Hollien, 1990; Scherer, 1981).

In order to test the validity of these predictions and discover new potential correlations the majority of empirical research studies have employed laboratory experimental designs to induce stress which is usually referred to as emotional stress, situational stress, laboratory stress, task-induced stress, cognitive stress and physical stress. Despite these different labels, all can be considered as factors on the psychological stress spectrum. Some authors differentiate between situational stress on the one hand and laboratory stress or task-induced stress on the other (Hicks, 1979; Meinerz, 2010). Laboratory stress refers to stress induced in a highly controlled laboratory setting. Examples include solving cognitive tasks under various conditions (Fernandez and Picard, 2003; Hecker et al., 1968; Jessen, 2006; Ruiz et al., 1996; Tolkmitt and Scherer, 1986), being subjected to environmental factors such as noise and temperature extremes (Vilkman and Manninen, 1986) or receiving electric shocks (Hicks, 1979). Situational stress, in comparison, is not caused by the experimenter inducing a stressor but rather by the naturally stressful nature of the setting itself. Examples of the latter include making a public speech (Hicks, 1979), completing an oral examination (Fuller et al., 1992; Sigmund, 2006), completing a simulated job interview (Meinerz, 2010), deceiving an interviewer (Streeter et al., 1977), watching gruesome slides (Ekman et al., 1976) and completing a dangerous mission (Johannes et al., 2000).

A number of studies have been conducted using recordings of situations involving real life stress, mostly catastrophic events such as aeroplane crashes or emergencies (Benson, 1995; Brenner et al., 1983; Hausner, 1987; Kuroda et al., 1976; Ruiz et al., 1996; Williams and Stevens, 1969). Williams and Stevens (1972) analysed the speech of a radio reporter during the Hindenburg crash and Streeter et al. (1983) examined the recording of a system administrator and his superior during the New York power cut in 1977. Few studies have concentrated on the investigation of the acoustic changes that are perceived to be indicative of stress (Protopapas and Lieberman, 1997). The following presents a selection of key research with the focus being on the acoustic characteristics of stress; a parameter centred description is adopted. A more comprehensive overview can be found in Jessen (2006) and Kirchhübel et al. (2011).

4.2.1 Fundamental frequency (f_0)

The rate of vocal fold vibration is the parameter most widely studied in research on stress in speech. Increase in mean f_0 due to stress was reported by Benson (1995), Brenner et al. (1983), Kuroda et al. (1976), Ruiz et al. (1996) and Williams and Stevens (1969) who investigated real life stress displayed by control tower operators and/or pilots. Streeter et al. (1983) found an increase of mean and variability of f_0 in one of their participants recorded during the New York power cut. Notably, however, for the other speaker, a decrease in f_0 was measured which underlines the interpersonal variance with respect to the nature of the reactions to the same stressors.

When inducing psychological stress using cognitive stressors, such as solving various cognitive tasks under time pressure, noise or physical strain, an increase in mean f_0 measures was reported (Brenner et al., 1983; Jessen, 2006; Ruiz et al., 1996; Scherer et al., 2002). Using recordings from oral examinations Sigmund (2006) concluded that there was an increase in f_0 in the stress condition compared to control recordings and motivationally aroused participants in Streeter et al. (1977) showed significant increases when being deceptive compared to when telling the truth. Subjecting participants to electric shocks, Hicks (1979) noted a slight yet non-significant increase in mean f_0 for the men; however, a significant elevation in f_0 was present during a public speaking condition. If an increase in f_0 is taken as indicative of stress, it is surprising that subjects found the public speaking condition more stressful than the threat of receiving electric shocks. It could have been the case that the increase in f_0 during the public speaking was a result of an increase in overall speaking intensity. People tend to speak more loudly and clearly during presentations or speeches. However, as described in one of the subsequent sections below, intensity measurements do not support this conclusion. An alternative explanation which is adopted by Hicks (1979) claims that situational stress, in contrast to laboratory stress, is in closer contact to real life stress as participants' fear of negative consequences is greater. Johannes et al. (2000) reported an increase in mean and variability of f_0 during situations of cognitive load and a similar result of an elevation in average f_0 was obtained when their participants performed a dangerous simulation task (space craft docking) and a dangerous real life task (jumping into a river).

There are studies however, which did not show the expected rise in f_0 mean. Tolkmitt and Scherer (1986) did not observe an elevation in mean f_0 when presenting their participants with emotionally arousing slides nor when asking them to solve difficult cognitive tasks. Asking participants to complete meter readings under time pressure Hecker et al. (1968) also failed to illustrate an increase in f_0 with stress. When looking at the extent of the rise in f_0 interesting differences can be observed between studies that have investigated real life stress and those that have induced stress in the laboratory where, typically, the f_0 increase is much greater in the former compared to the latter.

For example, Williams and Stevens (1969) documented an increase in mean f_0 from 208 Hz before flight difficulties to 432 Hz during flight difficulties for a pilot in an emergency situation. Ruiz et al. (1996) reported an increase in mean f_0 of over 60 Hz between a neutral and stress condition. Investigating pilot speech before and during an emergency, Brenner et al. (1983) recorded a rise in mean f_0 from 149 Hz to 264 Hz for one of their pilots. Laboratory induced stress on the other hand yielded much lower f_0 increases. Analysing the speech of 82 participants Mendoza and Carballo (1998) reported a mean f_0 of 195 Hz for the control condition as compared to 208 Hz, 211 Hz and 210 Hz for the three stress conditions. Jessen (2006) observed an increase of 5 Hz to 6 Hz for untrained police officers and 1 Hz to 6 Hz for trained police officers between neutral and stress conditions. Calculating the average mean f_0 for seven males, Brenner et al. (1983) recorded an increase in mean f_0 of 4 Hz (109 Hz to 113 Hz) between a non-stressed and stressed condition in their laboratory experiment. It appears that mean f_0 increases more drastically when people experience real-life stress as compared to laboratory induced stress. This difference between laboratory and real-life data could also be observed in relation to deception as detailed in Vrij (2008), described in Chapter 2.

4.2.2 Amplitude

The intensity of the overall acoustic output can also be seen as a common parameter of interest despite the difficulties involved in ensuring that reliable measurements are obtained. Some studies reported measurements of overall amplitude of the speech data but the results from these are inconsistent. Streeter et al. (1983) observed a rise in overall amplitude for one speaker but no such tendency for a second. The findings of Hecker et al. (1968) revealed a significant decrease in overall amplitude in a stress condition for three participants, a significant increase for one and no correlation for another six participants. Hicks (1979) observed a slight increase in a physical stressor condition but a decrease in a condition where participants were asked to make a speech on stage.

More recently, rather than measuring the overall amplitude, studies have tended to determine the amplitude of different frequency bands. Although different techniques were employed to achieve these calculations, there is general agreement that amplitude measurements are greater in higher frequencies, i.e. above 1000 Hz in stress conditions compared to control conditions (Benson, 1995; Fuller et al., 1992; Mendoza and Carballo, 1998; Scherer et al., 2002). Ruiz et al. (1996) analysed spectral balance frequency in vowels and tentatively suggested that this might be a promising parameter that is sensitive to stress.

4.2.3 Temporal parameters

Speaking tempo was found to increase as stress increased in recordings of a pilot's speech during an emergency landing (Hausner, 1987). While one employee in Streeter et al. (1983) reduced his speaking tempo, another showed no remarkable change. The majority of laboratory studies converged on the finding that speaking tempo (as measured in AR) had a tendency to increase as a result of stress (Hecker et al., 1968; Hollien et al., 1993; Karlsson et al., 2000; Siegman, 1993). Nevertheless, studies exist that did not identify a correlation between stress and AR (Jessen, 2006; Scherer et al., 2002). Meinerz (2010) reported a decrease in SR when participants had to complete a mock interview. Similarly, a slight but not statistically significant reduction in SR occurred when participants had to make a speech in the study of Hicks (1979). In that same study, however, no change in SR occurred when participants were given electric shocks.

Relatively little investigation has been directed towards how stress influences pauses and hesitation patterns. One reason for this could be that often there is insufficient data in order to quantitatively measure these phenomena. Studies that have attempted to examine the former conclude that the number of pauses decreases with stress (Benson, 1995; Hollien et al., 1993; Siegman, 1993). Hicks (1979) observed an increase in the number of pauses in his laboratory stress condition (i.e. receiving electric shocks) but a decrease in his situational stress condition (i.e. making a public speech). In terms of hesitations Hicks (1979) and Siegman (1993) observed an increase whereas Benson (1995) reported the absence of these in a stress condition. Jessen (2006) argued that it may be crucial to differentiate between different types of pauses; whilst the number of respiratory pauses increased in the stress condition, 'acoustic pauses', i.e. pauses that did not show any or very little acoustic energy, decreased when participants were stressed.

4.2.4 Formants

Formants have received relatively little attention in comparison to other parameters. Providing, amongst other things, an indication of the precision of articulation this parameter has potential to contribute to the understanding of the effects of stress on speech. The few studies that have reported on formant measurements show contradictory findings. Hecker et al. (1968) observed that under stress vowel targets were often not reached by some participants and Karlsson et al. (2000) reported a centralisation of the vowel space in stressful conditions. Hecker et al. (1968) suggested increased speaking tempo as the cause for target undershoot. Benson (1995) observed an increase in F_1 in high vowels which again could be caused by imprecision of articulation. Having analysed formant frequencies as well as formant bandwidths, Sigmund (2006) concluded that only the former showed significant changes. The paper by Sigmund noted an increase in F_1 and F_2 but, unfortunately, it was not specified which vowel phonemes were analysed. Similar results were reported by Meinerz (2010) whose analysis showed an increase in F_1 and F_2 in three different stress conditions as compared to modal speech. An increase in F_2 could be indicative of a fronting of the vowel space and would be in contrast to previous findings of target undershoot. The method of formant extraction employed in Meinerz (2010) does not allow for the classification of vowel categories however, and therefore no assumptions can be made with regard to precision of articulation. F_3 was not significantly affected by stress. In their comparison of speech from a control condition to speech produced under laboratory-induced stress as well as real life stress, Ruiz et al. (1996) also found significant variation in formant values. These variations were vowel dependent however, and did not apply to F_1, F_2 and F_3 to the same extent. A more complex picture is presented in Tolkmitt and Scherer (1986) in that women tended to show target overshoot or hyper-articulation in a cognitive stress condition whereas articulation precision was reduced in a condition involving emotional stress. For men, the variation in formant values as a function of stress was insignificant. The authors also looked at possible individual differences due to personality and divided the participants into groups of high anxious, low anxious and anxiety deniers. Changes were most prevalent in anxiety-denying subjects who, it was concluded, were most susceptible to stress.

4.2.5 Voice quality

Evidence of an auditory assessment of voice quality can be found in Hecker et al. (1968) who observed the presence of low frequency vibration during a stress condition which could be indicative of ‘creaky voice’. Van Lierde et al. (2009) observed female voices to be more breathy and strained under stress. Analysing voice quality by means of acoustic methods can entail the use of sophisticated techniques. Several authors have used inverse filtering as a means of separating the glottal source from the acoustic speech signal by eliminating the vocal tract transfer function from the latter. Using this method Waters et al. (1995) observed that there are effects on different measures of the glottal pulse when speech is produced under increased task-load. Without naming the direction of change they reported significant modifications in pitch, rising slope and closing slope of the glottal pulse. In addition, they discovered a great amount of inter-speaker variability between the relationship of various glottal parameters and stress. Hecker et al. (1968), Hollien (1990), Jessen (2006) and Williams and Stevens (1972) have all remarked that voicing irregularities or ‘voicing breaks’ occur when people speak under stress.

4.2.6 Perceptual studies

An alternative approach to quantifying the acoustic correlates of stress can be taken by asking listeners to rate samples of speech in terms of how stressed the speaker sounds. Streeter et al. (1983) inferred from their perceptual study that listeners’ ratings of stress were significantly correlated with f_0 for both speakers investigated. Protopapas and Lieberman (1997) used f_0 contours from the speech of a helicopter pilot during a real-life stress and a non-stress condition and overlaid those contours onto synthesized fixed-vowel utterances. They then manipulated various f_0 parameters such as mean, maximum, minimum and standard deviation without changing the overall intonational contour. The results indicated that increased mean and maximum f_0 tended to be associated with stress while f_0 range failed to do so. In order to test whether the association between mean f_0 and perceived stress also holds across speakers, Protopapas and Lieberman (1997) performed a subsequent experiment in which they recorded four speakers (two male, two female) uttering the word ‘top’. The vowel portion of each speaker’s utterance was then used in conjunction with twelve different f_0 tracks to synthesize a set of stimuli consisting of 48 tokens. As was found in the previous experiment the stress ratings correlated highly with mean and maximum f_0 for the two male voices. The ratings for the female voices showed less correlation however. The authors also suggest that perceived stress may be related to formant frequencies as a higher stress rating was given to voices with higher formant frequencies. Rather than testing specific acoustic factors and how they correlate with stress, Jessen (2006) designed more of an exploratory study in order to investigate which phonetic as well as linguistic parameters might contribute to the perception of stress. One of the more consistent findings was the correlation between vocal or linguistic expressiveness and stress. Linguistic expressiveness was rated worse in the stress conditions compared to the control condition. What exactly contributes to vocal/linguistic expressiveness ratings is not clear but it is likely to be a combination of phonetic, syntactic, lexical and stylistic factors (Jessen, 2006, p. 204).

4.3 Research on speech under cognitive Load

The same challenge that was faced with determining a general concept of stress or psychological stress applies equally to cognitive load or cognitive workload. As Baber et al. (1996) state, "given the range of demands which can have a bearing upon workload, it would be difficult to provide a unified definition of the term; indeed there are many definitions of workload." (p. 38). Cognitive load/workload has often been confounded with psychological stress especially in those studies that have employed a task induced stressor such as solving cognitive tasks under time pressure or performing multiple tasks under life–threatening conditions. In these instances it is difficult to differentiate between whether changes, if present, were generated by the effects of stress or by cognitive loading. Cognitive load/workload may well lead to stress in some contexts and some individuals but nevertheless, cognitive load and stress are individual entities and should not be confused. In the same way as a narrow reading was taken for the concept of speech under stress, the discussion on cognitive load will primarily focus on those studies that have clearly differentiated between task/time-induced cognitive stress and cognitive load. This differentiation is necessary in order to make predictions, based on the cognitive theory of deception, of possible changes in speech when people are being deceptive. The reader is directed to Berthold (1998) and Khawaja (2010) for a comprehensive discussion encompassing the many different aspects of the concept of cognitive load.

Different methods have been employed to induce cognitive load in the laboratory. One of the most popular techniques is the multiple task experiment. These commonly involve participants completing a primary task such as real or simulated flying or driving, playing computer games or talking to an interviewer, as well as simultaneously completing a secondary task such as memorising words or digits, solving mathematical equations or holding constant eye contact with an interlocutor. At times, an additional time pressure factor is included to achieve even higher levels of cognitive load. As mentioned, studies involving time-constrained tasks are likely to produce cognitive/psychological stress as well as cognitive load and therefore have largely been excluded from the present review. Span tests have also been used to raise cognitive load levels. In reading span experiments, for example, subjects were required to read fluently while answering comprehension questions (Daneman and Carpenter, 1980). An example of a speaking span experiment entailed testing the maximum number of words for which participants could successfully compose a grammatically correct sentence (Daneman, 1991). Word or digit span tests are similar to the multi-task experiments already mentioned in that they intersperse a primary word/digit memory task with a demanding secondary task (Conway et al., 2005). The 'Colour Word test' developed by Stroop (1935) has proven to be an effective means of inducing cognitive load without simultaneously stimulating stress. Participants are presented with words naming colours; however, there is a mismatch between the ink of the word and the colour it refers to. Subjects are asked to direct their attention to naming rather than reading, i.e. name the colour of the font used.

Moving outside of the laboratory a number of studies have analysed recordings of real life events containing high and even extreme levels of cognitive load. Examples include communication of fighter controllers and fighter pilots during combat flights and communicative exchanges during Australian Bushfire management (Khawaja et al., 2009; Khawaja, 2010). As pointed out above, findings from real-life studies need to be treated with caution as they are likely to demonstrate a confounding of stress and cognitive load.

Several databases are available which contain speech influenced by emotions, workload and the general affective states pertinent to human life. A majority of these have been assembled in the HUMAINE database (Douglas-Cowie, 2007). One data example particularly relevant to speech under cognitive load is the DRIVAWORK corpus which has been collected at Erlangen, Germany. Using a simulated driving task, 24 participants were recorded during three types of episodes: relaxed, driving normally, or driving with an additional mental arithmetic task. In addition to the audio data, video recordings, participant self-ratings, and physiological measures (electrocardiography, galvanic skin response, skin temperature, breathing, electromyography and blood volume pulse are also available. A second database, SUSAS (Hansen and Bou-Ghazale, 1997), predominantly includes speech under varying psychological stress conditions. However, it does present a limited amount of speech samples under two levels of workload difficulties, i.e. 50% and 75% workload.

4.3.1 Fundamental frequency (f_0)

The majority of studies that have investigated the effects of cognitive load on speech have reported a tendency for an increase in mean f_0 (Griffin and Williams, 1987; Huttunen et al., 2011; Scherer et al., 2002). In the case of Lively et al. (1993) only one of the five speakers showed a significant increase in mean f_0 and Huttunen et al. (2011) detail a mean f_0 increase ranging from an average of 7 Hz to 12 Hz in the most extreme cases. All studies agree on a reduction in f_0 variability (Johnstone and Scherer, 1999; Lively et al., 1993) and Huttunen et al. (2011) specify this decrease to be 5 Hz on average in their investigation.

4.3.2 Amplitude

Not as many studies could be located that investigated the intensity of the speech output. When producing speech under cognitive workload as compared to a control condition, Lively et al. (1993) observed an increase in amplitude and amplitude variability. Similarly, Griffin and Williams (1987) and Huttunen et al. (2011) stated an increase in mean intensity which amounted to 1dB in the results of the latter study. The significance of this increase in terms of perceptibility is not clear. Scherer et al. (2002) described an increase in higher frequency energy in their findings.

4.3.3 Temporal parameters

A major focus of analysis of speech under cognitive load has been the investigation of temporal parameters. There is agreement across the published investigations into cognitive load that SR and AR decrease in cognitive demanding situations (Bromme and Wehner, 1987; Huttunen et al., 2011; Kowal and O'Connell, 1987). As one might expect, those studies which have looked at the concept of cognitive load in terms of cognitive or time-induced stress noted a rise in speech tempo (Baber et al., 1996; Lively et al., 1993; Scherer et al., 2002; Silberstein and Dietrich, 2003). Rather than accrediting the elevation in SR to cognitive stress, Baber et al. (1996) provide an alternative explanation based on the observed increase in speech errors. All studies agree that frequency and duration of unfilled/silent pauses increases under cognitively demanding conditions (Baber et al., 1996; Bromme and Wehner, 1987; Goldman-Eisler, 1968; Greene et al., 1990; Jou and Harris, 1992; Kowal and O'Connell, 1987; Müller et al., 2001; Roßnagel, 1995). In addition to a rise in silent pauses, Rummer (1996) and Oomen and Postma

(2001) registered a higher amount of filled pauses/hesitations in the cognitive load group. Yap et al. (2011) is one of the surprisingly few studies that investigated phoneme duration concluding that vowels were longer with cognitive load. Lively et al. (1993) on the other hand noted a reduction in phrase and segment durations. Analysis of response onset time (ROT) is also relatively sparse but the general tendency appears to point to an overall increase when speaking under cognitive load (Greene et al., 1990; Roßnagel, 1995). Rochester (1973) provided an early review into the significance of pausing as a function of task difficulty. Interestingly, he also discussed a number of studies which have looked at pausing behaviour as a function of personality. He summarised the findings as follows: "Subjects scoring high in an audience sensitivity test paused more frequently when addressing an audience than did low scorers" and, "pause frequency remained constant but duration increased when utterances of subjects scoring high in concern for approval … and extroversion … were compared with the vocalizations of low-scoring subjects" (Rochester, 1973, p. 75).

4.3.4 Formants

Spectral vowel characteristics appear to have been rather less at the centre of investigation for research of speech under cognitive load. The findings of Lively et al. (1993) and Yap et al. (2011) correspond in that both studies did not uncover changes in mean formant frequencies F_1, F_2 and F_3. Huttunen et al. (2011) detect a slight tendency for F_1 and F_2 to centralise in their analysis. It is clear that a considerable amount of work needs to be invested into this aspect of cognitive load research.

4.3.5 Voice Quality

Even less investigated are parameters related to voice quality. Yap et al. (2010) reported that cognitive load can be related to a more creaky voice quality. However, they also draw attention to the presence of individual variation. It was also revealed that a few speakers adopted a more breathy glottal setting when speaking with cognitive load.

4.4 Research on clear speech and hyper-articulation

No clear definition exists in the research literature of what exactly constitutes controlled or planned speech. In fact, references to the notion of speech under 'behavioural control' are rare. As a consequence less strict classification criteria need to be applied when investigating this type of speaking context. 'Clear' speech (Picheny et al., 1986) or 'hyper-speech' (Johnson et al., 1993; Lindblom, 1990) could present possible points of departure. Lombard speech, which refers to speech produced under high background noise, may also be viewed as potentially relevant. Indubitably, speech under background noise is connected to clear speech but for the purposes of the present review, it was considered too far removed from the speech situation under question. People may choose to speak 'more clearly' for a variety of reasons and in a number of contexts. For example, clear speech may be employed when speaking in noisy environments, when communicating with foreigners or hearing impaired persons or when addressing small children. People may also pay increased attention to their speech when communicating with computerised spoken language systems, such as car navigation systems or automated transport reservation systems. In light of this, clear speech is usually regarded as an attempt to increase communicative intelligibility.

Experimentally, clear speech tended to be elicited by instructing participants to speak clearly or asking them to speak as though they were conversing with a person who had a hearing loss or who was not a native speaker of the language (Bradlow et al., 2003; Krause and Braida, 2002; Picheny et al., 1986; Smiljanić and Bradlow, 2005). Referring back to Lindblom's hyper-hypo theory of speech production, Hazan and Baker (2011) highlight the importance of communicative intent. Future studies, they stress, should distinguish between clear speech that is elicited via formal instruction and more spontaneously produced clear speech used to ensure successful and effective communication with a conversation partner. Other methods involved degrading the communication channel between interlocutors by introducing various types of noise over headphones (Cooke and Lu, 2010; Granlund et al., 2011; Grynpas et al., 2011). As explained above, studies that used binaurally applied masking noise in order to elicit clear speech production were largely ignored from the present review as they were considered too broad. In accordance with the previous discussion on speech and stress and speech and cognitive load, it is reasonable to assume that there are different types of clear speech. Krause and Braida (2004) briefly discuss this aspect but additional publications regarding this issue were difficult to locate. The literature on the acoustic and phonetic characteristics of clear speech is relatively sparse compared to other areas of affective speech research. The following overview attempts to cover most findings but more comprehensive discussions can be found in Smiljanic and Bradlow (2009) and Uchanski (2005).

4.4.1 Fundamental frequency (f_0)

Every study that measured f_0 in connection with clear speech arrived at the same result - an increase in the average frequency of vocal fold vibration (Bradlow and Bent, 2002; Chen, 1980; Clark et al., 1988; Krause and Braida, 2004; Köster, 2001; Picheny et al., 1986). In conjunction with an increase in the mean value of f_0, greater f_0 variability, usually expressed in terms of f_0 range, was also reported (Bradlow and Bent, 2002; Krause and Braida, 2004; Picheny et al., 1986).

4.4.2 Amplitude

Just as there was relatively little investigation into the intensity parameters for speech under stress and cognitive load research, reports on intensity modifications associated with clear speech are sporadic. Despite this infrequency however, the results are comparable in terms of the observed trends. Generally, there is an increase in amplitude when people speak ‘clearly’ (Chen, 1980; Clark et al., 1988; Picheny et al., 1986). Picheny et al. (1986) go even further and specify the overall increase to be 5-8 dB.

4.4.3 Temporal parameters

As was observed with research on cognitive load, the temporal aspects of speech seem to have received foremost attention in the research on clear speech, too. This time, it is the durational domain that comprises the core interest. A number of studies inform the reader of increased speech segment durations, be it vowels or consonants (Bond and Moore, 1994; Chen, 1980; Clark et al., 1988; Ferguson and Kewley-Port, 2002; Ferguson and Kewley-Port 2007; Köster, 2001; Lindblom, 1990; Picheny et al., 1986). In addition to phoneme duration, Chen (1980) also commented on an increase in voice onset time (VOT) and aspiration for ‘clear’ speech. Liu et al. (2004) illustrates the effect of speech element lengthening in terms of longer sentence duration for clear speech. With regards to SR, the general tendency of a reduction in rate becomes visible in a range of studies (Bradlow et al., 1996; Bradlow and Bent, 2002; Moon and Lindblom, 1994; Picheny et al., 1986; Smiljanić and Bradlow, 2005; Uchanski et al., 1996).

Caution needs to be employed when comparing across studies as it transpires that different methodologies were used when measuring rate, some calculating words per minute (Picheny et al., 1986), others determining average sentence duration (Liu et al., 2004) and others counting phones per second (Krause and Braida, 2004). There appears to be a lack of findings on the effect of clear speech on AR.

Pausing behaviour in relation to clear speech corresponds to that observed with cognitive load. When speaking clearly, people tend to increase the amount and length of silent periods in their speech (Bradlow and Bent, 2002; Cutler and Butterfield, 1990; Cutler and Butterfield 1991; Krause and Braida 2004; Picheny et al., 1986). Both Picheny et al. (1986) and Krause and Braida (2004) addressed the connection between an increase in speech rate and the occurrence and nature of pauses and indeed, when rate was controlled, the difference in pausing behaviour between baseline speech and clear speech disappeared (Krause and Braida, 2004).

4.4.4 Formants

Compared to the research on speech under stress/cognitive load described above, investigations into clear speech have devoted a substantial amount of attention to vowel formant frequencies. Using different methods and verbal descriptions, almost all of the studies agree on the fact that vowels tend to be more distinct in clear speech. In the majority of cases this conclusion is based on vowel space area measurements deduced from centre formant frequencies for F_1 and F_2. Clear speech was reported to show vowel space expansion in various studies including Bradlow et al. (2003), Bradlow and Bent (2002), Chen (1980), Ferguson and Kewley-Port (2002; 2007), Johnson et al. (1993). Picheny et al. (1986) verbalise the observed increase in vowel space dimension as a tendency for F_1 and F_2 to move closer to their respective target values and Lindblom (1990) expresses the phenomenon in terms of vowels becoming more distinct. Bond and Moore (1994) approached the issue from a slightly different angle comparing the speech characteristics of the least intelligible talkers, assumingly producing ‘unclear’ speech, to the most intelligible talkers producing ‘clear’ speech. They concluded that unclear speech showed the least differentiated and most constricted vowel space while clear speech was characteristic of markedly distinguished vowels. In addition to the peripheralisation of vowels, Chen (1980) also remarks on a tighter clustering of formants which he rationalises to be a possible result of the consonant-vowel-consonant stimulus data used. Monosyllabic wordlists, he argues, are likely to have provided less room for contextual variability. Ferguson and Kewley-Port (2002; 2007) reveal a little more detail stating that the overall expansion was not uniform across vowels. More specifically, it was the F_2 values of the front vowels that were most heavily involved while the back vowels failed to show a significant difference. Furthermore, it was the F_1 of the low vowels that tended to have a more substantial increase, the low vowels being more resistant to change. In contrast to this, Köster (2001) only observed a periperalisation in the back vowels, the front vowels being less affected. Krause and Braida (2004) was the single study which did not concur with the finding of an overall enlargement of vowel space area, reporting that there was no significant change in formant frequencies. The authors did perceive a narrower formant bandwidth however, which could be viewed as supporting the notion that vowels are more distinct in clear speech.

4.4.5 Voice quality

Given that the amount of research on voice quality in relation to speech under stress/cognitive load was sparse, it comes to no surprise that no literature could be located which addressed changes or lack of changes in view of this parameter from the perspective of clear speech production.

4.5 Research on stylistic variation

4.5.1 Audience design and Accommodation theory

Bell (1984) provides a skilfully laid out essay in which he addresses the dimension of style in relation to intra-speaker variation. The framework which he proposes in this paper moves beyond the Labovian notion of defining style in terms of the amount of attention paid to speech (Labov, 1966; Labov 1972). 'Audience Design', as the theory is called, collates and describes a multitude of factors that incite style shifts within speakers, and at its core the framework proposes that speakers tailor their speech mostly in response to other persons, i.e. their audience. Members of the audience take on different roles depending on whether they are directly addressed, whether their presence is known and whether or not their participation in the interaction is approved by the speaker. Being known, approved and addressed, the 'addressee' would be the chief character in the interaction followed by auditors who are known and approved but not directly attended to. Overhearers and eavesdroppers are neither addressed nor ratified and in case of the latter their presence is not even known by the speaker. Typically, addressees are physically closest to the speaker while auditors and overhearers are more distant and it is expected that due to their increased salience, it is addressees who will influence style shifts to the greatest extent.

Speech accommodation theory or Communication Accommodation theory as it was later refined was first developed by Giles in 1973 (Giles and Coupland, 1991). It aims to account for and explain speech style modifications that occur in spoken interactions. Accommodation Theory complements Audience Design in that besides the ‘person’ variable other factors are suggested to have an influence on style. The theory centres on a number of strategies including *convergence* (making one’s speech style more similar to that of the interlocutor), and *divergence* (diverging away from the interlocutor’s speech). A number of variables have been identified that influence speech style shifting including speaker external factors as, for example, place, topic, and, as already mentioned, interlocutor. In addition, more speaker internal factors such as self-monitoring, cognitive and emotional involvement, and wish for approval have been put forward (Giles and Coupland 1991). Convergence and divergence may not just be unimodal phenomena but it is likely that adjustments are made in multiple dimensions of speech simultaneously. Neither should the two concepts be regarded as mutually exclusive; speakers may come together on some parameters yet drift apart on others. Some influential empirical investigations can be found in Brown and Fraser (1979), Coupland (1984), Giles and Coupland (1991), Rickford and McNair-Knox (1994) and Tauroza and Allison (1990). For an overview of current issues within research on stylistic variation the reader is directed to Schilling-Estes (2008) and Schilling (2013).

Speech accommodation has been linked to impression formation and convergence may be used to achieve conversational rhythm which, in turn, fosters harmony between the speakers (Miller, 2007). Street (1984) discovered that failure to accommodate to the interlocutor was perceived unfavourably by listeners. The speech behaviour of forty dyads was examined in a series of fact-finding interviews and parameters measured included response latency, speech rate, and turn duration. Regarding speech evaluation, there was some evidence that greater convergence on speech rate and response latency was positively related to the competence and social attractiveness judgments of participants. Amongst other things, Dixon et al. (1994) examined the attribution of guilt as a function of speech accommodation and concluded that those suspects who showed convergence were rated as significantly less guilty than those who diverged. It follows that asynchronous conversational rhythm may alert speakers to the fact that the interaction is affected in some way. A range of factors could lead to asynchrony in conversation and deception is one of them (Miller, 2007). It is conceivable that participants may be concerned with making a good impression in order to cover up their deceit. By extension, they may work towards conversational harmony.

4.6 Summary

Table 1 – Vocal tendencies related to the three affective states discussed including psychological stress, cognitive load and hyper-control

	Arousal Theory	Cognitive Theory	Attempted Control Theory
	Psychological Stress	Cognitive Load	Hyper-articulation
Mean f_0	↗	↗	↗
f_0 variability	↗	↘	↗
Mean Intensity	↗	↗	↗
Formants F_1 and F_2	↘	?	↗
Formant F_3	?	?	?
Vowel Space Area	↘	?	↗
LTFD	?	?	?
Speaking Rate	↘	↘	↘
Articulation Rate	↗	↘	?
Silent Pauses	↘	↗	↗
Filled Pauses	↗	↗	?
Response Latency	?	↗	?
Voice Quality	creaky, tense, breathy	creaky, breathy	?

As mentioned at the outset of this chapter, psychological stress, emotional arousal, cognitive load and behavioural control have been shown to exert effects on speech. Although the exact nature of these effects has not been completely established, tendencies can be drawn out. Taking the three theoretical processes as possible models of deceptive behaviour, Table 1 lists the changes that could be expected in the speech of liars in relation to a number of acoustic and temporal speech parameters. An upward arrow '↗' indicates an increase and a downward arrow '↘' a decrease. A question mark signifies that the findings are inconclusive or that the research is insufficient in making generalisations. The predictions are, at times, contradictory across theories. For example, while cognitive theory suggests a decrease in articulation rate, the emotional approach predicts an increase. At the same time we may find overlap, however; all three theories predict an increase in mean f_0 and a decrease in speaking rate.

There are a range of situational and individual factors that could influence the behaviour displayed in connection with the act of lying. Research on stylistic variation tended to focus on the lexical/linguistic/phonological level. Rather than investigating acoustic/temporal parameters of speech, attention was given to phonological or morphological variables. Therefore, it is difficult to predict how exactly intra-speaker stylistic changes would manifest themselves on the acoustic or temporal domain of speech, and a parameter-centred description analogous to the research on stress, cognitive load and control cannot be offered. However, it may be that deception causes liars to pay more attention to their speech and the speech of their interlocutors and in turn adapt their style of speaking.

As Vrij (2008) points out, behavioural - in the present case vocal - cues predicted by the various theoretical processes may suggest that deception is occurring. The emphasis is on *may* however as people could also experience heightened arousal, cognitive load or the impulse to consciously control, manage and monitor their behaviour in connection to contexts or situation devoid of deception. Conversely, people may be lying without displaying any of the signs associated with these theories.

This chapter utilised the deceptive processes introduced in Chapter 2. It applied these to the domain of speech in order to make predictions of possible voice and speech changes associated with deception. As part of this, a detailed review of the literature surrounding ‘speech and stress’, ‘speech and cognitive load’, ‘speech and behavioural control’ and ‘stylistic variation in speech’ was presented. Chapter 5 will narrow the focus even further to describing and evaluating those studies that have investigated vocal characteristics of deceptive speech and, in particular, acoustic as well as temporal parameters.

5 Deceptive speech

5.1 Empirical studies of deceptive speech

Chapter 5 presents a detailed overview of the results of the empirical deceptive speech studies to date. In view of research questions 1 and 2 the decision was made to present the findings in terms of temporal and non-temporal/acoustic parameters. The non-temporal category includes frequency based features such as f_0, intensity, and features of the glottal waveform while the temporal categorisation covers findings on hesitations, pauses, speech errors and response onset times. The research conducted by Julia Hirschberg and Aldert Vrij and their respective colleagues will be reported on in a separate fashion as they are viewed as key drivers within deception research. Given the increasing attention on investigating the merit of response latencies/reaction time as a cue to deception these studies have also been listed in a separate fashion. Although only a few, the studies that have investigated different interviewing techniques in relation to cues to deception also received a separate heading.

5.1.1 Acoustic investigations

Anolli & Ciceri (1997) devised a rather controlled yet inventive procedure to elicit truthful and deceptive speech by asking 35 participants to describe an ambiguous picture in a dyadic communication setting. In comparison to many other studies which have tended to ignore the role of the person being deceived, the authors controlled for level of suspicion in the hearer differentiating between non-suspicious hearer (L1) and suspicious hearer (L2). All of the three f_0 parameters investigated (mean, range and variability) indicated a measureable increase in the L2 condition whereas the values of the L1 condition were fairly similar to the Truth condition. The analysis of mean, range and variability of speaking intensity also did not result in any significant differences between the three conditions: truth, L1 (Lie to non-suspicious hearer) and L2 (lie to suspicious hearer).

Buller & Aune (1987) investigated truthful and deceptive interactions based on responses to personality scales between intimates, friends and strangers. Analysis was performed on a perceptual level using fourteen human coders and amongst the vocal behaviours measured were pitch level and variety, volume and pleasantness. Rather than reporting the ratings for each individual vocal cue the different voice measures were collapsed into an overall rating of vocal activity. As far as the reader can tell from the description of the results, there do not seem to have been many significant changes in vocal behaviour as a function of deception. The only notable finding related to the observation that strangers, but not friends or intimates, were judged as less vocally pleasant when lying. There are obvious shortcomings with analysing voice based features only using perceptual rating scales and this methodological weakness is enhanced by the fact that the coders do not appear to have had any substantial training in phonetics. Added to this is the fact that the study employed a between-subjects design rather than a within subjects analysis which ought to be a fundamental requirement in the light of the inherent and learnt voice differences between speakers.

In a study by **Ekman et al. (1976)**, sixteen student nurses were interviewed after watching two short films - one pleasant film clip and one unpleasant one. After watching the pleasant film, they were asked to honestly describe their feelings while watching. However, in relation to the unpleasant imagery of the second movie they were required to conceal any negative feelings. Mean f_0 was found to increase significantly during the deceptive part of the experiment. One of the shortcomings of the experimental set up employed in the study relates to the confounding of being deceptive and experiencing negative emotions. It is impossible to differentiate whether the increase in average f_0 was due to the act of deceiving or a result of the stress incited by the unpleasant scenes of the film. In order to test the effect of deception there would have had to have been a condition in which the participants truthfully described the unpleasant film, for example. Ekman (1988) addressed this particular criticism however and stated that the motivation lay in the wish to model a particular real-life lying scenario, i.e. that of suicidal patient concealing their true feelings in order to be discharged from hospital. Further to this the author argues that the majority of social lies will be of this two-fold nature. **Ekman et al. (1991)** replicated this experiment with a bigger sample size and achieved a similar result of an increase in mean f_0.

Using the Idiap Wolf corpus described in Chapter 1, **Hung & Chittaranjan (2010)** approached the analysis of deceptive speech from the perspective of automatic detection. Amongst others, they extracted pitch and energy related features and concluded that the distribution of these is higher for liars than non-liars indicating greater speaker activity for the former. One part of the analysis that calls for caution is the fact that the results are based on a between speaker comparison. Analogous to social psychology research, the authors calculated the standard mean differences of features for each of the separate groups, in this case liars and non-liars, and then compared them against each other. As already pointed out on several occasions, speech research should normally be performed on within-speaker designs in order to account for the idiosyncratic nature of voices and speech. The database does offer the potential of performing within speaker analyses as participants were recorded during the initial rule familiarisation process prior to actually commencing the game. In view of this, there is scope for more speech research on the corpus.

An inventive card game, in which deception had to be employed in order to win, formed the basis for the study by **Liu (2005)**. Motivated to further add to the literature evaluating VSA based technologies the research focused on features related to pitch and jitter and how powerful these are in detecting deception. Using Bayesian Hypothesis testing, classification of deceptive and non-deceptive speech was performed on a speaker dependent and independent level. Although pitch proved to be somewhat useful in detecting deception in the experiment, overall, the findings highlight the flawed basis of the claims underlying VSA.

Rockwell et al. (1997a) analysed the truthful and deceptive responses to interview type questions and did not find a change in mean f_0 or f_0 range with deception. They did discover an increase in what they call 'pitch variety' which, it is assumed, refers to the f_0 variability as measured in standard deviation (SD). The authors also measured three intensity parameters- mean, range and variance, and reported a wider intensity range in the deceptive speech samples. In addition to performing an acoustic analysis of the data, perceptual coding of the three f_0 parameters was completed. This resulted in a significantly higher rating of average pitch level for the deceptive samples. The reported voice quality measures were purely based on perceptual analysis and it was concluded that deceptive voices sounded less pleasant than truthful ones. Articulatory precision was mentioned as one of the vocal behaviours coded but the paper failed to report on these findings. Given that the acoustic analysis did not demonstrate a significant change in mean f_0 between the deceptive and non-deceptive responses, there appears to be a discrepancy between the two modes of measurement, i.e. acoustic analysis and perceptual coding. This is not necessarily surprising and in a follow-up paper Rockwell et al. (1997b) addressed this issue further by raising important considerations for human- versus machine-based deception detection. The reader may like to refer back to Chapter 3 in which an overview is given of deception detection, both machine- and human-based.

During an interview situation in which participants had to tell the truth and lie about different topic areas, **Streeter et al. (1977)** noted an increase in mean f_0 by an average of 3.3 Hz for the lie condition. This elevation in f_0 mean in the lie condition rose to 5.4 Hz for those participants who experienced an arousal manipulation as opposed to the 'non-aroused' sample.

Suzuki and colleagues investigated the usefulness of f_0 for deception detection by measuring the responses of three criminals to polygraph testing. The authors concluded that the relationship between f_0 and deception was not 'clear' (Suzuki et al., 2002, p. 131). In the same way that f_0 did not show a meaningful pattern, neither did the intensity calculations prove to be for deception detection. The paper failed to give a transparent description of the method and instrument of analysis and therefore the results need to be treated with caution.

A paper which is frequently cited in the literature surrounding empirical studies of acoustic correlates of deceptive speech relates to the research of Motley (1974). Unfortunately, numerous attempts to retrieve the paper failed and therefore the study can only be reported based on the description included in secondary sources. Given its extensive citations, it was decided to include the study as part of the current literature overview, nevertheless. Cestaro and Dollins (1994) reported that participants in Motely (1974) were subjected to the 'playing card paradigm' as commonly employed in fMRI studies described in Chapter 3. The monosyllabic 'no' responses of twenty participants were analysed for fundamental frequency, intensity, duration, formant structure and harmonics. It is stated that only the duration measure resulted in a significant difference between the truthful and deceptive 'no' responses. The direction of the change was not specified however. In a similar fashion, **Cestaro and Dollins (1994)** analysed truthful and deceptive monosyllabic 'no' responses for changes in mean fundamental frequency, mean intensity and response duration. Following an anagram task, 28 subjects performed a peak of tension polygraph test in which half of the participants had to deny completing the number search task. The authors concluded that none of the voice measures analysed reliably differentiated between the truthful and deceptive responses. The reader is alerted to the fact that Cestaro and Dollins (1994) employed a between subjects paradigm and therefore the findings need to be viewed with caution.

5.1.2 Temporal investigations

The results of the acoustic analysis in the study by Anolli and Ciceri (1997) were already summarised but the research also included analyses of temporal aspects of speech. The study showed an increase in frequency of unfilled pauses when subjects were deceiving a non-suspicious interlocutor (L1) but no significant pattern emerged with regards to duration of filled and unfilled pauses, Speaking Rate or Articulation Rate. The authors discussed the findings by drawing on the simultaneous involvement of emotional, cognitive and social factors which shape may shape liars' behaviour.

Arciuli et al. (2009) adopted a narrower investigative focus by specifically looking at the role of the hesitation marker 'um' in low-stakes laboratory elicited lies and real life high-stakes deception scenarios. The laboratory setting employed a false opinion paradigm. Participants were required to give either their true or false opinions on a selection of topics that were salient to them. The results met the authors' expectations in that there were significantly less 'ums' in the deceptive statements compared to the truthful ones. The real-life deceptive data were taken from telephone conversations that provided case evidence for a murder trial (*the People of the State of California v. Scott Peterson*). Akin to the results of the laboratory study, the frequency of 'um' significantly declined in the deceptive utterances of the convicted murderer compared to his truthful statements. Two explanations are theorised to account for the observed decrease in 'um' during deception. It may be that the decline in hesitations is a sign of attempted behavioural control. Alternatively, it could be argued that the attenuation of 'um' during lying is a subliminal phenomenon generated by the lack of authenticity inherent in deceptive messages. Increasingly, research has shown that 'ums' are not merely signs of hesitation or production error in speech but, rather, signal special lexical meaning and form an important aspect of natural communication (Clark and Fox Tree, 2002).

Working within the theory of avoidance-avoidance conflict developed by Bavelas and colleagues (Bavelas et al., 1990), **Bello (2006)** approached deception from a slightly different angle using the concept of equivocation. Equivocation refers to intentionally ambiguous, non-straightforward communication. The study investigated several 'dysfluency' variables as a function of equivocated and non-equivocated messages and results revealed an increase in speech errors and pause frequency when people were being equivocal. Contrary to the author's expectation, hesitations were more prominent in the non-equivocated samples which, he speculates, could be the result of the operation of behavioural control mechanisms during equivocation.

Asking participants to tell the truth and lie about their previous job, **Bond et al. (1985)** reported that deceptive answers tended to contain fewer hesitations, a finding which is in accordance with that of Bello (2006). Frequency of speech errors, on the other hand, did not show a significant change as participants moved from truthful to deceptive responses. The authors acknowledged the inconsistency in the deception literature with regards to non-verbal behaviours. They explained the discrepancies by highlighting the need for liars to possess behavioural flexibility in order to succeed with their lies. A later study by Bond et al. (1990) considered cultural differences in deceptive behaviour by examining groups of Americans and Jordanians who were asked to give truthful and deceptive descriptions of familiar persons. The only significant finding that emerged related to Jordanians tending to use more hesitations when lying. No such correlation was found for the American participants. The authors succinctly state that universal cues to deception did not surface in their study.

The study by Buller and Aune (1987) has already been mentioned under the section on frequency-based parameters. The authors also included a measurement of speaking rate in their vocal activity ratings and analogous to the rest of their speech findings, no significant pattern arose. In subsequent years, **Buller et al. (1989)** completed a further deception related study, this time looking at the effects of probing. Having completed a few personality scales participants were then interviewed by a suspicious or a non-suspicious receiver about their answers to the questionnaires. While half the participants were required to tell the truth, the other half was asked to lie. Vocalic behaviour was scored by human judges and analysis revealed that deceivers showed fewer speech errors and silent pauses compared to truth-tellers. No difference was observed with regards to response latency. Behavioural control and monitoring were discussed as possible reasons for the noted changes. Furthermore, the authors commented on the general effect of probing in that it led to an increase in speech errors and pausing irrespective of truthfulness. This result is of particular interest in light of research question 3 of the present study which addresses the effect of different interviewing techniques on deceptive speech cues. In a similar line to Bello (2006), **Buller et al. (1994)** deviated from the conventionally studied falsification type of deception and focused on the alternative strategy called equivocation. Revisiting data previously collected by Bavelas et al. (1990) human raters coded several verbal and non-verbal behaviours of which silent pauses, hesitations, speech errors and response latency were the most relevant for the purposes of this work. There was no significant effect on the frequency of speech errors or silent pauses but equivocating senders used more hesitations and showed shorter response latencies compared to truth-tellers. Another finding amongst their results is the increase in vocal tension following equivocation. Unfortunately, the authors do not specify what exactly constitutes 'vocal tension' and, indeed, how accurately lay listeners could have measured this concept. Some of the data provided the opportunity to perform a three way comparison between falsification, equivocation and truth-telling. This concluded in the observation that equivocal and falsified messages share similar behavioural profiles when evaluated against truthful communication. The authors appraise the findings with reference to interpersonal deception theory.

Cody et al. (1984) decided to pay closer attention to the effect of preparation time on a number of behavioural aspects when being deceptive. Subjects participated in dyadic conversations in which they had to respond to twelve pre-assigned questions about their personal details and social interests. The group allocated to the deceptive condition was required to respond dishonestly to one of the twelve questions. Members of the group were then encouraged, by the communication partner who was a confederate, to follow up their deceptive answer with two more deceptive responses which were more elaborate narratives. The truth-telling group was merely asked to respond honestly to all the questions including the two elaborated responses. The data consisted of three types of responses for truth-tellers and liars. Namely, a prepared response to the critical question, a spontaneous narrative response immediately following the critical response and lastly, a delayed narrative response produced once the twelve question checklist had been completed. A range of temporal variables were coded including response latency, speech errors, speech rate and frequency as well as duration of pauses. Speech rate was measured by dividing the number of words spoken by the message duration and filled and unfilled pauses were collapsed under the umbrella term 'pauses'. With regards to the prepared critical response, liars exhibited shorter response latencies and slightly more speech errors compared to truth-tellers. Both the immediate and delayed narrative response showed an increase in pausing duration during lying. In addition to this, when lying, the immediate response demonstrated longer latency periods while the delayed response exposed more speech errors. None of the three types of responses, whether critical, immediate or delayed, resulted in a difference between deceivers and non-deceivers with regards to pausing frequency or speech rate. One noteworthy aspect which the authors highlight is that truth-tellers as well as liars required added processing time for the delayed response, as inferred by the latency period. This underlines the importance of taking into consideration the influence of retrieval time when assessing and interpreting hesitation in speech. On a broader level this indicates that truth-tellers and liars may show the same or similar behaviour, yet for different reasons. This observation will be of use when evaluating research question 4 of the present study which focuses on the behaviour of truth-tellers when placed in a similar scenario as that of liars.

Davis et al. (2005) secured access to real life suspect recordings which they then used to investigate a number of deception related subject matters, behavioural cues being one of them. Two of the four temporal speech parameters analysed displayed significant differences between truthful and deceptive utterances. While frequency of hesitations was negatively correlated with deception, number of speech errors showed the opposite effect. No difference occurred for speech rate or pausing frequency. One novelty of the methodological design of the study can be seen in the attempt to control for the incriminating potential of the topic that is discussed. In other words, the interview questions were classified for their level of potential threat to the suspect and then analysed according to the former categorisation. Of note is the authors' contention that their study offered little evidence of the presence of cues indicative of behavioural control or cognitive load during the deceptive excerpts. This not only reflects the intricate nature of deceptive behaviour in general, but also underlines how small methodological differences can lead to considerable differences in findings.

In **DePaulo et al. (1982)**, participants were asked to give descriptions of several persons they were familiar with, with some depictions being truthful and others not truthful. Frequency of hesitations, speech rate and speech errors were coded. Only the latter revealed differences between honesty and dishonesty; deceptive utterances included more speech errors. In addition to the actual cues, the researchers also investigated the characteristics of what participants perceived to be signs of deception in speech. People were believed to be deceptive when they made more speech errors and hesitations and when they spoke with a slower tempo. This demonstrates that lay judges do place weight on temporal aspects of speech when making sincerity evaluations.

As indicated in Chapter 2, deception takes different guises and research designs should aim to move beyond dichotically dividing statements into truths and lies. Rather, a multi-dimensional approach should be adopted that accounts for the variation found within deceptive communication. McCornack (1992) expressed this in terms of information manipulation theory which conceives of deception as a departure from Grice's conversational maxims (Grice 1989).

Ebesu and Miller (1994) embraced this framework by examining the behavioural profiles associated with deceptive messages that differ along the dimensions of truthfulness as well as directness. More specifically, in four separate interview scenarios participants were required to produce outright as well as implied truths and lies. Following Buller and Burgoon's (1996) intercultural deception theory, the dependent measures were divided into strategic and non-strategic cues. Perceptual analysis revealed no differences between deceivers and non-deceivers for frequency of pauses and hesitations. Only one of the four scenarios showed a significant truthfulness effect for speech rate - deceivers spoke with a slower tempo. Regarding directness, the results illustrated that when people responded with vague and unclear language they spoke faster and used less pauses. The authors account for the findings by drawing on the appropriateness of the distinction between strategic/non-strategic cues, and they further reason that deceivers will vary in the way they combine and mix those features depending on psychic state and social context. The paper finishes by pointing out the need for controlling methodological aspects such as the characteristics of the deceptive scenario used and the type of deception elicited.

Miller and Stiff (1993) first introduced the concept of sanctioned and unsanctioned lies in experimental research. The sanctioned/unsanctioned distinction refers to lying that has been approved by the experimenter and lies that are produced of participants' own volition (see Chapter 2). **Feeley and DeTurck (1998)** investigated how this type of motivational difference affects deceptive behaviour. The experiment was designed akin to the 'Exline' procedure already described, and participants were assigned to three veracity conditions- truth, sanctioned lie, and unsanctioned lie. Participants worked in groups of four to solve as many anagram tasks as possible after which they were interviewed about the strategies that they used for solving the word puzzles. Unlike the truth-telling group, the liars were exposed to a cheating manipulation during the experimental task. While the truth-tellers could honestly report on their strategies used, the sanctioned liars had to lie about their tactics in solving the anagrams. The unsanctioned deceivers were placed in the position where they had to decide whether to convey the truth or deceive. Behavioural coding was performed on five vocal parameters of which three (speech rate, response latency and frequency of silent pauses) did not show any significant variation between the three conditions. However, compared to the truth-telling condition, frequency of hesitations and number of speech errors increased when lying was endorsed. The opposite effect occurred for unsanctioned lies which displayed less speech errors. The authors acknowledge the importance of communication appraisal and how individuals' behaviour will be shaped by weighing up situational difficulty against their perceived ability for coping with it.

Using an original experimental design involving a staged event **Granhag and Strömwall (2002)** were interested in the consistency of truth-tellers' and liars' behaviour over multiple interview sessions. Having witnessed a mock murder scenario participants were required to either truthfully recall what had happened or to distort the real state of affairs and provide a false account. Both deceivers and non-deceivers were subjected to three interrogations spanning over eleven days. The single significant difference between truth-tellers and liars occurred in pausing behaviour; liars used pauses less frequently. As to speech rate, response latency and frequency of hesitations veracity did not cause an effect. A predictable yet noteworthy observation related to effect of elapsed time on behavioural features. For truth-tellers and liars alike, speech rate steadily decreased over the three sessions and both showed considerably more hesitations in the last interview compared to the preceding two. The authors draw on the theory of attempted behavioural control to explain the liars' more rigid and stiff appearance. It is furthermore argued that 'time elapsed' seems to strengthen deceivers' need for increased behavioural monitoring.

In a similarly imaginative experiment **Strömwall et al. (2006)** investigated nonverbal behaviour during mock police interrogations. Truth-tellers could honestly describe their immediately preceding activities (making a purchase at a convenience shop) but liars had to distort their true actions (buying/selling drugs) and convince the interviewer that they were doing what the truth-tellers were, i.e. buying a sandwich or soft drink. Analysis took the form of coding a range of temporal speech features including speech rate, number of hesitations and frequency and duration of silent pauses. None of these resulted in significant differences between the truth-tellers and liars. The researchers present an illuminating discussion of the findings. Contrary to the common conception that high-stakes settings will result in more salient differences between liars and truth-tellers, the paper argues that both truth-tellers and liars will experience the situational pressure to the same extent resulting in an increased self-presentational awareness for both. It follows that, in order to appear credible and sincere, deceivers and non-deceivers alike will adopt similar strategies which moderate possible non-verbal behavioural differences. In view of research question 4, these findings and related discussion are worth keeping in mind.

Heilveil and Muehleman (1981) explored deception related cues in a psychotherapy setting rather than the commonly used criminal interrogation context. In two consecutive interviews, one of which was with a therapist, participants responded honestly and dishonestly to sets of five questions. Coding of the data showed frequency of speech errors and response latency to be positively correlated with deception. Frequency of silent pauses, on the other hand, turned out to be not significant as a function of veracity.

Klaver et al. (2007) were interested in the influence of personality factors on deceptive behaviour and, in particular, the relationship between psychopathy and nonverbal indicators of deception. A sample of incarcerated offenders was asked to provide true and false accounts of two different criminal transgressions. Truthful statements were elicited by asking participants to describe the offence that led to their current conviction; for the deceptive samples, participants had to provide a falsified account of a crime re-enactment video. Behavioural coding revealed an increase in the number of speech errors and hesitations as a function of deception which, the authors point out, is consistent with the cognitive theory of deception. Psychopathy as such did not lead to any significant main effects for the parameters investigated; however there was a tendency for psychopathic individuals to speak faster while lying. In summary, the results failed to convince that deception manifests itself differently in psychopathic individuals compared to non-psychopaths, at least on the behavioural plane. Nevertheless, the merits of studying offender populations are paramount.

One of the earliest studies into behavioural cues of deception can be found in **Knapp et al. (1974)**. As well as completing a scale assessing Machiavellian personality traits, participants were videotaped while voicing their true and false beliefs about an issue salient to them. The results regarding the vocal features analysed were unremarkable in that there was no difference in frequency of speech errors or silent pauses between deceivers and non-deceivers. Most interestingly, no significant interactions could be observed between deception and Machiavellianism.

A similarly explorative study was performed by **Kraut (1978)** who videotaped participants telling the truth and lying in simulated job interviews. Neither frequency of speech errors nor number of hesitations correlated with deceptiveness. In a follow up experiment, the author further studied the role of pausing in relation to deception. The stimulus material comprised a recording of a mock job interview in which one of the interviewee’s answers was manipulated experimentally with regards to response onset time. The manipulation took the form of a long silent period of seven seconds inserted between the end of one of the interviewer's questions and the interviewee's corresponding answer. Judges were then asked to rate the truthfulness of the interviewee and the results turned out to be rather complex in that the experimental manipulation affected honesty judgements in two directions. On the one hand, the increased response latency fuelled judges' suspicion if they were already suspicious that the interviewee was not forthright. On the other hand, if judges believed the interviewee was telling the truth, the longer response onset time further strengthened that opinion. The authors concluded with highlighting the necessity of paying close attention to contextual factors when making honesty judgements based on behavioural observation. Two years later Kraut & Poe (1980) extended their research into a setting which is closer to forensic or criminal reality. Participants were videotaped in mock airport customs inspections and were required to tell the truth or lie about the possession of illegal imports. The data was then presented to judges who were to rate the truthfulness of the passengers. The judges, indeed, rated some passengers as more suspicious than others; however, behavioural analysis including frequency of speech errors and response latency did not reveal and differences between smugglers and non-smugglers. Nonetheless, the judges' decisions and the coded behavioural data from the truthful and deceptive passengers provided for correlation analysis which offered insight into which cues were regarded as indicators of honesty or dishonesty regardless of the actual veracity of the passengers. It transpired that passengers were more likely to be deemed suspicious if they hesitated before answering, i.e. the longer the response onset time the higher the impression of dishonesty.

Concerned with testing the effect of behavioural training on deception detection accuracy, **Levine et al. (2005)** performed four research studies, one of which was a behavioural coding study. The stimulus data involved truthful and deceptive attitudinal responses by one male and one female speaker. Vocal parameters analysed included speech rate, frequency of speech errors as well as silent pauses and response latency. In comparison to the truthful statements, the deceptive utterances demonstrated shorter response onset times, fewer speech errors and a smaller amount of silent pauses. The authors highlight the variability in research findings in relation to pausing behaviour as a function of deception and instead of minimising this variability the authors take the crucial step in emphasising its significance. Given that deceptive behaviour is affected by contextual, social and psychological factors the authors stress the challenges for the viability of deception detection training based on assessments of non-verbal behaviour.

In a series of three experiments, **Mehrabian (1971)** studied the behavioural concomitants of deceitful communication. Experiment 1 required participants to give their true and false views on a topic about which they had strong feelings (abortion) and to convince the listener that they were communicating their true beliefs in both cases. Calculations of speech error rate did not lead to any significant differences. Experiment 2 employed a similar methodological procedure to experiment 1 in that participants had to tell the truth and lie about their opinion on a variety of social issues in front of a listener judge. In addition to the actual lying and truth-telling, participants also had to role play what they understood to be a deceitful and truthful communication, i.e. telling the truth in a way as to make the listener believe one is lying and vice versa. Results indicate that speech rate decreased when participants were actually lying and when they were role-playing deceit about an actual truth. The methodology of the third experiment was based on the 'Exline' procedure (described in Chapter 2) in that some of the subjects were implicated in cheating during the performance of the experimental task (communicating by extrasensory perception). In a following interview subjects were asked about their strategies to achieve successful completion of the task resulting in half responding honestly and the other half, if they decided to agree to the cheating, lying. In contrast to experiment 2, speech rate was measured to be higher when communicators lied.

Miller (2007) expanded the line of research that was used in the study by Kraut (1978) and included a second parameter, namely speech rate, which would become subject of experimental manipulation. Two actors were recorded during their rendition of a scripted dialogue between a role-played dating couple. Timings of response latencies were controlled to sound 'too early' or 'too late' and speech rate was adjusted to a faster tempo. Listeners were then asked to evaluate the sincerity of the question-answer pairs. Responses that carried late latencies were consistently perceived as lies; no other significant effect occurred. The author concluded that response time carries more weight over speech rate when making veracity judgements.

Porter and Yuille (1996) deviated from the widespread focus on the non-verbal cues to deception to take a closer look at the verbal aspects of deceptive statements. Participants performed a mock-theft or an innocuous task and then completed a security interview in which they were questioned about the theft. Those who engaged in the innocent activity were required to truthfully recall their actions. Those who stole the money however were told to provide a) a partially deceptive account, b) a completely false alibi or c) a truthful confession. Reality Monitoring (RM) and Statement Validity Analysis (SVA) were used to score the interviews on a range of verbal criteria. The single cue related to the vocal parameters of interest to the present study was frequency of hesitations and it showed no significant differences between truthful and deceptive accounts. The authors stress the limited generalisability of their findings to real forensic settings and argue for the need for cross validation studies using diverse methods of enquiry. **Porter et al. (2008)** presented an example of this by comparing offenders and non-offenders on behavioural cues shown during deception and truth-telling. Samples of convicts and students provided videotaped accounts of true and fabricated emotional events which were then coded on verbal and non-verbal criteria. Neither frequency of hesitations and silent pauses nor speech rate reached statistical significance as a function of veracity. In terms of differences between sample types, there was a slight tendency that offenders increased their speech rate when deceiving while such pattern was absent from the student data. The results support the authors' hypothesis that offenders would feel comfortable with the act of lying which would partly manifest itself in vocal behaviour devoid of noticeable changes in speech rate or pausing frequency between truths and lies. The findings in relation to the behaviour of the students did not meet the author's predictions however. Rather than showing signs of increased cognitive load or behavioural control, the student sample did not differ from the offenders on the majority of cues investigated.

Riggio and Friedman (1983) encouraged participants to honestly and dishonestly describe a set of pictures. In order to avoid the confounding factor of emotional involvement, pictures were chosen as they were thought to be sufficiently neutral stimuli. In their methodology section the authors list speech rate, and frequency of hesitations and speech errors as part of their behavioural analysis. As an alternative to reporting the findings of each individual cue, the authors performed factor analysis by combining discrete cues into larger behavioural units. While this might provide for a better representation of the strategies used by lay observers when attempting deception detection, it impedes the presentation of the results for each of the cues in isolation. Consequently, the reader cannot be certain how the three temporal parameters were affected. Since the three measures were not named as significantly distinguishing truthful from deceptive descriptions in the discussion of the results, it may be assumed that they behaved similarly across conditions.

The research of Rockwell et al. (1997a) has already been introduced above. In addition to the various frequency based parameters measured, the authors also examined speech rate, latency period and fluency. The methodology section also lists 'internal pauses' as a parameter under investigation but the paper did not report on these findings. The results substantiated the influence of increased cognitive processing mechanisms in connection with deception as lying resulted in a slower speech rate and longer response onset times. It was also reported that liars appeared less fluent but which parameter was associated with fluency, i.e. whether speech errors, hesitations or silent pauses, was not clearly defined. In interpreting the findings the authors drew on the concept of cognitive load as well as interpersonal deception theory. For example, speech rate, it was argued, could be a side effect of the increased cognitive difficulty associated with lying. On the other hand, it could also be a consequence of liars' conscious control of their image and behaviour projection.

As briefly described in Chapter 3, the guilty knowledge test and related paradigms tend to rely on global comparisons between behaviours shown prior to and succeeding a critical segment. **Stiff et al. (1994)** deviated from these traditional pre-post models arguing that they fail to account for peoples' skills in adapting to communicative changes. Rather than showing abrupt and permanent changes in behaviour after the introduction of a critical probe, communicators are more likely to exhibit momentary departures before returning to their baseline behaviour. In order to capture these temporary changes, the authors employed time series analysis in conjunction with an impulse intervention model to investigate response latencies as a function of veracity. The data comprised videotaped interviews with students about academic integrity. Prior to these interviews, the students were given an exam they had taken to mark for themselves and while some of the students graded their tests honestly, others cheated by inflating their actual scores. The interview protocol contained three critical questions directly addressing the potential cheating incident. Although analysis failed to reveal an effect of message veracity on response onset time, several informative findings were discovered. The study converged with previous findings that response latencies are correlated with cognitive effort; the higher the cognitive involvement the longer the onset of replies. Response latencies to the critical questions were longer but, over time, participants adapted to the high intensity nature of the more challenging probes and returned to their baseline latencies. Hence, the authors encourage the use of time series models over pre-post models in order to capture speakers' communicative adaptability. More generally, it emerged that response latencies were highly variable between individuals regardless of the experimental manipulations.

5.1.3 Julia Hirschberg and colleagues

In a six year funded project, Julia Hirschberg and her team were working on automatic methods of distinguishing deceptive from non-deceptive speech. For the purposes of this work they collected and annotated a large corpus: the CSC deception corpus which was introduced in Chapter 1. A number of publications emerged from the grant addressing various questions in relation to deceptive speech. These are described in turn below.

In line with approaches employed for automatic speech recognition, **Hirschberg et al. (2005)** extracted and modelled a range of pitch, energy and duration related features which they classify as acoustic/prosodic. In addition to this they also computed lexical features and speaker dependent cues. They then ran machine learning experiments on the different feature sets, firstly for each in isolation and then in combination. The accuracy of the acoustic/prosodic feature set in discriminating between deceptive and non-deceptive speech amounted to 61.5% and the predictive power of the lexical set in classifying deceptive speech correctly summed to 61%. Combining the acoustic/phonetic and lexical feature sets the classification accuracy could be increased to 62.8% but the largest improvement in accuracy was achieved by adding the speaker dependent feature set resulting in a value of 66.4%. The authors conclude that, compared to the worse than chance accuracy displayed by human judges in differentiating between deceptive and non deceptive speech, the automatic method seems to carry potential.

Enos et al. (2006) subsequently compared deception detection using automatic methods to the deception detection ability of human judges. The study corroborates previous findings in that judges performed worse than chance and as a consequence the authors offer an encouraging outlook for automatic methods. They also highlight evidence suggesting individual differences in judges' abilities to detect deceit and offer personality differences as a possible influential factor.

Interesting results emerged from an investigation into silent and filled pauses as a function of deceptive speech **(Benus et al. 2006)**. The frequency of both silent pauses and hesitations were significantly higher in truthful utterances compared to the deceptive ones. On the other hand, the analysis did show longer response latencies and longer durations of silent pauses in the deceptive samples. Furthermore, differentiating between the nasal (um) and non-nasal (uh) hesitation variant, the results illustrated a correlation between number of 'um' and deception. The authors refer to the interrelationship between cognitive load and behavioural control in regulating and influencing speech when people are being deceptive.

Torres et al. (2008) presents the single study that investigated glottal waveform features for deceptive speech. Using the CSC corpus, the authors extracted thirteen glottal parameters including, for example, amplitude quotients as well as open- and closing quotients from nineteen speakers. Overall, the analysis revealed a remarkable amount of variation across individual speakers for the type of glottal features selected as a function of deception. However, the authors emphasised that the pitch parameter carried one of the lowest selection frequencies suggesting that pitch alone might not be as useful a feature as often thought. In addition, inter-speaker analysis proved H_1-H_2 to be the one parameter most consistently selected. The authors take on a relatively encouraging outlook and encourage future work on the glottal source feature in relation to deceptive speech.

5.1.4 Aldert Vrij and colleagues

Aldert Vrij is one of the key figures in deception research and his publications have proven to be influential in both theory and practice. Considering the field from a social psychology perspective his main explorations focus on the visual cues to deception and human deception detection ability. However, his research efforts have also ventured into the speech domain by analysing temporal features including rate, response latencies, speech errors and hesitations.
When interpreting the following results, it is recommended that the reader keeps in mind some of the drawbacks in relation to a) the use of 'between subject designs' and b) perceptual coding of speech parameters rather than taking objective measurements.

Vrij and Winkel (1991) investigated cultural differences in non-verbal indicators of deception in simulated police interviews by analysing a sample of Dutch and Surinam subjects. One group of participants had to pretend to have shoplifted a set of headphones which they were to hide in their pockets. The other half did not 'steal' the headphones and, consequently, was not required to conceal anything. Subsequent to this, both groups were interviewed by a uniformed police officer, either of Dutch or Surinam origin, who they had to convince that they did not possess the 'stolen' headphones. Analysis revealed an increase in number of hesitations and a faster speech rate for the deceptive group. No change was observed in frequency of speech errors or pitch as a function of deception. While Surinam and Dutch speakers differed in a number of vocal characteristics, ethnic origin and deception did not show a significant interaction effect. The authors acknowledge that the increase in speech rate is inconsistent with the results of a number of other studies and ascribe it to the fact that participants were able to pre-plan their lies.

With a slightly different research angle in mind, **Vrij (1995)** modified the design employed in **Vrij and Winkel (1991)**. In addition to a brief baseline recording, participants underwent two interviews, for one of which they were given a set of headphones to carry in their pockets. In both interview a uniformed police officer questioned the participants about the possession of the headphones and it was the participants' task to convince the officer that they did not carry the object with them. This resulted in each participant providing one baseline recording, one honest interview and one deceptive interview. Based on verbatim transcripts, two analysts coded the tapes for frequency of hesitations and speech errors as well as changes in pitch of voice. Compared to their baseline speech participants showed more changes in voice pitch in both interviews. The changes in pitch were more striking in the deceptive condition which also revealed a decrease in frequency of hesitations compared to the baseline and truth. The paper does not spell out what exactly is meant by 'change in pitch' but it is assumed that the authors refer to an increase in mean pitch. No differences emerged with regards to frequency of speech errors. The authors interpret the findings by drawing on the theory of behavioural control. In this regard, the decrease in hesitations is seen as a result of over-control. The changes in voice pitch do not necessarily fit the picture of hyper-control however, but are more in line with emotional involvement. Deception in this study came in the form of concealments rather than falsifications and the paper highlights the moderating effect of 'lie difficulty' in eliciting behavioural changes during deception.

Pursuing the idea of lie complexity introduced in Vrij and Winkel (1991) and Vrij (1995), **Vrij and Heaven (1999)** investigated non-verbal behaviour in relation to easy as opposed to more difficult lies. Participants were shown a video and were then asked questions varying in degree of complexity about aspects of the video. The 'easy' questions merely inquired about physical characteristics of the people who featured in the film, whereas the more complex questions required participants to use their imagination in interpreting the event in the video clip. The design was structured in such a way that participants had to produce an honest and a deceptive response to each type of lie, resulting in two truthful and two deceptive answers per participant. Frequency of hesitations and speech errors were measured by coding the verbatim transcripts. The results support the authors' hypothesis that lie complexity will have an impact on speech disturbances. Difficult lies generated more hesitations and speech errors than those that were more easily produced. When the question was cognitively more demanding liars showed more hesitations and speech errors compared to truth-tellers. Interestingly however, when the question was easy, truth-tellers showed the highest occurrence of hesitations. It could be that liars over-controlled their speech behaviour in those instances whereas truth-tellers were not as concerned about portraying a 'clean' demeanour. This interpretation would be supported by the additional finding that truth-tellers more frequently raised doubts about their own memory by saying 'I think' and 'I guess'. It could further be theorized that when the lie became more complex, liars had to dedicate more cognitive effort on the lie itself rather than their behavioural appearance resulting in an increase in hesitations. The authors do underline the potential impact of individual differences and, in particular, what is perceived to be an 'easy' vs. a 'difficult' lie.

Vrij et al. (2000) performed a systematic analysis on verbal and non-verbal behaviours as indicative of deception. As part of the experiment a sample of nursing students watched a film and then answered three open ended questions about its content. Half of the participants were asked to report truthfully whereas the other half was instructed to lie. The participants were video as well as audio taped in the process. Based on the verbatim transcripts the researchers analysed speech disturbances including hesitations and errors as well as latency periods and speaking rate. Liars were found to have longer latency periods and use more hesitations compared to the non-liars. No difference between the groups emerged with regards to speech errors or speech rate. The authors tentatively suggested the potential usefulness of paying attention to signs of cognitive load when attempting to detect deception.

Recognising the slight methodological imperfections of Vrij et al. (2000), **Vrij et al. (2001a)** made transformations to the analysis paradigm by examining deception as a within-subjects factor. This time, the truthful as well as the deceptive interviews of each of the participant student nurses were analysed. The findings are identical to the ones reported in Vrij et al. (2000); there was an increase in hesitations and response latency for the liars. In addition, the within-subject analysis also revealed an increase in speech errors and a faster speech rate in conjunction with lying. In light of the cognitive theory of deception the authors hypothesised a slowing down of speech and therefore, the result regarding speaking rate came as a surprise. Not only is it contradictory to the idea of 'thinking hard', it also does not harmonize with the results of the remaining temporal features analysed or participant's self-assessment following the task. The authors do not offer an alternative explanation that would account for the increased rate for the liars but it may be envisaged that contextual and/or methodological factors, i.e. method of analysis, played a role.

Vrij et al. (2001b) elaborated on the study of Vrij et al. (2001a) by using the same methodology but analyzing a bigger sample size. The results of the two studies only corresponded with regards to latency period which was found to be longer for liars in both cases. No effect of deception emerged for speech rate or frequency of hesitations and speech errors. The authors recognise the absence of consistent behavioural cue to deception and emphasise that what might be true at a group level analysis may not be applicable when it comes to analysing the individual.

Vrij and Mann (2001) had the opportunity to complement the overwhelming number of deception-related laboratory studies with an investigation of the real-life deceptive account of a convicted murderer. Initially denying his involvement in the killing of the victim, forensic evidence later reinforced his connection to the murder forcing him to make a confession. From the police interviews, the authors extracted comparable fragments of undisputed truths and lies for analysis. Parallel to previous studies, perceptual coding of a range of non-verbal behaviours was performed by two judges. Amongst the vocal characteristics addressed were speech rate, frequency of hesitations and speech errors, and number as well as duration of silent pauses. Speech rate was assessed categorically using a 3-point Likert Scale. It is not entirely clear how the authors measured the length of pauses using auditory means only and, unfortunately, the paper does not give a satisfactory description of the method used. Overall, the findings support the notion that lying placed a greater cognitive strain on the suspect's speech than truth-telling. The extent of the cognitive involvement during the deceptive parts transpired through the presence of longer and slightly more frequent pauses, more speech errors and a slower rate of speech. Of interest also is the result that number of hesitations was not affected by the deception factor and, in actual fact, the suspect used slightly less hesitations during lying than truth-telling. This, again, might be evidence for attempted behavioural control. The authors conclude that based on the present findings, high-stakes lying seems to be influenced by processes of increased cognitive demand and behavioural control rather than signs of emotional stress.

The investigation into deception as it occurs in real life continued with a study by **Mann et al. (2002)** who analysed the behaviour of a number of suspects during their police interview. The offences comprised lower level transgressions such as theft as well as more serious crimes involving arson, attempted rape and murder. Only cases in which ground truths could be established were included in the analysis which took the shape of perceptual coding of the video tapes by two independent analysts. Pausing duration and frequency of speech disturbances were measured; however, similarly to Vrij and Mann (2001), the methods employed for the vocal analysis could be viewed in a controversial light. First of all, hesitations and speech errors were conflated and treated as one category. Given that the causes and functions of hesitations and speech errors can be very variable, a distinction should be made between the two phenomena. The second potential weakness, once again, relates to the accuracy in measurement of the durational aspects of the silent pauses. Results exposed an increase in the length of pauses during the deceptive clips but no change in frequency of speech disturbances as compared to the truthful fragments. As was the case in the previous studies, the authors tentatively postulate the involvement of increased cognitive processes during deception. They also point to the concept of individual differences and the importance of establishing an adequate starting point, for example in the form of collecting baseline behaviour for the individual, before attempting deception detection.

Vrij et al. (2004) devised an inventive design in order to study the verbal and non-verbal characteristics displayed by children and adults when deceiving. Truth-tellers participated in a staged scenario whereas liars were only given a scripted account of the same event. While the group of truth-tellers were allowed to honestly recall what had happened, the liars were asked to pretend to have lived through the scripted version of the event they were given. The standard set of temporal characteristics including frequency of hesitations, pauses and speech errors, latency period and speech rate was coded by two independent analysts based on the verbatim transcripts and an auditory assessment of the video tapes. Overall, cues to deceit were astonishingly similar across the different age groups; however, with regards to the vocal features, none demonstrated a significant interaction effect with the veracity factor and this was the case for children as well as adults. A fully convincing explanation for the absence of any significant findings for the vocal features could not be given but it highlights the variability inherent in people's behaviour when being deceptive.

The concept of heightened cognitive processes in relation to deception has repeatedly been alluded to in the studies described above. Due to the absence of an actual manipulation of cognitive load the conclusions could only be presented in speculative format. **Vrij et al. (2008)** addressed this apparent gap by empirically testing the influence of different levels of cognitive load on behaviours shown during deception. With the help of a modified and enhanced version of the design used in Vrij et al. (2004), the study collected interview data from a group of truth-tellers and a separate group of liars. Truth-tellers participated in a staged event which concluded in them being accused of stealing a £10 note. The group of liars, on the other hand, did not experience this event but were asked to steal the £10 note and then given a scripted version of the staged incident. Following this, both groups were interviewed by a uniformed police officer, who asked about their activities immediately prior to the interview. It was the participants' task to convince the officer that they did not take the money. In order to introduce the cognitive load manipulations, half of the truth-tellers and half of the liars were instructed to recall their activities in reverse order. A number of vocal features were coded including latency period, length of pauses, frequency of hesitations and speech errors and, finally, speech rate which was measured in words per minute. Latency period was defined as 'a noticeable pause of a second or more'; the same limitations that were already mentioned with regards to the analysis of pausing also apply here. Taking the results of the control condition first, perhaps surprisingly given the findings of previous studies, none of the vocal features analysed differed between the truth-tellers and liars. The authors explain the non-significance with the fact that liars had ample time to rehearse their false account which in itself was not too complex a story. In the reverse order interview, on the other hand, liars spoke slower and included more hesitations and speech errors in their account in comparison to the truth-telling group and when measured against the liars in the control group. It appears that the cognitive load manipulation was successful in that it resulted in more speech signs indicative of cognitive load and by extension more cues to potential deceit. One issue which deserves attention is the fact that the truth-tellers of the control group did not differ significantly from the truth-tellers of the reverse order group with regards to the vocal cues analysed. One might have expected a pattern parallel to that of the liars, i.e. the truth-tellers in the reverse order recall showing more signs of cognitive load than those in the control condition. Unfortunately, the authors do not present their rationalisation on this question. The study concludes with the recommendation that cognitive load manipulation may be advantageous in facilitating lie detection in suspect police interviewing.

5.1.5 Response time techniques

As explained in Chapter 2, the cognitive theory of deception rests on the assumption that lies involve more cognitive processing demands and thereby are likely to take longer in their formulation than truths. In relation to this, and given the relative ease of measurement, reaction time and response latencies have been attractive parameters for deception related research investigations since the early 20th century. At first, the idea was that word associations in response to critical stimulus words would require longer reaction times for guilty as opposed to innocent subjects. Goldstein (1923) provides a brief account of the initial thoughts on reaction time measures in relation to deception and also supplies one of the earliest studies testing the relationship. Increased reaction time was found to be a symptom of conscious deception in these early investigations. However, later discoveries revealed that reaction time is not as constant as initially thought but that it is susceptible to a number of influencing variables unconnected to lying itself. As a consequence, its initial attraction faded and it has only been in recent years that the concept has received renewed attention (Sheridan and Flowers 2010). Some of the studies described under the heading 'temporal' above included response time measures and so the reader will have already obtained an insight into the pattern of the results. Given its growing status within deception research, it was decided to devote a bit more detail to this parameter. The following will give an overview of the studies that have primarily investigated reaction time measures in connection with deception. A modified version of the Guilty Knowledge Test (GKT) or Concealed Knowledge Test (CKT), both described in Chapter 3, has been used by the majority of these.

Baskett and Freedle (1974) approached research into deception from a slightly different angle. Rather than focusing on the actual production of behaviours during deception they analysed how certain behaviours, in this case response latency, affect judges' perceptions of whether deception was taking place. Participants were asked to evaluate their own characteristic traits against a set of aurally presented adjectives by responding with either 'yes' or 'no'. The latencies of the monosyllabic responses were varied to appear shorter or faster and it was discovered that listeners were more likely to judge both experimental manipulations as instances of deception compared to the unchanged stimuli.

Almost 30 years on and using a similar methodological set-up **Boltz (2005)** replicated this finding and summarises*:* "Truthfulness was communicated through response times of an intermediate duration that varied little across different participants, whereas dishonesty was conveyed through latencies that were significantly shorter or longer" (p. 114). The author theorised that veracity judgements of interactional statements are guided by expected conversational norms as conjectured in Burgoon's expectancy violation theory (Burgoon 1993). Responses that were too quick or too slow defied what would have been expected in the given interactional setting and consequently were perceived as negative.

Engelhard et al. (2003) explored the feasibility of using the emotional Stroop task, a modified version of the well known paradigm used to study attentional processes, as a lie detection tool. The original Stroop paradigm has already been explained in Chapter 4 but to reiterate, participants are required to name the colour in which words are printed rather than recite the words themselves. The modified version aims to exploit preferential processing of word meanings by introducing 'crucial' words alongside more neutral ones and the theory is that the former will show interference effects which could be manifested, for example, in longer response times. The 'crucial' words will depend on the nature of the research topic in question and may span from words relevant to patients clinical condition to words pertaining to a crime. Accordingly, Engelhard and colleagues (id.) hypothesised that guilty participants would show slower colour-naming latencies to a crime-relevant Stroop task compared to a control Stroop version and therefore would be differentiable from innocent participants. The experimental procedure employed the by now familiar mock crime paradigm with a 'guilty' group performing a mock theft and an innocent group merely being informed about the crime. Both groups then received a guilty knowledge test as well as the two versions of the Stroop task. The results failed to support the authors' propositions in that guilty participants did not show increased interference in colour-naming latencies compared to the innocent group.

Farrow et al. (2003), too, investigated whether there is a difference between the time taken to lie or tell the truth. The experiment involved subjects answering honestly and dishonestly to sets of questions which were either presented in visual or in auditory format. Regardless of mode of presentation, response onset times were significantly longer for deceptive answers.

Harrison et al. (1978) exemplifies one of the earliest exploratory empirical studies into response onset times as a function of deception. Using a question-answer paradigm participants were asked to tell the truth and lie about their likes and dislikes and their general interests. Given the era, response latencies were recorded with rather rudimentary means by way of manual time-keeping. Results signalled that temporal aspects of speech were associated with deception and, more specifically, that participants took longer to formulate a deceptive response than a truthful one. Indeed, not only were deceptive answers longer in response latencies, longer response onset times also tended to be judged as less believable by listeners.

In comparison, **O'Hair et al. (1981)** produced diverging results with regard to response onset time. Investigating various behavioural cues in connection with prepared and spontaneous lying, the results showed that anticipated lies carried shorter response latencies compared to truthful utterances. Spontaneous lies, in contrast, did not conform to this pattern and likewise, personality did not produce a significant interaction with veracity.

Seymour et al. (2000) attempted to assess the viability of using response time measures in order to detect the possession of guilty knowledge by replicating and extending the methodological paradigm of an earlier study by Farwell and Donchin (1991) who tested the usefulness of event-related brain potentials (ERP) in exposing guilty knowledge. The reader may remember that deception detection using ERP was introduced in Chapter 3. Participants first completed a crime scenario learning task in which they familiarised themselves with crime specific information such as alias name, password and content of the forthcoming message exchange. This was then followed by the actual execution of the mock crime itself which involved the illegitimate use of a computer account to send a criminal message. Following this, participants were given a new list of 'target' phrases similar to the ones presented in the initial learning task and were asked to memorise them. The final part of the experimental procedure comprised a phrase classification task in which participants were presented with different blocks of phrases including examples from the 'target' list as well as 'irrelevant' phrases. The 'irrelevant' phrases contained three types of stimulus words: firstly, 'probe' items which participants had encountered in the earlier crime scenario learning task, secondly 'innocent-probe' items taken from a mock-crime which participants had not committed or been subjected to, and thirdly, 'non-probe' items which participants had not encountered before. During the classification test participants had to identify the 'target' words only, rejecting all other phrases as 'irrelevant'. The expectation was that respondents would be slower at rejecting the probe items than the 'probe-innocent' or 'non-probe' items because of a 'guilty knowledge effect' and, indeed, empirical support could be gained. Participants required, on average, 300 ms longer to correctly reject probe items in comparison to the other 'irrelevant' items. In two additional experiments, the authors showed that reaction times are relatively resistant to strategic control and, in that way, they weakened the concern raised in Farwell and Donchin (1991) about the utility of reaction time as a measure for detecting guilty knowledge. Although the study did not directly probe deception, its findings are certainly of relevance and support the expansion of deception related research into this direction.

Sheridan and Flowers (2010) published one of the more recent investigations into response time measures in relation to deception. Completing a series of three experiments, the authors discovered that, irrespective of the level of cognitive complexity of the response or the mode of response (manual vs. spoken), deception resulted in significantly longer response latencies.

Moreover, when answering truthfully participants were quicker to make confirmatory responses than negations. The authors provide a comprehensive discussion on the findings and offer an original explanation for the observed increase in response times when lying. The study suggests that rather than adding cognitive load on the message generation per se, as stipulated by the cognitive theory of deception, lying involves an *extra* cognitive operation which is referred to as the 'lying constant' by the investigators.

Vendemia et al. (2005) acknowledged the multifaceted nature of deception and decided to focus their research on the role of the central nervous system (CNS). They performed a series of experiments in which they manipulated CNS measures such as attention, workload and memory to test the efficiency of reaction time measures in relation to deception. Initially, participants were required to truthfully complete a set of autobiographical statements. These were then mixed with a random set of false responses generated by the experimenter and this comprised the first of the two stimulus sets. Stimulus 2 came in the form of the words 'true' or 'false' and it was the participants' task to honestly or deceptively evaluate the baseline truth-value of the stimulus 1 sentences against the stimulus 2 words by pressing corresponding keys. The main finding that emerged was that deceptive responses required longer reaction time latencies than truthful responding. The authors also assessed the effects of practice or planning on deceptive responding and, rather surprisingly perhaps, discovered that rehearsal did not seem to be effective in lowering the response latencies of deceptive answers to match the level of the truthful reaction times. The findings are promising and promote continued research into the neuro-cognitive mechanisms involved during deception.

In a series of three experiments, **Verschuere et al. (2004a)** also tested for the presence of a 'guilty knowledge effect' in relation to demands on attention as measured by reaction time latencies. Participants were given a set of ten pictures to memorise and were then told that an examiner would attempt to detect five of the memorised pictures by performing a polygraph test. The five to-be-detected pictures comprised the 'guilty knowledge' category, whereas the remaining five fell into a 'mere knowledge' group. The aim in distinguishing between 'guilty' and 'mere' knowledge stemmed from an attempt to test the underlying mechanism of the GKT. Specifically, it served to test the value of the orienting response by disentangling the effects of stimulus familiarity and stimulus significance. The polygraph examination, in actual fact, was purely a feigned set-up as the authors were primarily interested in the attentional demands of guilty knowledge rather than the physiological reactions to it. In the real experimental task, participants were presented with the ten previously memorised pictures (five 'guilty knowledge' and five 'mere knowledge') in addition to a set of unfamiliar pictures ('neutral knowledge') and were asked to respond to the visual stimuli by way of a dot probe classification task (Bradley et al., 1998). The results of all three experiments confirm the orienting hypothesis, i.e. the argument that guilty knowledge elicits an orienting response, as participants systematically responded slower on trials containing 'guilty knowledge' as opposed to 'neutral knowledge'. Findings regarding the difference in responding to 'guilty' and 'mere' knowledge trials were a little more ambiguous. If enhanced orienting really is the underlying mechanism of the guilty knowledge concept, more attention and consequently, longer reaction times would be expected during trials containing the to-be-detected pictures compared to the merely familiar pictures. Although only one of the three experiments substantiated this prediction significantly, promising trends were observed in the other two investigations which render further research valuable.

In order to get a better understanding of the underlying processes of the guilty knowledge test, **Verschuere et al. (2004b)** combined cognitive measurements (reaction time) with physiological recordings (heart rate, electrodermal activity). In particular, the authors were interested in whether the guilty knowledge effect is attributable to the popular orienting reflex or whether it is more appropriately framed in terms of the defence reflex. Although the importance of the study's theoretical focus is acknowledged, the primary interest for the current literature review lies in the reaction time measures. The methodological approach involved participants carrying out one of two mock crimes (theft or exam fraud). This was followed by a polygraph examination. Pictures related and unrelated to the executed crime were presented and participants had to conceal recognition of the crime-relevant images. In order to get reaction time measures, examinees were instructed to press the space key on a standard keyboard whenever they heard a tone during the picture presentation. In light of previous findings, longer reaction times were expected for crime-relevant pictures; however, this prediction did not manifest itself in the present data as reaction times did not differ significantly between the two stimuli. Methodological factors are suggested in order to account for the non-significance. In particular, the fact that stimuli and probe came from different modalities, i.e. visual vs. auditory, could have caused the absence of a consistent guilty knowledge effect on reaction times.

Drawing on established constructs of cognitive psychology, Walczyk et al. (2003) are the first to systematically develop a cognitive model of lie production which they term the Activation-Decision-Construction Model (ADCM). As the name indicates, the model is based on a sequence of three phases: a) an activation component, b) a decision component and c) a construction component. Once a question is heard semantic and episodic memory in relation to the question is activated during a). Generally, the truth and all the semantic and episodic information linked to it will be automatically activated. Steps b) and c) only come into play if, from the outset, respondents believe that solely telling the truth might not be in their best interest. If telling the truth is anticipated to result in detrimental effects, a decision to lie is likely to follow which will lead to the third phase of the ADCM- construction of the deceptive message. Constructing a plausible lie will require attention and cognitive effort; not only is it necessary to inhibit the utterance of the truth, social cognitive knowledge and social context constrains also need to be considered for permissible lie generation to be successful. Given then that, according to the model, lying involves these extra cognitive processes it is predicted that response time may be a generic cue to deception, and specifically that lying requires longer response latency. Furthermore, the model encourages the distinction between yes/no and open-ended questions as the latter require more cognitive involvement.

In two experiments, **Walczyk et al. (2003)** evaluated the value of their proposed theoretical concept. Experiment 1 directly tested the construction component by asking participants to either lie ('lie' condition) or tell the truth ('truth' condition) to a mixture of yes/no and open-ended questions about personal information. Between-subjects analysis showed that lying took on average 230 ms longer than truth-telling. Not directly related to deception as such, the authors also discovered that, irrespective of veracity, participants took longer to answer questions probing potentially embarrassing information as opposed to neutral issues. A third finding emerged from participant's self-report data immediately following the lie production and indicates that in the majority of cases the truth was automatically activated even if participants were pre-cued to lie. Empirical support for this anecdotal finding can be seen in Spence et al. (2001) who reported increased activation of brain areas that are linked to response inhibition when participants were lying which, the authors hypothesised, was an indication that liars had to suppress the automatically activated truth.

In their follow-up experiment the authors introduced a third condition of 'truth-lie' on top of the two initial categories 'truth' and 'lie'. Participants in the 'truth-lie' group could answer questions honestly apart from those which referred to embarrassing subject matters to which they had to respond deceptively. As well as assessing the effect of the decision component, the addition of the third group also enabled within subject analysis of the latency effect in relation to lying that arose from the between subject comparisons in experiment 1. The within-speaker comparisons revealed that it took approximately 336 ms longer to answer open-ended questions deceptively than honestly. Despite the fact that the yes/no comparisons did not reach significance the difference between veracity conditions was in the expected direction. Corroborating the results of the first experiment, lying was found to take significantly longer than truth-telling and moreover, the decision to lie increased the response time to open questions by a further 166 ms. Amongst the response time measure the authors also quantified participants' verbal efficiency levels with a view to explain potential individual differences in the act of lying. In both studies, those who showed higher levels of verbal skill produced faster lies. Overall, the experiments indicate that response time may be a viable cue to deception but, the authors concede that further testing in a more ecologically valid setting is necessary.

In accordance, **Walczyk et al. (2005)** moved their research closer to the real-life context by testing participants face-to-face rather than via a computer interface. Over two experiments comparable findings to that of Walczyk et al. (2003) were obtained with liars responding significantly slower than truth-tellers to questions about recent episodic memories. Again, this was true for both within and between subject comparisons. Face-to-face interviews produced a more striking effect for the 'decision to lie' component as it augmented the response time by over twice as much as was reported for the computer-based testing in Walzcyk et al. (2003). Yet again, potential individual differences were controlled by measuring participants' social skills. It transpired that people possessing high levels of social skills as measured by the Social Skills Inventory (Riggio, 1986) were the quickest liars. The authors conclude by tentatively proposing a cognitive framework for lie detection which is based on the use of response time and response inconsistencies as cues to deception.

Using a similar design to their previous studies, **Walzcyk et al. (2009)** continued to test their cognitive approach to deception by exploring the effects of rehearsal on response times during lying. Similarly to earlier designs, participants were required to tell the truth or lie in response to a mixture of yes/no and open-ended questions tapping into topics of relevance to general student existence e.g. academics, current residence. Rehearsed lies were elicited by providing participants with the questions shortly prior to the actual testing. As well as replicating their previous findings that lying carries longer response times, the study discovered that rehearsal reduced response times in comparison to unrehearsed lies but not as far as to the level of truth-telling. Thinking in terms of the ADCM, rehearsal should lessen the cognitive load involved during lie construction and therefore the finding met expectations.

Complementary to the commonly employed deception paradigms involving analysis of behavioural correlates, **Parliament and Yarmey (2002)** introduced an alternative method based on recognition memory of eyewitnesses. Participants watched a video of a staged crime and were then asked to make identifications of the perpetrator and victim based on photo line-ups. The task of perpetrator identification was manipulated in such a way that some participants had to protect the wrong-doer while still appearing credible and honest to the police. In essence, they were told to successfully protect the perpetrators identity. Participants who were part of the control group could perform the eyewitness identification task as truthfully and accurately as possible. As anticipated, when mock-eyewitnesses were told to lie in the identification task in order to protect the perpetrator they tended to opt for avoiding an identification altogether rather than accusing an innocent person. More strikingly, however, deceptive eyewitnesses were significantly quicker in making their identification decisions compared to the truthful ones. The authors conclude that the timing of decision making in identification tasks could potentially be a diagnostic of attempted deceit.

Similarly to Parliament and Yarmey (2002), **Zhou and Zhang (2006)** endeavoured a new approach to investigating deception by analysing computer mediated communication rather than face to face encounters. In addition to shifting the communication setting to a virtual environment, the study also attended to the effect of group size on deceptive behaviour by comparing dyads to triadic groups. The authors devised a laboratory experiment in which participants had to collaborate in a group decision task about desert survival strategies via the medium of synchronous instant messaging. Truth-tellers could honestly communicate what strategies they would employ in the imaginary scenario whereas deceivers had to lie about the true nature of the strategic approach they would follow. A text analyzer was used to derive the values of the verbal and non-verbal cues under investigation and the two variables of relevance to the temporal domain were response latency (time between receiving and sending a message) and pause interval (time between sending two consecutive messages). The results indicated that in dyads as well as triads deceivers showed shorter gaps in between sending successive messages. Only the dyadic interactions revealed an effect on response latency in that deceivers took longer to reply to a message. At first sight, the decrease in pause duration might be viewed as incompatible with the longer response latencies but the authors present a convincing interpretation. It is suggested that it may take a deceiver longer to respond to an incoming message because of the increased cognitive complexities surrounding the formulation of a deceptive argument. Once the core of the line of reasoning is in place, it is easier for deceivers to elaborate on their claims.

The last two studies highlight the fact that research into the behavioural cues of deception could benefit from a broader outlook incorporating methods and procedures from different, yet analogous, modalities.

5.1.6 Interviewing styles

Research on how different interviewing techniques affect the behaviour of truth-tellers and liars is sparse. **Vrij et al. (2006b)** explored this subject matter. Utilising a similar methodological design to Vrij et al. (2004) groups of truth-tellers and liars were interviewed with two varying styles of questioning across three interview phases. The two styles under consideration were an information gathering style as opposed to an interrogation/accusatory approach. From the vocal domain, only the frequencies of speech errors and hesitations, which were collapsed into one category labelled 'speech disturbances', were analysed. No difference was found between liars and truth-tellers; however, an effect was obtained for interview style in that participants, regardless of veracity, showed less speech disturbances in the interrogative part of the interview. The study makes an important contribution in highlighting the effect context has on behaviour.

Continuing the investigation of the viability of interviewing techniques, **Vrij et al. (2006a)** tested the Behaviour Analysis Interview (BAI) (see Chapter 3). The design used to collect the data was based on a revised version of the procedure employed in Vrij et al. (2004) and Vrij et al. (2006b). Analysis of the video data was conducted by means of perceptual coding. The single feature relevant to speech that was measured concerned the duration of time before participants answered questions which, it is assumed, relates to response latency. A second aspect that was evaluated and which appears to be associated with the vocal domain came in form of the question 'Does the interviewee sound sincere?' Unfortunately, the paper fails to specify what aspects of the voice or speech contributed to the vocal impression of sincerity. The BAI interview did not trigger a meaningful variation in either response latency or sounding sincere between truth-tellers and liars. More broadly, the authors emphasise that the assumptions on which the BAI is based could not be supported with their experimental results. Rather, they advocate an approach grounded in the cognitive theory of deception.

5.2 Summary

A summary of the findings is presented in Tables 2 and 3. Analogous to the verbal review, frequency and temporal investigations are tabulated separately. The described research studies are listed in alphabetical order together with the experimental design and method of analysis used. The findings in relation to deceptive speech are briefly summarised in verbal format. Given that each study was described in the body of the current work, the experimental designs and analysis methods were described in broad terms only (e.g. 'false opinion paradigm', 'human coders'). An exception to this are the studies that have used one of the corpora described in Chapter 1, in which case the corpus is named. If a parameter was not measured as part of the investigation the cell was filled with an '-'. There are a few investigations which were described in the body of the text but which do not feature in the tables. These are studies for which the full reference could not be located or which were not transparent enough in relation to the methodology employed. Also not included are those studies which reported subjective cues to deception, i.e. which behaviours are perceived to be indicating deception rather than objective cues, i.e. actual deceptive behaviour.

Table 2 – Overview of previous research into the acoustic characteristics of deceptive speech

Author	Method	Analysis	f_0	Amplitude	Voice quality	Remarks
Anolli and Ciceri (1997)	False information paradigm	Acoustic analysis	Increase in mean and variability in L2 only	No change in mean or variability in L1 or L2	-	Passive listener(L1), Suspicious listener (L2)
Buller and Aune (1987)	False opinion paradigm	Human coders	-	-	Less vocally pleasant	Between-speaker
Cestaro and Dollins (1994)	False information paradigm	Acoustic analysis	No change in mean	No change in mean	-	Between-speaker
Ekman et al. (1976)	False information paradigm	Acoustic analysis	Increase in mean	-	-	
Ekman et al. (1991)	False information paradigm	Acoustic analysis	Increase in mean	-	-	
Hung and Chittaranjan (2010)	Idiap Wolf corpus	Automatic feature extraction	Higher distribution of f_0 related features	Higher distribution of energy related features	-	Between-speaker
Liu (2005)	False information paradigm	Automatic feature extraction	f_0 marginally useful in automatic deception detection	-	-	
Rockwell et al. (1997a)	False information paradigm	Acoustic analysis	No change in mean, increase in variability	Wider range	Less vocally pleasant	
Streeter et al. (1977)	False information paradigm	Acoustic analysis	Increase in mean	-	-	
Torres (2008)	CSC corpus	Automatic feature extraction	-	-	Potential value of H_1-H_2	
Vrij (1995)	False information paradigm	Human coders	Increase	-	-	

Table 3 – Overview of previous research into the temporal characteristics of deceptive speech

Author	Method	Analysis	Speaking Rate	Pauses	Hesitations	Response Onset Time	Remarks
Anolli and Ciceri (1997)	False information paradigm	Speech analysis software	No change	Increase in frequency in L1, no change in duration	No change in duration	-	Passive listener(L1), Suspicious listener (L2)
Arciuli et al. (2009)	False opinion paradigm/real life data	Human coders	-	-	Decrease in frequency of 'um'	-	
Bello (2006)	False information paradigm	Human coders	-	Increase in frequency	Decrease in frequency	No change	
Benus et al. (2006)	CSC corpus	Automatic extraction	-	Increase in duration	-	Increase	
Bond et al. (1985)	False information paradigm	Human coders	-	-	Decrease in frequency	-	
Bond et al. (1990)	False information paradigm	Human coders	-	No change in frequency	Increase in frequency for J only	-	Jordanians (J) vs. Americans (A)
Buller and Aune (1987)	False opinion paradigm	Human coders	No change	-	-	-	Between-speaker
Buller et al. (1989)	False opinion paradigm	Human coders	-	Decrease in frequency	-	No change	Between-speaker
Buller et al. (1994)	False information paradigm	Human coders	-	No change in frequency	Increase in frequency	Decrease	
Cody et al. (1984)	False opinion paradigm	Human coders	No change	Increase in duration in S and P, no change in frequency	-	Increase in S	Spontaneous (S), Planned (P), Between-speaker

Table 3 (cont.) – Overview of previous research into the temporal characteristics of deceptive speech

Author	Method	Analysis	Speaking Rate	Pauses	Hesitations	Response Onset Time	Remarks
Davis et al. (2005)	Real-life data	Human coders	No change	No change in frequency	Decrease in frequency	-	
DePaulo et al. (1982)	False information paradigm	Human Coders	No change	-	No change in frequency	-	
Ebesu and Miller (1994)	False information paradigm	Human coders	Decrease	No change in frequency	No change in frequency	-	
Engelhard et al. (2003)	False event paradigm	Computer-based	-	-	-	No change	
Farrow et al. (2003)	False opinion paradigm	Computer-based	-	-	-	Increase	
Feeley and DeTurck (1998)	False information paradigm	Human coders	No change	No change in frequency	Increase in frequency for S, no change for U	No change	Sanctioned (S), Unsanctioned (U), Between-speaker
Granhag and Strömwall (2002)	False information paradigm	Human coders	No change	Decrease in Frequency	No change in frequency	No change	Between-speaker
Harrison et al. (1978)	False opinion paradigm	Manual time-keeping	-	-	-	Increase	
Heilveil and Muehleman (1981)	False opinion paradigm	Human coders	-	No change in frequency	-	Increase	
Klaver et al. (2007)	False event paradigm	Human coders	-	-	Increase in frequency	-	

Table 3 (cont.) – Overview of previous research into the temporal characteristics of deceptive speech

Author	Method	Analysis	Speaking Rate	Pauses	Hesitations	Response Onset Time	Remarks
Knapp et al. (1974)	False opinion paradigm	Human coders	-	No change in frequency	-	-	
Kraut (1978)	False information paradigm	Human coders	-	-	No change in frequency	-	
Kraut and Poe (1980)	False information paradigm	Human coders	-	-	-	No change	
Levine et al. (2005)	False opinion paradigm	Human coders	No change	Decrease in frequency	-	Decrease	
Mann et al. (2002)	Real-life data	Human coders	-	Increase in duration	No change in frequency	-	
Mehrabian (1971)							
Experiment 2	False opinion paradigm	Human coders	Decrease	-	-	-	
Experiment 3	False information paradigm	Human coders	Increase	-	-	-	Between-speaker
O'Hair et al. (1981)	False information paradigm	Human coders	-	-	-	Decrease in 'P', No change in 'S'	Spontaneous (S), Prepared (P)
Parliament and Yarmey (2002)	False information paradigm	Manual time-keeping	-	-	-	Decrease	Eyewitness identification
Porter and Yuille (1996)	False event paradigm	Human coders	-	-	No change in frequency	-	Between- speaker
Porter et al. (2008)	False opinion paradigm	Human coders	No change	No change in frequency	No change in frequency	-	

Table 3 (cont.) – Overview of previous research into the temporal characteristics of deceptive speech

Author	Method	Analysis	Speaking Rate	Pauses	Hesitations	Response Onset Time	Remarks
Rochwell et al. (1997a)	False information paradigm	Speech Analysis Software	Decrease	-	-	Increase	
Seymour et al. (2000)	False event paradigm	Computer-based	-	-	-	Increase	
Sheridan and Flowers (2010)	False information paradigm	Tachistoscope and timer	-	-	-	Increase	
Stiff et al. (1994)	False event paradigm	Computer-based	-	-	-	No change	
Strömwall et al. (2006)	False event paradigm	Human coders	No change	No change in frequency or duration	No change in frequency		Between-speaker
Vendemia et al (2005a)	False information paradigm	Computer-based	-	-	-	Increase	
Verschuere et al (2005a)	False information paradigm	Computer-based	-	-	-	Increase	
Verschuere et al. (2005b)	False event paradigm	Computer-based	-	-	-	No change	
Vrij and Winkel (1991)	False event paradigm	Human coders	Increase	-	Increase in frequency	-	Between-speaker
Vrij and Heaven (1999)	False information paradigm	Human coders	-	-	Increase in frequency for 'D', Decrease in frequency for 'E'	-	Easy (E), Difficult (D)
Vrij et al. (2000)	False information paradigm	Human coders	No change	-	Increase in frequency	Increase	Between-speakers

Table 3 (cont.) – Overview of previous research into the temporal characteristics of deceptive speech

Author	Method	Analysis	Speaking Rate	Pauses	Hesitations	Response Onset Time	Remarks
Vrij et al. (2001a)	False information paradigm	Human coders	Increase	-	Increase in frequency	Increase	
Vrij et al. (2001b)	False information paradigm	Human coders	No change	-	No change in frequency	Increase	
Vrij and Mann (2001)	Real-life data	Human coders	Decrease	Increase in frequency and duration	Decrease in frequency	-	
Vrij et al. (2004)	False information paradigm	Human coders	No change	No change in frequency	No change in frequency	No change	Between-speakers
Vrij et al. (2006a)	False information paradigm	Human coders	-	-	-	No change	
Vrij et al. (2008)	False event paradigm	Human coders	No change in 'C', Decrease in 'R'	No change in frequency in 'C' and 'R'	No change in frequency in 'C', Increase in frequency in 'R'	No change in 'C' and 'R'	Chronological (C), Reverse (R), Between-speakers
Walczyk et al. (2003)	False information paradigm	Computer-based	-	-	-	Increase in O, No change in Y	Open-ended (O), Yes/No (Y)
Walczyk et al. (2005)	False information paradigm	Computer-based	-	-	-	Increase	
Walczyk et al. (2009)	False information paradigm	Computer-based	-	-	-	Increase	
Zhou and Zhang (2006)	False information paradigm	Computer-based	-	Decrease in duration	-	Increase	Written language rather than speech

A summary of the overall trends is presented in Table 4 below. By convention, a '↗' denotes an increase with deception whereas a '↘' signals a decrease in the relevant parameter when people were lying. In cases where no change occurred between deceptive and non-deceptive speech '=' is used. A question mark is used to denote that research in relation to the relevant parameters could not be located.

Table 4 – Summary of the findings of previous research into the acoustic and temporal characteristics of deceptive speech

Parameter	Findings in relation to deceptive speech
f_0	↗, =
f_0 variability	?
Amplitude	=
Formants F_1, F_2 and F_3	?
Vowel space area	?
Glottal features	?
Voice quality	?
Articulation rate	?
Speaking rate	↗ , ↘, =
Frequency of silent pauses	↗ , ↘, =
Frequency of hesitations	↗ , ↘, =
Duration of silent pauses	↗ , ↘, =
Duration of hesitations	=
Response onset time	↗ , ↘, =

Several observations can be made from Tables 2 – 4. Firstly, there are considerably more studies investigating temporal aspects of deceptive speech than frequency based parameters. On the one hand, this may be because temporal parameters provide a better environment for deception cues to surface than frequency based parameters. On the other hand, this disproportion may be related to the fact that temporal analysis of speech is 'easier' in the sense that specialised speech analysis software is not necessarily needed.

Secondly, the reader may have noted the discrepancies between the temporal findings of deceptive speech with regards to direction of change. On the one hand, there are those studies that report an overall slowing down of speech and on the other hand there are investigations that describe an acceleration of speech. For response time investigations, a similar disparity is apparent; some studies report an increase in response time as a function of deception, others report a decrease. This conflict in findings may be the result of differences in research designs and/or analysis methods. It may also stem from varying confounding factors such as interaction environment or, perhaps, style of interviewing and type of questions asked.

Thirdly, individual variability is paramount.

In view of the scarcity of previous findings, research into the acoustic aspects of deceptive speech is of value. There is also justification for further work along the temporal domain, even though it has been more widely explored:

a) The majority of studies listed above that have reported on temporal aspects of deceptive speech have been carried out by psychologists, who may not have had formal training in phonetics or speech analysis. Often the vocal aspects were not the focus of the experimental investigation but part of a wider study into other behavioural correlates of deception. In the majority of cases a between-subjects design was employed. While this is frequently used in the Psychology discipline, it is rarely found in speech research as it presents methodological weaknesses. Speaker-specificity is an important concept that cannot be overlooked and therefore any conclusions based on between subjects comparisons need to be treated with caution.

b) In the case of the temporal investigations, analysis was almost exclusively performed by perceptual coding of the audio recordings. While frequency counts are possible using this method (e.g. number of hesitations and speech errors), accuracy in durational measurements would be very difficult to achieve with auditory assessment only. Therefore, those studies that have reported on latency periods, pausing behaviour and pitch using perceptual coding analysis ought to be interpreted carefully. Auditory judgements of voice pitch, especially when performed by non-phoneticians, are likely to be tainted by inexactness and subjectivity.

In light of this, further research using a more refined and skills-based approach is justified.

In summary, not only is it necessary to pay close attention to the methods of analysis used, it is also of importance to take into closer consideration contextual factors and 'the individual' when evaluating deceptive speech. With these factors in mind, the next part of the thesis presents the empirical research conducted.

6 Methodology

The empirical research comprising the present work consists of two laboratory experiments. The present chapter describes, in detail, the methodology of data collection and analysis for each of the two experiments including characteristics of the participants/speakers[5], the experimental procedure, the equipment used and analysis techniques. As discussed in Chapter 1, a handful of corpora are in existence which contain samples of deceptive speech. For various reasons already explained only one of these was believed to be suitable for the present investigation. The 'smell of fear' dataset forms the basis of experiment 2. As this one set of data was not sufficient for the purposes of the present work, the decision was made to collect a new database of deceptive speech which forms the core of experiment 1.

6.1 Experiment 1

Experiment 1 took the form of a laboratory experiment which aimed to elicit deceptive and non-deceptive speech. The primary goal of the experimental methodology was to replicate real-life police interviews as closely as possible and specific attention was paid to the style of interviewing used, i.e. investigative interview.

6.1.1 Participants

The data consist of an opportunity sample of ten male native British English speakers between the ages of 20 and 30. The majority were drawn from the student population at the University of York where the experiment took place. Most of the speakers were from the Northern part of England and none had any self-reported voice, speech or hearing disorders.

6.1.2 Experiment procedure

In the context of deception in a criminal setting it may be expected that detection apprehension and motivation to successfully deceive is elevated. The current experiment has been designed with this in mind. Part of the procedure was modified from Porter and Yuille

[5] The term participant and speakers will be used interchangeably in the present work.

(1996) 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 to which wardens would be able to differentiate between guilty and non-guilty participants. Further to this, the volunteers were informed that the experiment was also part of a communication research project and therefore audio data would be collected. The information that was given to the participants can be found in Appendix A. Having given written consent that they were prepared to continue with the study participants had to complete three tasks:

<u>Task 1:</u> sitting in a quiet office room ('preparation room'), participants were asked to disclose demographic details such as age, hometown and occupation, and complete a set of personality scales. The scales were employed in order to control for the fact that possible individual differences in personality and level of anxiety could affect behaviour shown during the experiment. Included in the pack of questionnaires were a) Goldberg's Big Five Scale (Goldberg, 1992) and b) a short form of the State Trait Anxiety Inventory (STAI) questionnaire (Marteau and Becker, 1992). The Big Five scale consists of 100 adjectives which describe what are believed to be the five fundamental dimensions of personality, i.e. extraversion, agreeableness, conscientiousness, neuroticism and openness. The STAI is a short self-evaluation questionnaire (6-item scale) concerning feelings of cognitive and emotional tension. On completion of the forms, which can be found in Appendix B, they were taken to an interview room where they were involved in a baseline interview. This took the form of a brief conversation between the participants and one interlocutor (the present author). The spontaneous speech elicited served as a means of control data to enable within-speaker comparisons. The interlocutor was a female native German speaker in her twenties. She was familiar with all ten participants. Topics of conversation included hobbies and interests,

academic studies, and university life in general. An example dialogue can be seen in the following exchange:

Baseline Interviewer: *'Oh, I see you like Taekwondo'. Where do you practice Taekwondo?*

Participant: *'Um, actually, just here at the sport centre. There is a guy who comes in during term time.'*

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 an alibi. Each participant had ten 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 were to fail on either, however, they would lose the extra payment and be asked to write a report about what had happened[6] (Vrij, 2008). On completion of the tasks and prior to debriefing, participants were given post-interview rating scales in which they were asked to indicate their level of nervousness and motivation during both the truth-telling and lying phases (Appendix B).

Task 3 involved an investigative interview. The interviewer[7] was a middle-aged male native British English speaker and he had no previous familiarity with any of the ten participants. He was not a professional interviewing officer but he had a background in acting which was highly

[6] The experimenter told all the participants that they convinced the security warden successfully so nobody, in fact, failed the interview. The threat of having to write a report was merely used as an added motivational incitement on top of the financial incentive.

[7] Acknowledgement and a special 'thank you' to Francis Newton for his time and efforts.

beneficial to the execution of his role. Prior to the actual experiment, he was informed of the main aspects involved in investigative interviewing and given chance to practice with a mock participant. During the interview, he encouraged the participants to describe in detail their activities in the one hour prior to the interview. The following gives an illustration of the format of questioning used:

Interviewer: *I'm investigating two items that have gone missing and I'm asking questions to all sorts of people.*

Participant: *Ok.*

Interviewer: *I'm interested in the time between 1pm and 1.30pm.*

Participant: *Right.*

Interviewer: *Can you tell me where you were at that time, what you saw, um, what rooms you were in, any details that you can think of between that time.*

Participant: *Eh, yes, um, I was in this building actually, um ...*

The interviewer was blind to whether the interviewee was telling the truth or fabricating a story. Due to the small sample number it was not feasible to randomise the order of accusations and therefore all the participants commenced the interview by telling the truth.

6.1.3 Recording setting and equipment

The experiment was conducted in the Linguistics department at the University of York and all aspects of it were kept constant between conditions and participants. 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 25 m 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 sitting 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 and the recordings (.wav format) were made using a sampling rate of 44.1 kHz and bit depth of 16 bit. The final recordings are high quality and the excellent spectrographic resolution enabled for accurate acoustic and temporal analysis.

6.1.4 Measurement equipment and parameters analysed

'*Sound Forge^{TM} Pro 10'* software (Sony Creative Software) was used for initial editing of the speech files. The acoustic analysis was performed using *Praat 5.1.44* speech analysis software (Boersma and Weenink, 2011). Sennheiser HD 280 Pro closed cup style headphones were used to carry out the acoustic and auditory analysis. The selection of parameters for analysis was informed by the literature presented in Chapters 4 and 5.

These included:

- Mean fundamental frequency (f_0 mean)
- Standard deviation of f_0 (f_0 SD).
- Mean amplitude
- Mean vowel formant frequencies F_1, F_2 and F_3
- Vowel space area estimation (VSAe)
- Long term formant distributions (LTFD) for F_1, F_2, $F_{3,}$ and F_4
- Amplitude difference between first and second harmonic H_1-H_2
- Speaking rate (SR)
- Articulation rate (AR)
- Frequency of silent pauses (PR) and hesitations (HR)
- Duration of silent pauses and hesitations
- Response onset time (ROT)

6.1.5 Pilot testing

The experiment took the form of a within-subjects design and every participant produced three types of speaking conditions – control speech, truthful speech and deceptive speech. The truthful and deceptive data was elicited during the 'security interview' in task 3; control data was recorded during the baseline interview in task 1.

Before commencing with a full analysis of the data, it was decided to perform a smaller scale pilot test in order to evaluate the methodological design and identify any problematic issues in relation to the data collected as well as the methods of analysis used. As part of this pilot study, f_0 mean, mean intensity and vowel formants F_1, F_2 and F_3 were analysed from all ten speakers. The findings are presented in summary format below. A decision was made not to include a full description of the results in the main body of the present work as these were not relied upon in subsequent discussions. The full findings can be found in Kirchhübel and Howard (2013).

The results of the pilot test suggested that none of the three parameters analysed resulted in a consistent and reliable change between baseline speech on the one hand and truthful and deceptive speech on the other. f_0 mean varied little across conditions. The f_0 SD results also did not present any strong observable trends between participants' baseline speech, their truth-telling or their lying. With regard to overall amplitude changes, the findings, too, did not offer grounds for a reliable distinction between the three speaking conditions. The F_1 and F_2 differences between conditions were not statistically significant. For F_2 in particular, there was a considerable amount of variation across the truthful speech and the deceptive condition with values increasing, decreasing or not changing. The F_3 results pointed towards a negative correlation between truthful/deceptive speech and baseline speech but it was suggested that this was the result of the inclusion of outliers in the statistical computations. When thoroughly inspecting the figures, it transpired that F_3 remained very stable across the three conditions. Not only was there a lack of significant changes for the parameters investigated but also, if change was present, it failed to demonstrate consistencies between the speakers.

On multiple occasions the homogeny in results for the truthful speech and the deceptive speech was highlighted. This was particularly striking with regards to the f_0 SD and intensity measure and to some extent also the formant behaviour. Does truth-telling and lying result in corresponding acoustic changes in speech? Or could this result have been an artefact of the methodological design employed including the characteristics of the data elicited?

After the Baseline data had been recorded participants were informed that they were going to have a 'security interview' which would require them having to tell the truth as well as deceive. For reasons connected to practicality, the truthful and deceptive speech was collected from the same interview. Furthermore, the part of the interview that necessitated lying came second for all participants. It is possible that participants' awareness of the fact that they would have to lie had a global influence on their interview speech. Therefore, it may not be feasible to distinguish the truth telling condition from the deceptive condition as straightforwardly as has been done.

With this in mind, a revision of the experimental methodology (more specifically, the method of analysis) was thought to be necessary. It was decided to only compare the baseline condition to the interview, with the view that the interview will be representative of deceptive speech as it would occur in real-life. So, rather than analysing three different files per participant (baseline speech, truthful speech, deceptive speech) every speaker provided two recordings (baseline speech, deceptive speech) resulting in a total of twenty files for analysis. In total the duration of the original, unedited recordings ranged from about 2 minutes to 12 minutes (average around 7 minutes) in the baseline and from around 12 minutes to 20 minutes (average 13 minutes) in the interview. Having implemented the methodological adjustments the findings with regards to f_0, intensity and mean formant frequencies were re-examined before continuing with the analysis of the remaining acoustic and temporal parameters.

6.1.6 Measurement technique

One needs to choose a suitable size of interval over which a given parameter is determined which is usually referred to as global vs. local measurements. The present work aimed to assess the effect of lying on various parameters of speech on an inter- as well as intra-speaker level. In order to facilitate this, local measurements had to be computed. While global measurements can be processed rather straightforwardly by taking into account the entire duration of the edited recording, local measurements require a further methodological decision relating to the kind of speech interval used. Possible domains include the syllable, the nucleus, the inter-pause stretch (IP) or the intonation phrase, for example. The syllable may be seen as a suitable unit in view that speaking rate tends to expressed as syllables per second. However, evidence of the use of syllables rather than larger units such as phrases, for example, did not appear from the previous literature on deceptive speech. Indeed, it is assumed that in the majority of cases, investigations were based on global measurements. Therefore, largely for reasons of practicality, the inter-pause stretch was chosen as the unit of measurement in the current work. Given that the data was taken from spontaneous speech the number of inter-pause stretches was not uniform across speakers or conditions but considerations of practicality and the value in having spontaneous as opposed to read speech superseded this expected shortcoming. On average, the baseline provided 92 IPs (SD = 41) and an average of 118 IPs (SD = 42) could be collected from the interview. Inter-pause stretches were marked using text grids.

Mean f_0 was measured from each of the inter-pause stretches using the inbuilt function in *Praat,* i.e. the 'get f_0' command (Figure 4). When taking the f_0 measurements, attention was given to ensure that any non-linguistic speech phenomena such as laughter or coughing, or erroneous measurements did not affect the estimations. While it was possible to obtain the minimum and maximum values of f_0 from each of the IP phrases, *Praat* did not allow for a local extraction of f_0 SD values. Therefore, the f_0 SD values are based on the averages calculated from the IPs. It is acknowledged that *Praat* facilitates easy programmability and a script could have been written in order to extract local f_0 SD measurements.

Mean amplitude (measured in RMS in decibels) was determined for each IP phrase using the 'get intensity' function of the *Praat* software (Figure 4). This ensured that f_0 and amplitude were calculated from the same speech material. As described in the beginning of Chapter 4, reliable absolute amplitude measurements are difficult to obtain. Although headset microphones were used and the recording environment and levels were kept constant across participants, participant's chosen level of speaking could not be regulated. Therefore, rather than expressing the values in absolute form, the relative difference in mean amplitude between baseline and interview is reported for each speaker.

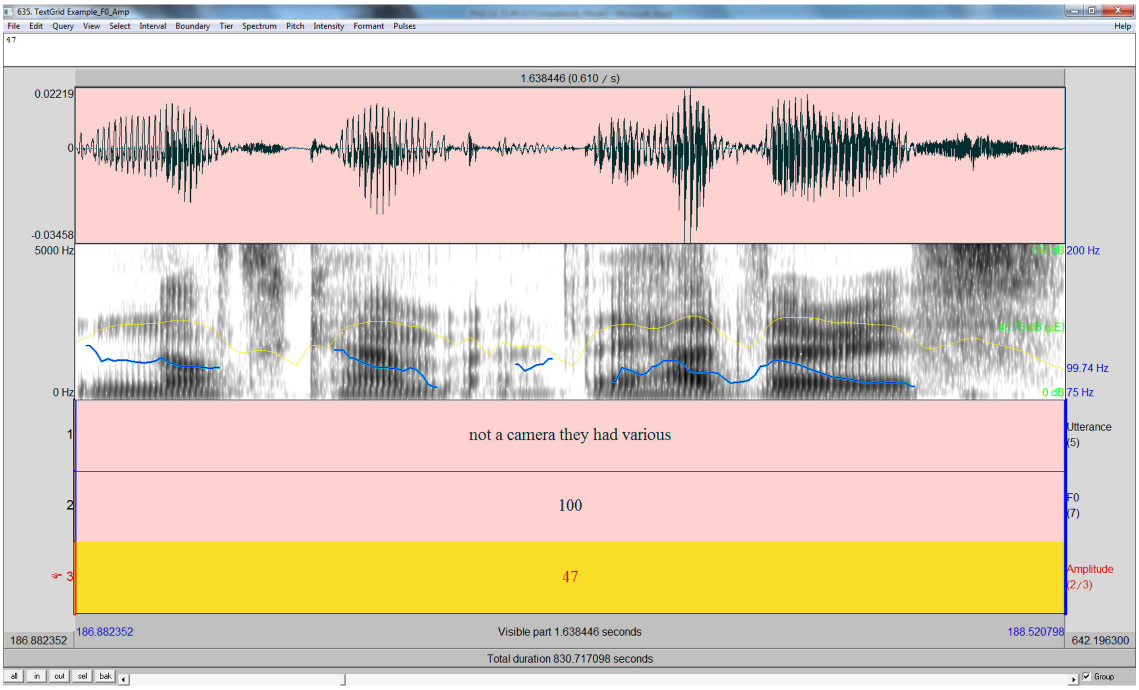

Figure 4 – Example of a textgrid in *Praat* used to measure f_0 and amplitude

Before commencing formant analysis the files were edited so as to only contain speech from the participants. As indicated in Table 5, samples were at least 90 seconds long. In some cases, the interview contained twice as much speech as the baseline which was likely to be a result of the methodological adjustments. Both speaking conditions provided sufficient quantity of material for reliable formant measurements for all ten speakers.

Table 5 – Length of speech extracted for all ten speakers in the baseline and interview

Participant	Baseline	Interview
1	119	492
2	91	346
3	161	375
4	312	624
5	205	193
6	334	264
7	350	181
8	214	320
9	444	313
10	211	262
Average	**244**	**337**

Vowel formant measurements were extracted from Linear Predictive Coding (LPC) spectra using *Praat's* inbuilt formant measuring tool. The default number of formants selected by the *Praat* algorithm was adjusted for each speaker in order to achieve greatest possible accuracy in the measurements. It has to be pointed out that the method offered by the *Praat* software is not infallible (Harrison, 2004; 2011). Therefore, its accuracy was always checked visually against the spectrogram and any errors were corrected by hand. 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 (Figure 5). To be counted in the analysis the vowels had to show relatively steady formants and be in stressed positions. For all speakers, eight vowel categories /iː, ɪ, ɛ, a, ɜː, ʌ, ɒ, ɔː/ (FLEECE, KIT, DRESS, TRAP, NURSE, STRUT, LOT and NORTH[8]) were measured using one to twenty tokens per category for each condition. In the case of /ɜː/ and /ɔː/ frequency of tokens tended to be lower for some speakers and in the case of /ɜː/ two speakers did not show any tokens in one of the two conditions. The vowel categories were still included in the general analysis but the low frequency aspect was kept in mind when interpreting the results. In total, around 2600 measurements were taken, 1000 from the baseline and 1600 from the interview. Timing, mean values for each of the three formants and the source word for each vowel were recorded using *Microsoft Excel*.

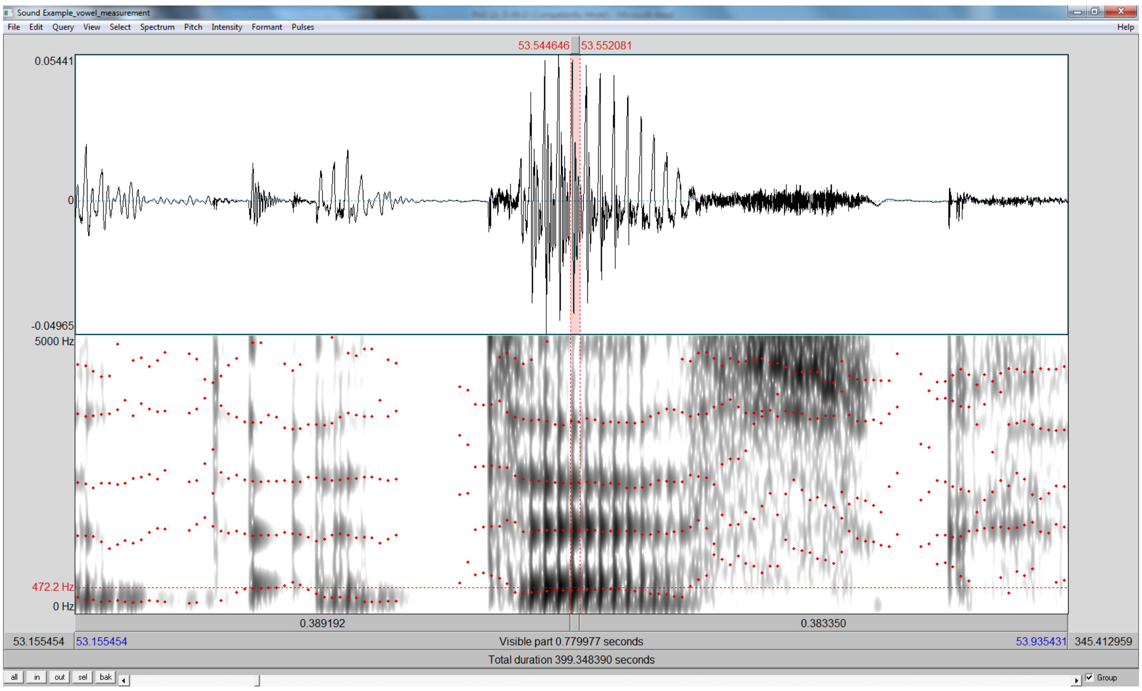

Figure 5 – Example of vowel formant measurement in *Praat*

[8] Standard Lexical sets for English developed by John C. Wells (1982). There are 24 lexical sets which represent words of the English language based on their pronunciation in Received Pronunciation (RP).

The mean formant values were also used to plot vowel space area estimations (VSAe) for each of the ten speakers. Vowel space area was estimated by calculating the area of vowel polygons using a modification of the formula written by Mark Adams (Fabricius et al., 2009). The points of the polygons were the eight monophthongs /iː, ɪ, ɛ, a, ɜː, ʌ, ɒ, ɔː/. Rhodes (2012) points out that one of the limitations of this measure is the fact it does not take into account speakers' natural variation in vowel targets. As mentioned in Chapter 5, formants, and in particular, vowel space area has only received very little (if any) interest in relation to deceptive speech. The method of polygon estimation was therefore judged to be sufficient for the purposes of the current work.

An automatic vowel extractor based on a combination of voiced frames and an amplitude threshold was applied to the edited recordings. On average, the baseline provided 59 seconds of vocalic material (SD = 32 ms) and an average of 82 seconds of vocalic stream (SD = 43 ms) was extracted from the interview. As can be seen the files exceed the 6 second minimum specified by Moos (2008) and therefore it may be assumed that the vocalic material collected will have had a minimal confounding effect on long-term formant results. The vocalic stream was subjected to J P French Associates implementation of the iCAbS (iterative cepstral analysis by synthesis) formant tracker developed by Frantz Clermont (Clermont et al., 2007). F_1 – F_4 measurements were obtained every 5 milliseconds which facilitated the calculations of the long-term formant distributions.

A script developed by Bert Remijsen was applied to the vocalic material previously extracted (Remijsen, 2004). The script identified H_1 and H_2 from a 10 ms window from the centre of a vocalic segment and computed the values for H_1-H_2. On average, 917 (SD = 485) measurements were taken from the baseline and 1086 (SD = 384) measures from the interview.

Speaking rate (SR) and articulation rate (AR) measurements can be operationalised in different ways. Only a brief dialogue will be provided in the present work; a more thorough discussion can be found in Jessen (2007), Künzel (1997) and Trouvain (2004). One has to consider the unit of measurement, i.e. whether to 'count' words, syllables or sounds. If the decision has been made to use the syllable the next question revolves around whether to use its phonological/canonical form or whether to express it in its actual realisation, i.e. its phonetic form. The phonetic syllable was chosen as it is the most optimal for measuring tempo (Jessen, 2012).

As discussed above, one also needs to choose between taking global or local measurements, or both. In the case of SR local measurements were calculated using the speaking turn and, as was the case for the f_0 measurements, the inter-pause stretch formed the speech interval for AR calculations. It may be argued that the inter-pause stretch does not present the most optimal unit as it is influenced by speakers' pausing behaviour which, paradoxically, the AR measure attempts to eliminate (Jessen, 2007). Furthermore, Dankovičova (1997) had shown that it is the intonation phrase that provides for the best representation of the duration of phonological words compared to the inter-pause stretch and the clause. The inter-pause stretch was chosen as the interval of measurement in the current work, as it still presents the most common domain in the literature. It is also substantially easier to identify than the intonation phrase and therefore offers practical advantages.

AR was calculated using the IP stretches previously marked for the analysis of f_0. The marked intervals facilitated timing measures and syllables were determined purely auditorily (Figure 6). To be counted in the analysis a speech interval had to include at least four syllables in order to control for any confounding effects caused by phrase-final lengthening or shortening often found in short stretches of speech (Jacewicz et al., 2010). Filled pauses such as 'um' or 'uh', if occurring within a clause as opposed to turn initially were considered as instances of ordinary speech articulation; speech dysfluencies such as speech errors or slips of the tongue were generally not counted in the analysis.

AR measures could have been performed on the edited recordings used to perform the formant analysis. However, in order to accurately measure SR and pausing behaviour it was important that analysis was performed on the original recordings which included the full dialogue between participants and interlocutors. For example, silent periods can be instances of pausing but they can also be related to the closure period during plosive consonants or the silence in between turn-taking. In order to accurately classify silent periods they need to be analysed in the context in which they occur. As with the IP phrases, speaking turns were marked up on text grids and syllables were identified using auditory means (Figure 6). Once again, participants often showed more speaking turns in the interview as a result of the experimental methodology. In addition, speaking turns were not equal across speakers. As already mentioned, the value in collecting spontaneous speech outweighs this anticipated drawback. On average, the baseline provided 25 (SD = 12) speaking turns and the interview 36 (SD = 8).The raw measurements including the actual text of the speech interval, its duration, the number of syllables and the SR/AR in syllables per second were recorded using *Microsoft Excel.*

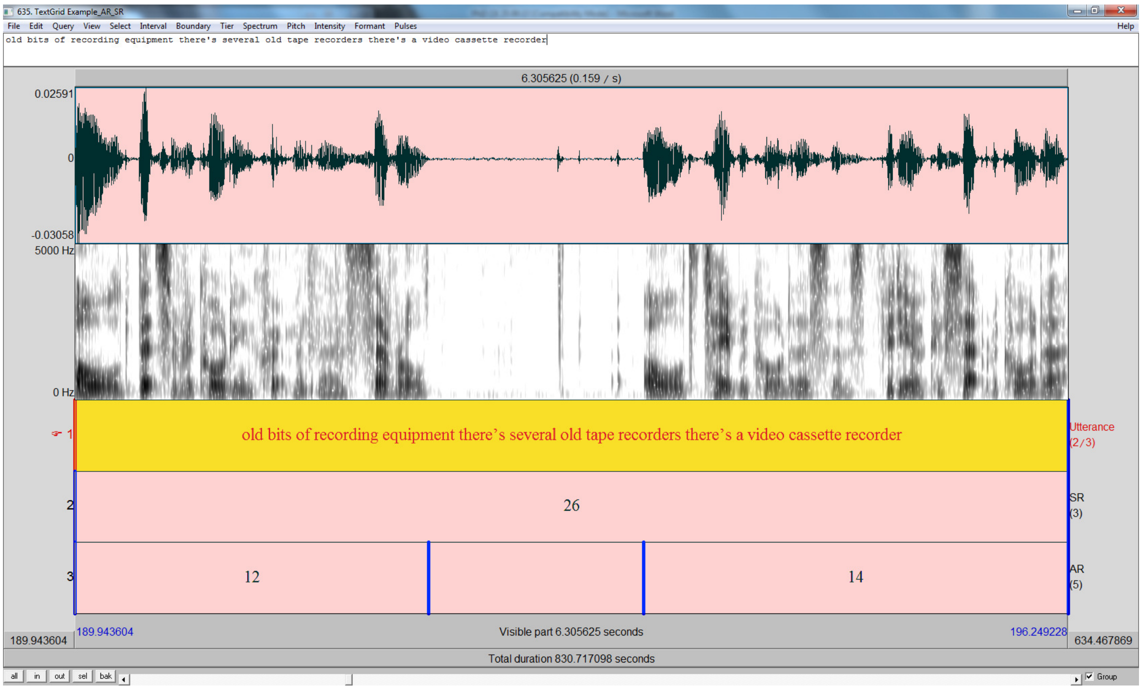

Figure 6 – Example of a text grid in *Praat* used to measure AR and SR

As defined in Chapter 5, AR refers to intervals of speech excluding pauses. In order to exclude pauses it is necessary to define a pause including its threshold value or, in other words, the minimum duration necessary for a silent interval to be classed as a pause. In general, there is no agreement as to what this minimum durational limit should be and studies have used different benchmarks ranging from 60 ms (Kendall, 2009) and 100 ms (Künzel, 1997) over 200 ms (Duckworth and McDougall, 2012) and 250 ms (Goldman-Eisler, 1968) to 270 ms (Kowal and O'Connell, 1980). Künzel (1974) has shown that silent periods of less than 60 ms are not consistently perceived as instances of pauses and in light of this, it was decided to employ a silent threshold of 100 ms in the present study. Reports of high thresholds are rare in the literature and therefore no upper cut-off was employed. The closure phases of plosive consonants can complicate exact pause measurement, in particular when occurring at the beginning of an inter-pause stretch. In these instances the closure phases were included in the pause measurement. Admittedly, this method is far from accurate. It was nevertheless decided to opt for this strategy as it seemed to be the most practical method in the given circumstances.

Goldman-Eisler (1968) differentiates between three types of pauses. The first category relates to periods of silence that are the results of articulatory processes e.g. closure periods of plosives. The second category includes pauses that result from respiratory activity and the third group contains pauses that are not related to articulation or physiological activity but which are due to cognitive, pragmatic and discourse processes. She refers to the latter as "hesitation pauses" (Goldman-Eisler 1968, p. 12). It is relatively straightforward to identify and exclude the first type of pauses by setting a minimum duration threshold as has been done in the current work. Differentiating between pauses that serve respiratory function and hesitation pauses is more difficult however. The recordings were of high enough quality that would have allowed for the identification of audible in-breaths and out-breaths; however, respiration pauses may not have audible respiratory activity or conversely hesitation pauses may contain breathing activity. For example, it was not unusual that a silent pause without any audible noise was followed by breathing which was then followed by another completely silent stretch. Categorising this stretch into three different pauses, i.e. hesitation pause – respiratory pause – hesitation pause, would have seemed counter-intuitive and therefore the decision was made to conflate breath and hesitation pauses and count them as one silent pause (Figure 7).

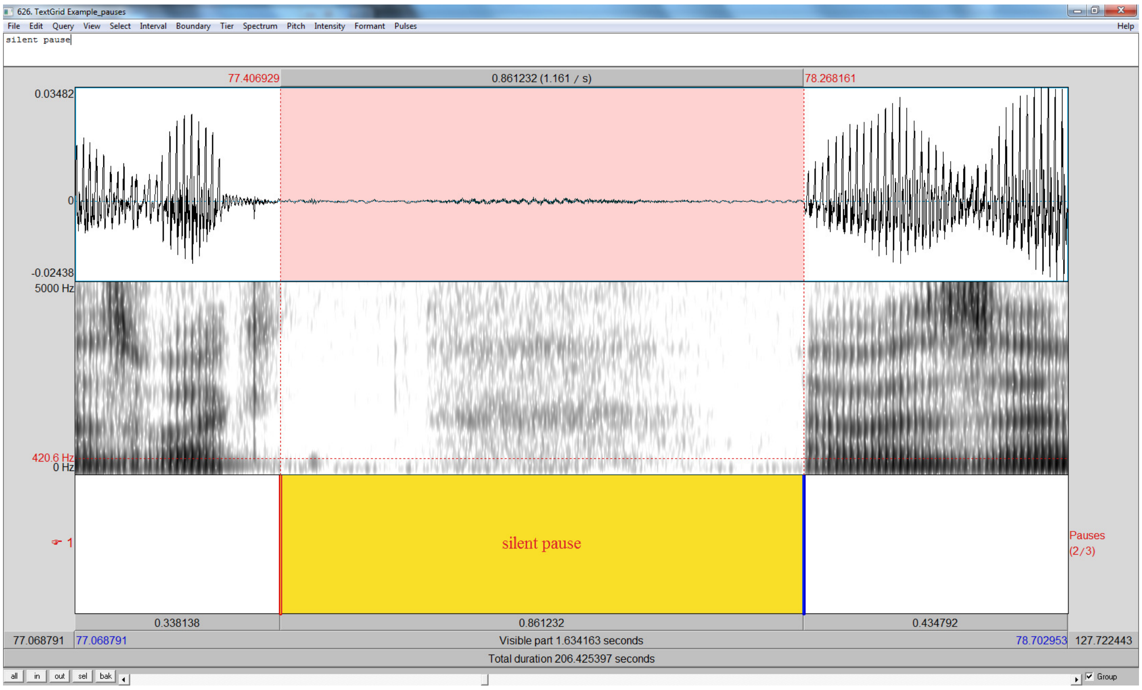

Figure 7 – Example of a text grid in *Praat* used for pause categorisation

Filled pauses are commonly regarded as hesitation or discourse markers and, in English, they typically appear in the form of 'um' or 'uh'. Hesitations have strong speaker-specific power and as Künzel (1997) memorably stated: "individuals tend to be quite consistent in using 'their' respective personal variant of the hesitation sound, in particular with respect to the optional addition of a bilabial nasal consonant" (p. 51). Moreover, research has shown that 'um' and 'uh' are distributed differently across speakers and, therefore, it could be illuminating to take a look at the different types of hesitations separately (Duckworth and McDougall, 2012). Both, the 'vocalic only' and the 'vocalic +nasal' variants were analysed in the present work.

As with SR and AR, silent pauses and hesitations were marked up in *Praat* using text-grids which facilitated easy retrieval of duration and frequency calculations (Figure 8). Table 5 showed the variances in length of speech extracted for each speaker and each of the two conditions. In order to control for the difference in length of speech, the frequency aspects of pauses and hesitations was conveyed in the form of a pause rate (PR) and hesitation rate (HR) measurement. This corresponds to the number of pauses or hesitations per minute and was calculated using the following formula:

PR/HR = *Frequency of pauses/hesitations x 60/length of speech*

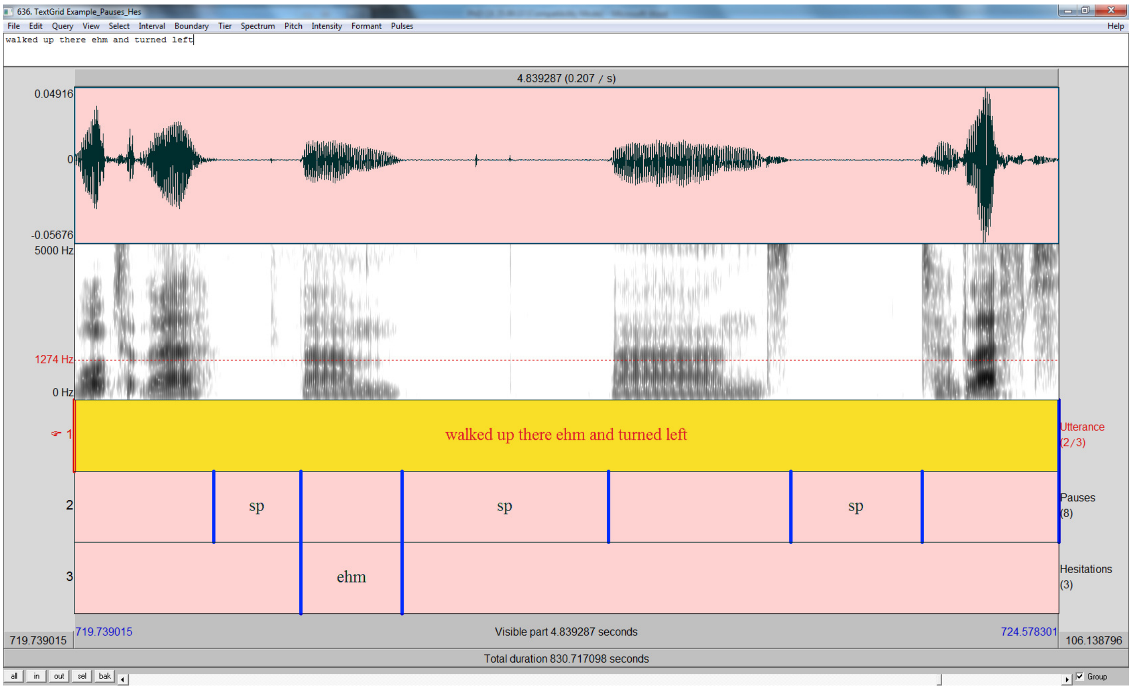

Figure 8 – Example of a text grid in *Praat* used to measure pauses and hesitations

Response latency, also referred to as response onset time in the speech literature, signifies the time in between the end of a speaker's turn and the beginning of that of a second speaker. In the present case, it was measured as the silent period between the end of the interviewer's turn and the start of the participant's speech. As outlined in Chapter 5, this parameter has been especially popular amongst researchers looking into deceptive behaviour. Once more, text grids generated by *Praat* facilitated the ROT measurements (Figure 9).

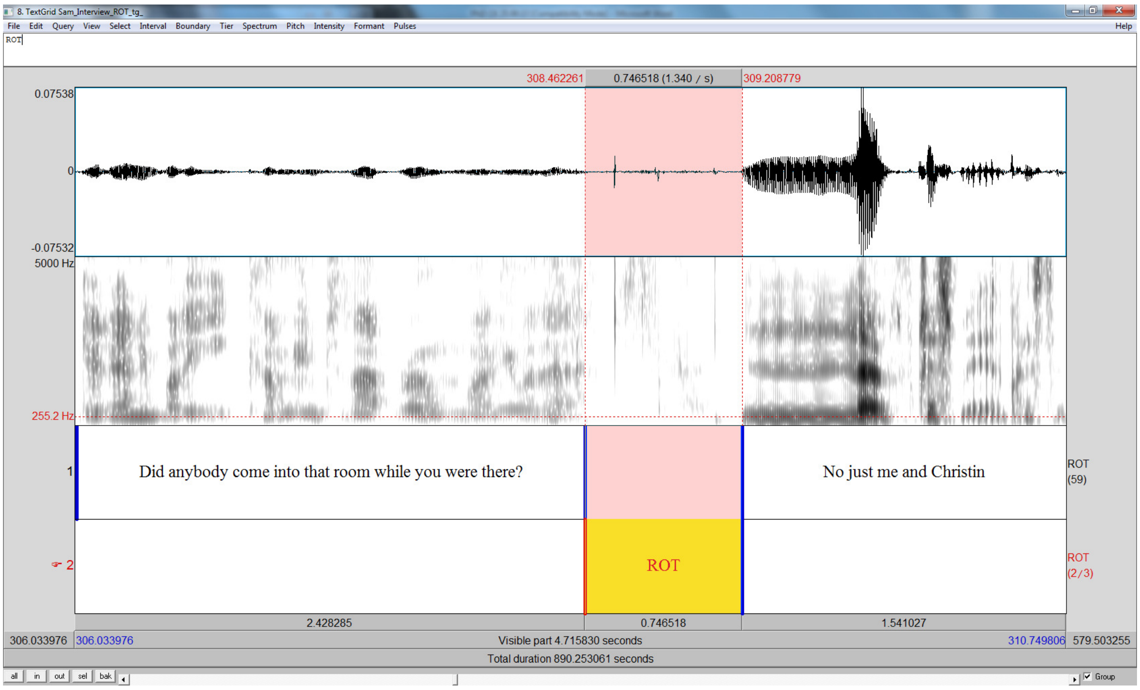

Figure 9 – Example of a text grid in *Praat* used to measure ROT

A score was derived from the STAI questionnaire by summing the ratings for each of the six scale items. Positive items, e.g. I feel calm were negatively scored. The scores were expressed as Percent of the Maximum Possible score (POMP) scores (Cohen et al., 1999). This method of scoring reflects a participant's score in terms of the maximum achievable score on the scale. It was calculated as follows:

*[(observed score – minimum score)/(maximum score – minimum score)] * 100*

The observed score relates to participants' scores for each of the six items scales. For example, one participant may have given a score of '3' to the statement 'I am relaxed' (item 4 on the STAI scale). The minimum and maximum scores are, respectively, the minimum and maximum possible scores on the scale. So, for a 4 point Likert scale, as was used in the STAI, the minimum possible score would be 1 and the maximum possible score would be 4. The STAI scores were correlated with the differences across conditions for the measured speech parameters. During the analysis of the Big Five scale several limitations became obvious. Firstly, it became clear that the scale was too broad for the purposes of the present work. Each of the five traits contains a number of sub traits. For example, neuroticism is associated with being tense, moody and anxious. A more fine-grained classification of tense vs. not tense would have been more appropriate in relation to deception. Secondly, the scale failed to measure aspects of personality which are related to motivations, emotions, attitudes, abilities, self-concepts and social roles. These would have presented a better reflection of the participant's personalities and, perhaps, provided more appropriate aspects of personality measures for deception research. For these reasons, only the findings of the STAI results are presented as part of the current work.

6.2 Experiment 2

As introduced in Chapter 1 and the beginning of this chapter, the data for experiment 2 was taken from the 'smell of fear' corpus. The corpus formed part of a broader research study which looked at various human responses when being deceptive including biological, physiological, psychological and behavioural characteristics. The following will only provide a partial description of the experimental methodology focusing on those aspects that were relevant to the speech analysis. A full description of the aims and results of the project can be found in Eachus et al. (2013).

Again the methodological design took the form of a laboratory experiment in which participants had to deceive in return for a financial reward; rather than being subjected to an investigative interview, liars and truth-tellers were presented with an accusatory style of questioning. Chapter 2 described how deception can be achieved by various means including outright falsifications and simple concealment, for example. While experiment 1 elicited narrative falsifications, the methodological design of experiment 2 generated concealment type messages. In addition, a tighter control of some of the contextual factors that were not kept constant in experiment 1, i.e. interviewer, type of questions asked, was in place.

6.2.1 Participants

The data consist of an opportunity sample of 37 male native English speakers with a mean age of 24 years. Participants were drawn from the staff and student population at the University of Nottingham and none had any self-reported voice, speech or hearing disorders.

6.2.2 Experiment procedure

Unlike experiment 1, which oriented itself on investigative interviewing practices, experiment 2 used a scalable interrogation paradigm analogous to the more accusatory interrogation techniques described in Chapter 3. After collection of baseline data, nineteen of the participants were given a yellow token containing a picture of the Eiffel Tower and the word avocado which they were asked to conceal when interviewed. These nineteen speakers will be referred to as 'liars' for the remainder of this work. The remaining eighteen speakers did not receive a token and therefore were not required to conceal any knowledge. They are the 'truth-tellers'.

Truth-tellers and liars were interviewed in three different speaking conditions- baseline, interrogation 1 and interrogation 2. Designed to elicit control data, the baseline interview of experiment 2 was comparable to the baseline condition in experiment 1 in that it contained neutral and relaxation based questions. Examples include:

- *'Have you ever travelled abroad and if so where did you last go?'*
- *'What is your favourite type of movie?'*

The interrogations aimed to enhance participants' state of arousal by asking more probing and penetrating questions. Interrogation 1 provoked a low level of emotional involvement by posing questions about social desirability and concealing information in general. Example questions are:

- *'Would you consider yourself to be a generally honest person?'*
- *'Would you generally find it easier to keep a secret from someone else close to you such as a family member or a partner?'*

Interrogation 2 was more provocative by directly challenging participants' truthfulness. The accusatory nature of the questions can be seen in the following examples:

- *'Are you giving us truthful answers to our questions?'*
- *'From the way you are answering these questions we think you are withholding some information. I will ask you again. Are you concealing any information from us?'*

The questions employed in experiment 2 were a mixture of yes/no and open ended format. Some of the yes/no questions mirror the type of questions common to control question tests in polygraph examinations (see Chapter 3). The three conditions contained twenty questions and a full list of these is included in Appendix C.

Prior to the each of the three conditions participants completed the STAI questionnaire which featured in experiment 1. In addition, they rated their experiences of the two interrogations using a subjective experiences questionnaire. On a 4-point Likert scale, participants rated how difficult they found it to lie/tell the truth and how confident they were that the investigators believed them. Liars were told that they would receive a financial reward if they manage to convince the investigators that they are not in the possession of the token in question. They had to produce deceptive messages. Truth-tellers, on the other hand, could honestly answer the questions. Originally, the truth-telling condition was included in the experimental design of the study by Eachus et al. (2013) in order to have a control group. As was explained between subject comparisons are not desirable for the type of speech analysis performed in the present context. No attempts were made to compare the liars and truth-tellers against each other. Rather, the data of the truth-tellers were treated separately and interpreted in an attempt to address research question 4.

In contrast to experiment 1 which used human interlocutors, the questions of the three speaking conditions in experiment 2 were pre-recorded and then presented via loudspeaker from a standard laptop computer. This ensured that the presentation of questions was kept constant across subjects and conditions thereby eliminating possible differences due to interactional context. In order to avoid participants anticipating the end of the interrogations, the questions were not numbered in a serial fashion. The order of speaking conditions was kept the same for all participants starting with the baseline interview which was followed by interrogation 1 and, finally, interrogation 2. Overall, the experiment took 75 minutes and on completion of interrogation 2, participants were debriefed and received their £30 participation reward.

6.2.3 Recording setting and equipment

The experiment was conducted at the University of Nottingham. The recordings were made using a standard tape recorder placed in the middle of a table. Despite attempts, it was not possible to ascertain the specifics of the recording equipment, i.e. type of recorder and recording level. However, on inquiry, it was ensured that all aspects of the setting and equipment were kept constant between conditions and speakers. The stereo recordings were received in mp3 format with a sampling rate of 48k Hz and a bit depth of 16 bit. *Sony Sound Forge* was used to convert the files into standard (.wav) audio files before commencing analysis.

Unfortunately, there were limitations with the quality of the recordings. Firstly, participant's speech appeared to be at a distance to the microphone resulting in very low levels of speech. Secondly, there was considerable background noise (tape hiss, computer scanning) present throughout the recording as indicated in Figure 10. Thirdly, mp3 is a lossy compression algorithm and it has been shown to affect the acoustic properties of voice signals including formant frequency values (Gonzales et al., 2003; Meinerz and Masthoff, 2010; 2011). As mentioned above, the 'smell of fear' corpus formed only part of a bigger scale research study looking into multiple aspects of deceptive behaviour including bio-chemical reactions. The speech recordings were of secondary importance and, therefore, attention to the data collection procedure was not as careful and thorough as it could have been. It would have been elegant to perform temporal as well as acoustic analysis on the data of experiment 2 providing the opportunity to compare the results from both experiments. Efforts were made to include some form of acoustic analysis but after repeated consideration of the qualitative shortcomings it was decided to accept an incomplete data analysis over a flawed data analysis.

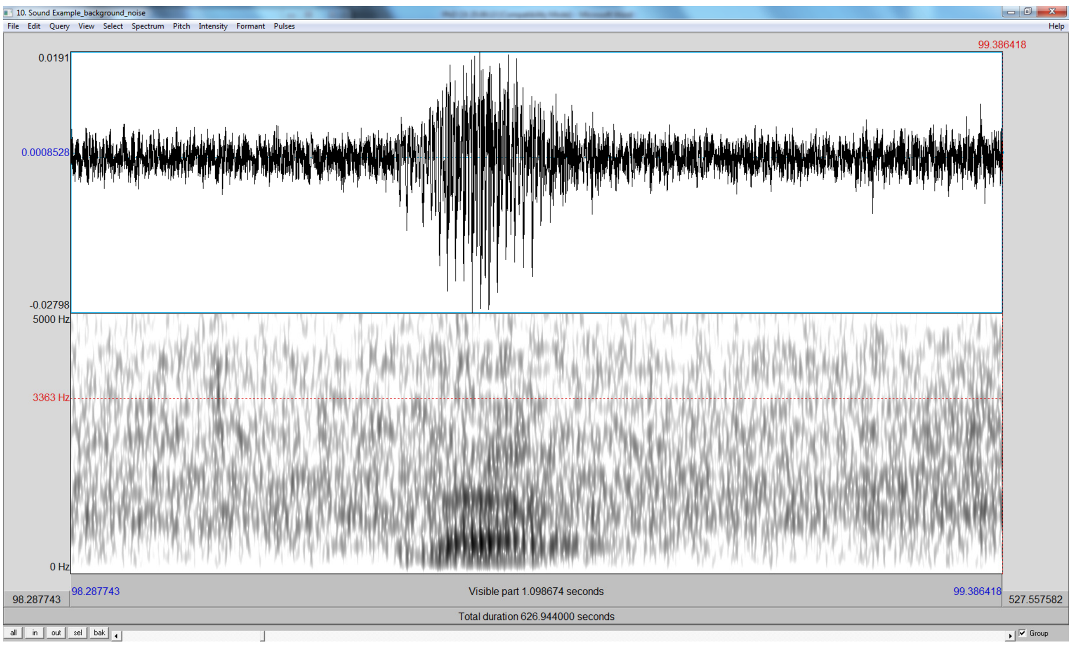

Figure 10 – Illustration of the presence of background noise in the recordings of experiment 2

6.2.4 Measurement equipment and parameters analysed

As was the case with the data from experiment 1, *Sound Forge* was used for initial editing of the speech files and the temporal analysis was performed using *Praat*. The motivation for merely performing temporal analysis and not attempting to investigate the frequency based parameters was largely based on data inadequacy as explained above. Given the nature of the data, i.e. yes/no and short answer open questions, only a selected number of temporal parameters could be investigated. Included in the analysis were:

- Articulation rate (AR)
- Speaking rate (SR)
- Frequency of silent pauses (PR) and hesitations (HR)
- Duration of silent pauses and hesitations
- Response onset time (ROT)

Every speaker provided one file for each of the three speaking conditions, resulting in 114 files for analysis. The duration of the original, unedited recordings were comparable across the three conditions and groups. On average, the files were in between 10 and 13 minutes long.

6.2.5 Measurement technique

The two levels of interviewing differed in that the second interrogation was more coercive. The questions were more direct in targeting participants' knowledge and possession of the object thereby eliciting explicit lies from the liars and undisputed truths from the truth-telling group. The majority of questions in interrogation 1 were more implicit focusing on trustworthiness in general rather than the possession of the object in question. It could be argued that the majority of responses to the interrogation 1 questions were not overt lies as such. Of course, the discussion in the preceding chapters highlighted that lying does not just involve straightforward falsifications; far from it. It is intuitive to think that liars would want to use as many truths as possible in their deceptive account in order to keep the potential for getting caught to a minimum. Indeed, the deceptive speech data from experiment 1 exemplified the fusion of local truths and lies in the global deceptive narrative. Analysis has been completed for all three conditions; the results are complex however, and have raised a number of issues which were beyond the scope of this project. For example, it addresses the effect of different levels of interrogation as well as the impact of context conditioning on behaviour. Therefore, it was decided to narrow the focus to a two-way comparison between baseline and interrogation 2 as this was thought to be of most value for the overall aim of the thesis. The analysis of all three conditions can be found in Kirchhübel et al. (2013).

The technique for measuring the various temporal parameters mirrors that used in experiment 1. Text grids generated by *Praat* facilitated easy access to durational calculations which were then transferred into *Microsoft Excel* for further examination. The nature of the data being short answer responses, participants often produced little speech and at times only 'yes' or 'no' responses. As pointed out in section 6.1.6 above, there are limitations of measuring speech tempo using short or even monosyllabic utterances. Phrase final lengthening or shortening may unjustly influence the measurements. Given the significance of rate measures in experiment 1, it was decided to perform AR and SR calculations on the data of experiment 2 under the provision that no monosyllabic responses would be included. In addition, special attention was given to exclude lengthened or shortened phrases or high frequency phrases such as 'I don't know'. Judgements of what was believed to be excessive lengthening or shortening were based on auditory assessment. The AR and SR results are interpreted bearing in mind these caveats.

SR was measured based on the number of phonetic syllables in participant's speaking turns. On average, the baseline provided 18 turns (SD = 2) for the liars and 19 (SD = 2) turns for the truth-tellers. The interrogation produced an average of 12 turns (SD = 4) for the liars and 13 (SD = 4) turns for the truth-tellers. AR was calculated in phonetic syllables per second using the IP phrase as a domain. The liars provided, on average, 27 IPs (SD = 7) for the baseline and 14 IPs (SD = 7) for the interrogation. The truth-telling group produced, on average, 27 IPs (SD = 8) and 16 IPs (SD = 8) for the baseline and interrogation, respectively. ROT was measured as the time in between the end of a question and the beginning of the participant's response. Similarly to experiment 1, hesitations were analysed taking into account the 'vocalic' 'uh' as well as 'vocalic + nasal' 'um' variants. In addition, a further subdivision was made between hesitations which occurred turn-initially and those that occurring during a speaker's turn.

Table 6 – Length of speech extracted for all 37 speakers in the baseline and interview

	Liars		Truth-tellers	
Participant	**Baseline**	**Interrogation**	**Baseline**	**Interrogation**
1	49	15	70	27
2	37	42	48	11
3	26	14	45	20
4	28	6	158	41
5	62	25	56	24
6	44	18	27	14
7	82	49	109	61
8	108	42	33	22
9	30	16	47	18
10	26	12	140	96
11	90	15	186	44
12	92	121	48	53
13	57	44	44	18
14	43	32	20	12
15	42	19	85	71
16	54	20	35	35
17	74	45	50	25
18	61	21	163	109
19	55	18	-	-
Average	**56**	**30**	**76**	**39**

As in experiment 1, in order to control for differences in length of speaking time across conditions and speakers, frequency of hesitations was expressed in the form of a hesitation rate (HR) measurement calculated as number of hesitations per minute. As demonstrated in Table 6, there are differences in the length of speech extracted for liars and truth-tellers and across conditions. Some of these differences in lengths may be attributed to variation in speakers' verbosity; some were more talkative and prepared to go beyond monosyllabic responses than others. Although some of the speakers provided a small amount of speech, overall, the quantity was judged to be adequate for the temporal analysis.

6.3 Statistical analyses

Descriptive statistics including graphical representation of the data were generated in *Microsoft Excel*. Statistical tests were employed where possible to assess the effect of deception on the various speech parameters analysed. *SPSS* (v.20) was used to perform the inter-speaker and intra-speaker comparisons. The baseline and interview/interrogation were treated as independent variables and the various speech parameters as dependent variables. Inter-speaker comparisons were performed using the Wilcoxon signed rank test. Apart from the amplitude measurements, inter-speaker comparisons were performed for all analysed parameters. Intra-speaker comparisons were performed using independent t-tests. No intra-speaker comparisons were performed for f_0 SD, amplitude, VSAe, PR and HR as only one global value was available for these parameters per speaker per condition. Unless specified, correlations were assessed using the non-parametric Spearman's rho. The rationale behind choosing non-parametric over parametric tests was based on the fact that the between speaker comparisons contained a small sample size, i.e. ten speakers only.

This chapter presented the reader with a thorough introduction to the nature of experiments 1 and 2 including the methodologies employed in the data collection and analyses techniques. Chapter 7 will shed light on the results of the acoustic and temporal analysis of experiment

7 Experiment 1

A detailed description of the experimental data employed in the current study was presented in Chapter 6. Chapter 7 presents the findings of the acoustic and temporal analysis of experiment 1 thereby addressing research question 1: Can deception be identified based on acoustic and temporal characteristics of speech? The results are presented in visual and summary format. The absolute values and detailed statistical findings can be found in Appendix D.

7.1 Acoustic analysis

7.1.1 f_0

As described in Chapter 2, deception may result in increased levels of stress, cognitive load or behavioural control. All three emotional and cognitive states have resulted in an increase in f_0. The few studies that exits which investigated mean f_0 in relation to deceptive speech reported an increase or no change.

The f_0 mean and f_0 SD values obtained for each speaker and for each condition are described in numerical and visual formats below.

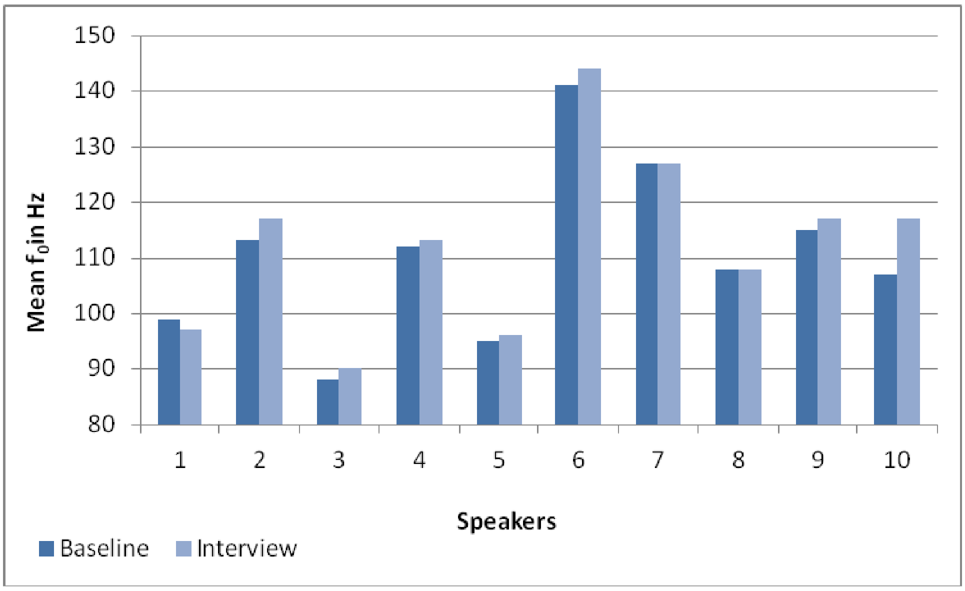

Figure 11 – Mean f_0 in baseline and interview for all speakers

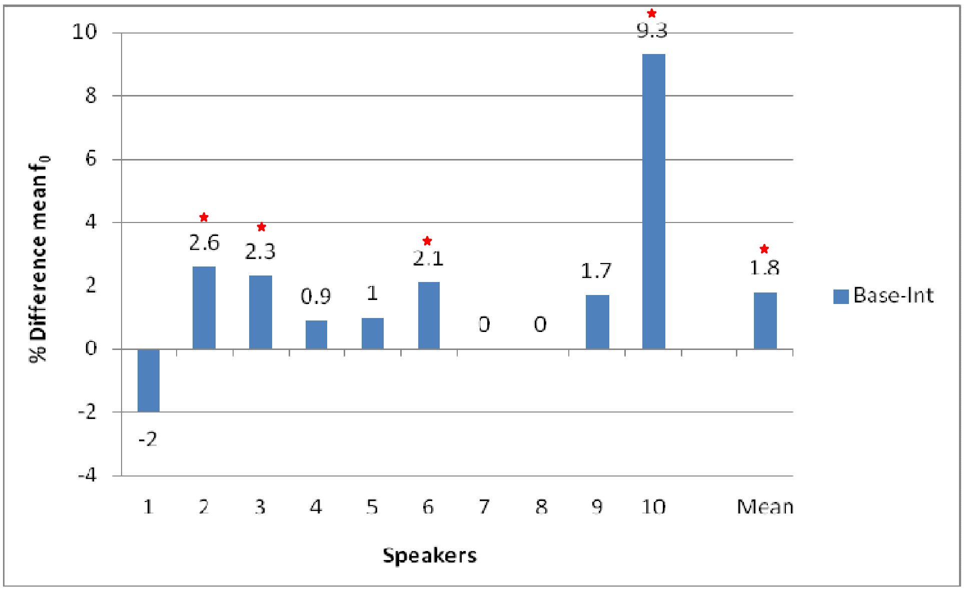

Figure 12 – % difference in mean f_0 between baseline and interview for all speakers

To obtain the percentage difference between mean f_0 in the baseline and mean f_0 in interview (Figure 12) the following calculation was used:

$$\% \text{ difference} = (f_0 \text{ Interview}) \times (100 / f_0 \text{ baseline}) - 100$$

A '% difference' of 0 would mean that there was no difference in mean f_0 across conditions suggesting that deception had no effect on mean f_0. Positive values would indicate that the average f_0 values measured in the interview are higher compared to those of the baseline condition and negative values indicate the opposite, i.e. a decrease in f_0 in Interview.

Judging by the close approximation of the bars in Figure 11,the mean f_0 results did not display an extraordinary change between conditions. Still, a Wilcoxon signed rank test showed a significant difference between f_0 in baseline (mean = 110 Hz) and in interview (mean = 113 Hz) ($T = 4$, $p = .05$, $r = -.44$). The majority of speakers, if changing, show a slight increase in mean f_0 in the interview. The increase ranged from 0.9% (1 Hz) to 9.3% (10 Hz). Only speaker one demonstrates a reduction which amounted to 2% (2 Hz). The increase in mean f_0 reaches statistical significance for four of the speakers and this is marked by the red asterisks. Speakers 2, 3, 6 and 10 show increases in mean f_0 of 3 Hz, 2 Hz, 3 Hz and 10 Hz, respectively. Speaker 10 stands out from the remaining speakers in that his increase in mean f_0 is more substantial. Indeed, when removing this speaker from the analysis, the difference between mean f_0 in baseline and interview is not significant ($T = 4$, $p > .05$, $r = -.40$); however, a medium effect size still remains. These values correspond to those of earlier laboratory research on speech under stress, cognitive load, and clear speech which tended to observe average increases of 1 - 15 Hz. They also correspond to those studies which have observed a slight increase in mean f_0 with deception.

Two questions may be asked. Why did speaker 1 show diverging trends to the rest of the sample, and why did speaker 10 show such a sizeable increase in mean f_0? Factors connected to the linguistic content of the baseline and interview speech were considered but nothing in the baseline or interview conditions differed for the two speakers as compared to the remaining eight. Assessment of the STAI did not show extremely diverging anxiety scores for these speakers either as illustrated in Figure 13. Was the act of lying responsible for these results? Informal listening suggested that speaker 1 was more interactive in the baseline as compared to the interview during which he appeared to be more monotone. Perhaps, this speaker was attempting to control his behaviour during the interview in order to come across as truthful. Speaker 10, on the other hand, was more talkative and conversational during the security interview based on informal auditory judgements. It could be that this speaker was trying to make a good impression on the interviewer in order to conceal his deceit. These two speakers highlight the presence of individual variability in deceptive behaviour. As will become apparent throughout this thesis the concept of speaker idiosyncratic behaviour is important within the domain of deception as it is in so many other spheres.

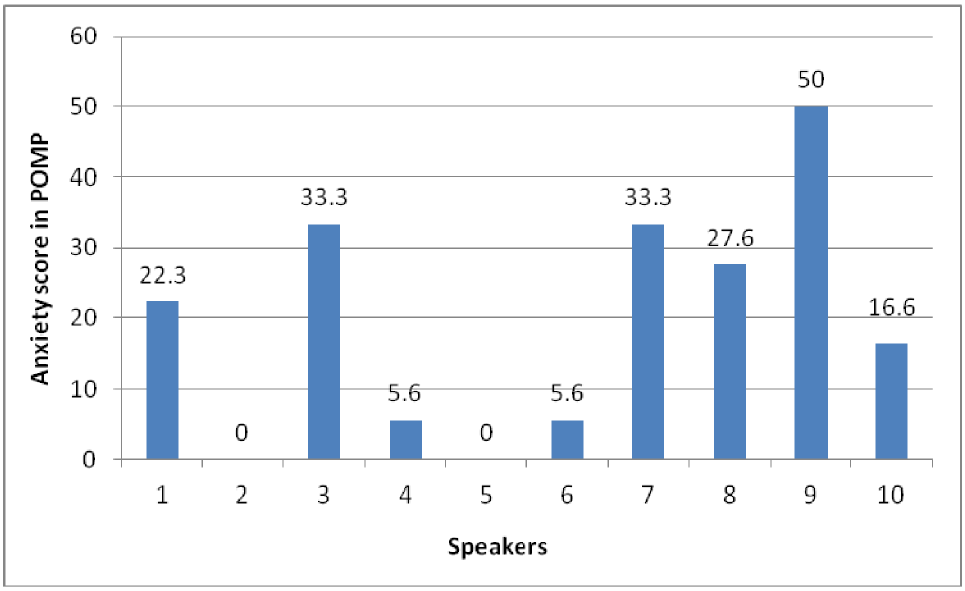

Figure 13 – Average stat-trait anxiety scores for all speakers

Mean is the most common and most widely reported measure of central tendency in relation to f_0. Other measures that could have been employed to describe the f_0 distribution are median, mode, or skew and kurtosis, for example, which may have painted the results in a different light (Nolan, 1983; Papp, 2008). However, in order to offer comparison with previous research the mean f_0 values were deemed to be most suitable for the current work. From a perspective of pitch perception, it could be argued that the f_0 measurements would have been more appropriately expressed in logarithmic units (e.g. semitones) rather than Hz. Hz was nevertheless chosen in the present case as it allowed for comparability with previous studies. For the interested reader, the difference in mean f_0 between baseline and interview was 0.46 semitones.

7.1.2 f_0 variability

The overview in Chapter 4 illustrated that f_0 variability tended to increase with stress and clear speech but reduced in speech under cognitive load. Research into f_0 variability in connection with deceptive speech could not be located.

f_0 variability as measured in the current work was expressed by the standard deviation of the mean f_0 values.

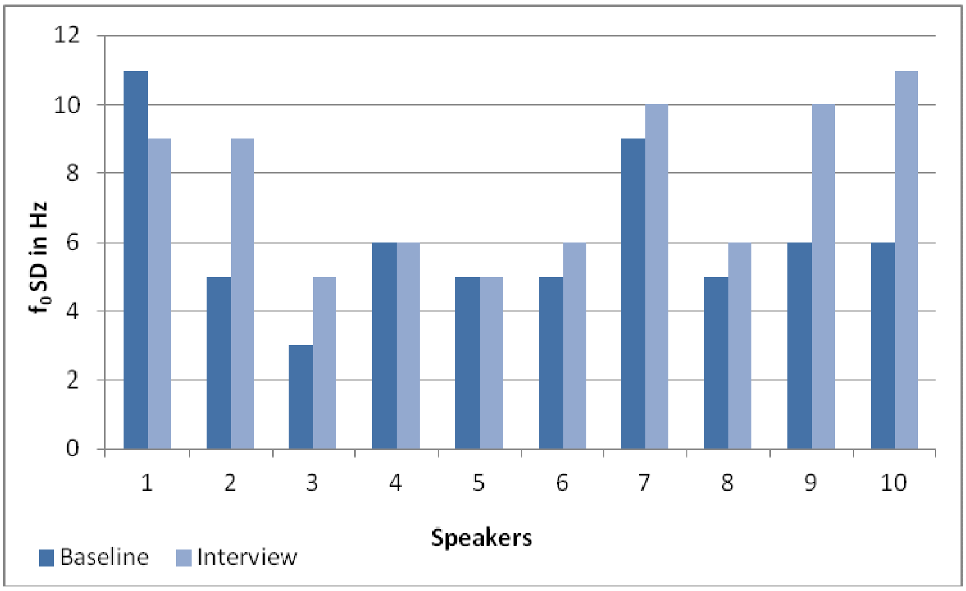

Figure 14 – f_0 SD in baseline and interview for all speakers

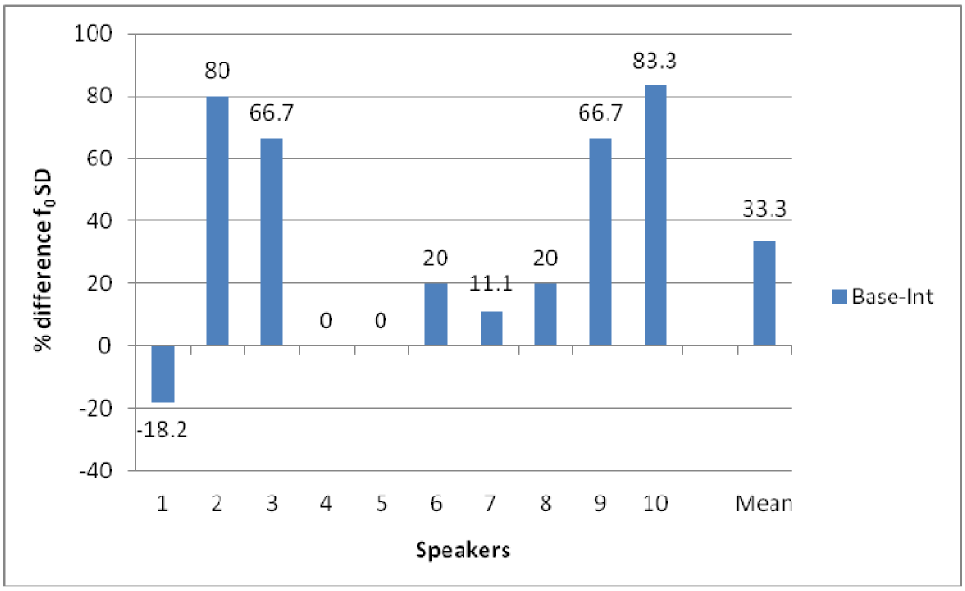

Figure 15 – % difference in f_0 SD between baseline and interview for all speakers

Seven of the speakers show an increase in f_0 SD (Figure 14). The magnitude of increase ranged from 11.1% (1 Hz) to 83.3% (5 Hz) as displayed in Figure 15. Two speakers do not show any change, however these are not the same two speakers as seen in the mean f_0 results. Speaker 1 had a slight reduction in variability of f_0 amounting to 2 Hz and closer inspection reveals that this is the same speaker who had the decrease in mean f_0. Not surprisingly, the speaker who presented the largest increase in f_0 SD (5 Hz) was the one who had the largest increase in mean f_0, i.e. speaker 10. Between speaker comparisons showed the increase in f_0 SD to be not significantly different in the interview (mean = 8 Hz) as compared to the baseline (mean = 6 Hz) ($T = 4.5$, $p > .05$, $r = -.42$). However, removing speaker 1, who showed a reduction in f_0 SD, results in a significant increase in f_0 variability in the interview as compared to baseline ($T = 0.00$, $p \leq .05$, $r = -.56$). Unfortunately, as explained in section 6.1.6, statistical testing could not be extended to examine intra-speaker differences in relation to f_0 SD.

As illustrated in Figure 16 there was a significant positive relationship between the change in mean f_0 and the change in f_0 SD between baseline and interview ($r_s = .79$, $p \leq .01$). The blue dots represent the change in f_0 mean and f_0 SD for all ten speakers (four of the speakers have corresponding changes and therefore their values are overlapping each other).

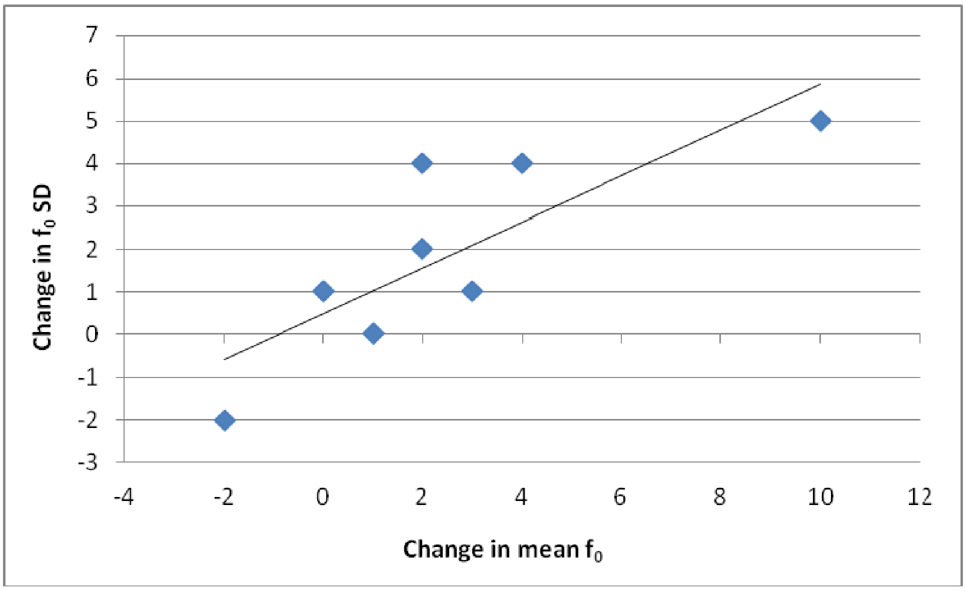

Figure 16 – Correlation between change in mean f_0 and change in f_0 SD

As explained in section 4.1, f_0 is linked to vocal fold physiology and anatomy whereas f_0 variability is more reflective of speaker-specific behaviour in terms of monotonous as opposed to melodic speech. Therefore, the two parameters should be kept separate (Jessen et al., 2005). By expressing the f_0 SD in absolute values, as has been done, it is not independent of the mean values; however, conveying the f_0 SD in terms of 'coefficient of variation' (CV) allows for a separation of the two. Even when expressing f_0 SD in relative terms, the correlation between increase in mean f_0 and f_0 SD still remains positive and significant (r_s = .71, $p \leq .05$) as illustrated in Figure 17.

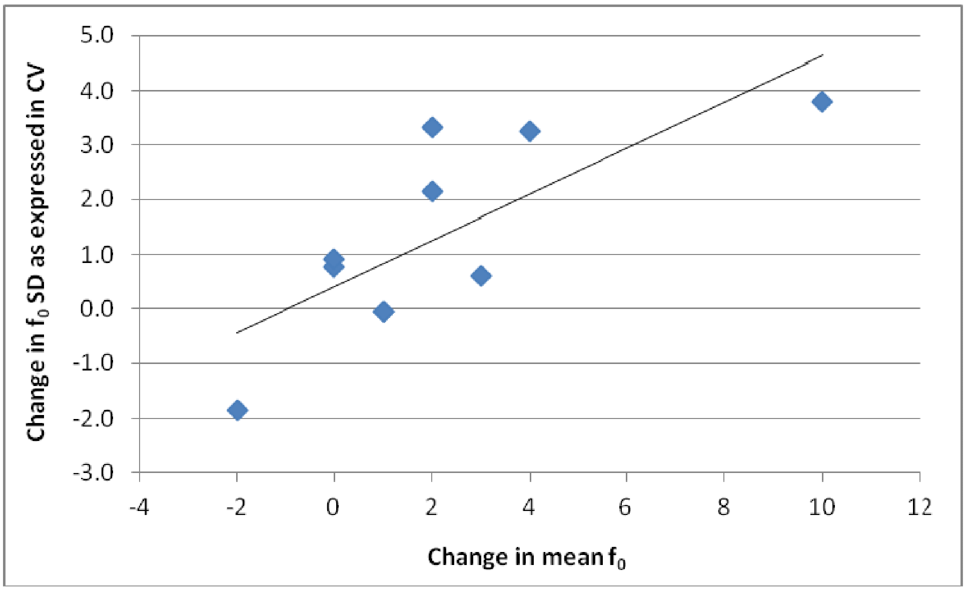

Figure 17 – Correlation between mean f_0 and f_0 SD (SD expressed in coefficient of variation)

It was noted that the f_0 SD values tended to be low overall. Low f_0 SD could have been a sign that speakers were not engaging in the interaction in both the baseline and interview; that they were employing a passive and subdued style of speaking. This was not the case however. In general, speakers actively connected with the experimental tasks. A possible explanation may be related to the method of measuring. The f_0 SD values represent the standard deviation of the mean f_0 values across all IP stretches. An averaging process will have already taken place when calculating the mean f_0 from each IP stretch and this could be reflected in the lower f_0 SD values. Had f_0 been measured based on a smaller unit, i.e. the syllable or vocalic element, the values may have been slightly higher.

7.1.3 Amplitude

In the research literature, speaking amplitude was found to increase with stress, cognitive load and when people are speaking more clearly. The amplitude findings in relation to deceptive speech are unclear but with some studies reporting a wider range in amplitude.

As explained in section 6.1.6, only relative amplitude measures will be reported in the current work. No consistent trend emerged. There was variability in direction of change in amplitude with half the speakers increasing and the other half decreasing (Figure 18). Magnitude of change was also variable across speakers. For some speakers (speakers 1, 2, 3, 4 and 5) the change was relatively small whereas for others it was somewhat larger (speakers 6, 7, 8, 9 and 10). Speaker 6 and speaker 9 showed the largest changes (around 7.5 %) but in opposite directions each. While speaker 6 increased by 4 dB, speaker 9 decreased by 4 dB. Overall, the extent of change ranged from -2 dB to +4 dB. The values are similar to Huttunen et al. (2011) who reported an increase of 1 dB with increased cognitive load. They are less comparable to those reported by the literature into clear speech who reported increases of 5-8 dB. However, for factors discussed in Chapter 6, caution has to be employed when comparing amplitude measures across studies.

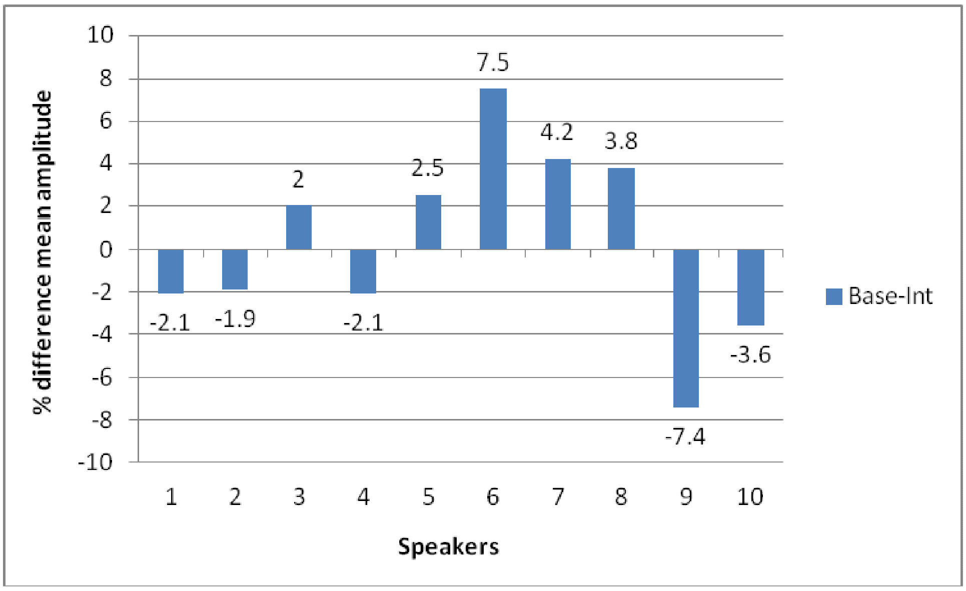

Figure 18 – % difference in mean amplitude between baseline and interview for all speakers

There is a large body of research that has demonstrated that increases in amplitude can lead to increases in f_0. Jessen et al. (2005) provide a detailed overview of the literature. There was no correlation between the change in f_0 mean and amplitude ($r_s = -.30$, $p > .05$) in the current data. The blue dots in Figure 19 represent the changes for each of the ten speakers (two of the speakers have corresponding changes and therefore their values are overlapping each other). Speakers who did not have a remarkable change in amplitude between baseline and interview nevertheless showed f_0 mean differences. Likewise, change in mean f_0 did not always result in overall amplitude changes. For example, speaker 10 who displayed a mean f_0 increase of 10 Hz actually decreased in overall amplitude by 2 dB.

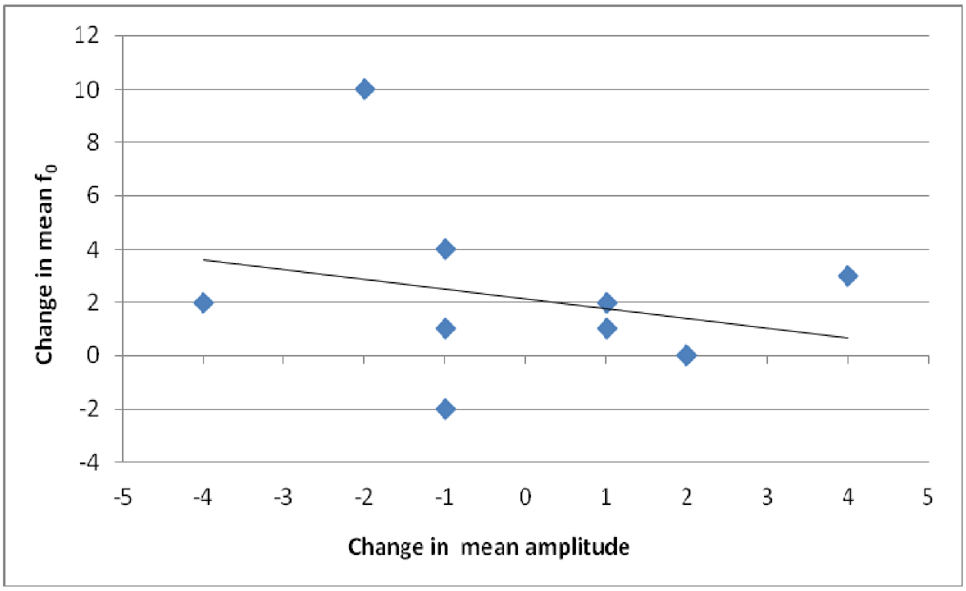

Figure 19 – Correlation between changes in mean amplitude and changes in mean f_0

As reported in Chapter 4, recent studies on speech and stress have started to report on the amplitude of different frequency bands rather than the amplitude of the whole of the frequency spectrum. Analysing the current data using this approach may reveal more consistent trends.

7.1.4 Vowel formant frequencies F_1 and F_2

Previous research on vowel formants in relation to speech under stress reported an overall decrease in formant frequencies and a reduction in overall vowel space area. Clear speech on the other hand resulted in an expansion of speakers' vowel spaces. However, this expansion was not uniform across vowels and different studies reported different vowels to be most affected. The effect of cognitive load on formant frequencies and vowel space area is unclear. The investigation of formant frequencies appears to present a gap in the literature on deceptive speech.

For F_1-F_2, a 'global' mean across the eight measured vowel categories was calculated for each individual. In addition, a mean for F_1-F_2 was also computed for each vowel category.

7.1.4.1 F_1 by speaker

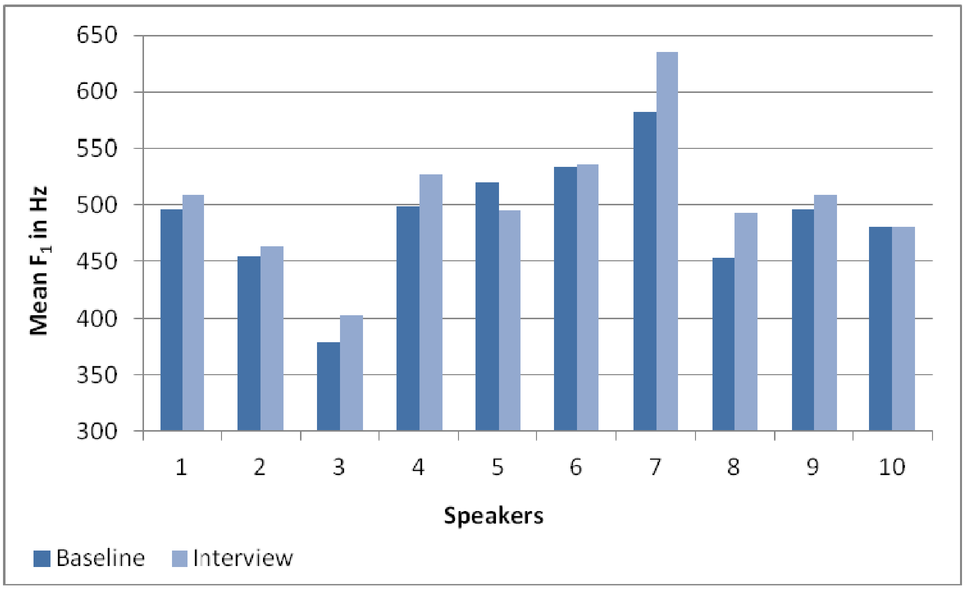

Figure 20 – Mean F_1 in baseline and interview for all speakers

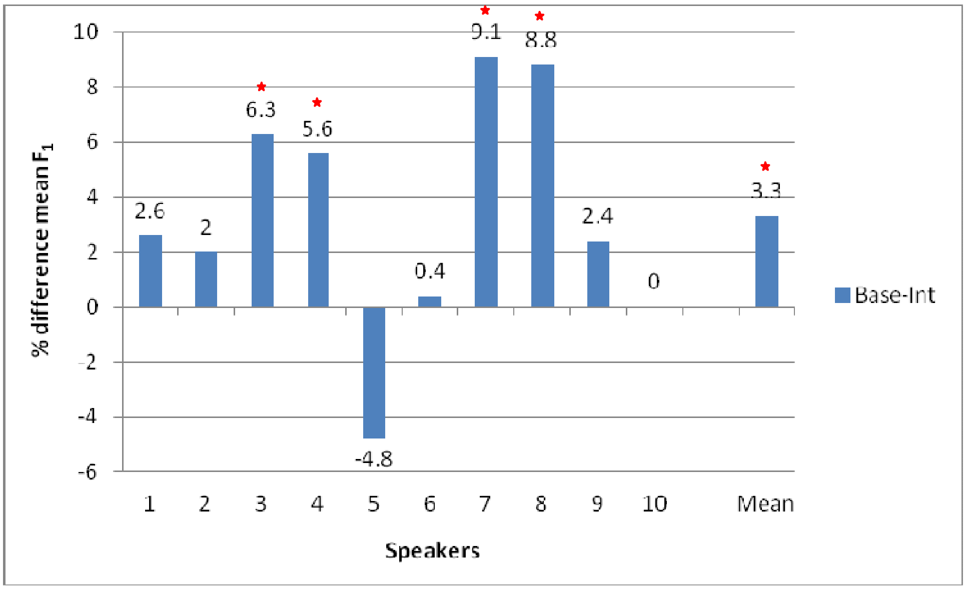

Figure 21 – % difference in mean F_1 between baseline and interview for all speakers

Figures 20 and 21 demonstrate that, if changing, the majority of speakers exhibited an increase in F_1 in the interview. The increase differed in magnitude and ranged from 0.4% (2 Hz) to 9.1% (53 Hz). Speaker 5 was the only participant who showed a reduction in mean F_1 in the interview which averaged 4.8% (25 Hz). Independent t-tests showed this increase in average F_1 to be significant for four of the speakers (speakers 3, 4, 7 and 8). Overall, the increase in F_1 from Baseline (mean = 489 Hz) to interview (mean = 505 Hz) was significant ($T = 6$, $p \leq .05$, $r = -.44$).

Could the increase in mean F_1 in the interview be a result of a peripheralisation of vowels? The standard deviation (SD) values of F_1 may provide an answer to this. If vowels are more peripheral, their SD values are expected to be larger. Comparing the SD values of F_1 in baseline (mean = 118 Hz) and interview (mean = 124 Hz) showed that there was no significant change across conditions ($T = 10.5$, $p \geq .05$, $r = -.39$). Although seven speakers showed larger SD values in the interview as compared to the baseline, this only amounted to an average increase of 11 Hz. The SD values do not give strong support to the notion that speakers were producing more precisely articulated vowels when they were deceiving.

Could the increase in mean f_0 be another explanation for the increase in F_1? Researchers have argued that changes in formant values, particularly F_1, may be due to measurement errors on the part of the acoustic analysis method used. Formant estimation by LPC modelling is said to be affected by a bias towards the harmonic that is closest to the resonant peak. If the f_0 increases, the spacing between the harmonics will become wider and they could potentially be farther away from the resonant peaks. As a result, the estimation of the resonant peak by the LPC based method will be less reliable (Fitch, 1989). There was no significant relationship between change in mean f_0 and change in F_1 (r_s = -.62, p ≥ .05).

Research on 'Lombard speech' has also shown that an increase in amplitude results in an upward shift in F_1 (Kirchhübel, 2009). In view of this it was thought interesting to test whether amplitude changes may offer an explanation for the formant behaviour of F_1. The correlation between change in overall amplitude and change in F_1 was weak and not statistically significant (r_s = .27, p ≥ .05).

7.1.4.2 F_1 by vowel

The reader may wonder why vowel categories are compared as it appears, at first sight, difficult to predict differences between vowel types as a function of deception. However, Ruiz et al. (1996) discovered that the effect of stress on speech in relation to formant frequencies was vowel dependent. Furthermore, research on clear speech showed that the F_1 of low vowels tended to be most affected.

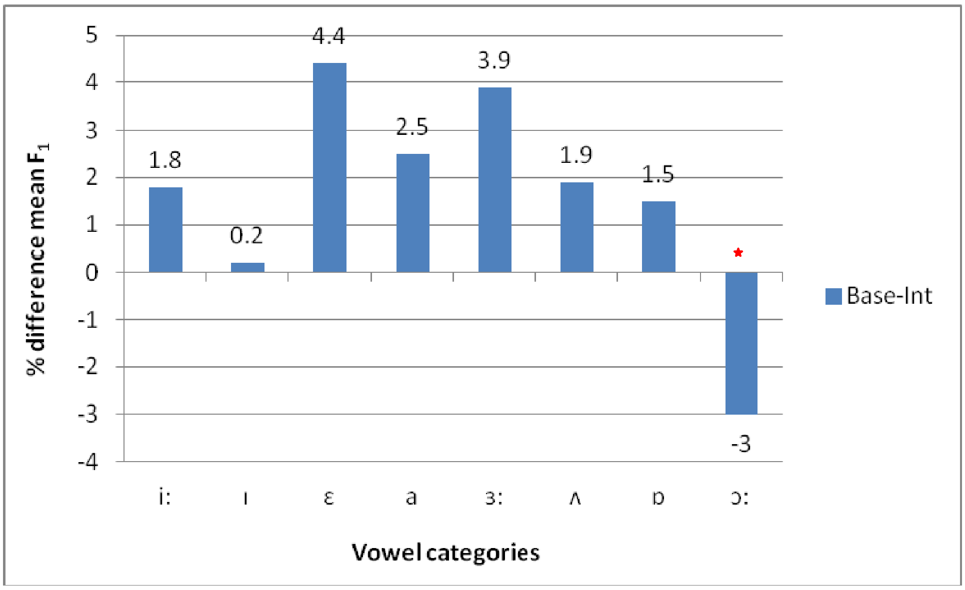

Figure 22 – % difference in mean F_1 between baseline and interview for all vowel categories

Figure 22 shows that the change in F_1 that did occur ran in the same direction across vowel categories. With the exception of /ɔ:/, F_1 was slightly higher in the interview speech. This increase ranged from 0.2% (1 Hz) - 4.4% (25 Hz) and was largest for /ɛ/ and /ɜ:/. The /ɔ:/ vowel was the only category that showed a lower mean F_1 value in the interview and this reduction, which averaged 3% (13 Hz) turned out to be statistically significant ($T = 3$, $p \leq .05$, $r = -.52$). Taking a closer look at /ɛ/ and /ɜ:/ it could be observed that the F_1 %-differences were bigger compared to /ɔ:/, yet the shifts were not statistically significant. Probing the raw measurements it became evident that individual speakers showed noticeable within-speaker variation in F_1 frequencies. The reason for this may be connected to the lexical type of /ɛ/ and /ɜ:/ tokens collected or co-articulatory processes may have played a role. No significant correlation existed between F_1 in the baseline condition and the change in F_1 in the interview ($r = .08$, $p \geq .05$). However, one may still wonder why /ɛ/, /ɜ:/ and /ɔ:/ appeared to be affected most. They do not form a cohesive unit. In fact, /ɛ/ is a front vowel, /ɔ:/ a back vowel and /ɜ:/ is often realised as a central variant. Regrettably, no reasonable explanation could be pinpointed. One argument which is admittedly far from satisfactory could be the mere difference in quantity of lexical tokens for each of the vowel categories in both conditions. Although attempts were made to achieve comparable numbers of word tokens for every vowel category the nature of the data did not always allow for this which, in turn, could have affected overall formant distributions.

7.1.4.3 F_2 by speaker

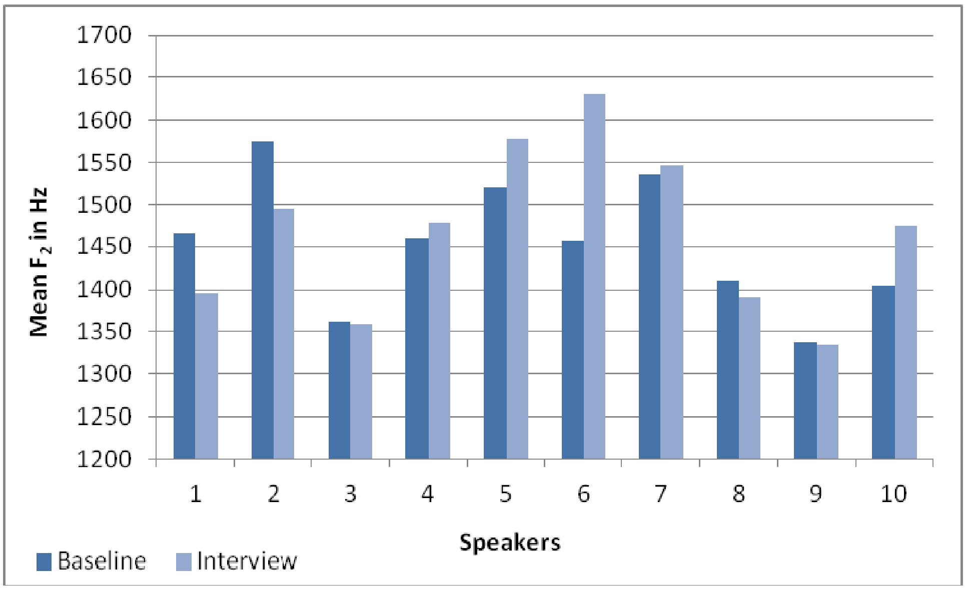

Figure 23 – Mean F_2 in baseline and interview for all speakers

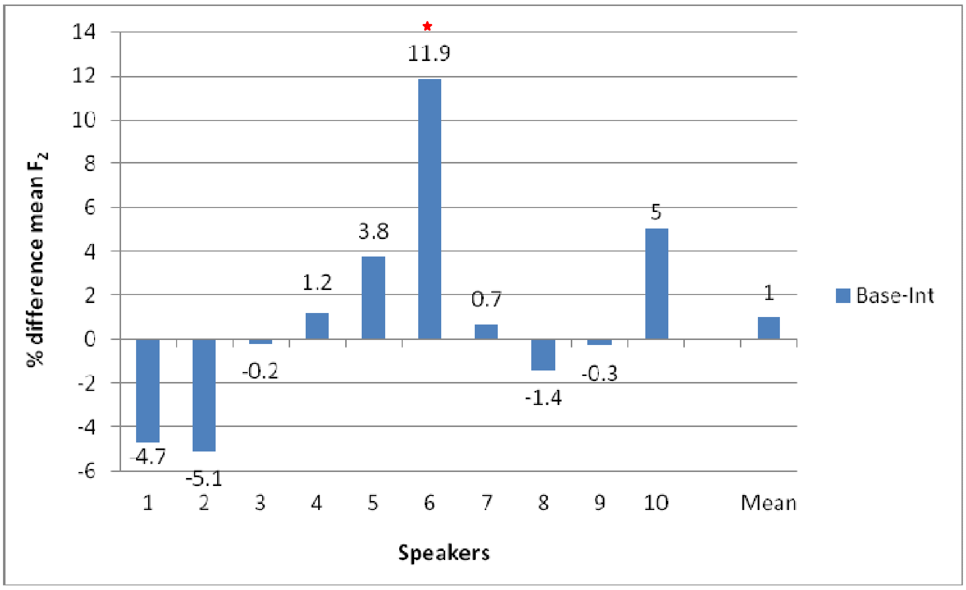

Figure 24 – % difference in mean F_2 between baseline and interview for all speakers

The pattern for the F_2 results is less consistent than that observed for F_1. Figures 23 and 24 show that some speakers increase (speakers 5, 6 and 10), some decrease (speakers 1 and 2) and others do not change extensively (speakers 3 and 9). Furthermore, compared to F_1, the change in F_2 was less extreme. Overall, the change in F_2 from baseline (mean = 1453 Hz) to interview (mean = 1468 Hz) was not significant ($T = 24$, $p \geq .05$, $r = -.08$). Speaker 6 was the only participant who demonstrated a significant increase which summed to 11.9% (173 Hz) in total. Although the mean F_2 values did not change considerably across conditions, a difference in the SD values may still be present which could indicate vowel peripheralisation. Results show no significant difference in SD between baseline (mean = 368 Hz) and interview (mean = 360 Hz) ($T = 15.5$, $p \geq .05$, $r = -.27$). In fact, the majority of speakers showed lower F_2 SD values (average of 16 Hz) in the interview.

7.1.4.4 F_2 by vowel

Research on the effects of clear speech on F_2 has resulted in diverging findings in relation to which vowels were most involved in the change. Some reported that front vowels were most heavily affected whereas others stated that clear speech effects were present for back vowels only.

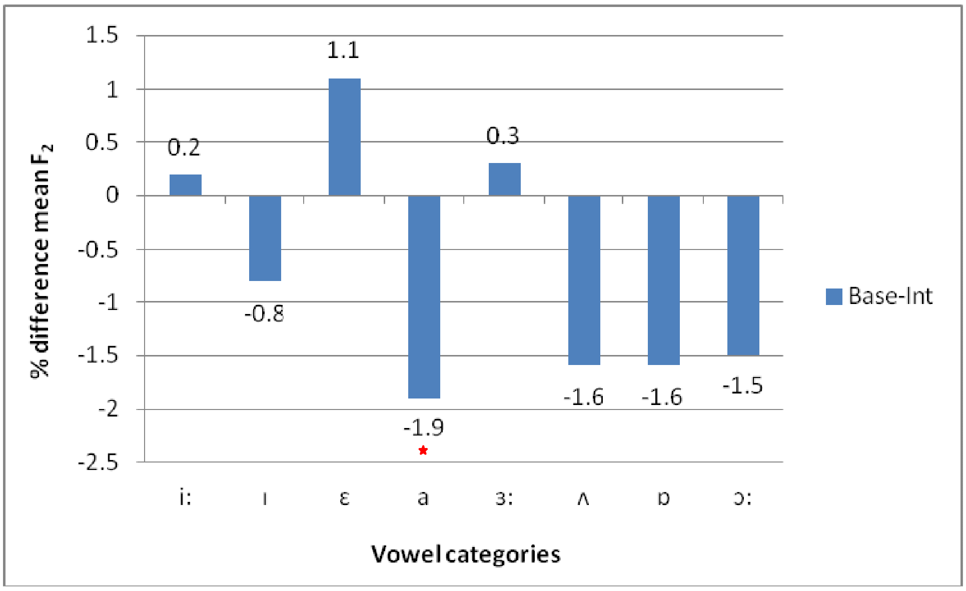

Figure 25 – % difference in mean F_2 between baseline and interview for all vowel categories

Figure 25 reveals that there were observable differences between vowel categories in relation to F_2 change. However, none of the vowel categories change by more than 2%; this is less compared to that observed with F_1. Pearson's correlation revealed no significant relationship between F_2 in the baseline condition and the change in F_2 in the interview ($r = -.06$, $p \geq .05$). Despite this, there appears to be a uniform effect on the low and back vowels. The F_2 of /a/, /ʌ/, /ɒ/ and /ɔː/ decreased from baseline to interview by an average of 1.6 % but this decrease was only significant for /a/ ($T = 0$, $p \leq .01$, $r = -.63$). Scrutinizing the nature of the lexical token words collected for this vowel category for each condition did not provide for a satisfactory explanation. In both instances, content words comprised the majority of tokens.

The conclusions that can be generated from the mean formant frequencies are weak. There appears to be a slight increase in F_1; the change in F_2 is more variable however. Moreover, different vowels seem to be affected to different degrees. The SD values did not lend support to the notion that vowels were more peripheral when speakers were lying. However, a more fine-grained examination may be fruitful. A better way of evaluating whether peripheralisation of vowels is occurring may be seen in vowel space area estimations.

7.1.5 Vowel space area estimation

While an increase in size of vowel space area could be an indication of peripheralisation of vowels, i.e. hyper-articulation, a decrease may point to target undershoot. Vowel space areas were calculated for all speakers in both baseline and interview in order to see whether the extent and size of speakers' working vowel space (in the F_1 x F_2 plane) was affected by the act of lying.

A visual representation of the results is provided in Figures 26 - 35 which show the vowel space areas in the F_2 x F_1 plane for all ten speakers in both baseline and interview conditions[9]. Also included in the figures (bottom right) are histograms which illustrate the mean VSAe values for both speaking conditions[10]. Even though it is not a straightforward task to judge the effect of lying on vowel space area from the scatter plots alone, the visual mode does bring to light that the size of vowel spaces differed across conditions for some speakers.

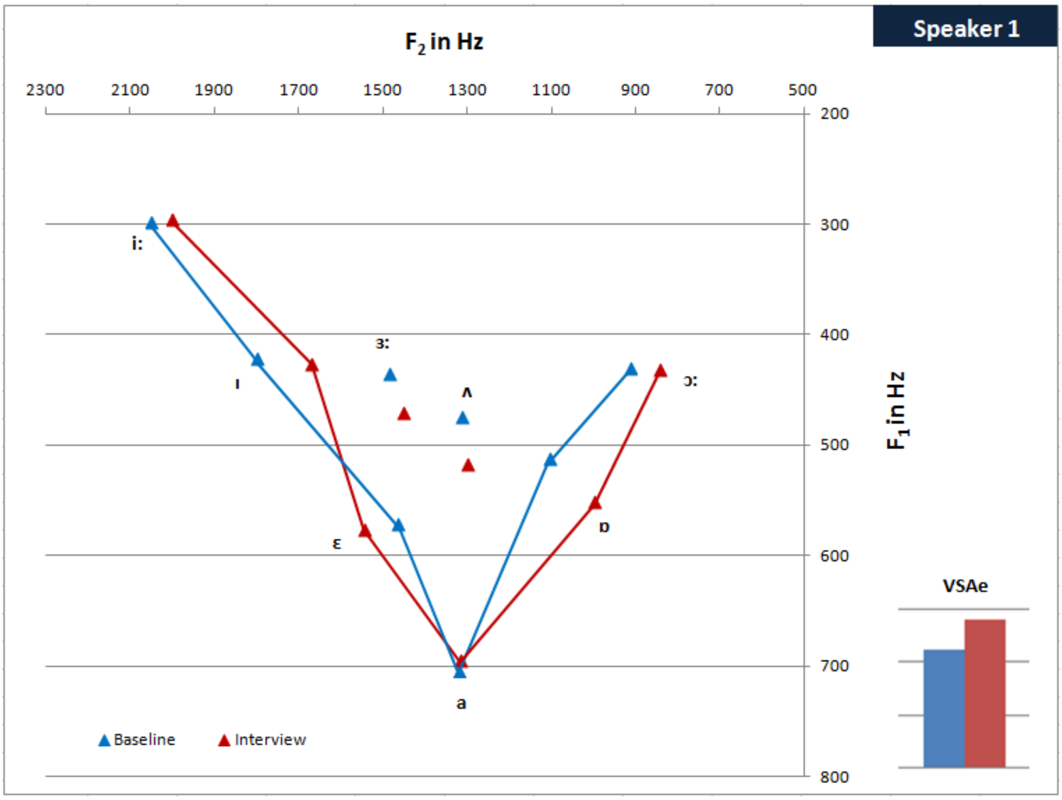

Figure 26 – Scatterplot of mean F_1 and F_2 values for all vowel categories in baseline and interview for speaker 1, connected to represent vowel space area. Histogram of vowel space area estimation for baseline and interview

[9] For the visual representation only the vowels /i:, ɪ, ɛ, a, ɒ, ɔ:/ were connected. However, for the vowel space area estimation all eight vowel categories were included.

[10] Acknowledgement goes to Richard Rhodes for the visual ideas of representing the VSA data.

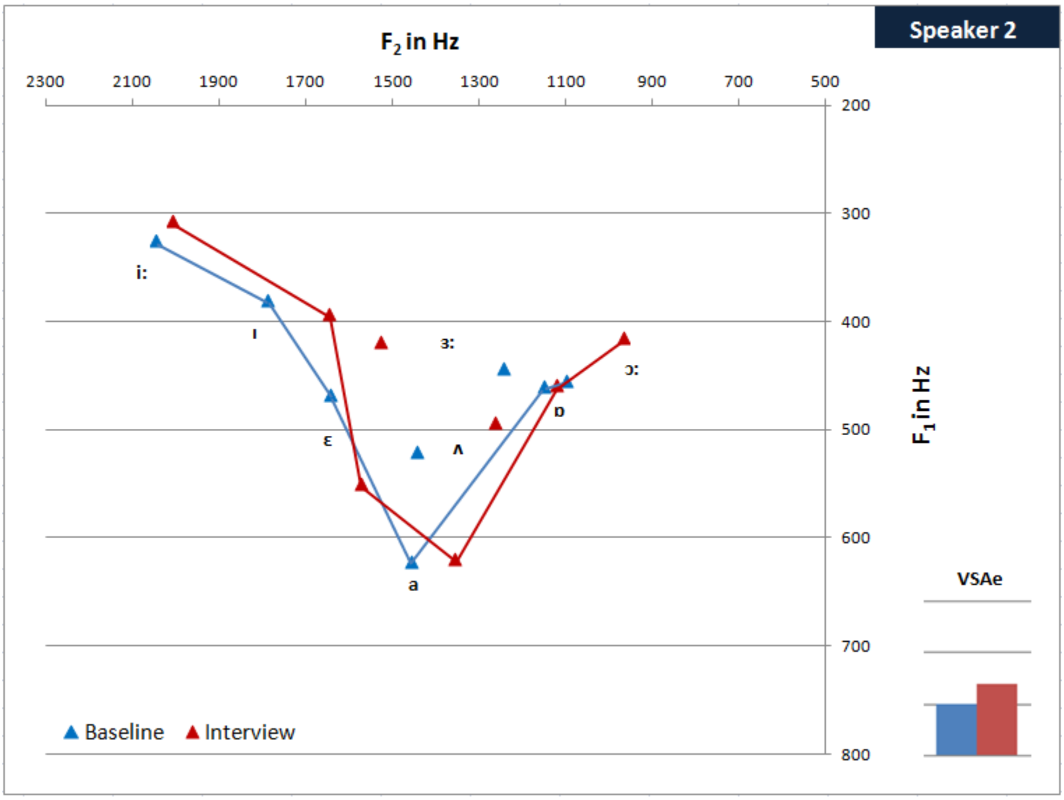

Figure 27 – Scatterplot of mean F_1 and F_2 values for all vowel categories in baseline and interview for speaker 2, connected to represent vowel space area. Histogram of vowel space area estimation for baseline and interview

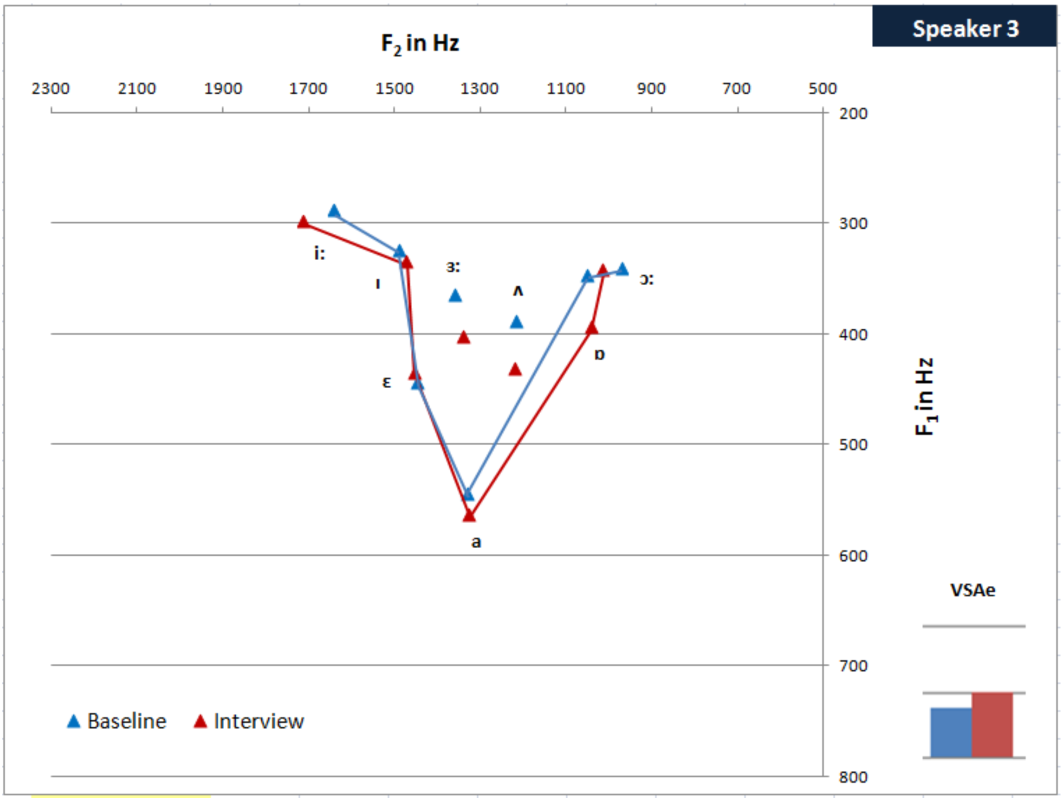

Figure 28 – Scatterplot of mean F_1 and F_2 values for all vowel categories in baseline and interview for speaker 3, connected to represent vowel space area. Histogram of vowel space area estimation for baseline and interview

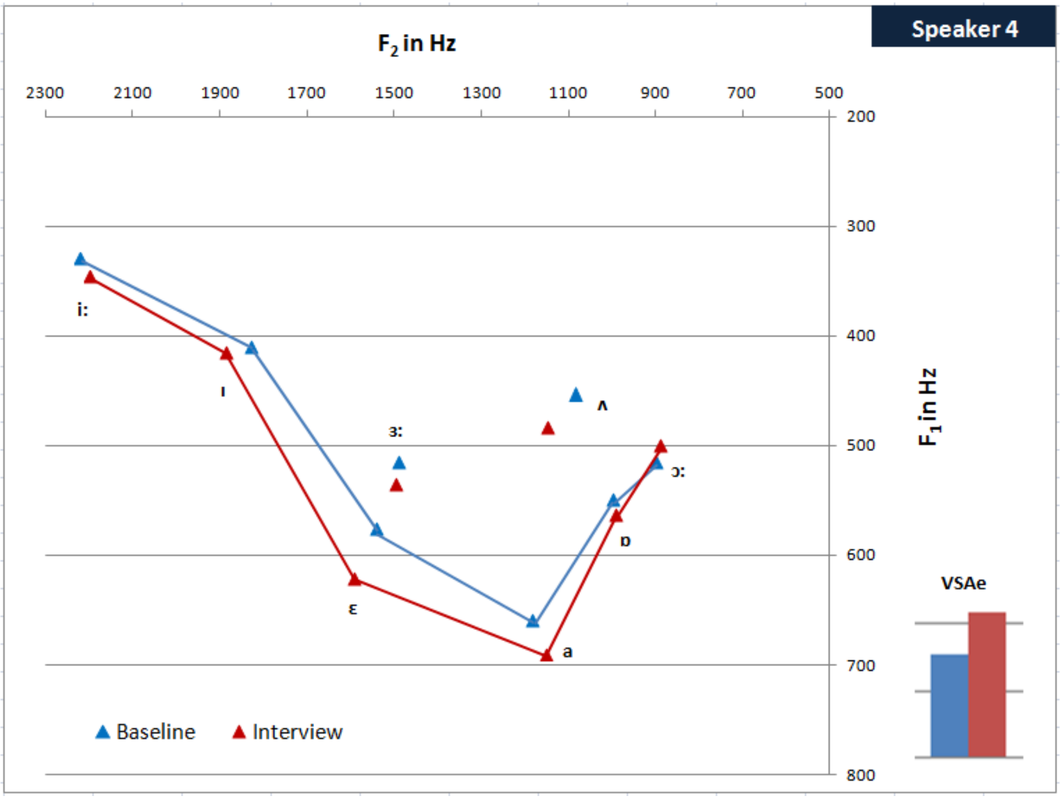

Figure 29 – Scatterplot of mean F_1 and F_2 values for all vowel categories in baseline and interview for speaker 4, connected to represent vowel space area. Histogram of vowel space area estimation for baseline and interview

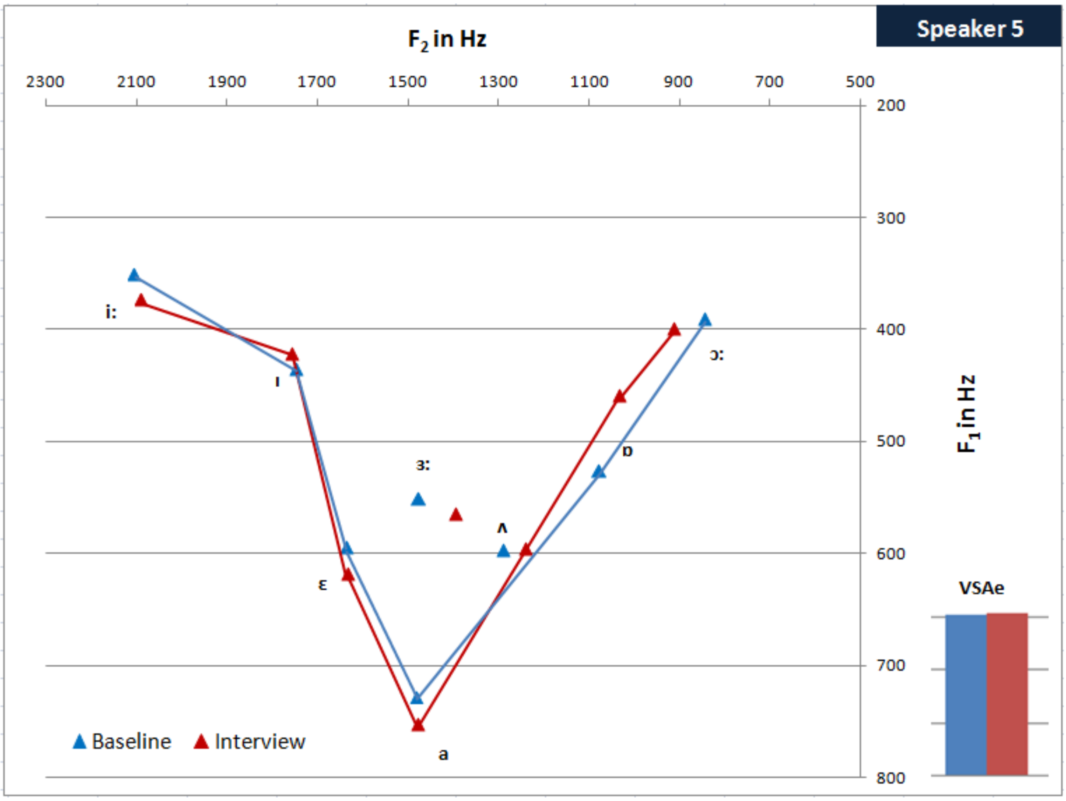

Figure 30 – Scatterplot of mean F_1 and F_2 values for all vowel categories in baseline and interview for speaker 5, connected to represent vowel space area. Histogram of vowel space area estimation for baseline and interview

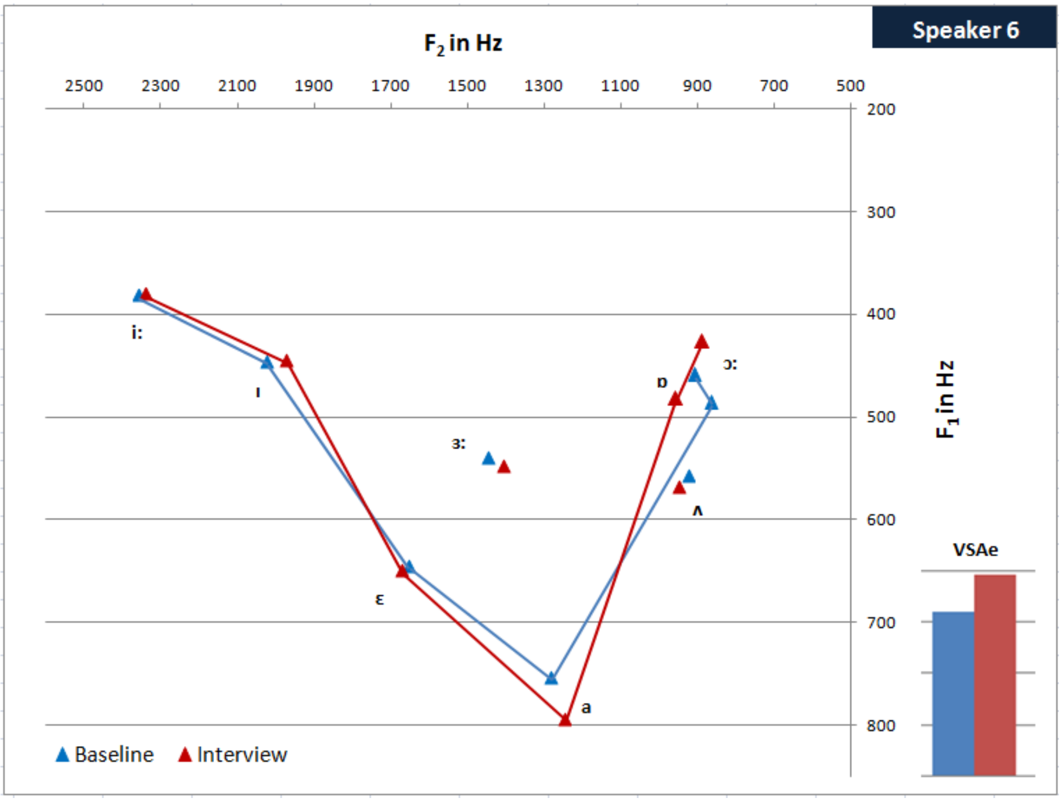

Figure 31 – Scatterplot of mean F_1 and F_2 values for all vowel categories in baseline and interview for speaker 6, connected to represent vowel space area. Histogram of vowel space area estimation for baseline and interview

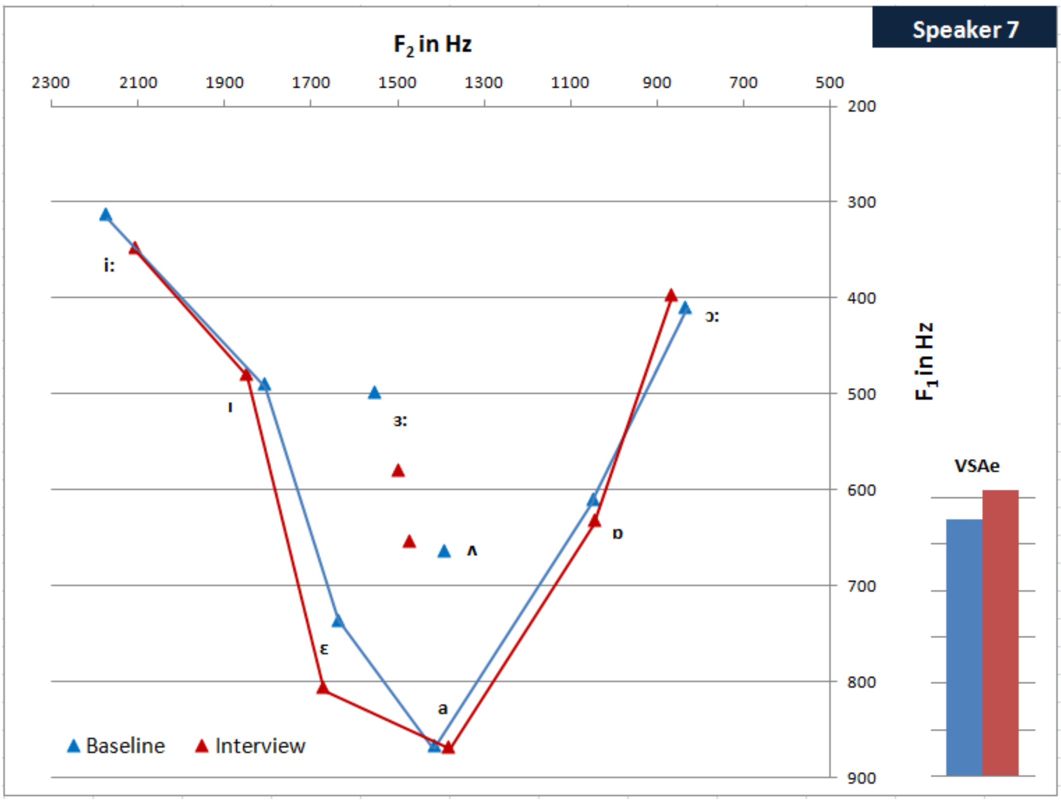

Figure 32 – Scatterplot of mean F_1 and F_2 values for all vowel categories in baseline and interview for speaker 7, connected to represent vowel space area. Histogram of vowel space area estimation for baseline and interview

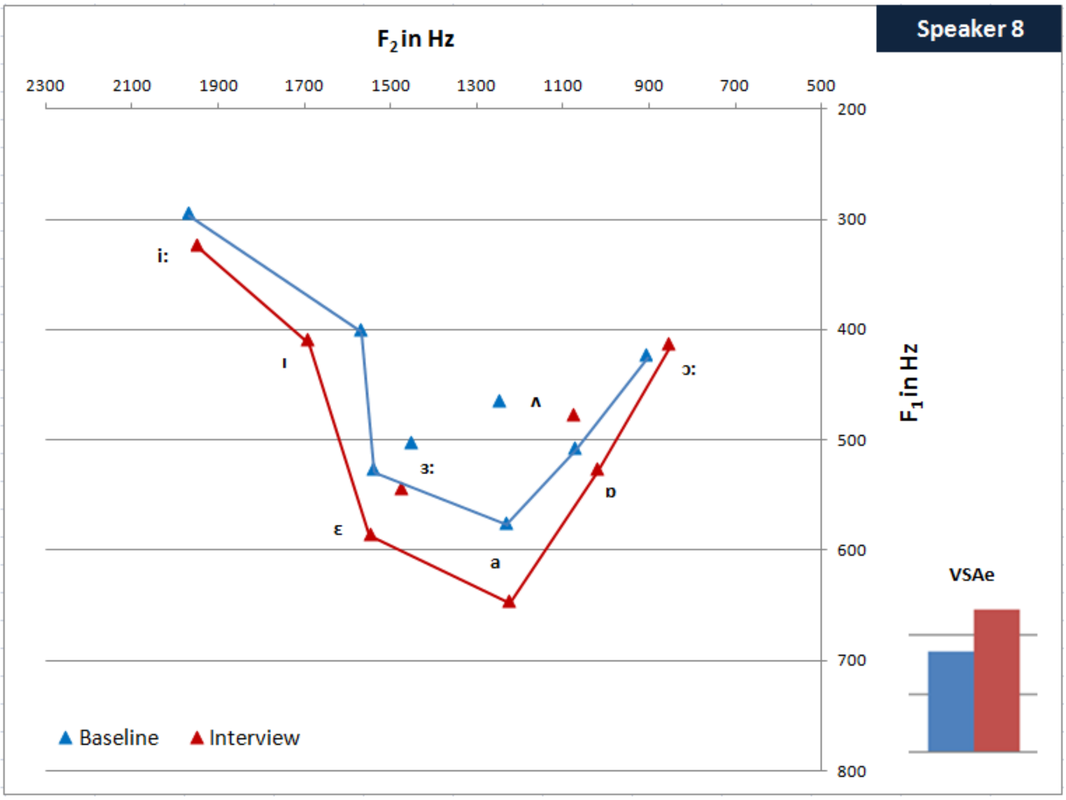

Figure 33 – Scatterplot of mean F_1 and F_2 values for all vowel categories in baseline and interview for speaker 8, connected to represent vowel space area. Histogram of vowel space area estimation for baseline and interview

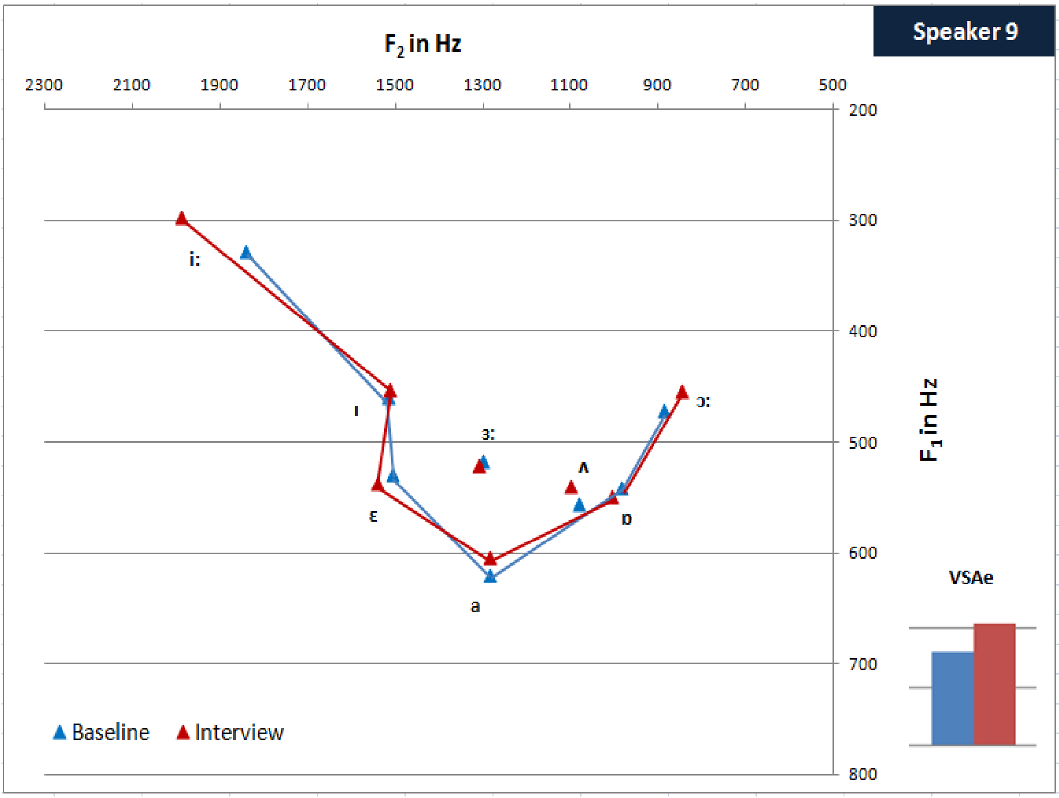

Figure 34 – Scatterplot of mean F_1 and F_2 values for all vowel categories in baseline and interview for speaker 9, connected to represent vowel space area. Histogram of vowel space area estimation for baseline and interview

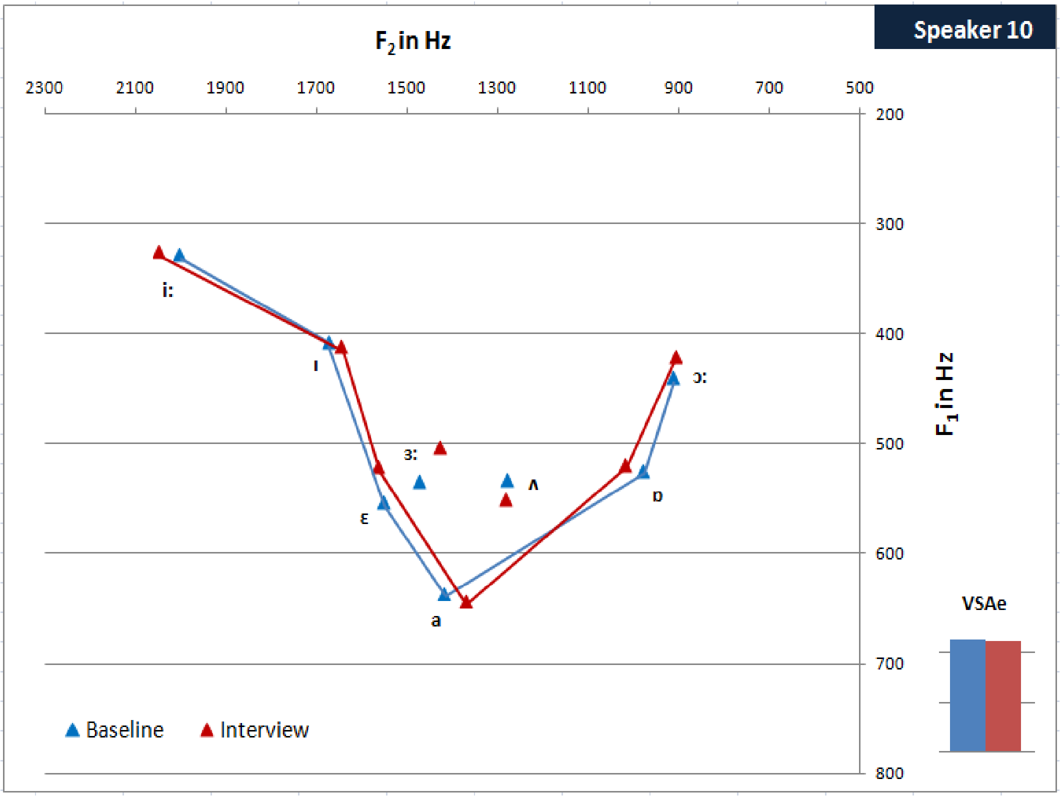

Figure 35 – Scatterplot of mean F_1 and F_2 values for all vowel categories in baseline and interview for speaker 10, connected to represent vowel space area. Histogram of vowel space area estimation for baseline and interview

Figures 36 and 37 illustrate that there appeared to be a uniform effect on vowel space area. Nine out of ten speakers showed an increase in vowel space area in the interview condition ranging from a mere 1% (1553 Hz^2) to a more substantial 41.6 % (36237 Hz^2). The increase in vowel space area from baseline (mean = 114534 Hz^2) to interview (mean = 136402 Hz^2) was significant ($T = 2$, $p \leq .01$, $r = -.58$). Participant 10 was the single speaker who displayed a reduction in his vowel space. Compared to the extent of the enlargement shown by the other speakers this decrease (1.6%) was minimal. In light of the general increase in working vowel space in the interview condition, the question arose whether some form of hyper-articulation was involved when speakers were being deceptive. Considering speakers' vowel space areas on an individual basis did not immediately provide for a straightforward answer. Hyper-articulation in the sense of 'high vowels' getting higher, 'low vowels' lower, 'front vowels' more front and 'back vowels' further back could only be observed in a minority of cases (Johnson et al., 1993, p. 519; Picheny et al., 1986, p. 444). Speaker 7 and speaker 8 could be regarded as producing vowels that were acoustically more extreme in the interview but the remaining speakers did not show a uniform stretching of their acoustic vowel space. Rather, expansion appeared to be variable with regards to which one of the two vowel dimensions- F_1/F_2- was involved and which vowel category was affected. For speakers 1 and 3, for example, stretching of the vowel space area mainly resulted from an increase in the F_1 of the /ɜː/, /ʌ/ and /ɒ/ vowels. While speaker 4 showed expansion in the front and central vowels /ɪ/, /ɛ/ and /a/, speakers 2 and 9 hyper-articulated vowel targets belonging to the /iː/ and /ɔː/ vowel categories. Speakers 5, 6 and 10 showed inconsistent expansion and reduction along F_1 and F_2 affecting variable vowel categories. What, at first sight, appeared to be a uniform pattern turned out to be more variable when considered on a speaker by speaker basis. Individual variability with regards to deceptive behaviour (including speech) was also one of the factors that emerged from the previous literature.

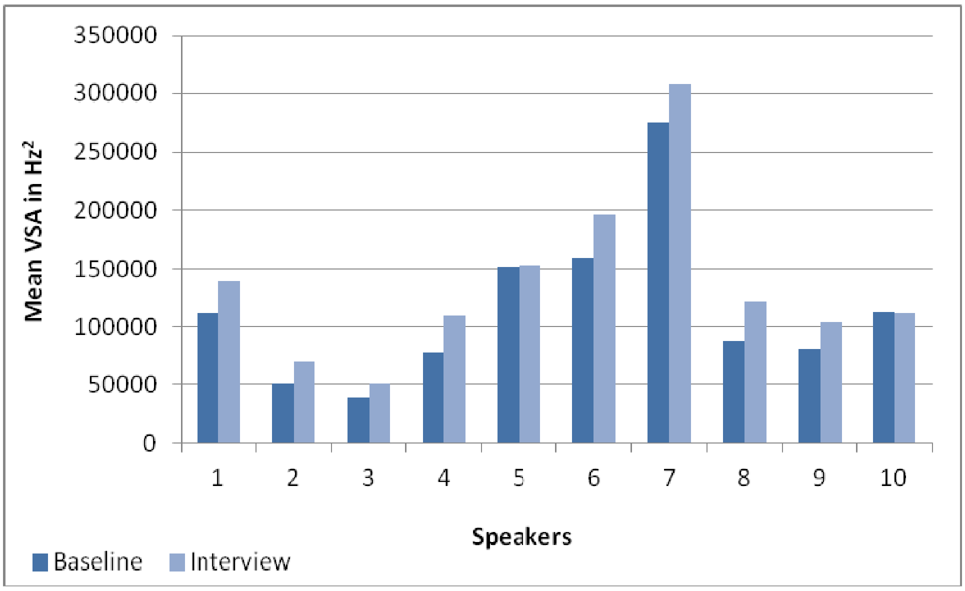

Figure 36 – Mean vowel space area in baseline and interview for all speakers

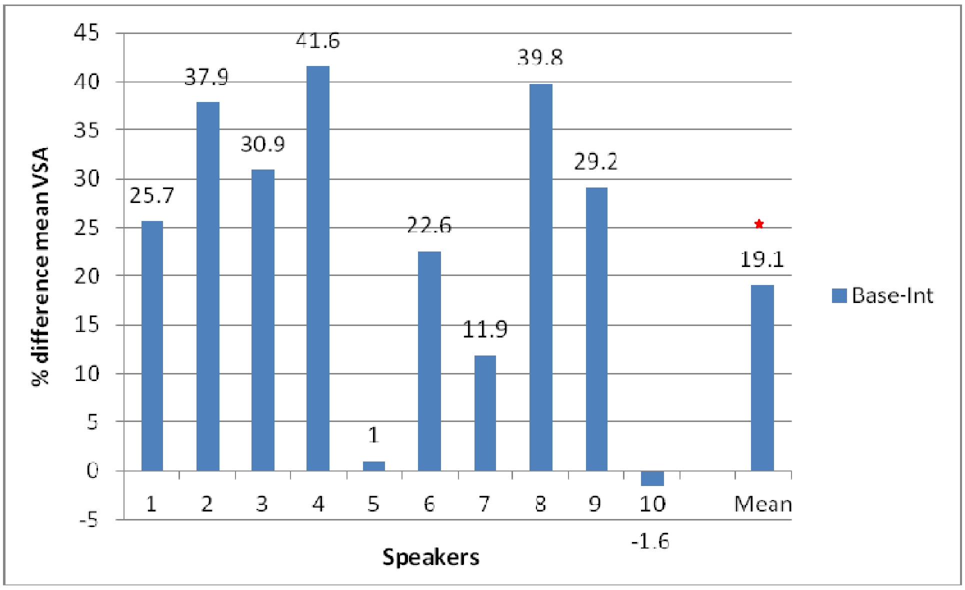

Figure 37 – % difference in mean vowel space area between baseline and interview for all speakers

The reader may have noticed the relatively large % increases for many of the speakers. A note of warning needs to be made in relation to interpreting results from geometric area measures such as the one employed here. Even limited changes along the F_1 or F_2 of dimension can result in large changes overall. Furthermore, it is acknowledged that both formant frequencies and vowel duration measurements need to be taken into account when assessing vowel space reduction or expansion (Turner et al., 1995). Although, it is believed that relevant observations can be made from the present vowel space analysis, future work could adopt a more refined analysis method by taking durational measurements into account.

7.1.6 Vowel format F_3

F_3 has received even less attention than F_1 and F_2 in the research literature on speech under stress and cognitive load. It is also rarely reported in studies investigating clear speech. Empirical findings in relation to F_3 behaviour when speakers are lying are non-existent. F_3 is less affected by the vocalic quality of vowels and, as described in Chapter 5, it provides information on speakers' voice quality or "personal voice characteristics" (Fant, 1970, p. 48). A few studies have reported changes in voice quality when speakers are stressed or experience cognitive load. Some report an increase in creaky voice, others show increased breathiness. DePaulo et al. (2003) reported that liars sound 'more tense'. These findings are mainly based on auditory assessments of voice quality. Indeed, in the case of Depaulo et al. (2003) the judgment of liars sounding 'more tense' was, for the most part, based on the evaluation of lay persons, i.e. non-phoneticians. Given limited amount of previous research, the analysis of voice quality in relation to deceptive speech, whether by means of auditory or acoustic measurements, is an area worth exploring. A change in F_3 may be indicative of a change in speakers' phonatory or laryngeal settings. Do speakers display such changes when lying and if so, are these consistent across speakers?

7.1.6.1 F_3 by speaker

Figures 38 and 39 suggest variability in direction and magnitude of F_3 change across speakers. An equal number of speakers increased and decreased and the average extent of increase (1.4%) was similar to that of decrease (1.3%). In terms of magnitude of change the F_3 values correspond to those observed for F_2. Speaker 2, 3 and 8 were the only speakers who showed a significant change in F_3 values across conditions. However, the decrease of 2.7% (61 Hz) and increases of 1.4% (32 Hz) and 2.5 % (60 Hz) were not very sizeable differences. Overall, the change in F_3 from baseline (mean = 2420 Hz) to interview (mean = 2421 Hz) was not significant ($T = 27$, $p \geq .05$, $r = -.01$).

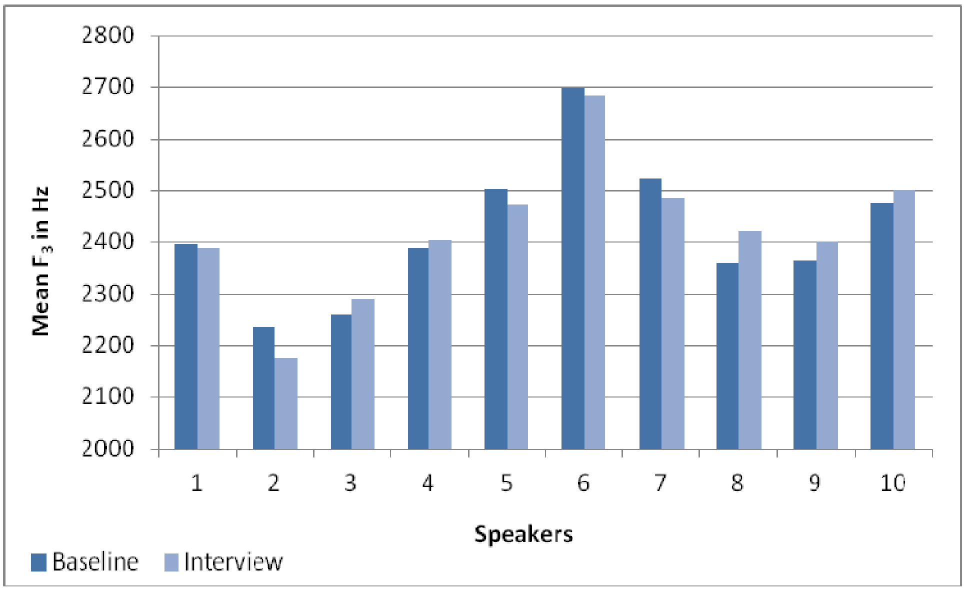

Figure 38 – Mean F_3 in baseline and interview for all speakers

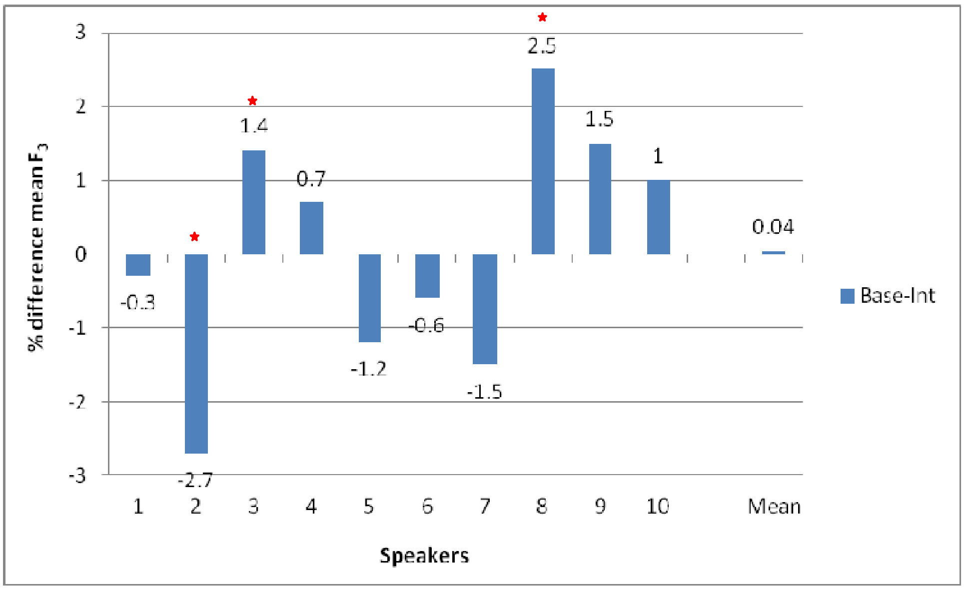

Figure 39 – % difference in mean F_3 between baseline and interview for all speakers

7.1.6.2 F_3 by vowel

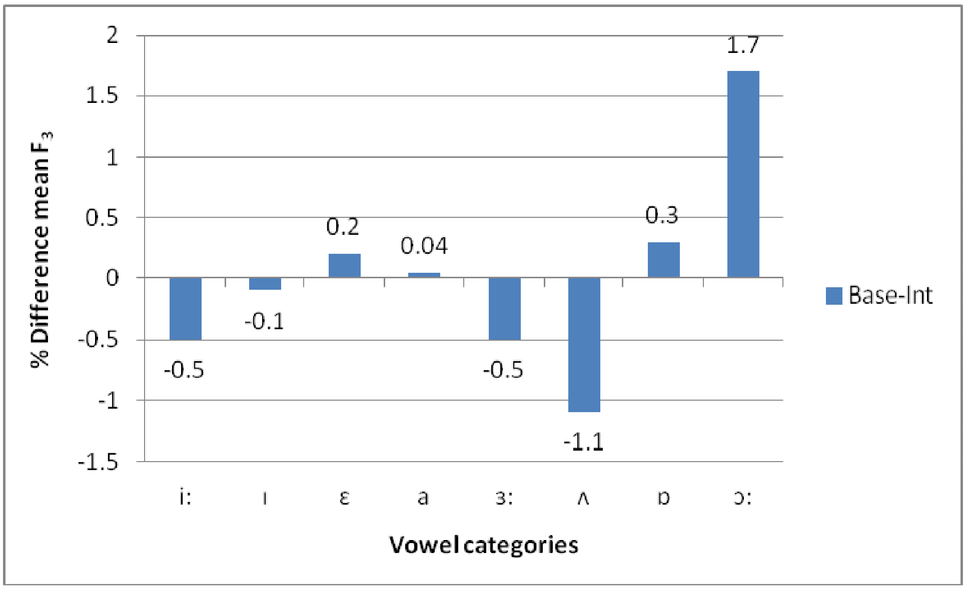

Figure 40 – % difference in mean F_3 between baseline and interview for all vowel categories

With the exception of /ʌ/ and /ɔ:/, the F_3 values of baseline and interview seemed to correspond very closely to each other (Figure 40). The largest difference occured in /ɔ:/ and amounted to a 1.7% (39 Hz) increase but even this change could be considered negligible. There was no significant difference in F_3 across the two conditions for any of the vowel categories. There was a tendency that high baseline F_3 values were more likely to be subject to an increase in the interview condition than lower F_3 values as signified by the negative trend line in Figure 41. The overall correlation between F_3 in baseline and F_3 decrease in the interview was statistically significant ($r = -.37$, $p \leq .001$); however, on careful consideration of the raw values it surfaced that the figures were rather equal and that the statistical significance should not be overvalued.

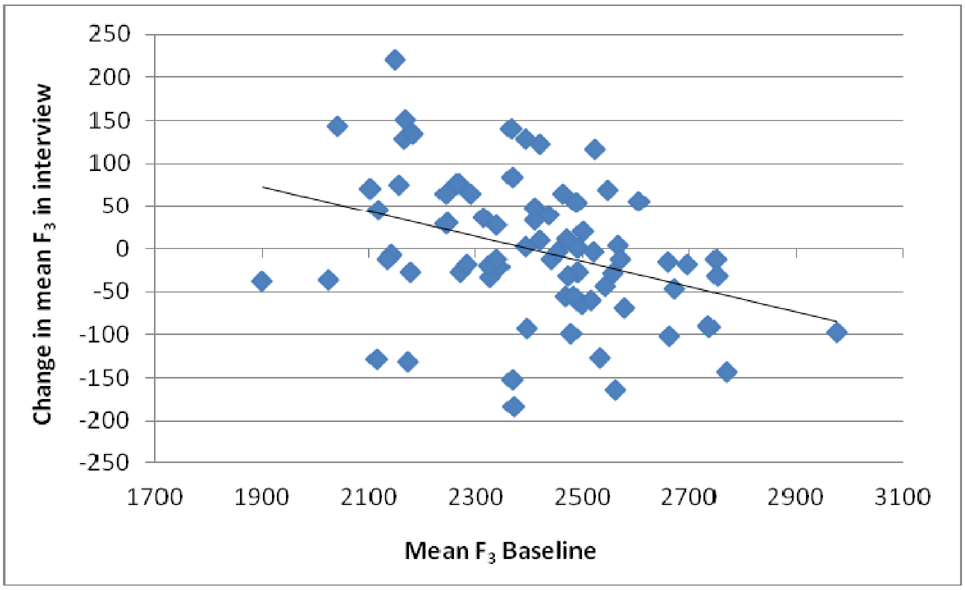

Figure 41 – Correlation between mean F_3 in baseline and change in mean F_3 in interview

Of course, although F_3 did not change considerably when speakers were lying, it does not mean that voice quality changes were not present. Nolan and Grigoras (2005) argue that "the cues which arise from the vocal tract ... are of course not immune from contingent variation ... however ... such variation is usually less severe in its effect than that to which laryngeal activity is sensitive" (p.168). F_3 may be seen as a supra-laryngeal cue to voice quality and the question arises whether a laryngeal cue to voice quality would provide more insightful findings. Chapter 4 briefly introduced acoustic measures of the glottal source signal which have been used to characterise different phonation types. One of those measures involved comparing the amplitude of harmonics H_1 and H_2 which represents an indirect measure of open quotient, i.e. the interval of the glottal cycle during which the glottis is open.

7.1.7 H_1-H_2

The difference between the amplitude of the fundamental frequency (H_1) and its first harmonic (H_2), has been used primarily to distinguish between breathy and modal phonation (Klatt and Klatt, 1990; Wayland and Jongman, 2003). When the vibration of the vocal folds has a large open quotient, the fundamental (H_1) is the most prominent in the spectrum of a vowel. If H_1 has a markedly higher amplitude than H_2, H_1-H_2 will be large and positive. A large positive H_1-H_2 value has been linked to breathy voice, while a small or negative H_1-H_2 may be indicative of creakiness (Ni Chasaide and Gobl, 1997). The single study investigating glottal waveform features in relation to deception can be seen in Torres et al. (2008). The authors reported that

H_1-H_2 was the parameter with the highest discriminatory ability of deceptive and non-deceptive speech.

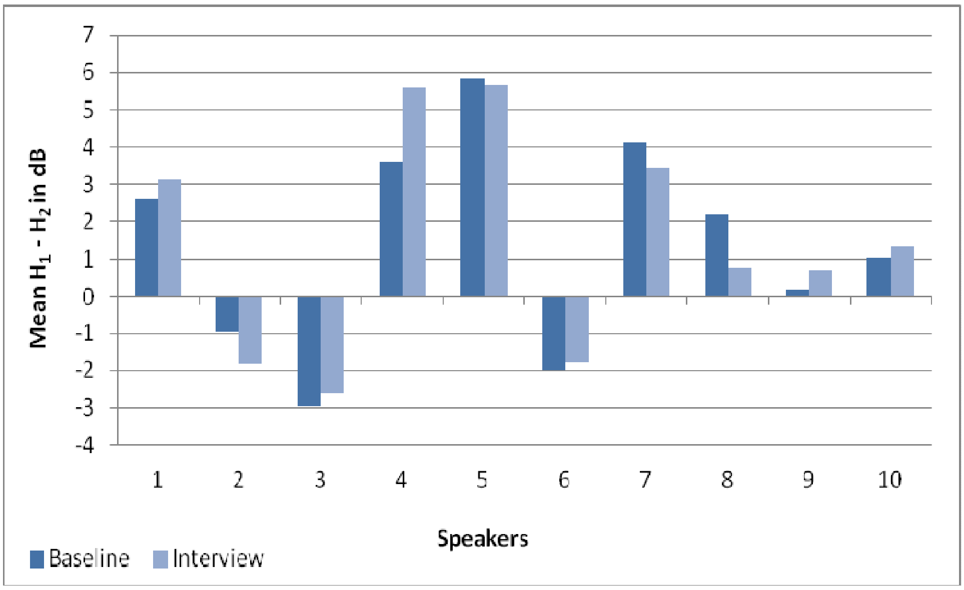

Figure 42 – Mean H_1-H_2 in baseline and interview for all speakers

The first observation that can be made from Figure 42 is that the majority of speakers have positive H_1-H_2 values and three speakers show negative H_1-H_2 values. This might suggest that, in the current sample, more speakers have a breathy type of phonation than creaky phonation. The more interesting question of whether speakers become more or less breathy can be seen in Figure 43. There appears to be variability in mean H_1-H_2 change across speakers with half the speakers increasing and half decreasing. Judging by the magnitude of the % difference values, the increases and decreases appear to be substantial. However, when considering the absolute differences as illustrated in Figure 44 the change in mean H_1-H_2 across conditions becomes less striking. Overall, the change in H_1-H_2 from baseline (mean = 1.37 dB) to interview (mean = 1.44 dB) was not significant ($T = 25$, $p \geq .05$, $r = -.06$). Although there does not appear to be a systematic pattern across conditions on an intra-speaker level, the difference in H_1-H_2 was significant for five of the speakers. It could be that this is a result of large sample sizes and indeed, the invariably small effect sizes for almost all speakers ($r = \leq .10$) suggested that the changes in H_1-H_2 within speakers had little empirical meaning.

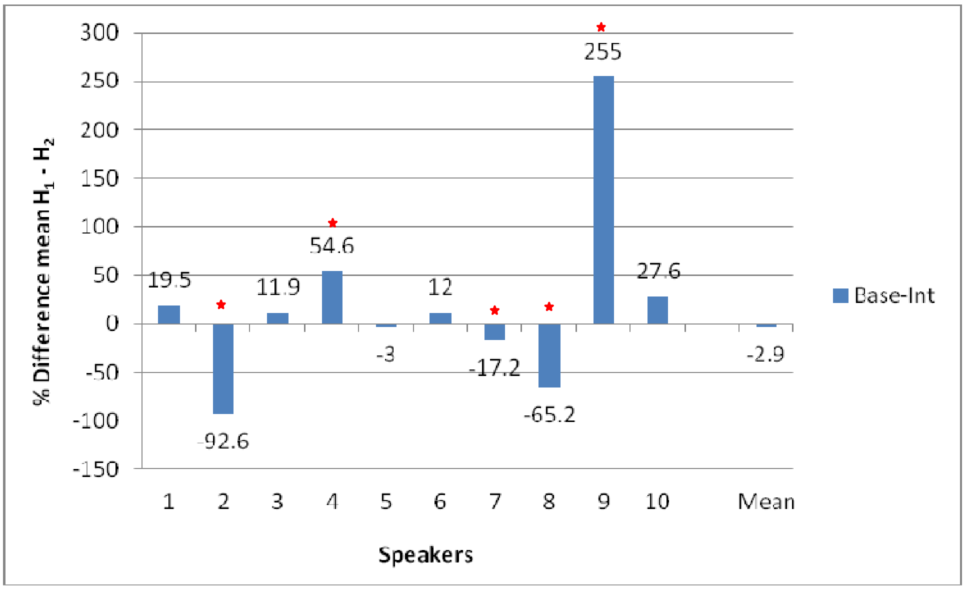

Figure 43 – % difference in mean H_1-H_2 between baseline and interview for all speakers

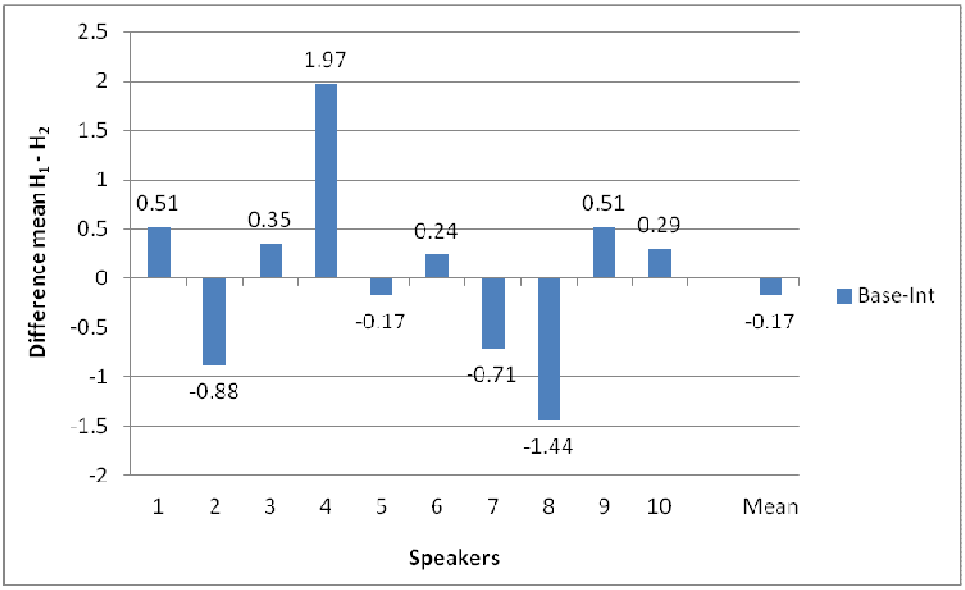

Figure 44 – Absolute difference in mean H_1-H_2 between baseline and interview for all speakers

Breathiness is expected to increase as f_0 decreases (Wayland and Jongman, 2003). Therefore, the differences in H_1-H_2 across conditions that could be observed may be attributable to a change in f_0. The correlation between the change in H_1-H_2 and the change in f_0 was not significant ($r_s = -.21$, $p \geq .05$).

There are limitations with the method used to estimate H_1-H_2 in the current study which could have affected the results. Firstly, potential confounding effects of low F_1 values on H_1-H_2 were not corrected for. Secondly, H_1-H_2 is based on spontaneous speech rather than sustained vowel productions. Nevertheless, the present attempt to analyse deceptive speech in relation to glottal source characteristics may be seen as a stepping stone for future research.

The current investigation employed acoustic analysis in order to assess possible voice quality and phonation changes as a function of deception. Complementary to the acoustic analysis, an auditory analysis of voice quality features conducted by trained phoneticians may have provided valuable insight.

7.1.8 Long–term formant distribution

One of the advantages of short term analyses of vowels as demonstrated in the mean formant frequency analysis presented above enables the discrimination between different vowels and thereby provides information on possible sound-by sound variation. However, this method of analysis also carries disadvantages. It is time-consuming. Even if the actual formant measurements were extracted automatically by way of a script, the task of categorising and marking vowels according to which set they belong to, i.e. /ɛ/ or /a/ still remains. Given the time-consuming nature, only a selected number of vowel categories tend to be analysed. In the current work, mean formant measurements were taken from eight monophthongs only; formant values of vowels which were not part of this set of eight monophthongs did not enter the analysis. It may be argued that a long-term analysis of formants which takes into account all available vocalic segments would have provided for more consistent results. As introduced in Chapter 4, long-term analysis of formants, i.e. calculations of long-term formant distribution (LTFD) provides information on speakers' habitual speaking patterns, e.g. tendency for lip-rounding. Perhaps, such a method which takes into account the long-term behaviour of formants may be better suited to analysing intra-speaker differences that could result from the act of lying. In view of this it was decided to further analyse formant frequencies by way of LTFD analysis. No previous work on long term behaviour exists in relation to deceptive speech. Below are the exploratory findings for the long-term analysis of formants F_1, $F_{2,}$ F_3 and F_4.

7.1.8.1 LTF_1 and LTF_2

As with the mean formant frequencies, a change in LTF_1 or LTF_2 across conditions may be indicative of vowel peripheralisation and/or a change in speakers' articulatory settings.

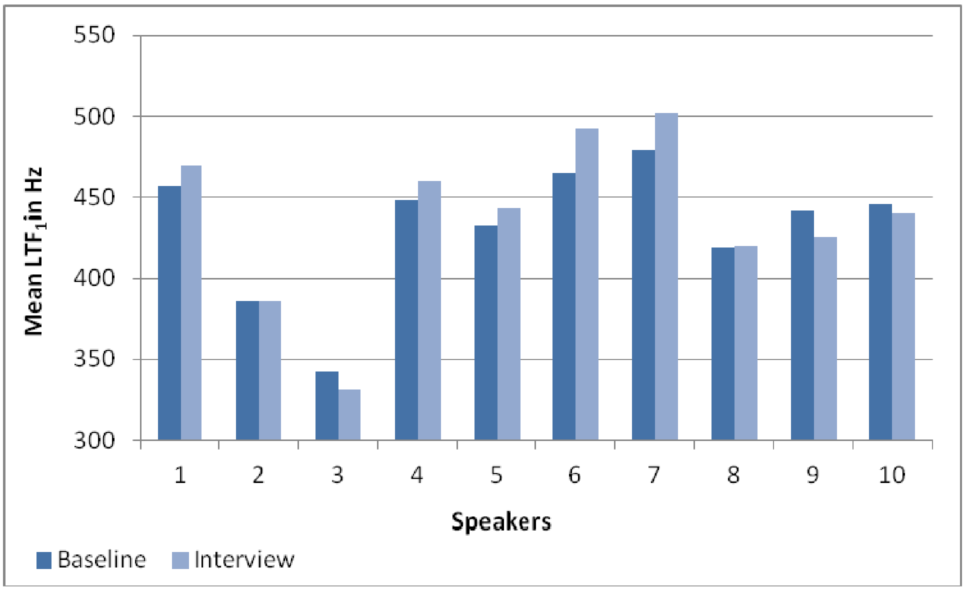

Figure 45 – Mean LTF_1 in baseline and interview for all speakers

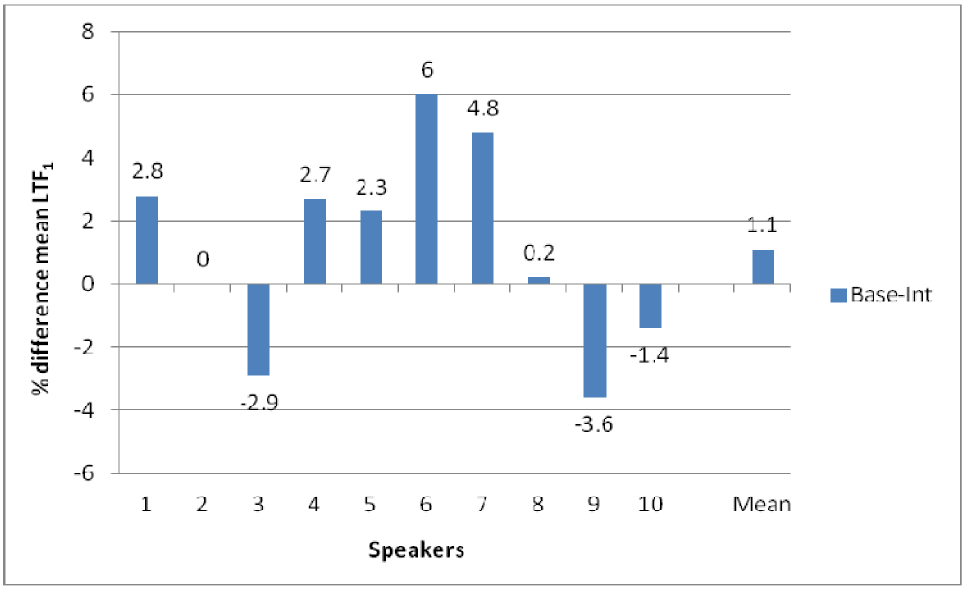

Figure 46 – % difference in mean LTF_1 between baseline and interview for all speakers

Figures 45 and 46 illustrate that there was variability in the direction of change in mean $LTF_{1.}$ Some speakers increase, some decrease and others do not change. The average increase amounts to 3.7 % and the largest increase of 6% (28 Hz) can be seen in speaker 6. With 2.6 %, the average decrease is slightly less than the increase. Overall, the change in LTF_1 from baseline (mean = 432 Hz) to interview (mean = 437 Hz) was not significant ($T = 12.5$, $p \geq .05$, $r = -.26$). Even though the mean LTF_1 values do not change across conditions, it may be the case that speakers show a change in variability of F_1 in interview which would be indicative of more or less precise articulation when lying. Indeed, for the majority of speakers LTF_1 SD increases in the interview; however it is only two speakers who show a notable change of around 30 Hz. On a group level, the increase in LTF_1 SD values from 132 Hz in baseline to 139 Hz in interview was not significant (T = 12, p ≥ .05, r = -.28).

There was no significant correlation between change in LTF_1 and change in mean f_0 (r_s = -.43, p ≥ .05). Wagner (2010) reported that LTF_1 was higher when speakers employed increased vocal effort. Indeed, as illustrated in Figure 47, there was a strong and positive correlation between change in LTF_1 and change in amplitude (r_s = .64 p ≤ .05). The value of the coefficient of correlation (R^2 = .41) indicates that around 41% of the variation in LTF_1 can be accounted for by the difference in amplitude measures. However, it cannot be said that the increase in LTF_1 was caused by the increase in amplitude; other factors may have contributed to the change in LTF_1.

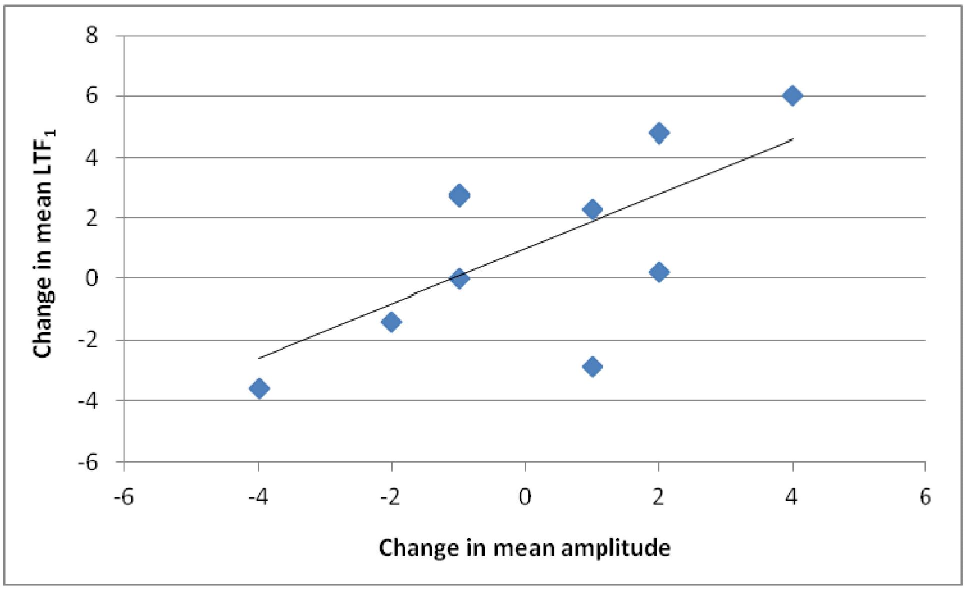

Figure 47 – Correlation between change in mean LTF_1 and change in mean amplitude

As was observed with LTF_1, the change in LTF_2 from baseline to interview was not consistent across speakers. Figures 48 and 49 demonstrate that around one third of speakers increased, one third decreased and the remainder did not change. The extent of change was similar for increasing and decreasing values; this was also observed for LTF_1. Overall, the difference in LTF_2 across conditions (0.4%) was negligible. Statistical tests confirmed that there was no significant difference between LTF_2 in baseline (mean = 1474 Hz) and LTF_2 in interview (mean = 1469 Hz) (T = 26, p ≥ .05, r = -.03). The increase in LTF_2 SD from baseline (mean = 317 Hz) to interview (mean = 326 Hz) was small (9 Hz) and not significant (T = 26, p ≥ .05, r = -.03). Looking at individual speakers, there was variability with half showing larger LTF_2 SD values in interview and others reducing their values. Speaker 8 was the only speaker who may be regarded as showing a substantial increase in LTF_2 SD of 124 Hz.

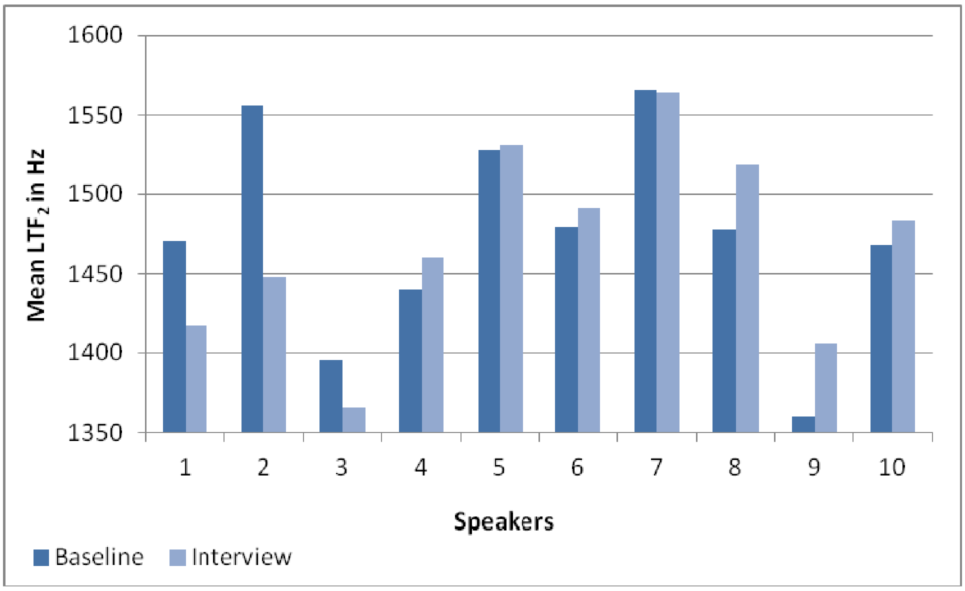

Figure 48 – Mean LTF_2 in baseline and interview for all speakers

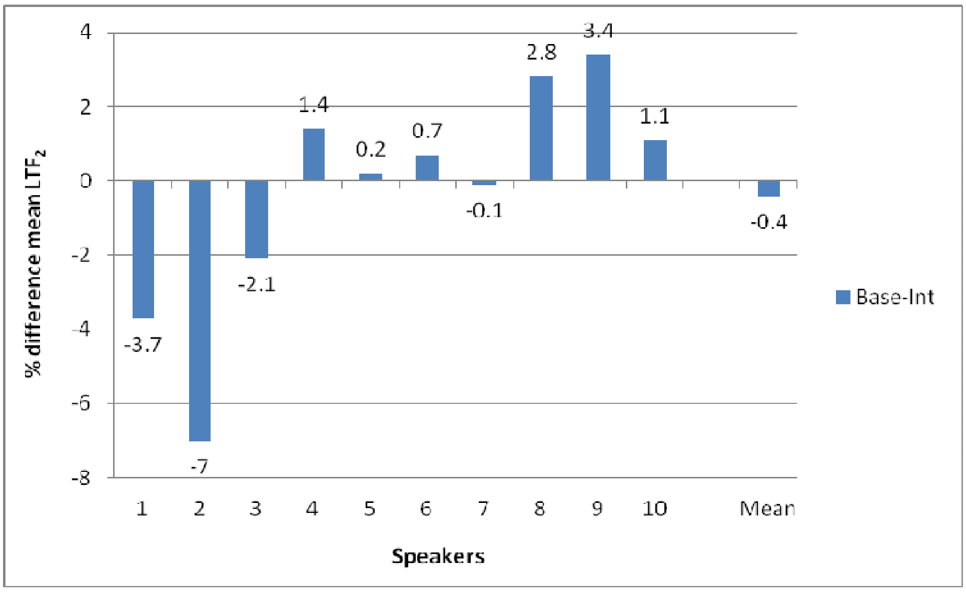

Figure 49 – % difference in mean LTF_2 between baseline and interview for all speakers

Significant differences in LTF_1 and LTF_2 did not occur as a function of deception. There was a tendency that speakers showed slight increases in LTF_1 in interview; however, LTF_2 appeared to be relatively stable across conditions. There was a significant correlation between change in LTF_1 and intensity which suggest that, at least, some of the observed variability in LTF_1 may be accounted for by an increase in speakers' vocal level.

7.1.8.2 LTF_3 and LTF_4

As indicated in Chapter 4, higher formants may provide insight into aspects such as voice quality. In view of this the mean formant frequencies of F_3 were analysed and the findings did not reveal a substantial change across baseline and interview. F_4 may also provide information on voice quality; however it has received even less attention in the research literature than F_3. A change in LTF_3 or LTF_4 between baseline and interview may be indicative of a change in voice quality when speakers were lying.

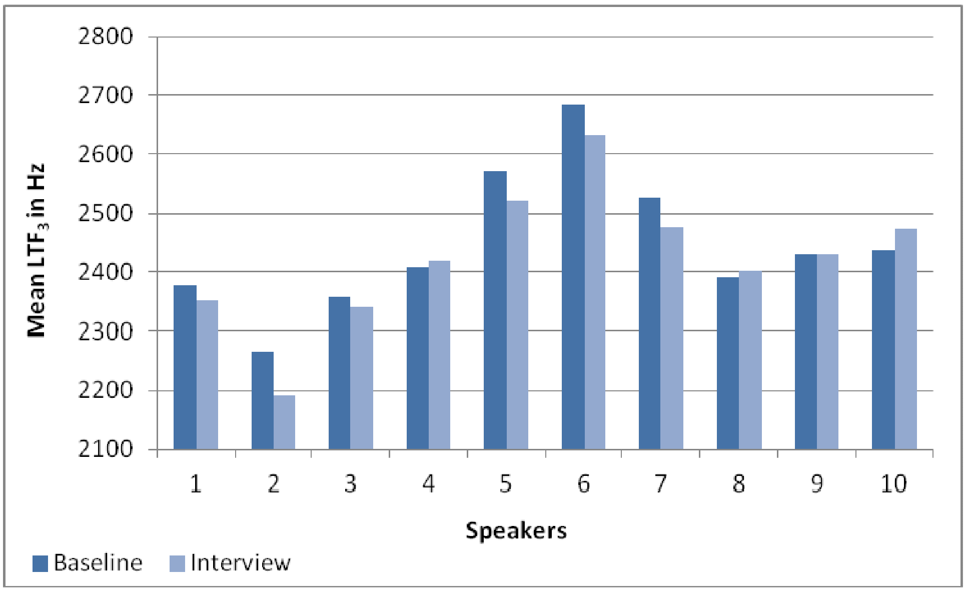

Figure 50 – Mean LTF_3 in baseline and interview for all speakers

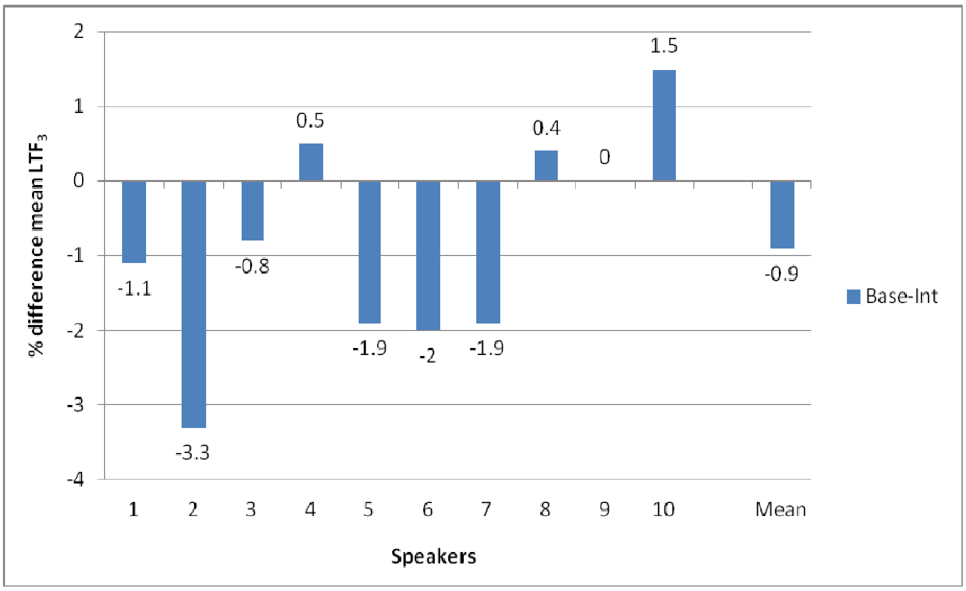

Figure 51 – % difference in mean LTF_3 between baseline and interview for all speakers

Figures 50 and 51 suggest that there was a tendency for LTF_3 to decrease in the interview. Six speakers follow this trend. The magnitude of decrease was relatively small however ranging from 0.8% (18 Hz) to 3.3% (75 Hz). Out of the three speakers who showed higher LTF_3 values in interview, speaker 10 had the largest increase of 1.5% (38 Hz). The decrease in LTF_3 from baseline (mean = 2445 Hz) to interview (mean = 2424 Hz) was not statistically significant ($T = 8$, $p \geq .05$, $r = -.38$).

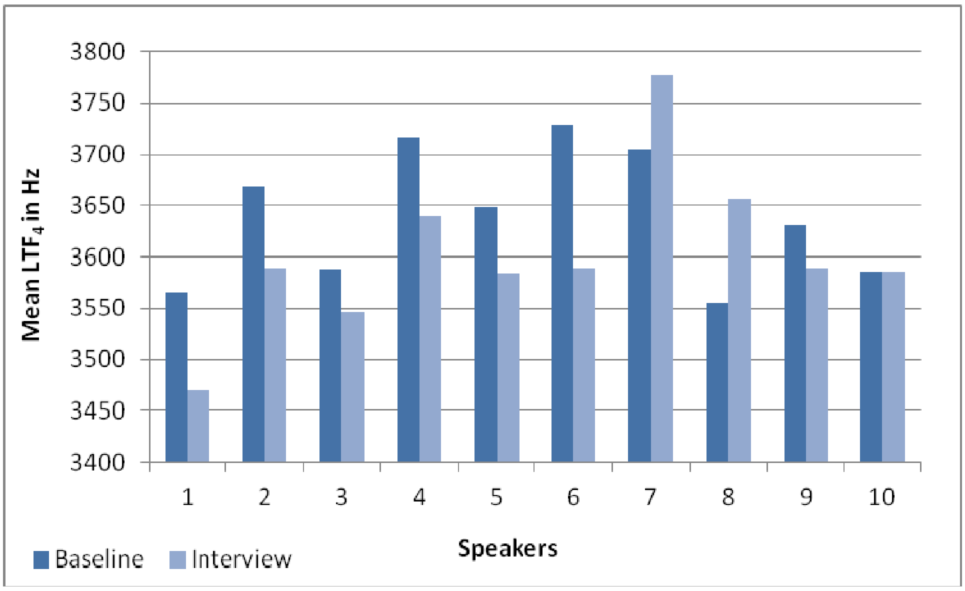

Figure 52 – Mean LTF_4 in baseline and interview for all speakers

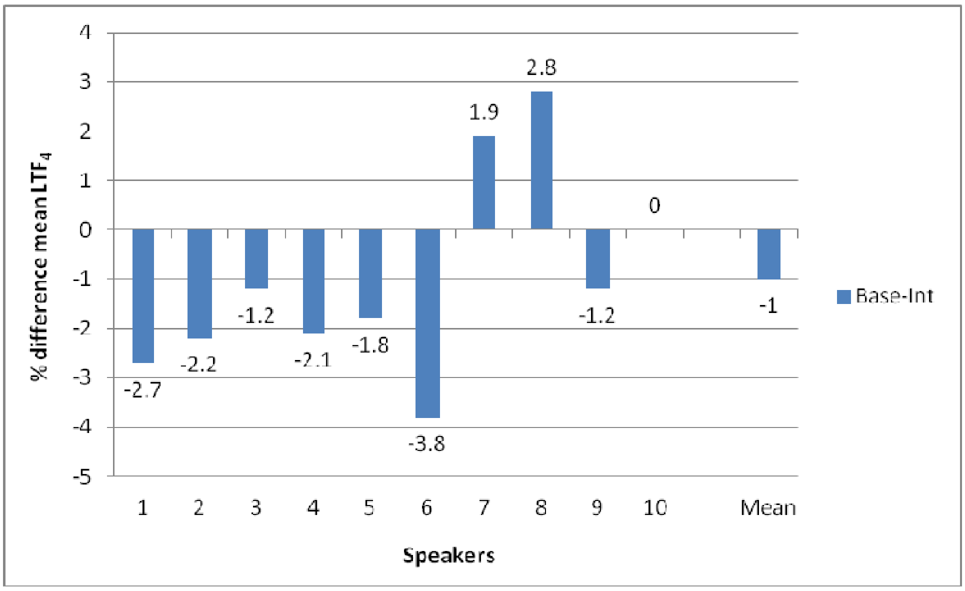

Figure 53 – % difference in mean LTF_4 between baseline and interview for all speakers

The results in relation to mean LTF_4, displayed in Figures 52 and 53, mirror those of mean LTF_3. There was a tendency for mean LTF_4 to be lower in the interview. Seven speakers conformed to this decreasing trend and the average reduction across these was 2.1%. Two speakers increased by an average of 2.3 %. The overall decrease of LTF_4 from baseline (mean = 3639 Hz) to interview (mean = 3603 Hz) was small and not statistically significant ($T = 12$, $p \geq .05$, $r = -.28$).

In general, there was no statistically significant difference between LTF_3 and LTF_4 across conditions. Although there was a slight tendency for both formant distributions to be lower in the interview, the reductions were negligible. Once again, there were differences between speakers in relation to direction as well as magnitude of change. Statistical testing was confined to the inter-speaker level only. As could be observed with the H_1-H_2 calculations, large sample sizes are likely to result in statistical significance even if the differences are minute. In the case of the LTFD data, sample sizes averaged 11745 measurements in the baseline and 16324 measurements in the interview. Intra-speaker comparisons were likely to have been significant but it is questionable how meaningful any significance would have been.

McDougall et al. (2012) investigated the relationship between LTF analysis and standard vowel formant measurements for fifteen speakers of Southern British English. They discovered that for the LTF distributions, mode was a more effective measure of central tendency than the mean. Using the mode, would have perhaps presented a more accurate picture of the distribution of the data.

In the current work, formant frequencies were analysed using different methods. The first method involved static measurements from the middle portion of a particular vowel. These static measures were then used to compute vowel space area estimates. The second approach estimated the long-term distributions of the first four vowel formants LTF_1-LTF_4. A third way in which the vowel data could have been analysed relates to the dynamic aspects of the vocalic elements (McDougall, 2006).

7.1.9 Summary of findings from acoustic analysis

Before continuing to describe the results of the temporal analysis of experiment 1, a summary of the findings of the acoustic analysis is presented (Table 7). As before, '↗' indicates an increase while '=' denotes no change.

Table 7 – Summary of the results of the acoustic analysis

Parameter	Result	Significance
Mean f_0	↗ (slight tendency)	ns
f_0 variability	↗ (slight tendency)	ns
Amplitude Difference	=	ns
F_1	↗ (slight tendency)	ns
F_2	=	ns
F_3	=	ns
LTFD	=	ns
Vowel Space Area	↗	**
H_1-H_2	=	ns

Statistical significance: ns = not significant, ** = significant at .01 level

Mean f_0 was relatively stable across both conditions. There was a tendency for speakers to have slightly higher mean f_0 values in the interview. However for only one of the speakers was this increase substantial. The f_0 SD data provided somewhat more consistency in that the majority of speakers showed an increase in f_0 variability in the interview. The lack of a systematic pattern inherent in the amplitude results was not entirely unexpected given the difficulties and impracticalities that are involved in order to obtain reliable and accurate measurements. With regards to vowel formant frequencies the results were complex. F_1 appeared to increase in the interview but there was no significant trend for F_2 or F_3. Moreover, there was variability between speakers and certain vowel categories were more affected than others. The results for vowel space area estimation were more interesting in that all but one speaker increased their working vowel space in the interview. However in the majority of cases the expansion was not a uniform acoustic stretching of the vowel space but varied according to vowel category and across speakers. Neither H_1-H_2 nor LTFD analysis showed any significant or consistent changes across conditions.

So far, it appears that speakers did not differ between baseline and interview in relation to the acoustic parameters of speech analysed. Are there changes in the temporal domain?

7.2 Temporal analysis

7.2.1 Articulation rate

Research on stress reports that speakers increase their rate of articulation when stressed; findings in relation to cognitive load suggest the opposite, i.e. a decrease in AR with cognitive demanding situations. There are no deceptive speech studies which reported findings on AR.

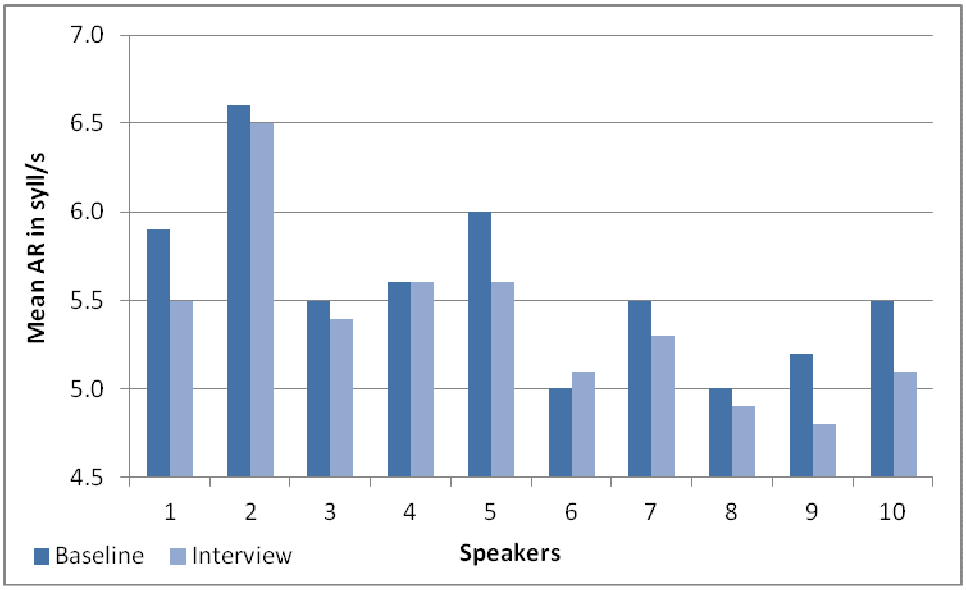

Figure 54 – Mean AR in baseline and interview for all speakers

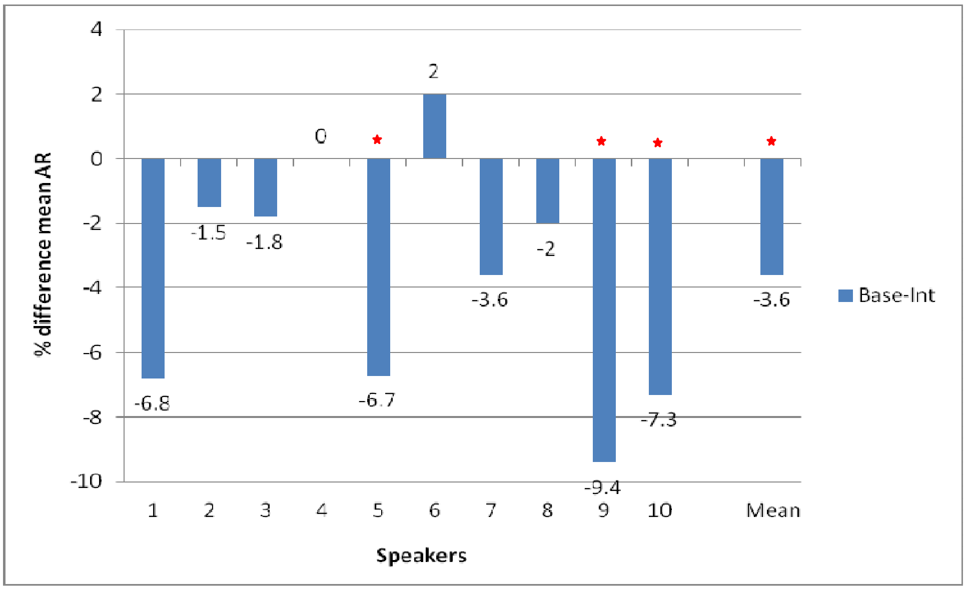

Figure 55 – % difference in mean AR between baseline and interview for all speakers

It becomes immediately obvious from Figures 54 and 55 that for the majority of speakers mean AR values were decreasing in the interview condition. The magnitude of decrease ranged from 1.5% (0.1 syll/s) to 9.4% (0.5 syll/s). Speaker 6 was the single speaker who showed an increase in mean AR; this increase was only small however amounting to a mere 2% (0.1 syll/s). The difference in AR between baseline (mean = 5.6 syll/s) and interview (mean = 5.4 syll/s) was significant ($T = 2.5$, $p \leq .05$, $r = -.54$). On an intra-speaker level, the decrease was significant for three speakers as marked by the red asterisks. An observation that was of a more fascinating nature was the apparent consistency between speakers. For the majority of parameters analysed so far, individual variability tended to prevent systematic patterns from emerging.

AR has often been regarded as highly speaker-specific and "remarkably invariant" (Goldman-Eisler, 1968, p. 25). However, more recent studies have shown that this may not be the case and that AR is subject to intra-speaker variation (Gold, 2012). One factor which has been shown to have an effect on AR is utterance or phrase length as expressed in number of syllables. Quene (2008) have demonstrated that longer phrases, i.e. phrases containing more syllables tend to be spoken with a faster rate. Jacewicz et al. (2010) on the other hand report that shorter phrases were spoken faster in their sample of American English speakers. Figure 56 confirms that the interview tends to contain shorter utterances, i.e. less syllables in an IP phrase, compared to the baseline ($T = 4$, $p \leq .05$, $r = -.53$).

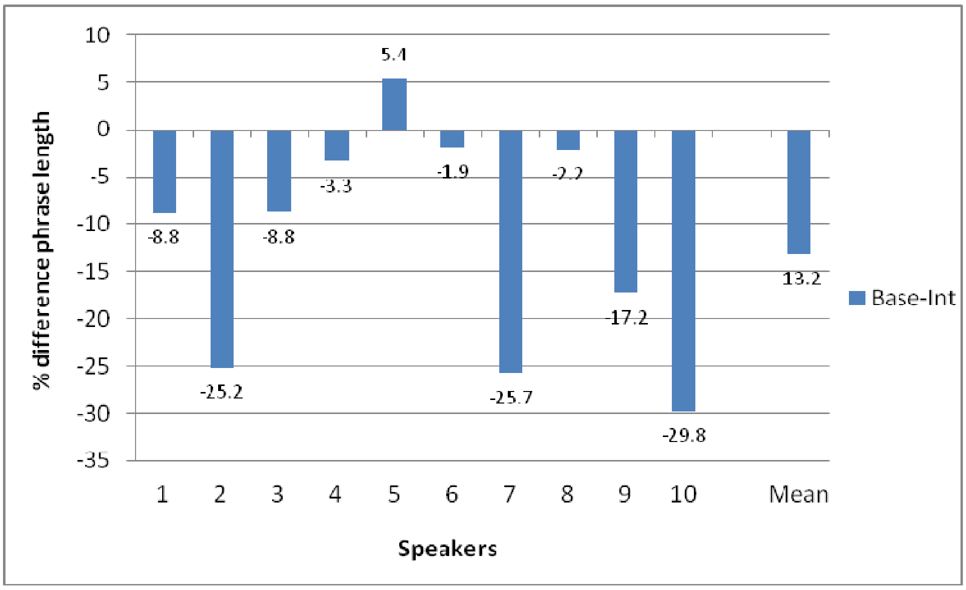

Figure 56 – % difference in length of utterance between baseline and interview for all speakers

However, when correlating the difference in mean AR across conditions with the difference in utterance length, no significant or strong trend appeared ($r_s = .33$, $p \geq .05$). Figure 57 illustrates that speakers who had similar phrase length in baseline and interview nevertheless showed marked differences in mean AR across conditions. Similarly, speakers who had notable differences in phrase length did not show corresponding differences in mean AR. If, for example, 'anticipatory shortening' had been responsible for the difference in mean AR across conditions one would have expected there to be a strong positive correlation between decrease in phrase length and decrease in mean AR.

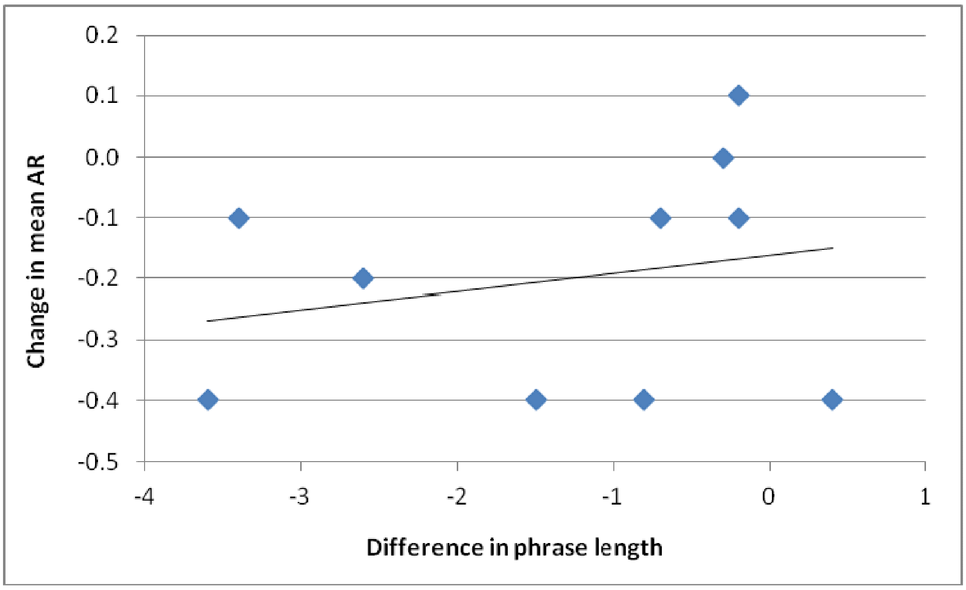

Figure 57 – Correlation between difference in phrase length and change in mean AR

The findings correspond to those of speech under cognitive load; they do not match the predictions of the effect of stress. Could the act of lying have been responsible for the decrease in mean AR?

7.2.2 Speaking rate

Research has shown that speakers reduce their speaking tempo in all three domains, i.e. when stressed, when experiencing cognitive load and when speaking more clearly. The literature in relation to changes in speaking tempo as a function of deception shows a disparity in that some studies report increases in tempo, others decreases, and yet others no change.

Figures 58 and 59 show that mean SR decreased in interview by 4.3 % (0.2 syll/s). Five speakers showed a reduction in mean SR in the interview and for four of these this decrease, which averaged 13.9%, was significant. Only two speakers demonstrated an appreciable increase with 8.2% (0.4 syll/s) and 5.3% (0.2 syll/s), respectively; however, these increases in mean SR were not statistically significant. The remaining speakers did not differ substantially between conditions. Although the decrease in SR from baseline (mean = 4.2 syll/s) to interview (mean = 4.0 syll/s) was not statistically significant ($T = 8$, $p \geq .05$, $r = -.38$) a strong tendency for a slower rate of speech in the interview was nevertheless apparent.

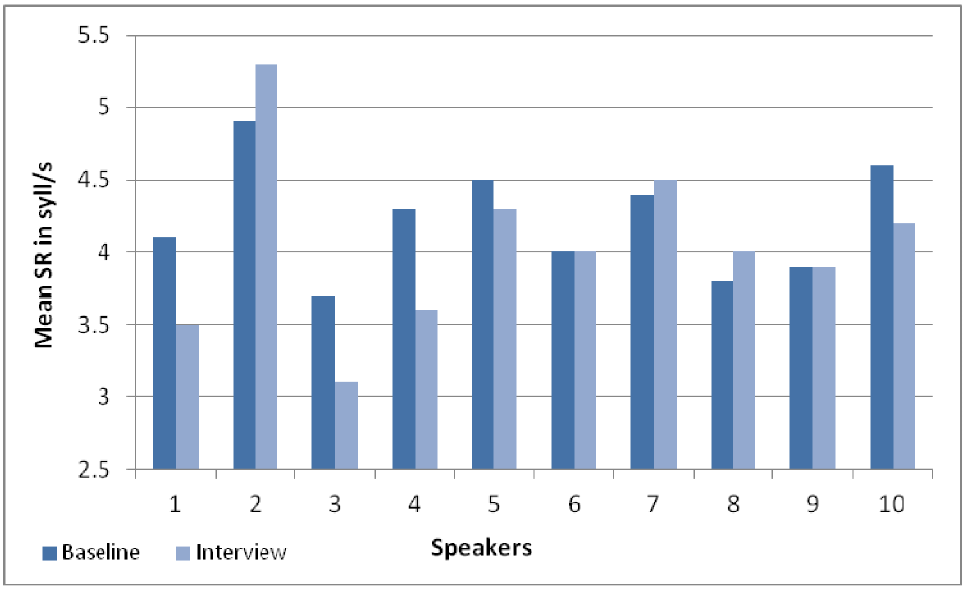

Figure 58 – Mean SR in baseline and interview for all speakers

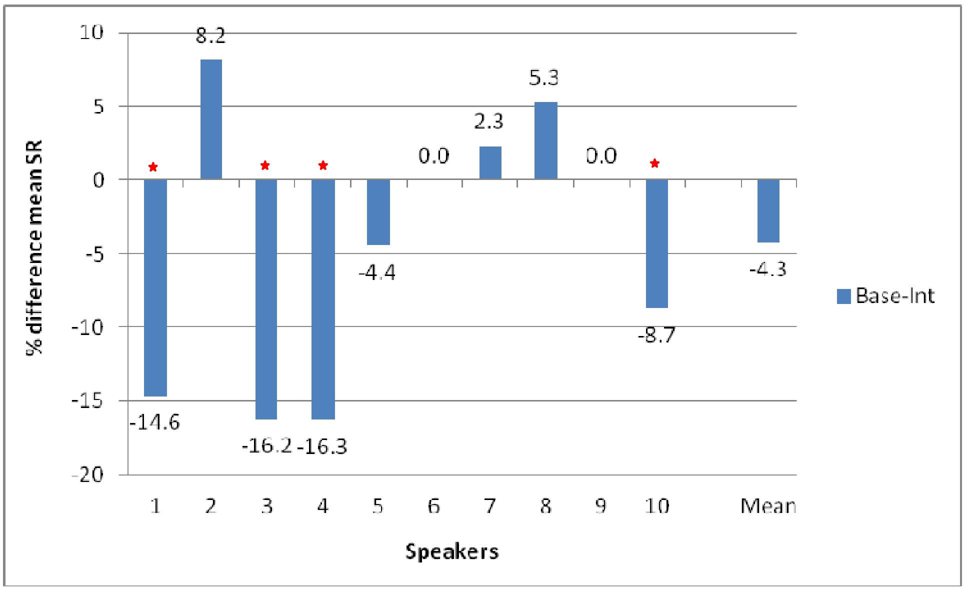

Figure 59 – % difference in mean SR between baseline and interview for all speakers

For four speakers the decrease was significant (red asterisks). These speakers are not the same that showed the significant decrease in mean AR; this does not come as a surprise, however, as SR and AR measure different aspects of speech tempo.

The task of lying in the interview may have increased speakers' level of cognitive load and they may have perceived the situation as more stressful than the baseline. It is also possible that speakers were speaking more clearly in the interview in order to present a sincere and trustworthy image. The decrease in SR may have been a result of the combined occurrence of all three processes. On the other hand, could the decrease in mean SR be connected to factors unconnected to the deceptive act?

7.2.3 Silent pauses (frequency and duration)

Two aspects of pausing behaviour were examined- frequency and duration. Unfortunately, statistical analysis of the frequency results could only be performed on an inter-speaker level. Research has reported a decrease in pausing with stress; however, the findings of cognitive load and clear speech research tend to suggest an increase in occurrence and duration of pauses. An increase as well as a reduction in pausing has been reported in relation to deceptive speech.

7.2.3.1 Frequency

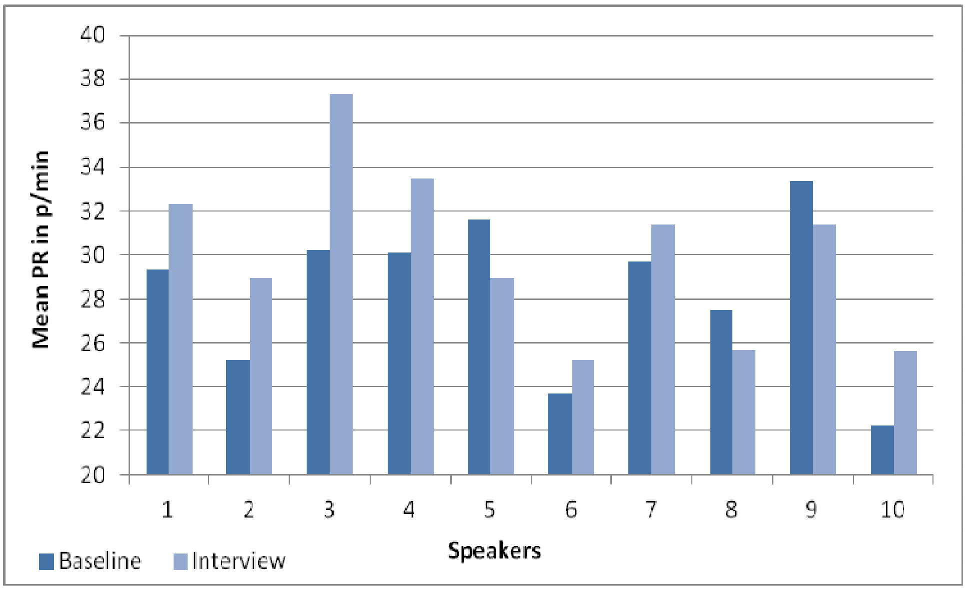

Figure 60 – Mean PR in baseline and interview for all speakers

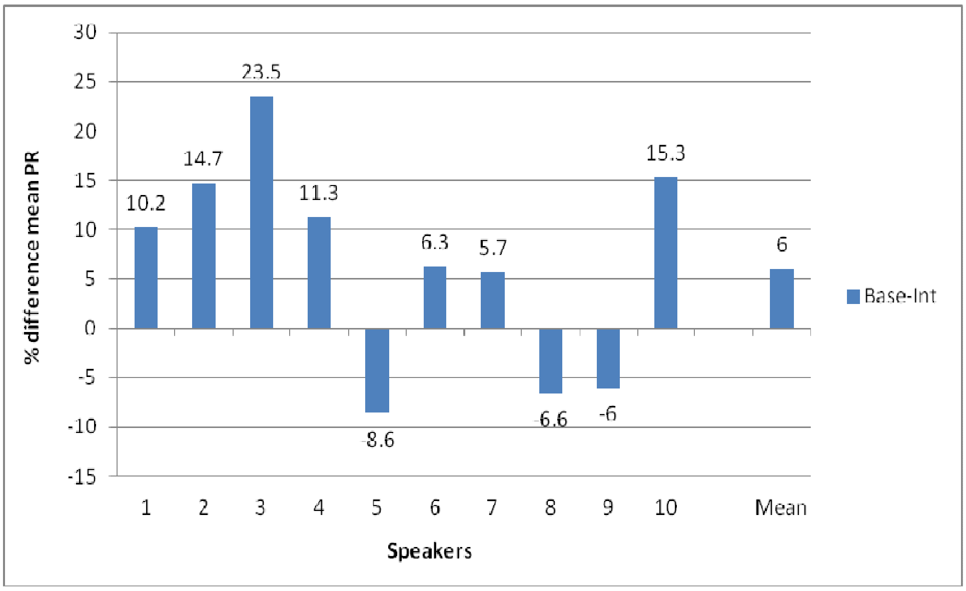

Figure 61 – % difference in mean PR between baseline and interview for all speakers

Although statistical testing did not produce a significant difference in the number of silent pauses between baseline (mean = 28.3 p/min) and interview (mean = 30.0 p/min) ($T = 12$, $p \geq .05$, $r = -.35$), there was a clear tendency for the quantity of pauses to increase during the interview. This is confirmed by the medium sized effect size ($r = -.35$). Seven out of the ten speakers conform to this general tendency showing an average increase of 12.4%. Three speakers showed a reduction in pause frequency in the interview which averaged 7.1 % (Figures 60 and 61).

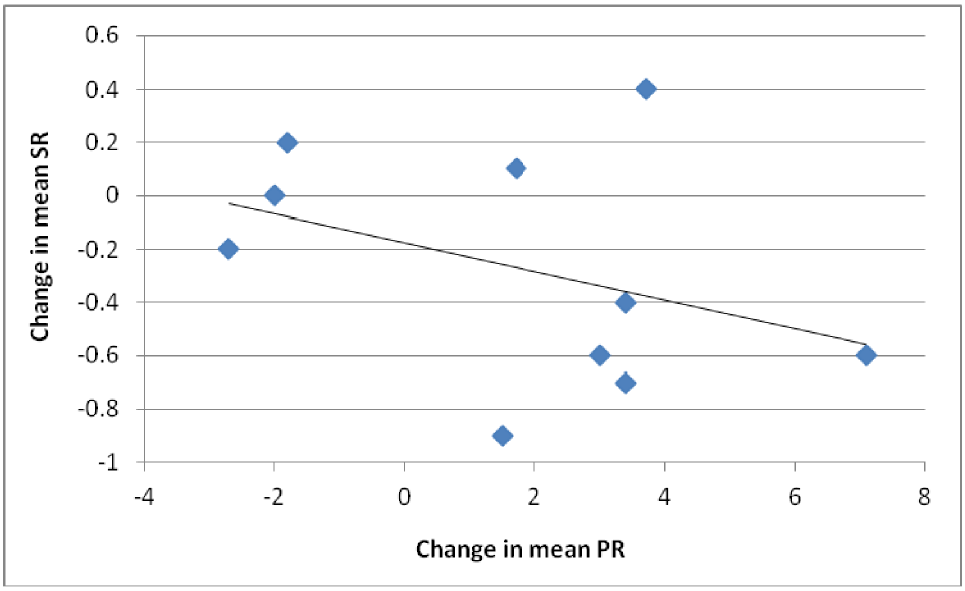

Figure 62 – Correlation between change in mean PR and change in mean SR

As explained in Chapter 4, SR is a measure of speaking tempo which is heavily influenced by pausing behaviour. It is very conceivable that the increase in pause frequency is related to the decrease in mean SR. Indeed, Figure 62 demonstrates the negative relationship. As number of pauses increased in interview, mean SR measure reduced also. The correlation was not significant but, nevertheless, had a medium effect size (r_s = -.32, p ≥ .05).

7.2.3.2 Duration

A noticeable consistency appears when inspecting the pause duration measurements (Figures 63 and 64).

With an average increase of 10.4%, silent pauses were significantly longer in the interview (mean = 518 ms) compared to the baseline (mean = 572 ms) (T = 2, p ≤ .01, r = -.58). All but one speaker followed this trend. The durational increase ranged from 1.6% (10 ms) to 30.0% (133 ms). Only 1 speaker showed a reduction in duration of pauses in interview; however this only amounted to 1.2% (8 ms). Perhaps surprisingly, the increase was statistically significant for speakers 4 and 9 only and, based on the visual impression, one might have expected speaker 2 to reach similar statistical significance. The raw measurements do suggest noticeable variability in duration of pauses within speakers which may provide an explanation for the lacking statistical significance. Table 8 gives the SD values for pause duration for all speakers in baseline and interview.

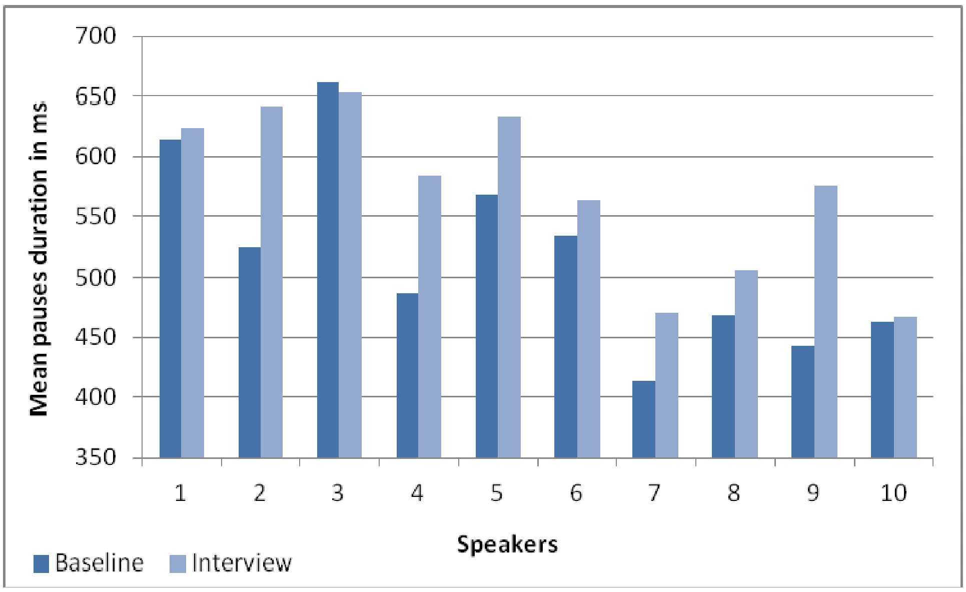

Figure 63 – Mean pause duration in baseline and interview for all speakers

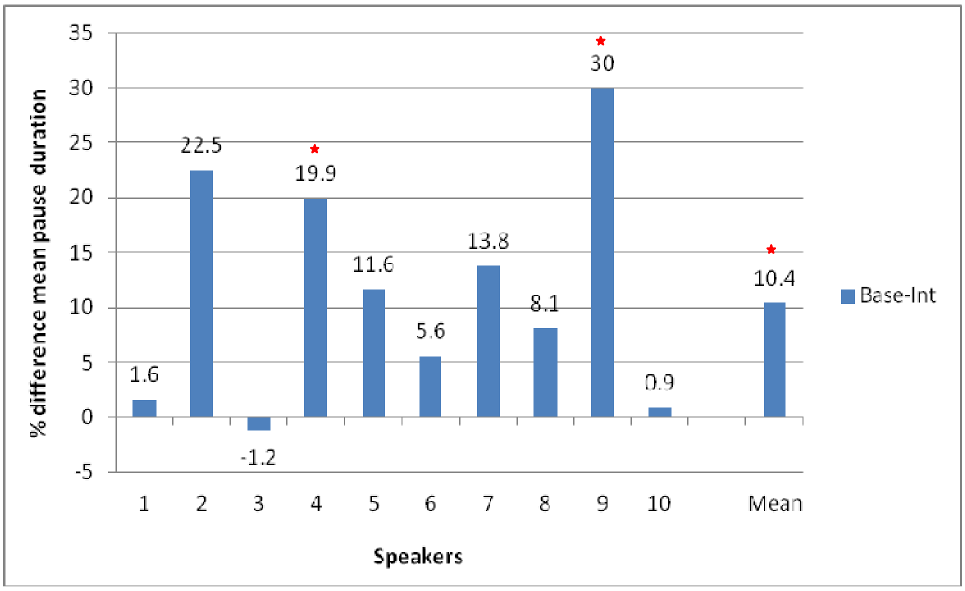

Figure 64 – % difference in mean pause duration between baseline and interview for all speakers

Table 8 – SD values (in ms) of the mean pause duration for all speakers in baseline and interview

Participant	Baseline	Interview
1	413	382
2	341	668
3	396	636
4	331	457
5	383	488
6	366	378
7	252	349
8	294	306
9	296	442
10	320	367
Mean	**339**	**447**

More generally, Table 8 proves that SD values were relatively large. It is conceivable that the absence of an upper threshold for pause duration (maximum length) could have introduced outliers. In view of this, the mean, perhaps, was not the best measure of central tendency as it is heavily affected by extreme values. Indeed, Campione and Veronis (2002) discovered that the mean failed to generate a stable measure of central tendency for their data. Rather the median proved to be a more useful measure. In line with this observation the median values were computed for the pause duration data of experiment 1 (Figures 65 and 66).

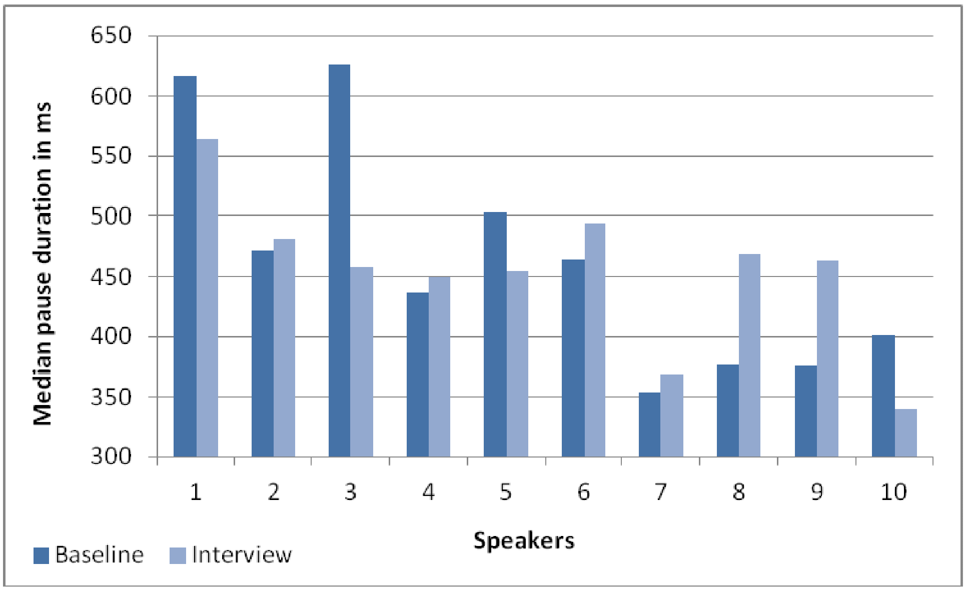

Figure 65 – Median pause duration for all speakers in baseline and interview

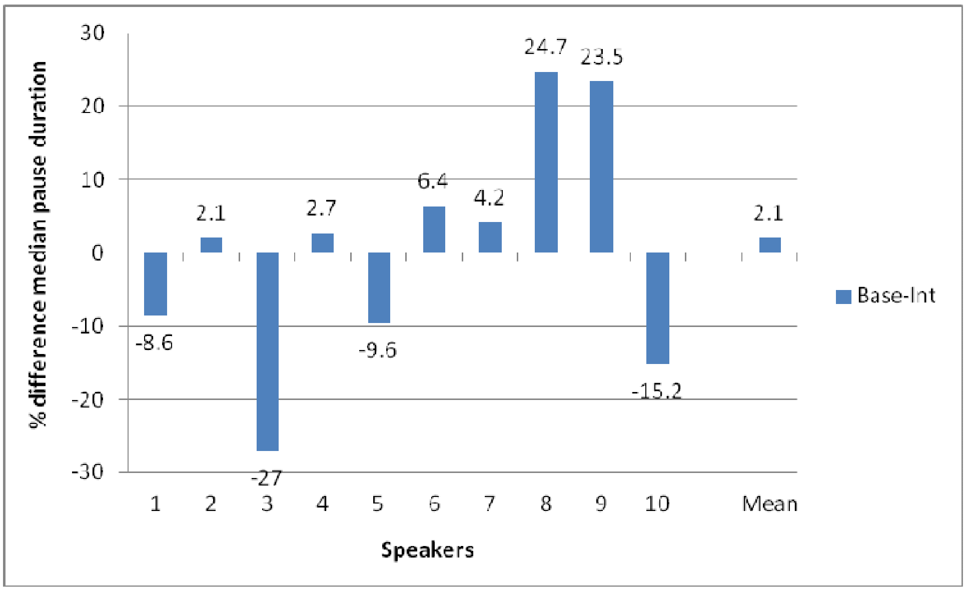

Figure 66 – % difference in median pause duration between baseline and interview for all speakers

Compared to the mean values, the median values show more variability across speakers in relation to changes in pause duration across conditions. There is still a substantial amount of speakers who use longer pauses in interview; however, the magnitude of increase is smaller. Furthermore, four as opposed to only one speaker demonstrate shorter pauses in interview. Overall, the increase in pause duration from baseline (median = 451 ms) to interview (median = 461 ms) was not significant ($T = 27$, $p \geq .05$, $r = -.01$). Figure 66 also highlights the presence of individual variability. While speaker 3 shows a marked decrease in pause duration, speakers 8 and 9 show notable increases.

It has been shown that the decrease in mean SR was related to an increase in frequency of pauses; in addition, pause length may also have affected the mean SR results. There was no significant correlation between the change in mean SR and the change in mean (r_s = .44, $p \geq .05$) and median (r_s = .30, $p \geq .05$) duration of pauses. However, the effect sizes were higher than .30 (as signalled by the r_s values) suggesting that the correlation showed a medium size effect.

As a measure of actual speed of articulation, the AR value does not incorporate pausing behaviour and is therefore independent of the former. Kendall (2009) observed an inverse relationship between pause duration and AR; as duration decreased, AR increased. He concluded that this was evidence for the “complex coordination of cognitive, physiological (i.e. articulatory and respiratory), and discourse processes” (p. 165). The correlation between change in mean AR and change in number (r_s = .32, $p > .05$) and duration (r_s = .30, $p > .05$) of pauses was not significant. However, judging by the medium effect size of r = .32 and r = .30, it appears that a linkage may have been present which would strengthen the assumption that the interview condition did affect the generation and production of speech. Did the act of deceiving generate those changes?

The comparison between the mean and median values of the pause duration data has exemplified two things. Firstly, speakers may not have produced longer pauses in the interview after all. Secondly, comparable methods of analysis are paramount in order to compare findings across different studies. Could the discrepancies in previous research on deceptive speech be explained by different investigators using diverging methods and measures of analysis?

7.2.4 Hesitations (frequency and duration)

Research on stress and cognitive load has reported an increase in hesitations but the findings in relation to hesitation phenomena in clear speech are unclear. Deceptive speech research tended to only report frequency results. As with the other temporal parameters results are diverging with some concluding an increase in hesitations and others a reduction as a function of deception.

Similar to the analysis of silent pauses, hesitations were examined with regards to frequency and duration. Again, statistical analysis was restricted to the inter-speaker level for the frequency results. Initially, hesitation markers were analysed by combining ‘um’ and ‘uh’ but following this, a detailed look will be taken at the behaviour of ‘um’ and ‘uh’ separately.

7.2.4.1 Frequency

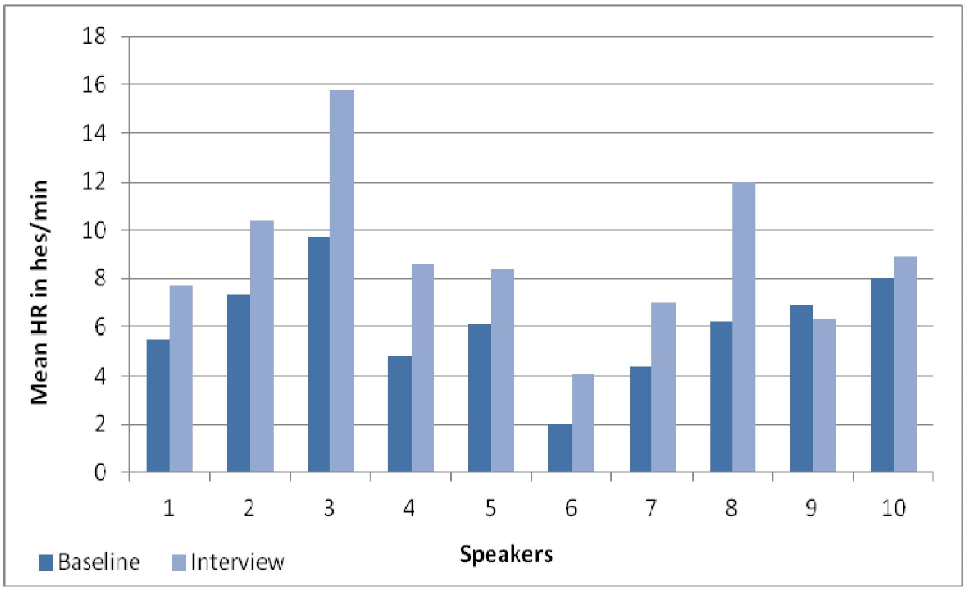

Figure 67 – Mean HR in baseline and interview for all speakers

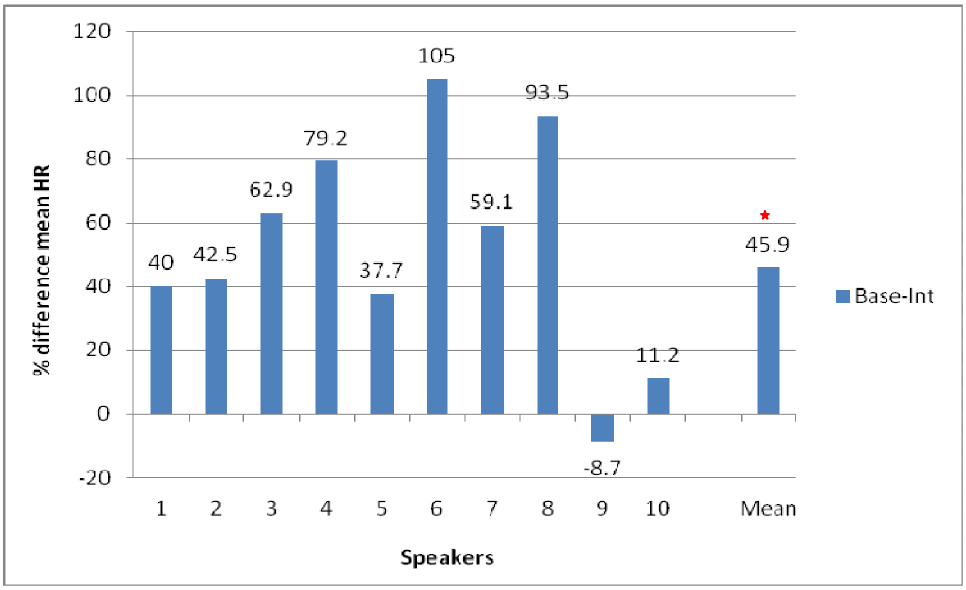

Figure 68 – % difference in mean HR between baseline and interview for all speakers

Figures 67 and 68 clearly indicate that speakers used considerably more hesitations during the interview as compared to the baseline. The increase in HR from baseline (mean = 6.1 hes/min) to interview (mean = 8.9 hes/min) was significant (T = 1, p ≤ .01, r = -.60). The magnitude of increase averaged 59%. Only one speaker showed a reduction in hesitations in the interview; this decrease merely amounted to 8.7% however.

7.2.4.2 Duration

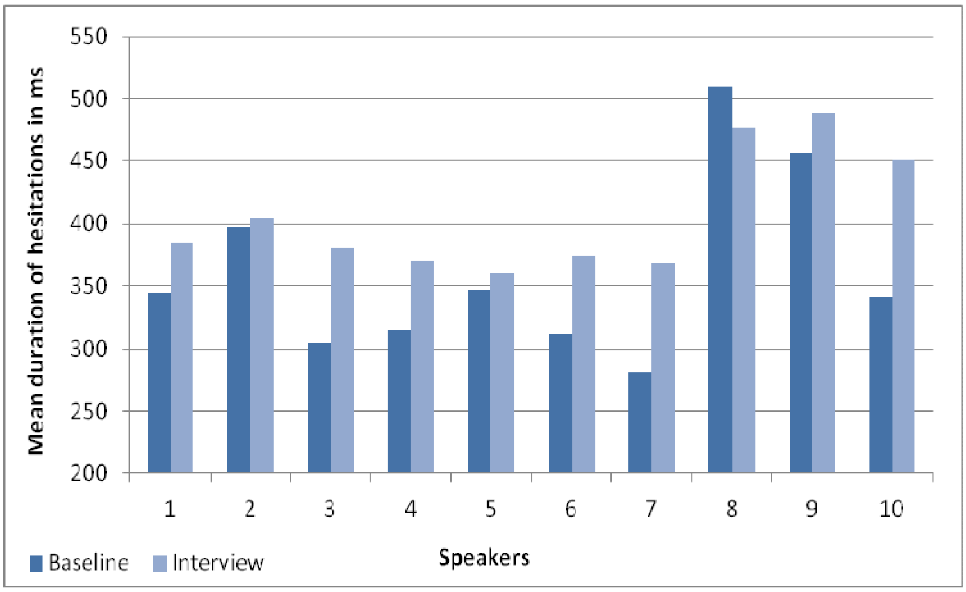

Figure 69 – Mean duration of hesitations in baseline and interview for all speakers

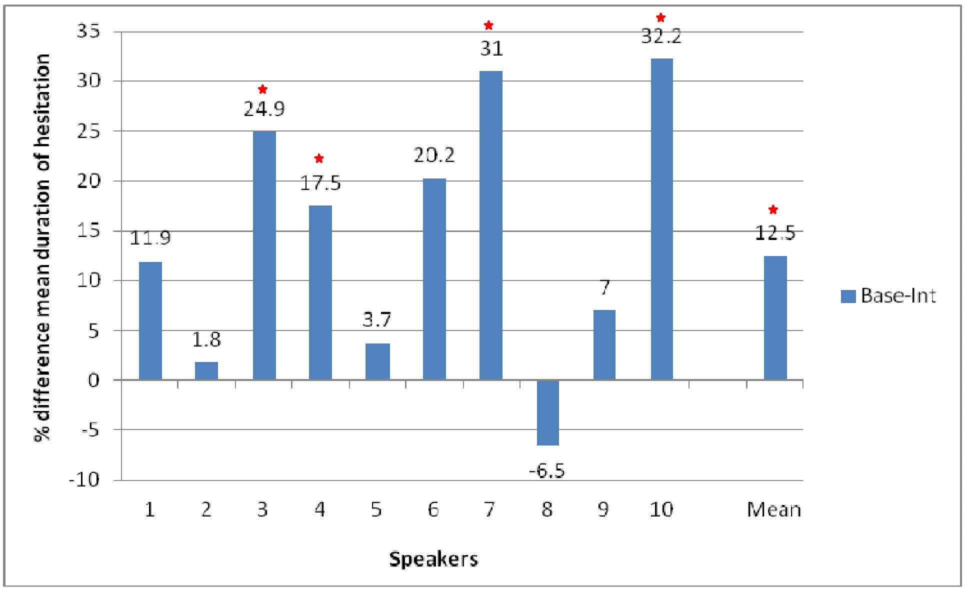

Figure 70 – % difference in mean duration of hesitation between baseline and interview for all speakers

Judging by Figures 69 and 70, hesitations were significantly longer in the interview compared to the baseline ($T = 4$, $p \leq .05$, $r = -.53$). Almost all speakers conformed to this trend and the magnitude of increase ranged from 1.8% to 32.2%. For speakers 3, 4, 7 and 10 the increase was statistically significant; however for speaker 6, who showed a higher proportional increase than speaker 4 (19.3% vs. 15.6%) the difference did not reach statistical significance. This may be due to a relatively small sample size; looking back at Figure 67 speaker 6 portrayed the lowest mean HR in both conditions (2 hes/min vs. 4.1 hes/min).

The analysis of the duration of silent pauses showed that the results were affected by the measure of central tendency used. There is no a priori reason why median should provide a better measure for the current data on duration of hesitations. Indeed, exploratory analysis of the data using the median did not reveal any significant differences to those obtained from the mean.

Mean SR may be affected by a change in hesitation phenomena. There was no correlation between change in number of hesitations ($r_s = .14$, $p \geq .05$) and change in mean SR; however, there was a strong tendency for a correlation between reduction in mean SR and increase in duration of hesitations ($r_s = -.50$, $p = .07$) as illustrated in Figure 71.

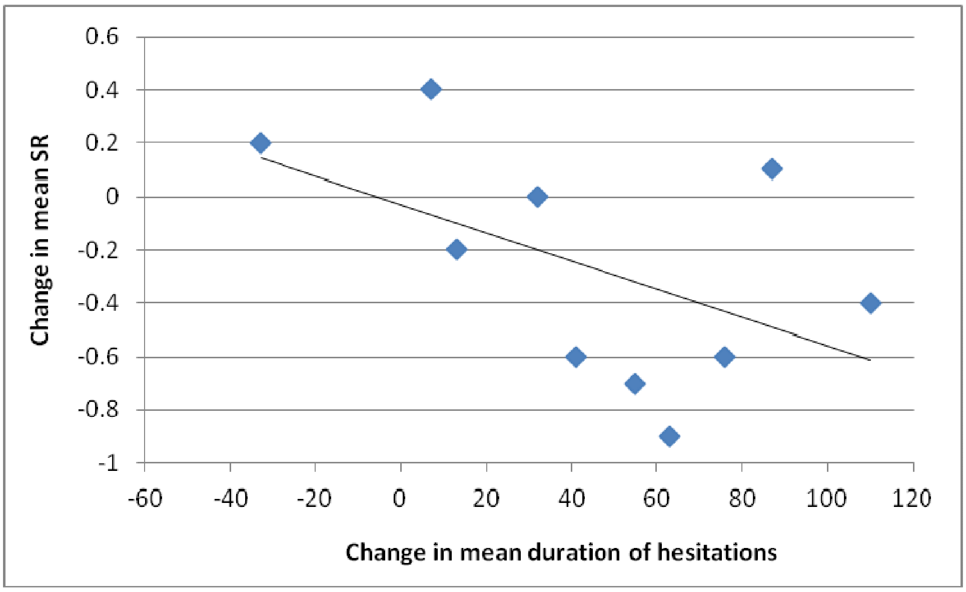

Figure 71 – Correlation between change in duration of hesitations and change in mean SR

The increase in number and duration of hesitations in the interview could be manifested in different ways. On the one hand, it could be that the duration of the individual segments of the hesitations increased e.g. longer vocalic segment for 'um' and 'uh' or, in the case of the latter, a longer nasal. On the other hand, it may be that the different types of hesitation markers increase to differing degrees. Phonetically, ‘um’ contains two phonemes- a vowel and a nasal consonant; ‘uh’ only contains one vocalic segment. Putting aside possible segment lengthening, ‘um’ hesitations can be expected to be longer than ‘uh’ hesitations. Benus et al. (2006) reported an increase in ‘um’ but not ‘uh’ when speakers were lying. Was the increase in duration of hesitations as seen in Figure 70 a result of an increase in ‘um’? By extension, did speakers use more ‘um’ than ‘uh’ when they were lying?

7.2.5 Hesitations by type ‘um’ and ‘uh’

The frequency and duration aspects of ‘um’ and ‘uh’ are expressed in exactly the same way as was done for all hesitations combined. Frequency is expressed as number of ‘um’/’uh’ per minute and statistical analyses were performed using the Wilcoxon signed rank test (inter-speaker comparisons) and independent t-tests (intra-speaker comparisons). The frequency counts for the different types of hesitations were often considerably low and, as a result, the percentage values were judged to be not helpful in illustrating and interpreting the results. Therefore, only averages are presented for the frequency data of ‘um’ and ‘uh’.

7.2.5.1 Frequency and duration of ‘um’

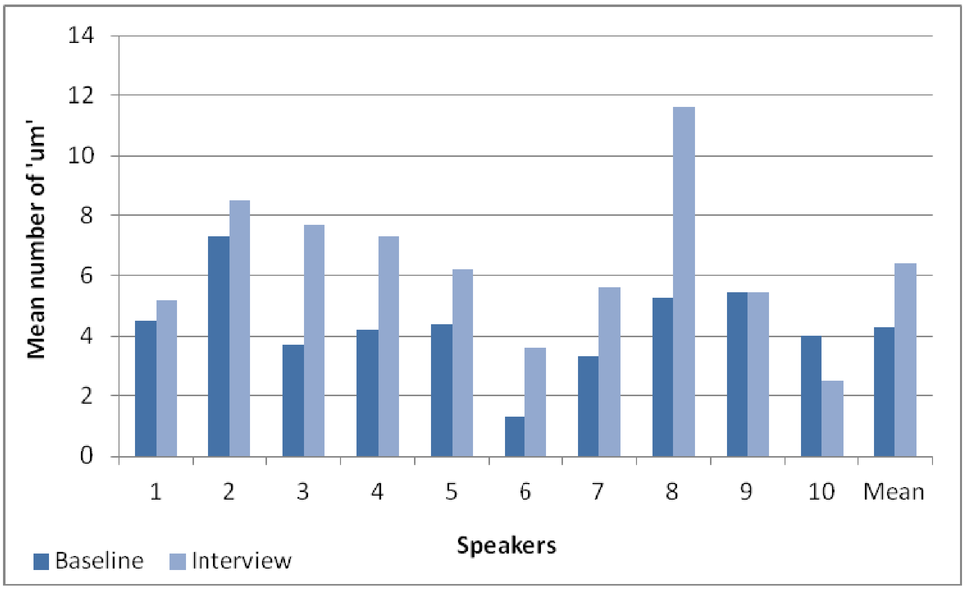

Figure 72 – Mean number of ‘um’ in baseline and interview for all speakers

The behaviour of ‘um’ mirrored that observed for all hesitations seen in section 7.2.4. Frequency of ‘um’ increased from baseline (mean = 4.3 ‘um’/min) to interview (mean = 6.4 ‘um’/min) by an average of 48.8% (Figure 72). This was statistically significant ($T = 3$, $p \leq .05$, $r = -.52$). The magnitude of increase ranged from 15.5% as seen in speaker 1 to quite noteworthy changes up to over double the baseline value as demonstrated by speakers 3, 6 and 8. Speaker 10 was the only example of a decrease in the number of ‘um’ in the interview.

As indicated by Figures 73 and 74 ‘um’ was significantly longer in the interview (mean = 405 ms) as compared to the baseline (mean = 448 ms) ($T = 7$, $p \leq .05$, $r = -.47$). This increase was almost uniform across speakers averaging 17%. For two out of the eight speakers the durational increase was significant (marked by red asterisks). Speaker 8 was the single speaker who showed a reduction in length of ‘um’. Compared to the rest, speaker 6 portrayed a minimal difference in durational values across the two conditions. One reason for this could be seen in the fact that this speaker used very few hesitations (1.3 hes/min vs. 3.6 hes/min) and therefore the scope for variability in the duration of ‘um’ tokens may have been reduced.

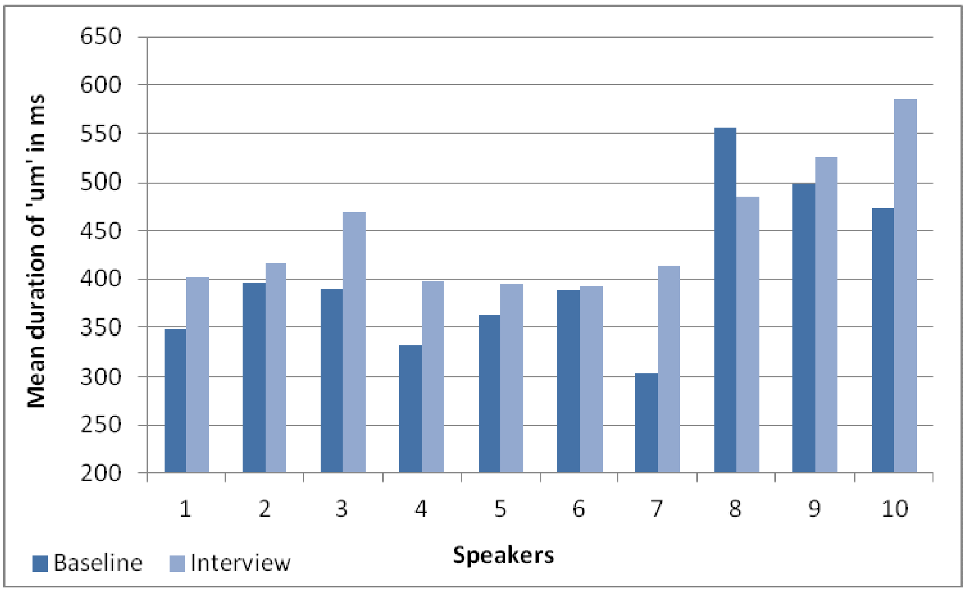

Figure 73 – Mean duration of ‘um’ in baseline and interview for all speakers

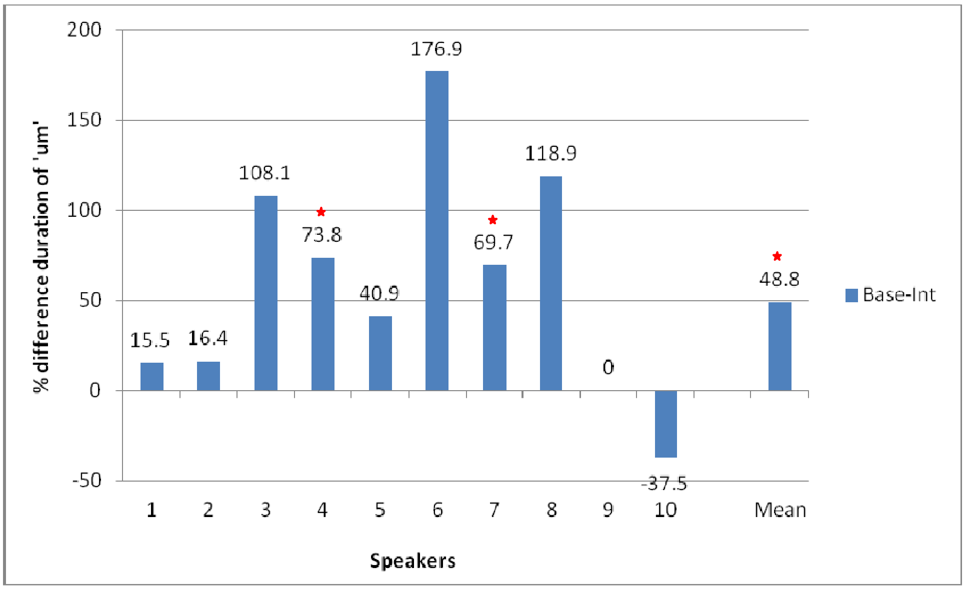

Figure 74 – % difference in mean duration of 'um' between baseline and interview for all speakers

7.2.5.2 Frequency and duration of 'uh'

Figure 75 illustrates the frequency of 'uh' expressed in rate of 'uh' per minute for both conditions.

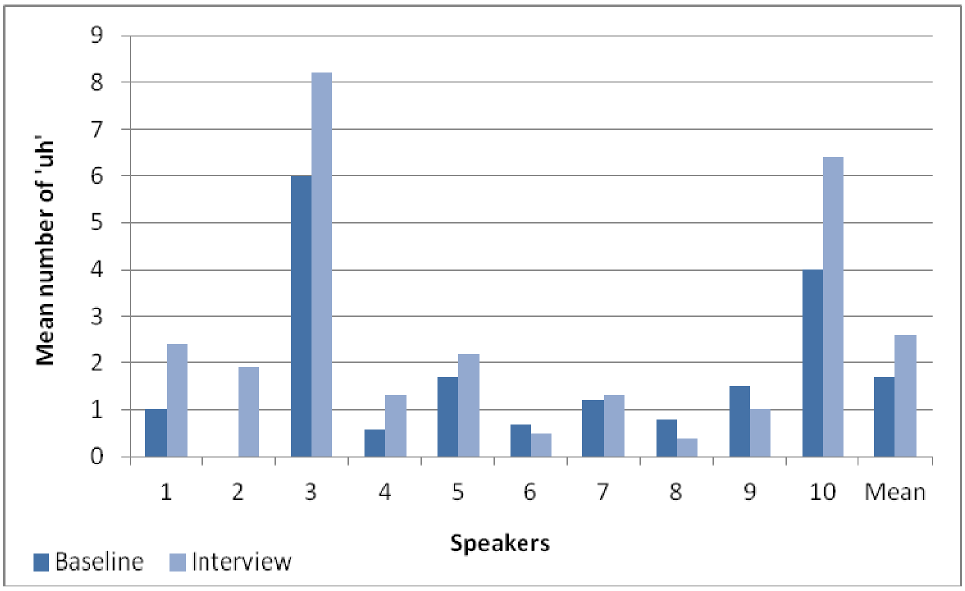

Figure 75 – Mean number of 'uh' in baseline and interview for all speakers

In general, the majority of speakers show low instances of 'uh'. When comparing the rate measures of 'um' and 'uh' illustrated in Figures 72 and 75, respectively, it becomes obvious that speakers tended to use the hesitation marker with the bilabial nasal to a greater extent than the simple vocalic version. The change in 'uh' from baseline (mean = 1.7 'uh'/min) to interview (mean = 2.6 'uh'/min) was negligible and not statistically significant ($T = 9.5$, $p \geq .05$, $r = -.41$). Speaker 3 and 10 present some evidence for the claim that speakers have personal preferences for the type of hesitation markers used.

Speaker 2 did not exemplify any instances of 'uh' in his baseline speech and therefore was not included in the display of the mean duration of 'uh' (Figures 76 and 77).

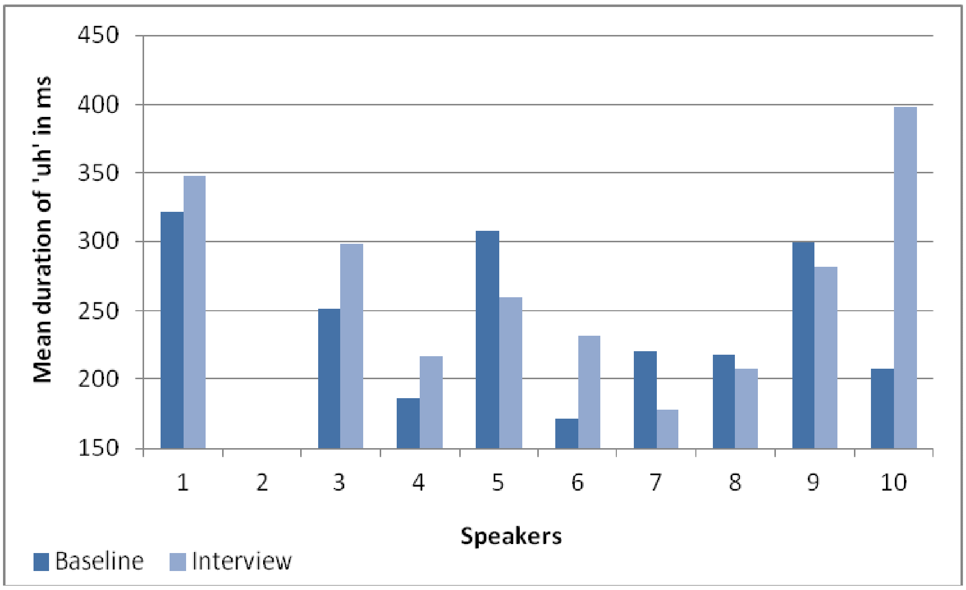

Figure 76 – Mean duration of 'uh' in baseline and interview for all speakers

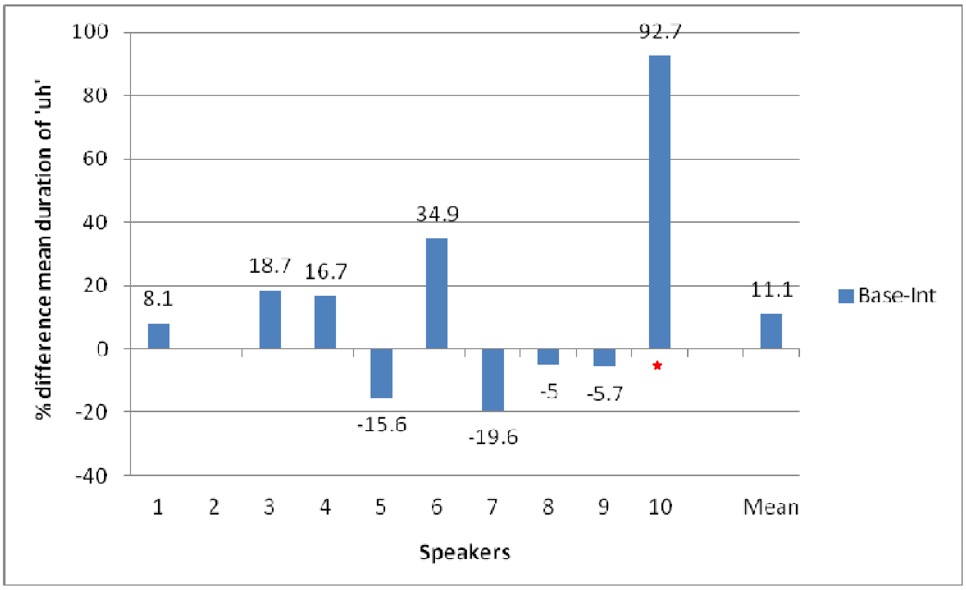

Figure 77 – % difference in duration of 'uh' between baseline and interview for all speakers

No significant difference in duration of ‘uh’ in the baseline (mean = 242 ms) compared to the interview (mean = 269 ms) was apparent ($T = 15$, $p > .05$, $r = -.20$). The change was not as consistent across speakers as that observed for the duration of ‘um’. Some speakers showed a lengthening of ‘uh’, others a reduction. Speaker 10 is the only example of a significant increase in length of ‘uh’ in the interview amounting to 192 ms.

The current results corroborate the findings of Benus et al. (2006). Overall, the increase of hesitations in the present data, which was the result of an increase in ‘um’ rather than ‘uh’, may be indicative that speakers were experiencing cognitive load or psychological stress.

7.2.6 Response onset time

Response onset time has been found to be affected by increased cognitive load; heightened demand on mental processing results in longer response latencies. A disparity in results for ROT can be seen in the research on deception with some studies reporting increases and others shorter ROT with lying.

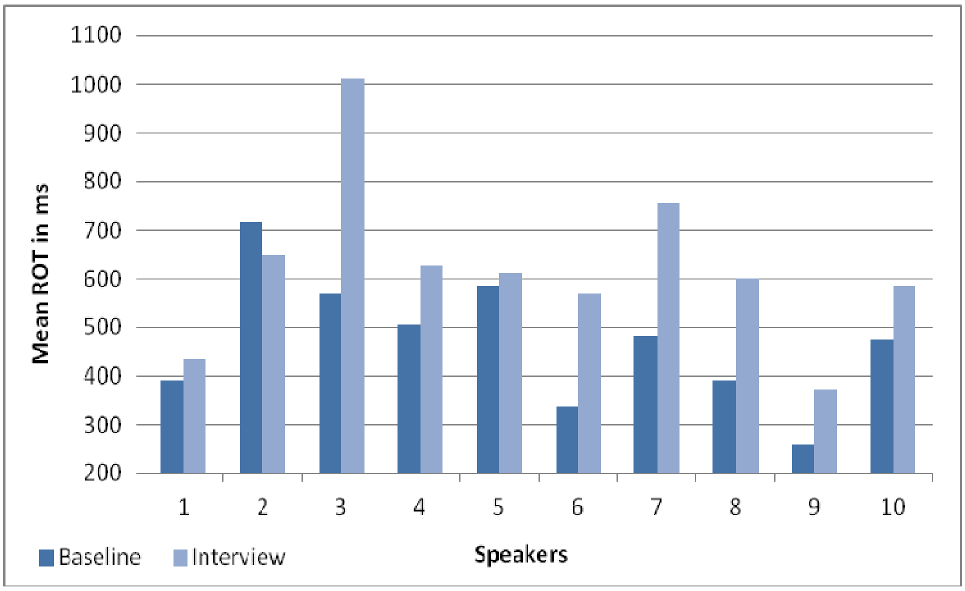

Figure 78 – Mean ROT in baseline and interview for all speakers

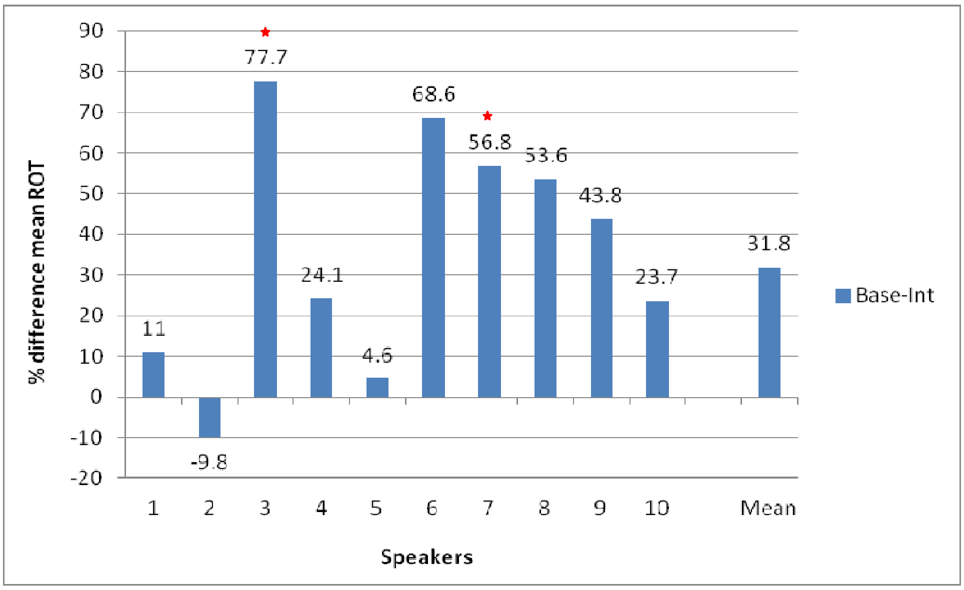

Figure 79 – % difference in mean ROT between baseline and interview for all speakers

Speakers were significantly slower to respond to the interview questions than to those presented in the baseline condition (Figures 78 and 79). The difference in ROT between baseline (mean = 472 ms) and interview (mean = 622 ms) reached statistical significance ($T = 3$, $p \leq .05$, $r = -.56$). All speakers but one speaker conform to this directional trend but the extent of the increase differs ranging from 4.6% to 77.7%. This increase is significant for speakers 3 and 7 only which is a little surprising perhaps given that speaker 6 shows a similar increase in length of mean ROT. The reason for the lack of statistical significance might be traced back to the relatively large standard deviations.

The increase in ROT is in line with the assumption that speakers were experiencing heightened cognitive load during the interview. The difference in ROT between baseline and interview amounted to 150 ms. This is less than the value reported in Walczyk et al. (2003) who compared truthful and honest responses to questions and found that deceptive responding took on average 336 ms longer than honest responding.

An interesting observation emerged when correlating the change in ROT across conditions with speakers self-reported levels of anxiety. Figure 80 demonstrates that speakers with a higher level of anxiety show larger increases in mean ROT when lying in the interview. Although the correlation was not significant ($r_s = .61$, $p = .06$), a trend was apparent.

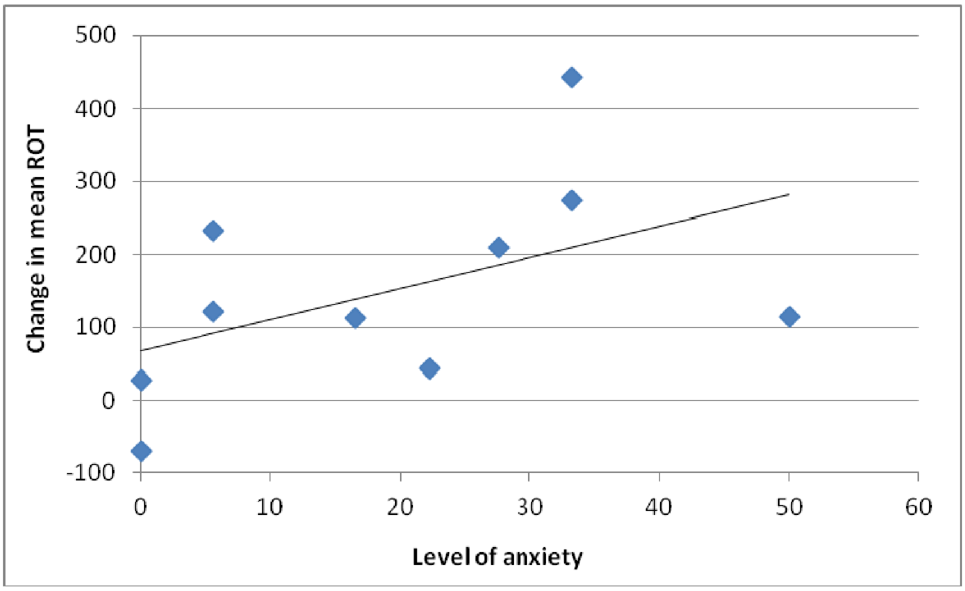

Figure 80 – Correlation between level of anxiety and change in mean ROT

7.2.7 Summary of findings from temporal analysis

A summary of the results for each of the temporal parameters investigated is presented in Table 9.

Table 9 – Summary of the findings of the temporal analysis

Parameter	Result	Significance
Speaking Rate	↘ (tendency)	ns
Articulation Rate	↘	*
Frequency of Silent Pauses	↗ (tendency)	ns
Duration of Silent Pauses	↗	(*)
Frequency of Hesitations	↗	*
Duration of Hesitations	↗	*
Response onset Time	↗	*

Statistical significance: ns = not significant, * = significant at .05 level

The findings seem promising with respect to highlighting differences between the baseline data and the interview condition. There was a significant decrease in mean AR for the deceptive speech produced during the interview. Mean SR also showed a tendency for reduction during the interview but unlike with mean AR this did not reach statistical significance. Silent pauses were significantly longer during the interview when calculating the mean but the significance disappeared when the median was used as a measure of central tendency. Although not statistically significant, there was a tendency for an increase in the frequency of silent pauses. Hesitations were significantly longer and occurred significantly more often in the Interview. When separating the hesitation data according to type of hesitation, the results confirm that there is a significant increase in number and duration of ‘um’ in interview; frequency and duration of ‘uh’ is less affected. Finally, there was a statistically significant increase in response latencies for the interview as opposed to the baseline.

7.3 Discussion of findings

The reader was presented with a thorough description of the findings of the acoustic and temporal analysis of experiment 1. Interesting issues have emerged. At points throughout the description of the results the question was raised as to whether the observations can be linked to the act of deceiving. The following attempts to address the primary question of interest: Does deception result in changes along the frequency or temporal domain of speech?

Perhaps the most striking observation relates to the fact that contrary to the temporal findings, the results for the majority of the acoustic parameters were not significant. With the exception of the f_0 parameters and vowel space area estimation, there was no consistent and significant change for mean intensity, vowel formant measurements, LTFD and H_1-H_2 across conditions. Admittedly, the change in some of the temporal parameters did not reach statistical significance; however, in these cases there were nevertheless strong directional tendencies. No such tendencies emerged for the frequency based parameters.

Admittedly, the fact that no significant changes occurred in the frequency domain of speech as a function of deception in the current experiment does not mean that such change does not happen:

1. It may be that change did occur but that the method used to quantify this change was not accurate enough. This is unlikely to be the case with the analysis of formant frequencies, as different methods were employed and all did agreed on similar

findings. In the case of H_1-H_2 however, it could be argued that the method was not refined enough as explained in section 7.1.7

2. It could be that lying does have an effect on the frequency domain of speech but that it manifests itself in parameters different to the ones measured in this study. As described in section 7.1.6, Nolan and Grigoras (2005) argued that laryngeal cues are less stable than supra-laryngeal cues and "highly sensitive to psychological state, environmental pressures on speech and paralinguistic needs" (p. 168). Perhaps a more detailed analysis of glottal parameters such as jitter, shimmer or spectral tilt, for example, would reveal differences between deceptive and non-deceptive speech.

3. The methodological design may have failed in eliciting deceptive behaviour as it would occur in real life or that the sample size of ten speakers was too small to reveal significant trends.

In any case, the current findings have important implications for the validity of any voice analysis based technologies which claim to measure the frequency components of speech in order to make inferences about whether the speaker is lying or not.

Speakers did show speech related changes when they were lying as opposed to speaking in the baseline. Some of these observable trends could be accounted for by taking into account acoustic, linguistic and methodological explanations. The increase in f_0 SD was related to the increase in mean f_0, even though minute. The decrease in mean SR in the interview condition was related to the observed tendency for number and duration of pauses to increase. Some of the changes in mean vowel formant values and overall vowel space area may have been the result of the linguistic context from which the tokens were taken from. More generally, unequal sample sizes of measurements across conditions may have also had a confounding effect. However, other changes such as the decreasing mean AR and the increase in ROT are not as easily explained by these mechanistic or methodological explanations. One can be confident that some of the observed changes between the baseline and the interview can indeed be attributed to the experimental manipulation. However, whether the act of lying itself was responsible for the increase in f_0/f_0 variability, the expansion of vowel space area, the slowing down of speaking tempo and the increase in silent pauses and hesitations is another question.

As pointed out in Chapter 2, there is no such thing as a behavioural trait that would consistently indicate that deception was taking place. Rather, it is behaviours connected to emotional, cognitive and social processes which may accompany the act of deceiving that would provide deceptive cues. Chapter 4 offered predictions (Table 1 in section 4.6) about possible changes in speech when people experience stress, cognitive load and behavioural control, all of which may be accompanying the deceptive act. For ease of reference this table is reproduced below.

Table 10 (reproduction of Table 1) – Vocal tendencies related to the three affective states discussed including psychological stress, cognitive load and hyper-control

	Arousal Theory	Cognitive Theory	Attempted Control Theory
	Psychological Stress	Cognitive Load	Hyper-articulation
Mean f_0	↗	↗	↗
f_0 variability	↗	↘	↗
Mean Intensity	↗	↗	↗
Formants F_1 and F_2	↘	?	↗
Formant F_3	?	?	?
Vowel Space Area	↘	?	↗
LTFD	?	?	?
Speaking Rate	↘	↘	↘
Articulation Rate	↗	↘	?
Silent Pauses	↘	↗	↗
Filled Pauses	↗	↗	?
Response Latency	?	↗	?
Voice Quality	creaky, tense, breathy	creaky, breathy	?

The table does not specify predictions for all the measured parameters and as emphasised in Chapter 4, the predictions are based on generalisations across different studies leaving aside some of the variation in findings.

The tendencies for increase in mean SR correspond to the predictions made with regards to all three- the effect of psychological stress, cognitive load and behavioural control. However, the results with regards to AR, pausing behaviour and vowel space area measurement are more aligned with the cognitive theory of deception and to some extent the behaviour control theory. The decrease in AR and SR, the increase in frequency and duration of pauses and hesitation, and the longer ROTs suggest that speakers may have been experiencing heightened cognitive load during the interview. Based on the theoretical concepts and empirical findings introduced in Chapters 2-5, it is conceivable that lying placed speakers under a heightened level of cognitive load. The findings also indicate that speakers may have been controlling and monitoring their behaviour when they were speaking in the interview in order to come across as sincere. Of interest is the fact that the arousal theory was not suitable for explaining the changes in speech observed with the current data. This has indirect bearing on the suitability of polygraph testing - how useful are arousal based polygraph examinations really in detecting liars?

It may very well be that speakers in experiment 1 did experience increased cognitive load and attempted to control their behaviour when lying which would have resulted in the observed changes in speech during the interview. Does this explain the whole story?

As explained in Chapter 2, lying may not just result in speaker-internal processes such as psychological stress or cognitive load alone. The social-interactional aspects in which deception occurs may also influence liars' behaviour. The present data offers some support for the argument that speakers were tailoring their speech around that of their interlocutors. Specifically, from the f_0 SD data and the AR measures it appears as though participants may have been converging their speech towards that of the interviewer. Examination of the speech of the two interviewers revealed some marked differences with regards to f_0 variability. There was a tendency that the baseline interviewer had a more monotone style of speech than the interviewer conducting the security interview who showed greater pitch movement (Figures 81 and 82). The difference in f_0 SD values was significant ($T = 4$, $p \leq .05$, $r = -.50$).

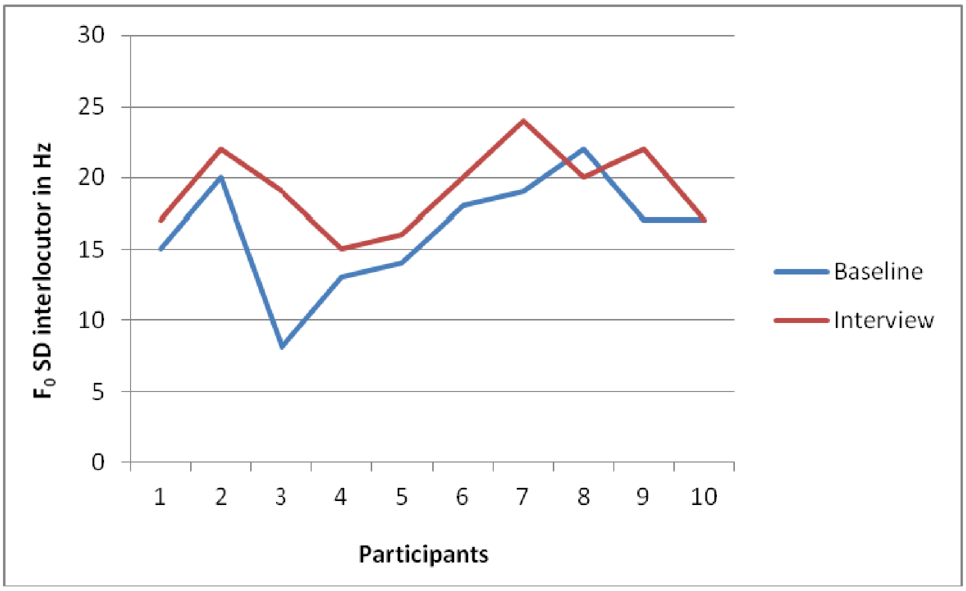

Figure 81 – Mean f_0 SD for interlocutors in baseline and interview across recordings

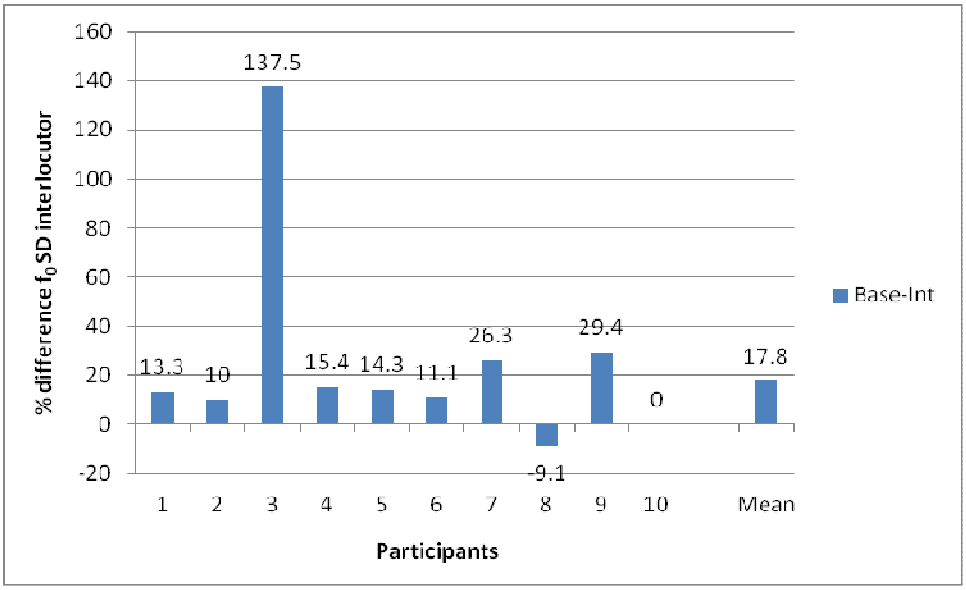

Figure 82 – % difference in f_0 SD between interlocutors in baseline and interview across recordings

Further to this, the interlocutor in the baseline condition spoke faster than the interlocutor in the interview condition (Figure 83). The difference in mean AR[11] between the two interlocutors was significant ($T = 0$, $p \leq .01$, $r = -.63$)[12] and consistent across participants (Figure 84).

[11] There was a significant difference in utterance length between the baseline data (mean = 9.9) and interview material (mean = 7.3); however, there was no significant correlation between difference in utterance length and change in mean AR ($r_s = -.26$, $p > .05$).

[12] The same methodology was applied as when calculating the AR of the participants. Phonetic syllables formed the unit of measurement and the inter-pause stretch functioned as the interval of measurement.

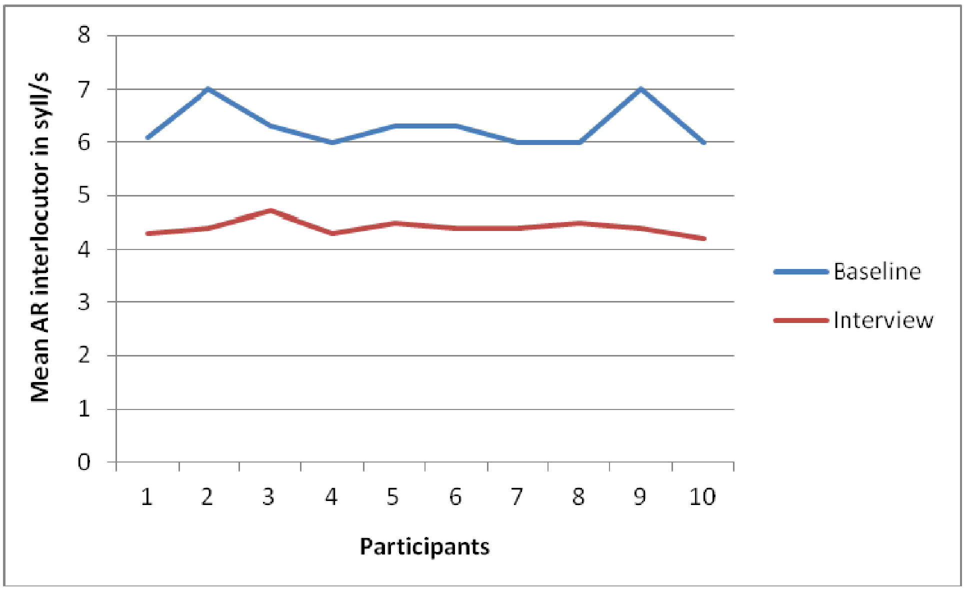

Figure 83 – Mean AR for interlocutors in baseline and interview across recordings

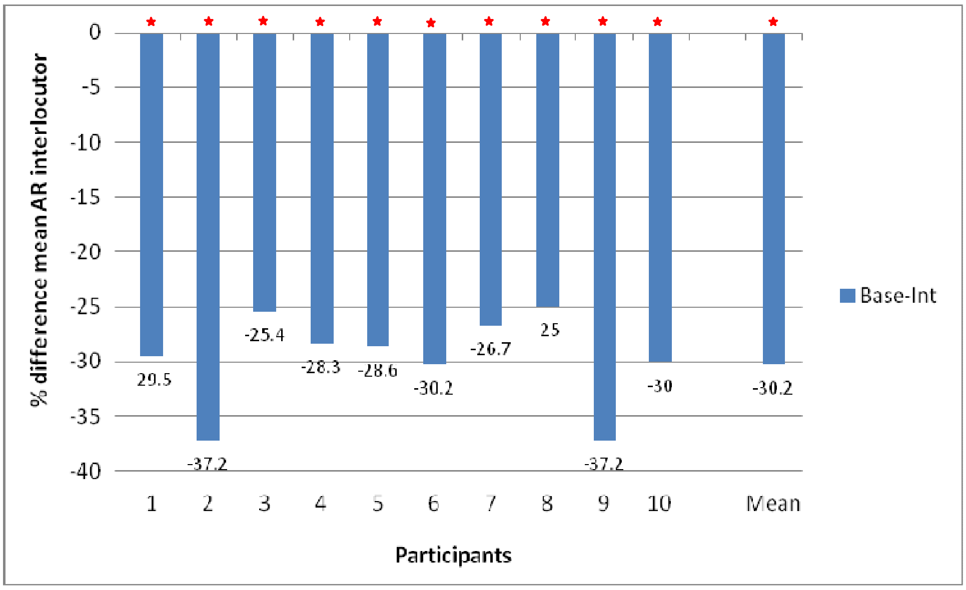

Figure 84 – % difference in mean AR between interlocutors in baseline and interview across recordings

Given that one interviewer was female and the other male, mean f_0 values will be strikingly different and would not offer insightful information and, similarly, vowel formant measurements would not have provided meaningful data in this regard. It would have been interesting however to examine the interviewers' speech with regards to vowel space area to see whether there was a marked disparity in the extent of tongue body movement. The increase in frequency and duration of pauses as well as the longer response latencies could also have been related to the addressee's speech. It would be interesting to investigate these parameters to further substantiate that convergence was indeed taking place; perhaps further investigation may have even revealed divergence. It is fair to say that in light of the f_0 SD and

AR values, it has been shown that the speech of the addressee could have had some effect on participant's speech.

Now that it has been established that speakers did appear to tailor aspects of their speech towards that of the interviewer the next step would be to find explanations for this. As described in Chapter 4, liars may have an increased self-awareness and awareness of their surroundings. They may attempt to counter a potentially image threatening situation, such as the 'security interview' of the present experiment for example, by creating a positive atmosphere and appearing cooperative. As has been pointed out in Chapter 4, conversational harmony including speech convergence may be one way of achieving cooperativeness and friendliness. Accommodating could have occurred for reasons other than deception however. Baseline and interview involved two differing communicative settings including differences in interviewers, topics and style of questioning. Regardless of the deceptive act, participants may have adopted style shifts purely as a result of these situational differences.

It could be that participants perceived the setting of the baseline speech interview as more informal and less threatening than the 'security interview' irrespective of the deceptive act. The Labovian notion of *attention paid to speech* may provide an explanation (Chapter 4). Accordingly, more formal registers are said to contain longer pauses and a slower rate of speaking. A recent study by Kendall and Wolfram (2009) illustrates the impact of different register types on temporal aspects of three local leaders.

Furthermore, the type of questions used in the interview differed to those employed in the Baseline condition so that the former largely contained a narration of previous events while the latter featured more informal and conversational speech. Setting aside the speakers' intent to deceive, it could be that the narrative process itself required a higher amount of mental activity in the first place. It may be the case then that similar changes in speaking behaviour would have occurred had the participants been telling the complete truth. As described in Chapter 6, the idea of the original experimental design included a three-way differentiation between baseline, truth and lie; however, unfortunately, there were potential flaws with the implementation of this.

Overall, the data seems to suggest that speakers were experiencing heightened cognitive load in the interview and that they were attempting conversational harmony. However, the reasons for increased cognitive involvement and speech convergence are wide-ranging and it is at this stage impossible to affirm or negate the possibility that the deceptive act was responsible for this. Moreover, when comparing individual speakers the picture was more complex and uniform trends across all speakers were rare. In order to find general patterns and tendencies the discussion has somewhat abstracted away from this observed individual variability. The reader may experience a sense of dissatisfaction with these attempts to simplify and categorise the data, to make them fit a certain straightforward model. Throughout Chapters 2-5, it was highlighted how complex and variable deceptive behaviour is and one may argue that it is exactly its complexity that is the most interesting aspect. Rather than squeezing the data into a certain framework, one should, perhaps, simply highlight its compound nature and see the scientific value in this.

Having said this, the potential significance of temporal parameters for detecting deception in speech is recognized. In order to make claims with regards to the viability of the temporal parameters in highlighting deception, it is necessary to assess more and varied data. Do the findings of experiment 2 corroborate or refute the findings of experiment 1?

8 Experiment 2

Chapter 7 has shown that deception has the potential of manifesting itself in the temporal domain of speech but alternative explanations of the observed changes could not be discarded. Additional testing with a different set of data was believed to be indispensable in order to satisfactorily attend to the aims of the present research. Chapter 8 presents the results of the temporal analysis of experiment 2. The absolute values for each speaker are attached in Appendix E.

Chapter 6 provided a detailed description of the experimental design which involved a scalable interrogation paradigm in which participants progressed from a baseline interview through two levels of interrogation (e.g. baseline, interrogation 1 and interrogation 2). One group of participants were given a token, possession of which they were asked to conceal during the two interviews. These are the liars. A different group of participants were not given a token and therefore did not have to conceal anything during the two interrogatory phases. This was the truth-telling group. As explained, the present analysis focused on the comparison of the baseline and 2nd interrogation session only.

8.1 Liars

8.1.1 Articulation rate

A decrease in mean AR was observed when speakers were lying in experiment 1.

Overall, speakers spoke with a faster rate when lying during the interrogation in experiment 2 (Figures 85 and 86).[13] This increase in AR from baseline (mean = 5.1 syll/s) to interrogation (mean = 6.0 syll/s), which amounted to 17.6%, was significant ($T = 2$, $p \leq .001$, $r = -.61$). With only one exception, this increase in rate was uniform across speakers and ranged from 1.8% to 46%. For eleven speakers the increase was statistically significant (red asterisks).

[13] Speaker 4 only produced monosyllabic yes/no answers to all but one question in the interrogation. As a consequence, a meaningful AR could not be computed for this speaker.

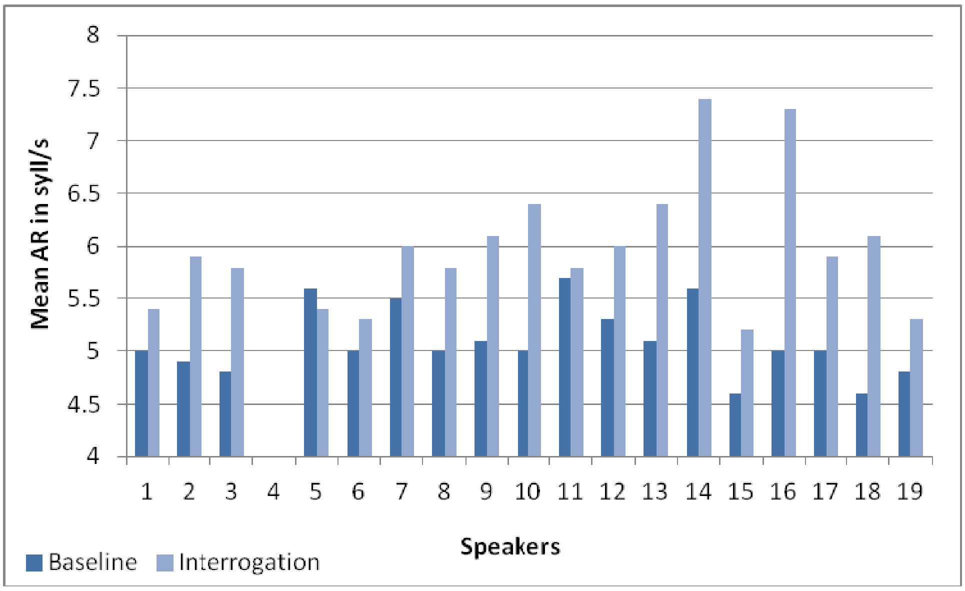

Figure 85 – Mean AR in baseline and interrogation for all liars

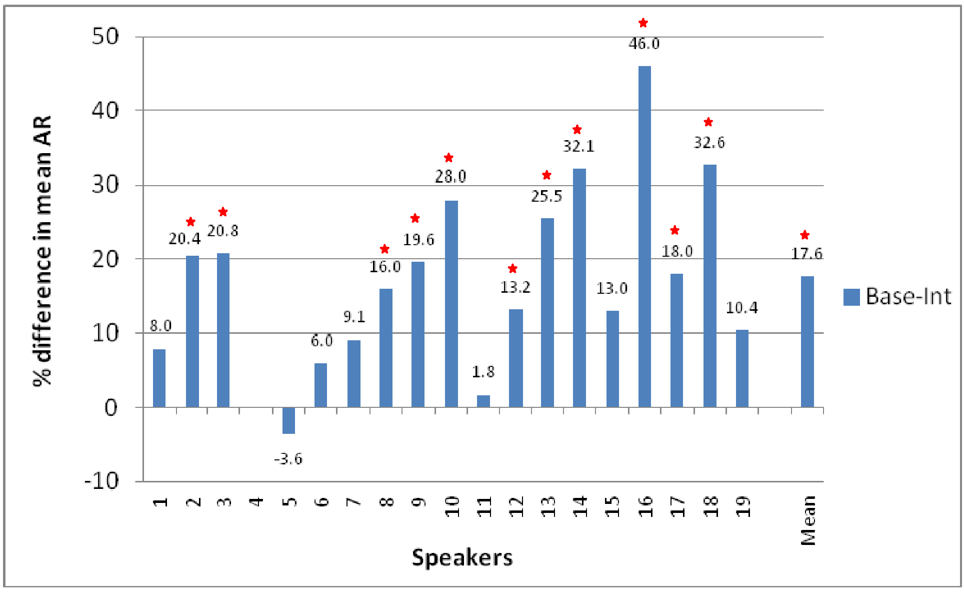

Figure 86 – % difference in mean AR between baseline and interrogation for all liars

Did the act of lying make participants speak faster? If so, then the findings are in stark contrast to those of experiment 1. In terms of measurement technique, mean AR was calculated in almost exactly the same way across the two experiments and therefore this does not provide a viable explanation for the difference. Could phrase length have played a role? On average, the baseline contained shorter phrases (mean = 4.8 syllables) than the interrogation (mean = 6.0 syllables) and this difference was significant ($T = 27$, $p \leq .01$, $r = -.42$). When correlating the difference in phrase length with the observed change in mean AR across conditions, it did transpire that there was a tendency for longer phrases to be spoken faster (Figure 87).

However, the correlation was not significant and only had a medium effect size (r_s = .33, $p \geq .05$). If the observed increase in mean AR cannot be fully explained by phrase length or measurement technique, what other possible factors are there? Perhaps, instead of experiencing cognitive load as suggested in experiment 1, liars in experiment 2 were more stressed which would be compatible with the increase in mean AR. Or maybe, credit needs to be given to contextual factors, particularly the type of questions asked.

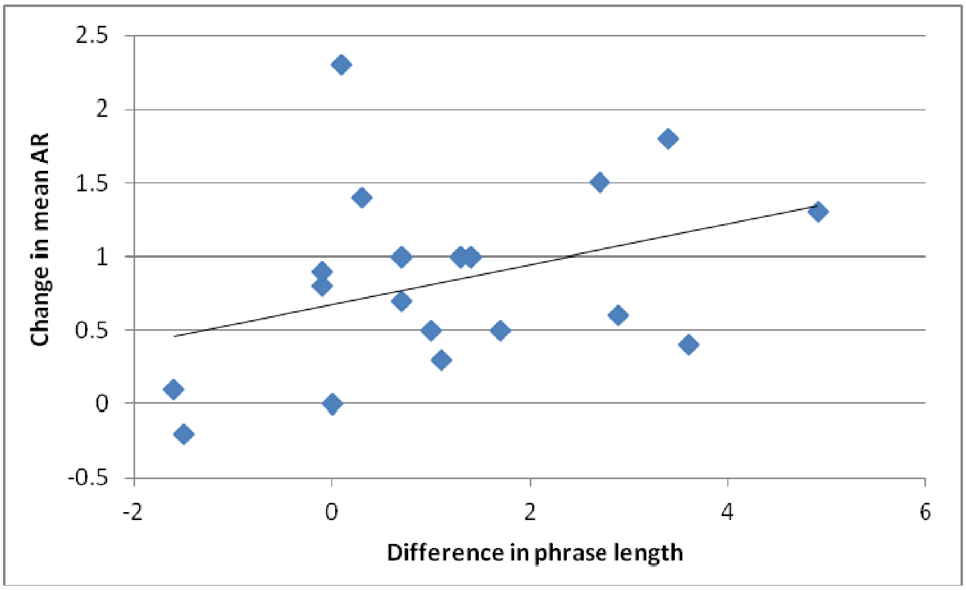

Figure 87 – Correlation between difference in phrase length and change in mean AR

8.1.2 Speaking rate

Experiment 1 resulted in a decrease in mean SR when speakers were deceiving the interrogator (Figures 88 and 89). Again, speaker 4 was excluded from the SR measurements for lack of polysyllabic responses.

Comparing Figures 85 – 89 suggests that the change in mean SR mirrors that observed for mean AR. Overall, there was a significant increase in SR from baseline (mean = 3.9 syll/s) to interrogation (mean = 5.0 syll/s) ($T = 2$, $p \leq .001$, $r = -.59$). Almost all speakers adhered to this trend but, as with mean AR, there was variability in magnitude of increase. On an intra-speaker level, the increase was statistically significant for ten speakers (red asterisks).

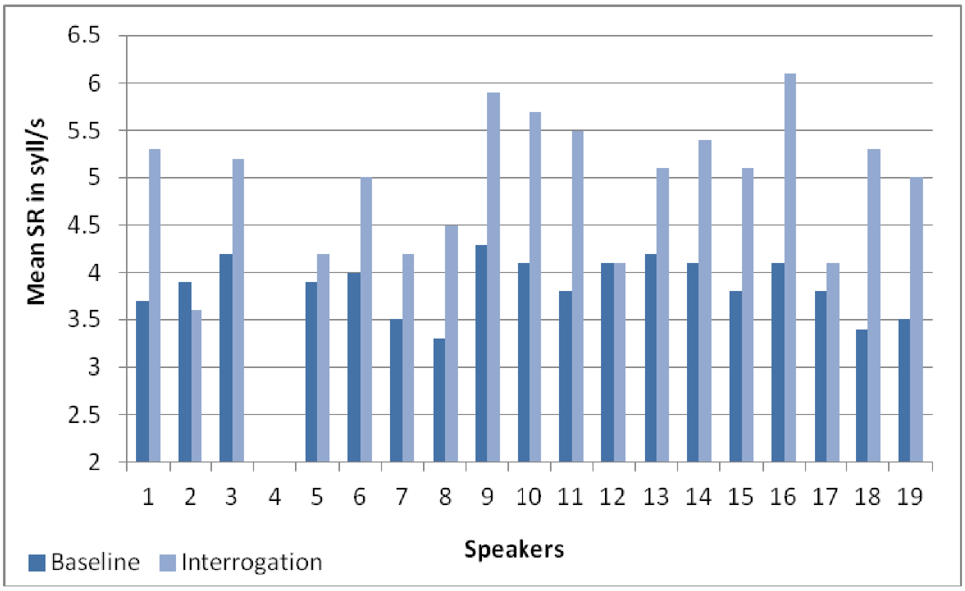

Figure 88 – Mean SR in baseline and interrogation for all liars

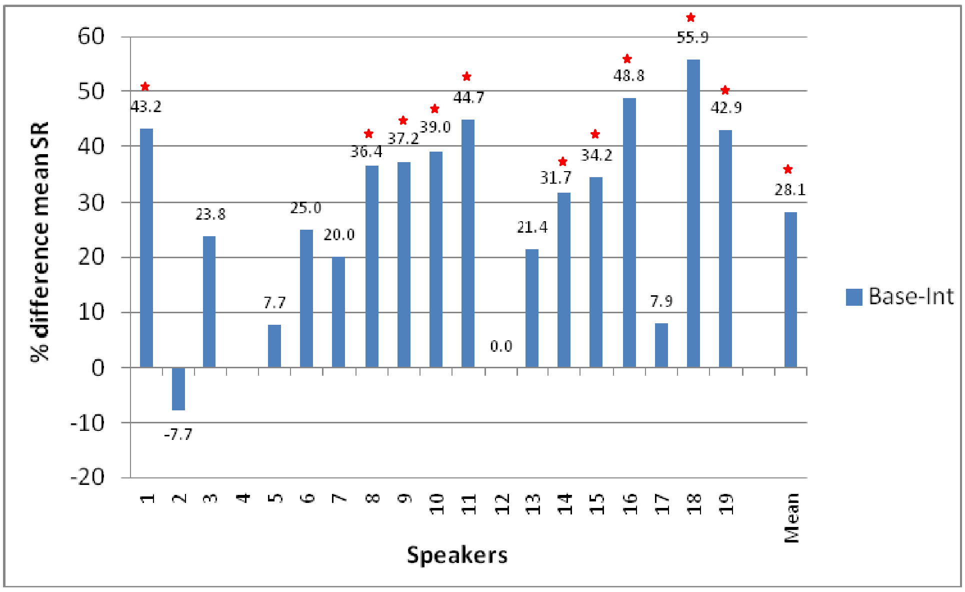

Figure 89 – % difference in mean SR between baseline and interrogation for all liars

There is nothing mechanistic to suggest that mean SR and mean AR should be correlated as they measure different aspects of speaking tempo. However, as pointed out under section 7.2.3 above, a correlation may be the result of a combination of cognitive, articulatory and discourse processes. Having said that, the correlation between change in mean SR and mean AR was weak and not statistically significant ($r_s = .26$, $p \geq .05$).

An increase in mean SR is not in line with the assumption that speakers were experiencing cognitive load or stress during the interrogation. In section 7.2.3, it was established that pausing behaviour played an important role in the calculation of AR and SR. Perhaps these results can be used to explain some of the speech rate observations observed in experiment 2.

8.1.4 Silent pauses (frequency and duration)

As was the case for experiment 1, silent pauses were analysed according to frequency of occurrence and length. Frequency was expressed using the already familiar pause rate (PR) measure and as the data only allowed for a global PR value statistical testing was restricted to the inter-speaker level. Section 7.2.3 demonstrated that the median was a potentially better measure of central tendency for pause duration than the mean. In view of this, the median was used to express pause duration with the data of experiment 2.

There was a tendency for liars in experiment 1 to employ more and longer pauses. Can the same be observed for experiment 2?

8.1.4.1 Frequency

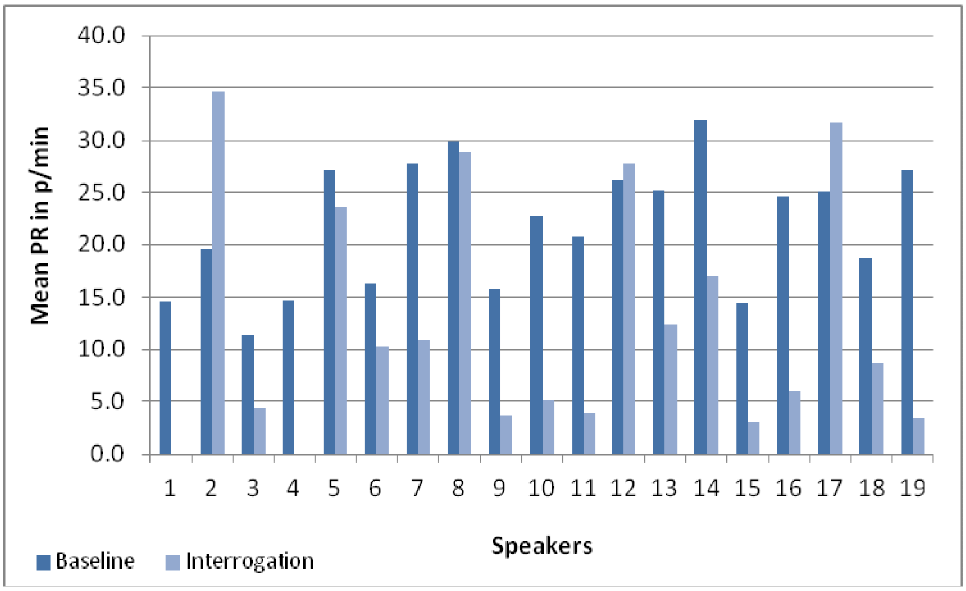

Figure 90 – Mean PR in baseline and interrogation for all liars

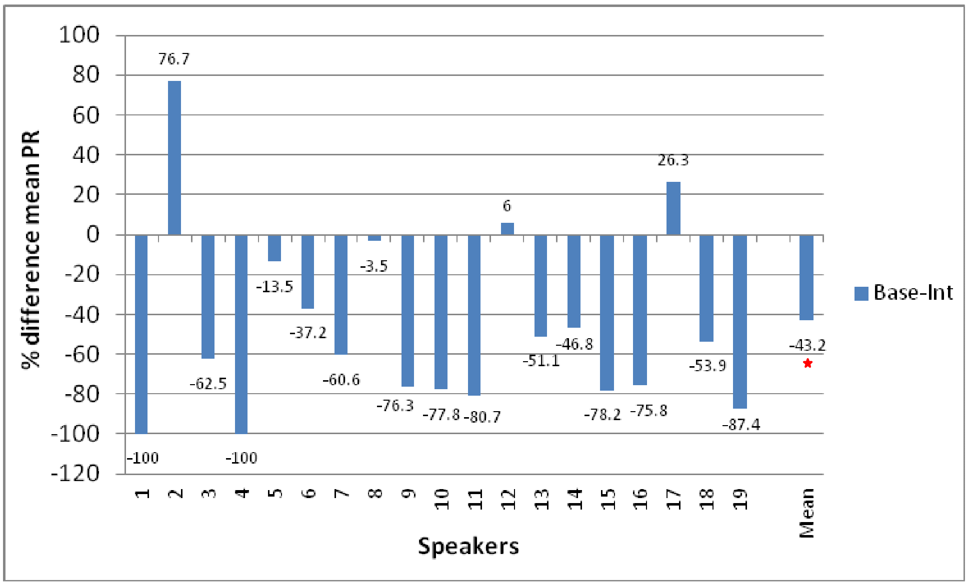

Figure 91 – % difference in mean PR between baseline and interrogation for all liars

It becomes immediately obvious when looking at Figures 90 and 91 that there was an almost uniform decrease in number of pauses from baseline (mean = 21.8 p/min) to interrogation (mean = 12.4 p/min). The decrease which averaged 43.2% was significant (T = 21, $p \leq .01$, r = -.48). The magnitude of decrease differed with values ranging from 3.5% to 100%, i.e. small decrease vs. no use of pauses at all in the interrogation. Speakers 1 and 4 provide examples of the latter.

As has been established, pausing behaviour is linked to SR in that a decrease in pausing is expected to result in an increase in SR and vice-versa. Figure 92 illustrates that, for the current data, the relationship between the decrease in pausing and the increase in mean SR is strong and highly significant ($r_s = -.68$, $p \leq .001$).

The interaction between pausing and SR is demonstrated on an individual level by speaker 2. According to Figure 89 this speaker was the only participant who showed a decrease in mean SR in the interrogation and Figure 91 shows that the number of pauses increased in interrogation for this speaker. Speaker 4 did not produce any pauses when interrogated. This corresponds to the fact the repertoire of this speaker's answers mainly revolved around either 'yes' or 'no'.

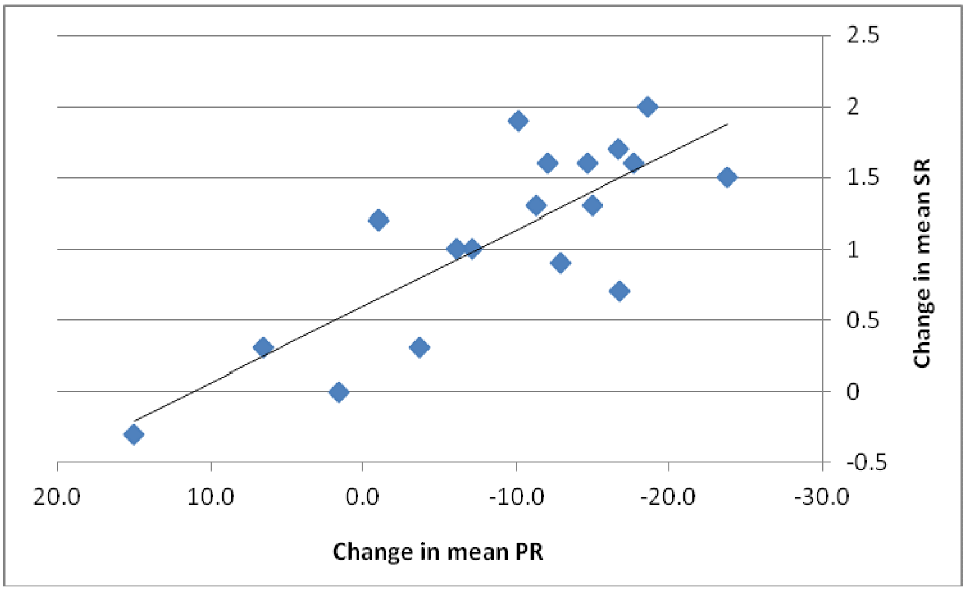

Figure 92 – Correlation between change in mean PR and change in mean SR

8.1.4.2 Duration

As illustrated in Figure 90, speakers 1 and 4 did not use any pauses in the interrogation. Furthermore, some speakers only employed one or two pauses during the interrogative part. A sample size of one or two was judged to be misrepresentative and therefore speakers who showed less than three pauses in either the baseline or interrogation were not included in the pause duration calculations. In total, ten speakers qualified for pause duration measures (Figures 93 and 94).

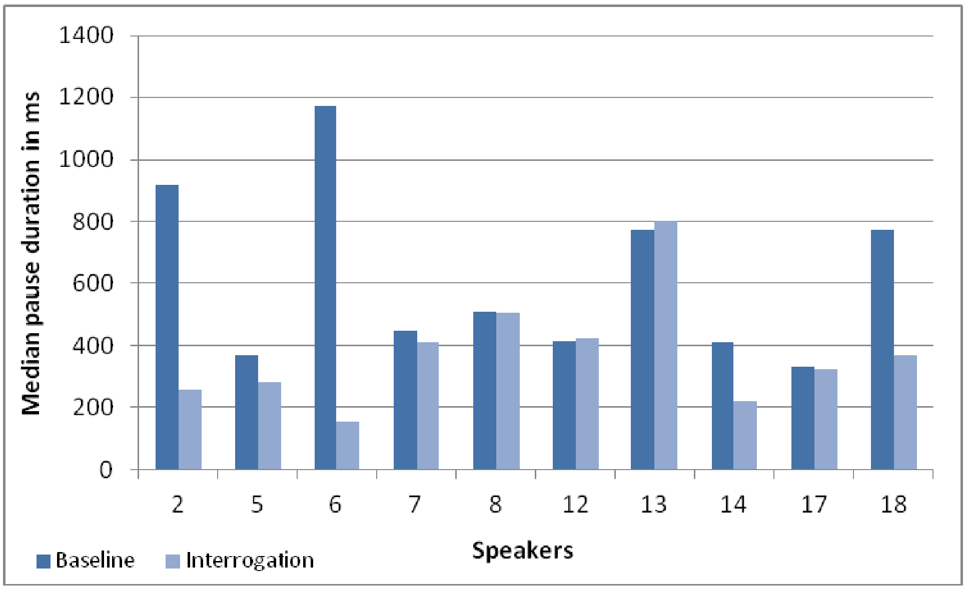

Figure 93 – Median pause duration in baseline and interrogation for ten liars

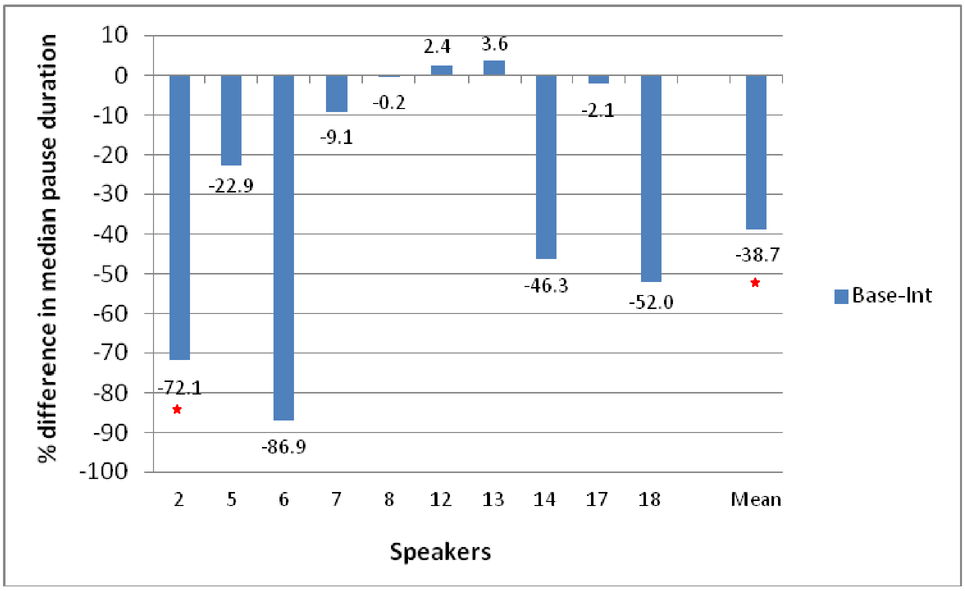

Figure 94 – % difference in median pause duration between baseline and interrogation for ten liars

Overall, speakers showed shorter pauses in the interrogation (mean = 375 ms) than the baseline (mean = 612 ms). The average decrease of 38.7% was significant (T = 7, $p \leq .05$, r = -.47). There was slightly more variability across speakers than observed so far. While the majority show shorter pauses in the interrogation, two speakers show slightly longer pause durations and another two do not vary notably in their length of pausing across conditions. Only for speaker 2 was the decrease in duration significant. The remarkably high SD values are the likely cause for the non-significance shown by the remaining speakers. On average, the SD values of the pause durations were 915 ms in the baseline and 841 ms in the interrogation. The reason for these variations in pause duration is likely to be connected to the nature of the 'communicative interaction'. This point will be taken up again in the following discussion.

There was no significant correlation between duration of pauses and mean SR (r_s = -.16, $r \geq .05$) which suggests that the decrease in mean SR in interrogation was more strongly influenced by the decrease in number of pauses rather than the reduction in length of pauses. Similarly, the correlation between duration of pauses and mean AR was weak and not significant (r_s = .02, $p \geq .05$).

Once more, it appears that liars in experiment 2 behaved differently to those in experiment 1. The decrease in pausing correlates with the observed increase in mean SR. It may also relate to the fact that speakers spoke with a faster AR when lying. Of note are speakers' average pause duration values in the baseline. Compared to experiment 1, the values tended to be considerably higher. Speakers 2 and 6, in particular, show remarkably long pausing behaviour.

Could the context of experiment 2, i.e. the style of questioning as well as the type of questions asked be responsible for some of these findings?

8.1.5 Hesitations (frequency and duration)

Analogous to silent pauses, hesitations were analysed with regards to frequency of occurrence, as conveyed by the hesitations per minute (HR) measurement, and duration. As with silent pauses, the calculation of HR resulted in only one global value per speaker per condition limiting statistical testing to the inter-speaker level. Not every speaker employed hesitation markers in all conditions and even if they were used, the number tended to be very low at times. For the durational measurements only those speakers who offered at least three hesitations per condition were considered.

Experiment 1 demonstrated that speakers were using more and longer hesitations when lying during the interview. In view of the findings so far, the opposite may be expected for the liars in experiment 2.

8.1.5.1 Frequency

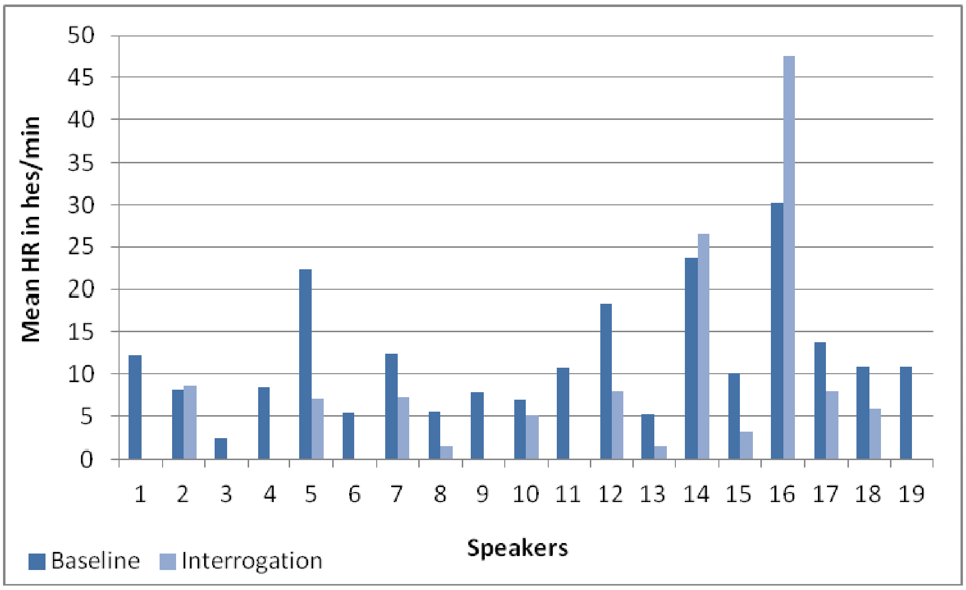

Figure 95 – Mean HR in baseline and interrogation for all liars

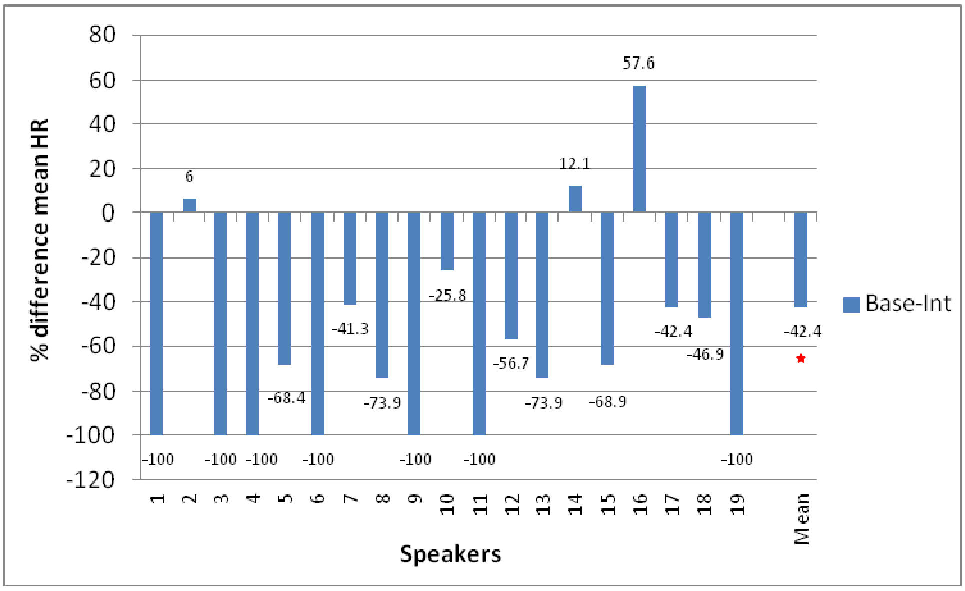

Figure 96 – % difference in mean HR between baseline and interrogation for all liars

Indeed, the findings in relation to hesitation phenomena correspond to the impression gained so far that liars in experiment 2 were quicker and less hesitant in their speech (Figures 95 and 96). There was a significant decrease in number of hesitations from baseline (mean = 11.9 hes/min) to interrogation (mean = 6.8 hes/min) ($T = 24$, $p \leq .01$, $r = -.46$). The overwhelming majority of speakers followed this trend, only three speakers showing the opposite (i.e. an increase in hesitations). The magnitude of decrease varied but for seven of the speakers the reduction resulted in zero number of hesitations in the interrogation. For the three speakers who showed an increase in mean HR in interrogation, this was only really noteworthy for speaker 16.

Can the increase in mean SR be connected to the decrease in number of hesitations? The correlation was weak and not significant (r_s =-.19, $p \geq .05$). However, when removing speaker 16, who appeared to be an outlier, a medium effect size was reached (r_s = -.41, $p \geq .05$). This suggests that the decrease in hesitation may have had some contribution towards the decrease in mean SR.

As an interesting side point, it can be seen from comparing the baseline values across speakers, that there is variability with regards to the number of hesitations speakers produce with some employing a rather substantial amount yet others showing relatively few. This highlights the importance of employing a within-speaker design when analysing speech.

8.1.5.2 Duration

Only seven of the nineteen speakers showed more than three hesitation markers in both baseline and interrogation.

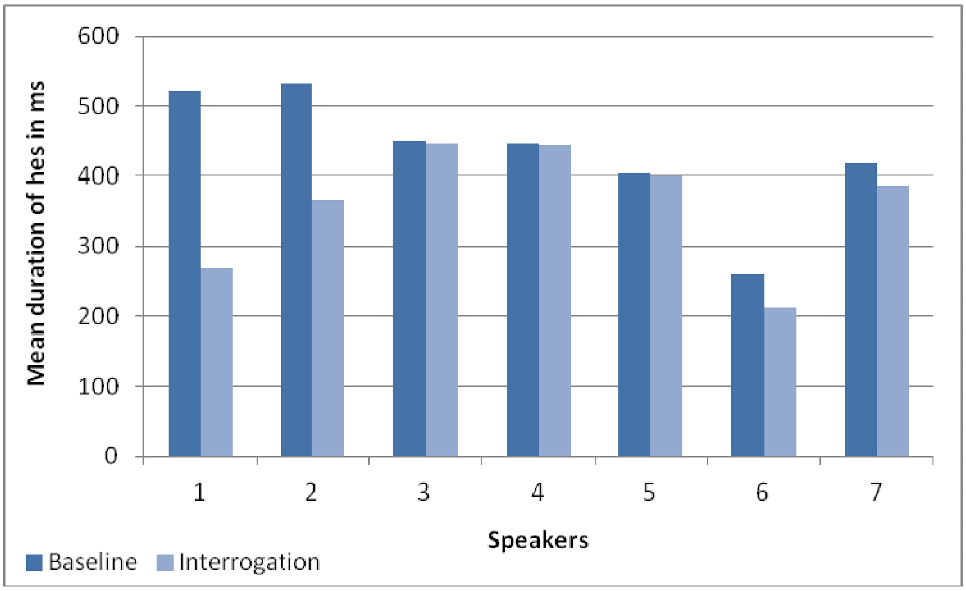

Figure 97 – Mean duration of hesitations in baseline and interrogation for seven liars

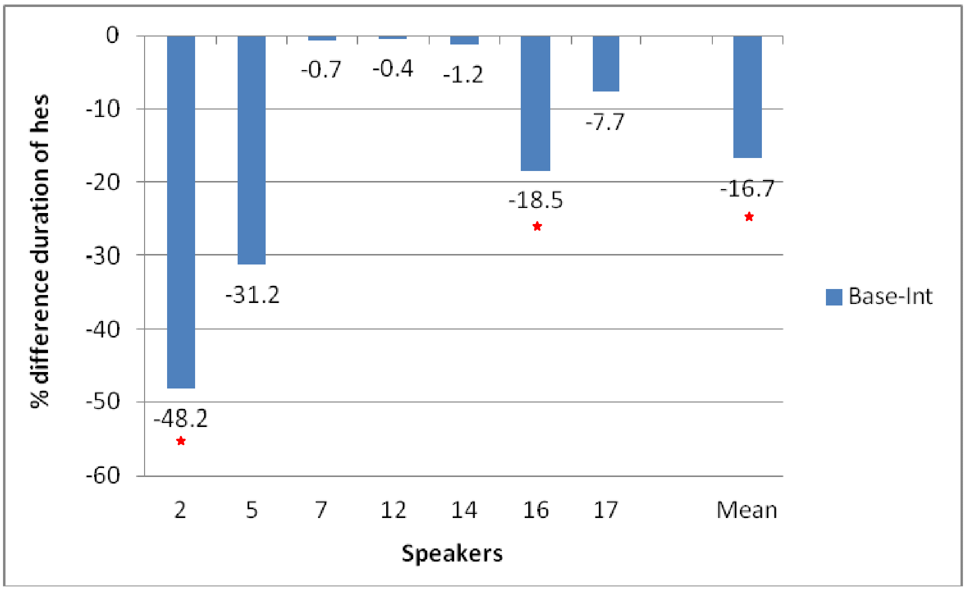

Figure 98 – % difference in mean duration of hesitation between baseline and interrogation for seven liars

Figures 97 and 98 indicate that there was a reduction in the duration of hesitations from baseline (mean = 433 ms) to interrogation (mean = 360 ms). The average decrease of 16.7% was significant ($T = 0$, $p \leq .05$, $r = -.63$). The decrease was only really noticeable for about half of the speakers. The demonstrated decreases from speakers 7, 12, and 14 were negligible. For two of the speakers the decrease was statistically significant.

To summarise, while all nineteen speakers used hesitations in the baseline condition, only twelve provided hesitation tokens in the interrogation which suggests that there was a general reduction in hesitations when speakers were lying. Generalisations in relation to how duration of hesitations was affected are more limited owing to the small sample size. What is clear however is that speakers did not increase their durations of 'um' and 'uh' in the interrogation. This decrease in hesitation is difficult to combine with either of the three behavioural states posited, i.e. cognitive load, stress, or clear speech; a clear indication that factors other than speaker-internal processes are likely to be at work.

8.1.5.3 Hesitation by type and location

In Chapter 7, it was highlighted that for various reasons, it may be of interest to distinguish different types of hesitation markers and analyse them separately. In experiment 1, a distinction was made between 'um' and 'uh' and this was also implemented with experiment 2. As well as differentiating between hesitations according to their phonetic composition, a further demarcation can be seen in distinguishing hesitations based on their position within the syntactic phrase. Grosjean and Collins (1979) discovered that syntactic nature of the pause location has an effect on pause duration and frequency. In view of this, a further differentiation was made taking into consideration where in the syntactic phrase the hesitation marker occurred. Specifically, a division was made between 'turn-initial', i.e. those hesitations which occurred at the beginning of a speaker's turn and 'turn-internal', i.e. hesitations that happened during the speaker's response.

Given the nature of the data, hesitation phenomena were scarce in the interrogation and, at times, even the baseline. This severely limited any generalisation and, therefore, it was decided to only present a summary finding of the initial analysis. When examining the frequency of 'um' and 'uh', it appeared that there was a reduction in both types of hesitations during the interrogation when measured against the baseline. As could be observed in experiment 1, there seemed to be a preponderance of 'um' over 'uh'; individual variability was nevertheless present still with some speakers preferring the nasal variant 'um' and others the vocalic marker 'uh'. Duration of 'um' and 'uh' was not significantly affected by deception and showed a high amount of inter-speaker variability. The results in relation to syntactic location were only exploratory in nature. Analogous to the general pattern of a reduction, there appeared to be fewer turn-initial and turn-internal pauses in the interrogation. However, it became clear that the nature of the responses elicited, in both baseline and interrogation, limited the opportunity for speakers to realise turn-internal pauses and therefore the current data may not have been the most adequate for this type of differentiation.

8.1.6 Response onset time

An increase in mean ROT with deception could be observed in experiment 1.

Although the decrease in ROT from baseline (mean = 2302 ms) to interrogation (mean = 1967 ms) was not statistically significant ($T = 55$, $p \geq .05$, $r = -.26$), a general tendency for a reduction could be observed (Figures 99 and 100). So far, the direction of change across speakers tended to be uniform for the majority of parameters analysed as part of experiment 2. The mean ROT values illustrate more variability between individual speakers with about two thirds showing shorter ROTs and one third of the group producing longer response onsets. For three speakers this change was significant. The reason that many of the intra-speaker comparisons did not reach statistical significance despite relatively substantial changes in mean ROT could once again be related to excessive variability in ROT values. The ROT SD values averaged 1567 ms in the baseline and 1181 ms in the interrogation.

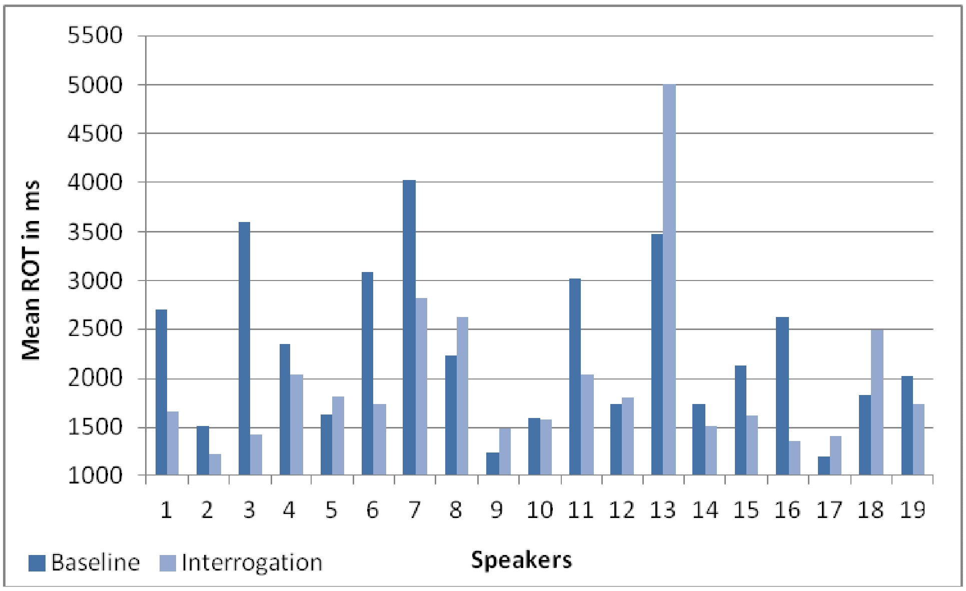

Figure 99 – Mean ROT between baseline and interrogation for all liars

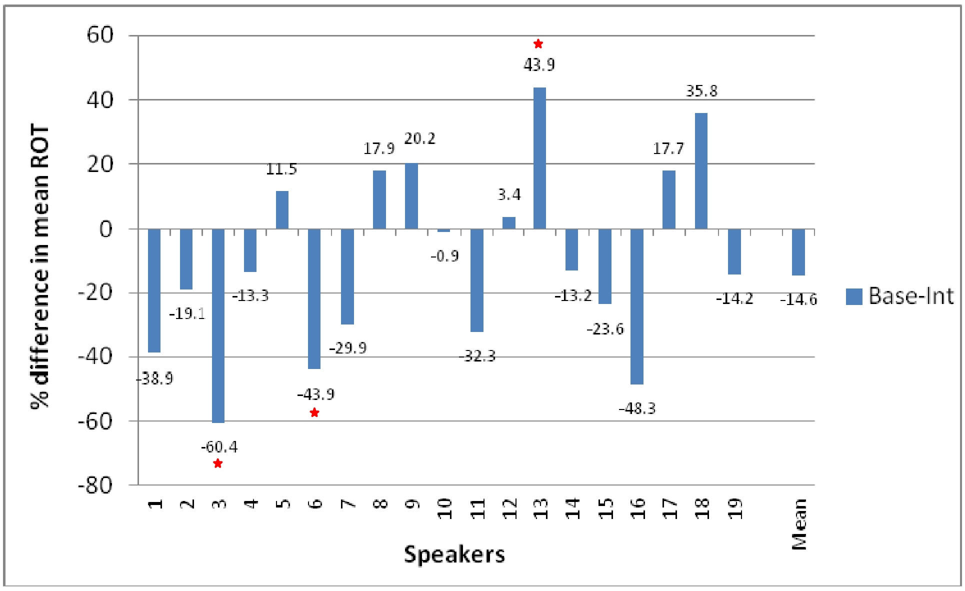

Figure 100 – % difference in mean ROT between baseline and interrogation for all liars

The decrease in mean ROT corresponds to the general impression of a speeding up of tempo when speakers were lying in experiment 2. Taking a look at Figure 99 it can be noticed that the average ROT values are extremely high with several speakers reaching 2 seconds and over. Compared to the average values of experiment 1 (around 500 ms), this is a marked difference in length. The longer ROTs correspond to the unusually long silent pauses described in section 8.1.4. What gave rise to these exceptionally long durations? As already stated, an answer may be found when considering the nature of the questioning style used.

8.1.8 Summary and discussion

A summary of the results is presented in Table 11. A significant increase in mean AR and mean SR was observed for the interrogation when contrasted with the baseline. Speakers used less silent pauses and hesitations when answering the interrogator as compared to the baseline and there was a tendency for liars to respond quicker when interrogated as seen in the tendency for mean ROT to decrease. Although there was a trend for pauses and hesitations to be shorter during the interrogation, this was subject to variability across speakers and the relatively small sample sizes limit the scope for generalizing. Dividing hesitations further down by taking into account phonetic and syntactic form revealed that while frequencies were significantly lower in the interrogation when evaluated against the baseline, durations did not alter significantly.

Table 11 – Summary of findings of temporal analysis of the liars in experiment 2

Parameter	Result	Significance
Articulation Rate	↗	*
Speaking Rate	↗	*
Frequency of Silent Pauses	↘	*
Duration of Silent Pauses	↘ (tendency)	*
Frequency of Hesitations	↘	*
Duration of Hesitations	↘ (tendency)	*
Response Onset Time	↘ (tendency)	ns

Statistical significance: ns = not significant, * = significant at .01 level

As already mentioned, unfortunately the data from experiment 2 did not allow for the analysis of frequency based parameters such as f_0 and vowel formants. Therefore comparison across experiments was limited to the temporal parameters only. The findings of experiment 2 in relation to the liars are intriguing and offer a multitude of points for discussion. In addition to giving further thought to the question of whether deception can result in consistent changes in speech (research questions 1 and 2) the following discussion highlights the importance of context, in particular the style of interviewing, when analyzing deceptive behaviour, thereby addressing research question 3.

The results of experiment 1 demonstrated a slowing down of tempo when being deceptive. It was concluded that it was possible that lying resulted in increased cognitive load which in turn manifested itself in the temporal domain of speech. However, other possible factors which could have triggered the slowing down in speech could not be discarded. As pointed out in Chapter 6, experiment 2 attempted to control some of the confounding factors present in experiment 1 in order to better test the merit of speech analysis as a way of identifying deception. Speaker external factors such as difference in interviewer and type of questions asked were minimised. The results of experiment 2 do not correspond to those of experiment 1. The increase in mean AR and SR, the reduction in number of pauses and hesitations, the shorter ROTs, all this indicates a general acceleration of speaking tempo when being deceptive and does not support the predictions made by the cognitive theory. Why did liars speak faster and show less hesitancy in experiment 2?

At first sight, the mean AR results and the decrease in pausing seem to partly corroborate the arousal theory. Given that the interrogation was informed by a coercive style of questioning it may very well be that speakers experienced an increased sense of physiological arousal and psychological stress. Physiological and psychological measurements do offer some verification for this. As is detailed in Eachus et al. (2013), body temperature significantly increased between baseline and interrogation and participants showed higher self-reported levels of stress during the interrogatory phase (mean = 10.9) as opposed to the baseline (mean = 9.1). On the other hand however, the results from the remaining physiological measures do not agree with the proposition that the speakers were more stressed during the interrogation. Heart rate remained relatively stable across the two conditions and, perhaps more intriguingly, respiration rate decreased in the interrogation as compared to the baseline. Anecdotal evidence indeed showed that participants were consciously trying to control their breathing and heart rate during the interrogation session. This would suggest that speakers were aiming to control their appearance when lying. As well as controlling their physiology, it is conceivable that liars were attempting to monitor behavioural aspects including speech. These physiological findings in connection with the decrease in hesitations, the increase in silent pauses and overall speech tempo give some substance to the attempted control theory (Table 1, reproduced as Table 10) Vrij and Heaven (1999) also 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.

In the current case, explanations based on the attempted control theory seem plausible; however, it is believed that purely 'behavioural control' is insufficient in fully explaining the observations. Two reasons can be offered that would encourage further interpretative work. Firstly, as pointed out in Chapter 4, more empirical research is needed to assess the relationship between behavioural control and speech. For example, Table 1 (and 10) does not offer any predictions with regards to the effect of behavioural control on response onset times. Moreover, the concept of 'behavioural control' is even more loosely defined than that of 'stress' or 'cognitive load'. The predictions made in relation to the possible changes in speech when speakers were controlling their behaviour were largely based on the findings of research into 'clear' or 'hyper-speech'. Hyper-speech may not adequately reflect the type of behavioural control possibly involved when lying. Secondly, although, attempts were made to keep potential confounding factors to a minimum, complete elimination is often impossible even in laboratory based research settings. In light of this, are other explanations possible, explanations beyond speaker internal processes, in order to account for the increase in overall rate of speech production when speakers were lying?

Vrij and Heaven (1999) suggest that lie complexity affects pausing behavior. In their research the authors illustrate 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). Indeed, neurobiological evidence exists that spontaneous and rehearsed lying may engage similar areas of the brain but to different degrees. Ganis et al. (2003) discovered that spontaneous lying activated regions associated with semantic and episodic memory retrieval, working memory engagement and response inhibition to a greater degree than rehearsed lies did.

In experiment 2, 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. The reader may remember that participants in experiment 1 were also able to semi-plan their answers; they were given 10 minutes to prepare a convincing statement prior to the 'security interview'. However, as stated above, a concealment type of lie may be 'easier' than a 'falsification'. Participants in experiment 1 had to invent a false account which may be seen as cognitively more demanding than 'simply' denying the possession of an object.

Beyond lie complexity, what other factors may have contributed to the disparity in temporal findings of experiment 1 and experiment 2? The two experiments featured two different styles of interviewing. An investigative style was employed in experiment 1 while experiment 2 contained an accusatory style of questioning. Buller et al. (1989) discovered that probing resulted in an increase in speech errors and hesitations. As part of his investigation, the interviewer in experiment 1 employed open-ended as well as follow up questions. He did not pursue an agenda but adjusted his questioning according to the interviewee's responses; he was probing participants' activity throughout. The interview scenario in experiment 2 used a more direct form of questioning. The questions were preset so participants' responses were not subject to follow-up questions. Admittedly, the questions were blunt and put participants on the spot but probing in the sense of searching for detailed information was not the case. It is possible that the difference in degree of probing between the two interviews lead to the diverging results with regards to number of hesitations.

Vrij et al. (2006b) also compared these two styles of interviewing and reported similar findings. In their study the authors observed that liars used fewer hesitations during the accusatory interview than the information gathering one. Moreover, information gathering interviews were perceived as more cognitively demanding than accusatory ones. The speech behaviour of liars in experiment 1 and experiment 2 seems to reflect this. However, caution has to be applied to not oversimplify the data or try and explain it by only resorting to the experimental manipulation, i.e. the act of deceiving. Compared to the baseline which aimed to create a positive and warm atmosphere, the interview 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 limited information sharing and general detachment. The results of the truth-telling group will provide more evidence whether the setting itself, i.e. a coercive and accusatory environment, would be sufficient to provoke changes in speaking behaviour.

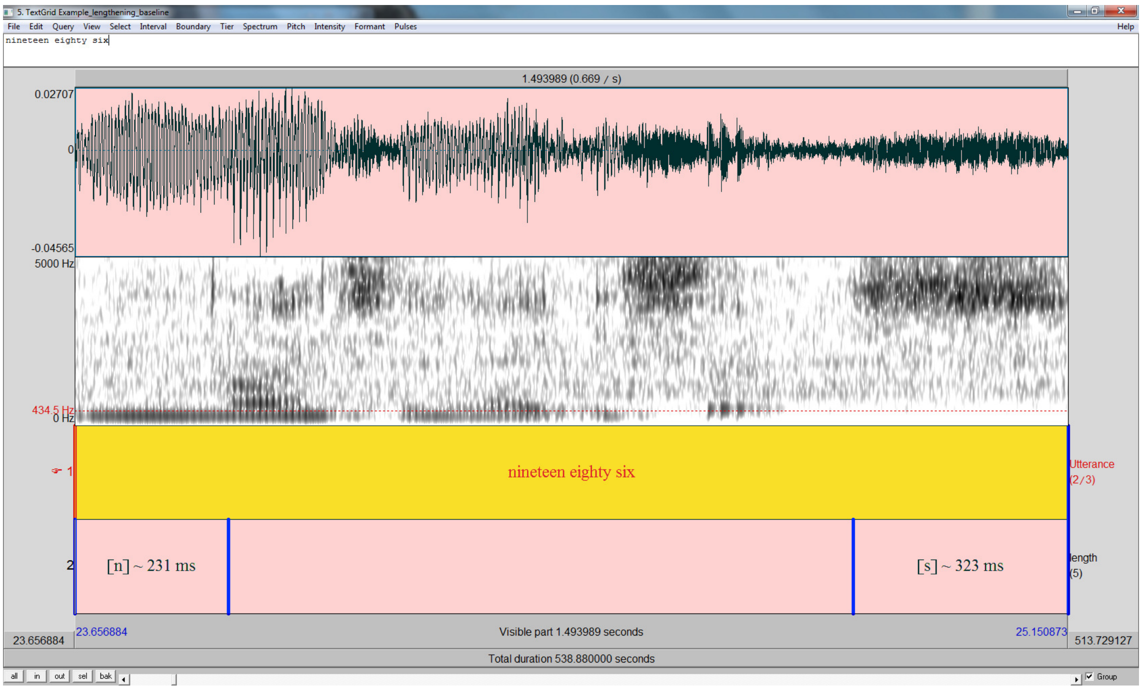

Figure 101 – Example of sound lengthening in the baseline in experiment 2

During the analysis process, it was repeatedly noted that speakers tended to speak more precisely when answering certain questions during the baseline interview. Occasionally, it was noted that speakers lengthened sounds beyond their canonical durations as illustrated in Figure 101.

Often, participants would even repeat the question before giving their answer. The opposite could be observed during the interrogation. Rather than giving a detailed response, speakers often appeared to be more succinct, vague and less direct in their answers. ‘Yes’ and ‘no’ responses were, by far, more common in the interrogation than the baseline. In addition, some speakers used the same response to answer multiple questions. For example, one participant answered the majority of the questions with ‘*I don’t have a token*.’ Participants may have resorted to some form of automated responding, akin to a suspect's ‘no comment’ strategy in PACE interviews when interrogated. They may have decided to say as little as possible in order not to slip up or give the investigators a trail which they could follow up. Goldman-Eisler (1968, pp. 58-59) has shown that routine and automated speech results in a decrease of pausing behavior and Ebusu and Miller (1994) discovered that vague and indirect language resulted in a faster rate and less pausing.

It was noted during the presentation of the results that the average length of pauses and ROT appeared to be rather high. Similarly, speakers average baseline SR values are lower than the normative SR values reported in previous studies. In his doctoral dissertation on tempo variation in speech production, Trouvain (2004, pp. 7-8) includes a survey of studies investigating SR in different languages and accents. The values for spontaneous speech across Dutch, English, French, German and Swedish ranged from 4.17 syll/s to 5.85 syll/s. Speakers’ mean SR of the baseline condition in the current study was 3.9 syll/s. The reason for this exceptionally low mean SR and long pausing may be seen in the experimental procedure employed and particularly in the format of question presentation. While prerecording the questions ensured consistency in delivery across participants, it also resulted in a less natural communicative setting. In a way, it could be compared to the automated speech systems people engage in when completing certain bank transfers or travel arrangements over the telephone. Based on personal experience, speakers tend to employ a slower speaking tempo and longer pauses when communicating with such speech recognition software.

While it is likely that the ‘unusual’ communicative environment of experiment 2 contributed to a slowing down of tempo in the baseline, it, apparently, did not affect the interrogation speech. The observed mean SR of 5.0 syll/s is closer to the normative values of the various studies reported in Trouvain (2004). One might ask the question how adequate the baseline condition was in eliciting participants' habitual ways of speaking. Did the act of lying result in faster and less hesitant speech or could it be that, in actual fact, speakers were slower and more hesitant in the baseline condition as compared to their usual speech?

Some of the baseline questions required participants to go back in memory e.g. '*Have you ever travelled abroad and where did you last go?*' or enquired about their personal preferences e.g. *'What is your favourite music'?* Speakers seemed to give a lot of thought to these questions. For example, one participant responded in the following way: '*Um, that's a tough question, I guess. I like, I guess, I like rock music but it depends what I'm doing. When I'm studying I prefer classical music but for fun I'd say like alternative rock music.'* It may be that the baseline condition presented a cognitively more demanding situation compared to the interrogation and rather than speeding up during the interrogatory phase, it could it be that speakers were slowing down during the baseline. In future work it is planned to analyse the data on a question by question basis which would provide more insight into this issue. This highlights that what may be assumed to be baseline behaviour e.g. control speech, in fact, may not be representative.

It cannot be discounted that methodological limitations may have slightly skewed the results. 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. As pointed out in Chapter 6 there are limitations with measuring tempo-related speech characteristics using these types of utterances. Firstly, they limit the opportunity for the occurrence of hesitation markers if speakers did decide to keep their responses as minimal as possible. Secondly, although, monosyllabic responses and high-frequency utterances such as 'I don't know' were not included in the analysis of AR and SR, the problem of sound prolongation or syllable shortening may not have been erased completely. Furthermore, learning effects or context conditioning may have affected participant's speech across conditions. As the interrogation always occurred after the baseline, it could be that speakers had gained familiarity with the speaking environment during the baseline recording and resorted back to their habitual speaking tempo when it came to the interrogation.

This chapter has answered questions but simultaneously created new ones. Firstly, the cognitive load theory of deception was not supported with the present data. While changes in the temporal domain did occur these were in the opposite direction as that proposed by the theory. Both from the speech analysis and the physiological domain, the findings were more adequately explained by making reference to the attempted control theory which predicts that liars consciously monitor their behaviour including speech. Secondly, it has been shown that external factors cannot be eliminated and will influence speaker internal processes even in experiments with tightly controlled methodological procedures. Type of interview, i.e. information gathering vs. accusatory may lead to completely diverging results when analysing deceptive speech in the temporal domain.

So far, analysis has focused on the behaviour of liars in two different interview situations. As pointed out in Chapter 1, the analysis of the behavior of truth-tellers was thought to be of interest, especially when placed in a similar experimental environment as liars. The questions that could be asked are: Will truth-tellers change their speaking behaviour? If, yes, what will these changes be?

8.2 Truth-tellers

As explained, experiment 2 also involved the collection of truthful speech. This was achieved by having a group of speakers, who did not receive a token and consequently did not have to conceal anything, complete the two interview sessions and answer the same questions as the liars. As pointed out in Chapter 6, the experimental design relied on a between–subjects paradigm and therefore the deceptive and truthful speech samples did not stem from the same speakers. This limited the scope for direct comparison of features and no attempt was made to directly compare the liars against the truth-tellers.

8.2.1 Articulation rate

The liars in experiment 2 displayed a significant increase in articulation rate. What about the truth-tellers (Figures 102 and 103)?

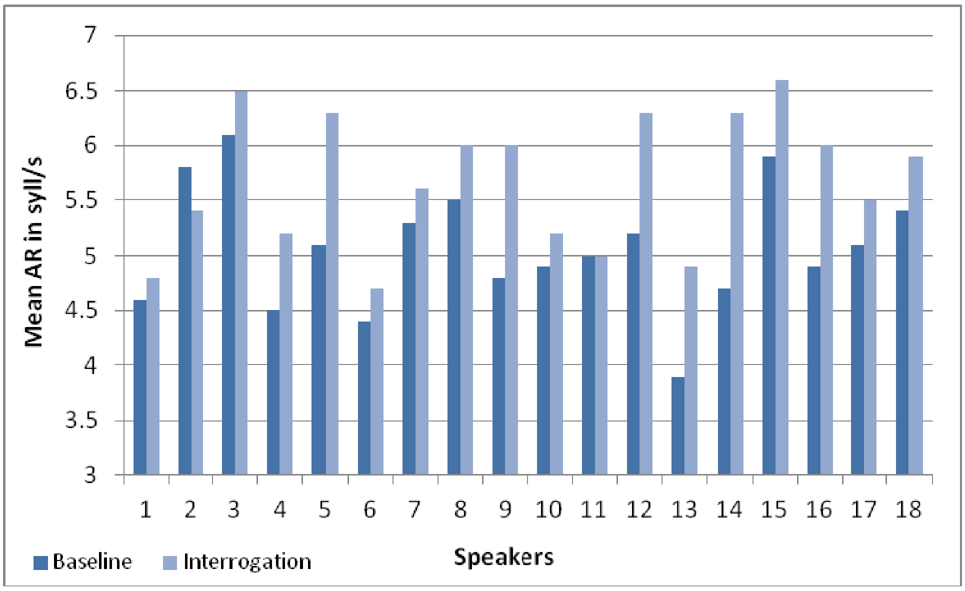

Figure 102 – Mean AR in baseline and interrogation for all truth-tellers

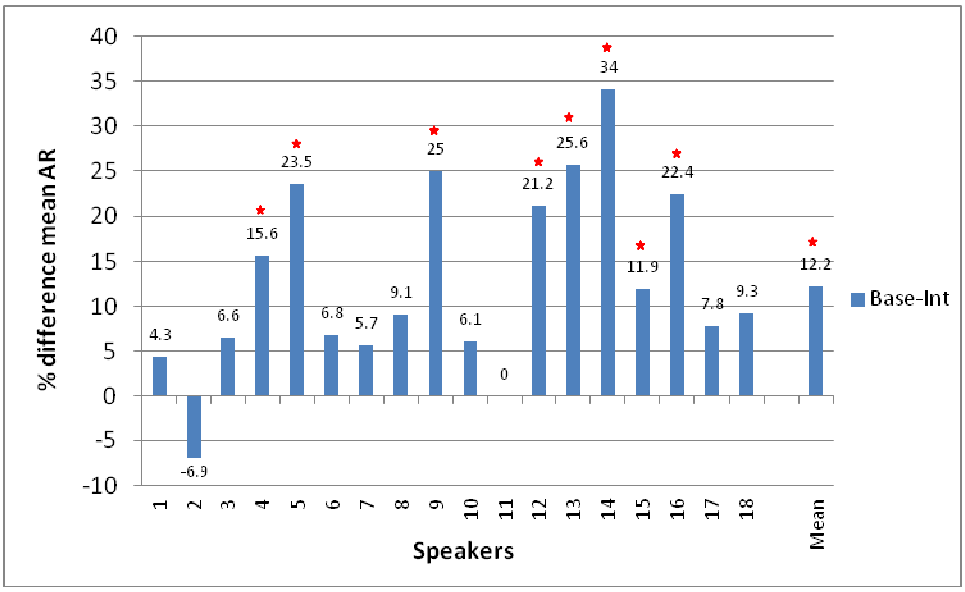

Figure 103 – % difference in mean AR between baseline and interrogation for all truth-tellers

The average AR increased from baseline (mean = 5.1 syll/s) to interrogation (mean = 5.7 syll/s) and this increase was significant ($T = 6$, $p \leq .001$, $r = -.56$). This was an almost uniform trend across all participants with speakers 2 and 11 presenting the only exceptions to this. The magnitude of increase differed for speakers but was most marked (and statistically significant) in speakers with a red asterisk. The change in mean AR was evaluated against the difference in phrase length across baseline and interrogation. As with the group of liars, the baseline condition contained significantly shorter phrases than the interrogation (5.4 syll vs. 6.3 syll) which could have affected the AR values. There was a medium strength correlation between the increase in phrase length and the increase in mean AR ($r_s = .38$, $p \geq .05$); however, this did not reach statistical significance. It follows that, while anticipatory shortening may have contributed to the faster mean AR in the interrogation, other factors cannot be dismissed.

8.2.2 Speaking rate

The group of liars showed an increase in mean SR when interrogated. Figure 104 and 105 detail the mean SR results for the truth-telling group.

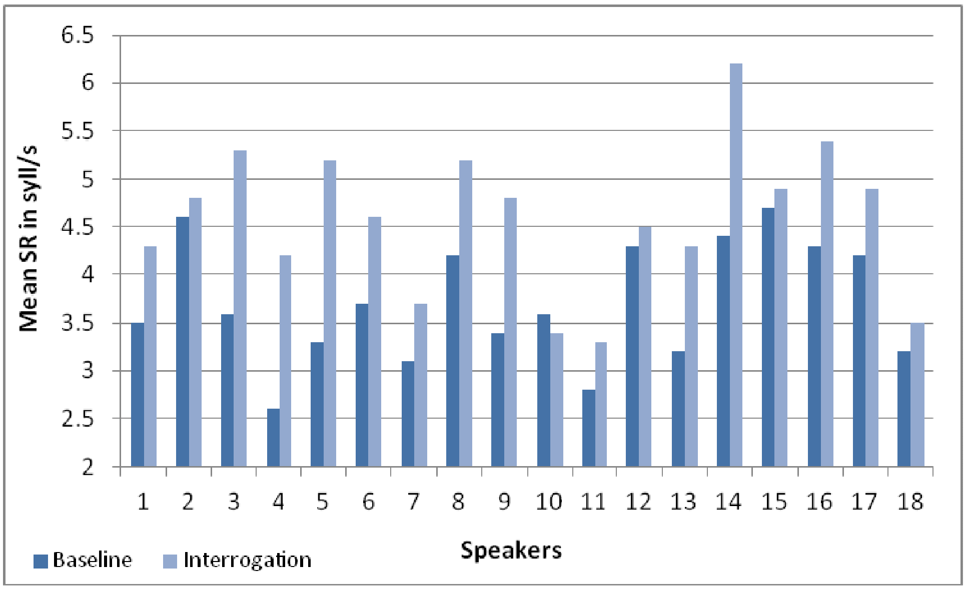

Figure 104 – Mean SR in baseline and interrogation for all truth-tellers

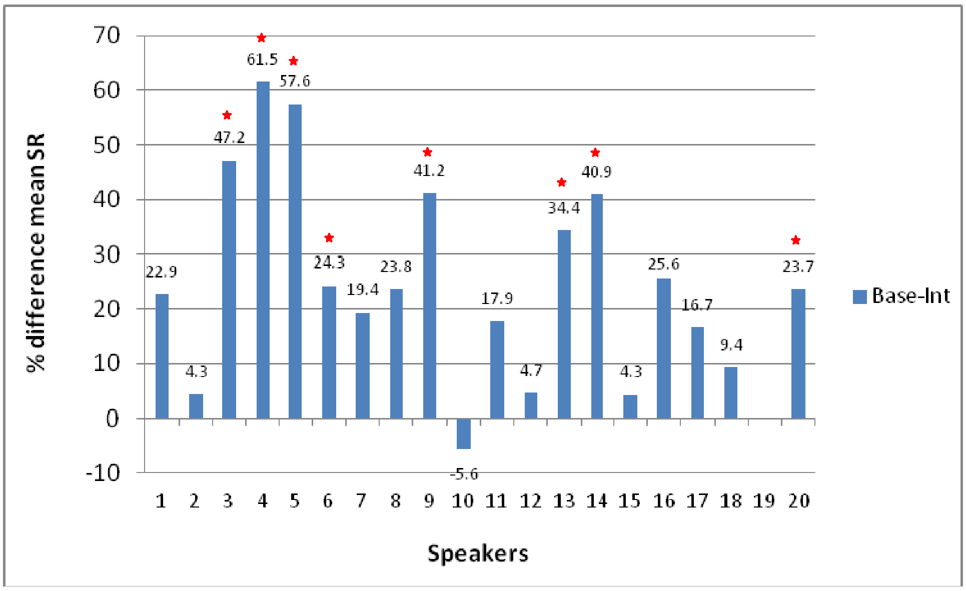

Figure 105 – % difference in mean SR between baseline and interrogation for all truth-tellers

The results correspond to those of AR in that, with the exception of one speaker, SR was shorter when participants were interrogated as opposed to when they were providing speech during the baseline. The increase from baseline (mean = 3.7 syll/s) to interrogation (mean = 4.6 syll/s) was significant (T = 2.5, $p \leq .001$, r = -.60). On the intra-speaker level, the increase was statistically significant for seven of the speakers as indicated by the red asterisks. As observed with the liars, some of the mean SR values were noticeably low and this was particularly striking in speakers 4, 7, 10, and 11.

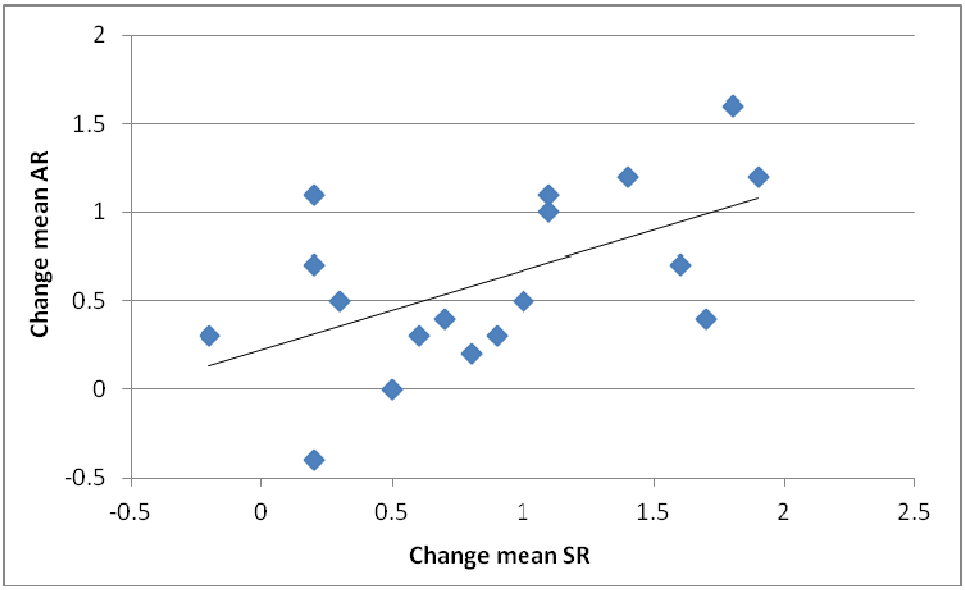

Figure 106 – Correlation between change in mean SR and change in mean AR

Interestingly, there was a positive and significant correlation between the difference in mean AR and mean SR ($r_s = .56$, $p \leq .05$) as illustrated in Figure 106.

Is the increase in mean SR and mean AR indicative that speakers were experiencing stress despite the fact that they were telling the truth?

8.2.3 Silent pauses (frequency and duration)

Liars employed less pauses during the interrogation and there was a tendency for pauses to be shorter also. Could the same be observed for truth-tellers (Figures 107 - 110)?

8.2.3.1 Frequency

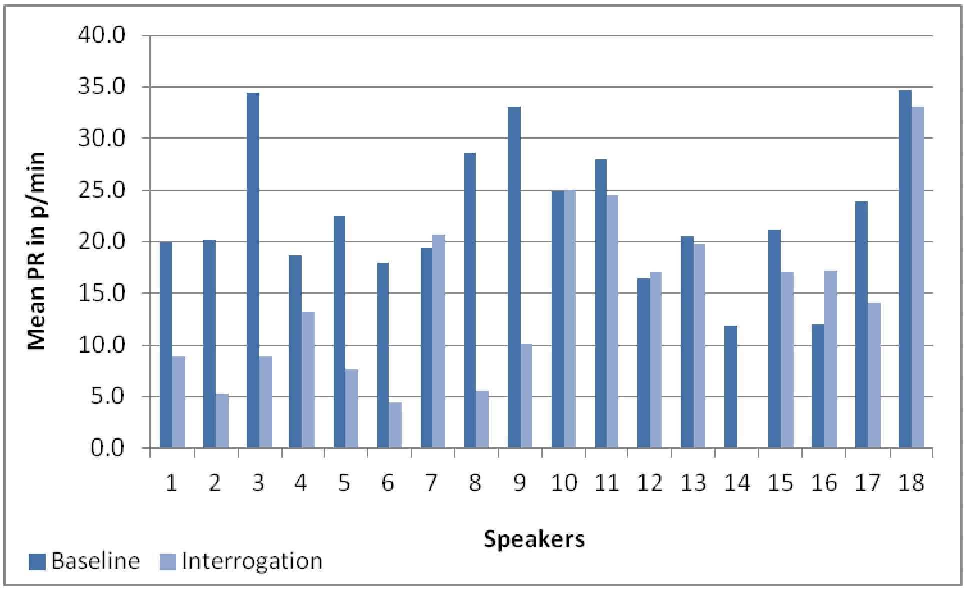

Figure 107 – Mean PR in baseline and interrogation for all truth-tellers

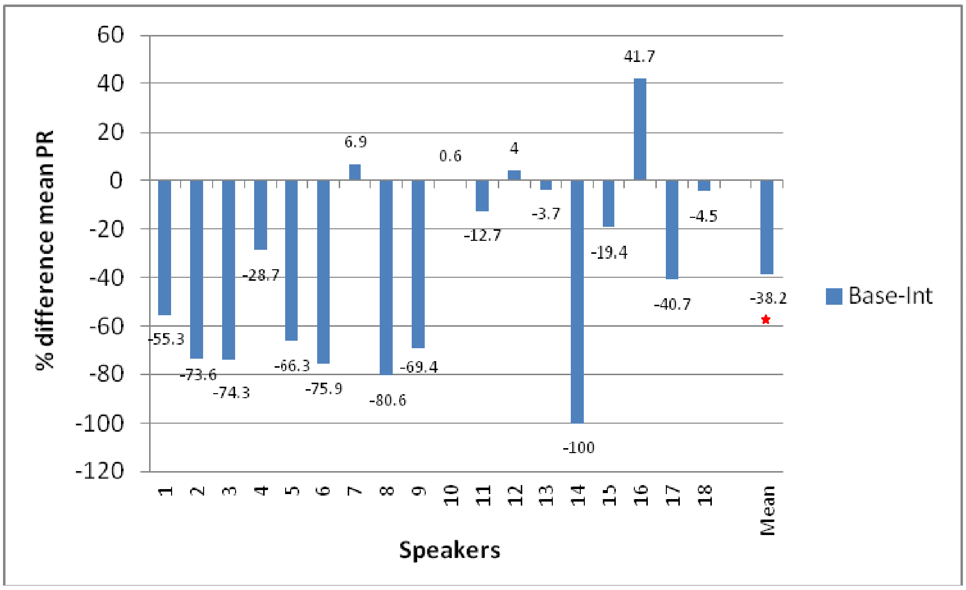

Figure 108 – % difference in mean PR between baseline and interrogation for all truth-tellers

Truth-tellers, as liars, made use of less pauses when they were being interrogated. The decrease in HR from baseline (mean = 22.7 p/min) to interrogation (mean = 14.0 p/min) was significant ($T = 15.5$, $p \leq .05$, $r = -.51$). The reductions were substantial for a number of speakers and in the case of speaker 14 resulted in no ocurrence of even a single instance of a pause during the interrogation. As expected, there was a significant correlation between decrease in number of pauses and increase in mean SR ($r_s = -.48$, $p \leq .05$).

8.2.3.2 Duration

As with the liars, pauses appeared to be less frequent in the speech of the truth-tellers particularly during the interrogation. Only speakers which employed three or more pauses were included in the pause duration calculations.

Overall, truth-tellers appeared to show longer pauses during the interrogation (median = 579 ms) as compared to the baseline (median = 541 ms); however, this increase was not significant ($T = 33$, $p \geq .05$, $r = -.23$). Unlike for the other parameters which showed uniformity across speakers, the data on pause duration suggests that there was strong variability in direction of change across speakers. Speaker 13 was the only one who showed a statisticaly significant change between baseline and interrogation. As with the liars, within-speaker variation in pause length was high (around 800 ms in baseline and interrogation) which may explain the lack of intra-speaker significance.

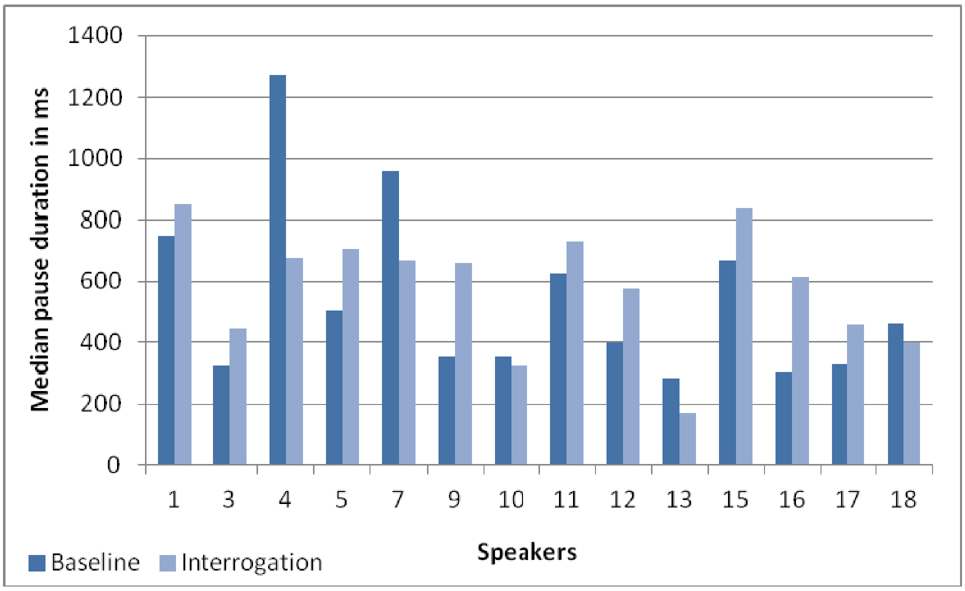

Figure 109 – Median pause duration in baseline and interrogation for fourteen truth-tellers

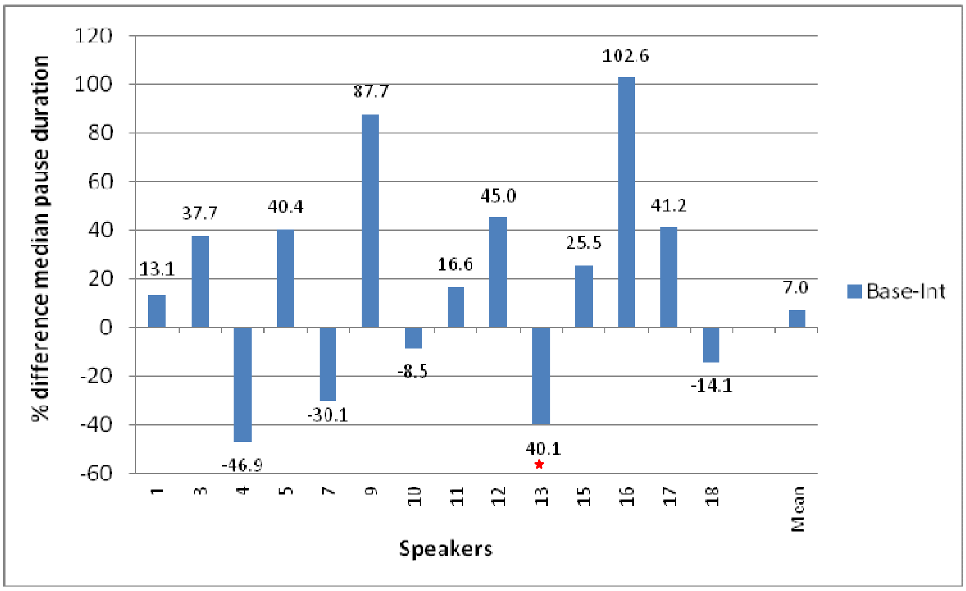

Figure 110 – % difference in median pause duration between baseline and interrogation for fourteen truth-tellers

Although truth-tellers showed a decrease in number of pauses when interrogated, their pauses became longer when accused of lying. This contrasts with the liars who showed a reduction in number as well as duration.

8.2.4 Hesitations (frequency and duration)

Liars used less hesitation when concealing their knowledge of the token during the interrogation. There was also a tendency for hesitations to be shorter, however owing to the reduced sample size, that conclusion was less strong. What effect did the accusatory interview have on the speech hesitations of truth-tellers (Figures 111 - 114)?

8.2.4.1 Frequency

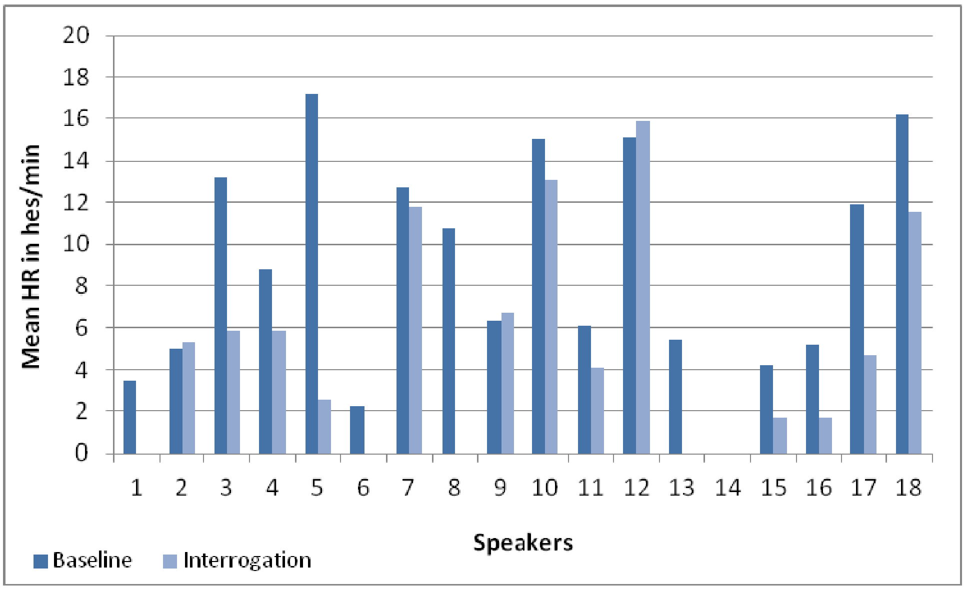

Figure 111 – Mean HR in baseline and interrogation for all truth-tellers

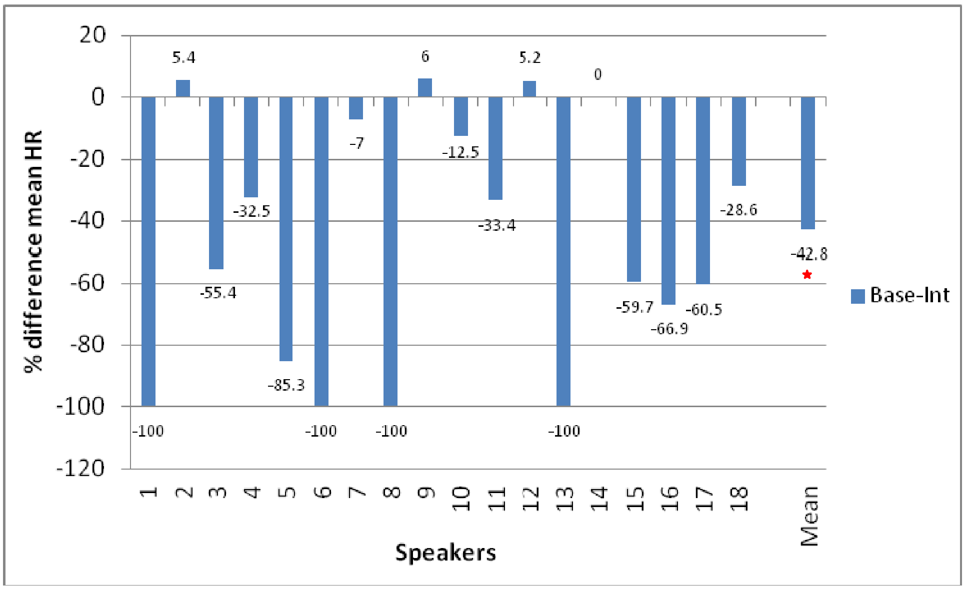

Figure 112 – % difference in mean HR between baseline and interrogation for all truth-tellers

For almost all speakers there was a decrease in the number of hesitations during the interrogative part of the experiment. The reduction in HR from baseline (mean = 8.8 hes/min) to interrogation (mean = 5.1 hes/min) was significant ($T = 6$, $p \leq .001$, $r = -.56$). The magnitude of decrease differed across speakers but could result in a complete absence of hesitations in the interrogation for some.

The idiosyncratic nature of hesitations can, once more, be seen in the variability in HR in the baseline with some employing a sizeable amount but others relatively few. Of particular interest is speaker 14 as he is the single speaker who does not employ any hesitation marker in the baseline condition.

8.2.4.2 Duration

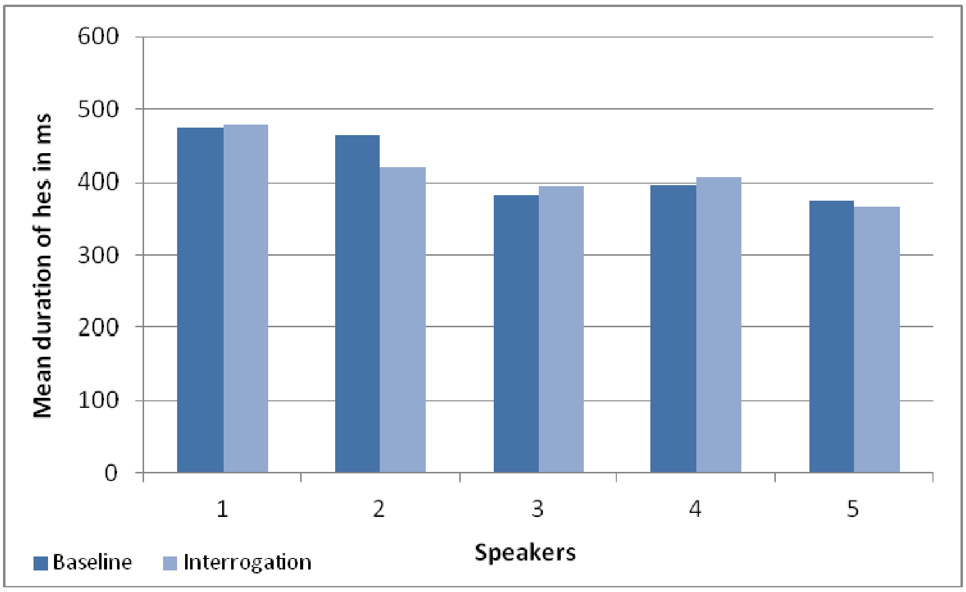

Figure 113 – Mean duration of hesitations in baseline and interrogation for five truth-tellers

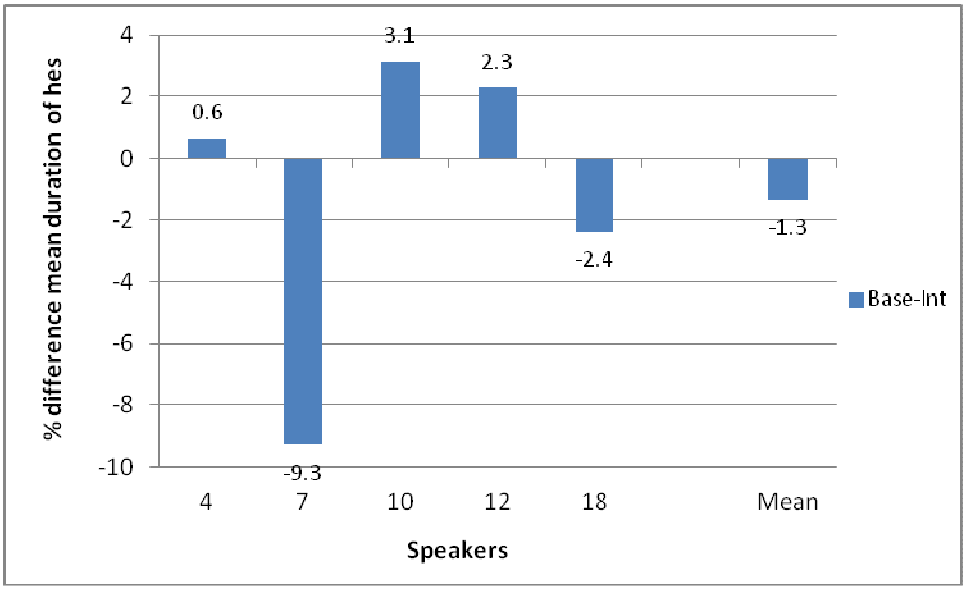

Figure 114 – % difference in duration of hesitations between baseline and interrogation for five truth-tellers

Only five speakers provided more than three hesitation markers in both the baseline and the interrogation. The difference in duration of hesitations between baseline (mean = 418 ms) and interrogation (mean = 413 ms) was not significant ($T = 7.5$, $p > .05$, $r = 0$). Indeed, Figures 113 and 114 confirm that, with the possible exception of speaker 12, the values are very close to each other across conditions. For none of the speakers was the change significant. Needless to say, a sample size of five does not allow for strong generalisations.

Truth-tellers, like liars, appeared to be less hesitant when they were interrogated about the possession of an object as compared to their baseline speech.

8.2.4.3 Hesitation by type and location

As with the liars, some exploratory analysis was performed in relation to different types and functions of hesitations. Informal observations suggest that frequency of both 'um' and 'uh' decreased during the interrogation. The general finding of a predominance of 'um' over 'uh' can also be noted for the group of truth-tellers. Neither duration of 'um' nor 'uh' was significantly affected; however small sample sizes were small due to a lack of hesitation tokens. Almost every speaker displayed a higher number of turn-initial and turn-internal hesitations in the baseline as compared to the interrogation. As was the case for the liars, the nature of the responses elicited may have limited the usefulness of such fine-grained differentiation for the current data.

8.2.5 Response onset time

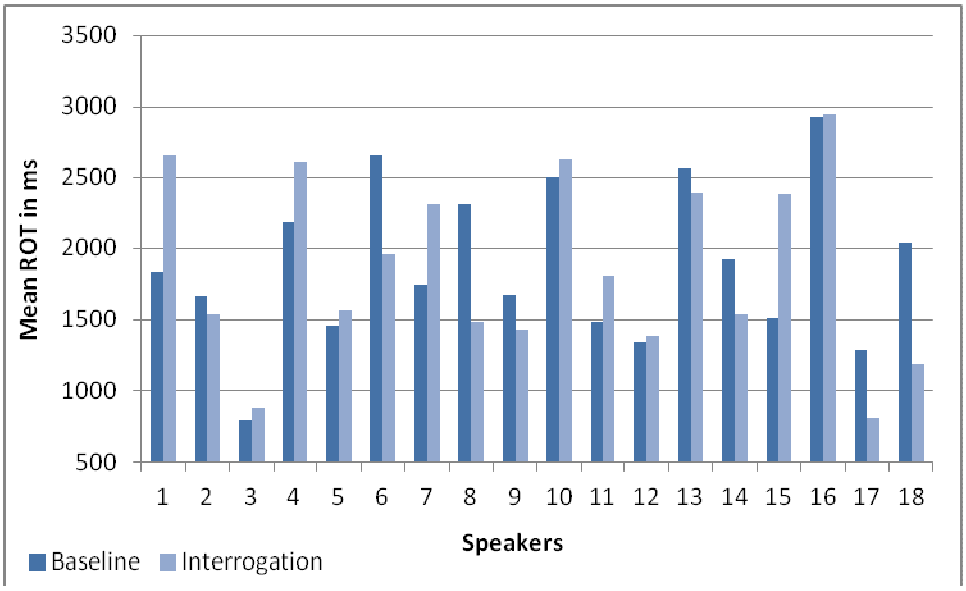

Figure 115 – Mean ROT in baseline and interrogation for all truth-tellers

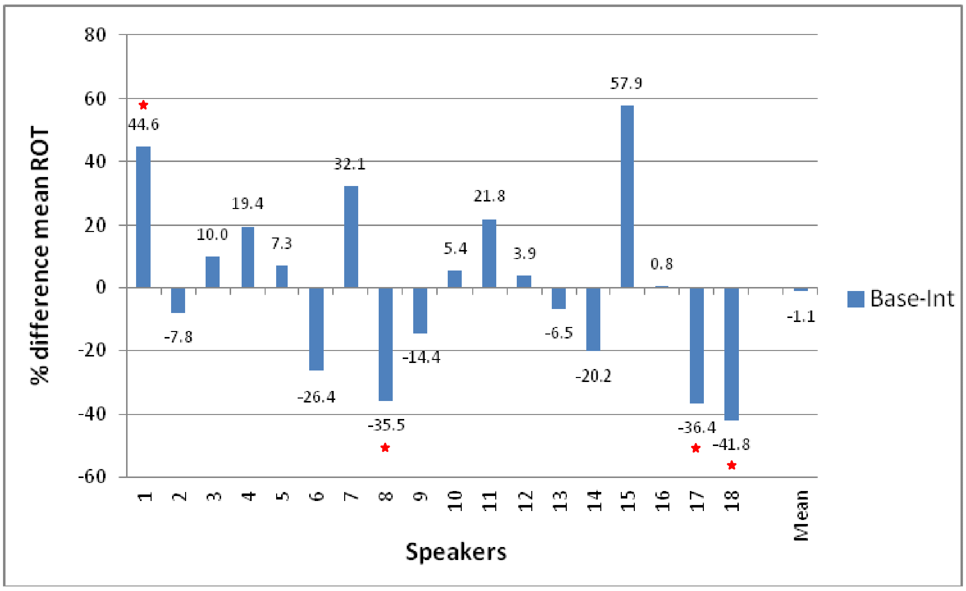

Figure 116 – % difference in mean ROT between baseline and interrogation for all truth-tellers

Figures 115 and 116 suggest that there was a high degree of variability across speakers. About half of the truth-tellers had longer response onsets during the interrogation, whereas the other half appeared to be responding quicker when interrogated. Overall, the change in ROT between baseline (mean = 1885 ms) and interrogation (mean = 1864 ms) was not significant ($T = 82.5$, $p \geq .05$, $r = -.02$). For four of the speakers the change in mean ROT between conditions was statistically significant. As with the liars, there was high variability in mean ROT within speakers with values averaging 948 ms in the baseline and 1251 ms in the interrogation.

8.2.6 Summary and discussion

Table 12 lists an overview of the results for the truth-tellers. When speaking the truth during the interrogation, participants showed a significant increase in mean AR and SR. Furthermore, there was a significant reduction in frequency of silent pauses and hesitations when truth-tellers were interrogated as opposed to their baseline speech. While duration of hesitations was not significantly affected, there was a tendency for pauses to be longer during the interrogation. When analysing hesitations considering type and syntactic position, a reduction in number was apparent during the interrogation as opposed to the baseline. There was no change in duration however. Mean ROT did not change significantly across the two conditions but there was a high degree of variability across speakers.

Table 12 – Summary of the findings from the temporal analysis of the truth-tellers in experiment 2

Parameter	Result	Significance
Articulation Rate	↗	*
Speaking Rate	↗	*
Frequency of Silent Pauses	↘	*
Duration of Silent Pauses	↗ (tendency)	ns
Frequency of Hesitations	↘	*
Duration of Hesitations	= (tendency)	ns
Response onset Time	=	ns

Statistical significance: ns = not significant, * = significant at .05 level

The first question that comes to mind relates to the fact that the speech behaviour of this group of speakers changed across conditions. Why did this change occur given that this group was not being deceptive? Secondly, the results of the temporal analysis of the speech of the truth-tellers were similar in nature to that of the deceptive group. The liars as well as the truth-tellers showed corresponding changes in almost all parameters measured. Therefore, it is not only of interest why a change occurred but also why this change was, to a large extent, compatible between the two groups.

To address the former question first, as mentioned in the discussion on the results of the liars (section 8.1.6) the interrogation setting itself could have had an influence on participant's behaviour. The truth-telling group was not manipulated to deceive the interviewer but they nevertheless had to undergo the same questioning procedures as the lying group. Just as for the liars, the interrogation setting could have been perceived as a hostile environment and, even though speakers were not expected to deceive they may still have felt uncomfortable, intimidated and perhaps even annoyed. Ofshe and Leo (1997) have shown that participants incorrectly accused of lying do experience fear of not being believed and this may result in similar behaviour as portrayed by a liar who experiences detection apprehension. When examining the physiological measures, higher respiration rates were recorded for the truth-tellers during interview as compared to baseline. In addition, as indicated by the STAI questionnaire truth-tellers reported greater levels of stress when interrogated (mean = 10.2) than when speaking in the baseline (mean = 9.1). The following dialogue illustrates that speakers did feel anxious despite the fact that they were telling the truth:

<u>*Interviewer:*</u> *Does this make you feel anxious? (Question 16)*

<u>*Participant:*</u> *I'm anxious that I'm being led down this path...*

Would the fact that speakers were subjected to coercive questioning strategies also explain why the change along the temporal domain was equal between the liars and the truth-tellers? Strömwall et al. (2006) argued that in high stakes situations liars and truth-tellers will experience the situational pressure to the same extent resulting in an increased self-presentational awareness for both. This is often referred to as the context overshadowing effect (Vrij, 2008, p. 88). It follows that, in order to appear credible and sincere, deceivers and non-deceivers alike will adopt similar strategies which moderate possible non-verbal behavioural differences. Although the setting of experiment 2 may not conform to what one might think of as a high stakes environment, it may still have elicited situational pressure. It was argued that liars may have resorted to automated responding as a form of self-protection-protection from being exposed as liars. The truth-tellers may have chosen a similar method of responding but for different reasons. Rather than protecting a lie from being revealed, truth-tellers may have aimed to defend their integrity and self-esteem. Some may have considered the questioning to be unjustified and unwarranted. By detaching themselves from the interrogation setting they may have intended to underline their disapproval and dissatisfaction with the situation they were being placed in. The detachment may be seen in the fact that truth-tellers, like liars, often only responded in indirect and monosyllabic language. An example response to the question: '*We know you were given an envelope before this interview. Can you tell us what was in it?*' was '*Uh, I can.*'

The results corroborate the findings of the study of Vrij et al. (2006b) which was already introduced in section 5.1.6. The authors reported that liars as well as truth-tellers showed less speech disturbances (hesitations and speech errors) during accusatory interviews. Moreover, the authors reported that there was no effect of veracity, i.e. no difference could be established between liars and truth-tellers on any of the behavioural parameters analysed. Vrij et al. (2006a) provide more evidence for an absence of variation between truth-tellers and liars with regards to behavioural cues shown during an accusatory interview.

In relation to the group of liars, it was discussed that the communicative setting may have affected the findings. It may further be possible that the unnatural environment prevented or masked the occurrence of a difference between liars and truth- tellers. Perhaps, the lack of a two-way interaction has had a stronger impact on speakers' behaviour than the act of lying. The purpose of presenting participants with pre-recorded questions was to minimize the effect of contextual variability, i.e. manner of presentation. Rather than just assuming that this was the case, it was decided to test whether the style of presentation was indeed comparable across the baseline and interrogation. The speaker presenting each of the twenty questions in baseline and interrogation was analysed with regards to AR, SR and pausing behaviour. It transpired that mean SR (5.1 syll/s for both baseline and interrogation) and frequency of pauses (19.2 p/min vs. 19.6 p/min) were indeed stable across the two conditions. However, rather unexpectedly, both the speaker's mean AR and duration of pauses significantly increased in the interrogation as compared to the baseline. AR increased from an average of 5.4 syll/s in the baseline to 5.9 syll/s in the interrogation ($T = 218.5$, $p = .05$, $r = -.23$). With a median value of 268 ms in the baseline to a median value of 435 ms in the interrogation, the lengthening of pauses was more striking ($T = 5$, $p \leq .001$, $r = -.58$). This finding highlights several points. Firstly, it cannot be discounted that speech accommodation, even if subconscious, may have still played a role in experiment 2. Liars as well as truth-tellers showed an increase in mean AR in the experiment. The increase in duration of pauses in the interrogation suggests that truth-tellers may have been more susceptible to accommodation than liars. Secondly, it may have been too optimistic to believe that variation in the contextual domain can be completely eliminated.

Most of the results of the truth-telling group matched that of the lying group. However, a few differences did emerge nevertheless. Specifically, differences could be indentified with respect to mean ROT and to some extent duration of silent pauses. Liars showed a strong tendency for a reduction in response onset time whereas no consistent pattern appeared for the truth-tellers. Similarly, while liars appeared to produce shorter pauses during the interrogation, truth-tellers generated slightly longer pauses. Chapter 5 highlighted that response time measurements appear to have gained a special standing within deception research. Perhaps it is the one aspect of speech which has the strength to differentiate between deceptive and non-deceptive speech? Amongst other things, Table 3 (Chapter 5) summarised the previous literature on response time as a function of deception and shows that the findings are mixed with some studies reporting longer ROTs and others shorter. The results of the current study strengthen this disparity; experiment 1 demonstrated an increase in mean ROT with lying whereas experiment 2 revealed that, when lying, mean ROTs were shorter.

Chapter 1 introduced the reader to the aim of the present research including the three research questions which guided the investigation. A comprehensive and critical review of the literature was presented in Chapter 2, 3, 4 and 5. This was followed by a description of the two experiments which formed the basis of the empirical work in Chapter 6. Chapter 7 described the acoustic and temporal analysis of experiment 1 and Chapter 8 detailed the results of the temporal analysis of experiment 2. The analysis findings were described in visual and verbal format and summaries and discussions were included at the end of each major section. Interesting and thought provoking observations were made. Chapter 9 will place these in the larger context which guided the current work.

9 Conclusions

The following part of the thesis combines all empirical findings described across Chapters 7 and 8 and relates them back to the four research questions introduced in Chapter 1. Each of the research questions will be addressed in turn and some of the identified methodological issues are addressed. In closing, practical implications of the empirical findings are spelled out and future directions are suggested which are believed to be valuable in strengthening the present understanding of deception and how it is manifested in speech.

9.1 Discussion

1. **Can deception be identified based on acoustic and temporal characteristics of speech?**

Table 13 – Summary of the results from the acoustic and temporal analysis of experiment 1

Parameter	Result	Significance
f_0	↗ (slight tendency)	ns
f_0 SD	↗ (slight tendency)	ns
Amplitude	=	ns
F_1	↗ (slight tendency)	ns
F_2	=	ns
F_3	=	ns
Vowel Space Area	↗	**
LTFD	=	ns
H_1-H_2	=	ns
Speaking Rate	↘ (strong tendency)	ns
Articulation Rate	↘	*
Frequency of Silent Pauses	↗ (strong tendency)	ns
Duration of Silent Pauses	↗	*
Frequency of Hesitations	↗	*
Duration of Hesitations	↗	*
Response onset Time	↗	*

Statistical significance: ns = not significant, * = significant at .05 level, ** = significant at .01 level

A summary of the results of the acoustic and temporal analysis of experiment 1 described in Chapter 7 is presented in Table 13. An upward arrow (↗) indicates that there was an increase in the parameter in question in the deceptive speech as compared to the baseline speech. A downward arrow (↘) indicates the opposite- a decrease, and the '=' sign signifies that no consistent change occurred. The frequency parameters, which are listed in the top half of the table, did not provide for consistent or reliable changes between the deceptive and non-deceptive speech samples. Some slight tendencies occurred but often these could be explained by factors other than the deceptive act. Vowel space area was the single parameter which showed a significant change during deception. When lying, speakers appeared to use larger working vowel spaces which could be indicative of hyper-articulation. However, as pointed out in section 7.1.5, this increase was not consistent across speakers or vowel categories. Moreover, the vowel formant measurements, both in form of mean formant frequencies and long-term formant distribution, did not support the view that hyper-articulation was taking place when speakers were deceiving. Overall, frequency based parameters of speech were not useful in differentiating deceptive from non-deceptive speech in the present experiment.

The temporal parameters on the other hand, listed at the bottom half of Table 13, offered more uniformity. In general, the results indicated a slowing down of speech when being deceptive. Comparing the results to the predictions made in Chapter 4 (Table 1, reproduced in Chapter 7 as Table 10), which were based on the three different theories of deception, it became clear that one theory in particular seemed to be most adequate in explaining these results. The overall trend of a slowing down of speech was in agreement with the cognitive theory of deception. In light of this, it appears as though there is potential merit in the temporal analysis of speech as a way of differentiating between deceptive and non-deceptive speech.

As became obvious in Chapter 5 when probing the available literature, research into the temporal aspects of deceptive speech has resulted in conflicting observations. The majority of studies tended to observe an overall slowing down of speaking tempo and, in particular, an increase in hesitation phenomena. Indeed, the conclusions from those studies strongly favour the 'cognitive approach' to lie detection. However, a number of studies exist which suggest the opposite, namely, an increase in speaking rate and a decrease in speech disturbances. Intrigued by this discrepancy, the present work aimed to further investigate the temporal domain of speech using a different set of deceptive speech data. Research question 2 addressed this issue of how robust temporal analysis really is for the purposes of deception detection.

2. **Do temporal parameters of speech provide a better means of detecting deception than frequency-based parameters?**

Taking into account the findings of experiment 2 the proposition that temporal parameters provide a better means of detecting deception than frequency based parameters is weakened. Table 14 evaluates the two experiments against each other. As before an upward arrow indicates an increase in the parameter when people were lying as compared to baseline speech and a downward arrow indicates a decrease when lying. Diverging changes with regards to the analysed temporal parameters can be observed across experiments. As already pointed out, the liars in experiment 1 showed a slower rate of speaking (both AR and SR), an increase in frequency and duration of pauses and hesitations and longer response onset times. Experiment 2, on the other hand, resulted in the opposite – when lying participants demonstrated a decrease in AR and SR, a decrease in number and duration of pauses and hesitations, and shorter response onset times. In combination, the findings from experiment 1 and experiment 2 reflect the disparity in results of previous research. The results of experiment 2 were interpreted in light of the behavioural control theory. It was suggested that rather than experiencing increased cognitive load, as may have been the case in experiment 1, participants in experiment 2 attempted to control their behaviour in order to come across as sincere.

Table 14 – Comparison of the results of the temporal analysis of experiment 1 and experiment 2

Parameter	Experiment 1	Experiment 2
Articulation Rate	↘	↗
Speaking Rate	↘	↗
Frequency of silent pauses	↗	↘
Duration of silent pauses	↗	↘
Frequency of Hesitations	↗	↘
Duration of Hesitations	↗	↘
Response Onset Time	↗	↘

Undoubtedly, some of the differences across both experiments are related to varying levels of cognitive load and behavioural control. However, attributing the observed changes to speaker internal processes only would have been too cursory and ignorant and not a true reflection of the complexities involved. In particular, speech accommodation theory was a driving theme when discussing the findings of experiment 1. It was shown that speakers may have tailored their speech towards that of the interviewers. As a result, it was difficult to discern whether the observed changes were due to speakers' experiencing cognitive load or a product of the social interaction. Further, even if it could be ascertained that speakers did converge, it could not be established whether this was related to the act of deceiving or connected to more general discourse strategies.

Experiment 2 attempted to control some of the confounding factors related to the context of the interaction (e.g. addressee and type of questions) in order to get a clearer picture of the deceptive process. Despite this, the observed changes could not be solely accredited to the act of lying. It transpired that although the influence of discourse-driven variables on speech was eliminated, other factors emerged which could have confounded the results. Based on this, it was argued that internal as well as external factors are likely to influence speakers' behaviour when being deceptive.

3. **Is the use of information gathering interview techniques more effective in eliciting deceptive speech cues than interrogation type interviews?**

Table 15 – Comparison of the two interviewing techniques in relation to the temporal findings

Parameter	Investigative Interview	Accusatory interview
Articulation Rate	↘	↗
Speaking Rate	↘	↗
Frequency of silent pauses	↗	↘
Duration of silent pauses	↗	↘
Frequency of Hesitations	↗	↘
Duration of Hesitations	↗	↘
Response Onset Time	↗	↘

The contradictory findings of experiment 1 and 2 were related to the different interviewing techniques used. As the reader will recall, experiment 1 used an investigative style of interviewing while experiment 2 was designed to employ an accusatory style. Table 15 compares the two interviewing techniques with regards to the observed changes in temporal parameters when speakers were deceptive. The first observation that can be made is that both styles of interviewing elicited a change in speaking behaviour along the temporal domain. Secondly, while both interview types resulted in temporal changes of speech, these changes went into the opposite direction. The investigative interview seemed to result in an overall slowing down of speaking tempo when speakers were deceptive. The accusatory interview, on the other hand, caused an overall acceleration in liars' speech. Whether the temporal changes in experiment 1 and 2 were cues to deception per se is more controversial an issue and was addressed in the discussion of research question 2. What has become clear, however, is that different interviewing techniques draw out contrasting effects on speech. This, once more, highlighted the complex nature of deception and underlined the difficulties in locating consistent cues of deceptive behaviour. The picture was further complicated by the fact that the observed changes in the speech of liars were not solely restricted to deceptive interactions. Truthful interactions, as proven by the data of experiment 2, led to similar adjustments in speech.

4. **To what extent does the speech of truth-tellers differ to that of liars when placed in a similar experimental setting?**

As spelled out in Chapter 1, previous research into deception, particularly investigations into speech related characteristics, often focused on liars only. Truth-tellers tended to be regarded as the norm against which the liars were compared. Indeed, support for this notion seemed to be found in the neurological domain. Initial fMRI studies showed that 'truth' seems to function as a 'baseline' because truthful responding was not associated with areas of increased brain activation (Spence, 2004). However, this was revoked by later studies which proved that truthful and deceptive responding may result in the activation of similar regions of the brain, i.e. prefrontal-parietal region (Langleben et al., 2005). The current work recognised that truth-tellers are just as likely to deviate from their normal speaking behaviour as are liars when the situational demands are similar. Experiment 2 offered the opportunity to investigate the speech behaviour of truth-tellers and liars and compare it to their respective baseline or control speech. It transpired that truth-tellers and liars showed corresponding changes in the temporal domain of speech for the majority of parameters analysed. Both groups displayed faster AR and SR, and used fewer hesitations and pauses. Strömwall et al. (2006) argued that in high stakes situation it is not only liars who will experience the situational pressure. Truth-tellers, too, will be aware of the consequences should they not be believed and, as a consequence, will resort to similar measures. The speaker internal processes connected to the emotional, cognitive and attempted control theories are not exclusive to deception but may accompany truth-telling as well. Furthermore, the context overshadowing effect ensures that external factors such as type of interview and interviewer are likely to have an effect on truth-tellers just as they have on liars. The findings underline Ekman's caution regarding the Othello error described in Chapter 2. Whenever attempts at lie detection are made, it needs to be remembered that truth-tellers and liars may show similar behavioural signs, e.g. corresponding changes in speaking behaviour.

9.2 Methodological issues

Empirical research is fallible to methodological shortcomings and the current study is no exception to this. No attempts are made to hide or cover up the fact that various limitations exist with the present research. As well as pointing out the flaws of the study, the following also highlights some of the advantages in comparison to previous research.

9.2.1 Experimental design

The overriding weakness of the work is that laboratory-induced deception 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. Attempts were made to verify participants' motivation to deceive by way of post-interview rating scales. These confirmed that all participants were highly motivated to succeed in the deceptive act (score of 6 or higher on a 7-point Likert scale).

The reader may further criticise the ecological validity of the study in light of the tightly controlled experimental setting, particularly for experiment 2. It is acknowledged that, ideally, one would carry out field experiments that reflect real life settings as closely as possible. However, as stated in the beginning of this work, research into a relatively unexplored area, such as speech and deception, needs to start off with fully controlled experiments, where variables can be manipulated more strictly. Clean, high quality recordings must provide the starting point and if robust differences between deceptive and non-deceptive speech are found in these ideal conditions, research can then move on to investigating less controlled data in the field or even real-life samples. Despite the high degree of laboratory control, the empirical studies did attempt to follow real-life suspect interviewing practices as closely as the circumstances could allow unlike some previous research designs, which asked participants to deceive during game situations, for example.

The reader will remember that based on the pilot testing of experiment 1, it was decided to revise the experimental methodology. Instead of dividing the security interview into two separate speaking conditions – truth and lie, it was considered as one speaking event, only reflecting deceptive speech. This methodological adjustment resulted in unequal sample sizes for the baseline and interview conditions in that the latter was considerably longer. Despite this, it became clear that revising the methodology significantly improved the experimental design overall. Having the truthful and deceptive speech in the same speaking event better reflects the practical setting. Considering real life police interviews, for example, suspects who are indeed guilty may contain a mixture of questions; some eliciting truths (e.g. biographical questions) and others lies. In addition, guilty suspects may be producing truths as well as lies when interviewed about the specific criminal event(s) (Turner et al., 1975; Maclin and Maclin, 2004; Wang et al., 2004). Having said this, it would have been insightful to have two separate speaking events, one reflecting truth-telling and one producing deceptive speech. Experiment 2 has shown that truth-tellers and liars show little differences in speaking behaviour when undergoing accusatory interviews. One of the potential reasons for this was attributed to the lack of face to face interaction between interviewer and interviewee. Would similar outcomes be observed with information gathering interviews which place more emphasis on inter-personal interaction?

Regularities tend to surface in a group rather than in individual speakers. A sample size of ten speakers could be seen as too small to produce any consistent trends even if there had been any. Experiment 1 contained various confounding factors including the use of different interviewers for the baseline and interview conditions and the type of questions used in both were different. Whilst this may reflect real-life practices it makes the differentiation of deception related processes and factors related to the stylistic domain of speech more complicated. A slightly larger sample was used as part of experiment 2 which also controlled some of these external factors by erasing differences in interviewer, question presentation and type of questions asked between conditions. The downside of the more stringent control seen in experiment 2 relates to the creation of a more clinical experimental environment. Experiment 2 only featured an analysis of selected temporal parameters. Reasons for restricting the analysis to the temporal domain were specified. An investigation that would have featured acoustic as well as temporal analysis would have provided for a better and more far-reaching comparison with experiment 1.

Amongst the limitations, the present work also has specific assets. One of these concerns the separation of stress and deception in that the latter was not inferred from the former. Harnsberger et al. (2009), for example, only included those subjects 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 dissociate deception and stress.

The vast majority of previous studies on deception focused on the behaviour of groups of subjects and tended to fail to report on the single individuals. Although analysis based on an individual affects statistical power, it provides for a more realistic picture when translating laboratory findings to real-world scenarios. The present study took into account the individual. Admittedly, this resulted in, at times, slightly disordered patterns but more importantly, it highlighted the fact that individual variability strongly underlines deceptive behaviour.

9.2.2 Measurement and analysis technique

Statistical analysis as it was performed on the current data was based on a linear model. Taking the AR data as an example, mean AR was computed for each speaker in each of the two conditions based on speakers' IP stretches. For the intra-speaker comparisons the IP phrases from the baseline were compared to the IP phrases of the interview. As it was spontaneous speech data, the number of IPs across conditions was not equal, i.e. the sample size was unbalanced. Therefore, a repeated measures design could not be used. Instead, the different IP stretches for baseline and interview were treated as independent measurements and a Mann-Whitney test was performed for the significance testing. Of course, the IP phrases are not truly independent of one another; they are from the same speaker. Accounting for these 'random effects' has been problematic; however, recent research has found a solution to this problem in the form of mixed-effect modelling which incorporates ‘random effects’ such as speakers (Hay, 2011). A mixed-effect model would have presented a more appropriate statistical technique for the current data.

There are possible methodological refinements in relation to the analysis of silent pauses and hesitations as conducted in the current study. Pauses could have been differentiated according to their lengths. For example, based on their large scale study of pause duration Campione and Veronis (2002) investigated circa 6000 pauses in reading and spontaneous speech from five different languages. The discovered a trimodal distribution and classified pauses according to brief (< 200 ms), medium (200 ms – 1000 ms), and long (> 1000ms). Such a three-way differentiation might have modelled the current data more accurately and possibly may have revealed more consistent and significant trends. Pause differentiation could have been extended to include location. For example, a division could be made between pauses that occur at grammatical boundaries and pauses which are not at grammatical junctures (Goldman-Eisler, 1968, p. 13). Kendall (2009), for example, reported that grammatical pauses were significantly longer than non-grammatical pauses.

Clark and Fox Tree (2002) differentiated between four different contexts in which hesitations may occur e.g. whether they are surrounded by silent pauses or whether they occur in between stretches of speech. Jessen (2012) reports findings which show that hesitations which are surrounded by silence are longer than hesitations which occur within speakers speaking turns. Informal attempts were made to differentiate hesitations according to whether they occurred turn-initially or during a speaker’s turn as part of the analysis of experiment 2. A more refined classification such as employed by Clark and Fox Tree may be of value. No hypothesis is offered at this stage in how far such fine-grained differentiation could be relevant to the study of deception and speech. However, hesitation and pausing in general are not just indicators of central planning processes; they also reflect social processes and carry important discourse functions (Crown and Feldstein, 1985; Kendall, 2007). In view of this, testing along this avenue might reveal as yet unexpected findings.

Articulatory precision was measured by calculating speakers’ vowel space areas. It was hypothesised that more careful speech would result in an enlargement of vowel space area whereas unclear and less precise articulation may result in smaller working vowel space. The VSAe measurements revealed that, when lying, speakers showed significantly larger vowel space areas then when speaking in the baseline. The mean formant frequency and LTFD data did not support this conclusion however. Another measure that could have given insight into whether speakers were speaking more clearly or not when deceiving could be seen in AR. In the present case AR was used in order to calculate the speech at which the articulators were moving. Phonetic syllables were measured to achieve this. However, in addition to giving an insight into how fast speakers articulate, AR could also be used as a measure of articulatory precision. Koreman (2006) discovered that measurements of AR using the phonological instead of the phonetic syllable may be indicative of style of speaking, i.e. speaking clearly as opposed to speaking unclearly.

9.3 Practical and theoretical implications

In the beginning of this work the popularity of voice stress analysis based technologies was mentioned. As explained, these devices are based on the assumption that liars will show more emotional arousal, i.e. will experience more stress than truth-tellers. The current investigation did not provide support for these predictions. Arousal theory did not adequately explain the observed changes in speech displayed by liars in the present study. In view of the present findings, it is fair to argue that the theoretical foundation and legitimacy of voice stress analysis based technology and, to some extent, polygraph testing is questionable. As mentioned in the introduction, lay persons tend to have confidence in the detective power of these tools. Indeed, one of the truth-tellers exemplified this by saying: *"I'm telling the truth and I hope that the data will reveal that."* The scientific community should strive to continue to inform misguided layperson perception in order to prevent the unethical use of scientifically unfounded methods and technologies.[14] As described in Chapter 1, the forensic speech science community has already established a professional code with regards to the forensic application of speech analysis. Perhaps, guidelines or standards could be set up which directly address the issue of using voice stress analysis based technology to detect deception.

The hypothesis that deception is a complex task involving cognitive processing demands which possibly exceed those required in truth-telling constituted an important part of this work. Support for this notion was partly gained from the results of experiment 1. Experiment 2

[14] Although their scientific claims are unfounded, VSA systems have been somewhat successful in that they have been shown to prevent people from telling lies (Damphousse et al., 2007). People will respond more honestly if they believe that a 'lie-detector' is used to assess the veracity of their statements. This is referred to as the 'bogus pipeline' effect (Jones and Sigall, 1971).

seemed to disfavour the cognitive theory of deception, however. Rather than discrediting the theory as such, the findings were explained by the nature of the experimental data. Specifically, attention was drawn to the fact that external factors such as type of questions and difficulty of the lie had important effects on speaker internal processes. Whilst complex fabrications indeed heighten cognitive load, easy concealment type lies may lead to the opposite, a reduction in cognitive involvement. Bearing in mind these caveats, the importance of cognitive load theory is acknowledged. In particular, it is likely to be appropriate in contexts which require lengthy responses such as police interviews.

The current work has shown that type of interview is a strong moderating variable affecting the temporal dimension of deceptive speech. The direction and magnitude of this effect may differ across the different interviewing techniques however. Police interviews typically try to elicit narrative responses while polygraph examinations elicit short answer yes/no utterances. The current data has shown that narrative deceptive responses may differ from short answer deceptive responses with regards to the temporal characteristics of speech. In terms of the practical setting, this would suggest that police interviews and polygraph examinations may not be used in conjunction when trying to locate deceptive behaviours based on temporal characteristics of speech only. Moreover, the findings suggest that there is not one tool or one set of guidelines that can be applied in order to detect liars. Detecting deception from police interviews may be a completely different task to detecting deception during brief airport security checks. When reviewing the literature, researchers and practitioners who are involved in detecting deceptive and suspicious behaviour need to be conscious not to misinterpret and/or misapply the findings. In order to move from an abstract laboratory based research tool to a real life diagnostic of deception, the method or technique needs to meet certain criteria and one of these is reproducibility or inter-rater reliability. Strong inter-rater reliability is based on two things: a high degree of accuracy of the technique and low individual variability. If 'speech analysis of temporal parameters' is regarded as a technique in deception detection, the present research has shown that accuracy will vary depending on situational factors such as type of interview. Furthermore, individual variability was a striking feature of deceptive responding. If a recommendation can be made then that is to accept that there is no single tool, no universal guidelines or strategies that will diagnose deceit. Whilst there are methods or techniques that are of potential value, the application of these need to be considered in light of the situational factors involved.

9.4 Deceptive speech research in future

Speech and deception has research potential. However, future investigations should be informed and guided by findings of previous research. In the words of DePaulo and Bond (2012) “we encourage future researchers to have a good reason for doing what has been done more than 200 times” (p. 120). Separation should be made between the types of deceptive speech analysed. For example, one should draw a distinction between deceptive narrative responses elicited from interview settings and concealment type messages from short answer protocols.

With regards to fabrications elicited from interviews, various cognitive based methods could be tested. To give an example, future research could explore the effectiveness of increasing cognitive load in uncovering differences between deceptive and non-deceptive speech. Increased cognitive load could be achieved by asking participants to perform a secondary task simultaneous to producing the deceptive account (Lancaster et al., 2013). Other measures that have been used in past research involved asking participants to maintain constant eye contact or instructing them to retell their deceptive account in reverse as opposed to chronological order (Vrij, 2008). Recalling events in a variety of temporal orders is a measure that is employed in cognitive interviewing (Shepherd, 2007, p. 219; Milne and Bull, 1999). Designed to help cooperative witnesses remembering details of an event, cognitive interviewing has gained increased acceptance. The application of aspects of the cognitive interview could be extended to include suspects as demonstrated by Vrij et al. (2008). Empirical studies could further test the efficacy of the cognitive interview in differentiating deceptive from non deceptive accounts taking into consideration speech related measures.

A paper by Vrij and Granhag (2012), which received a number of critical as well as supportive responses, argues for the need of a change in direction within deception research. As part of the novel approach, which is referred to as ‘interviewing to detect deception’, the authors highlight the importance of the type of questions asked. It is proposed that future research should focus on developing interview protocols that are effective in highlighting deception, not only within a police interview room but also in other settings, i.e. during airport security checks or general intelligence gathering. For example, lying about past events tended to be the experimental paradigm generally employed. While this is of practical relevance to real-life police interviews with suspect, it does not reflect the day to day task of border control/security agents who need to identify deceptive intentions. Scholars have started to tailor their research around this new approach. For example, Lancaster et al. (2013) and Vrij et al. (2009) have tested the merit of using anticipated and unanticipated questions in differentiating truthful and deceptive answers and Warmelink et al. (2012) investigated lying about future intentions. So far, these new and progressive research paradigms have failed to incorporate analysis of speech and voice related characteristics, something which could be addressed in future. As highlighted in experiment 2, it is essential that, when developing these novel interviewing techniques, attention is also given to the nature of the baseline questions asked.

In order to raise the ecological validity of deception research, varying contextual factors should also be investigated. Both Evans et al. (2012) and Vrij and Granhag (2012) stress the importance of research which is simultaneously informed by theory and practice. Whilst studying deception within the criminal setting is of paramount importance, DePaulo and Bond (2012) emphasise that future research should not fail to pursue to study deception within the many other aspects of life e.g. interpersonal relations, business and politics. Moreover, as well as performing empirical studies, researchers should not dismiss the value of studying the psychology and philosophy of deception. Above and beyond studying the speech of liars, future research should further focus on the complex relationship between truth-telling and lying. In addition, research designs could take a broader scope and investigate the characteristics of speech behaviour which is viewed as ‘suspicious’. The Idiap wolf corpus described in Chapter 1 would provide an ideal starting point for this.

9.5 In final closing

The current study offers a thorough and detailed investigation into the acoustic and temporal aspects of deceptive speech. Its novelty is multi-fold:

Firstly, consideration was given to parameters of speech not previously examined in connection with deception, including formants and articulation rate. Frequency based parameters in particular have so far received little or no attention in deceptive speech research and the present work contributes towards filling this gap.
Secondly, as speech analysis formed the primary goal of the work, the method of data collection and data analysis was informed by knowledge of phonetics and speech production. In the majority of cases, previous research was conducted by psychologists who relied on

methodologies common to their discipline, i.e. between-subjects design. This study was based on within-speaker comparisons using high quality and clean recordings (experiment 1).

Thirdly, the study differentiated between different types of deceptive contexts. Narrative type falsifications elicited from longer information gathering interview sessions (experiment 1) were compared to monosyllabic/short-answer deceptive responses produced during short interrogation protocols (experiment 2). Previous studies rarely took into consideration different communicative and contextual settings.

Fourthly, the thesis incorporated an analysis of aspects of the speech of truth-tellers when placed in similar setting to liars. 'Speaking the truth' has often been conflated with baseline speech behaviour. However, it may be expected that 'truth-telling' in the context of interrogatory pressure may differ from people's baseline speaking behaviour and therefore it is of value to study truth-telling in contexts of denial (Danielewicz-Betz and Ogasawara, 2013) in its own right. This study may be seen as pioneering the analysis of the speech of truth-tellers when rightfully denying the possession of an object.

The findings have demonstrated that deception is a complex mechanism involving the simultaneous interplay of speaker-internal and speaker-external factors. In the words of the British Psychological Society working group on polygraph analysis:

> "We must not deceive ourselves into thinking that there will ever be an error-free way of detecting deception" (BPS working group, 2004, p. 30).

Appendix A Experiment 1 - Information

A.1 Participant Information Sheet

You are being invited to take part in a research study. Before you decide if you want to participate it is important for you to understand the purpose of the research and what it will involve. Please take time to read the following information carefully. Please ask me if there is anything that is not clear or if you would like more information. Take your time to decide whether or not you wish to take part.

What is the purpose of the study?
The study looks into the effectiveness of a security campaign proposed by the University. To this end, it will test the ability of non-uniformed security wardens to differentiate guilty from non-guilty suspects. In addition, the present research looks into how people tell a convincing story. For this reason auditory data will be collected during the experiment.

Why have I been invited to take part?
We are seeking male adults who are native English speakers and who do not have any known speech or hearing impairments.

What is involved?
You will be asked to perform 3 tasks. Task 1 will require you to fill in personality questionnaires as well as biographical details. On finishing the questionnaires you will be asked to record some audio data with the experimenter. Task 2 involves the performance of a mock-theft and task 3 consists of an interview in which you have to convince a security warden of your innocence. There will be occasions during task 3 where you have to lie to the interviewer. The study will take approximately 1 hour.

What will happen to my data?
The data will be analysed and written up in project reports. Your data will always be treated anonymously; it will be impossible to identify you from the data. The raw data will be stored confidentially and only authorized researchers will have access to it.

Do I have to take part?
No. It is up to you to decide whether or not to take part. If you do, you will be given this information sheet to keep and be asked to sign a consent form. If you later change your mind you are still free to withdraw without giving a reason.

What are the possible benefits and risks of taking part?
You have the chance of earning £10 if you are successful in convincing the interviewer. In addition to the financial benefits we hope that you will find the experiment interesting and stimulating.

There are no anticipated hazards, risks or after affects associated with participating in this study. However, you will be asked to lie during the completion of task 3. Furthermore, if you fail to convince the interviewer of your innocence you may be asked to write a report about what happened.

Who is organising the study?
The study is being organised by Professor David Howard and Christin Kirchhübel of the Audio Lab in the Electronics Department. The study is being funded by the Engineering and Physical Sciences Research Council (EPSRC).

Who has reviewed of the study?
The study has passed the Physical Science Ethics committee review process.

A.2 Participant Informed Consent Form

Investigating the effectiveness of a novel security campaign

You have been invited to take part in an experiment. The experimenter will have informed you about the nature and the purpose of the research and you will have been given an information leaflet which you can keep. You have been invited to ask questions or request further information if any aspect was unclear.

In order for the experiment to commence we need your written consent. Please read the statements printed below carefully and feel free to ask questions if there is anything you do not understand.

In this experiment you are asked to perform a mock-theft about which you will be interviewed. The interviews will be audio recorded. During the interview you need to convince your interviewer that you are innocent of the theft by fabricating a false alibi. The study will take approximately 1 hour.

I understand the nature of the study and the procedures involved.

I understand that my participation is completely voluntary. I am free to withdraw at any point without giving a reason.

I understand that only the researchers conducting the study will have access to the data which, when not in use, will be confidentially stored.

I understand that my data will be analysed and will possibly appear in journal/conference publications. However, my data will be treated anonymously and no identifying information will be attached to it.

By signing the form you fully and freely consent to take part in the present experiment and agree for your data to be analysed.
Participant's name (BLOCK CAPITALS)

…………………………………………….

Participant's signature

……………………………………………. Date: ……………...

A.3 Debrief form

Your participation in this study is much appreciated. Thank you.

Occasionally, experimental research does not reveal the actual goals of the research study until participants have completed the experiment. The present study misled you into believing that the aims of the research were to test the effectiveness of a security campaign which the University is thinking of implementing. No such security scheme exists; it was just a cover story. There are no non-uniformed security wardens patrolling the building. The person who interviewed you participated in the study on a voluntary basis.

You were also told that being able to successfully deceive is correlated with possessing desirable social traits which in reality has not been verified empirically. Misinforming you in these subtle ways was thought to be essential in order to create a realistic scenario and to raise your motivation to succeed in the experiment.

The study seeks to look at the phonetic and acoustic correlates of deceptive speech and whether liars and truth-tellers can be differentiated on the basis of speech. For this purpose, you were audio recorded telling the truth and lying in an interview setting. We will analyse the truthful and deceptive utterances on a range of speech parameters including pitch, speaking intensity, speaking tempo, articulation and voice quality. Previous deception research has not really focussed on the speech signal and therefore the results from this research will provide relatively novel insights.

If you feel uncomfortable and do not want to participate any further in the study please let the researcher know. You are free to withdraw and your data would not be analysed. If you are interested in obtaining more details about this study or if you have any other questions please contact the researcher. As the data collection is still ongoing we would like to ask you not to talk about the nature of this study with others.

Appendix B Experiment 1 - Questionnaires

B.1 Biographical questions

Participant ID Number ……………….. **Date ………………….**

Please fill out this short biographical questionnaire.

1. Gender:
2. Date of Birth:
3. Place of Birth:
4. What is your current occupation?
5. If a student, what course are you studying and what year are you in?
6. What is your favourite module and why?
7. Do you have a part-time job?
8. Do you play sports? Are you part of a sports team?

B.2 Big Five scale

How accurately can you describe yourself?

Participant ID number……………… **Date…………………**

Please use this list of common human traits to describe yourself as accurately as possible. Describe yourself as you see yourself at the present time, not as you wish to be in the future. Describe yourself as you are generally or typically, as compared with other persons you know of the same sex and of roughly your same age. ***If there are any words that you do not know or understand, please circle them.*** Before each trait, please write a number indicating how accurately that trait describes you, using the following rating scale:

Inaccurate					Accurate			
Extremely	Very	Quite	Slightly	Neither	Slightly	Quite	Very	Extremely
1	2	3	4	5	6	7	8	9

1. ___Active	38. ___Inconsistent	75. ___Touchy
2. ___Agreeable	39. ___Inefficient	76. ___Trustful
3. ___Anxious	40. ___Inhibited	77. ___Unadventurous
4. ___Artistic	41. ___Innovative	78. ___Uncharitable
5. ___Assertive	42. ___Insecure	79. ___Uncooperative
6. ___Bashful	43. ___Intellectual	80. ___Uncreative
7. ___Bold	44. ___Introspective	81. ___Undemanding
8. ___Bright	45. ___Introverted	82. ___Undependable
9. ___Careful	46. ___Irritable	83. ___Unemotional
10. ___Careless	47. ___Jealous	84. ___Unenvious
11. ___Cold	48. ___Kind	85. ___Unexcitable
12. ___Complex	49. ___Moody	86. ___Unimaginative
13.___Conscientious	50. ___Neat	87. ___Uninquisitive
14. ___Considerate	51. ___Negligent	88. ___Unintellectual
15. ___Cooperative	52. ___Nervous	89. ___Unintelligent
16. ___Creative	53. ___Organized	90. ___Unkind
17. ___Daring	54. ___Philosophical	91. ___Unreflective
18. ___Deep	55. ___Pleasant	92. ___Unrestrained
19. ___Demanding	56. ___Practical	93.___Unsophisticated
20.___Disorganized	57. ___Prompt	94. ___Unsympathetic
21. ___Distrustful	58. ___Quiet	95. ___Unsystematic
22. ___Efficient	59. ___Relaxed	96. ___Untalkative
23. ___Emotional	60. ___Reserved	97. ___Verbal
24. ___Energetic	61. ___Rude	98. ___Vigorous
25. ___Envious	62. ___Self-pitying	99. ___Warm
26. ___Extraverted	63. ___Selfish	100. ___Withdrawn
27. ___Fearful	64. ___Shallow	
28. ___Fretful	65. ___Shy	
29. ___Generous	66. ___Simple	
30. ___Haphazard	67. ___Sloppy	
31. ___Harsh	68. ___Steady	
32. ___Helpful	69. ___Sympathetic	
33. ___High-strung	70. ___Systematic	
34. ___Imaginative	71. ___Talkative	
35.___Imperceptive	72.___Temperamental	
36.___Imperturbable	73. ___Thorough	
37. ___Impractical	74. ___Timid	

B.3 Short form of the state-trait anxiety questionnaire

Participant ID number………………. **Date …………**

A number of statements which people have used to describe themselves are given below. Read each statement and then circle the most appropriate number to the right of the statements to indicate how you feel **right now, at this moment**. *There are no right or wrong answers. Do not spend too much time on any one statement but give the answer which seems to describe your present feelings best.*

		Not at all	Somewhat	Moderately	Very much
1.	I feel calm	1	2	3	4
2.	I am tense	1	2	3	4
3.	I feel upset	1	2	3	4
4.	I am relaxed	1	2	3	4
5.	I feel content	1	2	3	4
6.	I am worried	1	2	3	4

B.4 Post interview rating scale

Participant ID number ……………. **Date………………..**

We would like to ask you a few more questions about your experience of the research study. This information is completely anonymous.

How did you feel about the tasks? Describe your feelings about some of the tasks you have just completed. Please rate on the scale below by circling the corresponding number.

1. Did you feel nervous when you were performing task 2, i.e. taking the money?

Not nervous	1	2	3	4	5	6	7	Extremely nervous

2. Did you feel it was difficult to complete task 2, i.e. taking the money?

Not difficult	1	2	3	4	5	6	7	Extremely difficult

3. Did you feel nervous when you were lying to the interviewer in task 3?

Not nervous	1	2	3	4	5	6	7	Extremely nervous

4. Did you feel nervous when you were telling the truth to the interviewer in task 3?

Not nervous	1	2	3	4	5	6	7	Extremely nervous

5. Did you feel it was difficult to convince the interviewer in task 3

Not difficult	1	2	3	4	5	6	7	Extremely difficult

6. How confident are you that you convinced the interviewer?

Not confident	1	2	3	4	5	6	7	Extremely confident

7. To what extent were you motivated to be successful in convincing the interviewer in task 3?

Not motivated	1	2	3	4	5	6	7	Extremely motivated

8. Did you perceive the interviewer as authoritative?

Not authoritative	1	2	3	4	5	6	7	Extremely authoritative

Appendix C Experiment 2 - Questions

Table 16 – List of questions used for each of the three conditions in experiment 2

Question	Baseline	Interview 1	Interview 2
1	Can you start with telling me how old you are?	Ok, for our records can you state what year you were born?	Please state for our records your nationality?
2	Which year were you born in?	Do you like going on rides at theme parks?	What are the first four letters of your surname?
3	What nationality are you?	Have you ever been formally interviewed by the police?	Have you ever told a lie to anyone in the past?
4	Can you tell me whether you're a student, member of staff or not connected to the university?	Would you consider yourself to be a generally honest person?	We know you were given an envelope before this interview. Can you tell us what was in it?
5	Have you ever had or do you now have a pet of any kind?	Have you ever told a lie to anyone in the past?	Do you have any information which you've been told to keep secret from us today?
6	When writing are you left-handed or right-handed?	Do you generally give money to people in the street collecting for charities?	Are you giving us truthful answers to our questions?
7	What did you have for breakfast today?	Do you generally give money to beggars in the street?	Apart from your answers do you think you're giving us any other clues as to whether you're concealing anything from us?
8	Do you have a best friend and if so what is their first name?	If you found some money in the street and you did not know who it belonged to would you be tempted to pick it up and keep it rather than hand it in to someone?	Did your envelope contain a blue token?
9	What was your favourite subject at school?	If you damaged someone else's property such as a car in a car park and no one else saw you do it would you leave your details so the owner could contact you?	If you have a blue token then we know it had the word banana on it. You might remember this. There was also a picture of the statue of liberty on it. Was she facing left or right?
10	Do you have a full driving licence?	Are you giving us truthful answers to our questions?	Did your envelope contain a red token?
11	How did you travel onto campus today?	Would you generally find it easier to keep a secret from someone else close to you such as a family member or a partner?	If you have a red token then presumably you haven't seen a picture of the Eiffel Tower this morning?
12	In total how long did it take to travel onto campus today?	Would you generally find it easy to keep a secret from someone even if the security guard or police officer asked you about it in an official capacity?	Do you think that your answers up to now sound evasive or are you giving us full and honest answers to our questions?

Table 16 (cont.) – List of questions used for each of the three conditions in experiment 2

Question	Baseline	Interview 1	Interview 2
13	Have you ever travelled abroad and if so where did you last go?	Do you have any information which you've been told to keep a secret from us today?	If we are to believe you you're saying you don't have a blue or red token in that case you either don't have a token or you have one of the yellow tokens. Have you seen the word avocado since you got here this morning?
14	When you are relaxing on holiday are you more likely to be sunbathing on a beach, reading a good book or lounging by the pool?	Apart from your answers do you think you're giving us any other clues as to whether you might be concealing anything from us?	From the way you're answering these questions we think you're withholding some information. I'll ask you again, are you concealing any information from us?
15	What is your favourite type of music?	Did anyone give you anything before you came to this interview which you have with you now?	Let's assume, just for a moment, that you were given a token. You be trying your best not to let us know you've got it but you're giving us other signs that you're concealing some information from us. Do you think that if we sit here long enough we'll be able to tell if you have a token just by looking at you?
16	What is your favourite type of movie?	Has someone offered you a reward for concealing any information from us?	Ok, let's keep the idea that you have a token. From your responses so far we think you might have either a red or a yellow token. If we identify the token correctly you'll lose your reward for the experiment. Does this make you feel anxious?
17	What is your favourite food?	Has anyone given you anything which you now have in your pockets or in a bag which we shouldn't know about?	If you have the red token why don't you just make it easier for us and tell us that you have it?
18	What kind of activities do you do to de-stress after a difficult day?	Would you be prepared to empty your pockets and/or bag in front of us?	If you're really telling us the truth why should we believe you?
19	What is your idea of somewhere where you can go to get peace and quiet?	Would you feel anxious if we searched you now?	If you don't have a token then you've got nothing to hide or worry about. But if you do have one do you really think you're not going to be found out?
20	Describe what you consider to be the perfect relaxing day.	Are you worried that we might discover you have something on your person which you should not have?	Look, we know you have a token. Doesn't matter how much you deny it, we don't believe you. We just don't know what colour it is and the whole purpose of this interview is to reveal that you have a token and the colour of that token. Why don't you just give up and hand it over?

Appendix D Experiment 1 - Results

Table 17 – Average values for all speakers in baseline (Base) and interview (Int) for f_0 (Hz), f_0 SD (Hz), and mean amplitude (dB)

Speaker	f_0 Base	f_0 Int	f_0SD Base	f_0 SD Int	Amplitude Base	Amplitude Int	Sig f_0	Sig f_0 SD
1	99	97	11	9	47	46	ns	-
2	114	117	5	9	53	52	*	-
3	88	90	3	5	51	52	*	-
4	112	113	6	6	47	46	ns	-
5	95	96	5	5	40	41	ns	-
6	141	144	5	6	53	57	*	-
7	127	127	9	10	48	50	ns	-
8	108	108	5	6	52	54	ns	-
9	115	117	6	10	54	50	ns	-
10	107	117	6	11	56	54	***	-
Group	110	113	6	8	-	-	*	ns

Table 18 – Average values for all speakers in baseline (Base) and interview (Int) for F_1, F_2, LTF_1, LTF_2 (all in Hz) and VSAe (Hz^2)

Speaker	F_1 Base	F_1 Int	F_2 Base	F_2 Int	VSAe Base	VSAe Int	LTF_1 Base	LTF_1 Int	LTF_2 Base	LTF_2 Int	Sig F_1	Sig F_2	Sig VSAe	Sig LTF_1	Sig LTF_2
1	496	509	1465	1396	111368	140001	457	470	1471	1417	ns	ns	-	-	-
2	454	463	1575	1495	50677	69923	386	386	1556	1448	ns	ns	-	-	-
3	379	403	1362	1359	38532	50471	342	332	1395	1366	*	ns	-	-	-
4	499	527	1460	1478	77106	109208	448	460	1440	1460	*	ns	-	-	-
5	520	495	1520	1578	151472	153024	433	443	1528	1531	ns	ns	-	-	-
6	534	536	1457	1630	159935	196171	465	493	1480	1491	ns	**	-	-	-
7	582	635	1536	1547	275649	308396	479	502	1566	1564	*	ns	-	-	-
8	453	493	1411	1391	86916	121481	419	420	1478	1519	*	ns	-	-	-
9	497	509	1338	1334	80558	104069	442	426	1360	1406	ns	ns	-	-	-
10	480	480	1405	1475	113124	111274	446	440	1468	1484	ns	ns	-	-	-
Group	489	505	1453	1468	114534	136402	432	437	1474	1469	*	ns	**	ns	ns

Table 19 – Average values for all speakers in baseline (Base) and interview (Int) for F_3, LTF_3, LTF_4 (all in Hz) and H_1-H_2 (dB)

Speaker	F_3 Base	F_3 Int	H_1-H_2 Base	H_1-H_2 Int	LTF_3 Base	LTF_3 Int	LTF_4 Base	LTF_4 Int	Sig F_3	Sig H_1-H_2	Sig LTF_3	Sig LTF_4
1	2394	2388	2.62	3.13	2378	2352	3566	3470	ns	ns	-	-
2	2235	2174	-0.95	-1.83	2267	2192	3669	3589	*	***	-	-
3	2259	2291	-2.95	-2.6	2358	2340	3588	3546	ns	ns	-	-
4	2387	2403	3.61	5.58	2408	2420	3717	3640	ns	***	-	-
5	2502	2472	5.83	5.66	2571	2523	3648	3584	ns	ns	-	-
6	2699	2682	-2.01	-1.77	2684	2631	3729	3589	ns	ns	-	-
7	2522	2484	4.14	3.43	2526	2477	3705	3777	ns	*	-	-
8	2359	2419	2.21	0.77	2392	2401	3555	3656	**	***	-	-
9	2364	2400	0.2	0.71	2431	2431	3631	3589	ns	**	-	-
10	2476	2501	1.05	1.34	2437	2475	3586	3586	ns	ns	-	-
Group	2420	2421	5.83	5.66	2455	2424	3639	3603	ns	ns	ns	ns

Table 20 – Average values for all vowel categories in baseline (Base) and interview (Int) for F_1, F_2 and F_3 (all in Hz)

Vowel	F_1 Base	F_1 Int	F_2 Base	F_2 Int	F_3 Base	F_3 Int	Sig F_1	Sig F_2	Sig F_3
/i/	323	329	2041	2045	2617	2603	ns	ns	ns
/ɪ/	417	418	1726	1713	2482	2481	ns	ns	ns
/ɛ/	564	589	1564	1581	2452	2458	ns	ns	ns
/a/	670	687	1342	1316	2341	2329	ns	**	ns
/ɜ/	489	508	1430	1435	2363	2364	ns	ns	ns
/ʌ/	511	521	1227	1208	2286	2293	ns	ns	ns
/ɒ/	513	521	1041	1024	2400	2373	ns	ns	ns
/ɔ/	436	423	914	900	2271	2310	*	ns	ns
all vowels	490	499	1411	1403	2401	2401	ns	ns	ns

Table 21 – Average values for all speakers in baseline (Base) and Interview (Int) for AR and SR (syll/s), PR (pause/min), median duration of pauses (ms), HR (Hes/min), duration of hesitations (ms) and ROT (ms)

Speaker	AR Base	AR Int	SR Base	SR Int	PR Base	PR Int	Dur Pause Base	Dur Pause Int	HR Base	HR Int	Dur Hes Base	Dur Hes Int	ROT Base	ROT Int	Sig AR	Sig SR	Sig PR	Sig Dur Pause	Sig HR	Sig Dur Hes	Sig ROT
1	5.9	5.5	4.1	3.5	29.3	32.3	614	624	5.5	7.7	344	385	392	435	ns	*	-	ns	-	ns	ns
2	6.6	6.5	4.9	5.3	25.2	28.9	524	642	7.3	10.4	397	404	718	648	ns	ns	-	ns	-	ns	ns
3	5.5	5.4	3.7	3.1	30.2	37.3	661	653	9.7	15.8	305	381	570	1013	ns	*	-	ns	-	*	ns
4	5.6	5.6	4.3	3.6	30.1	33.5	487	584	4.8	8.6	315	370	507	629	ns	*	-	ns	-	*	ns
5	6.0	5.6	4.5	4.3	31.6	28.9	568	634	6.1	8.4	347	360	585	612	**	ns	-	ns	-	ns	ns
6	5.0	5.1	4.0	4.0	23.7	25.2	534	564	2	4.1	311	374	338	570	ns	ns	-	ns	-	ns	ns
7	5.5	5.3	4.4	4.5	29.7	31.4	414	471	4.4	7	281	368	482	756	ns	ns	-	ns	-	*	*
8	5.0	4.9	3.8	4.0	27.5	25.7	468	506	6.2	12	510	477	392	602	ns	ns	-	ns	-	ns	ns
9	5.3	4.8	3.9	3.9	33.4	31.4	443	576	6.9	6.3	456	488	260	374	***	ns	-	**	-	ns	ns
10	5.5	5.1	4.6	4.2	22.2	25.6	463	467	8	8.9	341	451	473	585	**	*	-	ns	-	*	ns
Group	5.6	5.4	4.2	4.0	28.3	30.0	518	572	6.1	8.9	361	406	472	622	*	ns		**		*	*

Table 22 – Average values for all speakers in baseline (Base) and interview (Int) for frequency and duration of ‘um’ and ‘uh’ (all in ms). Average values for both interlocutors across all recording sessions in baseline (Base) and interview (Int) for f_0 SD (Hz) and AR (syll/s)

Speaker	Um Base	Um Int	Uh Base	Uh Int	Dur Um Base	Dur Um Int	Dur Uh Base	Dur Uh Int	f_0 SD Interlocutor Base	f_0 SD Interlocutor Int	AR interlocutor Base	AR interlocutor Int	Sig Um	Sig Uh	Sig Dur Um	Sig Dur Uh	Sig f_0	Sig AR
1	4.5	5.2	1	2.4	349	402	322	348	15	17	6.07	4.34	-	-	ns	ns	-	***
2	7.3	8.5	0	1.9	397	416	-	-	20	22	6.95	4.41	-	-	ns	-	-	***
3	3.7	7.7	6	8.2	390	470	251	298	8	19	6.28	4.69	-	-	ns	ns	-	***
4	4.2	7.3	0.6	1.3	333	398	186	217	14	15	5.95	4.28	-	-	**	ns	-	***
5	4.4	6.2	1.7	2.2	363	395	308	260	14	16	6.28	4.52	-	-	ns	ns	-	***
6	1.3	3.6	0.7	0.5	389	392	172	232	18	20	6.29	4.41	-	-	ns	ns	-	***
7	3.3	5.6	1.2	1.3	304	413	220	177	19	24	5.99	4.35	-	-	**	ns	-	***
8	5.3	11.6	0.8	0.4	556	486	218	207	22	20	5.98	4.52	-	-	ns	ns	-	***
9	5.4	5.4	1.5	1	499	525	299	282	17	22	6.94	4.39	-	-	ns	ns	-	***
10	4	2.5	4	6.4	474	585	207	399	17	17	6.03	4.17	-	-	ns	***	-	***
Group	4.3	6.4	1.8	2.6	405	448	242	269	16	19	6.28	4.41			*	ns		**

Appendix E Experiment 2 - Results

E.1 Liars

Table 23 – Average values for all liars in baseline (Base) and interrogation (Int) for AR and SR (syll/s), ROT (ms), and PR (pause/min)

Speaker	AR Base	AR Int	SR Base	SR Int	ROT Base	ROT Int	PR Base	PR Int	Sig AR	Sig SR	Sig ROT	Sig PR
1	5	5.4	3.7	5.3	2707	1654	14.6	0.0	ns	*	ns	-
2	4.9	5.9	3.9	3.6	1516	1226	19.6	34.6	*	ns	ns	-
3	4.8	5.8	4.2	5.2	3595	1422	11.4	4.3	*	ns	*	-
4	-	-	-	-	2349	2037	14.8	0.0	-	-	ns	-
5	5.6	5.4	3.9	4.2	1628	1816	27.2	23.5	ns	ns	ns	-
6	5	5.3	4	5	3089	1734	16.3	10.2	ns	ns	*	-
7	5.5	6	3.5	4.2	4018	2815	27.7	10.9	ns	ns	ns	-
8	5	5.8	3.3	4.5	2228	2626	29.9	28.9	*	*	ns	-
9	5.1	6.1	4.3	5.9	1234	1483	15.8	3.7	*	*	ns	-
10	5	6.4	4.1	5.7	1597	1582	22.8	5.1	*	*	ns	-
11	5.7	5.8	3.8	5.5	3022	2047	20.7	4.0	ns	*	ns	-
12	5.3	6	4.1	4.1	1740	1799	26.2	27.8	*	ns	ns	-
13	5.1	6.4	4.2	5.1	3481	5009	25.2	12.3	*	ns	*	-
14	5.6	7.4	4.1	5.4	1740	1511	32.0	17.1	*	*	ns	-
15	4.6	5.2	3.8	5.1	2115	1615	14.4	3.1	ns	*	ns	-
16	5	7.3	4.1	6.1	2622	1355	24.6	5.9	*	*	ns	-
17	5	5.9	3.8	4.1	1193	1404	25.1	31.7	*	ns	ns	-
18	4.6	6.1	3.4	5.3	1836	2494	18.8	8.7	*	*	ns	-
19	4.8	5.3	3.5	5	2032	1744	27.2	3.4	ns	*	ns	-
Group	5.1	6.0	3.9	5.0	2302	1967	21.8	12.4	***	***	ns	**

Table 24 – Average values for all liars in baseline (Base) and interview (Int) for duration of pauses (ms), HR (hes/min), and duration of hesitations (ms)

Speaker	Dur Pause Base	Dur Pause Int	HR Base	HR Int	Dur Hes Base	Dur Hes Int	Sig Dur Pause	Sig HR	Sig Dur Hes
1	-	-	12.2	0	-	-		-	-
2	920	257	8.2	8.7	521	270	*	-	*
3	-	-	2.3	0	-	-		-	-
4	-	-	8.4	0	-	-		-	-
5	367	283	22.4	7.1	532	366	ns	-	ns
6	1170	153	5.4	0	-	-	ns	-	-
7	450	409	12.4	7.3	449	446	ns	-	ns
8	509	508	5.5	1.4	-	-	ns	-	-
9	-	-	7.9	0	-	-		-	-
10	-	-	6.8	5.1	-	-		-	-
11	-	-	10.7	0	-	-		-	-
12	415	425	18.4	7.9	446	444	ns	-	ns
13	773	801	5.3	1.4	-	-	ns	-	-
14	410	220	23.7	26.5	404	399	ns	-	ns
15	-	-	10.1	3.1	-	-		-	-
16	-	-	30.2	47.6	260	212		-	*
17	331	324	13.8	7.9	418	386	ns	-	ns
18	773	371	10.9	5.8	-	-	ns	-	-
19	-	-	10.9	0	-	-		-	-
Group	612	375	11.9	6.8	433	360	*	**	*

E.2 Truth-tellers

Table 25 – Average values for all truth-tellers in baseline (Base) and interrogation (Int) for AR and SR (syll/s), ROT (ms), and PR (pause/min)

Speaker	AR Base	AR Int	SR Base	SR Int	ROT Base	ROT Int	PR Base	PR Int	Sig AR	Sig SR	Sig ROT	Sig PR
1	4.6	4.8	3.5	4.3	1838	2658	19.8	8.9	ns	ns	*	-
2	5.8	5.4	4.6	4.8	1670	1540	20.1	5.3	ns	ns	ns	-
3	6.1	6.5	3.6	5.3	797	877	34.4	8.8	ns	*	ns	-
4	4.5	5.2	2.6	4.2	2188	2612	18.6	13.3	*	*	ns	-
5	5.1	6.3	3.3	5.2	1461	1567	22.5	7.6	*	*	ns	-
6	4.4	4.7	3.7	4.6	2660	1957	17.9	4.3	ns	*	ns	-
7	5.3	5.6	3.1	3.7	1748	2309	19.3	20.7	ns	ns	ns	-
8	5.5	6	4.2	5.2	2307	1487	28.7	5.6	ns	ns	*	-
9	4.8	6	3.4	4.8	1676	1434	33.0	10.1	*	*	ns	-
10	4.9	5.2	3.6	3.4	2498	2632	24.8	25.0	ns	ns	ns	-
11	5	5	2.8	3.3	1485	1808	28.1	24.5	ns	ns	ns	-
12	5.2	6.3	4.3	4.5	1337	1389	16.3	17.0	*	ns	ns	-
13	3.9	4.9	3.2	4.3	2564	2397	20.5	19.8	*	*	ns	-
14	4.7	6.3	4.4	6.2	1930	1540	11.9	0.0	*	*	ns	-
15	5.9	6.6	4.7	4.9	1512	2388	21.1	17.0	*	ns	ns	-
16	4.9	6	4.3	5.4	2927	2949	12.1	17.1	*	ns	ns	-
17	5.1	5.5	4.2	4.9	1285	817	23.9	14.2	ns	ns	*	-
18	5.4	5.9	3.2	3.5	2042	1188	34.6	33.1	ns	ns	*	-
Group	5.1	5.7	3.7	4.6	1885	1864	22.7	14.0	***	***	ns	*

Table 26 – Average values for all truth-tellers in baseline (Base) and interview (Int) for duration of pauses (ms), HR (hes/min), and duration of hesitations (ms)

Speaker	Dur Pause Base	Dur Pause Int	HR Base	HR Int	Dur Hes Base	Dur Hes Int	Sig Dur Pause	Sig HR	Sig Dur Hes
1	749	847	3.5	0	-	-	ns	-	-
2	-	-	5.0	5.3	-	-	-	-	-
3	324	446	13.2	5.9	-	-	ns	-	-
4	1276	677	8.7	5.9	475	478	ns	-	ns
5	502	705	17.2	2.5	-	-	ns	-	-
6	-	-	2.2	0	-	-	-	-	-
7	957	669	12.7	11.8	463	420	ns	-	ns
8	-	-	10.8	0	-	-	-	-	-
9	351	659	6.3	6.7	-	-	ns	-	-
10	353	323	15.0	13.1	382	394	ns	-	ns
11	626	730	6.1	4.1	-	-	ns	-	-
12	398	577	15.1	15.9	396	405	ns	-	ns
13	284	170	5.5	0	-	-	*	-	-
14	-	-	0	0	-	-	-	-	-
15	666	836	4.2	1.7	-	-	ns	-	-
16	302	612	5.2	1.7	-	-	ns	-	-
17	325	459	11.9	4.7	-	-	ns	-	-
18	462	397	16.2	11.6	376	367	ns	-	ns
Group	541	579	8.8	5.1	418	413	ns	***	ns

Reference List

Abe, N., Suzuki, M., Mori, E., Itoh, M., and Fujii, T. (2007). Deceiving others: distinct neural responses of the prefrontal cortex and amygdale in simple fabrication and deception with social interactions. *Journal of Cognitive Neuroscience,* 19(2), 287—295.

Adams, S. (2002). Communication under stress: indicators of veracity and deception in written narratives. *PhD thesis.* Virginia Polytechnic Institute and State University.

Aamodt, M. G. and Custer, H. (2006). Who can best catch a liar? A meta-analysis of individual differences in deception detection. *The Forensic Examiner,* Spring, 7-11.

Anolli, L. and Ciceri, R. (1997). The Voice of deception: Vocal strategies of naïve and able liars. *Journal of Nonverbal Behavior,* 21*(4)*, 259—284.

Arciuli, J., Villar, G., and Mallard, D. (2009). Lies, lies and more lies. *Proceedings: The Annual Meeting of the Cognitive Science Society*, Amsterdam.

Baber, C., Mellor, B., Graham, R., Noyes, J. M., and Tunley, C. (1996). Workload and the use of automatic speech recognition: The effects of time and resource demands. *Speech Communication,* 20, 37—53.

Baskett, G. D. and Freedle, R. O. (1974). Aspects of language pragmatics and the social perception of lying. *Journal of Psycholinguistic Research,* 3, 117—131.

Bavelas, J., Black, A., Chovil, N., and Mullett, J. (1990). *Equivocal communication*. Newbury Park. CA: Sage.

Beckford Wassink, A., Wright, R. A., and Franklin A. D. (2006). Intra-speaker variability in vowel production: An investigation of motherese, hyperspeech, and Lombard speech in Jamaican speakers. *Journal of Phonetics,* 35, 363—379.

Bell, A. (1984). Language style as audience design. *Language in Society,* 13, 145—204.

Bello, R. (2006). Causes and paralinguistic correlates of interpersonal equivocation. *Journal of Pragmatics,* 38, 1430—1441.

Benson, P. (1995). Analysis of the Acoustic Correlates of Stress from an Operational Aviation Emergency. In I. Trancoso and R. Moore (Eds.). *Proceedings ESCA-NATO Tutorial and Research Workshop on Speech under Stress*, Lisbon, 61—64.

Benus, S., Enos, F., Hirschberg, J., and Shriberg, E. (2006). Pauses in deceptive Speech. *Proceedings ISCA 3rd International Conference on Speech Prosody*, Dresden, Germany.

Berthold, A. (1998). Repräsentation und Verarbeitung sprachlicher Indikatoren für kognitive Ressourcenbeschränkungen. *PhD thesis*, Universität des Saarlandes.

Bhatt, S. and Brandon, S. E. (2008). Review of voice stress based technologies for the detection of deception. Retrieved 12th April 2010 from http://www.polygraph.org/section/review-voice-stress-based-technologies-detection-deception.

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

Boltz, M. G. (2005). Temporal dimensions of conversational interaction: The role of response latencies and pauses in social impression formation. *Journal of Language and Social Psychology,* 24, 103—138.

Bond, C. F., Kahler, K. N., and Paolicelli, L. M. (1985). The miscommunication of deception: An adaptive perspective. *Journal of Experimental Social Psychology,* 21, 331—345.

Bond, C. F., Omar, A., Mahmoud, A., and Bonser, R. N. (1990). Lie detection across cultures. *Journal of Nonverbal Behaviour,* 14, 189—205.

Bond, Z. S. and Moore, T. J. (1994). A note on the acoustic-phonetic characteristics of inadvertently clear speech. *Speech Communication,* 14, 325–337.

Bond, C. F. and DePaulo, B. M. (2006). Accuracy of deception judgements. *Personality and Social Psychology Review,* 10, 214—234.

Bradley, B. P., Mogg, K., Falla, S. J., and Hamilton, L. R. (1998). Attentional bias for threatening facial expressions in anxiety: manipulation of stimulus duration. *Cognition and Emotion,* 12(6), 737—753.

Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication,* 20, 255–72.

Bradlow, A. R. and Bent, T. (2002). The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America,* 112(1), 272—284.

Bradlow, A. R., Kraus, N., and Hayes, E. (2003). Speaking clearly for learning-impaired children: sentence perception in noise. *Journal of Speech, Language, and Hearing Research,* 46, 80–97.

Brenner, M., Shipp, T., Doherty, E. T., and Morrissey, P. (1983). Voice measures of psychological stress- laboratory and field data. In I. R. Titze and R. C. Scherer (Eds.). *Vocal fold physiology: biomechanics, acoustics and phonatory control*. Denver: Denver Center for Performing Arts, 239—248.

British Psychological Society working group (2004). *A review of the current scientific status and field of application of polygraphic deception detection.* Final Report from the BPS Working Party. Leicester, England: British Psychological Society.

Bromme, R. and Wehner, T. (1987). Zum Zusammenhang von Sprechgeschwindigkeit und Sprechfehlem mit der Aufgabenschwierigkeit beim lauten Denken. *Zeitschrift für Experimentelle und Angewandte Psychologie,* 34, 1—16.

Brown, P. and Fraser, C. (1979). Speech as a marker of situation. In H. Giles and K. Scherer (Eds.). *Social markers in speech*. Cambridge: Cambridge University Press, 33—62.

Buller, D. B. and Aune, R. K. (1987). Nonverbal cues to deception amongst intimates, friends, and strangers. *Journal of Nonverbal Behaviour,* 11, 269—289.

Buller, D. B., Comstock, J., Aune, R. K., and Strzyzewski, K. D. (1989). The effect of probing on deceivers and truth-tellers. *Journal of Nonverbal Behavior,* 13, 155—170.

Buller, D. B., Burgoon, J. K., Busling, A. L., and Roiger, J. F. (1994). Interpersonal deception: VIII. Further analysis of nonverbal and verbal correlates of equivocation from the Bavelas et al. (1990) research. *Journal of Language and Social Psychology,* 13, 396—417.

Buller, D. and Burgoon, J. (1996). Interpersonal deception theory. *Communication Theory,* 6, 203—242.

Burgoon, J. K. (1993). Interpersonal expectations, expectancy violations, and emotional communication. *Journal of Language and Social Psychology*, 12, 30—48.

Camden C., Motley, M. T., and Wilson, A. (1984). White lies in interpersonal communication: A taxonomy and preliminary investigation of social motivations. *Western Journal of Speech Communication,* 48, 309—325.

Campione, E. and Véronis, J. (2002). A Large-Scale Multilingual Study of Silent Pause Duration. In B. Bel and I. Marlien, (Eds.*). Proceedings of the Speech Prosody 2002 conference Aix-en-Provence*, France, 199—202.

Carson, T. L. (2010). *Lying and Deception*. Oxford: Oxford University Press.

Cestaro, V. L. and Dollins, A. B. (1994). *An analysis of voice responses for the detection of deception*. Department of Defence Polygraph Institute. Report No. DoDPI94-R-0001.

Chen, F. R. (1980). Acoustic characteristics and intelligibility of clear and conversational speech at the segmental level. *MA Dissertation*, Massachusetts Institute of Technology.

Clark, J., Lubker, J., and Hunnicun, S. (1988). Some preliminary evidence for phonetic adjustment strategies in communication difficulty. In R. Steele and T. Threadgold (Eds.). *Language Topics: Essays in Honour of Michael Halliday*. Amsterdam: John Benjamins, 161—180.

Clark, H. and Fox Tree, J. (2002). Using uh and um in spontaneous speaking. *Cognition,* 84, 73—111.

Clermont, F., Harrison, P., and French, P. (2007). Formant-pattern estimation guided by cepstral compatibility. *16th Annual Conference of the International association of Forensic Acoustics and Phonetics (IAFPA)*, Plymouth, UK.

Cody, M. J. and O'Hair, H. D. (1983). Nonverbal communication and deception: Differences in deception cues due to gender and communicator dominance. *Communication Monographs,* 50, 175—193.

Cody, M. J., Marston, P. J., and Foster, M. (1984). Deception: paralinguistic and verbal leakage. In R. N. Bostrom and B. H. Westley (Eds.). *Communication Yearbook* 8. Beverly Hills, CA: Sage, 464—490.

Coleman, L. and Kay, P. (1981). Prototype semantics: The English word lie. *Language,* 57(1), 26—44.

Cohen, P, Cohen, J., Aiken, L. S., and West, S. G. (1999). The problem of units and the circumstance for POMP. *Multivariate Behavioural Research,* 34(3), 315—346.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., and Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin and Review,* 12(5), 769—786.

Cooke, M. and Lu, Y. (2010). Spectral and temporal changes to speech produced in the presence of energetic and informational markers. *Journal of the Acoustic Society of America,* 128, 2059—2069.

Coupland, N. (1984). Accommodation at work: Some phonological data and their implications. *International Journal of the Sociology of Language,* 46, 49—70.

Crown, C. and Feldstein, S. (1985). Psychological correlates of silence and sound in conversational interaction. In D. Tannen and M. Saville-Troike, (Eds.). *Perspectives on silence*. Norwood, NJ: Ablex, 31—54.

Cutler, A. and Butterfield, S. (1990). Durational cues to word boundaries in clear speech. *Speech Communication,* 9, 485–95.

Cutler, A. and Butterfield, S. (1991). Word boundary cues in clear speech: A supplementary report. *Speech Communication,* 10, 335–353.

Damphousse, K. R., Pointon, L., Upchurch, D., and 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*. Retrieved 28th January 2011 from http://www.ncjrs.gov/pdffiles1/nij/grants/219031.pdf.

Daneman M. (1991). Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research,* 20, 445–464.

Daneman, M. and Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior,* 19(4), 450—466.

Danielewicz-Betz, A. and Ogasawara, N. (2013). Truth statements in denial context: A study of prosodic, paralinguistic, and discursive cues in Japanese speakers. Journal of Forensic Research S11: 001. doi:10.4172/2157-7145.S11-001, online at http://www.omicsonline.org/2157-7145/2157-7145-S11-001.pdf

Dankovičova, J. (1997). The domain of articulation rate in Czech. *Journal of Phonetics,* 25, 287—312.

Davatzikos, C., Ruparel, K., Fan, Y., Shen, D. G., Acharyya, M., Loughead, J. W., Gur, R. C., and Langleben, D. D. (2005). Classifying spatial patterns of brain activity with machine learning methods: Application to lie detection. *NeuroImage,* 28, 663—668.

Davis, M., Markus, K. A., Walters, S. B., Vorus, N. and Connors, B. (2005). Behavioral cues to deception vs. topic incriminating potential in criminal confessions. *Law and Human Behavior,* 29, 683—704.

Davis, M. and Hadiks, D. (1995). Demeanor and credibility. *Semiotica,* 106, 5—54.

Denes, P. B. and Pinson, E. N. (1993). *The speech chain: the physics and biology of spoken language*. Freeman: New York.

DePaulo, B. M. and Rosenthal, R. (1979). Telling lies. *Journal of Personality and Social Psychology,* 37, 1713—1722.

DePaulo, B. M., Rosenthal, R., Rosenkrantz, J., and Green, C. R. (1982). Actual and perceived cues to deception: A closer look at speech. *Basic and Applied Social Psychology,* 3, 291—312.

DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. *Psychological Bulletin,* 111, 203–243.

DePaulo, B. M. and Pfeifer, R. L. (1986). On-the-job experience and skill at detecting deception. *Journal of Applied Social Psychology,* 16, 249—267.

DePaulo, B. M., Kirkendol, S. E., Tang, J. and O'Brien, T. P. (1988). The motivational impairment effect in the communication of deception: Replications and extensions. *Journal of Nonverbal Behavior,* 12, 177—202.

DePaulo, B. M. and Kirkendol, S. E. (1989). The motivational impairment effect in the communication of deception. In J. C. Yuille (Ed.) *Credibility assessment*. Dordrecht: Kluwer Academic Press, 51—70.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M. and Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology,* 70(5), 979—995.

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

DePaulo, B. M. and Bond, C. F. (2012). Beyond accuracy: Bigger, broader ways to think about deceit. Journal of Applied Research in Memory and Cognition, 1, 120—121.

DeTurck, M. A. and Miller, G. R. (1985). Deception and arousal: Isolating the behavioural correlates of deception. *Human Communication Research,* 12, 181—201.

Dixon, J. A., Tredoux, C. G., Durrheim, K., and Foster, D. H. (1994). The role of speech accommodation and crime type in attribution of guilt. The Journal of Social Psychology, 134(4), 465—473.

Douglas-Cowie, E. (2007). Humaine: D5h Mid Term Report on Database Exemplar Progress. Accessed on 12.05.12 at http://emotion-research.net/.

Duckworth, M and McDougall, K. (2012). Developing disfluency profiles for individual speakers: A study of Standard Southern British English. *Presentation delivered to the 21st Annual conference of the International Association of Forensic Acoustics and Phonetics (IAFPA)*, Santander, Spain.

Eachus, P., Stedmon, A. W., and Baillie, L. (2013). Hostile intent in crowded places: A field study. Applied Ergonomics, 44(5), 703—709.

Ebesu, A. S. and Miller, M. D. (1994). Verbal and nonverbal behaviors as a function of deception type. *Journal of Language and Social Psychology,* 13, 418—442.

Edelmann, R. (1998). Non-verbal behaviour and deception. In D. Canter and L. Alison (Eds.). *Interviewing and deception*. Aldershot: Ashgate Dartmouth, 157—182.

Ekman, P. (1985). *Telling lies: Clues to deceit in the marketplace, politics and marriage.* W. W. Norton New York: NJ. (Reprinted in 1992 and 2001).

Ekman, P. (1988). Lying and nonverbal behaviour: Theoretical issues and new findings. *Journal of Nonverbal Behaviour,* 12, 163—176.

Ekman, P. and Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica,* 1, 49—98.

Ekman, P. and Friesen, M. (1974). Detecting deception from body or face. *Journal of Personality and Social Psychology,* 29, 288—298.

Ekman, P., Friesen, W. V., and Scherer, K. R. (1976). Body movement and voice pitch in deceptive interaction. *Semiotica,* 16, 23—27.

Ekman, P., O'Sullivan, M., Friesen, W. V., and Scherer, K. R. (1991). Face, voice, and body in detecting deceit. *Journal of Nonverbal Behavior,* 15(2), 125—135.

Ekman, P. and Frank, M. (1993). Lies that fail. In M. Lewis and C. Saarni (Eds.). *Lying and deception in everyday life*. New York: The Guilford Press, 184—200.

Engelhard, I. M., Merkelbach, H. and van den Hout, M. A. (2003). The guilty knowledge test and the modified stroop task in detection of deception: An exploratory study. *Psychological Reports,* 92, 683—691.

Enos, F. (2009). Detecting deception in speech. *PhD thesis,* submitted to Columbia University.

Enos, F, Benus, S, Cautin, R. L., Graciarena, M, Hirschberg, J. and Shriberg, E. (2006). Personality factors in human deception detection: Comparing human to machine performance. *Proceedings of the 9th International Conference on Spoken Language Processing*, Pittsburgh, USA.

Eriksson, A. and 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.

Exline, R., Thibaut, J., Hickey, C., and Gumpert, P. (1970). Visual interaction in relation to Machiavellianism and an unethical act. In P. Christie and F. Geis (Eds.). *Studies in Machiavellianism*. New York: Academic Press, 53—75.

Evans, J. R., Houston, K. A., and Meissner, C. A. (2012). A positive, collaborative, and theoretically-based approach to improving deception detection. *Journal of Applied Research in Memory and Cognition,* 1, 122—123.

Fabricius, A., Watt, D. and Johnson, D. E. (2009). A comparison of three speaker-intrinsic vowel formant frequency normalisation algorithms for sociophonetics. *Language Variation and Change,* 21, 413—435.

Fant, G. (1970). *Acoustic theory of speech production: With calculations based on X-ray studies of Russian articulations*. (1st edition published in 1960). The Hague: Mouton.

Farrow, T. F. D., Reilly, R., Rahman, T. A., Herford, A. E., Woodruff, P. W. R., and Spence, S. A. (2003). Sex and personality traits influence the difference between time taken to tell the truth or lie. *Perceptual and Motor Skills,* 97, 451—460.

Farwell, L. A. and Donchin, E. (1991). The truth will out: interrogative polygraphy ('lie detection') with event-related brain potentials. *Psychophysiology,* 28, 531—547.

Feeley, T. H. and DeTurck, M. A. (1998). The behavioural correlates of sanctioned and unsanctioned deceptive communication. *Journal of Nonverbal Behavior,* 22(3), 189—204.

Ferguson, S. H. and Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America,* 112, 25—71.

Ferguson, S. H. and Kewley-Port, D. (2007). Talker differences in clear and conversational speech: acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research,* 50, 1241—55.

Fernandez, R. and Picard, R. W. (2003). Modelling drivers' speech under stress. *Speech Communication,* 40, 145—159.

Fiedler, K. and Walka, I. (1993). Training lie detectors to use non-verbal cues instead of global heuristics. *Human Communication Research,* 20, 199—223.

Fitch, H. (1989). Comments on 'Effects of noise on speech production: Acoustic and perceptual analyses [Journal of the Acoustical Society of America, 84, 917—928 (1988)]. *Journal of the Acoustical Society of America*, 86(5), 2017—2019.

Frank, M. G. and Ekman, P. (1997). The ability to detect deceit generalizes across different types of high-stake lies. *Journal of Personality and Social Psychology,* 72(6), 1429—1439.

Frank, M. G. and Feeley, T. H. (2003). To catch a liar: Challenges for research in lie detection training. *Journal of Applied Communication Research,* 31, 58—75.

French, J. P. and Harrison, P. (2006). Investigative and Evidential application of Forensic Speech Science. In H. A. Armstrong, E. Shepherd, E. G. Gudjonsson and D. Wolchover, (Eds.). *Witness testimony: psychological, investigative and evidential perspectives*. Oxford: Oxford University Press, 247—262.

Fuller, B. F., Horii, Y., and Conner, D. A. (1992). Validity and reliability of nonverbal voice measures as indicators of stressor-provoked anxiety. *Research in Nursing and Health,* 15, 379—389.

Fussey, P. (2013 in press). Eternal Return? Social science perspectives of 'upstream' terrorist activities. J*ournal of Police and Criminal Psychology*, in press for publication in Vol. 28 (2).

Ganis, G., Kosslyn, S. M., Stose, S., Thompson, W. L., and Yurgelun-Todd, D. A. (2003). Neural correlates of different types of deception: An fMRI investigation. *Cerebral Cortex,* 13, 830—836.

Giles, H. and Coupland, N. (1991). *Language: contexts and consequences*. Open University Press: Milton Keynes.

Giles, H., Coupland, N., and Coupland, J. (1991). Accommodation theory: Communication, context and consequence. In H. Giles, J. Coupland, and N. Coupland, (Eds.). *Contexts of accommodation: developments in applied sociolinguistics*. Cambridge: Cambridge University Press

Global Deception Research Team (2006). A world of lies. *Journal of Cross-Cultural Psychology,* 37(1), 60—74.

Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. New York: Harper and Row.

Gold, E. (2012). Articulation rate as a discriminant in forensic speaker comparisons. *Proceedings University of New South Wales Forensic Speech Science Conference*. Sydney, Australia.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment,* 4, 26—42.

Goldman-Eisler, F. (1968). *Psycholinguistics-Experiments in spontaneous speech*. London: Academic Press.

Goldstein, E. R. (1923). Reaction times and the consciousness of deception. *The American Journal of Psychology,* 34(4), 562—581.

Gonzales, J., Cervera, T., and Jose Lau, M. (2003). Acoustic analysis of pathological voices compressed with MPEG system. *Journal of Voice*, 17, 126—139.

Granhag, P. A. and Strömwall, L. A. (2002). Repeated interrogations: Verbal and nonverbal cues to deception. *Applied Cognitive Psychology,* 16, 243—257.

Granlund, S., Baker, R., and Hazan, V. (2011). Acoustic-phonetic characteristics of clear speech in bilinguals. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)*, Hong Kong, 17-21 August 2011, 763—766.

Greene, J. O., Lindsey, A. E. and Hawn, J. J. (1990). Social goals and speech production: Effects of multiple goals on pausal phenomena. *Journal of Language and Social Psychology,* 9, 119—134.

Grice, P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Griffin, G. R. and Williams, C. E. (1987). The effects of different levels of task complexity on three vocal measures. *Aviation, Space, and Environmental Medicine,* 58(12), 1165—1170.

Grosjean, F. and Collins, M. (1979). Breathing, pausing and reading. *Phonetica,* 36, 98—114.

Grynpas, J., Baker, R. and Hazan, V. (2011). Clear Speech strategies and speech perception in adverse listening conditions. *Proceedings of 17th International Congress of Phonetic Sciences (ICPhS)*, Hong Kong, 17-21 August, 779—782.

Gudjonsson, G. H. (2003). *The psychology of interrogations and confessions*. Chichester: John Wiley & Sons Ltd.

Hammerschmidt, K. and Jürgens, U. (2007). Acoustical correlates of affective prosody. *Journal of Voice,* 21(5), 531—540.

Hansen, J. H. L. and Bou-Ghazale, S. (1997). Getting started with SUSAS: a speech under simulated and actual stress database. *Proceedings Eurospeech, International Conference on Speech Communication and Technology*, September, Rhodes, Greece.

Hanson, H. M. (1997). Glottal characteristics of female speakers: Acoustic correlates. *Journal of the Acoustical Society of America*, 101, 466—481.

Harnsberger, J. D., Hollien, H., Martin, C. A., and Hollien, K. A. (2009). Stress and deception in speech: Evaluating Layered Voice Analysis. *Journal of Forensic Science,* 54(3), 642—650.

Harrison, A. A., Hwalek, M., Raney, D. F., and Fritz, J. G. (1978). Cues to deception in an interview situation. *Social Psychology,* 41, 156—161.

Harrison, P. T. (2004). Variability of formant measurements. MA Dissertation, Department of Language and Linguistic Science, University of York, York, UK.

Harrison, P. (2011). Formant measurement errors from real speech. *Presentation delivered to the 20th Annual Conference of the International Association for Forensic Phonetics and Acoustics*, Vienna.

Hartwig, M. (2005). Interrogation to detect deception and truth: Effects of strategic use of evidence. *PhD thesis*, University of Gothenburg.

Hartwig, M., Granhag, P. A., and Vrij, A. (2005). Police interrogation from a social psychology perspective. *Policing and Society,* 15, 379—399.

Hartwig, M. and Bond, C. F. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin,* 137(4), 643—59.

Hausner, M. (1987). Sprechgeschwindigkeit als eine Funktion von Stress: Eine Fallstudie. *Language and Style,* 20, 285—311.

Hay, J. (2011). Statistical analysis. In M. Di Paolo and M. Yaeger-Dror, (Eds.). *Sociophonetics: A students' guide*. New York: Routledge, 198—214.

Hazan, V. and Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *Journal of the Acoustical Society of America,* 130(4), 2139—2152.

Hecker, M. H. L., Stevens, K. N., von Bismarck, G., and Williams, C. E. (1968). Manifestations of task-induced stress in the acoustic speech signal. *Journal of the Acoustical Society of America,* 44, 993—1001.

Heilveil, I. and Muehleman, J. T. (1981). Nonverbal cues to deception in a psychotherapy analogue. *Psychotherapy: Theory, research and practice,* 18, 329—335.

Hicks, J. W. Jr. (1979). An acoustical/temporal analysis of emotional stress in speech. *PhD Thesis*, University of Florida.

Hillenbrand, J., Cleveland R. A. and Erickson, R. L. (1994). Acoustic correlates of breathy voice quality. *Journal of Speech and Hearing Research,* 37, 769—778.

Hirsch, A. R. and Wolf, C. J. (2001). Practical methods for detecting mendacity: A case study. *The Journal of the American Academy of Psychiatry and the Law,* 29, 438—444.

Hirschberg, J., Benus, S., Brenier, J. M., F. Enos, S. F., Gilman, S., Girand, C., Graciarena, M., A. Kathol, L. M., Pellom, B., Shriberg, E., and Stolcke, A. (2005). Distinguishing deceptive from non-deceptive speech. *Proceedings Eurospeech*, Lisbon, Portugal.

Hocking, J. E. and Leathers, D. G. (1980). Nonverbal indicators of deception: A new theoretical perspective. *Communication Monographs,* 47, 119—131.

Höfer, E., Köhnken, G., Hanewinkel, R., and Bruhn, C. (1993). *Diagnostik und Attribution von Glaubwürdigkeit*. Kiel: Final report to the Deutsche Forschungsgemeinschaft, KO 882/4-2.

Hoffman, B. (2006). *Inside terrorism*. New York: Columbia University Press.

Hollien, H. (1990). *The acoustics of crime: The new science of forensic phonetics*. New York: Plenum Press.

Hollien, H., Saletto J. A., and Miller, S. K. (1993). Psychological stress in voice: A new approach. *Studia Phonetica Posnaniensia,* 4, 5—17.

Hollien, H. and Harnsberger, J. D. (2006). The use of voice in security evaluations. In A. K. Webb and J. C. Kircher, (Eds.). Proceedings of the Workshop on the Use of Autonomic and Somatic Measures for Security Evaluations. *The Journal of Credibility Assessment and Witness Psychology,* 7(2), 74—78.

Hopper, R. and Bell, R. A. (1984). Broadening the deception construct. *Quarterly Journal of Speech,* 70, 288—302.

Hung, H and Chittaranjan, G (2010). The Idiap wolf corpus: exploring group behaviour in a competitive role-playing game. In *MM '10 Proceedings: Proceedings of the International Conference on Multimedia*. ACM: New York, 879—882.

Huttunen, K., Keränen, H., Väyrynen, E., Pääkkönen, R., and Leino, T. (2011). Effect of cognitive load on speech prosody in aviation: Evidence from military simulator flights. *Applied Ergonomics*, 42, 348—357.

Iacono, W. G. and Lykken, D. T. (1997). The validity of the lie detector: Two surveys of scientific opinion. *Journal of Applied Psychology*, 82, 426—433.

International Association of Forensic phonetics and Acoustics (IAFPA). Available at http://www.iafpa.net/

Inbau, F. E., Reid, J. E., Buckley, J. P., and Jayne, B. C. (2001). *Criminal interrogation and confessions.* Fouth edition. Gaithersburg: Aspen.

Jacewicz, E., Fox, R. A. and Wei, L. (2010). Between-speaker and within-speaker variation in speech tempo of American English. *Journal of the Acoustical Society of America,* 128, 839—850.

Jessen, M., Köster, O., and Gfroerer, S. (2005). Influence of vocal effort on average and variability of fundamental frequency. *International Journal of Speech, Language and the Law,* 12(2), 174—213.

Jessen, M. (2006). *Einfluss von Stress auf Sprache und Stimme. Unter besonderer Beruecksichtigung polizeidienstlicher Anforderungen*. Idstein: Schulz-Kirchner Verlag GmbH.

Jessen, M. (2007). *Forensic reference data on articulation rate in German. Science and Justice*, 47, 50—67.

Jessen, M. (2008). Forensic phonetics. *Language and Linguistics Compass,* 2(4), 671–711.

Jessen, M. (2012). *Phonetische und linguistische Prinzipien des forensischen Stimmenvergleichs*. München: LINCOM.

Johannes, B., Salnitski, V. P., Gunga, H-C., and Kirsch, K. (2000). Voice stress monitoring in space- Possibilities and limits. *Aviation, Space, and Environmental Medicine,* 71, 58—64.

Johnson, K., Flemming, E., and Wright, R. (1993). The hyperspace effect: Phonetic targets are hyper-articulated. *Language,* 69(3), 505–528.

Johnson, M. K., Foley, M. A., Suengas, A. G., and Raye, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology: General,* 117, 371—376.

Johnson, M. K. and Raye, C. L. (1981). Reality monitoring. *Psychological Review,* 88, 67—85.

Johnstone, T. and Scherer, K. R. (1999). The effects of emotions on voice quality. In *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, 2029—2032.

Jones, E. and Sigall, H. (1971). The bogus pipeline: a new paradigm for measuring affect and attitude. *Psychological Bulletin*, 76, 349—364.

Jou, J. and Harris, R. J. (1992). The effect of divided attention on speech production. *Bulletin of the Psychonomic Society,* 30, 301—304.

Karlsson, I., Banziger, T., Dankovicová, J., Johnstone, T., Lindberg, J., Melin, H., Nolan, F. and Scherer, K. (2000). Speaker verification with elicited speaking styles in the VeriVox project. *Speech Communication,* 31, 121—129.

Kendall, T. S. (2009). Speech rate, pause, and linguistic variation: An examination through the sociolinguistic archive and analysis project. *PhD thesis*, Duke University.

Kendall, T. 2007. Listening to silence: Interpretation and transcription of pause in Deposition. In M. T. Turell, M. Spassova, and J. Cicres, (Eds.). *Proceedings of the Second European IAFL Conference on Forensic Linguistics / Language and the Law*, 323—332.

Kendall, T. S. and Wolfram, W. (2009). Local and external standards in African American English. *Journal of English Linguistics,* 37(4), 305—330.

Khawaja, A., Chen, F., Owen, C., and Hickey, G. (2009*).* Cognitive load measurement from user's linguistic speech features for adaptive interaction design. In *Human-Computer Interaction - INTERACT 2009. Lecture Notes in Computer Science: Information Systems and Applications*, Springer Berlin, Heidelberg, 485—489.

Khawaja, A. (2010). Cognitive Load Measurement using Speech and Linguistic Features. *PhD thesis*, University of New South Wales.

Kirchhübel, C. (2009). The effects of Lombard speech on vowel formant measurements. *MSc thesis*, University of York.

Kirchhübel, C., Howard, D. M., and Stedmon, A. W. (2011). Acoustic correlates of speech when under stress: Research, methods and future directions. *The International Journal of Speech, Language and the Law,* 18(1), 75—98.

Kirchhübel, C and Howard, D. M. (2013). Detecting suspicious behavior using speech: Acoustic correlates of deceptive speech – an exploratory investigation. *Applied Ergonomics*, 44(5), 694—702.

Kirchhübel, C., Howard, D. M., and Stedmon, A. (2013). Analysing deceptive speech. In D. Harris, (Ed.). *EPCE'13 Proceedings of the 10th International Conference on Engineering Psychology and Cognitive Ergonomics*. Springer-Verlag Berlin, Heidelberg, 134—141.

Klatt, D. H. and Klatt, C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America,* 87(2), 820—857.

Klaver, J. R., Lee, Z., and Hart, S. D. (2007). Psychopathy and nonverbal indicators of deception in offenders. *Law and Human Behavior,* 31, 337—351.

Knapp, M. L., Hart, R. P., and Dennis, H. S. (1974). An exploration of deception as a communication construct. *Human Communication Research,* 1, 15—29.

Köhnken, G. (1987). Training police officers to detect deceptive eyewitness statements. Does it work? *Social Behaviour,* 2, 1—17.

Koreman, J. (2006). Perceived speech rate: The effects of articulation rate and speaking style in spontaneous speech. *Journal of the Acoustical Society of America*, 119, 582—596.

Köster, S. (2001). Acoustic-phonetic characteristics of hyper-articulated speech for different speaking styles. *Proceedings International Conference on Acoustics, Speech and Signal Processing*, Salt Lake City, 873—876.

Kowal, S. and O'Connell, D. (1980). Pausological research at Saint Louis University. In. H. Dechert and M. Raupach, (Eds.). *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler*. The Hague: Mouton, 61—66.

Kowal, S. and O'Connell, D. (1987). Some temporal aspects of stories told while or after watching a film. *Bulletin of the Psychonomic Society,* 25, 364—366.

Kozel, F. A., Padgett, T. M. & George, M. S. (2004a). A replication study of the neural correlates of deception. *Behavioral Neuroscience* 118: 852—856.

Kozel, F. A., Revell, L. J., Lorberbaum, J. P., Shastri, A., Elhai, J. D., Horner, M. D., Smith, A., Nahas, Z., Bohning, D. E., and George, M. S. (2004b). A pilot study of functional magnetic resonance imaging brain correlates of deception in healthy young men. *Journal of Neuropsychiatry and clinical Neuroscience,* 16, 295—305.

Kozel, F. A., Johnson, K. A., Mu, Q., Grenesko, E. L., Laken, S. J., and George, M. S. (2005). Detecting deception using functional magnetic resonance imaging. *Biological Psychiatry,* 58, 605—613.

Krause, J. C. and Braida, L. D. (2002). Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *Journal of the Acoustical Society of America,* 112, 2165–72.

Krause, J. and Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *Journal of the Acoustical Society of America,* 115(1), 362—378.

Krauss, R. M. (1981). Impression formation, impression management, and nonverbal behaviors. In E. T. Higgins, C. P. Herman, and M. P. Zanna, (Eds.). *Social cognition: The Ontario Symposium*, Vol 1, Hillsdale: Erlbaum, 323—341.

Kraut, R. E. (1978). Verbal and nonverbal cues in the perception of lying. *Journal of Personality and Social Psychology,* 36, 380—391.

Kraut, R. E. (1980). Humans as lie detectors: Some second thoughts. *Journal of Communication,* 30, 209—216.

Kraut, R. E. and Poe, D. (1980). On the line: The deception judgements of customs inspectors and laymen. *Journal of Personality and Social Psychology,* 36, 380—391.

Künzel, H. J. (1974). Eine experimentelle Untersuchung zur Pausenperzeption. *MA thesis*, Universität Kiel.

Künzel, H. J. (1997). Some general phonetic and forensic aspects of speaking tempo. *Forensic Linguistics,* 4(1), 48—83.

Kuroda, I., Fujiwara, O., Okamura, N., and Utsuki, N. (1976). Method for determining pilot stress through analysis of voice communication. *Aviation, Space and Environmental Medicine,* 47, 528—533.

Labov, W. (1966). *The social stratification of English in New York City*. Washington DS: Center for Applied Linguistics.

Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Ladefoged, P. (2006). A course in Phonetics (5^{th} edition). London: Thompson.

Lancaster, G., Vrij, A., Waller, B., and Hope, L. (2013). Sorting the liars from the truth tellers: the benefits of asking unanticipated questions on lie detection. *Applied Cognitive Psychology,* 27(1), 107—114.

Langleben, D. D. (2008). Detection of deception with fMRI: Are we there yet? *Legal and Criminological Psychology,* 13, 1—9.

Langleben, D. D., Schroeder, L., Maldjian, J. A., Gur, R. C., McDonald, S., Ragland, J. D., O'Brien, C. P., and Childress, A. R. (2002). Brain activity during simulated deception: An event-related functional magnetic resonance study. *NeuroImage,* 15, 727—732.

Langleben, D. D., Loughead, J. W., Bilker, W. B., Ruparel, K., Childress, A. R., Busch, S. I., and Gur, R. C. (2005). Telling the truth from lie in individual subjects with fast event-related fMRI. *Human Brain Mapping,* 26, 262—272.

Laver, J. (1980). *The phonetic description of voice quality*. Cambridge: Cambridge University Press.

Laver, J. (1991). *The gift of speech: papers in the analysis of speech and voice*. Edinburgh: Edinburgh University Press.

Levine, T. R., Park, H. S., and McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the 'veracity effect'. *Communication Monographs,* 66, 125—144.

Levine, T. R., Feeley, T. H., McCornack, S. A., Hughes, M., and Harms, C. M. (2005). Testing the effects of nonverbal behavior training on accuracy in deception detection with the inclusion of a bogus training control group. *Western Journal of Communication,* 69, 203—217.

Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In W. J. Hardcastle and A. Marchal, (Eds.). *Speech Production and Speech Modelling*. The Netherlands: Kluwer Academic, 403—439.

Liu, S., Del Rio, E., Bradlow, A. R., and Zeng, F-G. (2004). Clear speech perception in acoustic and electric hearing. *Journal of the Acoustical Society of America,* 116(4), 2374—2383.

Liu, X. (2005). Voice stress analysis: detection of deception. *MSc thesis*, University of Sheffield.

Lively, S. E., Pisoni, D. B., Van Summers, W., and Bernacki, R. H. (1993). Effects of cognitive workload on speech production: Acoustic analysis and perceptual consequences. *Journal of the Acoustical Society of America,* 93, 2962—2973.

Lykken, D. (1998). *A tremor in the blood: Uses and abuses of the lie detector*. Reading, MA: Perseus Publishing.

Maclin, O. H. and Maclin, M. K. (2004). The alias advantage and perceptions of guilt. *Journal of Psychology,* 138, 339—349.

Mann, S. (2001). Suspects, lies and videotape: An investigation into telling and detecting lies in police/suspect interviews. *PhD thesis*, University of Portsmouth.

Mann, S., Vrij, A., and Bull, R. (2002). Suspects, lies and videotape: An analysis of authentic high-stakes liars. *Law and Human Behavior,* 26, 365—376.

Mann, S., Vrij, A., and Bull, R. (2004). Detecting true lies: Police officers' ability to detect deceit. *Journal of Applied Psychology,* 89, 137—149.

Marteau, T. M. and Bekker, H. (1992). The development of a six-item short form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology,* 31, 301—306.

Masip, J. (2006). The social psychology of the detection of deception. *Law and Human Behavior,* 30, 403—407.

Masip, J. (2005). Is the liar caught sooner than the cripple? Popular wisdom versus scientific knowledge on the non-verbal detection of deception. *Papeles del Psicólogo,* 26, 78—91.

Masip, J., Garrido, E., and Herrero, C. (2004). Defining deception. *Anales de psicología,* 20(1), 147—171.

McCornack, S. A. (1992). Information manipulation theory. *Communication Monographs,* 59, 1—16.

McDougall, K. (2006). Dynamic features of speech and the characterisation of speakers: towards a new approach using formant frequencies. *International Journal of Speech, Language and the Law,* 13(1), 89—126.

McDougall, K., Nolan, F., Harrison, P., and Kirchhübel, C. (2012). Characterising speakers using formant frequency information: A comparison of vowel formant measurements and long-term formant analysis. *Proceedings International association of Acoustics and Forensic Phonetics (IAFPA),* Santander, Spain.

Mehrabian, A. (1971). Nonverbal betraying of feeling. *Journal of Experimental Research in Personalit*y, 5, 64—73.

Meinerz, C. (2010). *Effekte von Stress auf Stimme und Sprechen: Eine phonetische Untersuchung auf der Grundlage ausgewählter akustischer und sprecherdynamischer Parameter unter Berücksichtigung verschiedener Stressklassen*. Norderstedt: Books on Demand GmbH.

Meinerz, C. and Masthoff, H. (2011). Effect of telephone-line transmission and digital ausio format on formant tracking measurements. *Presentation delivered to the 20th Annual Conference of the International Association of Forensic Phonetics and Acoustics* (IAFPA), Trier, Germany.

Meinerz, C. and Masthoff, H. (2012). Effect of telephone-line transmission and digital ausio format on formant tracking measurements - revisited. *Presentation delivered to the 21th Annual Conference of the International Association of Forensic Phonetics and Acoustics* (IAFPA), Santander, Spain.

Mendoza, E. and Carballo, G. (1998). Acoustic analysis of induced vocal stress by means of cognitive workload tasks. *Journal of Voice,* 12, 263—273.

Metts, S. (1989). An exploratory investigation of deception in close relationships. *Journal of Social and Personal Relationships,* 6, 159—179.

Miller, G. R. (1983). Telling it like it isn't and not telling it like it is: Some thoughts on deceptive communication. In J. I. Sisco, (Ed.). *The Jensen lectures: Contemporary communication studies*. Tampa: University of South Florida, 91—116.

Miller, G. R. and Stiff, J. B. (1993). Deceptive communication. Newbury Park, CA: Sage.

Miller, A. (2007). Was That the Truth? Temporal Vocal Cues and the Perception of Deception. *BA thesis*, Haverford College, US.

Milne, R. and Bull, R. (1999). *Investigative interviewing: psychology and practice*. Chichester: John Wiley & Sons Ltd.

Moghadam, A. (2003). Palestinian suicide terrorism in the second intifada: Motivations and organizational aspects. *Studies in Conflict and Terrorism, 26*(2), 65—92.

Montaigne, M. The Essays of Montaigne 1877, complete. Project Gutenberg's The Essays of Montaigne, Complete. Published in 2006. http://www.gutenberg.org/files/3600/3600-h/3600-h.htm#link2HCH0009.

Moon, S. J. and Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustical Society of America,* 96, 40–55.

Moos, A. (2008). Forensische Sprechererkennung mit der Messmethode LTF (long-term formant distribution). *MA thesis*, Universität des Saarlandes.

Moston, S. J. and Engelberg, T. (1993). Police questioning techniques in tape recorded interviews with criminal suspects. *Policing and Society,* 6, 61—75.

Motley, M. T. (1974). Acoustic correlates of lies. *Western Speech,* 38, 81—87.

Müller, C., Großmann-Hutter, B., Jameson, A., Rummer, R., and Wittig, F. (2001). Recognizing time pressure and cognitive load on the basis of speech: An experimental study. In J. Vassileva, P. Gmytrasiewicz, and M. Bauer, (Eds.). (2001). *UM2001, User modeling: Proceedings of the 8th International Conference*. Berlin: Springer.

National Research Council (2003). *The polygraph and lie detection*. Committee to review the scientific evidence on the Polygraph. Washington, DC: The National Academic Press.

Ni Chasaide, A. and Gobl, C. (1997). Voice source variation. In W. J. Hardcastle and J. Laver, (Eds.). *The Handbook of Phonetic Sciences.* Oxford: Blackwell, 427—461.

Nieuwenhuis, S., De Geus, E. J., and Aston-Jones, G. (2011). *Psychophysiology,* 48, 162–175.

Nolan, F. (1983). *The phonetic bases of speaker recognition*. Cambridge: CUP.

Nolan, F., McDougall, K., de Jong, G., and Hudson, T. (2009). The DyViS database: style-controlled recordings of 100 homogeneous speakers for forensic phonetic research. *International Journal of Speech, Language and the Law,* 16(1), 31—57.

Noaln, F. and Grigoras, C. (2005). A case for formant analysis in forensic speaker identification. *International Journal of Speech, language and the Law,* 12(2), 143—173.

O'Brien, K. A. (2008). Assessing Hostile Reconnaissance and Terrorist Intelligence Activities: The Case for a Counter Strategy. The *RUSI Journal,* 153(5), 34—39.

Ofshe, R. J. and Leo, R. A. (1997). The decision to confess falsely: Rational choice and irrational action. *Denver University Law Review,* 74, 979—1112.

O'Hair, H. D., Cody, M., and McLaughlin, M. L. (1981). Prepared lies, spontaneous lies, Machiavellianism and nonverbal communication. *Human Communication Research,* 7, 325—339.

O'Hair, D. and Cody, M. J. (1994). Deception. In W. R. Cupach and B. H. Spintzberg, (Eds.). *The dark side of interpersonal communication.* Hillsdale, NJ: Lawrence Erlbaum, 181—213.

Oomen, C. C. E. and Postma, A. (2001). Effects of divided attention on the production of filled pauses and repetitions. *Journal of Speech, Language, and Hearing Research,* 44, 997–1004.

Papp, V. (2008). The effects of heroin on speech and voice quality. *MSc thesis*, University of York, UK.

Parliament, L. and Yarmey, A. D. (2002). Deception in eyewitness identification. *Criminal Justice and Behavior,* 29, 734—746.

Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. Oxford: Oxford University Press.

Pedahzur, A., Perliger, A., and Weinberg, L. (2003). Altruism and fatalism: the characteristics of Palestinian suicide terrorists. *Deviant Behavior, 24*(4), 405—423.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (1986). Speaking clearly for the hard of hearing II: intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research,* 29, 434—446.

Plotkin, A. (1997). http://www.eblong.com/zarf/werewolf.html.

Police and Criminal Evidence Act (1986). Available at https://www.gov.uk/police-and-criminal-evidence-act-1984-pace-codes-of-practice

Porges, S. W. (2006). The Polygraph: One machine, two world views. In A. K. Webb and J. C. Kircher, (Eds.). Proceedings of the Workshop on the Use of Autonomic and Somatic Measures for Security Evaluations. *The Journal of Credibility Assessment and Witness Psychology,* 7(2), 47—73.

Porter, S. B. (1994). The language of deceit: are there reliable verbal cues to deception in the interrogation context? *PhD thesis*, The University of British Columbia.

Porter, S. and Yuille, J. C. (1996). The language of deceit: An investigation of the verbal clues to deception in the interrogation context. *Law and Human Behavior,* 20, 443—459.

Porter, S., Doucette, N. L., Woodworth, M., Earle, J., and MacNeil, B. (2008). 'Halfe the world knows not how the other halfe lies': Investigation of cues to deception exhibited by criminal offenders and non-offenders. *Legal and Criminological Psychology,* 13(1), 27—38.

Porter, S., Juodis, M., ten Brinke, L. M., Klein, R., and Wilson, K. (2010). Evaluation of the effectiveness of a brief deception detection training program. *The Journal of Forensic Psychiatry and Psychology,* 21(1), 66—76.

Post, J., Sprinzak, E., and Denny, L. (2003). The terrorists in their own words: Interviews with 35 incarcerated Middle Eastern terrorists. *Terrorism and Political Violence, 15*(1), 171—184.

Protopapas, A. and Lieberman, P. (1997). Fundamental frequency of phonation and perceived emotional stress. *Journal of the Acoustical Society of America,* 101, 2267—2277.

Remijsen, B. (2004). http://www.lel.ed.ac.uk/~bert/msr&check_spectr_indiv_interv.psc.

Rhodes, R. W. (2012). Assessing the strength of non-contemporaneous forensic speech evidence. *PhD thesis,* submitted to the Department of Language and Linguistic Science, University of York.

Rickford, J. and McNair-Knox, F. (1994). Addressee- and topic- influenced style shift: A quantitative sociolinguistic study. In D. Biber and E. Finegan, (Eds.). Sociolinguistic Perspectives on Register. New York: Oxford University Press, 235—276.

Riggio, R. E. and Friedman, H. S. (1983). Individual differences and cues to deception. *Journal of Personality and Social Psychology,* 45, 899—915.

Riggio, R. E. (1986). Assessment of basic social skills. *Journal of Personality and Social Psychology,* 51, 649—660.

Rochester, S. R. (1973). The significance of pauses in spontaneous speech. *Journal of Psycholinguistic Research,* 2(1), 51—81.

Rockwell, P., Buller, D. B., and Burgoon, J. K., (1997a). The voice of deceit: Refining and expanding vocal cues to deception. *Communication Research Reports*, 14(4), 451-459.

Rochwell, P., Buller, D. B., and Burgoon, J. K. (1997b). Measurement of deceptive voices: Comparing acoustic and perceptual data. *Applied Psycholinguistics,* 18, 471—484.

Rosenfeld, J. P. (2002). Event-related potentials in the detection of deception, malingering, and false memories. In M. Kleiner, (Ed.). *Handbook of Polygraph testing*. New York: Academic Press, 265—286.

Rosenfeld, J. P. (2011). P300 in Detecting Concealed Information. In B. Verschuere, G. Ben Shakhar, and E. Meijer, (Eds.). *Memory detection: Theory and application of the concealed information test. Cambridge:* Cambridge University Press, 63—89.

Roßnagel, C. (1995). Übung und Hörerorientierung beim monologischen Instruieren: Zur Differenzierung einer Grundannahme. *Sprache und Kognition,* 14, 16-26.

Ruiz, R., Absil, E., Harmegnies, B., Legros, C., and Poch, D. (1996). Time- and spectrum-related variabilities in stressed speech under laboratory and real conditions. *Speech Communication,* 20, 111—129.

Rummer, R. (1996). *Kognitive Beanspruchung beim Sprechen*. Weinheim: Beltz.

Sapir, A. SCAN. Available at www.lsiscan.com.

Scherer, K. R. (1979) Non-linguistic vocal indicators of emotion and psychopathology. In C. E. Izard (Ed.) *Emotions in Personality and Psychopathology*. New York: Plenum Press, 495—529.

Scherer, K. R. (1981). Vocal indicators of stress. In J. K. Darby, (Ed.). *Speech Evaluation in Psychiatry.* New York: Grune and Stratton, 171—187.

Scherer, K. R. (1985). Vocal cues to deception: a comparative channel approach. *Journal of Psycholinguistic Research,* 14(4), 409—425.

Scherer, K. R., Grandjean, D., Johnstone, T., Klasmeyer, G. and Bänziger, T. Acoustic correlates of task load and stress. *Proceedings, ICSLP*, 2002, Denver, USA, 2017—2020.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication,* 40, 227—256.

Schilling-Estes, N. (2008). Stylistic variation and the sociolinguistic interview: A reconsideration. In R. Monroy and A. Sánchez, (Eds.). *25 Años de Lingüística Aplicada en España: Hitos y Retos (25 Years of Applied Linguistics in Spain: milestones and challenges; Proceedings from AESLA 25*. Murcia: Servicio de Publicaciones de la Universidad de Murcia.

Schilling, N. (2013). Investigating stylstic variation. In. J. K. Chambers and N. Schilling, (Eds.). *The Handbook of Language Variation and Change, 2nd edition*. Oxford: Wiley-Blackwell.

Schlenker, B. R. and Leary, M. R. (1982). Social anxiety and self-presentation: A conceptualisation and model. *Psychological Bulletin,* 92, 641—669.

Seymour, T. L., Seifert, C. M., Shafto, M. G., and Mosmann, A. L. (2000). Using response time measures to assess 'guilty knowledge'. *Journal of Applied Psychology,* 85, 30—37.

Shepherd, E. (2007). *Investigative Interviewing: The conversation management approach*. Oxford: University Press Oxford.

Sheridan, M. R. and Flowers, K. A. (2010). Reaction times and deception - the lying constant. *International Journal of Psychological Studies,* 2(2), 41—51.

Siegman, A. W. (1993). Paraverbal correlates of stress: implications for stress identification and management. In L. Goldberger and S. Breznitz, (Eds.). Handbook *of stress. Theoretical and clinical aspects* (2nd edition). New York: The Free Press, 274—299.

Siegman, A. W. and Reynolds, M. A. (1983). Self-monitoring and speech in feigned and unfeigned lying. *Journal of Personality and Social Psychology,* 45, 1325—1333.

Sigmund, M. (2006). Introducing the database ExamStress for speech under stress. *Proceedings of the 7th Nordic Signal Processing Symposium*, Rejkjavik, 290—293.

Silberstein, D. and Dietrich, R. (2003). Cockpit communication under high cognitive workload. In D. Dietrich, (Ed.). *Communication in high risk environments*. Hamburg: Helmut Buske Verlag, 9—56.

Silke, A. (2003). Becoming a terrorist. In A. Silke, (Ed.). *Terrorists, victims and society: Psychological perspectives on terrorism and its consequences*. Chichester: Wiley.

Sip, K. E., Roepstorff, A., McGregor, W., and Frith, C. D. (2007). Detecting deception: the scope and limits. *Trends in Cognitive Science,* 12(2), 48—53.

Smiljanić, R. and Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *Journal of the Acoustical Society of America,* 118(3), 1677–88.

Smiljanić, R. and Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and Linguistics Compass,* 3, 236—264.

Sound Forge™ Pro 10. *Sony Creative Software*, © 2003-2011.

Sokolov, A. N. (1963). *Perception and the conditioned reflex*. Oxford: Pergamon Press.

Spence, S. A. (2004). The deceptive brain. *Journal of the Royal Society of Medicine,* 97, 6—9.

Spence, S. A., Farrow, T. F. D., Herford, A. E., Wilkinson, I. D. Zheng, Y., and Woodruff, P. W. R. (2001). Behavioural and functional anatomical correlates of deception in humans. *Neuroreport: For Rapid Communication of Neuroscience Research,* 12, 2849—2853.

Spence, S. A., Hunter, M. D., Farrow, T. F. D., Green, R. D., Leung, D. H., Hughes, C. J., and Ganesan, V. (2004). A cognitive neurobiological account of deception: evidence from functional neuroimaging. *Philosophical Transactions of the Royal Society of London,* 359, 1755—1762.

Sporer, S. L. and Schwandt, B. (2006). Paraverbal indicators of deception: A meta-analytic synthesis. *Applied Cognitive Psychology,* 20, 421—446.

Sporer, S. L. and Schwandt, B. (2007). Moderators of nonverbal indicators of deception: a meta-analytic synthesis. *Psychology, Public Policy, and Law,* 13(1), 1—34.

Steller, M. and Koehnken, G. (1989). Criteria Based Content Analysis. In D. C. Raskin, (Ed.). *Psychological methods in criminal investigation and evidence*. New York: Springer-Verlag, 217—245.

Stiff, J. B. and Miller, G. R. (1986). 'Come to think of it...': Interrogative probes, deceptive communication, and deception detection. *Human Communication Research,* 12, 339—357.

Stiff, J. B., Kim, H. J., and Ramesh, C. N. (1992). Truth biases and aroused suspicion in relational deception. *Communication Research,* 19, 326—345.

Stiff J. B., Corman, S., Krizek, B., and Snider, E. (1994). Individual differences and changes in nonverbal behaviour. Unmasking the changing faces of deception. *Communication Research,* 21, 555—581.

Street, R. L. (1984). Speech convergence and speech evaluation in fact-finding interviews. *Human Communication Research, 11*, 139—169.

Streeter, L. A., Krauss, R. M., Geller, V., Olson, C., and Apple, W. (1977). Pitch changes during attempted deception. *Journal of Personality and Social Psychology,* 35, 345—350.

Streeter, L. A., Macdonald, N. H., Apple, W., Krauss, R. M., and Galati, K. M. (1983). Acoustic and perceptual indicators of emotional stress. *Journal of the Acoustical Society of America,* 73, 1354—1360.

Strömwall, L. A., Granhag, P. A., and Hartwig, M. (2004). Practitioner's beliefs about deception. In P. A. Granhag and L. A. Strömwall, (Eds.). *Deception detection in forensic contexts*. Cambridge: Cambridge University Press, 229—250.

Strömwall, L. A., Hartwig, M., and Granhag, P. A. (2006). To act truthfully: Nonverbal behaviour and strategies during a police interrogation. *Psychology, Crime and Law,* 12(2), 207—219.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology,* 18(6), 643–662.

Suzuki, A., Watanabe, S., Taheno, Y., Kosugi, T., and Kasuya, T. (2002). Possibility of detecting deception by voice analysis. *Polygraph*, 31(2), 129—134.

Tauroza, S. and Allison, D. (1990). Speech rates in British English. *Applied Linguistics*, 11, 90—115.

Tolkmitt, F. J. and Scherer, K. R. (1986). Effect of experimentally induced stress on vocal parameters. *Journal of Experimental Psychology,* 12(3), 302—313.

Torres, J. F., Moore, E., and Bryant, E. (2008). A study of glottal waveform features for deceptive speech classification. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 4489–4492.

Torres, J. F. (2010). Estimation of glottal source features from the spectral envelope of the acoustic speech signal. *PhD thesis*, Georgia Institute of Technology.

Trouvain, J. (2004). Tempo variation in speech production: Implications for speech synthesis. *PhD thesis,* Universität des Saarlandes.

Trovillio, P. V. (1939a). A history of lie detection. *Journal of Criminal Law and Criminology,* 29(6), 848—881.

Trovillio, P. V. (1939b). A history of lie detection (concluded). *Journal of Criminal law and Criminology,* 30(1), 104—119.

Turner, R. E., Edgley, C., and Olmstead, G. (1975). Information control in conversations: Honesty is not always the best policy. *Kansas Journal of Sociology,* 11, 69—89.

Turner, G. S., Tjaden, K., and Weismer, G. (1995). The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis. *Journal of Speech, Language and Hearing Research,* 38(5), 1001—1003.

Uchanski, R. M., Choi, S. S., Braida, L. D., Reed, C. M., and Durlach, N. I. (1996). Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *Journal of Speech and Hearing Research,* 39, 494—509.

Uchanski R. M. (2005). Clear Speech. In D. B. Pisoni and R. E. Remez, (Eds). *Handbook of Speech Perception*. Blackwell Publishers, Malden, MA, 207—235.

Van Lierde, K., van Heule, S., De Ley, S., Mertens, E., and Claeys, S. (2009). Effect of psychological stress on female vocal quality- a multiparameter approach. *Folia Phoniatrica et Logopaedica,* 61, 105—111.

Vendemia, J. M. C., Buzan, R. F., and Green, E. P. (2005). Practice effects, workload, and reaction time in deception. *American Journal of Psychology,* 118, 413—429.

Verschuere, B., Crombez, G., De Clecq, A., and Koster, E. H. W. (2004a). Autonomic and behavioural responding to concealed information: Differentiating orienting and defensive responses. *Psychophysiology,* 41, 461—466.

Verschuere, B., Crombez, G., and Koster, E. H. W. (2004b). Orienting to guilty knowledge. *Cognition and Emotion,* 18, 265—279.

Vilkman, E. and Manninen, O. (1986). Changes in prosodic features of speech due to environmental factors. *Speech Communication,* 5, 331—345.

Vrij, A. (1995). Behavioural correlates of deception in a simulated police interview. *Journal of Psychology,* 129, 15—28.

Vrij, A. (2008). *Detecting lies and deceit: pitfalls and opportunities* (2nd edition)West Sussex: Wiley.

Vrij, A. and Winkel, F. W. (1991). Cultural patterns in Dutch and Surinam nonverbal behavior: An analysis of simulated police/citizen encounters. *Journal of Nonverbal Behavior,* 15, 169—184.

Vrij, A., Semin, G. R., and Bull, R. (1996). Insight into behavior displayed during deception. *Human Communication Research,* 22, 544—562.

Vrij, A., Akehurst, L., and Morris, P. (1997). Individual differences in hand movements during deception. *Journal of Nonverbal Behavior,* 21(2), 87—102.

Vrij, A. and Heaven, S. (1999). Vocal and verbal indicators of deception as a function of lie complexity. *Psychology, Crime and Law,* 5(3), 203—215.

Vrij, A., Edward, K., Roberts, K. P., and Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behaviour. *Journal of Nonverbal Behavior,* 24, 239—263.

Vrij, A., Edwards, K., and Bull, R. (2001a). Stereotypical verbal and nonverbal responses while deceiving others. *Personality and Social psychology Bulletin,* 27, 899—909.

Vrij, A., Edward, K., and Bull, R. (2001b). People's insight into their own behaviour and speech content while lying. *British Journal of Psychology,* 92, 373—389.

Vrij, A. and Mann, S. (2001). Telling and detecting lies in a high-stake situation: The case of a convicted murderer. *Applied Cognitive Psychology,* 15, 187—203.

Vrij, A., Edward, K., and Bull, R. (2001). Stereotypical verbal and nonverbal responses while deceiving others. *Personality and Social Psychology Bulletin,* 27, 899—909.

Vrij, A., Akehurst, L., Soukara, S., and Bull, R. (2004). Detecting deceit via analyses of verbal and nonverbal behaviour in children and adults. *Human Communication Research,* 30 (1), 8—41.

Vrij, A., Mann, S., and Fisher, R. (2006a). An empirical test of the behaviour analysis interview. *Law and Human Behavior,* 30(3), 329—345.

Vrij, A., Mann, S., and Fisher, R. (2006b). Information-gathering vs. accusatory interview style: individual differences in respondents' experiences. *Personality and Individual Differences,* 41(4), 589—599.

Vrij, A., Mann, S. A., Fisher, R. P., Leal, S., Milne, R., and Bull, R. (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order. *Law and Human Behaviour,* 32, 253—265.

Vrij, A., Leal, S., Granhag, P., Mann, S., Fisher, R., Hillman, J., and Sperry, K. (2009) Outsmarting the liars: the benefit of asking unanticipated questions. *Law and Human Behavior,* 33, 159—166.

Vrij, A. and Granhag, P. (2012). Eliciting cues to deception and truth: What matters are the questions asked. *Journal of Applied Research in Memory and Cognition,* 1, 110—117.

Wagner, K. (2010). Der Einfluss der Sprechlautstärke auf die ersten drei Vokalformanten in mobilfunkübertragener Sprache: Forensischer Stimmenvergleich anhand der LTF-Methode (long-term formant distribution). *BA thesis*, Universität Frankfurt/M.

Walczyk, J. J., Roper, K. S., Seemann, E., and Humphrey, A. M. (2003). Cognitive mechanisms underlying lying to questions: Response time as a cue to deception. *Applied Cognitive Psychology,* 17, 744—755.

Walcyyk, J. J., Schwartz, J. P., Clifton, R., Adams, B., Wei, M., and Zha, P. (2005). Lying person-to-person about live events: A cognitive framework for lie detection. *Personnel Psychology,* 58, 141—170.

Walczyk J. J., Mahoney K. T., Doverspike D., and Griffith-Ross D. A. (2009). Cognitive lie detection: response time and consistency of answers as cues to deception. *Journal of Business and Psychology*, 24, 33—49.

Wang, G., Chen, H., and Atabakhsh, H. (2004). Criminal identity deception and deception detection in law enforcement. *Group Decision and Negotiation,* 13, 111—127.

Warmelink, L., Vrij, A., Mann, S., Jundi, S., and Granhag, P. (2012). The effect of question expectedness and experience on lying about intentions. *Acta Psychologica,* 141, 178—183.

Waters, J., Nunn, S., Gillcrist, B., and Von Colln, E. (1995). The effect of stress on the glottal pulse. In I. Trancoso and R. Moore, (Eds.). *Proceedings ESCA-NATO Tutorial and Research Workshop on Speech under Stress*, Lisbon, Portugal.

Wayland, R. and Jongman, A. (2003). Acoustic correlates of breathy and clear vowels: the case of Khmer. *Journal of Phonetics*, 31, 181—201.

Webb, A. K. and Kircher, J. C. (2006). Proceedings of the Workshop on the Use of Autonomic and Somatic Measures for Security Evaluations. *The Journal of Credibility Assessment and Witness Psychology,* 7(2), 17—168.

Wells, J. C. (1982). *Accents of English I: An introduction*. Cambridge: Cambridge University Press.

Williams, C. E. and Stevens, K. N. (1969). On determining the emotional state of pilots during flight: An exploratory study. *Aerospace Medicine,* 40, 1369—1372.

Williams, C. E. and Stevens K. N. (1972). Emotions and speech: Some acoustical correlates. *Journal of the Acoustical Society of America,* 52, 1238—1250.

Williams, L., Brammer, M., Skerrett, D., Lagapolous, J., Rennie, C., Kozek, K., Oliveri, G., and Peduto, T. (2000). The neural correlates of orienting: An integration of fMRI and skin conductance orienting. *Brain Imaging,* 11(13), 3011—3015.

Yap, T. F., Epps, J., Choi, E. H. C., and Ambikairajah, E. (2010). Glottal features for speech-based cognitive load classification. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 5234–5237.

Yap, T. F., Epps, J., Ambikairajah, E., and Choi, E. H. C. (2011). Formant frequencies under cognitive load: Effects and classification. *EURASIP Journal on Advances in Signal Processing*,

Zhou, L. and Zhang, D. (2006). A comparison of deception behaviour in dyad and triadic group decision making in synchronous computer-mediated communication. *Small Group Research,* 37, 140—164.

Zuckerman, M., DePaulo, B. M., and Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz, (Ed.). *Advances in experimental social psychology* volume 14. New York: Academic Press, 1—59.

Zuckerman, M. Amidon, M. D., Bishop, S. E., and Pomerantz, S. D. (1982). Face and tone of voice in the communication of deception. *Journal of Personality and Social Psychology,* 43(2), 347—357.