COMMUNICATION SKILLS TRAINING INTERVENTION BASED ON AUTOMATED RECOGNITION OF HUMAN EMOTION AND NON-VERBAL BEHAVIOUR

A thesis submitted for the degree of Doctor of Philosophy

By

Monica Pereira

Department of Computer Science

College of Engineering, Design and Physical Sciences

Brunel University London

2019

Supervised by: Professor Kate Hone

Second Supervisor: Dr Hongying Meng

This PhD was jointly funded by BAE Systems and Engineering and Physical Sciences Research Council

Contents

ACKNOWLEDGEMENTS

Delivery of this thesis would not be possible without the support and guidance of my supervisor, Professor Kate Hone. The support provided is unparalleled to any that I have ever received or could have ever expected. Kate has helped me develop into a better researcher with a mind suited for the beginning stages of an academic career. Any words used to thank her could never be enough to show my appreciation for everything. Furthermore, observing Kates progression in the time that I have known her has been very influential and motivating to say the least. Kate is a very inspiring and patient person!

I would like to thank Dr Chris Chandler for his devoted help throughout my whole academic career. To me, Chris has been a lecturer, a supervisor and a friend. Throughout all the roles he has played in my life, he has been influential. You taught me how to love research and my decisions up until now have been a credit to you.

I would also like to thank my family and partner for all their support and patience which has helped me to the point of submission. Thank you to my husband to be, for understanding the endless hours I spent at my desk and for creating a safe environment for me to succeed. To my mother, for her endless advice, support and unconditional love. To my sister, my voice of reason, my strength and my light thank you for being everything I need you to be. To my father, for always listening, never judgmental and using humour to repeatedly about my research which has helped me to articulate my research in a way that is understandable to everyone. Finally, to my extended family thank you for your patience and understanding when I was not able to attend family events.

I would like to thank Dr Inmaculada Andres for her continued support and motivational words throughout my thesis. The nature of her character is timeless and highly influential, she is a credit to anyone who knows her. I would also like to thank Dr Federico Colecchia for all the guidance and support throughout my PhD, even after leaving Brunel University London. I would also like to thank Dr Helen Dudfield and the team at QinetiQ for research collaboration on this project. My thanks extend to the funders of my PhD (BAE Systems and Engineering, Physical Sciences Research Council) and the Defence, Science, Technology Lab. Finally, I would like to thank all the people who have been critical about my work, offered advice, offered opportunities to present my work and provided general support. Names that should be mentioned is Dr Crina Grosan, Professor Mark Perry, Professor Martin Shepperd, Bhaveet Nagaria, Evie Xinos and Dr Nour Ali. The list goes on…

Declaration / Date of Submission

I hereby declare that this thesis is my own work that was conducted in accordance with the University Code of Research Ethics. I have also completed the necessary training requirements that are associated with the PhD completion: Research Methods, Research Development Series 1, 2 and 3.

Monica Pereira

21st October 2019

LIST OF FIGURES

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

LIST OF TABLES

ABSTRACT

Introduction: Across multiple sectors training programmes aim to help learners improve their communication skills. It is well recognised that non-verbal 'social signals' play an important role in effective communication. Previous research in the social signalling domain meticulously observed hours of videos and conducted observational studies to identify these social signals, an approach which is subjective and does not scale with large datasets. Technological developments hold promise to automate observation with possible practical application to training interventions. **Objective**: Therefore, the aim of the current research is to investigate whether communication skills can be improved using recently developed commercial off-the shelf technology to capture facial expression, voice emotion recognition, hand gestures and *honest* signals. **Methods**: Four stages of research were conducted. The first stage was to establish the relevant signals for performance appraisal in media interviews. The second stage was to identify the most appropriate method of providing feedback to trainees that is actionable and understandable. The third stage was to compare whether the designed social signal feedback method was more effective than standard methods of communication skills training in the context of media interviews. Finally, the fourth stage was to assess whether the skills gained in stage three was maintained after 6-months. Performance ratings were collected by an audience who were blind to experimental condition and conversational partners / trainers. **Results**: Performance ratings collected from the experiment and follow-up stages suggest that the social signal feedback group were more effective communicators compared to the traditional feedback group. The social signal feedback group displayed a significant reduction in frowning in the experiment stage and more positive emotions in the follow-up stage. However, the traditional feedback group exhibited more positive engagement during interviews. **Conclusion:** The social signal feedback method presented has some benefit over already existing methods of communication skills in media interviews.

Keywords: Emotional Communication, Social Signals Processing, Automated Emotion Recognition, Communication Skills Training, Affective computing, commercial off-the-shelf affect recognition technology

CHAPTER 1. SCOPE OF RESEARCH

1.1. Introduction

This thesis explores whether commercial-off-the-shelf automated recognition (COTS) technology can be used to improve communication skills training in the context of media skills training. This chapter provides the following information; an outline of communication and social signals and their role in human interaction research; the importance of communication in everyday life is also articulated along with a account of communication skills training in the U.K; the main objectives and contributions of this research, a brief description of the plan of this thesis and, finally, concludes with a summary.

1.2. Communication

Communication is a complex phenomenon that occurs in everyday life and is comprised of verbal and nonverbal communication. Research has shown that non-verbal / social signals makes up 65% - 95% of communication (Curhan & Pentland, 2007). It is theorised that they communicate an individual's unconscious emotions meaning that the use of nonverbal signals are significant in understanding social interactions (Liu, Mok, Wong, Xue, & Xu, 2007; Poggi & D 'errico, 2011; Reason, 1990; Tracy, Randles, Steckler, Crockett, & Cuddy, 2015; Vinciarelli, Pantic, & Bourlard, 2009; Vrij, Edward, & Roberts, 2000). Considering non-verbal communication accounts for most of the communicative content in a social interaction, it is important that researchers explore this further in the context of training. Typically, the focus of communication training research places emphasis on improving both verbal and nonverbal behaviours where performance evaluation is typically conducted by a single trainer. This limits the efficacy of training as research has shown that skills gained are highly reliant on the trainers' experience which often tends to be subjective (Aspegren, 1999). For this reason, it is beneficial for training programs to integrate accurate and objective observations of nonverbal cues to improve performance. Current tools are limited to support this requirement.

1.3. Affective Computing and Social Signal Processing in Human-Computer Interaction

Human-Computer Interaction (HCI) is a multidisciplinary practice concerned with the design, construction and implementation of human-centric interactive computer systems which uses a multitude of formal methods that include both quantitative and qualitative methods to capture information and ideas (MacKenzie, 2012).

Within the field of HCI is a topic called affective computing which can simply be described as a discipline which uses devices that can recognise, interpret and process human emotion (Picard, 1995). This is an interdisciplinary field that integrates computer science, cognitive science and psychology. Social Signals Processing (SSP) is a domain within affective computing that captures nonverbal signals using affect recognition technologies with the objective of modelling human behaviour. The reason this area hopes to model behaviour it to eventually develop emotionally intelligent computers.

The research conducted for this PhD thesis lies within the area of affective computing and SSP because affect recognition technology is used to capture social signals between two humans. Detection, analysis and interpretation of a dyadic interaction in this research is intended to achieve two outcomes: 1) to identify which signals predict communication performance and 2) this information could inform the design of an intervention that could support training.

1.4. Communication Training in the UK

Organisations make substantial investments to improve communication between employees (Brantley & Miller, 2007). Effective communication is the core of a well-run establishment as it results in better productivity which fundamentally benefits a country's economy (Crawford, Crawford, & Jin, 2013). However, in the UK, a report by McDonalds in 2015 identified a lack of training in soft skills and predicts that by 2020 over half a million UK workers will be significantly held back in their career as a result. This prediction is made for all sectors across the UK (Forte & Caan, 2015). The British Chambers of Commerce survey (2014) reported that 84% of companies look for communications skills training when recruiting new employees (Wiseman, Parry, & Baker, 2015). However, investment in skills training in the UK have fallen by £2.5 billion since 2011 across all sectors even though it is a well sought-after skill (Thornton, Sutton, Stanfield, & Leach, 2016). These reports emphasise a skills gap in the UK.

There are several factors that contribute to this decline, but the most important to highlight is that training is costly to run, and it is taxing on company production to spare employees for this training (Fallowfield, Jenkins, Farewell, & Solis-Trapala, 2003). This issue can be solved by providing easily accessible training at a lower cost. The current research investigates whether this is possible using off-the-shelf technology that is commercially available.

1.5. Communication Context

The context of a social interaction depicts the nature of the conversation and the nonverbal signals used. A unique communication context that is important to investigate is media interviews. This is because conversations are required to be concise and effective. Training an interviewee to effectively communicate in this context in beneficial as information relayed in an interview is likely to reach a large audience (Taylor, 2015). It is key to conveying a clear and concise message, particularly in hostile or crisis situations in order for the audience to trust the interviewee (World Health Organization, 2005).

Training in skilful communication is also a necessity in a variety of job roles and organisations across all sectors (Logan-Terry & Damari, 2015). Organisations make considerably large investments to train staff in communication skills to improve media interview performance. This is done to ensure that employees are positively perceived in media interviews because the way that employees are perceived in media interviews influences the audience's perception of the company that the interviewee represents (Taylor, 2015). There has been little or no research in this area in the field of SSP and is the focus of the current research will be to investigating communication skills training in the context of media skills training.

1.6. Appropriate Technology for a Fast-Paced Society

Common methods used by social psychologists during the early years of examining social interactions were to manually observe hours of pre-recorded video sessions of social interactions. This form of analysis is time-consuming which leaves room for subjectivity and is likely to result in different interpretations of meanings behind signals displayed. Video observations also do not scale with large amounts of data. An example of this would be that hours of video observations could result in some aspects being neglected, left without an annotation or interpretation. This method of video observation is still common today; however, some subjectivity is reduced with the inclusion of multiple observers. Recent developments in automated recognition technology have made it possible to objectively capture nonverbal signals in interactions.

The concept of 'appropriate technology' is the use of simple and 'easy to use' technology that makes it possible for people to use every day (Basu & Weil, 1998). Technologies developed in the area of social signals include affect recognition technology, examples include facial expression detection, voice affect recognition and movement detection. These technologies have recently been made commercially available, otherwise known as Commercial Off-The-Shelf (COTS) Technology. The use of this type of technology could enable a faster and more reliable method of improving communication skills in the UK and, as a consequence of this simple and cost-effective method of training, boost training turnout.

1.7. Thesis Statement

It was decided to use COTS technology to capture social signals, rather than use already existing open source software or develop bespoke solutions. This technology was used to provide rapid proof of concept for a variety of channels in the evaluation of communication skills performance and enable fast transferability to end users. The technology used can also be obtained commercially thereby assisting in narrowing the design space for future bespoke solutions. The focus of this research is on the functional applications of solutions developed by affective technology. The use of affective technology allows the user to process recordings locally when the classification of emotions and expression are produced (Dupré, Morrison, & Mckeown, 2018).

Largely, the benefits of this could be to allow for better access to training, more objective feedback during training and lowered costs of training courses. This would be independent of trainer's opinion. This research hopes to address the following research questions:

1) Can recently developed automated technology be used to help evaluate media skills performance?
2) Can this be used to provide feedback that helps people improve their communication?

1.8. Research Contribution

By utilising the recently developed automated affect technology in this research, a scalable alternative that gives rise to the possibility of faster and more objective measurement of social signals is proposed for the field of media skills training. Developing a fast and objective method of providing social signal feedback could improve media skills training by allowing trainers to focus on general reflection and improving self-awareness rather than viewing a video multiple times to observe trainees' behaviour which is time consuming.

Detection of social signals related to good performance in media interview training has several potential applications. Firstly, it could improve the quality of feedback during training to support human trainers, as trainers may not be able to observe all of the signals that could impact evaluation of trainees' communication skills. Secondly, it could objectively capture the signals required for effective communication to enhance feedback and evaluation of performance.

The intended contributions of this research are made in the field of HCI (Wobbrock & Kientz, 2016):

Empirical Contributions:

- Improved understanding of how signals are detected by current COTS which map onto human judgements of communication skills
- Understanding the short- and long-term impacts of training augmentation by social skills feedback
- Understanding the potential effectiveness of social signals training interventions through experimental evaluation

Methodological Contributions:

- A new approach to analysing social signal data

Artefact Contributions:

- User centred development of a training intervention based on detection of social signals through COTS

Practical Contributions:

- To make recommendations for training practice

1.9. Overview of the Thesis

The thesis can be divided into four main research stages. The first chapter provides the scope of this thesis' research to permit the reader to understand the reasoning behind this research. The second chapter provides a review of the relevant literature from earlier psychology theories of communication and emotional communication as well as research using technology to capture nonverbal signals during training. Throughout this chapter, a number of gaps in research will be identified. Based on these gaps, a research question and aims are detailed at the end of this chapter. The third chapter provides justification for the design of the research, the measures and materials used to evaluate communication skills performance and capture social signals during interactions. This chapter also provides details on population sampling.

The fourth chapter uses quantitative methods to explore social signals necessary for evaluating trainees' communication skills performance in a radio interview and an on-camera face-face interview. Briefly, a preliminary analysis was conducted using machine learning to investigate emergent patterns of social signals (dependent variable) across multiple COTS technologies with subjective ratings of performance. The signals identified were the target for social signal feedback provision during the intervention detailed in Chapter 6. This chapter also provides an estimated sample size calculation for this type of training intervention which informs the sample size for the experiment / intervention research stage.

The fifth chapter is aimed at identifying the most appropriate method of presenting socials signal feedback to trainees. The signals identified in the preliminary analysis for feedback were evaluated based on their explanatory power and their ability to relate the signals back to behaviours that can be highlighted and displayed to trainees in a way that is meaningful and actionable. Several different visual displays were presented and explained to participants. Insights were collected using semi-structured interviews and a usability rating scale. Information gathered at this stage was used to inform and refine the design choice of how feedback is presented in the intervention experiment stage.

The sixth chapter is an experimental evaluation of the feedback method developed by comparing its effectiveness in improving communication skills to that achieved with the standard feedback received by trainees in media interview training. Trainees took part in a baseline (pre-training interview) and post-training interview and were randomly assigned to either the standard feedback group or the intervention group to take part in a set of three practice interviews.

The seventh chapter presents a follow-up assessment six months later to evaluate whether the skills attained during initial training were maintained after 6 months and to compare performance between feedback types.

The eighth chapter includes a discussion about possible limitations of this research and proposes future research recommendations. The ninth chapter provides a conclusion for this thesis.

1.10. Summary

This chapter has presented a rationale for the need for research on the topic of communication skills and social signals processing considering the importance of nonverbal skills, the reasons why training may be limited, how use of social signals might help, it then presented the intended contributions of this research and the thesis overview. The next chapter reviews the literature in the field of affective computing and SSP.

CHAPTER 2. CAN AFFECT RECOGNITION TECHNOLOGY IMPROVE NONVERBAL COMMUNICATION SKILLS IN TRAINING? A Literature Review

2.1 Introduction

This chapter reviews the literature relevant to social signals processing (SSP) and communication skills training. This literature review discusses the main communication theories, the different models of emotional communication and the role of nonverbal cues in communication including earlier research on its function in communication. An introduction to SSP is then presented, an area within affective computing, and its function in social interaction. Next, this review introduces different methods of capturing nonverbal signals using sensors with criticisms of each system and current concerns within the domain. There is also discussion surrounding research conducted in the lab and in the wild. The review then presents the reader with research aimed at improving communication skills or social skills using various forms of training feedback methods. An introduction to the specific focus of the current research in communication skills training is then provided. Finally, based on the literature, this review then presents the research gap, the aim of this research as well as the overarching research question.

2.2. Communication

The study of communication is multidisciplinary and has two main schools of thought; the first focussed on the process of a message transmitted from a sender to a receiver and the second on the production and exchange of meanings (Fiske, 2010). The focus of this thesis is on the former in a dyadic interaction. Communication is studied by psychologists and sociologists, who emphasise how messages are encoded and decoded with efficiency and accuracy. They also emphasise how the behaviour of one individual influences the behaviour or state of mind of another. The messages and meanings conveyed have been modelled into a triangle of communication in Figure 2.1 (Fiske, 2010).

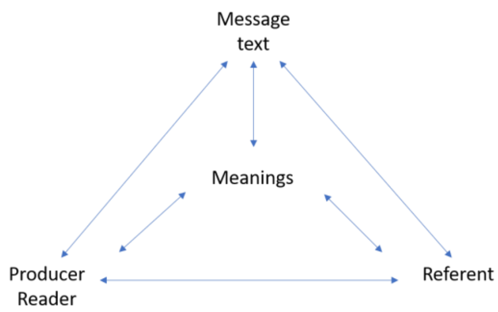


Figure 2.1. Fiske Model of Communication: Messages and Meanings

Several early models of communication have been developed; however, the Shannon and Weaver's Model is the most commonly used model of communication (Shannon & Weaver, 1949). The universality of this model suggests that it accounts for cultural differences in communication, particularly the use of nonverbal signals. Shannon and Weaver's (1949) model is displayed in Figure 2.2.

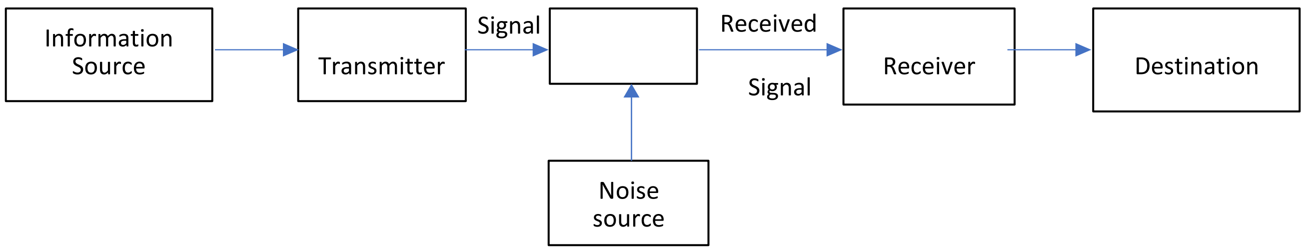


Figure 2.2. Shannon and Weavers model of Communication (1949)

Individuals send a message *(information source)* which is then *transmitted* (e.g. via the mouth or via non-verbal codes) and converted into a *signal* (e.g. sound waves) which is sent through a channel to a *receiver* (the ear) and eventually reaches its *destination* (e.g. the mind of the receiver) for processing. The way messages are conveyed are in the form of verbal and nonverbal codes such as facial expressions. Part of the meaning of a message can be delivered via nonverbal signals. Research has shown that nonverbal cues, also known as communication channels, are important for effective communication (Fiske, 2010), as noted by Shannon and Weaver (1949). There are various non-verbal channels that exist, each of which will be described later (section 2.4). Part of the reason nonverbal cues are important for communication is because these cues have been found to communicate unconscious emotions and aid in the delivery of verbal content (Kraut, 1978).

The reason why the Shannon and Weaver (1949) model is widely used is that it is generalisable, quantifiable and simplistic. While it was originally a linear communication model, the inclusion of backchannelling of signals was only added later (Fiske, 2010). However, this model has been criticised for its over-simplicity and is not context dependent (Chandler, 1994). Nevertheless, the use of this model to understand simple dyadic interactions are potentially useful.

2.2.1. Summary

The Shannon and Weaver Model of communication proposed that codes or nonverbal signals conveyed by various channels are important for effective communication and for understanding the meaning of a message conveyed. This model is the most widely used today, and while it is not without its limitations, it is possible to apply it to a simple dyadic interaction which is more suited to the analysis of nonverbal codes in an interaction than other models proposed which are beyond the scope of this research. There is often an overlap between emotions and nonverbal communication in social interactions. The next section discusses the literature on emotion and nonverbal communication.

2.3. Emotion and Nonverbal Communication

Nonverbal cues often overlap with discussion of emotion as there is a tight coupling between some nonverbal channels (particularly for facial expression) and some key theories of emotion (Knapp & Hall, 2009). This section introduces different models of emotion and a debate about the relationship between psychological constructs of emotion and human displays of affect. This section further discusses theories about communication of emotion which is also known as basic emotion theory.

2.3.1. Models of Emotion / Affect

2.3.1.1. Dimensional Models of Emotion

The view that emotions can be understood as a dimension is derived from Wundt who suggested that emotions are described along an affective continuum (Ekman, 2016). The principle of dimensional models is that a range of key dimensions may describe emotions (Feldman Barrett, 2011), i.e. relaxation/strain, pleasantness/unpleasantness and subdued / excited. The most referred to models are Russell's Circumplex Model (1980), Plutchiks' Eight Basic Emotions Model (1994) and Rolls Intensity Model (2005).

Russell's Circumplex Model (1980) was developed as he did not agree with the notion of basic emotions (discussed in section 2.3.1.3.). This model characterises emotion as a combination of activation (aroused - not aroused) and pleasantness ([pleasant – unpleasant] (Russell and Pratt, 1980). However, this model has faced criticism as there was substantial variability among affective states that failed to fall into their predicted regions of emotional responses (Remington, Fabrigar and Visser, 2000).

Similarly, Plutchik's Model Of Eight Basic Emotions vary in intensity and offer paired dimensions (e.g. joy - sadness, trust - disgust, surprise - anticipation, anger - fear etc) (Plutchik, 2001; Plutchik, 1994). Each emotion on this dimension make up further combinations of emotions, such as joy and trust can be love or fear and surprise can also be defined as awe. Combinations of these emotions are made based on the many emotions in this model. These combinations include optimism, aggressiveness, contempt, remorse, disapproval, awe, submission and love.

Another dimensional model is Rolls (2005) Intensity Model. Rolls (2005) attempted to explain emotion using a similar model of emotion and took into consideration that reward affects the intensity of the emotion experienced. Essentially, different emotional responses thought to be a product of different reinforcement contingencies.

2.3.1.2. Physiological Arousal and Emotion

In the early 20th century, James-Lange theory of emotion postulated that perception of an event results in physiological arousal. This model suggests that different physiological changes result in different emotions. However, this notion was criticised by Cannon-Bard (1931) who developed a theory of emotion that proposed that arousal did not change the emotion experienced, i.e. humans feel fear and, as a result of this emotional experience, run (Cannon, 1927, 1931). This view was backed by a later study conducted by Chwalisz, Diener, & Gallagher (1988) who found that severing of spinal cord as a result of an injury did not affect experienced emotions (i.e. physiological changes did not result in different emotional experienced). Schachter and Singers' (1962) two-factor theory of emotion poses that experienced emotions are a by-product of cognitive and physiological processes. More recent studies show a distinct physiological process which change during different emotional states (Gross & Levenson, 1993; Picard, Vyzas, & Healey, 2001). In addition to these studies, Ekman, Levenson and Friesen (1983) found that an increase in physiological arousal is associated with fear, surprise and happiness. However, a distinction between these emotions based on arousal have not yet been established.

2.3.1.3. Appraisal Theory

Appraisal theory posits that emotions felt are a consequence of one's evaluation of situations and events. In other words, our perception of an emotion causes an emotional response. This theory contrasts with the James-Lange theory of emotion discussed earlier (section 2.3.1.2). Originally, this theory was proposed by Magda Arnold (1960) who proposed that emotions depend on how we appraise objects and emotions. Arnold (1960) proposed four aspects of emotional appraisal: differences between perceptions and evaluation, immediacy of emotional evaluation, tendency to action and certainty.

An influential theorist who distinguished between primary appraisals (emotional reactions) and secondary appraisals (ability to cope with the situation) was Lazarus (1966). Lazarus (1991) proposed that emotions are a continuous process in that the same event can be reappraised and the initial response changes over time. This is commonly shared among appraisal theorists. Ultimately, Lazarus (1966) proposed that one's thought should precede your arousal and emotion. An example of this is when you are about to give a public speech and your simultaneously experience an increase in heart rate, sweaty palms and legs begin to shake and experience the emotion fear.

Appraisal theory has been critiqued in its ability to capture the dynamic nature of emotion. Two models of appraisal have emerged as a result of this. The first is the two-process model of appraisal proposed by Smith and Kirby (2000) which disguises between slow appraisals based on extensive reasoning from fast appraisals that are associative, or memory based. Fast and slow appraisal processes work simultaneously and are integrated to appraise the overall event.

The idea that nature of emotional experiences changes each time a new appraisal is added was originally proposed by Scherer (1984). Later, Scherer (2001) proposed the multi-level sequential check model which is made up of three levels of appraisal processing: innate (sensory-motor), learned (schema-based) and deliberative (conceptual) (Scherer, 2001). The model also posits sequential ordering of appraisals, such as a step-by-step (Marsella & Gratch, 2009). The step-by-step check process includes a relevance (novelty and relevance to goals), implication (cause and urgency), coping a potential check (control and power), and finally check normative significance (compatibility with one's standards). This sequential ordering of appraisals forms part of emotional experience.

2.3.1.4. Basic Emotion Theory

Basic emotion theory states that there are a small number of primary emotions which are expressed by humans. The concept of basic emotions has been declared by a host of researchers (Arnold, 1960; Ekman & Friesen, 1986; Gray, 1999; Izard, 1971; Oatley & Johnson-Laird, 1987; Plutchik, 2001). This view began with Darwin's' who proposed that the expression of emotions via facial expressions are essential for survival and these emotions have been evolutionarily selected for this reason (Darwin, 1872). Ekman and Friesen (1986) conducted empirical research on the universality of emotions and produced evidence for 6 basic emotions which were expressed in the form of facial expression that were identified in a range of cultures, including New Guinea, Japan and the United States (see Table 2.1) (Ekman, Sorenson, & Friesen, 1969). However, this view was challenged by other researchers arguing that emotions can be viewed in different dimensions which are placed on a continuum of arousal (Russell & Pratt, 1980) or that emotions experienced are directly related to autonomic arousal. There seems to be a difference between the expression of emotion and the subjective experience of emotion. Ekman and Friesen (1986) work equates the two but this is not true of all frameworks. Each perspective will be presented in this section.

Ekman's six basic emotions were also identified by Ekman and Oster (1979) but its notion of universality has been questioned. Specifically, the addition of surprise and fear as these emotions can be experienced in combination with other emotions (Oatley & Johnson-Laird, 1987). Both fear and surprise are usually expressed in conjunction with other emotions. It has been suggested that surprise is a response to a stimulus that is expressed together with another more articulated emotion (Power & Dalgleish, 2016). For example, if a person attends a surprise party, they can experience both surprise and joy or surprise and anger (if they do not like to be surprised) (Ekman, 1992). Other theories have proposed different core emotions with different arguments for their inclusion. These theories are presented in Table 2.1.

Table 2.1: A list of theories on basic emotions

Reference	Basic Emotion	Basis for inclusion
Arnold (1960)	Anger, aversion, courage, dejection, desire, despair, fear, hate, hope, love, sadness	Action tendencies
Ekman, Friesen and Ellsworth (1982)	Anger, disgust, fear, joy, sadness, surprise	Universal facial expressions
Gray (1999)	Rage and terror, anxiety, joy	Hardwired
Izard (1971)	Anger, contempt, disgust, distress, fear, guilt, interest, joy, shame, surprise	Hardwired
James (1884)	Fear, grief, love, rage	Bodily involvement
McDoughall (1926)	Anger, disgust, elation, fear, subjection, tender emotion, wonder	Relation to instincts
Mowrer (1960)	Pain and pleasure	Unlearned emotional states
Oatley and Johnson-Laird (1987)	Anger, disgust, anxiety, happiness, sadness.	Do not require propositional content
Panksepp (1982)	Expectancy, fear, rage, panic	Hardwired
Plutchik (1994)	Acceptance, anger, anticipation, disgust, joy, fear, sadness, surprise	Relation to adaptive biological processes
Tomkins (1981)	Anger, interest, contempt, disgust, distress, joy, shame, surprise	Density of neural firing
(Watson, 1930)	Fear, love, rage	Hardwired
(Weiner & Graham, 1984)	Happiness, sadness	Attribution dependent

Source: Ortony & Turner (1990)

Even though there is an abundance of researchers that subscribe to the concept of basic emotions, there is limited evidence on the universality of basic emotions (Moors, 2013; Ortony & Turner, 1990; James A Russell, 1994). The notion of expression of emotion through facial expression has also been challenged. For example, Fridlund (1994) postulated that emotions initiate an action-based response based on another person’s observation expression which is culturally different and not universally experienced. Feldman Barret suggests that emotions are not ingrained in us but rather are a result of experience (Barrett, 2006). This is referred to as appraisal theory. Our appraisal of a situation results in an emotional response. This maps into the notion that individuals are primed for expressing emotion and detection of another person’s emotions are derived from our ability to express that emotion. Meaning that emotions that happen to you are made by you and this change in emotional states are called predictions.

More recently, a study assessing agreement within emotion researchers' views of emotional theory by Ekman found that 88% of researchers endorsed the existence of universal emotions (Ekman, 2016) with emotional label agreement of 91% for anger, 90% for fear, 86% for disgust, 80% for sadness, 76% for happiness and 40- 50 % for shame, surprise and embarrassment. This research suggests some level of consensus among researchers in emotion (Ekman & Davidson, 1994).

2.3.3. Summary

A brief review of theories on emotion and emotional expression suggests that there is not one single approach to identifying emotion. The debate about the relationship between psychological constructs of emotion and human displays of affect still remains. From the early work of Darwin and Prodger, (1998) and empirical work by Ekman (Ekman et al., 1969; Ekman & Friesen, 1971) there have been suggestions for the existence of universal basic emotions that are displayed in recognisable facial expressions. While there is debate over the relationship between felt and expressed emotion, extensive research shows that facial expression for basic emotions are well recognised across cultures and contribute to social interactions (Ekman, 1999). As a result, the current research approaches emotion based on basic emotion theory, largely Ekmans 6 basic emotions. Additionally, it adopts the notion that autonomic responses are related to emotion experienced and will be investigated.

While there remains debate around theories of emotions, emotional expression has been linked to nonverbal cues and in improving communication. This is because nonverbal cues are interpreted by other dialogue partners as emotional signals which influence communication. The next section investigates the role of nonverbal cues in communication.

2.4. Role of Nonverbal Cues in Communication

Non-verbal signals are defined as a method of communication which does not contain linguistic content. These signals are also known as secondary codes and are communicated via different communication channels, such as facial expression, eye contact, gestures, body positioning, physical appearance, type of clothing worn and paralinguistic speech characteristics (i.e. speaking rate, volume pauses and interruptions) (Argyle, 1988; Coulson, 2004; Knapp, Hall, & Horgan, 2013). These can be further divided into combinations of signals such as posture while listening, the length of a gaze, mutual gaze between two interlocutors, eye widening, looking while listening, frowning while listening and so on.

Based on early research in anthropology and social psychology several functions of nonverbal signals have been identified. These include expression of emotions, communicating interpersonal attitudes, accompanying and supporting speech, self-representation (appearance) and rituals (i.e. greeting others in everyday situations). The meaning of these signals varies greatly in social settings and across cultures (Argyle, 1988).

2.4.1. Functions of Nonverbal Cues in Communication

There is long-standing research into the functions of nonverbal cues in communication which are said to be central to social behaviour (Gifford, 2012). Early research conducted by social psychologists showed that such signals contribute to the functioning of an interaction (Pentland & Heibeck, 2010). Correct use of these signals suggests social competence and social skills which result in harmonious synchronisation of reciprocal exchange of signals between interlocutors. Competent social skills often result in trust (DeGroot & Motowidlo, 1999), clarity in conversation (Krys et al., 2016) and rapport between interlocutors (Hart et al., 2016). Meanings behind nonverbal signals can communicate an array of actions, e.g. turn-taking or prosodic features, which can contribute to the overall congruency of communication. In this section, earlier research about communication channels is presented individually.

2.4.1.1. Physical Body Signals

Some physical body signals include hand gestures, feet gestures, head gestures, postural stances and orientation. For example, gestures of feet and head have significant meaning in interactions as they coordinate speech and supplement verbal communication (Navarro, 2003). Argyle (1988) states that the use of hand gestures is highly informative as it improves delivery of messages and, as a result, improves communication by regulating interactions. This regulation of interactions is done by suggesting turn-taking and assists the delivery of verbal content by signifying punctuation or greeting (Argyle, 2013; Navarro, 2003; Vinciarreli, Pantic & Bourlard, 2009). Similarly, head movements or nods are mostly used for the management of social interactions which include turn-taking in speech or fast nods suggest a wish to speak (Cerrato, 2005).

Different movements of hand gestures have different meanings, an example is up and down hand gestures indicate a requirement to dominate (Ellyson & Dovidio, 1985) while continuous and rounded gestures indicate an attempt to explain or win sympathy (Argyle, 1988). In addition, there are also symbolic gestures which are culture specific, an example is a ‘V’ sign indicates peace (Adler, Rodman, & Kramer, 1991). The gesture used for beckoning communicates that the gesturer wishes the receiver to follow them or these can be used to give directions (Ellyson & Dovidio, 1985).

Postural stances such as sitting, standing and laying down positions also communicate different meanings. Mostly, these are related to interpersonal attitudes including friendliness, superiority or inferiority and hostility (Bull, 1978). Similarly, the orientation in which we sit (body position) can suggest attitudes such as either aggression or intimacy. Research has found that sitting at a 90degree angle suggests a co-operative stance (Kluger & Denisi, 1996). Relating to intimacy, proximity can also suggest a relationship. However, this varies across cultures (Argyle, 1988).

2.4.1.2. Face and Head Movement Meaning

Facial expression can be broken down into sub codes including eyebrow positioning, mouth shape, nostril size and eye shape (Izard, 1971). Combinations of these signals have different meanings and it has been shown that there is less cross-cultural variation than other channels of communication (Ekman, 1992). An example is that a blank expression suggests boredom, smiling is considered friendly and a frown suggests anger or confusion.

Another facial action that has meaning in conversation is *eye movements and eye contact.* These actions transmit messages relating to dominance, affiliative relationships, i.e. making eyes at someone indicates more affiliative relationship, a need for feedback or to see how the listener reacts (see Kleinke, 1986).

2.4.1.3. Vocal Behaviour and Conversational Features

Vocal behaviour is divided into prosodic codes and paralinguistic codes. Prosodic codes, such as pitch or stress, affect the meaning of words used, for example a statement can turn into a question or expression of shock by pitch of voice (Profita & Bidder, 1988). Paralinguistic codes are those that communicate information about the sender. Codes include tone, volume, accent, speech errors and speed suggest a sender's class, social status, personality and a way of viewing the listener (Crystal & Ardener, 1971). Social status is indicated by the amount a listener interrupts a speaker as these are indicative of annoyance, disagreement with what is being said and eagerness to speak (Li, 2001). There are also cultural differences. Multiple interruptions can also suggest dominance which has also been associated with personality (Ferguson, 1977).

2.4.2. Emotional Communication

An early explanation of nonverbal behaviour is that nonverbal signals are used to assist communication in an interaction. A different explanation for nonverbal behaviour is that it leaks our state of emotions, as proposed by Argyle (1988). Argyle (1988) proposed the paradigm seen in Figure 2.4, like Shannon and Weaver Model of communication:

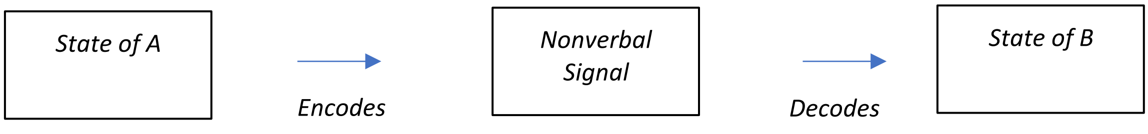


Figure 2.3 Argyle depiction of communication

Argyle (1988) posits that the possibilities of emotional communication in this instance that A does not want to communicate but their behaviour contains information that B can decode, i.e. perceiving aversion signal as disinterest. If A does not want to communicate then it could be possible that their behaviour is decoded by B incorrectly as they have incorrect information about the nonverbal signal and interprets this as deception.

The communication of emotions in social interactions is important to understand as others in the environment identify these signals and respond accordingly. Argyle (1988) has proposed three reasons for communicating emotions by humans:

1. Direct physiological reactions with no communicative intent (e.g. the facial expression of disgust when eating something unpalatable),
2. Spontaneous expressions of emotions which are thought to have evolved as social signals (e.g. expression of fear that informs others of the presence of a dangerous stimulus), and
3. Deliberately sent emotional expressions (not necessarily reflecting the emotional state which is experienced).

Research has found that deliberate transmission of emotional expressions is common and is dependent upon the presence of other people; for example, people are more likely to smile at other people than other stimuli, such as inanimate objects (Kraut & Johnston, 1979). People tend to conceal or emphasise spontaneous displays of emotions, so as not to reveal their emotional internal state which tend to be unconscious, implicit, low effort, rapid, automatic and contextualised (Kochanowicz, Tan, & Thalmann, 2016). This inhibition of revealing emotional states stems from societies' responses to emotional expression which can encourage us to dampen our emotions (Pentland & Heibeck, 2010).

Studies of emotional expression dates back to the work of Descartes who depicted facial expressions of the passions for French artists (Power & Dalgleish, 2016). In evolutionary terms, displaying emotions in interactions benefits both senders and receivers, i.e. facial expressions of fear or anger in threatening situations could result in seeking safety or fight for survival. Emotional signals are communicated via multiple channels; such as facial expressions, vocal behaviour (i.e. tone of voice and vocal bursts), gestures and posture (Argyle, 1988). The sections which follow consider how emotion is conveyed and how they are often interpreted.

2.4.2.1. Emotional Communication: Gestures and Body Movements

It has been suggested that a social interaction cannot occur without the use of body movements and gestures which make up 90% of an interaction (McNeill, 2008). Gestures and body movement can have both a planned communicative intent or be unplanned (i.e. unconscious) (Pentland & Heibeck, 2010); while the former is very often learned and culture-specific for some channels, the latter are often associated with emotions and levels of arousal (Knapp & Hall, 2009). A study by Van den Stock, Righart, and de Gelder (2007) investigated body postures and movements associated with emotions. Stock, Righart and de Gelder (2007) conducted research to test recognition of body expressions by matching an emotion to pictures of body expressions.

Research has shown that displays of gestures differ among males and females; males display greater, more open and more broad gestures than females who also tend to clasp their hands (Argyle, 1988). Moreover, use of gestures has also been linked to personality types or emotional states such as anxiety or aggression (Argyle, 2013; Song, Demirdjian, & Davis, 2015). People who express boredom, stress or negative feelings during an interaction are known as adaptors and are recognised for expressing their emotions unconsciously. Unconscious expressions include self-protection gestures (e.g. moving legs, swaying, folding arms and turning away from others), self-manipulations (e.g. nose scratching, ear touching or biting of the lips) and fidgeting with small objects which can also be interpreted by listeners as nervousness or boredom (Gross & Levenson, 1997). However, no research has found consistency in use of gestures and meaning which suggests use of this channel may vary extensively.

Posture and body movement is another form of nonverbal communication and includes standing, sitting, squatting and kneeling and lying down (Argyle, 2013; Vinciarelli, Pantic, & Bourlard, 2009). Postures are also associated with an unconscious relay of emotions and are suggestive of people's attitudes toward a situation or a person and / or tension in an interaction (Scheflen, 1964). Posture has been classified into three criteria, the first defines inclusive and non-exclusive postures, the second includes face-face orientation of interaction (e.g. facing the other person during an interaction or facing away from them) and, finally, congruent and incongruent postures signpost psychological involvement and is also known as mirroring (Chartrand & Bargh, 2002). An example of congruent postures is if a person is leaning toward another person suggests they are involved. However, if the two interlocutor postures are dissimilar this suggests no involvement. Congruent and incongruent postures have also been found to suggest relationships, dominance, status and affective state between two interlocutors (McArthur & Baron, 1983). Weisfeld and Beresford (1982) suggest that postures have been found to be less controlled than facial expressions, an example is that anxiety will not be revealed on the face but will be revealed in posture.

2.4.2.2. Emotional Communication: Facial Expression

Research by Darwin and Prodger (1998) and Ekman and Friesen (1971) have hypothesised that emotions are expressed via bodily expressions as they contain components that are necessary for survival (i.e. senses such as smell or sound for detecting danger) that enable humans to communicate on a social level which is hypothesised as the primary basis for our evolution. Facial expressions convey experienced emotions which can also be displayed intentionally or unintentionally.

Early work, motivated by Darwin, was conducted by Ekman who proposed that there are 6 basic emotions which are universally experienced (Ekman & Friesen, 1978), which include sad, happy, fear, surprise, angry and disgust. These are translated into 7 facial expressions with the addition of contempt. Agreement in the field is that expressions tend to be recognised consistently across most cultures from still photographs.

Ekman and Friesen (1978) developed a set of action units (AU) that describe features that make up each of the 7 facial expressions. These can be coded by humans to manually label each facial action (for AU descriptions and their associated facial muscles see Appendix 2.1). AUs are not solely associated with the facial expressions posed by Ekman and can therefore be interpreted using higher-order decision-making regarding emotions expressed (EmFACS) such as confusion, intrigue and posed expressions (Ekman, Friesen, Hager, & Human Face, 2002). AU have been used to produce realistic and recognisable facial expressions including intrigue (Cunningham, Kleiner, Bulthoff, & Wallraven, 2004; Ekman et al., 2002).

A widely used automated system used for labelling and capturing facial expressions is the Facial Action Coding System (FACS) which uses different AUs mapped by facial muscles. This system has discovered new AU patterns associated with shame, passion and friendliness (Smith-Lovin, Lewis, & Haviland, 1995). This facial emotion recognition system forms the basis of action units. Research has also suggested FACS can differentiate between telling the truth and lying more accurately than subjective human judgement (Frank & Ekman, 2004).

2.4.2.3. Emotional Communication: Vocal Behaviour

Vocal cues that have been extensively studied and include tone of voice as well as pauses/turn taking latencies (Knapp and Hall, 2009). Vocal nonverbal behaviour involves five components that include voice quality, non-linguistic and linguistic vocalisations, silences and turn-taking patterns (Vinciarelli et al., 2009).

Prosodic features can communicate using voice quality, i.e. vocal energy are suggestive of emotions such as anger and fear (Scherer, 2003) and fluent speaking rate suggests lack of hesitation, persuasiveness and competence (Scherer & Giles, 1979). Vocal pitch has been found to be a personality marker (Scherer & Giles, 1979) and suggests structure during dialogue (Hirschberg & Grosz, 1992). Additionally, a sudden increase in pitch when saying a word places emphasis (Hirschberg, 1993).

Non-linguistic vocalisations are vocal outbursts that can provide details about the social situation such as scream (fear), crying (sad / happy), whispering (secretive / considerate), groaning (pain / sad) and laughing (happy) (Keltner & Haidt, 1999). Crying in response to someone else's pain is often seen between people that are connected by social bonds and signifies empathy and is representative of mirroring (Keltner & Kring, 1998). Psychological research has demonstrated that vocal outbursts such as laughing, yawning, coughing and sighs are considered to be matched up with the basic emotions proposed by Ekman & Friesen (1967) (Russell & Dolls, 1997). Linguistic vocalisations are non-words that include "aha" and "uhm" and are used when someone cannot answer a question, or they are used when embarrassment is experienced (Glass, Merluzzi, Biever, & Larsen, 1982).

Research has also shown that posture, body movements, facial expression and voice prosody are expressive of emotions which are typically judged by observers or mapped onto experienced emotions (Coulson, 2004). Coulson (2004) found that angry and fearful bodily expressions are often incorrectly recognised than sad which is often correctly recognised. Researchers also found that facial expressions assist in the recognition of emotions expressed by bodily expressions and that a perceived whole-body expression influenced the recognition of voice prosody (Coulson, 2004).

In summary, research shows that many nonverbal cues are interpreted by other dialogue partners as emotional signals which influence communication. The next section introduces more recent research in the area of affective computing and nonverbal signals.

2.5. Affective Computing and Social Signal Processing

Affective computing can be defined as the study and development of devices and systems which are capable of interpreting, recognising, processing and simulating human emotions (Picard, 1995). Systems which have been developed to recognise human emotion can potentially be criticised because of a potential lack of mapping from displayed signals and felt or experienced emotion. However, there is evidence that signals are functionally relevant to effective human communication and one branch of affective computing specifically focusses on this. This branch is called SSP which has arisen as a multidisciplinary area applying computing technology to aid the understanding of role of social signals in communication with possible application to the design of computer systems that can be more applicable to human users (Damian, Dietz, Gaibler, & André, 2016).

A significant facet of human intelligence is social intelligence as it is hypothesised to be our primary reason for survival in evolution terms and the complexity of human communication is hypothesised as a by-product of our need for social interaction (Darwin & Prodger, 1998; Ekman, 1992). Modelling of social interactions can potentially be used to train an unsocial machine to interact with humans using the appropriate nonverbal cues to ensure the interaction is more natural and, as a result, trusted by users (Vinciarelli et al., 2012).

Research in the SSP domain typically focuses on detection of a signal in isolation such as a self-contained smile or frown (what) (Vinciarelli & Pentland, 2015); however, to note the significance of a smile, researchers must simultaneously identify *where* (context) the individual is (at home, at work or at a family gathering). It is also important to understand *when* the smile was displayed (timing of interaction) and *who* the smiler is (identity and age). This is known as the W4 quadruplet of social interactions (Vinciarelli et al., 2009). However, Rudovic, Nicolaou, & Pavlovic (2017) added *why* and *how* as research is required to understand the reason behind this expression (e.g. a TV show), thereby relabelling it the W5+. Rudovic, Nicolaou, & Pavlovic (2017) argue that that using W5+ for modelling human-human behaviour could enable the generalisation of findings across all interactions.

2.5.1. Functional Roles of Social Signal Processing

In contrast to earlier research briefly detailed in section 2.4.1, recent research in social interactions using SSP methods has developed functions that provide new insights into the role of nonverbal social signals in social interactions (Damian, Tan, et al., 2015; Tanaka, Sakti, Neubig, Toda, & Nakamura, 2015a). These functions contribute to understanding the reciprocal exchange of information between two interlocutors and include; thin slices of behaviour, dialogue management, communication of emotion, social relationships, judgements of personality and honest signals. Each can contribute to understanding the intricacies surrounding the back and forth exchange of information between two interlocutors. This understanding can contribute to the synthesis of social interactions that can inform the development of socially intelligent computers. This review now looks at each of these functions in turn.

2.5.1.1. Thin Slices of Behaviour

The ‘thin slices of behaviour’ can be described as the conclusion about attitudes and emotions based on the first 5 minutes of a social interaction (Ambady & Rosenthal, 1992). In the initial stages of any interaction, the way in which the sender of a message uses social signals depicts how they are perceived by the receiver (Vinciarelli et al., 2009). Impressions formed initially will decide how the receiver of a message responds to the initial message sent, i.e. the back channelling of signals. This is also known as ‘impression formation’. A large part of human interactions is forming impressions of others so that we infer whether an interaction with a person that may impact our lives in a positive or negative way, and this coincides with the notion of human survival (Ambady & Rosenthal, 1992).

A study by Nguyen and Gatica-Perez (2015) investigated impression formation in an employment context by collecting 939 conversational English-speaking video resumes from YouTube. These videos were annotated by Amazon Mechanical Turk crowdsourcing platform. Results showed that social / communication skills (as well as extraversion) are important for the formation of first impressions.

Recent research using automated recognition software to capture nonverbal signals in a job interview found that positive nonverbal behaviour in response to a first question predicts successful entry into university (Naim et al., 2016). Later, Muralidhar (2017) studied the relationship between nonverbal communication and first impressions in job interviews. Capturing of nonverbal signals were done using wearable devices. Preliminary results indicated that there is a positive correlation between speaking time and overall impression, and silent events were negatively correlated. Overall, the use of nonverbal behaviour was more predictive of forming all first impressions. This is also shown in the existing literature (Frauendorfer, Schmid Mast, Nguyen, & Gatica-Perez, 2014; M. L. Knapp et al., 2013; Muralidhar et al., 2016; Nguyen & Gatica-Perez, 2015).

2.5.1.2. Dialogue Management

To manage a social interaction, the appropriate use of nonverbal cues, such as facial expression, interruptions and vocal behaviour, is essential for flow and appropriate interruption/intervention during a conversation (Psathas, 1995; Yule, 1996). A synchronous and harmonious interaction between two people is typical when both have interacted with one another on many occasions and depends on non-verbal cues (facial expressions, gestures and gaze) which direct flow in a two-person discourse (Yule, 1996). Using SSP research, capturing of pitch and vocal energy using a sociometer (discussed later in section 2.5.1.6) have been found to request turn-taking (Choudhury & Pentland, 2004). An interaction is considered smooth if there are no interruptions or long silences and consistent turntaking between interlocutors (Choudhury & Pentland, 2004). Research using a multiagent system (a user and their smart phone) has shown that interruptions during discourse are consistent with an undesirable interactions and detection of mood (Ferreira, Lefevre, & Lefèvre, 2013).

A review of the literature on emotion expression using nonverbal behaviour in doctor-patient visits found that appropriate use of these signals is important for patient satisfaction, treatment adherence and care outcomes (Roter et al., 2014). Signals were monitored on the Roter Interaction Analysis System which is used for understanding conversation and nonverbal signals in medical settings.

2.5.1.3. Communication of Emotion

Research suggests that detection of social signals communicates emotion and that the expression of emotion improves the quality of the interaction (Chaby, Chetouani, Plaza, & Cohen, 2012). Chaby and colleagues (2012) presented a multidisciplinary method to study multimodal socioemotional behaviours in children aged from 6 – 12 years old that meet the criterial for Autism Spectrum Disorder. Emotional signals which were captured include speech, prosody, facial expression and postures. In additional, conversational features were captured that included synchrony and engagement. This research showed that it is possible for researchers to capture emotions.

2.5.1.4. Social Relationships

Nonverbal cues are known to influence how senders of information about their relation to the receiver of information (Gatica-Perez, 2009). Research conducted by Zancanaro, Lepri, and Pianesi (2006) showed that it is possible to classify the role of group members in a face-face interaction using automated recognition systems. The dataset analysed was an annotated Survival Corpus. Group member roles included socio-emotion area roles which include ‘neutral’, ‘gate-keeper’, ‘supporter’, ‘protagonist’ and ‘attacker’. They also include Task Area roles which include ‘follower’, ‘orienteer’, ‘giver’, ‘seeker’ and ‘recorder’. To classify social and task area roles, the social signals captured included speech and body fidgeting. Machine learning using Multilevel Support Vector Machine results from obtained from this study suggest that capturing signals are not effective enough for a real application in predicting roles, but rather that it is feasible to do so.

2.5.1.5. Judgements of Personality / Traits / Friendliness

In the literature, there are limited studies using SSP methods that investigate signals associated with relational messages, what is known is that relational messages strengthen relationships between individuals as a result of the reciprocal exchange of a dialogue (Vinciarelli et al., 2009). A recent study by Zhang, Luo, Loy, and Tang, (2018) investigated the fine-grain high level interpersonal relation traits (dominance, competitiveness, trusting, warm, and friendly, involved, demonstrative and assured) in static images in the wild. In contrast to focusing on a single channel of communication (facial expression) researchers devised a multitask network that can detect facial expression, gender, head pose and age. This study found that the inclusion of multiple features is a good method of identifying and predicting interpersonal relations.

It has also been found that smiling in combination with grand gestures are associated with confidence or dominance (Fang, van Kleef, & Sauter, 2018; Kim, Chang, Holland, & Pentland, 2008; Rasipuram & Sowmya, 2016), and smiling in combination with gaze demonstrates friendliness (Cafaro et al., 2012). In the context of a job interview, Naim and colleagues (2016) found that smiling was highly rated if displayed after the first interview question.

2.5.1.6. Pentland’s Honest Signals

Pentland and Heibeck (2010) states that nonverbal cues / social signals are evolutionarily inherited as a result of the primary function being survival. This explains why they are understood across most cultures. Pentland and Heibeck (2010) term these as *honest signals* as emotions are said to be relayed in a second string of communication. This second string of communication is said to be conveyed in the form of facial expression/micro facial expression, hand gestures, body movements, posture, distal and proximal proximity and tone of voice. Recognition of these signals are a recognition of the communicator’s true emotions (Pentland & Heibeck, 2010). Pentland describes the four ‘honest signals' that are consistently present in all interactions: mimicry, activity, proximity and consistency. These four honest signals can be observed in a dyadic interaction or multiple-person interaction in a number of different circumstances, i.e. deception (Sung & Pentland, 2005), negotiation (Curhan & Pentland, 2007) and within an organisation (Rutkin, 2014). Pentland also proposed that people can accurately judge others' emotions at above chance levels based on surprisingly small amounts of behavioural information, often called ‘‘thin slices’’ of behaviour (Curhan & Pentland, 2007). Sometimes, the ‘‘thin slices’’ that are investigated are less than 1 second in duration, but more often they are several seconds to several minutes long. This was discussed earlier in section 2.5.1.1.

2.5.2. Summary

There are multiple functions of the SSP domain, each extending on earlier research presented in 2.3.2. The development and design of the systems used to detect emotions in interactions vary, some are developed by academics and some are developed for commercial use. Each will be briefly discussed in the next section.

2.6. Technology for Affect Recognition and SSP

There are a number of approaches for capturing emotions using sensors. The categorical approach (basic emotion theory), dimensional approach and the appraisal-based (dimensions of meaning associated with particular emotions) approach. Each theoretical approach is discussed in section 2.3.

Sensors have been developed to automatically detect social signals and researchers manually interpret these. There are mainly two methods of capturing signals; bespoke developed systems (open source software) or commercially available off-the-shelf technology. Each present with advantages and disadvantages. The method in which signals are captured and systems are designed will be discussed in this section according to systems designed for different communication channels.

Sensors that sense and respond to affect are evolving and the domain of multimodal emotion recognition has made significant advances from ‘proof of concept’ systems. Multimodal detection includes automatic detection of a combination of facial expression, vocal prosody, text, body movement, eye gaze, peripheral signals, such as hand and head gestures, and central physiological signals during an interaction. These systems have also been found to capture contextual cues. Unimodal detection includes each signal isolation rather than in combination. In a review by D'Mello and Kory (2015) it was identified that multimodal affect detection systems were consistently more accurate than unimodal detection systems in the laboratory.

2.6.1. Evaluation Process

Data handling and coding in the SSP domain is labour intensive and relies on machine learning methods for training of data in predicative analyses. There are six steps in capturing emotions; data source identification (e.g. social interaction, email, etc), pre-processing (exclude irrelevant information), emotion model development (e.g. corpus based, machine learning based, or knowledge based), post-processing (fine tune output), results and evaluation of results (Binali & Potdar, 2012).

2.6.2. Media Technology to Capture Social Interactions

A multitude of sensors are used to capture signals and synthesise human-human interactions, most of which include cameras and voice recordings (Vinciarelli & Odobez, 2006), cellular phones (Pentland, 2007), physiological detectors (Gunes, Piccardi, & Pantic, 2008), functional magnetic resonance imaging (fMRI) (Montague et al., 2002) and EEG signals (Uddin, Iacoboni, Lange, & Keenan, 2007).

2.6.3 Open Source Software / Bespoke Software

Researchers often use open source software or develop their own systems to capture social signals. Both approaches use machine learning techniques that are based on classification and prediction methods to label emotions. A number of machine learning methods are used to process these recordings to investigate the W5+ model proposed by Rudovic et al., (2017) and will be noted briefly in this section.

For voice analysis, the pre-processing of speech features of a dyadic discourse include conversational elements such as the segmentation of speaker turns of around 2-3 ms and separating the data into speech and non-speech features (‘noise’) using methods such as Artificial Neural Networks (Ajmera, McCowan, & Bourlard, 2003), *k* nearest neighbour (Lu, Zhang, & Jiang, 2002) and Gaussian Mixture models (Gauvain, Lamel, & Adda, 1998). Clustering is then used to automatically recognise when which individual is talking (Vinciarelli et al., 2009b). Huang and colleagues (2001) state that for speech recognition features, typical methods include Mel Frequency Cepstrum Coeficients (MFCC) and Linear Predictive Coding (LPC). The analysis of speaking transitions between individuals in an interaction include splitting speaking segments into short intervals and timing the difference between intervals. The highest value resembles speaker changes (Vinciarelli et al., 2009b).

Classification is a supervised learning approach whereby the computer program learns from data input and uses this information to learn and as a result classify a new observation (Witten, Frank, & Hall, 2011). For instance, the classification of facial expressions is usually conducted using framebased and sequence-based methods (data input). Frame-based methods are typically used for the classification of the six basic emotion categories and have been useful in the static classification of facial expression of pain (Gholami, Haddad, & Tannenbaum, 2009; Lucey, Cohn, Prkachin, Solomon, & Matthews, 2011). Images used should be labelled and annotated and if all images are not annotated, this becomes a problem for machine learning methods and missing data influences training (Zhang, Luo, Loy, & Tang, 2018).

Classification algorithms that have been effective in correct classification of signals include Support Vector Machine (SVM), Relevance Vector Machine (RVM), Neural Networks (NN) and Bayesian Networks (BN) (Rudovic et al., 2017). However, a problem with frame-based methods are that they ignore the dynamic classification of facial expressions or action units. Methods used in dynamic classification are more focussed on these such as Dynamic Bayesian Networks (DBN) such as Conditional Random Fields (CRF) (Chang, Liu, & Lai, 2009; van der Maaten & Hendriks, 2012) and Hidden Markov Models (HMM) (Cohen, Sebe, Garg, Chen, & Huang, 2003; Lifeng Shang & Chan, 2009; Otsuka & Ohya, 1997).

There is several open source software which captures facial recognition. Some of which include FaceNet (Schroff, Kalenichenko, & Philbin, 2015) and OpenFace (Amos, Ludwiczuk, & Satyanarayanan, 2016). Each require an advanced understanding of how to develop a working system. Open source software often used for voice analysis often used is Praat. The next section presents COTS Technology.

2.6.4. Commercial Off-The-Shelf Technology (COTS)

COTS technology has been developed and made available to the public for use in commercial settings such as detecting vocal emotion of callers for call centres to improve call interactions (Nemesysco / Cogito / VoiceIQ) (Nemesysco.com / www.cogitocorp.com / www.voiceiq.ai), facial recognition software to unlock smart phones or detect emotions of film watchers (FACET / Affectiva), Microsoft Kinect to detect full body motion for improving rehabilitation (Chang et al., 2012), Fitbits provide of personalised health data to improve or maintain a lifestyle behaviour by changing initial behaviour through tracking movements (Piwek, Ellis, Andrews, & Joinson, 2016) and sociometric badges used for discovering the social networks of a company for better staff management (Fischbach et al., 2010).

There are technologies which have been developed in the lab which are now commercially available which include Affectiva and sociometric badges. Affectiva facial recognition software are used for academic research as well as used for commercial purposes (Dupré et al., 2017). Sociometric have been designed to detect physical proximity, conversational time, body movements, relative location and vocal features (Pentland & Heibeck, 2010). These are described as *honest signals* and are discussed in section 2.5.1.6. This technology was designed by Dr Sandy Pentland at Massachusetts Institute of Technology (MIT) and has been useful in the interpretation and investigation of social interactions in an unobtrusive manner (Olguín, Pentland, Olguin, & Pentland, 2007). Sociometric badges have been effective in understanding social interactions in studies attempting to understand a company's social network dynamics by digitally tracking their employees (Rutkin, 2014). This has been found effective in evaluating an employee's productivity (Rutkin, 2014). This not only illustrates that this wearable type of technology is beneficial in understanding a company's social network but also ways of improving this social network to enhance productivity. Additionally, these badges have also been effective in identifying how a person functions on a day-to-day basis (Chandrasegaran et al., 2016). This information was used to identify healthy habits and assist in adjusting habits to a healthier lifestyle. Sociometric badges have been discontinued and are only for sale to large corporations.

2.6.5. Criticisms

The problems with developing individual technologies are that these take time, are costly and prone to error. In addition, they also require technical skills. A possible alternative to consider using in capturing social signals in interactions are COTS technologies as they provide a much faster method of capturing signals in an interaction. However, these technologies present their own concerns. A down-side to these technologies are that algorithms used are patented and therefore the accuracy of these technologies is not known (Piwek et al., 2016). This produces many criticisms regarding the accuracy and reliability of these systems (Horvath, Mccloughan, Weatherman, & Slowik, 2013). Nevertheless, COTS technology provides proof of concept regarding capturing signals in an interaction.

2.6.6. Challenges for SSP

The SSP domain is still evolving and often faces analytical challenges about the perception of information conveyed in an interaction between the sender and receiver and if it is socially relevant. This challenge demonstrates the complexity of social interactions. Vinciarelli et al. (2012) state three major problems; the first is that there is little known about how social information is sent daily between two people. Little is also known about what makes these interactions natural; however, research has been done for emotion-related communication which suggests that this naturalness is a result of emotions and that everyday interactions are multimodal which are not often explored / captured in research (Hashimoto, Yamano, & Usui, 2009).

The second problem is that the data analysed should represent the context in which the interaction is automatically captured. The meanings of signals cannot be derived from observing the signals in isolation of the context or setting as the environment influences the signals shown. Additionally, the formality of the situation influences the signals displayed; such as relationship between interlocutors, personality and affective state.

The third problem is synchrony of the timing of signal displays in relation to other displays by other communicators. This potentially allows for a better synthesis of the interaction. Timing also extends to synchronized displays of multimodal cues that allow for the avatar to be perceived as natural and not strange or unfamiliar (Seyama & Nagayama, 2007). There is also no consensus about whether multimodal cues should be synchronized as some cues may be redundant and a single cue may compensate for the lack expressiveness in other modalities (de Gelder & Vroomen, 2000). Nevertheless, it is likely that multimodal cues do contribute to understanding of what constitutes these unconscious perceptions that humans develop without knowing why.

Overall, the aim of the SSP is to improve the interaction between humans and an emotionally intelligent computers, such as a conversational avatar. The challenges presented in this section revolve around developing a conversational avatar which can successfully converse with a human simulating a natural interaction. The issues presented in this section prove the complexity achieving this natural interaction.

2.6.7. Multimodal Fusion

Studies use sensors to identify social signals present in dyadic interactions which can inform the development of multimodal embodiment or a conversational virtual agent. As noted previously for effective modelling of social interactions, multimodal detection of social signals is essential as combinations of signals have different meanings when compared to their use in isolation (Adams & Kveraga, 2015; Cid, Manso, & Núñez, 2015; Yang, Metallinou, & Narayanan, 2014).

A study by Yang, Metallinou, & Narayanan (2014) investigated interlocutor dynamics in a dyadic interaction which assessed how interactors adapt their nonverbal cues in response to the other based on the goals and context of the interaction. Signals captured were speech features (i.e. pitch and energy) and body movement (i.e. looking at one another, turning away, approaching, touching and hand gestures). Results suggested that behavioural coordination between interlocutors in an interaction depends on the stances assumed in the initial interaction and are dependent upon the type of interaction. It was also found that body language is influenced by a combination of behavioural cues further implicating the importance of multimodal behavioural analysis. However, this study was limited to bimodal analysis and did not include expressions of the face or conversational features which could inform the given situation and, as a result, lack crucial information about the interaction.

Naim and colleagues (2016) conducted a study to identify the verbal and nonverbal behaviours which are predictive of good job interviews. A total of 138 interviews were included in the study and analysis included facial expression, eye gaze, language and prosodic language which includes pauses and tone. Interviews were rated by Amazon Mechanical Turkers and a weighted average was calculated which was the basis for the ground truth labels. This study revealed that speaking fluency, less filler word usage, use of more unique words and smiling more are predictive of a successful job interview. The results of this study show the importance of multiple modalities when analysing social interactions. However, this study did not include additional signals such as hand gestures and body movements and therefore lacks potentially important information about the social interaction.

It is noted that nonverbal communication between two interlocutors depend on the goal and the context of the interaction. For instance, non-verbal signals that have been detected and identified as potentially important in a job interview are to smile more (Naim et al., 2016), whereas in a healthcare setting turn-taking, speaking ratio, volume, pitch, smiling, frowning, head tilting, nodding, shaking and overall body movements were extracted using automated recognition technology and are assistive in clinician-patient interactions (Liu et al., 2016). In the classroom, non-verbal cues commonly captured during presentations are prosody, voice quality and gesturing activity (Cheng et al., 2014). Signals extracted from these contexts demonstrate that there are differences in how individuals communicate differently given the context.

2.6.8. Lab vs 'in the wild'

Much of the research to date has only been conducted in a controlled environment where the set-up is carefully designed allowing researchers to obtain full control of the session, location and instruments. This can be limiting as social interactions are typically conducted in everyday life 'in the wild' and results obtained in the lab may not be reliable in this context. A limited amount of research has been conducted to investigate behaviour in the wild (Gunes & Hung, 2015; Gunes et al., 2008), which is more natural and less posed than in a lab setting. Reasons for limited research in this area are privacy issues in obtaining video recordings of interviews and the effectiveness of recognition technology.

A study by Zhang and colleagues (2018) investigated whether warmness, friendliness and dominance can be predicted by a single facial image which are categorised by gender, age and head pose. Images were extracted from the Static Facial Expressions in the Wild (SFEW) 2.0 dataset which includes natural facial images extracted from movies. Researchers formulated a two-step method for recognising facial expression using a novel training method (see paper for more details). Using this method of prediction analysis, researchers found that facial images captured 'in the wild' can be predictive of interpersonal relations and facial expressions.

An additional study that captures social signals 'in the wild' using a predictive framework is a study by Nguyen and Gatica-Perez (2016) in forming first impressions in the context of a job interview. Researchers evaluated YouTube videos for speaking activity, prosody (pitch, energy and rate), proximity, head motion and facial expression using recognition technology and were able to effectively predict first impressions, communication skills and extraversion.

Nevertheless, research conducted in the lab allows for the precise measurement of the effects of independent variables on dependent variables in isolation. In the lab, researchers are also able to control for extraneous variables. This, in turn, enables a cause and effect relationship between variables to be established (Mitchell, 2012). It is for this reason that novel research should be conducted in a laboratory setting.

2.6.9. Summary

The use of sensors to detect social signals has proven to be effective; however, many of the studies in the literature only include a few channels of communication which limits researchers' ability to truly synthesise a social interaction as the meaning of combinations of signals can differ. In addition, there is little research that focuses on a reciprocal exchange or back channelling of multimodal signals in a fast and objective manner in media skills training which is important to understand relational messages central in a professional context (Kim & Suzuki, 2014). Furthermore, while interactions 'in the wild' are more effective in analysing interaction for simulation, novel research should be conducted in the laboratory as it is a good starting point for evaluating a context not previously modelled. The current focus of the current research is to investigate the back and forth exchange of messages by detecting multiple communication signals that include facial expression, body movement, hand gestures and voice emotions simultaneously in a controlled experiment. The next section discusses research on improving social skills using sensors.

2.7. SSP Interventions to Improve Human Communication

Researchers are beginning to investigate the potential of interventions based on technology to improve communication skills. Technology enhanced feedback is typically given to trainees in real time (Damian, Baur, & André, 2016), post-hoc (Fung, Jin, Zhao, & Hoque, 2015) or a combination of the two (Hoque, Courgeon, Martin, Mutlu, & Picard, 2013). Research has also included tutor feedback during training which has proven effective (Liu, Scott, Lim, Taylor, & Calvo, 2016). Each of these will be discussed in turn in this section of this thesis.

Ruiz, Chen and Oviatt (2010) have argued that presenting multimodal feedback reduces cognitive load as it mimics that of everyday interaction and understanding of the world. Advantages include robustness (refining imprecision through improved understanding of holistic behavioural actions), naturalness (increases communication about performance), flexibility (perceive and structure their communication), minimising errors (understanding errors).

2.7.1. Real-time Feedback

One approach to improving performance that has been proposed is the augmentation of social interactions. It is possible to use sensors and displays to provide real-time feedback about the users'

nonverbal behaviours (Damian, et al., 2015). Using social augmentation to improve social behaviour whilst engaging in an interaction is a means of improving social skills in those who have difficulties in engaging in one-to-one interactions (i.e. those on the autistic spectrum) (Tanaka et al., 2015a) or those without practice in certain contexts (i.e. public speaking) (Damian, et al., 2016).
The goals of a social augmentation system are to ultimately make them aware of one's own body and to improve the quality of their own behaviour, this includes the self-awareness.

Research conducted by Tanveer, Lin and Hoque (2015) developed an intelligent user interface which provided real-time feedback during a presentation with 30 native English Speakers using Google Glasses. Speakers presented three speeches with differing feedback types. Feedback types included continuous feedback, sparse feedback and baseline where they had received no feedback during presentations. The results showed that participants were more pleased with the sparse feedback strategy. This could be because the speaker will lose eye contact with the audience during the presentation which is most likely to negatively impact their performance. Also, sparse feedback throughout training could be effective but could also be distracting.

This section presents various frameworks and methods of providing real-time feedback during training to improve performance. These include virtual audience feedback as a means of practicing and improving confidence of speakers and the behavioural feedback loop which is commonly used in public speaking (Schneider, Börner, Rosmalen, & Specht, 2015). This section details each by presenting how each have been used and their significance in improving training performance as well as disadvantages of each.

2.7.1.1. Implicit - Virtual Audience Feedback

An important part of giving a presentation or public speaking is the ability to gauge the audiences' collective overt / behavioural responses (Radbourne, Glow, & Johanson, 2013). An example of this would be if a public speaker makes a joke and some of the audience smiles or laughs, this allows the speaker to know that the audience found the statement funny, or an unexpected silence, suggesting the statement might not been as successful as the speaker had hoped. Audience responses depend on the type of audience, genre of the talk and the context. Technology has been developed to capture audience reactions to performances in real-time which is not as intrusive wearing an EEG cap and can be more accurate than gathering retrospective data (Batrinca, Stratou, Shapiro, Morency, & Scherer, 2013; Xu & Plataniotis, 2016).

In 2013, Cicero explored the possibility of using an interactive virtual audience for training in public speaking using nonverbal features (Batrinca et al., 2013). Later, research by Chollet and colleagues (2016) investigated how efficient a virtual audience feedback is in improving public speaking during training. Training incorporated a combination of explicit feedback provided by visualised performance measures and implicit feedback provided by a virtual audience in real-time. Researchers compared three methods of feedback including a passive non-interactive audience (control group; n= 15), a passive audience with explicit feedback with a visual background (G1; n = 14) and an interactive virtual audience that provides feedback using nonverbal cues (G2; n = 16). Behaviours that participants were trained for were eye contact and avoiding pause fillers in two of the feedback presentations. Researchers compared learning outcomes based on a pre-training and a post-training test paradigm where improvement was measured on audio-visual nonverbal behaviour, the structure of the presentation and overall performance of presenters.

Presentations were assessed by experts based on pre and post training videos presented side-by-side for a direct comparison. Expert ratings revealed improvement in performance across all conditions. The highest improvement was observed for the interactive virtual interactive audience. However, when this was compared with improvements rated for the direct feedback group, this was not significant (p = .059). When comparing all groups, the virtual feedback group and the control group had better expert ratings compared to the direct feedback group suggesting that the feedback may be distracting. Objective ratings of pre-post session performances using filler words and eye contact revealed improvement across all conditions with no significant differences between each condition. Chollet and colleagues (2016) also found that those who performed well in the pre-training session did not benefit from the extra training (ceiling effect). Researchers suggest that future research should include a post-hoc feedback session.

2.7.1.2. Implicit – Avatar Reactions

One of the aims of SSP research is to mimic these perceptions using automated technology when developing embodied conversational agents or virtual agents. Essentially, multimodal embodiments are virtual characters or embodied conversational agents which interact with humans in a socially intelligent way by recognising multichannel signals and responding naturally. These systems interact with humans by recognition of their nonverbal skills and should match these behavioural cues (Cassell & Tartaro, 2007). There are several approaches in designing systems to accommodate learners' nonverbal cues, many studies have accounted for this by enabling the avatar to use back channelling cues such as nodding.

Research has found that monitoring social signals of learners can have a positive effect on learning (Lepper & Woolverton, 2002). The development of avatars has aimed to replicate this approach when designing adaptive educational systems and intelligent tutors (Baldassarri, Hupont, Abadía, & Cerezo, 2015).

Role-play avatars have also been known to assist learners in how to manage situations. For a project called eCIRCUS, Aylett, Vala, Sequeira, & Paiva (2007) aimed to improve empathy through interactive role-play with virtual agents by presenting the students with story dramas in a virtual school where the embodied conversational agents play the role of bullies, helpers and victims. After each episode, the student interacted with the agent who provided advice on how to manage the situation. This type of research promotes reflective thinking about social interactions. Other settings that have proved effective in understanding social situations have included enabling students to cope with bullying (Sapouna et al., 2009) and understanding of other cultures in children (Aylett et al., 2014). A natural interaction is important as this permits users to interact with the avatar and enables the user to trust the system. Another system was developed, in mental health, a system called MultiSense has been designed to interpret nonverbal behaviours to infer psychological distress (Stratou & Morency, 2017). However, these interactions are not directly aimed at improving communication skills using social signals.

There has also been research conducted as a means of improving job interview performance using technology-enhanced methods that uses automated recognition of signals for training. A study by Hoque, and colleagues (2013) developed an embodied conversational coach called MACH which captures facial expression and speech as well as generates speech and nonverbal behaviours in response by participant behaviour. An experimental design included three experimental groups (*n* = 90) which were gender-matched to avoid any gender-variability in behaviour. The control group (G1) watched educational videos on job interviews, the first experiment group (G2) practiced interviews with MACH and then watched themselves on video and the second experiment group (G3) practiced interviews with MACH and then watched themselves on video and received feedback about behaviours. All groups had taken part in an initial interview with a career counsellor where only experimental groups 2 and 3 were brought back for an hour-long intervention for a few days. All participants were brought back into the lab for a final interview with the career counsellor who was blind to the study conditions. Results revealed that counsellors' ratings of performance were significantly higher for those who had received MACH training intervention with feedback (G3) and without feedback (G2) than those in the control group (G1). Open-ended interviews revealed that participants felt that watching their videos were uncomfortable but great for learning. This study also found that smiling, pauses, speaking rate and filler words were the top attributes for visual feedback. MACH was rated as 80 using the System Usability Scale which is above the benchmark of usability at 60. However, this study only investigated facial expression and head poses which is limiting for understanding a holistic interaction.

Research by Damian and colleagues (2015) investigated the effectiveness of job interview training for teenagers using an interactive avatar. The study spanned three days: day 1 both the control and the experiment group took part in a baseline interview (pre-training), on day 2 the control group engaged in traditional training (training with book) and the experiment group interacted with the training system

(TARDIS) which reads and responds to users' nonverbal behaviours. On day three, students also took part in a 7-minute mock interview (post-training). Practitioners filled in questionnaires rating about students' performance which revealed no significant difference between groups on day 1. In contrast, a significant difference between groups was observed on day 3 where the experiment group significantly improved in their overall performance, whether practitioners would recommend students for jobs, in their appropriate use of smiles, eye contact and nervousness which the control group had not improved. Students also filled in a questionnaire where the experiment group rated themselves as less nervous than students who had received traditional training. While this study suggests that technology enhanced training was more effective in improving job interview performance than traditional methods, researchers did not say whether the practitioners were blind to trainee conditions. Additionally, the control condition relied on a single rating of a practitioner rather than obtaining ratings from neutral observers not present on the day to avoid any interaction influence.

2.7.1.3. Explicit – Behavioural Feedback Loop (BFL)

Previous research has shown that conversational aids (secondary information) presented during social interactions improve communication (Scherl & Haley, 2000). A particular framework for providing secondary feedback is the Behavioural Feedback Loop (BFL) which has been used as a method of augmenting social interactions and is required to suit the user, context and the scenario (Damian, Tan, et al., 2015; Damian et al., 2016; Damian, Dietz, & André, 2018). Research has shown that memory, motivation, decision making, and mood are important cognitive domains in personal augmentation (Xia & Maes, 2013). The feedback loop framework includes three learning paradigms such as observational learning (Shettleworth, 2009), operant conditioning (Skinner, 1938) and social cognitive theory (Bandura, 1986). It also includes two major components of training such as include reflection for improving self-awareness and action / practice which is important for learning and transferring skills (Damian, Baur, et al., 2016). Trainees tend to only improve in performance following training as real-time feedback is provided during the interaction.

Schneider, Börner, van Rosmalen, and Specht (2015) developed a Presentation Trainer (PT) which is designed to assist users who would like to improve their nonverbal skills in public speaking. Realtime feedback was provided though visual and haptic feedback which included body posture, use of gestures, voice volume, use of pauses, phonetic pauses, and ability to stay grounded without shifting while presenting which resemble dancing. In this quasi-experiment, 40 participants engaged in a five-minute lecture about nonverbal communication for public speaking and then took part in five successive training sessions. The control group were shown a version of PT which only contained a mirror image of themselves and the treatment group received haptic feedback (vibrations from a wristband) and visual feedback showing a mirror image of themselves which interrupted the presenter if the behaviour was too severe. Subsequently, participants were asked to fill in a questionnaire which assessed naturalness, invasiveness, boredom vs motivation, unlikelihood of free time use, learning perception and practice using tool vs classroom learning. Results showed that the treatment group were more motivated, they found the tool to be less invasive and learned more than the control group. Finally, results also showed that the treatment group produced fewer mistakes by the fifth session whereas the control group were consistent throughout sessions. However, users expressed that they initially found it difficult to pay attention to the feedback as well as give a speech; however, this improved by the fifth session.

A follow-up study conducted by Schneider, Börner, van Rosmalen, and Specht (2016) explored the use of PT outside of a laboratory setting by exploring to what extent an audience agrees with the PT that a presentation has improved as well as introduced PT to established training practices. It was reported that participants using PT were more confident using PT than the control group. Participants had presented a pitch to the PT and then to their peers who evaluated their performance and filled in a presentation assessment questionnaire. The PT was also used to assess these pitches. Results found that PT creates a more comprehensive learning environment for the acquisition of speaking in public in combination with already developed training practices.

Barmaki and Hughes (2018) provided feedback to trainees about their gestures using an avatar mediated interactive virtual training system using real-time visual and haptic feedback. All participants experienced the visual feedback and the haptic feedback systems. The researchers found that feedback of body language and gestures had a positive impact and participants preferred the system more than traditional feedback. Participants also expressed their enjoyment of the vibration feedback method.

The use of automated recognition and feedback in real-time may assist in improving communication skills in a variety of contexts. A study investigated improvements using wearable technology to provide constant real-time feedback about openness, body energy and speech-rate during public speaking (Damian, Tan, et al., 2015). This real-time feedback was demonstrated using a threshold intervention where if participants' behaviour exceeded the threshold, this would be flagged. Results found there was a significant effect for speech rate but not for openness and body energy and there is a need for feedback personalisation as some participants did not cross the predetermined threshold. Participants seemed to be reacting differently to the feedback in that some participants adapted behaviour gradually and others completely ignored it. In post-hoc interviews participants felt that it is important to know about their performance in relation to the predetermined threshold. Overall, participants found BFL helpful in adjusting their behaviour, it was not distracting, and it was a good point of reference for understanding how they were performing. However, the wearable was quite bulky and, as a result, participants were aware of the tech which influences their behaviour. A concern using these methods are that the feedback of social signals is often literature driven rather than data driven for context.

A study conducted by Bahreini, Nadolski and Wester (2017) investigated whether detailed feedback using a bespoke software that provides real-time feedback of facial and vocal emotional expression improved communication performance. Twenty-five participants engaged in an interactive game where they were asked to mimic the seven basic facial expressions and vocal emotions in a variety of situations including a dentist visit, a visit to a restaurant and a traffic accident. A within-participants design exposed all participants to feedback and no-feedback conditions. Results revealed that facial expression performance and vocal performance was significantly improved in the feedback condition. These findings are suggestive that providing feedback during training can lead to improvement. However, the paper focuses on ability to mimic rather than spontaneously generate relevant expression.

While these studies have shown efficacy for the BFL framework, it has been found to be distracting during training (Damian, Baur, et al., 2016). Some of these studies do not evaluate which could have impacted the results obtained. Users can absorb information if it is presented to them in batches compared to when presented to them sequentially. Key information provided by researchers are that the presentation of auditory cues while the user is speaking was also found to be very distracting (Tanveer, 2016). To solve these issues, research by Ofek, Iqbal and Strauss (2013) found that secondary information presented during a conversation can be absorbed by the user without others noticing and users need feedback during a task to improve behaviour.

2.7.1.4. Explicit - Post-Hoc Feedback

An alternative method of providing feedback is a summative (summary) and focused feedback technique detailed by Hoque, Courgeon and Martin (2013) in communication training. This method has proved effective in other research (Ali & Hoque, 2017; Tanaka, Sakti, Neubig, Toda, & Nakamura, 2015b). Summative feedback provides trainees with a summary of their interview performance. However, trainees felt that this type of feedback did not allow them to observe their behaviours throughout the interview and how it changed overtime. Subsequently, trainees were given the choice for focused feedback which enabled them to watch their own video for reflection and view their nonverbal behaviours and how they change over time on a dashboard as a function of time which allows participants to view their behaviours across multiple modalities which are in sync with one another. This technique of providing feedback for improving skills could be useful for reflection of trainee and discussion with a trainer. Research has shown that feedback provided by a peer in combination with a tutor has been effective in improving performance (Mitchell and Bakewell, 1995).

Zhao and colleagues (2017) collected data from an online platform (ROC Speak) that enables anyone access to communication skills training for job interviews with feedback on smile, body movement, filler words and voice modulation. Individuals were also provided with an overall assessment and comments by peers (positive or negative) and the possibility to playback their videos for reflection. Participants were randomly assigned to either the experimental group that received full feedback from ROC Speak system or the control group that received feedback from peers online. Results revealed that those in the experimental group had significantly improved in speaking skills, friendliness, vocal variety and articulation where the control group did not improve. These results are further evidence for the efficacy of multimodal feedback of social signals in improving communication skills in comparison to traditional methods. However, the ROC Speak system is sensitive to environmental factors which prevents them from being evaluated in a real-world setting. Furthermore, peers in this context are not experiential trainers which is needed for effective training.

More recently, Chollet, Ghate and Scherer (2018) developed an adaptive virtual agent for training social skills which reacts to the users' (medical students) performance which was automatically detected by the system. The system captures verbal and nonverbal signals (facial expression) and provides an after-action report. This system has yet to be evaluated for its training efficacy.

2.7.2. Summary

The BFL is particularly relevant for real-time feedback during performance of a task; however, a concern of this method is the identification of the feedback by users, processing and corrective action whilst engaging in the task which may result in a cognitive overload resulting in performance decline. There is are no studies investigating methods of feedback that enable personalised, summative and focused feedback in the context of media interviews which would provide more personalised skills training and provide better training outcomes targeted at an individual level. A table summary of the literature that is related to this these can be seen in Appendix 2.2.

Overall, a method which provides feedback post-event in summative form which is focused could be more useful in a training context that includes multiple training sessions. The next section investigates interactive training avatars and their efficacy of managing social interactions and improving social skills.

2.8. Specific Focus of PhD

Communication skills training is based on defining a target skill, modelling it, role-playing it, providing feedback and improving self-awareness which should be conducted in multiple sessions (Liu, Huang, Gao, & Cheng, 2017). Research shows that effective implementation of communication skills training programmes depends on the experience and expertise of trainers, students' willingness to learn, the programme's ability to improve self-awareness of emotional communication (as emotional intelligence is imperative for self-awareness) and in-depth feedback provision (Aspegren, 1999; Bahreini et al., 2017; Roter, 2004).

Implementation of training assists users in understanding their skills prior to training and guide their development throughout training which maintains positive motivation and allows trainees to be exposed to an interaction which could enable them to be calm in that setting. This could extend beyond the training session. Similarly, social cognitive theory proposes that trainees / students acquire skills through practice and feedback (Mann et al., 2011). Kolbs' (1984) experiential learning cycle theory states that learning is an integrated process of experience, observation that is reflective, conceptualisation and experimentation. An effective training program enables trainees to maintain their skills over time (Aspegren, 1999).

Many communication skills programmes have similar training techniques. In 1969, Sidney and Argyle developed a training technique to improve communication skills / social skills which includes roleplaying mock interview, playback and by reflection with the trainer (see Argyle, 1988). This has been found to be the most effective method for improving general communication skills (Argyle, 1988). This allows trainees to be aware of behaviours, such as their vocal behaviour (from voice recordings), their facial expression (from video recordings) and their use of gestures (from video recordings).
Furthermore, performance has been shown to be highly dependent on the trainer's experience as feedback is more valuable (Aspegren, 1999; Damian, et al., 2015).

Self-awareness is particularly important for communication skills improvement as those who are self-aware provide accurate accounts of their behaviour and their behaviours are consistent (Wicklund, 1979). This is also often seen in leaders (Aspegren, 1999; Pentland & Heibeck, 2010; Roter & Hall, 2004; Wicklund, 1979). In communication skills training, feedback of trainees' performance during mock interactions is used to increase self-awareness and have a better self-perception of their behaviour (Aspegren, 1999). Effective ways of improving self-awareness are to recite a script in front of a mirror. Mimicking pictures has been found to be successful for improving facial expression while voice expression can be improved by recording oneself speak and evaluating the recording (Aspegren, 1999).

There are several settings investigated which have shown to change behaviour. These include job interviews (Carl, 1980; Hollandsworth, Hazelskis, Stevens, & Dressel, 1979), information-sensitive conversations (Liu, Huang, Gao, & Cheng, 2017) and public speaking (Damian, Baur, et al., 2016; Damian, Tan, et al., 2015; Schneider, Börner, et al., 2016). Several systems have made use of mobile phones to provide feedback during face-face conversations (Lee et al., 2013), public speeches (Saket, Yang, Tan, Yatani, & Edge, 2014), group meetings (Lee et al., 2013) and presentations (Damian, et al., 2015). Research has also explored video conferencing using a laptop monitor for presentation of feedback (Tan, Schöning, Luyten, & Coninx, 2014).

2.8.1. Facilitators of Learning

Emotion has also been found to be a learning facilitator and an influencer in training and success (Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011). This success is a result of attention that is mediated by cognitive mechanisms of processing, storage of information and the reinstitution this information (Pekrun et al., 2011). Performance in learning has also been hypothesised to be mediated by intrinsic and extrinsic motivations associated with academic work (Pekrun, 1992).

Another factor that can influence training is the environment which has been found to impact communication / social skills training, for example negative environments prior to a job interview can affect job interview performance (Gebhard et al., 2018; Schneeberger, 2018). Positive emotions are perceived as positive in most instances; however, negative emotions can be ambivalent. In contrast, a learning cycle model proposed by Kort and Reilly (2002) has stated that there are six bi-polar emotional dimensions that arise during learning. These are frustration-euphoria, dispirited enthusiasm, humiliated-proud, terror-excitement, ennui-fascination and anxiety-confidence which are all experienced by students. These studies suggest that environment and emotions affect learning which could influence learning in training.

2.8.2. Skill Maintenance

Skills that have been maintained or transferred from training to the workplace are evidence of effective training. In general, examination of training effects over time is also relevant to considering the so called “transfer problem” in training research (Psychology, 2015). It is estimated that only 10% of training results in behavioural changes (Burke & Hutchins, 2007). A survey suggested that 40% of trainees fail to transfer skills immediately after training and 70% lost momentum after 1 year after training sessions (Saks, 2002). Given these results, skills transferred from training to practice is often low.

A meta-analysis on the efficacy of communication skills training revealed a moderate effect of the efficacy of communication skills which is promising, particularly training health professionals (Barth & Lannen, 2011). However, studies included were non-randomised and some had no specific training intervention in the control group. This efficacy has been shown to be moderate for communication skills training programs (Davis, Thomson, Oxman, & Haynes, 1995; Roter et al., 1995; Smith et al., 2000). This could be a result of exercising this skill after training in everyday life and the workplace.

The transference of skills post training could be influenced by the perceived utility or value of the training. This is influenced by the trainees’ evaluation of a recognised need to improve in a specific area that the training offers, the belief that applying the new skill will improve performance, the credibility of the skills and the practicality of the new skills for ease of transfer (Burke & Hutchins, 2007; Yelon, Sheppard, Sleight, Ford, 2004). Additionally, Campbell and Stanley (1963) have recommended that there should be test-retest intervals that include one month, six months and a year after initial training which assess long-term effects. This is because trainees have time to engage with the information provided during training and could apply it in real-life scenarios. To date, contexts investigated are limited to communication in the classroom, in those with social impairments, in the healthcare setting, in job interviews and in public speaking. There are no which assess the skills gained during training after a few months, this is taken for granted in the augmented training domain as this is a key component of effective training.

2.8.3. Media Interview Skills

The nature of communication in this context is unique; while similar to job interviews, in some respects, interviewees should communicate in a concise manner and reach a wider audience from all professional and cultural backgrounds. Media training manuals typically suggest some behaviours which should be avoided in media interviews. These are lack of vocal conviction, lack of eye contact, fast speaking rate, monotone voice and hesitation. These signals are an indication of nervousness, uncertainty and boredom and influence how the interviewee is perceived by the audience (Taylor, 2015). An additional behaviour that can be interpreted as boredom is excessive movements such as swaying and rocking, particularly when the other person in speaking (Tao & Tan, 2009).

The specific focus of this thesis is on media skills training. The use of a combination of nonverbal signals could be important for a good media interview as combinations of signals have different meanings. This includes mirroring interviewer’s movements, maintaining eye-contact and smiling. Together, these actions suggest that the interviewee is friendly (Ho et al., 2015; Taylor, 2015). Smiling has also been found to indicate confidence, honesty and dominance (Lapidot-Lefler and Barak, 2012; Knutson, 1996).

There are a limited number of empirical studies that explore the relationship between observable nonverbal behaviours and observer subjective judgments within the context of media interviews. Studies generally focus on small samples of interviews with high profile interviewees such as politicians. For example, Babad (1999) correlated observer judgement of global impression (positive/negative) created in a media interview with a set of observer judgements in relation to observable behaviour. This study focussed on the behaviours of five interviewers conducting televised political interviews and found several universal patterns across these individuals. The behaviours which appeared to create a positive impression included smiling, a relaxed face, nodding and round hand movements. Conversely, the behaviours associated with negative judgements included beating hand movements, leaning forward and blinking. Studies such as this have typically been small scale given the challenge of hand coding the non-verbal communicative behaviours under study.

2.8.4. Summary

No studies investigate signal displays in media interview within the SSP domain. Communication skills training has shown to improve nonverbal communication in the long term. This is because a training program is effective and is a result of attaining skills through raising self-awareness and not solely through practice. For training to be successful it is down to the experience of the trainer in combination of the willingness of the trainee and whether the training improves trainee self-awareness. While there seems to be low maintenance of skills for training in general, there is a moderate effect of skills maintenance for communication training skills. Research in the context of media interview skill training is limited, particularly in the maintenance of skills. Context is vital for a deeper understanding of interactions as opposed to shallow interactions. The focus of this thesis is to investigate improving communication skills training in media interviews. An important component of training is the method of feedback during training as this improves self-awareness which this section shows as a major component in improving communication performance. The next section presents research using technology to improve performance.

2.9. The Research Gap

A review of the literature revealed that there are little to no research conducted in social signal processing in identifying the signals important for communication skills in media interviews, which is limiting as conversations in media interviews are unique and are important to understand. Research is often limited to a few modalities in communication which is restrictive in understanding the complexities in communication. The preferred social signal feedback technique in the literature is the behavioural feedback loop; however, this technique has been found to be distracting resulting in an increase in cognitive load. Summative and formative / post-hoc methods seem to be more effective in improving communication skills as they allow for reflection. It was also identified that based on early research in social psychology and recent research in social signals processing that the first 30 seconds are when first impressions (thin slices) are formed in a social interaction. Based on these research gaps the premise of this research is posed in one overarching research question (RQ):

Can communication skills be enhanced using commercial automated technology in the context of media interview skills training?

Using the simple model of communication, the Shannon and Weaver model of communication, participants send an expressive message via different communication channels (facial expression, vocal expression, hand gestures, body movements and so on) to the recipient who decodes the message and responds according to how they respond to the message.

The next chapter compares different methods for collecting data for this thesis, justifies the research design of each stage of this PhD research and provides evidence and details of adherence to ethical guidelines. The next section also offers justification of the use of different statistical packages used for data analysis which is useful for understanding data handling.

CHAPTER 3. RESEARCH DESIGN AND METHODOLOGY

3.1. Introduction

This thesis includes four research stages. The first stage was to investigate the relevant social signals for effective communication in media interviews. The second stage was to assess the most appropriate method for providing social signal feedback to trainees that is both actionable and understandable. The third stage was to investigate whether provision of the signals identified in the first stage using the method of feedback identified in the second stage was more effective than traditional methods of communication skills training, such as recording and taking part in mock interviews then discussing trainee's performance with trainer. The fourth stage was developed to assess whether skills gained in the third stage were maintained after 6 months.

This chapter justifies the methods used to accomplish the objectives of this research. This chapter is divided into nine sections; 1) philosophical research paradigm; 2) research questions; 3) research context and target populations; 4) research design 5) communication skills evaluation 6) usability assessment; 7) technology used; 8) data handling software and 9) ethical approval and procedure.

3.2. Philosophical Research Paradigm

Scientific research philosophy is the approach that researchers adopt to design, conduct and interpret their research to obtain new and reliable knowledge. Typical paradigms include positivist (quantitative), interpretivist (qualitative), transforming (discrimination prevention) and pragmatist (mixed methods) (Žukauskas, Vveinhardt, & Andriukaitienė, 2018). The approach taken to develop the methodology for this research is from the standpoint of the pragmatist paradigm. This approach is focused on problems and its application in the real-world and uses mixed methods but is ultimately focused on 'what works' (Brierley, 2017). It draws on methods used from both positivist (experimentation) and interpretive (interpretivist) paradigms. In this thesis, two of four research studies include mixed methods.

3.3. Research Questions

The overarching research question is:

Can communication skills be enhanced using commercial automated technology in the context of media interview skills training?

This section details and justifies the design of the research that is focused on answering the research aims of each research question. Each research stage has its own aim which can be seen in Figure 3.1.

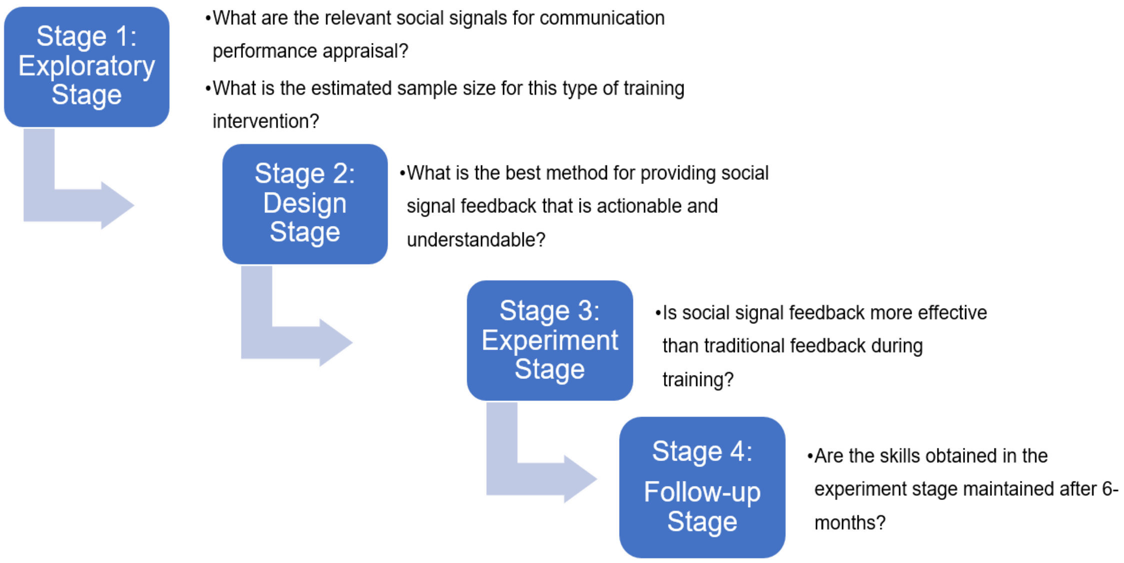


Figure 3.1. Four research stages and research questions

3.4. Research Context and Target Populations

The overarching context of study was derived from real world media skills training course which were delivered in the context of large organisations. Detailed observations of these exemplar courses were conducted to understand the typical training steps. This was then used to design a model of the training under controlled conditions in a university setting.

The observed training consisted of a lecture style introduction followed by simulated interviews. Simulated interviews were recorded and played back to participants. The recordings and playback of simulated interviews were done to enable the trainer and the participant to reflect on their performance and to provide participants with feedback to improve self-awareness. Research has shown that improving self-awareness improve communication effectiveness and is a skill that is observed in leaders (Hass & Eisenstadt, 1990; Wicklund, 1979).

3.4.1. Trainer Experience

Successful communication skills training relies on the expertise of the journalist / trainer (Aspegren, 1999). An experienced trainer has the ability to identify areas for improvement during training. Therefore, the studies reported in this thesis used trained / experienced journalist to conduct simulated interviews with participants.

3.4.2. Sampling Population

The sample selected for communication skills training were research staff and students. Method used to recruit participants was purposive sampling. An advantage of this method of population recruitment is that is a fast method and allows for a large sample size to be recruited (Field, 2013). However, it can be a subjective method of recruitment. To ensure no bias was introduced to recruitment of participants, researcher accepted all participants who responded to the recruitment email expressing an interest in training.

The academic / researcher / research student population are relevant to this training context because research is often reported on the news; moreover, they are representative of professionals who are likely to engage in media interviews (Reed, 2001). This population are also literate and educated and obtain skills to undertake a media interview. Finally, it is valuable to run this research at a university as it provides easy access to recruiting possibles trainees. The target sample size recruited in each research stage is discussed in the following section.

3.5. Research Design

This section details the research design for the exploratory stage, design stage, experiment stage and the follow-up stage. Description of the methods used and the intended sample size of each stage.

3.5.1. Exploratory Stage

The aim of this stage is to explore the social signals transmitted between two individuals and identify which of these are key for estimating successful communication in a media interview. Affect recognition COTS affect recognition technology was used to capture signals exhibited in interviews between the journalist and the trainee. Data was collected in context of an already existing training programme. This bespoke training programme typically included a 2 - hour interactive lecture on communication in media interviews with examples of a good and bad interviews. The duration of the workshop typically lasted 8 hours. Questions and discussions of content were made possible throughout training. This is an important step in communication skills training and in understanding the procedure of media interviews. The content of the lecture cannot be shared as the design of the workshop is the protected property of the media trainers. The workshops included two simulated interviews which were recorded; a radio interview and an on-camera interviews. Both mock interviews were watched back by the trainees and trainers and trainees' performance were discussed.

This training programme was then rearranged to meet the aims of this research; the length of the lecture was reduced significantly to 1-hour, which was feasible according to the trainers, and participants were provided an example of effective and poor communication in media interviews. Participants also took part in a radio interview followed by an on-camera interview.

In order to study the signals associated with good and bad media interview performance in the context of media training, the decision was taken to use an observational research method. This was done to collect data in a situation as close to a real environment as possible and therefore provide some basis for ecological validity. Thus, this research used the simulated context of a real training class /

workshop but with the addition of technology to record social signals. For full details about this stage see Chapter 4.

To collect ground truth data to classify good and bad communication skills it was decided to use standardised / validated questionnaires to collect subjective observer judgements. An exploratory approach taken to data analysis and it was decided to use k-nearest neighbour. K Nearest Neighbour (*k*-NN) is a nonparametric supervised machine learning algorithm and is often used for pattern recognition (Witten & Frank, 2005). There are several advantages of using *k*-NN; firstly, it can be simply implemented; secondly, it is a nonparametric algorithm which has no assumptions about the dataset and, as a result, is applicable to real-world data and; finally, has been used in multiple contexts including finance, video recording, image recognition, political science and healthcare (Witten & Frank, 2005). Furthermore, *k*-NN is an instance-based classifier which generates a prediction based on the "similarity of the query to its nearest neighbour in the training set" (Utgoff et al., 2011). This contrasts with decision trees or neural networks which creates an abstraction from data instances. Instance-based learning refers to a class of procedures for solving prediction problems based on past issues. Use of an instance-based classifier in this research stage is beneficial as results produced is based on the dataset (social signals captured) and labels (good and bad interview performance). This approach to prediction classification is based on the dataset itself and is ideal for contexts which do not have predetermined predictions about performance or a large dataset (i.e. a data corpus).

3.5.1.1. Evaluation of Performance

Evaluation of participants performance was important as this would reduce bias when identifying effective and poor communicators. This was also done to identify relationships between patterns of emotional / nonverbal signals and trainee performance evaluations, as rated by humans.

To obtain objective judgements of trainees' performance, participants interviews were rated by a trainer and three neutral observers using a communication evaluation questionnaire. There were several approaches taken to reduce the subjectivity in the ratings of performance. Firstly, because trainers had interacted with the trainees on the day of training which would have likely influenced their scores as a result of an interaction impression (Meissel, Meyer, Yao, & Rubie-Davies, 2017), additional ratings were obtained by three neutral observers who were not present on the day of training (Landman et al., 2012). Ratings obtained from three neutral observers were intended to act as an audience by being able to review both interviews multiple times for a more thorough rating as well as provide more realistic ratings (Naim et al., 2016).

Secondly, to further reduce the potential to rating bias the neutral observers were blind to the ratings provided by the trainer. Collection of this data was also used to measure the inter-rater and intra-rater reliability of the ratings obtained and the internal consistency of the CSRS in the context of media interviews.

3.5.1.2. Thin Slices of Behaviour

Social signal data was captured for the duration of the media interview and an exploratory analysis was conducted to decide a time window size for analysis (i.e. the first 30 seconds or the last 30 seconds of mock / simulated interviews). The first 30 seconds proved effective in predicting social signals that are mapped onto ratings by human judgements.

3.5.1.3. Sample Size

The intended sample size for the exploratory stage was chosen based on a compromise power analysis for linear multiple regression and taking into account the practical constraints on training group size that can be achieved per session (n = 20). The exploratory stage included two mock interview sessions which would result in a sample size of (n = 40).

However, when conducting exploratory research there were many factors which affected recruitment of 40 participants to consider, such as the suitability of statistical tests (power), time available for the study, recruitment and funding. The nature of the study did not allow for a large sample size to be recruited because: 1) only 5-6 participants could be successfully trained in a single media skills training workshop; 2) the costs for running the training day were beyond the originally planned budget and 3) there were time constraints for this project overall set by funders. As a result, three training days were conducted resulting in a total of 17 participants recruited. This pilot study informed the development of the feedback technique / method.

3.5.2. Design Stage

The intervention design stage is the second stage of this research. The overall aim was to determine the most understandable method of providing social signal feedback based on the signals identified in the exploratory stage as necessary for performance appraisal. A user-centred design method was used to collect participants views to inform and refine the design choice of feedback in the intervention evaluation stage.

Mixed methods approach was used to gather data in this stage as it allows the researcher to understand any contradictions between quantitative and qualitative findings. For the qualitative component semi-structured interviews were used. There are some advantages and disadvantages of conducting semi-structured interviews. An advantage is that it allows users to express their views beyond questions asked that acts like an extension tool as interviewees can ask questions which allows for content to be analysed later. This is in contrast to structured interviews in which content is restricted to the questions asked and is typically used for clinical diagnosis (Opdenakker, 2006). The second advantage is that it provides reliable data as it is more focused on the research aims in comparison to open interviews where there is a possibility of gathering useless information. The third advantage is that it establishes a two-way communication in comparison to structured interviews which is very linear. Finally, this interview technique can also let interviewees discuss concerns which the survey could not have highlighted in isolation or in questions originally developed.

In contrast, there are some caveats of using this technique, the first is that the researcher conducting the interviews should have some interview skills as well as follow a correct procedure to analysis the data and the questions should be carefully planned out (Cohen & Crabtree, 2006). Focus groups could have also been conducted; however, this technique is usually focused on group feelings, perceptions and opinions and some may not feel the need to participate.

The method used to collect opinions about feedback choice was a User-Centred Design (UCD) which is an iterative design process that centres on the users' needs in each stage to develop and improve the design (Gulliksen et al., 2003). Each of the methods of feedback designed considered the context of use, the requirements of the user, the design solutions and the evaluation of requirements. Each design was evaluated in the form of semi-structured interviews and usability rating scale (see section 3.7) to gather qualitative information about the look and feel of the interface. Interview questions developed for this were aimed at gathering information about both good and bad features of the participants preferred interface and less preferred interface.

Participants who had volunteered in the exploratory stage were invited back to share their thoughts on several design alternatives that were developed in prototype form. The recruitment of these participants allows for the evaluation of the different versions of feedback without running a new study to collect new social signals data. Additionally, participants could view their performance feedback using multiple methods resulting in a more realistic evaluation of feedback sessions. See Figure 3.2 for the UCD process used. Full details about this research stage can be found in Chapter 5.

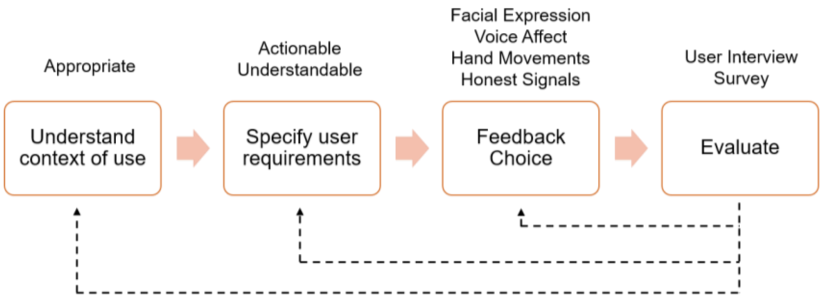


Figure 3.2 User-Centered Design Process and Methods

3.5.2.1. Sample size

A study by Virzi (1992) found that 5-6 participants were enough to assess usability of a system in usability studies. The distribution of the data at a sample size of 8 looks the same as 30 (Bangor, Kortum, & Miller, 2008). It is proposed that the 'magic number is five for early stage usability studies' (Virzi, 1992). Researchers found no additional details by including more participants (Virzi, 1992). There is some debate over this in the HCI literature, but for this research there was a limited pool of potential participants. As a result of this, the current research stage aimed to achieve at least five participants to get insight into design choice.

3.5.3. Experiment Stage

The experimental stage of this research is the third stage. The aim was to investigate whether communication skills can be improved using the training intervention designed compared to the standard feedback training received in media skills training workshops. An advantage of using experimental design for this research stage is that experiments are objective i.e. the views of the researcher do not affect the results of the study resulting in a more valid study than a pseudo experimental design where manipulation of an independent variable without the random assignment of trainees to a condition. Control of extraneous variables is also an advantage as this also results in a less biased dataset (Field, 2013). Experiments are replicable and extraneous and independent variables are controlled which enables a cause and effect relationship to be recognised (Saldaña, 2014).

A comparison of pre-training and post-training design was implemented in this research study. This experimental design compares a change which occurs between groups on a dependent variable (outcome measure) at two time periods. Figure 3.3 demonstrates the analysis that can be performed this type of analysis.

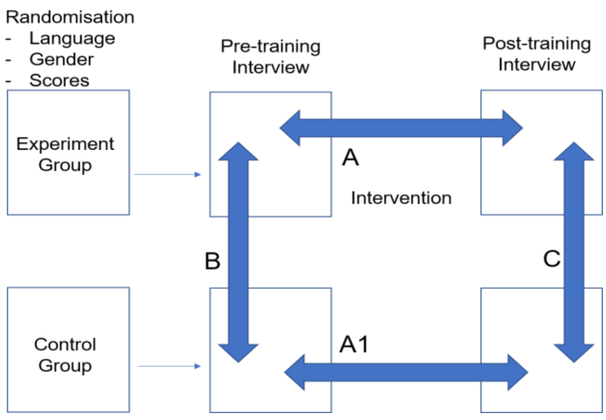


Figure 3.3. Experimental design. Analysis which can be done include how groups have changed from pre-test to post-test (A and A1), compare final post test results between groups suggests overall effectiveness of the intervention (C) and compare skills in the pre-test to ensure that randomisation was effective (B).

Participants were randomly assigned to be receive either social signal feedback or standard feedback after each practice/ mock interview. Randomisation was done following the baseline test, the group was split into pairs matched closely as possible by gender, first language and average score on baseline communication skill ratings. One member of each pair was allocated to the experiment / control at by the toss of a coin (random) (Allen, 2017). This aims to reduce the variability between conditions which may be attributable to extraneous variables such as gender differences and language concerns. For full details about this research stage can be found in Chapter 6.

3.5.3.1. Sample Size

A power calculation was conducted based on the training improvement from one session to another in the exploratory stage. This was done to reduce the risk of oversampling or under sampling the study (Erdfelder, Faul, & Buchner, 1996) and to establish an estimated sample size for this type of training intervention.

3.5.4. Follow-up Stage

The follow-up stage of this research is the final stage of this research. The aim was to determine whether skills obtained in the intervention evaluation stage were maintained after 6 months as skill maintenance is a hallmark of successful training (Aspegren, 1999). A 6-month time frame was selected as research has proposed that skills be investigated at 3- and 6-months post training to evaluate successful training. The point at which ethics was submitted and approved allowed for the 6-month mark to be evaluated.

The overall approach of this research was a post-experimental follow-up as participants had remained in the groups originally assigned to them in the experiment stage. A comparison analysis was conducted to observe any differences between groups. Full details surrounding this research stage can be found in Chapter 7.

3.6. Communication Skills Evaluation

There are many measures of communication skills effectiveness which can be seen in Table 3.1 along with issues surrounding their use. The Communication Skills Rating Scale (CSRS) was selected for use in this research project. Reasons will be discussed in this section.

3.6.1. Communication Skills Evaluation

The complexity of communication competence is illustrated by the number of questionnaires developed. To identify the most appropriate questionnaire for your research, Spitzberg and Adams (2007) proposed several approaches to identifying which one is relevant according the aims of the research. These include: “which competence domain will be assessed?”, “what will the relation be to valid social outcomes?", "will assessments be dispositional (trait) or episodic (state)?", "will skill and competence be framed within a specific context?", "will assessment of competence and skills be made by interactants, third parties or both?" and "how will assessment be operationalised?". These were all taken into consideration when selecting a communication skills evaluation questionnaire. An answer to these questions were; the context which will be assessed is communication in media interviews, there should be no social relation between trainees and trainers, assessments will be ratings of nonverbal signals, the assessments will be made by interactants and third parties (neutral observers) and assessment will be operationalised by discussion of communication skill ability following performance evaluation. Table 3.1 presents published communication assessments and concerns for use.

Table 3.1 Published Communication evaluation assessments and concerns

Scale	Developer	Items	Current Research Issue
Communicative Competence Scale (CCS)	(Wiemann, 1977)	36	Limited research Only used in student populations Limited correlation between peer perceived and expert competence Focused on one sided than dyadic interaction
Interpersonal Communication Competence Scale (ICCS)	(Rubin & Martin, 1994)	17	Lacks construct validity
Interpersonal Communication Skills Inventory (ICSI)	(Bienvenu, 1971)	54	Personality traits rather than behavioural
Communication Skills Questionnaire (CSQ)	(Takahashi, Tanaka, & Miyaoka, 2006)	29	Personality traits rather than behavioural
Communicative Effective Index (CEI)	(Lomas et al., 1989)	16	Atypical population
Communicative Activity Log	(Pulvermüller & Berthier, 2008)	38	Atypical population for comprehension
Communication Flexibility Scale (CFS)	(Martin & Rubin, 1994)	12	Needs more validation
Communication Functions Questionnaire (CFQ)	(Burleson, Kunkel, Mortenson, Samter, & Xu, 2003)	30	no basis for comprehensiveness of skills and scores differ between gender
Communication Adaptability Scale (CAS)	(Duran, 1983)	25	too abstract to make inferences about skills
Conversational Appropriateness and Effectiveness (CAE)	(Spitzberg, 1991)	20 / 20	Appropriateness is not the aim of this research and is abstract
Interpersonal Competence Questionnaire (ICQ)	(Buhrmester, Furman, Wittenberg, & Reis, 1988)	40	Trait measure
Social Performance Survey Schedule (SPSS)	(Lowe & D'Ilio, 1985)	100	Too long to administer and relatively undifferentiated
Social Skills Inventory (SSI)	(Riggio, 2005)	105	Too abstract leaving room for inference
Communication Patterns Questionnaire (CPQ)	(Futris, Campbell, Nielsen, & Burwell, 2010)	11	Focuses on marital communication
Conversation Skill Rating Scale	(Spitzberg & Adams, 2007)	25 / 5	Limited research assessing cultural differences

For this research, the researcher has chosen the CSRS (Spitzberg & Adams, 2007). The CSRS is a simple 30 item questionnaire which is an instrument for assessing self or other interpersonal skills in the context of a conversation. It combines both verbal and nonverbal behaviours and has been found to be valid in a variety of contexts (i.e. instructional contexts). This scale was developed to address several assumptions about the nature of communication competence.

There are versions of the CSRS that have been developed for self-report, instructors and observers. (Spitzberg & Adams, 2007). The different versions can be seen in Appendix 3.1 – 3.2. It can be applied to a specific conversation or in general. The CSRS provides feedback on interpersonal skills in the classroom and in a variety of populations (Spitzberg, Canary, & Cupach, 1994). The questionnaire can be broken into two parts, the first consists of 25 behavioural items and the second consists of 5 general communication impression items. The latter are called molar items and they are used to validate the behavioural items. Molar scores are rated using a Likert scale from 1-7 on the following: poor conversationalist to good conversationalist, socially unskilled to socially skilled, incompetent communicator to competent communicator, inappropriate communicator to appropriate communicator and ineffective communicator to effective communicator. The behavioural items are subdivided into four clusters: attentiveness, expressiveness, composure and coordination. The behavioural scale has been found to be more reliable when added all together to produce a single composite score (Spitzberg et al., 1994). The scale is measured on a 5-point competence continuum (inadequate, fair, adequate, good and excellent). It is a fast questionnaire to administer (~ 7 minutes) and has good validity and reliability consistently above $r = 0.80$.

Spitzberg and colleagues (1994) developed this questionnaire as they argued that the majority of existing questionnaires had a lack of evidence for validity. Most questionnaires are developed to measure traits or involve technology and activities which cannot be applied to many contexts (Rubin & Windahl, 1986). This also restricts their use in the context of the current research project.

A concern of using the CSRS is that it has been questioned as to whether it takes cultural differences into consideration; however, a study by Matsufuji (1993) found that there are some similarities between cultures in communication (ranging from $r = 0.19 – 0.73$) (see Spitzberg & Adams, 2007). However, the CSRS is more focused on conversational behaviours and contextually specific in application. As a result of this, this questionnaire will be used to evaluate communicative performance in media interviews for this research. More specifically, the molar scores produce a very high validity coefficient ($r = 0.91$) in comparison the behavioural item score (see Spitzberg & Adams, 2007). The molar scares were used to measure how an interviewee communicates as it provides information regarding the general impression of communication ability which is useful for this training context.

A recent study used the CSRS to evaluate participants conversation skills using the CSRS in the context of improving social skills in speed-dating (Ali et al., 2018). However, there has been no research in the context of media skills training using this questionnaire. As there has been little to no studies in communication skills in the context of media skills training, validity and reliability research on the CSRS suggests that it could be useful for evaluating communication performance in media interviews.

3.7. Usability Assessment

Usability assessment was conducted as part of the design stage. There are a number of usability evaluation questionnaires designed to measure the perceived usability of interactive systems. These include the system usability scale (SUS) (Brooke, 1996), Questionnaire for User Interface Satisfaction (QUIS) (Chin, Diehl, & Norman, 1988) and Computer System Usability Questionnaire (CSUQ) (Lewis, 1995).

To assess the usability of designs developed, the SUS was chosen as it has been found to be the simplest and yields the most reliable results across samples with high Cronbach's Alpha of 0.92 (Bangor et al., 2008; Tullis & Stetson, 2004) which hold true in relatively small sample sizes such as 12 or in as little as 5 user ratings (Tullis & Stetson, 2004; Virzi, 1992). It is also flexible in assessing usability and has been used for assessing usability of desktop applications, mobile phones, interactive-voice response systems and websites (Bangor et al., 2008). Similarly, research has found that while it is limited in what it explores, it is possible to add questions to clarify to make it more specific. The SUS is a 10-item Likert Scale questionnaire ranking each question from 1 – 5 where 5 suggests users completely agree with the item statement and 1 suggests users completely disagree (Brooke, 1996). A score below the benchmark of 68 suggests there are substantial concerns with the usability of the interactive interface which should be addressed while the above suggests a working interface. A score of 80 or higher is a very high SUS score (Kortum & Bangor, 2013).

Contrastingly, there limitations of using the SUS. One of these is that it is a subjective measure of perceived usability. It has been noted that SUS should not be the only method of evaluating usability and researchers should include a more objective method of extracting views of usability. In a retrospective paper Brooke (2013) stated that usability is a combination of effectiveness, efficiency and satisfaction. Both effectiveness and efficiency deal with pass vs fail and time of task. This was Brooke's justification for adding an objective measure when using the SUS in usability studies as it more powerful when used alongside other tools. Additionally, the SUS is not a diagnostic tool in as much that it does not assess ways of improving the usability score by identifying problems with the system.

In this research, the SUS is used in conjunction with semi-structured interviews which identifies which of the multiple displays the users prefer as well as ways in which we can improve the rating (Appendix

3.3). The semi-structured qualitative interview probing questions were designed in a way that would assess both the preferred and less preferred method in order to obtain a well-rounded amount of information on both types of feedback. Details about questions can be seen in Chapter 5.

To identify themes in interviews, recordings were transcribed verbatim and themes were identified using thematic analysis (Braun & Clarke, 2006). This method of analysis is commonly used for examining and identifying themes within data and is purely exploratory in contrast to other methods of analysis for qualitative interview datasets (i.e. grounded theory where themes identified are theory driven). Themes are described as patterns in datasets which are important in answering the research question.

The advantages of using thematic analysis are that it is flexible and exploratory by nature (Braun & Clarke, 2006), the interpretation of these are supported by data (Saldaña, 2014), themes are not related to individuals experience (Saldaña, 2014) and identifies themes to emerge from the data (Braun & Clarke, 2006). However, there are a few disadvantages in using TA; reliability is of concern as data can vary between researchers and situations (Guest, MacQueen, & Namey, 2012), it can disregard some data which could be useful in hopes of developing a theme, it has limited interpretive power if analysis excludes a theoretical framework (Braun & Clarke, 2006) and does not allow researchers to make claims about use of language. While this is often said to be a limitation, this could also be interpreted as an advantage as use of language can often be misinterpreted. TA will be used in this research as it is often used with a wide variety of research questions and is exploratory (Braun & Clarke, 2006).

3.8. Technology Used

The use of COTS technology allowed all social signals to be tested as developing bespoke systems would not be feasible in time constraints of thesis. Multimodal signals captured for modelling human interaction in this research were facial expression, physiological processes (somatic markers), hand gestures, voice affect recognition and ‘honest signals’. COTS technology was used to capture these signals and were synchronised to one second. This was done to assess which combinations of signals are shown simultaneously to synthesise the interaction more effectively (Vinciarelli et al., 2011).

Accuracies of each software will be reported using Receiver Operating Characteristic (ROC) measures as these demonstrate the diagnostic ability of a system which is based on a curve which is created by its true positive rate against the false positive rate. A ROC score ranges from 0 to 1 where a score of 1 demonstrates a perfect classifier (Chawla, Bowyer, Hall, & Kegelmeyer, 2002). In addition to the COTS technology used in this PhD and their accuracy, this section also introduces media technology used to capture media interviews.

3.8.1. Media Technology

Media technology was used to record media interview. Technology used included a Sony PJ220 camera (facial expression in exploratory study [Chapter 4] and follow-up study [Chapter 7]), Manfrotto photographic tripod (for camera), external Logitech HD Webcam C525 (facial expression detection in experiment study [Chapter 6]), Lenovo ThinkPad t460p laptop, a Zoom H4n Digital Voice Recorder (voice recording for exploratory study and follow-up study) and a Lishuai LED 312 Panel light diffuser (facial expression). These can be seen in Figure 3.4.

A B C D E F

Figure 3. Manfrotto tripod, C) Logitech Camera, D) ThinkPad Laptop, E) Voice Recorder and F) Light Diffuser4. Media technology used to capture and simulate a media interview. A) Sony PJ Camera, B)

The literature review identified the following nonverbal cues as potentially relevant to an effective social interaction – facial expressions, voice emotion analysis, hand movements and arousal. This section details the software selected to use in this research and the selection process.

3.8.2. IMotions: Facial Expression, Somatic Markers and Hand Movements

IMotions is a relatively novel biometric platform which captures and detects emotions on an individual level. Biometrics is the automated detection of an individual's physical and behavioural characteristics (Jain & Ross, 2015). Facial expressions and physiological processes provide insight into an individual's emotional state (Scherer, 2005). The use of iMotions allow for facial expression, hand gestures and physiological processes to be synchronised over multiple channels, i.e. this software has an integration advantage. A disadvantage of this software is that they did not offer a voice solution at the time of this work, so a separate software had to be selected (see next section).

IMotions version 6.3 was installed on the laptop along with the following two modules: Affdex by Affectiva (facial expression) and the Shimmer3 Unit Kit which contains a PPG-to-Heart Rate Ear Clip and skin conductance are known for physiological feedback to stress.

3.8.2.1. Facial Expression

To detect facial expressions and emotions Affectiva AFFDEX was used and its classifier include a total of 18 expressions. The ROC score for AFFDEX by Affectiva is 0.8 for joy, disgust, contempt and surprise (www.developer.affectiva.com/determining-accuracy/). (See Figure 3.5 for pictorial examples and Appendix 2.1 for AU codes). These expressions are equivalent to the Action Units (AU) described by Facial Action Coding System FACS).

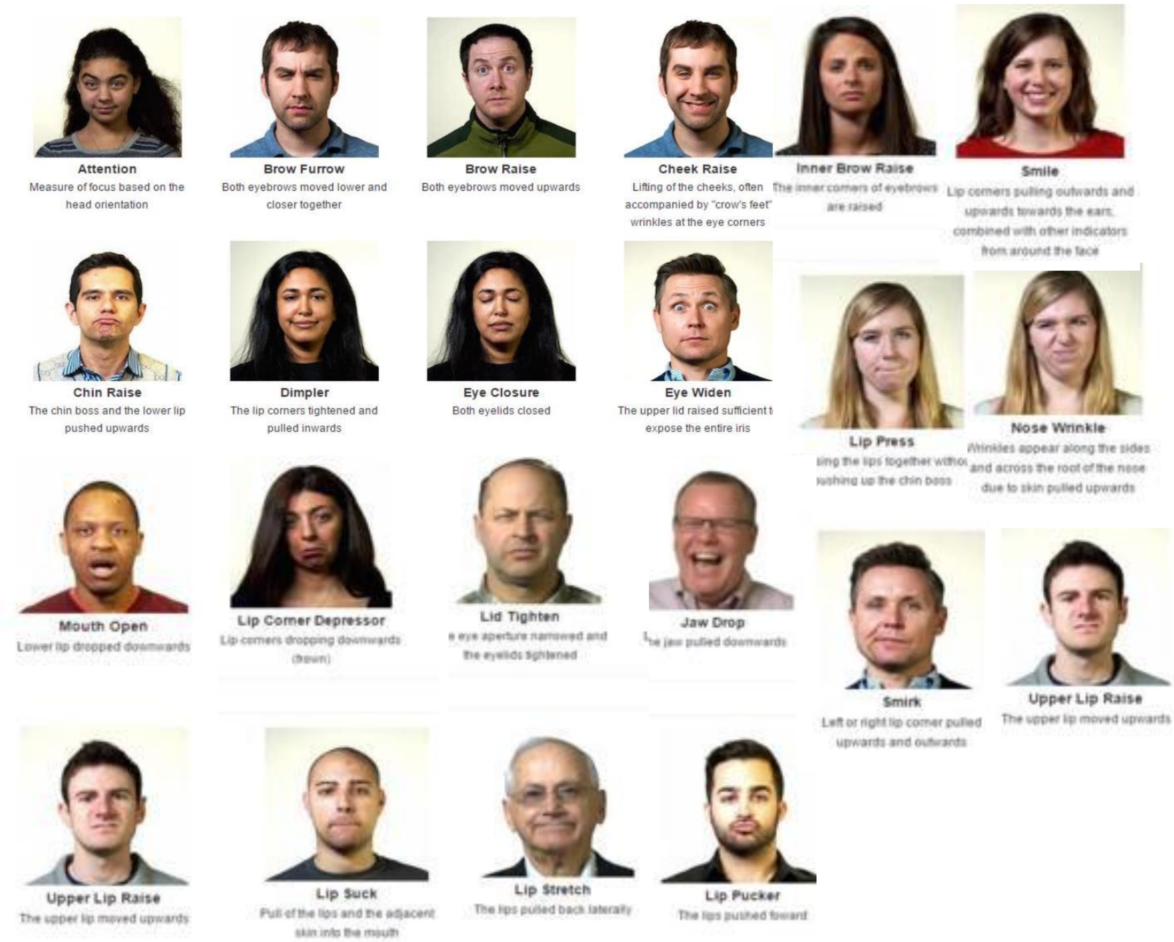


Figure 3.5. AU captured by FACS examples (Image source: [https://imotions.com/blog/facial-action-coding-system/.](https://imotions.com/blog/facial-action-coding-system/) Date accessed: February 2017)

The classifier categorises 7 emotions based on Friesen and Ekmans' Emotional Facial Action Coding System (Friesen & Ekman, 1983). These emotions can be seen in Figure 3.6.

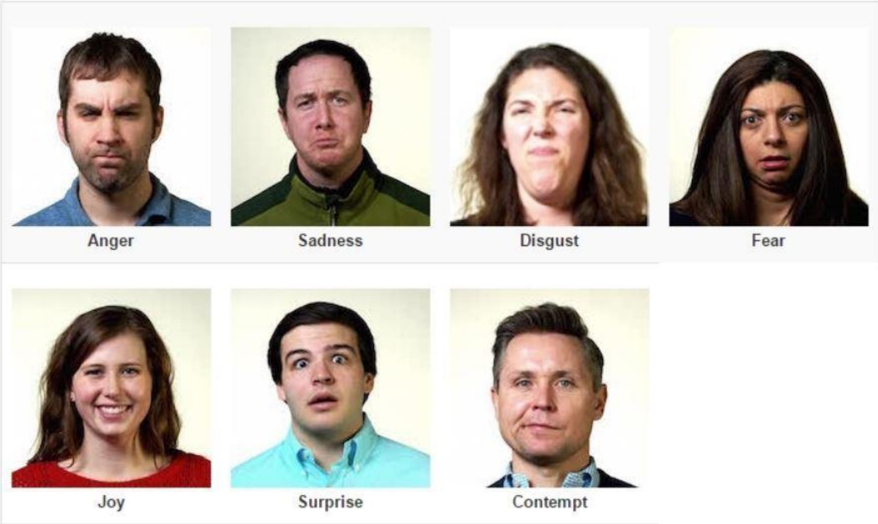


Figure 3.6. EMFACS: Universal Facial Expressions (Image source: https://imotions.com/blog/facial-actioncoding-system/. Date accessed: February 2017*)*

The system also detects engagement which is an overall measure of all emotional expressions. Additional computations include valence, interocular distance and pitch, yaw and roll; but these were not included in the analysis due to missing data or previous versions originally used did not include these measures.

To ensure high quality recordings, interviews were conducted in rooms with natural light and where there this was not possible a light diffuser was used (see Figure 3.4).

As the interaction investigated is a person-person where a camera would not be positioned directly in front of the participant, a pilot study was conducted to investigate the best camera position to capture a person's facial expression using AFFDEX by Affectiva. It was decided to use two cameras to record the session one directly behind the journalist and the other positioned to simulate a media interview. In many instances, multiple cameras can be seen in a media interview and therefore this does not deviate from the aim of the research.

3.8.2.2. Somatic Markers and Hand Gestures

To detect stressful responses to interview questions, the Shimmer 3 was used to record photoplethysmography (PPG) which uses a light that senses the rate of blood flow controlled by pumping of the heart (Anderson, Hsiao, & Metsis, 2017). Galvanic skin response (GSR), otherwise known as skin conductance, was also recorded which measures a change in electrical resistance of the skin caused by emotional stress (Boucsein et al., 2012). GSR is characterised by two activities; slow tonic activity (skin conductance level) and fast phasic activity (skin conductance response). Phasic peaks were investigated in this study as they are sensitive to emotionally arousing stimuli (Benedek & Kaernbach, 2010).

Skin responses were recorded using a non-invasive direct current using two electrodes. This current is applied and reported as a galvanic skin response (GSR) signal. The data collected generally has a low conductance and therefore is expressed in micro(µ)-siemens. The GSR electrodes were proximally placed on the volar phalanges of the fingers on the non-dominant hand. This enables maximal responsivity (Scerbo, Freedman, Raine, Dawson, & Venables, 1992) (see Figure 3.7). The Optical Pulse Ear-Clip sensor was placed on the corresponding ear lobe to detect heart rate for best quality signal.

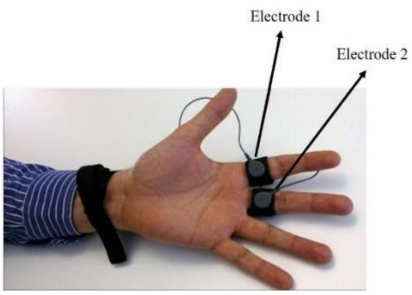


Figure 3.7. GSR electrode positioning on non-dominant hand

To extract phasic data from GSR signals a Mean Filter was applied to remove line noise from the data a low pass filter (default: 5Hz) was added. Peaks were detected by using an onset peak value of >0.01 micro siemens and an offset value of less than zero micro siemens. False positives were identified and removed with a signal jump threshold which was applied at 0.1 micro (µ) siemens.

Participants' data was calibrated to their baseline skin conductance and heart rate recordings. This baseline was a two-minute recording before the start of the media interviews (Benedek & Kaernbach, 2010).

The ear lobe optical pulse circuitry contains an amplifier and a filter to condition the heart rate signal. Participants were asked to relax and then hold their breath which would produce spikes in heart rate. This was done to investigate whether the device was functional. To reduce the amount of static noise introduced to the recording, the Shimmer 3 was at least 30 cm away from the laptop and all mobile and corded phones were removed from the room. The calibrated GSR signal µSiemens was used, the sample rate was calculated, and a median filter was applied to the data. To exclude noise in data a low pass filter was applied to the GSR signal.

The Shimmer 3 Unit+ contains an accelerometer which detects hand movement on a 3 point-axis (3 different directions). To investigate whether this was functional, prior to the start of the interview, participants were asked to move their hand and the dashboard was observed for peaks in data recordings. The Shimmer 3 GSR Unit was used rather than the Empatica E4 wristband, which has similar capabilities; however, Shimmer 3 was more easily accessible.

3.8.3. Voice Affect Recognition

Nemesysco^{Ltd} developed a commercially available voice recognition technology which is used to capture emotions from voice. The version used in this thesis is the voice affect recognition software QA5. This software uses proprietary signal processing algorithms to extract vocal parameters which is classified according to a range of emotions (Nemesysco.com) in real time or recordings can be imported and post-processed.

The reason why this software was used rather than other commercially available software's (e.g. Vokatori [vokaturi.com/]) was because they require the use of Python which is beyond the scope and aim of this thesis. Like the Shimmer 3 device, Nemesysco^{Ltd} was more easily accessible than other software available at the time. A caveat of this software is that it is proprietary software and the details about the algorithm for each label are inaccessible. For this reason, features were correlated with a widely used open source software called Praat (Boersma & van Heuven, 2001). The features produced by Nemesysco^{Ltd} were correlated with prosodic features extracted from Praat which is a voice extraction software that can be used to analyse, synthesize and manipulate speech.

A correlation analysis was conducted to validate the features collected by Nemesysco^{Ltd}. Vocal features extracted from Praat were pitch (mean and maximum), intensity (mean, energy, minimum and maximum) and formants 1, 2, 3, 4 and 5. Pitch is defined as the rate of the opening and closing of vocal folds, it is also known as fundamental frequency. Fundamental frequency and intensity are known to be important variables in communicating emotions in speech. The average pitch value for male speakers are typically found to be 100 Hz to 180 Hz and for females it is found to be 160 Hz to 300 Hz. A high mean pitch has been associated with stress and arousal (Sondhi, Khan, Vijay, & Salhan, 2015). Intensity is associated with the loudness of the voice and is associated with a variety of emotions including psychological stress (van Lierde, van Heule, De Ley, Mertens, & Claeys, 2009).

A formant is defined as a very high amplitude of the acoustic energy. They reflect natural resonance frequencies of the vocal tract and are changed by altering the shape of the vocal tract (Goudbreek, Goldman, & Scherer, 2009). Formants have shown the highest accuracy rate for anger (Mohanta & Mittal, 2016). Table 2 shows that stressed, upset, intensive thinking, imagination, energy, excited, EmoCog Ratio, concentration, extreme emotion is consistent with prosodic features extracted in Praat which are consistent with the literature as described. Table 3.2 details each emotion captured using QA5 and a definition of each label along with the Praat features correlation. This table can be seen on the next page.

Table 3.2. Vocal emotions said to be captured by Q A5 and vocal feature correlates

Emotion	Description	Praat Correlation
Energy	Indicates if speaker is sad, tired, boredom, comfortable or highly energetic.	Intensity minimum *(r = - .348, p = .044)* Mean pitch (r = .742, p = < .001) Format 4 *(r = .355, p = .039)*
Content	Indicates how pleased or happy a person is	--
Upset	Indicates how unpleased or sad a person is	Mean Intensity (r = .602, p = <. 001) Intensity (energy) (r = .520, p = .002)
Angry	Indicates how angry a person is	--
Stressed	Indicates how nervous a person is	Mean Intensity (r = .506, p = .002) Intensity (energy) (r = .502, p = .002) Maximum Intensity (r = .435, p = .010) Mean Pitch (r = .534, p = .001)
Embarrassment	Indicates how uncomfortable a person is	--
Intensive thinking	Indicates thinking intensity while speaking	Minimum Intensity (r = .352, p = .041) Pitch mean (r = - .622, p = < .001) Pitch maximum (r = -.369, p = .032) Formant 1 (r = -.348, p = .043) Formant 3 (r = -.355, p = .039) Formant 4 (r = - .362, p = .036) Formant 5 (r = - .369, p = .0.32)
Imagination Activity	Indicates whether the person is recalling information or visualising something	Minimum Intensity (r = .501, p = .003) Maximum Intensity (r = .379, p = .028) Pitch Mean (r = - .591, p = < .001)
Hesitation	Indicates how comfortable a person is when making the statement	Not measured
Uncertainty	Indicates how certain or uncertain a person is	--
Excitement	Indicates how positively or negatively excited a person is	--
Concentration	Indicates how concentrated the person is	Pitch Mean (r = .519, p = .002)
Arousal	Indicates deep and profound interest in the conversation	Pitch Mean (r = .471, p = .005)
Extreme Emotion	Indicates overall emotional activity	Formant 1 (r = - .342, p = .048) Formant 2 (r = - .473, p = .005) Formant 3 (r = - .485, p = .004) Formant 4 (r = - .479, p = .004) Formant 5 (r = -.545, p = .001)
Brain Power	Overall cognitive activity	
EmoCog Ratio	Indicates rationality	Minimum Intensity (r = - .388, p = .023) Pitch mean (r = .641, p = < .001)

The area under the Receiver Operating Characteristic (ROC AUC) curve score for Nemesysco^{Ltd} was 0.53 – 0.71 (Lacerda, 2009, 2012). While this score is poor- fair, the study investigating Nemesysco$^{Ltd'}$ accuracy did not specify which version was investigated. However, emotions in QA5 have been validated such as ‘embarrassment’ (Han & Nunes, 2010) as well as stressed and arousal (Konopka, Duffecy, & Hur, 2010). In addition to these validations, the QA5 has been used in the development of a conversational robot (Hashimoto et al., 2009; Usui, Kume, Yamano, & Hashimoto, 2008). This suggests the applicability of Nemesysco^{Ltd} in interactions. Nemesysco^{Ltd} user guide states that noise and environment could influence the results obtained. However, throughout the research stages noise and environment was controlled by ensuring interviews were conducted in a quiet area, ensuring sound was attenuated.

Recordings of interviews were edited using Audacity Software version 2.1.3 (audacityteam.org/) and post-processed in Nemesysco^{Ltd}. This method of processing was done for the exploratory stage and the follow-up stage but in the experiment stage recordings were conducted in real time using Nemesysco^{Ltd}. The laptop was placed close to the participant to ensure clear detection of their voice and limit detection of the journalists. All interviews were checked for journalist voice and noise. If this was found, this was removed using the audio editing using Audacity.

3.8.4. ‘Honest Signals’

To capture ‘honest signals’ researchers used sociometric badges. Honest signals are the social signals that are present in all interactions (Pentland & Heibeck, 2010). Sociometric badges were developed by Alex Pentland at his lab at Massachusetts Institute for Technology, the Human Dynamics Group (Pentland & Heibeck, 2010). These are detailed in sections 2.5.1.6. and 2.6.2. The badges were selected as they were designed solely for the purpose of capturing social signals in a multiperson interaction (two or more interlocutors) to understand social organisation in companies (Choudhury & Pentland, 2004) or to understand social organisation in a meeting (Kim et al., 2008). The badges have also been used to capture social signals in a two person interaction (Holding, Sundelin, Lekander, & Axelsson, 2019; Paxton, Rodriguez, & Dale, 2015; Zhang, Olenick, Chang, Kozlowski, & Hung, 2018).

To enable ‘honest signal’ detection, the badges contain four sensors; an infrared sensor receiver (captures face-face interactions), a Bluetooth detector (proximity of another badge detection), microphone (captures conversational features and not content) and a motion detector (captures movement) (Olguín et al., 2007). Data on the badges were stored on a microSD card. Furthermore, each badge is the size of an identity badge and is worn around the neck. It can be adjusted so that it is not intrusive or knocks on tables which can cause noisy data. See Figure 3.8.

Figure 3.8. Sociometric badge worn around the neck

The signals captured by the sociometric badges can be seen in Table 3.3. This table also includes the description of each honest signal which can be mapped into the four higher level descriptions of honest signals; mirroring, influence, activity and consistency.

Table 3.3 Honest signals and their descriptions captured by sociometric badges

HS*	Signals	Description
Activity	Body movement	Normalised acceleration magnitude over 3 movement axes
Activity	Body movement activity	Absolute value of the first derivative of the accelerometer's energy
Activity	Body movement rate	Indicates the direction of change in activity level (compared to first derivative)
Consistency	Body movement consistency	Movement consistency throughout interaction
Mirroring	Body movement mirroring	Mimicking of other badge wearers body movement
Mirroring	Body movement mirror lag	Delay in mimicking of body movement
Activity	Posture front back	Orientation of front back panel
Activity	Posture activity	Absolute angular velocity
Activity	Posture rate	Angular acceleration
Mirroring	Posture mirroring	Mimicking of other badge wearers posture
Mirroring	Posture mirror lag	Delay in mimicking of posture
Influence	Successful Interruptions	Number of successful interruptions made by the badge's wearer
Influence	Unsuccessful Interruptions	Number of unsuccessful interruptions made by the badge wearer
Consistency	Speed of turn-taking	Indicates speed of turn-taking in a conversation
Influence	Overlap	Total amount of speaking whilst someone else is also speaking
Influence	Total speaking	Total amount of combined speaking (speaking and overlap combined)
Influence	Volume Front	Average absolute value of amplitude of the front microphone
Consistency	Volume consistency front	Measurement of change in speech volume
Influence	Front pitch	Pitch of the voice from the front mic correlated with the fundamental frequency of the voice signal
Mirroring	Volume mirroring	Mimicking of other badge wearers volume
Mirroring	Volume mirroring lag	Delay in mimicking of other badge wearers volume

*(*HS = Honest signals) Extracted from (Mozos et al., 2017)*

The ROC score for these badges have been reported as 0.8 which is considered high (Zhang, Palo Alto Laboratory, et al., 2018). During interviews, badges were worn by both the participants and interviewers. Badges were synced by turning the badges on and waiting for a light blinking sequence indicating the badges were booting up. The light sequence that indicates that the badges were in sync were the following: blue and green lights turned on together, green light turned on, green lights turned off, blue lights turned on, blue lights turned off. If another sequence was observed the badges were connected to the laptop to sync the internal clocks and the microSD card was formatted. Once interviews were complete the badges were turned off and both synced on Sociometric Solutions Software which is a user interface specifically designed for sociometric badges (Sociometric Data Lab Enterprise Edition 3.1.2824). Badges were exported as structured meetings (as participants were facing each other and were in the same location throughout the meeting) with predefined default parameter settings resolution of 1 second intervals as .csv files (Sociometric Solutions Manual).

3.9. Data Handling Software

This section details the software for pre-processing of data before and conducting analysis. The representation of numerical data will be 3 decimal points as this is more precise when reporting p-values than the recommended 2 decimal points (Cole, 2015). For example, the rounding of a p-value of 0.0493 to 0.05 it suggests that the result is not significant; but if this value is reported as 0.049, this suggests significance. This method will not be applied when reporting machine learning accuracies where a whole number will be reported due to the nature of this type of analysis. Furthermore, in this thesis, statistically significant result will be interpreted as less than 0.05 (Field, 2013) unless stated otherwise.

3.9.1. Microsoft Excel

Microsoft Excel is a spreadsheet developed by Microsoft capable of performing calculations on data entered, visualisation tools, pivot tables and Visual Basic for Applications. Excel is a useful tool for creating datasets which can be easily imported into a range of statistical packages including the ones used in this thesis (SPSS and Weka, see below). Furthermore, clear graphs and figures can be created in excel fast and easily. This software (version 1902; Build 11328.20438) was used to normalise the dataset to the interval [0, 1] (also known as feature scaling) by using the maximum and minimum values in the dataset using the following formula:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

The above formula has been used in previous literature (Gao, Bianchi-Berthouze, & Meng, 2012). Using this method of data normalisation performs three pre-processing steps to the dataset; 1) it reduces the number of outliers, 2) it ensures the all the features / signals in the dataset are in the same range allowing for statistical tests to be conducted on the test and 3) involves linear transformation of the dataset (Patel & Mehta, 2011). Normalisation of data using this method was done for the exploratory stage, the experiment stage and the follow-up stage.

3.9.2. IBM Statistical Package for the Social Sciences

IBM SPSS Statistics (Statistical Package for the Social Sciences) is a software that is commonly used for statistical analysis on several different data formats (Ghasemi & Zahediasl, 2012). SPSS is a user-friendly software easy to use for large datasets, it can perform complex data manipulation with simple instructions, it can import data from almost any type of file and it can be used for qualitative and quantitative data. It is frequently used in the psychology research domain (Lowenthal & Lewis, 2018). SPSS was used for assumption testing, tests comparing the means of two groups (t-tests and their non-parametric equivalent) and multivariate analysis. An alternative to SPSS statistical package is 'R statistics'; this was not used in this research due to the time constraints of the research project imposed by the funders of the research. Advantages of SPSS is that is easily understood (General User Interface) and widely used. A limitation of using SPSS is that it has restrictions when attempting to conduct more advanced statistical tests (i.e. Cohen's kappa / Scotts pi or Fleiss's kappa).

To analyse qualitative data, interviews were transcribed and coded in Microsoft Word. An alternative to transcribing interviews is NVivo software; however, this software was not used as this was not available to the researcher at the time of analysis.

3.9.3. Waikato Environment for Knowledge Analysis 3– Machine Learning and Feature Selection

Waikato Environment for Knowledge Analysis (Weka) 3 is a machine learning software written in Java which contains a collection of machine learning algorithms which contains tools for classification, regression, clustering, visualisation and data preparation (Witten, Frank, & Hall, 2011). This software package was developed by the University of Waikato, New Zealand and is free to use. It is also easy to use for beginners in machine learning due to its graphical user interface. For this research the algorithms used in this thesis were conducted with their default parameters, unless specified (see Chapter 4 – k nearest neighbour). This programme was also used to select features / attributes for inclusion in analysis. Feature extraction and machine learning methods were conducted in Weka's main user interface, the Explorer. Files were originally created in Excel and imported as a .csv file and later saved as an .arff file.

An alternative to Weka is R statistical package which could have been used for machine learning algorithms used in this research. However, again, due to the time constraints imposed by the funders of this research. The use of R statistics would have been beneficial in this research as all statistical tests and machine learning algorithms could have been conducted on the same platform for consistency. However, the nature of the tests conducted on Weka and SPSS are different. Machine learning methods are designed to make accurate predictions whereas statistical models are designed for making inferences about the relationships between variables (Witten & Frank, 2005). It is noted because of this difference the outcomes of data analysis are not affected by the package used.

*3.9.4. G*Power 3.1 – A priori Power Analysis*

G*Power version 3.1 was used to calculate estimated sample size for the experiment stage. G*Power is a free software that is freely available (g-power.com) (Erdfelder et al., 1996). This is widely used by psychologists and is a simple method for conducting complex power and sample size calculations (Cunningham & McCrum-Gardner, 2007).

3.10. Ethical Approval and Procedure

This project was jointly funded by BAE Systems (British Multinational Defence, Security and aerospace company) and the Engineering and Physical Sciences Research Council (EPSRC). The project was conducted in collaboration with QinetiQ. The research stages conducted as part of the project were stages 1 – 3. Once the project was completed, the author decided to add an additional stage which is presented in Chapter 6. Ethical approval was given by the Ministry of Defence Research Ethics Committee and Brunel University Research Committee for Chapter 4-6. Ethical approval was sought for the follow-up stage from Brunel University Research Committee only as this study was an additional study following the collaboration with QinetiQ. Table 3.4 shows the ethical approval codes by each research committee.

Table 3.4. Research Ethics Committee Codes

	Research Stage	**Brunel University London**	**Ministry of Defence Research Ethics Committee**
1.	Exploratory Stage	3795-SS-Nov/2016-4310-1 (Appendix 3.4)	772/MoDREC/2019 (Appendix 3.5)
2.	Design Stage		
3.	Experiment Stage		
4.	6-month Follow-up Stage	11294-LR-Apr/2018-12578-1 (Appendix 3.6)	■

The sample recruited were research staff and students from Brunel University via email (Appendix 3.7). The sample does not include those under the age of 18 years old and any vulnerable participants. Participants were not asked to disclose information such as mental health issues but were asked to disclose whether they have a social impairment. Researcher ensured that there was no power relationship (i.e. students who feel obligated to take part as in a study for their teacher), because of how the participants were recruited (email) they had expressed an interest in taking part in the study and the researcher ensured that up until the end of the training day (closing statement) participants were able to withdraw their participation in the study.

Participant Information Sheets detailed the purpose of the study, requirements, duration and the risks associated with participation in this research (Appendix 3.8 – 3.11). Potential participants who responded to the participation invitation via email or posters distributed around the university were provided with a PIS to read at least 24 hours before being invited to sign a consent form. In conjunction with consent form details (Appendix 3.12) on the participants' right to withdraw, this was verbally reiterated by the researcher.

Following the distribution and explanation of the consent form, participants were asked to complete a demographics form where the following information was asked: about their experience in public speaking, media interviews, position (staff or student), gender (with the option to specify or prefer not to say), age ranges (with the option to select 'prefer to not to say'), ethnicity (with the option to select 'prefer to not to say'), place of birth, nationality (with the option to select 'prefer to not to say'), first language (with the option to select 'prefer to not to say') and disclose whether they have a social impairment (with the option to select 'prefer to not to say') (Appendix 3.13). This research complied with the Declaration of Helsinki as adopted at the 64th WMA General Assembly at Fortaleza, Brazil (World Medical Association, 2013). Participants were given the opportunity to ask questions. At the end of each study participants were provided with a closing statement (Appendix 3.14 – 3.17).

To ensure anonymity of participants, each participant was assigned a number in place of their name. This number was given to participants in the instance they want to withdraw their participation at a later stage so that the researcher can identify their data. Once identified, participants would be removed.

Data was stored according to the UKs Data Protection Act. The data management plan can be seen in Appendix 3.11. All hard copies of consent forms were sent to and will be stored at the Ministry of Defence Secretariat in accordance with extant UK Legislation and Ministry of Defence Policy (JSP 536) and electronic copies were stored at Brunel University London.

3.11. Summary

This section reviews the literature on each methodology used in this research and justifies the tools and approach taken. A summary of the methods selected can be seen in Table 3.5.

Table 3.5. Methodology of research stages

Stage	Method	Tools
Exploratory	Exploratory	CSRS Media and affect recognition technology
Intervention Design	User-Centered (mixed-methods)	SUS and Interview
Intervention Evaluation	Experimental (mixed methods)	CSRS Media and affect recognition technology Interview
6-Month Follow-up	Experimental	CSRS Media and affect recognition technology Interview

CHAPTER 4. EXPLORING SOCIAL SIGNALS ASSOCIATED WITH EFFECTIVE COMMUNICATION USING EMOTION RECOGNITION SYSTEMS

4.1 Introduction

This chapter details the first stage of this research. It investigates the social signals which are important for evaluating communication skills during a media interview. Detection of signals using COTS technologies described in Chapter 3 was used to predict good and bad performances based on judgements made by human raters using scales which are also described in Chapter 3.

The aim of this research stage is threefold: 1) to assess whether COTS technology can detect social signals in a person-person context; 2) to investigate the combinations of social signals important for evaluating trainee's communication skills performance and 3) to identify the estimated sample size to test the benefits of a social signals training intervention in this context.

The research questions for this exploratory stage is the following:

- *Can COTS technology detect and identify the relevant social signals for effective communication in the context of media interviews?*
- *What is the required sample size for this type of training intervention?*

This chapter details the methods and materials used to capture social signals, how data was collected, the approach used to analyse data, preliminary results, a more detailed analysis and a summary detailing the implications of the results for subsequent stages of research in this thesis.

4.2 Data Collection

4.2.1. Participants

Participants were recruited via email (Appendix 3.7) which resulted in 17 participants included in this study. University researchers were recruited as they are representative of professionals who are likely to take part in media interviews. The resulting sample included 11 males and 6 females whose age ranged from 18-56 years old). The language capability of participants was recorded and revealed that all participants spoke English fluently; however, 8 participants were native English speakers and 9 non-native English speakers. The nationalities of participants included British (5), Lithuanian (1), Iraqi (1), South African (1), Ghanaian (1), Nigerian (1), Malaysian (1), Korean (1), Greek (2), Dutch (2) and Italian (1). The roles that participants had within the university included research staff (5), professional staff (1), taught student (1) and research student (10).

Participants were excluded if they disclosed that they had a social disability; however, no participants had disclosed this information. Participants experience in public speaking ranged from ‘none’ to ‘extensive’ and media interview experience ranged from ‘none’ to ‘some’.

4.2.2. Trainer Experience

This workshop was run by two media training professionals with over 30 years of professional experience in journalism / training. The expertise of these journalists / trainers was valuable to data collection as it is likely to simulate the signals displayed in a real-world media interview than a journalist in training. Expertise is important for this research stage as the feedback and training are valuable as this stage is the building block for identifying relevant social signals for this type of context.

4.2.3. Procedure

Data was collected across total of 3 one day media skills training workshops which took place on Brunel Campus in Uxbridge across 3 months (April 2017 – June 2017). Each training day consisted of 5 – 6 participants. Workshops were conducted by media training professionals with over 20 years’ experience of journalism and 10 years of media skills training. Each workshop took place in a standard university seminar room with a projector, chairs and a table.

Before attending workshops, participants were instructed to provide a short and easy-to-understand summary of their research and describe their worst anticipated question in a media interview and the importance of their research. This was done to assist trainers in preparing for conducting tailored practice interviews for each of the participants. Some questions included: “what is the importance of your research?” and “What is the worst possible question they could expect during a media interview?”. Like a real-world setting, this helped trainers in preparing tailored training to suit individual research topics as journalists would be familiar with interviewees area and have already developed probing questions (Taylor, 2015). A schematic view of the procedure can be seen in Figure 4.1.

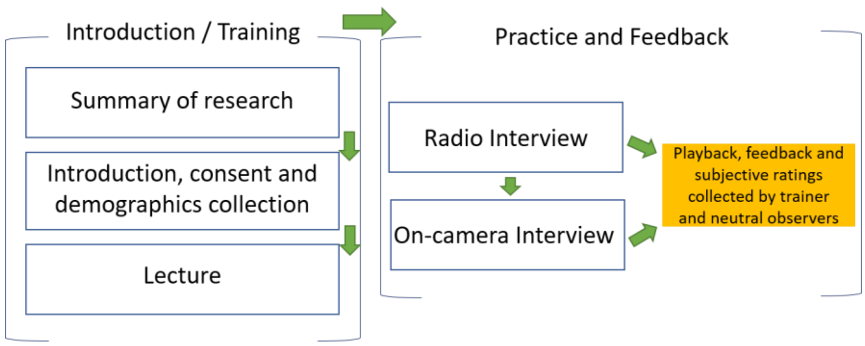


Figure 4.1. Schematic of the exploratory stage procedure

Commencement of the workshop included a full briefing of the study and formal consent was then collected (Appendix 3.12). Participants were verbally told that they have the right to withdraw their participation in this research at any point. Participants were then asked to provide demographic information (ethnicity, age, gender, job role, presence of social disability / communication disability and experience of presentation) (Appendix 3.13). Participants were also given the opportunity to take part in the workshop if they did not wish their social signals to be recorded without penalty. However, all participants gave consent to record signals.

After the introduction to the study, participants took part in a 1-hour lecture-style introduction to effective communication skills in media interviews in a group setting. This was proprietary training as normally delivered by trainers at Brunel University London. Slides are not included here as they are commercial confidential. Participants were then given individual time slots throughout the day to complete their practice interviews with the trainers. These interviews were conducted individually and included two practice interviews. The first interview was to simulate a radio interview where participant and trainer sat face-face with a table and chair with a voice recorder on the table. No cameras were turned on during this interview to avoid any influence this may have on performance. The second practice interview was a simulation of an on-camera interview with one camera placed behind the journalist to face the participant and a second camera placed beside the participant to ensure a more accurate post hoc face recognition. Participants were told that the camera layout was to mimic a broadcast. Interviews were ordered this way to incrementally introduce trainees to the interview process as research has shown that people are generally more nervous in on-camera interviews which was also stated by the media training experts (Taylor, 2015). In addition to incremental exposure, the difficulty of interview questions had also increased. Finally, to avoid social pressure of a peer group audience, interviews had taken place individually. The room layout can be seen in Figure 4.2. The professional media trainers acted as the interviewer for the practice media interviews.

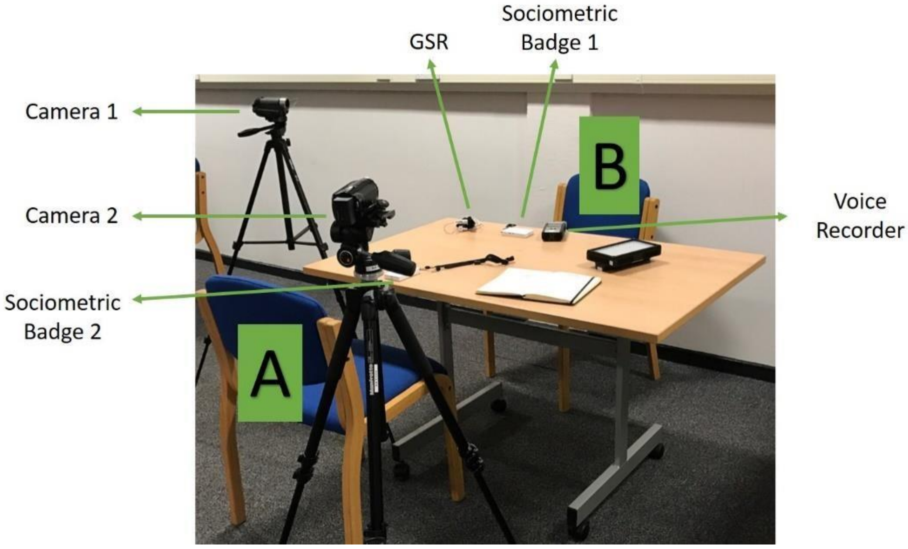


Figure 4.2. Study set-up for both practice interviews. A) Journalist and B) participant. The picture shows two cameras facing the participant for better detection of facial expressions and the voice recorder was included for better quality recordings

Prior to the commencement of each interview, the Shimmer 3 GSR device was attached to the participants and the sociometric badges were put on the journalist and the participant. The room included further recording equipment which can be seen in Figure 4.2. A two-minute baseline recording was conducted to obtain each participant's baseline heart rate and skin conductance (Benedek & Kaernbach, 2010). This was done to ensure an accurate measure of physiological alterations throughout the interview in response to different interview questions. Practice interviews included questions relevant to participants' research and difficulty of interview questions increased as interviews progressed. Each interview lasted between 5 to 8 minutes.

Interview recordings were played back to participants after each interview and they were then provided feedback about their performance by trainers. Trainers and participants were then asked to fill in the Communication Skills Ratings Scale (CSRS) after playback of interview to allow for self-reflection. Instructions for answering the questionnaire was explained to the journalist and trainee as instructed by (Spitzberg & Adams, 2007).

At the end of the workshop, participants were given a short closing statement reminding them of the purpose of the study and were reimbursed £5 per hour for their time. Again, participants were reminded that they have the right to withdraw their participation.

4.2.3.1. Post Hoc Evaluation of Performance by Neutral Observer

To remove potential bias by trainer-trainee interaction which could influence trainer ratings, practice interviews were rated by three neutral observers who were also able to pause and play back the videos as an audience would be able to in the real-world.

4.2.4. Measures and Materials

This section provides a summary of the materials used in this section. For details of the measures and COTS technologies refer to Chapter 3. For this stage of research, recordings were captured using media technology and later post-processed using emotion-recognition COTS technologies. Table 4.1 shows the channels of communication captured during practice interviews.

Table 4.1. The method in which COTS technology was used to capture the four different channels of communication during practice media interviews.

Social Signals	COTS Technology	Media Technology	Method
Voice Emotion Recognition	Voice Affect Recognition	Zoom H4N Pro Handy Audio Recorder Audacity software version 2.1.1 (trimming out journalists' voice)	Post-processed
Honest Signals	Sociometric Badges	■	Post-processed and exported as structured meetings with a resolution of 1 second intervals.
Facial Expression	iMotions (Affdex + Shimmer 3 GSR Unit+)	Sony PJ 220 Handycam camera Adobe Photoshop (video trims)	Post-processed and then imported into iMotions and postprocessed using the Affdex module to produce facial expression data.
Arousal and Hand Movement		■	Real-time processed to record heart rate and skin conductance. Accelerometer was used to infer hand gestures.

Participants' performance evaluation was evaluated by the trainer and three neutral observers using the molar rating section of the CSRS (Appendix 3.14).

4.2.5. Data Treatment

This section presents how the data was pre-processed and treated during data analysis. A preliminary analysis (see section 4.4.2.) was initially conducted on the data because of a small sample size. A more detailed and in-depth analysis was conducted (see section 4.4.3.) to validate the features selected for feedback in the preliminary data analysis.

4.2.5.1 Reliability of Communication Skills Rating Scale in Media Interview Context

The internal consistency of trainer ratings and self-report ratings of communication skills was conducted to investigate the reliability of the CSRS in the context of media interviews as this has not been investigated before. This was done using Cronbach's alpha (Tavakol & Dennick, 2011). In addition to this, interrater reliability was calculated using intraclass correlation which is often used to investigate the agreeableness between raters (Mandrekar, 2011)

4.2.5.2. Ground Truth Labelling

The subjective ratings of communication skills were calculated attributing 40% to the trainer and 20% to each of the neutral observers (Naim et al., 2016). This weighted average was done to ensure a well-rounded measure of performance from an audience and an expert. However, the 60% (20% for each observer) of the weighted average was attributed to the neutral observers as they produce a less subjective result as trainers could have been influenced by interacting with the trainee. In addition to this, the audience plays a bigger role in forming judgements of an interviewee as they are not as knowledgeable on the interviewees topic as the journalist and would be able to rate their performance in an unbiased manner (Taylor, 2015).

To separate the dataset to establish a ground truth which includes good and bad performances, the mid-point was located on a histogram and was used as a cut-off point. A high value suggests effective communication whereas a low value indicates poor communication.

4.2.5.3. Social Signal Data Pre-Processing

The exchange of social signals according to interview type is important to consider as they are different forms of communication (Taylor, 2015). Therefore, this was accounted for by normalising each dataset independently and merging the resulting dataset together. This coincides with the aims of the research to investigate the social signals which are predictive of effective communication and differences are accounted for by normalising the datasets separately. A significance test was conducted to investigate if there were any differences in the displays of social signals between the radio and on-camera interview datasets.

4.2.5.4. Thin Slices of Behaviour

Analysis was run on the first 30 seconds (Buchanan, 2009; Poggi & D ’errico, 2011) because it has been found that is at this stage that viewers form an initial impression of someone which is important for how the interviewee is perceived by the audience (Taylor, 2015).

4.3. Data Analysis

4.3.1. Preliminary Data Analysis

4.3.1.1. Social Signal Data Pre-processing

All communication channels in the preliminary data analysis were normalised using the *normalize filter* on WEKA (Witten, Frank, & Hall, 2011). This filter normalises the attributes / features in the dataset where the resulting values range from -1 to +1.

4.3.1.2. Feature Selection

To choose features which are effective in judging good or bad interview performance, *k*-Nearest Neighbour (*k*-NN) classifier is applied to each feature for producing the predictions. Features were selected by conducting (*k*-NN) on each feature where features with the highest classification accuracy were selected for inclusion of prediction accuracy analysis for the social signal channel. This method for feature selection accounts for the variance between effective and poor communicators (Abernethy, 2010).

4.3.1.3. Prediction Analysis

To predict performance based on the features selected, *k*-NN classifier was run where *k* =1 with a leave-one-out (L-O-O) cross validation. *K-NN* is a non-parametric learning algorithm used for both classification and regression. It calculates the distance between the test data and the training data and produces an accuracy according to this calculation. The Euclidean Distance Formula accounts for the number of dimensions and, as a result can be used with a number of features.

Equation 1. Euclidean Distance Formula

$$\mathrm{d}(\mathbf{p},\mathbf{q}) = \mathrm{d}(\mathbf{q},\mathbf{p}) = \sqrt{(q_1-p_1)^2 + (q_2-p_2)^2 + \cdots + (q_n-p_n)^2}$$
$$= \sqrt{\sum_{i=1}^{n}(q_i-p_i)^2}.$$

This algorithm uses the data which are separated into the different classes to predict the classification of a new sample point in the feature space and has been found to be effective in the real world as it does not obey any theoretical assumptions such as linear regression methods (Witten, Frank, & Hall, 2011).

A leave-one-out (L-O-O) cross validation method was used as it uses all the data to train and validate the predictive model for all possible combinations and is particularly useful for small sample sizes as small as 16 and applies to non-parametric methods such as *k*-NN (Fu, Carroll, & Wang, 2005). Figure 4.3 demonstrates the preliminary data analysis approach.

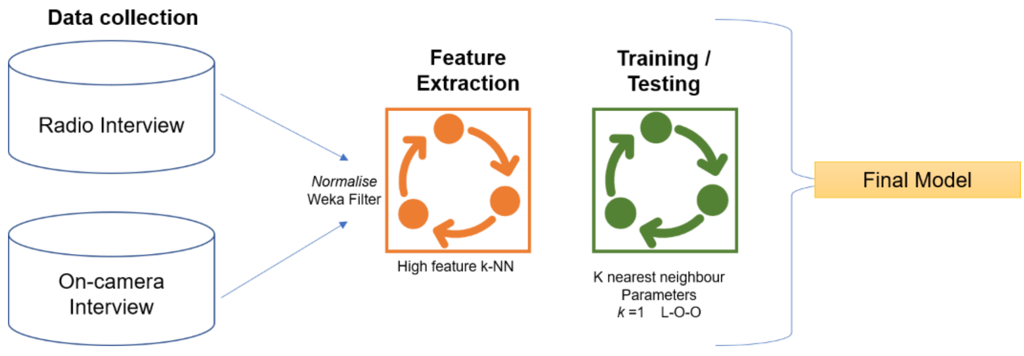


Figure 4.3. Preliminary data analysis indicating data collection, pre-processing, feature selection / extraction and k-NN parameters. F CV = 17-Fold Cross Validation; L-O-O = Leave-One-Out

4.3.2. Detailed Data Analysis 4.3.2.1.

Missing Data

Missing data because of low-quality recording or failure in recording were removed from the dataset. This was done as the *k*-NN is an instance-based classifier where the learning is based on the dataset (Witten & Frank, 2005). Instances in this case are referred to as all of participants recorded signals.

There were 3 participants where facial expressions were not captured due to low quality video recordings, 2 participants where sociometric badges did not record the interaction and 1 participant where the Shimmer 3 did not record which was used to record hand gestures, heart rate and skin conductance.

4.3.2.2. Data Pre-processing

The more detailed analysis used all four channels of communication (facial expression, honest signals, voice analysis and hand movements) which were normalised to the interval [0, 1] using the minimum of each feature (Gao et al., 2012) as described in Chapter 3. This is also known as Min-Max Feature Scaling.

Equation 2. Min-Max Feature Scaling Formula

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

4.3.2.3. Feature Selection

The radio and on-camera interview datasets were merged which created a larger the sample size (*n* = 34) which allowed for a correlation-based feature selection (CSF) method for feature selection to be conducted. The CSF method selects the features which are highly correlated with the labelled data (good vs bad communicators) and uncorrelated with each other (Hall, 1999). Features which produced a correlation value below 0.2 were excluded (Witton, Frank & Hall, 2011). The CSF was applied to each channel of communication dataset separately, i.e. facial expression only, sociometric badges only, voice affect only and hand movement only. This was used as a means of pre-processing the data for features to include for a classification analysis on the dataset.

4.3.2.4. Classification Analysis

Like the preliminary data analysis, *k*-NN was used for the same reasons. However, in the case where *k* = 1, two cross validation methods were used; 1) the data was partitioned into 17 training and 17 testing instances using a 17-fold cross validation and 2) L-O-O cross validation. Cross validation is a method for estimating the performance of a predictive model that partitions the dataset into a training set to train the model and a test to evaluate it (Kohavi, 1995). Folds of 17 were used in this instance to separate the data into 17 radio interviews and 17 video interviews which accounts for possible differences between the two types of interviews. The L-O-O method was used was it takes all the data into consideration and is the preferred method for small sample sizes, as is in this research stage (Witton, Frank and Hall, 2011). The dataset was further explored using k = 2, 3 and 4 where only the L-O-O method of cross-validation was used.

Bootstrap aggregating, also known as bagging, is an ensemble method which uses the minority class data, in this case it is poor communicators, without creating synthetic data or by changing the existing classification to create new training sets. This is in comparison to the “classifier balancer” function in Weka that creates synthetic data which is not characteristic of the data collected. This method is typically used on less stable algorithms to reduce variance and avoids overfitting of the model (i.e. decision trees) and not more stable algorithms (k-NN) and has often found to mildly degrade the performance (Breiman, 1996). However, prediction analysis has been conducted with and without bagging to compare accuracies. Furthermore, bagging has been proven effective in prediction analysis of imbalanced data groups (good vs bad performance; ground truth) and small sample sizes (Li, 2007). A step-by-step explanation of the prediction analysis can be seen in Figure 4.4.

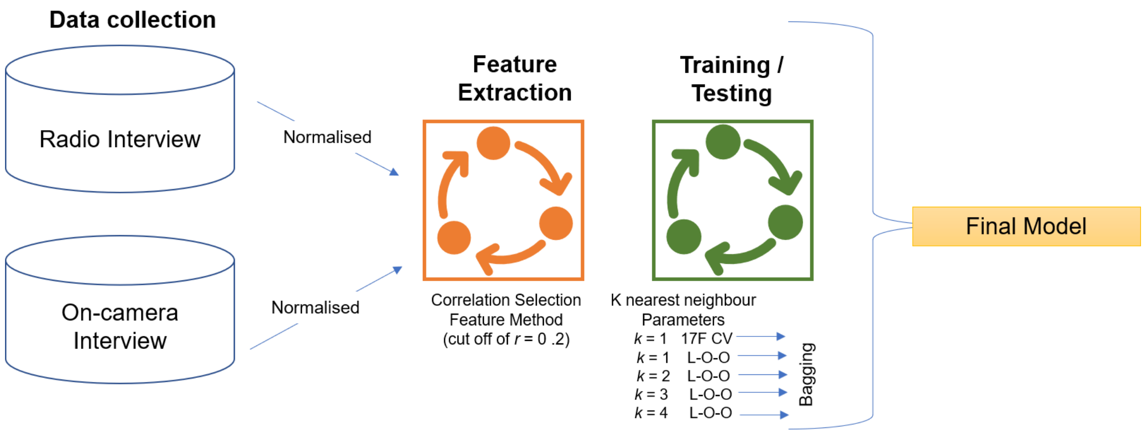


Figure 4.4. Detailed prediction analysis demonstrating the steps taken which include; independent interview dataset normalisation, feature extraction / selection and k-NN methods. F CV = 17-Fold Cross Validation; L-OO = Leave-One-Out

4.3.2.5. Autonomic Arousal

Autonomic arousal was measured to provide insight into internal physiological arousal levels (Schachter & Singer, 1962) which could predict communication skills. A baseline of two-minute was recorded to obtain participants baseline arousal (Boucsein et al., 2012). Both heart rate and phasic skin conductance were recorded. Subsequent peaks were then normalised to this baseline which indicate personal and accurate arousal. A univariate analysis was conducted to investigate if there were any significant differences in autonomic arousal between performance groups by interview type.

4.3.2.6. Behavioural Differences between Interviews

To assess whether there were any significant differences in displays of social signals between radio interview and the on-camera interview, a post hoc analysis was conducted using a Mann-Whitney U test as the data was not normally distributed (McKnight & Najab, 2010). Features included in the prediction analysis were the dependent variable and interview type was the independent variable.

4.3.2.7. Differences in Performance in Individual Features

A post hoc test was conducted to assess whether social signals differed between ratings of effective and poor communicators. A Man-Whitney U test was conducted where features included in the prediction analysis were the dependent variable and communication classification was the independent variable.

4.3.2.8. Sample Estimation for Experiment

To estimate a sample size for a training intervention an estimated sample size was conducted based on the first and second interview performance which is based on the subjective performance measures of communication skills. However, a rough indication of effect size is effective. A sample estimation calculation using G*Power will inform the sample size for the experiment study (Chapter 6) (Faul, Erdfelder, Lang, & Buchner, 2007).

4.4. Results

4.4.1. Subjective Ratings of Communication Skills

4.4.1.1. Radio Interview

Based on the mid-point located on the histogram, the cut-off for effective communicators for radio interviews was 23.70 where if a participant obtained a value higher than this they would be classified as an effective communicator (see Appendix 4.1). Internal consistency for trainer ratings for communication skills for the radio interview yielded a Cronbach's alpha of α = .960. Internal consistency was also calculated for neutral observers which yielded a Cronbach's alpha of α = .980. Finally, the internal consistency for self-report measures for communication skills during the radio interview produced a Cronbach's alpha of α = .961, suggesting use of the CSRS is valid for use in radio interviews.

The trainer and neutral observers had a high inter-rater agreement which yielded .694 with a 95% confidence interval from .370 to .875, ($F_{(16, 48)} = 3.324$, $p = .001$) as measured by an average measure of intraclass correlation (Mandrekar, 2011).

4.4.1.2. On-camera Interview

The mid-point on a histogram was located which recognised the cut-off between effective and poor communicators for on-camera interviews were 25.20. If a participant was rated higher than this then they would be classified as an effective communicator (see appendix 4.1). Internal consistency for trainer ratings for communication skills for the on-camera interview yielded a Cronbach's alpha of α = .973. Internal consistency was also calculated for neutral observers which yielded a Cronbach's alpha of α = .975. Finally, internal consistency for self-report measures of communication skills produced a high Cronbach's alpha of α = .974, suggesting use of the CSRS is valid for use in on-camera interviews.

The trainer and neutral observers had a moderate interrater agreement which yielded .640 with a 95% confidence interval from .275 to .852, ($F_{(16, 48)} = 2.939$, $p = .002$) as measured by an average measure of intraclass correlation.

4.4.2. Preliminary Data Analysis

A preliminary analysis was conducted on the data obtained for radio and on-camera interviews independently.

4.4.2.1. Vocal Affect Recognition

4.4.2.1.1. Radio Interview

The highest social signals identified were 'energy' (53%), 'upset' (65%) and 'hesitation' (47%) and were selected for inclusion in the prediction analysis. Inclusion of these signals suggests that they contributed the most to the overall model. Thereby suggesting that they are relevant for communication in radio interviews. These selected features produced an accuracy of 82% with a confusion matrix correctly classifying 9 out of 10 instances as effective communicators and 5 out of 7 instances as poor communicators.

4.4.2.1.2. On-camera Interview

The highest social signals that were identified were 'content' (65%), 'upset' (47%), 'hesitation' (59%) and 'extreme emotion' (65%). These four signals suggest relevance in expression of these emotions in communication during on-camera interviews. These selected signals produced a prediction accuracy of 82% with a confusion matrix correctly classifying 9 out of 11 instances as effective communicator and 5 out of 6 instances as poor communicators.

4.4.2.2. Honest Signals

4.4.2.2.1. Radio Interview

The honest signals selected were ‘movement rate’ (56%), ‘movement mirror’ (75%), ‘posture’ (63%), ‘posture mirroring’ (56%), ‘speed of turn-taking’ (50%), ‘volume’ (63%) and ‘volume mirror’ (63%).

These signals produced a 69% prediction accuracy. A confusion matrix showed the model correctly classified 8 out of 10 instances as a good performance; further correctly classifying 3 out of 6 instances as bad performance.

4.4.2.2.2. On-camera Interview

The social signals with the highest accuracy for training of good and bad performances were identified. These were ‘movement rate’ (63%), ‘movement mirror’ (56%), ‘posture’ (94%), ‘posture mirroring’ (50%) and ‘speed of turn-taking’ (75%). Further analysis revealed that the inclusion of volume (38%) and volume mirror (25%) did not affect the overall training accuracy and were, therefore, included. Inclusion of these signals were valuable as, according to the literature, were found to be important for media interviews (Taylor, 2015). The overall accuracy of these signals was 69%. A confusion matrix correctly classifying 8 out of 11 instances as a good performance; further correctly classifying 3 out of 6 instances as bad performance.

4.4.2.2.3. Facial Expression (on-camera interview only)

The social signals with the highest accuracy were identified as anger (71%), joy (64%), contempt (79%) and brow furrow (71%). These selected features produced an accuracy of 71%. Further analysis revealed that all facial expression features resulted in 71% accuracy and produced a confusion matrix correctly classifying 9 out of 10 instances as effective communicators and 1 out of 4 instances as poor communicators.

4.4.3. Detailed Analysis

In this analysis the data from the radio interview and on-camera interview were joined together and a similar analysis was conducted on the dataset. The results are shown in this section.

4.4.3.1. Social Signals Displayed During Interviews

The weighted average of human judgments of performance were calculated for both interviews a total of 13 interview cases were classified as poor communicators ($M = 20.739$; $SD = 2.818$) and 21 as effective communicators ($M = 28.581$; $SD = 2.652$). A Levene’s test (Gastwirth, Gel, & Miao, 2009) revealed equal variances between groups ($p = > .05$). An independent samples *t*-test indicated that there was a significant difference between the two groups in displays of social signals ($t_{(32)} = 8.183$, $p = < .001$, $d = 2.866$). Descriptive scores of each participant’s interviews and their classification of each interview performance see Appendix 4.1, suggesting difference in groups valid for exploration.

4.4.3.2. Vocal Affect Recognition

All interview cases were included in this analysis (*n* = 34). Features included in the analysis were stressed, extreme emotion, hesitation, concentrated and arousal. Table 4.2 details the analysis conducted.

Table 4.2. Accuracy of k-NN and bagging with number of iterations for vocal affect recognition data

k	Cross validation	Accuracy	Effective; Poor communicators	Bagging Accuracy (iterations)	Effective; Poor communicators
1	17	68%	15 of 21; 8 of 13	68% (100)	15 of 21; 8 of 13
1	Leave-one-out	68%	15 of 21; 8 of 13	68% (100)	15 of 21; 8 of 13
2	Leave-one-out	65%	19 of 21; 3 of 13	65% (100)	16 of 21; 6 of 13
3	Leave-one-out	59%	14 of 21; 6 of 13	59% (100)	14 of 21; 6 of 13
4	Leave-one-out	68%	18 of 21; 5 of 13	56% (100)	13 of 21; 6 of 13

A 17-fold cross validation was used for this data set where *k* = 1 where features selected produced an accuracy of 68%. L-O-O cross validation was used where *k* = 2 accuracy produced was 65%, *k* = 3 produced 59% accuracy and *k* = 4 produced an accuracy of 68%. Bagging results produced accuracies ranging from 56% to 68%. The highest accuracies were produced above chance (50%) for *k* = 1 and 4 and bagging accuracies for *k* = 1 and 4. Descriptive statistics can be seen in Figure 4.5.

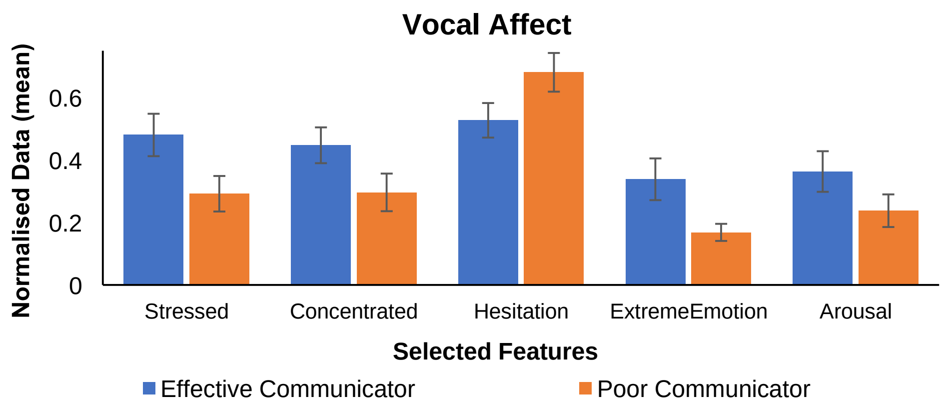


Figure 4.5. Means and standard errors for effective and poor communicators of selected features for vocal affect

Figure 4.5 suggest that poor communicators were more hesitant whereas effective communicators were more concentrated, stressed, aroused and emotional during both interviews. A Mann-Whitney U test revealed that these differences were not significantly different when investigated in isolation. Additionally, there was no significant differences between signals displayed in the radio interview and the on-camera interview. See Table 4.3.

Table 4.3. Mann Whitney U test results on selected features for vocal affect and interview difference

Feature	Mean Rank (Good, bad)	U Test Results	Interview difference
Stressed	19.98, 13.50	U = 84.50, p = .065	U = 139, p = .850
Concentrated	19.71, 13.92	U = 90, p = .099	U = 106, p = .185
Hesitation	15.24, 21.15	U = 89, p = .092	U = 98, p = .109
Extreme Emotion	19.38, 14.46	U = 97, p = .161	U = 130, p = .617
Arousal	18.93, 15.19	U = 106.50, p = .292	U = 131, p = .642

4.4.3.3. Honest Signals

A total of 32 interview cases were included in this analysis because of missing data. Features selected for inclusion in this analysis included ‘movement activity’, ‘movement rate’, ‘movement consistency’, ‘posture’, ‘posture activity’ and ‘unsuccessful interruptions’. Table 4.4 presents the analysis conducted.

Table 4.4. Accuracy of k-NN and bagging with number of iterations for honest signals

k	Cross validation	Accuracy	Effective; Poor communicators	Bagging Accuracy (iterations)	Effective; Poor communicators
1	17	78%	16 of 20; 9 of 12	78% (100)	16 of 20; 9 of 12
1	Leave-one-out	78%	16 of 20; 9 of 12	78% (100)	16 of 20; 9 of 12
2	Leave-one-out	81%	19 of 20; 7 of 12	78% (100)	16 of 20; 9 of 12
3	Leave-one-out	75%	15 of 20; 9 of 12	78% (100)	16 of 20; 9 of 12
4	Leave-one-out	75%	17 of 20; 7 of 12	81% (100)	17 of 20; 9 of 12

A 17-fold cross validation was used for this data set where k = 1 where features selected produced an accuracy of 78%. L-O-O cross validation was used where k = 2 accuracy produced was 81%, k = 3 produced 75% accuracy and k = 4 produced an accuracy of 75%. Bagging results produced accuracies ranging from 78% to 81%. The highest accuracies produced were above chance for k = 1 and 2 and bagging accuracies for k = 1, 2, 3 and 4. Descriptive statistics can be seen in Figure 4.6.

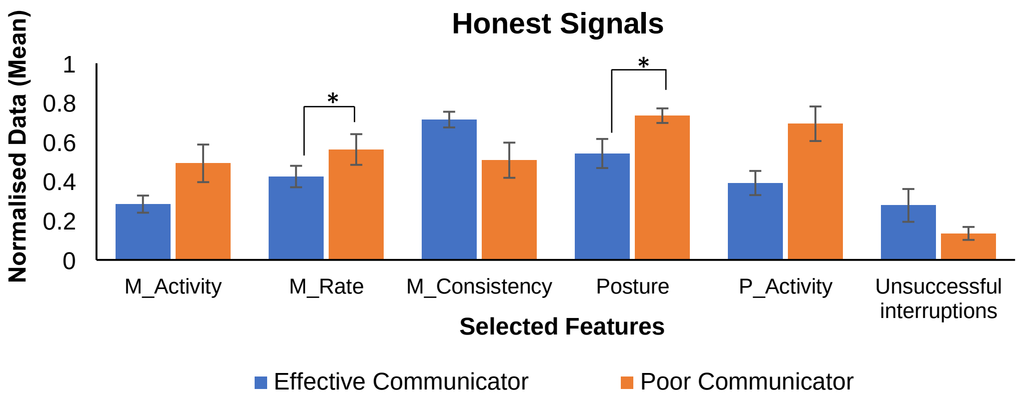


Figure 4.6. Means and standard errors for effective and poor communicators of selected features for signals

Figure 4.6 suggests that good communicators moved less and were consistent with their movements. These communicators also had a relaxed posture and less postural activity as well as more unsuccessful interruptions than those who were classified as poor communicators. A significance test comparing good communicators and poor communicators revealed that poor performers moved significantly faster than good performers. Additionally, those who altered their posture in the interview were more effective communicators than poor communicators who did not adjust their postures as often. An analysis investigating the differences between the radio and on-camera interview revealed a significant difference in between the movement consistency displayed in the radio interview and the on-camera interview. See Table 4.5.

Table 4.5. Mann Whitney U test results on selected features for Honest Signals. The dependent variables are the features selected and the independent variables are evaluation labels (good vs bad).

Feature	U Test Results	Interview difference
M activity	U = 88, p = .213	U = 121, p = .792
M rate	U = 65, p = .034	U = 94, p = .200
M consistency	U = 84, p = .161	U = 75, p = .046
Posture	U = 95, p = .330	U = 98, p = .258
P activity	U = 64.50, p = .031	U = 110, p = .497
Unsuccessful Interruptions	U = 101.50, p = .462	U = 113, p = .577

4.4.3.4. Facial Expression

A total of 14 interview cases were included in this analysis. A more detailed analysis was conducted as the data were normalised differently to the preliminary analysis. Features included in this analysis were ‘smile’, ‘smirk’, ‘sadness’, ‘joy’, ‘fear’, ‘contempt’ and ‘brow furrow’. Table 4.6 presents the analysis conducted.

Table 4.6. Accuracy of k-NN and bagging with number of iterations for facial expressions

k	**Cross validation**	**Accuracy**	**Effective; Poor communicators**	**Bagging Accuracy (iterations)**	**Effective; Poor communicators**
1	7	79%	10 of 10; 1 of 4	79% (100)	10 of 10; 1 of 4
1	Leave-one-out	79%	10 of 10; 1 of 4	79% (100)	10 of 10; 1 of 4
2	Leave-one-out	71%	10 of 10; 0 of 4	71% (100)	10 of 10; 0 of 4
3	Leave-one-out	71%	10 of 10; 0 of 4	71% (100)	10 of 10; 0 of 4
4	Leave-one-out	71%	10 of 10; 0 of 4	71% (100)	10 of 10; 0 of 4

A leave-one-out cross validation was used for this data set where k = 1 where features selected produced an accuracy of 79%, k = 2 accuracy produced was 71%, k = 3 produced 71% accuracy and k = 4 produced an accuracy of 71%. Bagging results produced accuracies ranging from 71% to 79%. The highest accuracies produced were above chance (50%) for k = 1 and bagging accuracies for k = 1. Descriptive statistics can be seen in Figure 4.7.

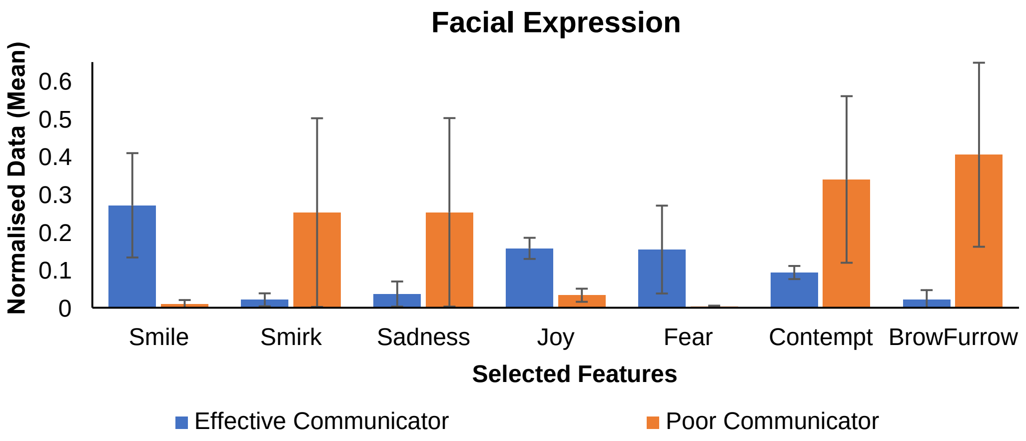


Figure 4.7. Means and standard errors for effective and poor communicators of selected features for facial expression

Descriptive statistics presented in Figure 4.7 suggest that effective communicators smiled more whereas those who were classified as poor communicators showed more contempt, frowned more, smirked more and showed more sadness. A Mann-Whitney U test revealed that these differences were not significantly different when investigated in individually. See Table 4.7.

Table 4.7. Mann Whitney U test results on selected features for Honest Signals. The dependent variables are the features selected and the independent variables are evaluation labels (good vs bad)

Feature	**U Test Results**
Smile	$U = 8.5, p = .102$
Smirk	$U = 15, p = .478$
Sadness	$U = 15, p = .480$
Joy	$U = 13, p = .322$
Fear	$U = 19.50, p = .944$
Contempt	$U = 7.5, p = .077$
Brow Furrow	$U = 9, p = .118$

4.4.3.5. Hand Movement

A total of 33 interview cases were included in the merged analysis. The missing data for 1 interview case was from the on-camera interview. Hand movements were not included in the preliminary analysis as the data only included a single signal where an analysis was not warranted for inclusion. Given that the use of gestures was solely visible to neutral observers in the on-camera interview, an analysis was conducted for radio and on-camera interviews together and then individually. Analysis for the radio interviews can be seen in Table 4.8, for the on-camera interview can be seen in Table 4.9 and the merged dataset can be seen in Table 4.10.

Table 4.8 *k*-NN analysis for Radio Interview

k	**Cross validation**	**Accuracy**	**Effective; Poor communicators**	**Bagging Accuracy (iterations)**	**Effective; Poor communicators**
1	9	35%	4 of 10; 2 of 7	35% (100)	4 of 10; 2 of 7
1	Leave-one-out	29%	4 of 10; 1 of 7	29% (100)	6 of 10; 1 of 7
2	Leave-one-out	58%	10 of 10; 0 of 7	18% (100)	3 of 10; 0 of 7
3	Leave-one-out	5%	0 of 10; 1 of 7	24% (100)	4 of 10; 0 of 7
4	Leave-one-out	52%	9 of 10; 0 of 7	24% (100)	4 of 10; 0 of 7

Table 4.9 *k*-NN analysis for On-camera Interview

k	**Cross validation**	**Accuracy**	**Effective; Poor communicators**	**Bagging Accuracy (iterations)**	**Effective; Poor communicators**
1	8	75%	8 of 10; 4 of 6	75% (100)	8 of 10; 4 of 6
1	Leave-one-out	75%	8 of 10; 4 of 6	75% (100)	8 of 10; 4 of 6
2	Leave-one-out	69%	9 of 10; 2 of 6	69% (100)	8 of 10; 3 of 6
3	Leave-one-out	63%	7 of 10; 3 of 6	69% (100)	8 of 10; 3 of 6
4	Leave-one-out	56%	9 of 10; 0 of 6	69% (100)	8 of 10; 3 of 6

Table 4.10. *k*-NN analysis for radio and on-camera interview (merged data)

k	**Cross validation**	**Accuracy**	**Effective; Poor communicators**	**Bagging Accuracy (iterations)**	**Effective; Poor communicators**
1	17	55%	13 of 20; 5 of 13	55% (100)	13 of 20; 5 of 13
1	Leave-one-out	55%	13 of 20; 5 of 13	55% (100)	13 of 20; 5 of 13
2	Leave-one-out	51%	17 of 20; 0 of 13	51% (100)	13 of 20; 4 of 13
3	Leave-one-out	36%	12 of 20; 0 of 13	48% (100)	14 of 20; 2 of 13
4	Leave-one-out	48%	16 of 20; 0 of 13	57% (100)	15 of 20; 4 of 13

The gesture data met the normality and homogeneity of variance assumption for running an analysis of variance (ANOVA) (p = > .05) (Tomarken & Serlin, 1986) and, as a result a 2 x 2 ANOVA was run on the data. Independent variables included performance (effective vs poor communicators) and interview type (radio vs on-camera interview).

Results from the ANOVA revealed that no main effect significant main effect between effective and poor performances ($F_{(1, 29)} = 4.567$, $p = .041$, $\eta^2 = .136$), no main effect between radio interview and on-camera interview ($F_{(1, 29)} = .461$, $p = .503$, $\eta^2 = .016$), but no interaction effect between performance and interview ($F_{(1, 29)} = 3.533$, $p = .070$, $\eta^2 = .109$). A post hoc analysis revealed no difference between good (M = 10.520; SD = 4.601) and bad (M = 10.149, SD = 4.267) performers in the radio interview ($t_{(15)} = .169$, $p = .868$, $d = 0.084$). However, there was a significant difference between those that performed well (M = 12.253; SD = 3.054) and poorly (M = 6.457; SD = 4.252) in the on-camera interview ($t_{(14)} = 3.181$, $p = .007$, $d = 1.560$). Descriptive statistics can be seen in Figure 4.8.

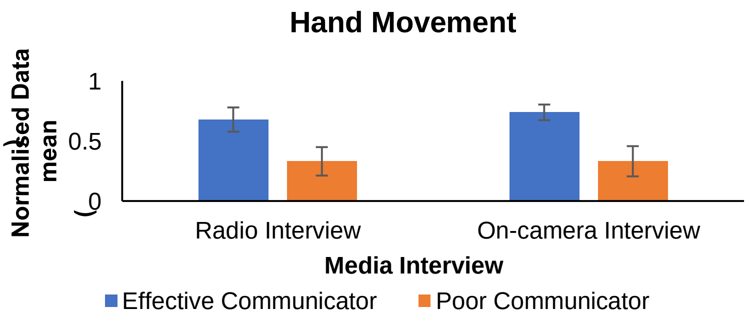


Figure 4.8 Means and standard errors for effective and poor communicators hand movements

4.4.3.6. Autonomic Arousal

A total of 33 participants were included in this analysis. The heart rate data met the normality and homogeneity of variance assumptions (p = > .05) and as a result a 2 x 2 ANOVA was conducted on the data. Results revealed no main effect for performance ($F_{(1, 30)} = 2.077$, $p = .354$, $\eta^2 = .065$), interview ($F_{(1, 30)} = .887$, $p = .354$, $\eta^2 = .029$) and no interaction effect between performance and interview type ($F_{(1, 30)} = .010$, $p = .921$, $\eta^2 = .000$). As a result, no post hoc tests were conducted.

The skin conductance data did not meet the normality and the homogeneity of variance assumptions were not met (p = < .05), a Kruskall Wallis test was conducted. There was no significant difference between performance ($X^2_{(1)} = .467$, $p = .495$, $r = 0.03$) with a *mean rank* of 8.30 for good communicators and 10.00 for bad communicators in the radio interview. Furthermore, there was no significant difference between good communicators (*mean rank* = 8.09) and bad communicators (*mean rank* = 10.67) in the on-camera ($X^2_{(1)} = 1.010$, $p = .315$, $r = 0.06$). Figure 4.9 shows descriptive statistics.

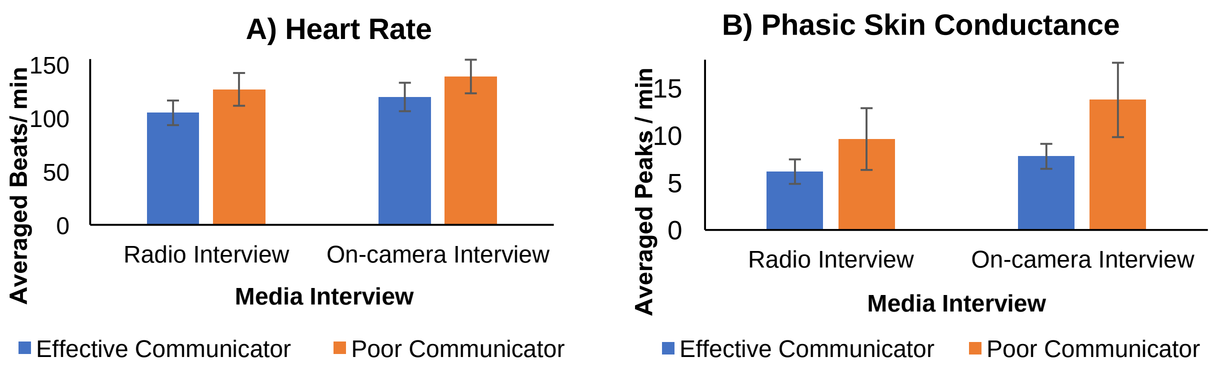


Figure 4.9 Means and standard errors for effective and poor communicators' heart rate (A) and phasic skin conductance (B)

4.5. Sample Estimation for Experiment Stage

To understand the sample size required for this type of training intervention, a sample estimation calculation was conducted. To do this, the ratings of communication skills ratings were summed using the weighted average previously described. The data was not normally distributed and therefore a non-parametric test was used to assess trainee improvement from the first session (radio interview) to the second session (on-camera interview). A Wilcoxon Signed Rank test revealed a significant difference between the first (voice) interview (*Mdn* = 24, *M* = 24.53, *SD* = 4.73) and the second (video) interview (*Mdn* = 27, *M* = 26.76, SD = 4.49), *Z* = -3.314, *p* = .001, *r* = 0.80. The effect size used in this instance, as the data was not normally distributed, was Rosenthal's effect size (Rosenthal, 1994):

$$r = \frac{z}{\sqrt{N}}$$

A sample estimation for the experiment stage was calculated using the large effect size (.80). An a priori test was carried out with G*Power to determine the required sample size which is required for this design. A large effect size of r = .80, a power of 0.95 and an alpha of .01 produced a sample size of 22.

4.6. Discussion

The aim of this stage was to identify the social signals necessary for evaluating communication performance in media interviews. Using the data collected a preliminary data analysis was conducted revealing the signals that will be used to provide feedback in the experimental stage (Chapter 6). A more detailed analysis was conducted using a larger sample size which was generated by using all interview cases as instances for analysis. The results from the analysis produced a range of prediction accuracies to identify the most relevant signals for based on machine learning analysis and a more formal analysis using mean comparison tests to identify clear differences between groups. Each communication will be discussed in turn as well as their meaning in relation to previous studies that attempt to predict human classification into good and bad communication based on signals.

4.6.1. Vocal Affect Data

Vocal affect COTS technology was able to detect signals in the context of media interviews as signals included in the analysis using a correlation method.

4.6.1.1. Preliminary Analysis

A total of 3 signals were able to predict communication skills in the radio interview with a prediction accuracy of 82%. According to software labels, interview cases rated as effective communicators showed more 'upset', less 'energetic' and were less 'hesitant'. This may suggest that effective communicators were more confident in media interviews (Skinner, Gordin, & Ed, 2013).

A total of 4 signals predicted performance with an accuracy of 82% in the on-camera interview. Those who were rated as good performance were more 'content', 'upset' and 'emotional' but were less 'hesitant'. These findings could suggest that good performers in on-camera interviews were more emotional. While both interview types produced a high accuracy, these results should be interpreted with caution as the sample size is small. Additionally, identification of 'upset' in this context could have been a result of a false positive (Pecchia, Chen, Nugent, & Bravo, 2014) or incorrect labelling of the data captured.

4.6.1.2. Detailed Analysis

When data was combined across interview types to produce more cases for analysis, a total of five voice signals were able to predict communication skill performance in media interviews with a range of 56 to 68% accuracy. Trainees who were categorised as effective communicators displayed more signals labelled as 'vocal arousal', 'extreme emotion', 'concentration' and 'stressed' while displaying less instances of vocal 'hesitation'. It is good to note that this model was less predictive with a larger sample size. However, this could be down to inaccurate labels of performance or noisy data.

Passion is defined by Nemesysco^{Ltd} from its vocal arousal readings in combination with the extreme emotion selected as a feature in this study could suggest that trainees rated as effective communicators were more passionate when discussing their research in the first 30 seconds of the interview. Previous research has found that passion is related to work performance and increases cognitive attention (Ho, Wong, & Lee, 2011). Similarly, passion for a topic has found to lead to better performance in public speaking (Morgan, 2008). The findings from this study supports the idea that passion and concentration are central for good communication in a media interview.

4.6.1.3. Vocal Affect – Common Signals

It is important to note that the signals selected for inclusion in both small and larger sample sizes were 'extreme emotion' and 'hesitation'. This suggests that these signals could be relevant for evaluating skills during media skills training.

4.6.2. Honest Signals

The COTS technology designed for capturing honest signals, sociometric badges, can detect honest signals in the context of media interviews.

4.6.2.1. Preliminary Analysis

A total of 7 signals predicted trainees' performance with a 75% accuracy in the radio interview. Those who were labelled as effective communicators displayed more posture mirroring, movement mirroring and volume mirroring. This is consistent with the literature on effective communication (Bilakhia, Petridis, & Pantic, 2013; Pentland & Heibeck, 2010; Spitzberg & Adams, 2007; Terven, Raducanu, Meza-de-Luna, & Salas, 2016).

The on-camera interview revealed that 7 signals were able to predict trainee's communication performance with a 69% accuracy. Good media interview performers exhibited more movement mirroring, posture mirroring, faster speed of turn-taking, a higher volume and more vocal mirroring. These are consistent with the convention of good media interviews (Taylor, 2015). A faster speed of turn-taking during conversation has been linked to engagement (Choudhury & Pentland, 2004) or irritation (Pentland & Heibeck, 2010).

4.6.2.2. Detailed Analysis

When data was combined across interview types to produce more cases for analysis, six honest signals channels were capable of categorising trainee's communication skills in media interview with an accuracy of 75 to 81%. Trainees who were categorised as effective communicators showed less movement activity, fewer postural changes, less movement rate, displayed more consistent movements, had a relaxed posture and had more unsuccessful interruptions. Clear differences were found between effective and poor communicators when interviewees posture changes as well as their rate of movement.

This study suggests that slower movements were considered important for effective communication during media interviews which is consistent with research and training guidance on media interviews (Taylor, 2015). Similarly, it was found that minimal changes in posture was rated as important for good performance in a media interview, which has also been found to be consistent with suggestions about what is considered as important for a good media interview (Taylor, 2015). Slow and minimal movement are suggested for on-camera interview as too much movement is often perceived by the audience as fidgeting, boredom or nervousness (Gross & Levenson, 1997). Gross and Levenson (1997) have also suggested that consistent movement is important for a good media interview, which was also found in this study. Signals detected are like previous discussions of important nonverbal features of a media interview and the convention for on-camera media interviews.

Occasions where trainees spoke before the journalist had completed an utterance, but then stopped, was rated as skilful communication. This indicates lack of congruity of conversation by the journalist and compliance by the trainee (Li, 2001; Vinciarelli et al., 2009). This has been suggested as important in understanding dyadic intercourse, particularly in intercultural communication (Li et al., 2005).

4.6.2.3. Honest Signals – Common Signals

Both movement rate and posture were found to be relevant in the preliminary analysis and the larger dataset analysis suggesting that these signals are appropriate for evaluating signals in the context of media interviews.

4.6.3. Facial Expressions

The preliminary analysis produced a 71% accuracy for all signal's captured. AFFDEX by Affectiva COTS technology can detect facial expressions in a dyadic media interview. Seven expressions included in the analysis could classify effective and poor communication ranging from 71% to 79% prediction accuracy. Effective communicators had exhibited more smiling, joy and fear. They also smirked less, showed more sadness, less contempt and frowned less.

Joy and smiling were found to be important in a similar study that used automated recognition technology to a reciprocal exchange of social signals to explore the best signals for good job interviews (Naim et al., 2016). A study found that smiling has been associated with high affiliation and dominance resulting in the person seeming more approachable (Knutson, 1996).

Additionally, fear emerged as one of the signals associated with highly rated communicators. This result seems counterintuitive; however, a previous validation study of Affdex found that the technology produces low accuracy for detection of fear and anger (Stöckli, Schulte-Mecklenbeck, Borer, & Samson, 2017) suggesting that the inclusion of fear may be ambiguous.

4.6.4 Hand Movements

Shimmer 3 was able to capture hand movements in media interviews. From this detection, researchers were able to infer use of hand gestures. Results revealed a prediction accuracy was initially low ranging from 48% to 57%; however, further analysis found that this accuracy improved in the on-camera interviews (ranging from 56% to 75%) compared to radio interviews (from 5% to 58%). This finding suggests that the increased hand gestures were associated with judgements of effective communicators in the on-camera interview only and not the radio interview. There was a difference between those that performed poorly to those that performed well in the on-camera interview where those that performed well used more hand movements.

The literature on hand gesture use has shown that use of this communication channel assists in the ability to communicate (Morgan, 2008). The results found in this study suggest that wearable technology could be utilised to support presenters as it is a low-cost intervention for detecting hand gestures. Interestingly, a study by Damian et al (2015) developed a system for providing real-time feedback during public speaking based on gesture detection. This design was motivated by a practical consideration of the capability of a noisy environment and did not consider a prediction of a good performance; however, the findings in this research provide some empirical support for their chosen approach.

4.6.5. Autonomic Arousal

Simultaneously, the Shimmer 3 was also able to detect autonomic arousal during media interviews. The motivation for inclusion of autonomic arousal was to assess whether there were any clear differences between physiological arousal associated with ratings of performance. The results showed that there was no significant difference in skin conductance and heart rate between skilled and unskilled communication. This suggests that the variance in performance is not solely driven by differences in arousal and that a combination of signals is more sensitive to measures of performance than arousal. Additionally, this difference may be a result of the experience in public and media interviews.

4.6.6. Bootstrapping Aggregation

The results from the bootstrap aggregation revealed very similar, if not the same accuracies produced by k-NN alone. This could be a result that bootstrap is more frequently used in conjunction with less stable algorithms such as decision trees and neural networks as opposed to *k*-NN which is a more 'stable' algorithm – meaning that its output typically changes less than that of decision trees and neural networks when the input data is perturbed (Gul et al., 2018). Bagging is usually employed with higher-variance learners, i.e. models that are less 'stable'. However, conducting bootstrap aggregation with unequal sample sizes attributed to equal groups demonstrates that the accuracies are consistent as this method has often been used in prediction analysis in this instance (Li, 2007).

4.6.7. Social Signal Inclusion: Preliminary vs detailed analysis

The signals identified in the detailed analysis included more than the preliminary analysis. This could have been because of an increase in sample size suggesting that a larger sample size increases the power of the analysis (Prajapati, Dunne, & Armstrong, 2010). Based on the results, the classification accuracy is generally lower. This could be a result of noise in the data but could also be a result of better precision in prediction accuracies (Gul et al., 2018)

4.7. Conclusion

The first aim of this stage was to investigate which social signals were important for evaluating trainee's communication performance in the context of media interviews:

Can COTS technology detect and identify the relevant social signals for effective communication in the context of media interviews?

The relevant social signals necessary for training intervention were identified in the preliminary data analysis. These will be fed back to participants in the experiment stage (Chapter 6). The core reason that the preliminary data was conducted was due to the tight deadline of the project. The funders required the intervention design to be completed when only the preliminary analysis had been done.

Table 4.11 shows the social signals identified in the preliminary analysis.

Table 4.11. Selected Social Signals for Feedback for improvement based on preliminary data analysis

Channel	Social Signals
Video Interview	
Facial Expression	Smile, smirk, anger, sadness, disgust, joy, surprise, fear, contempt and brow furrow.
Voice Emotion Recognition	Content, upset, hesitation and extreme emotion
Honest Signals	Movement Rate, movement mirroring, posture, posture mirroring, speed of turn, volume and volume mirroring.
Voice Interview	
Voice Emotion Recognition	Energy, upset and hesitation
Honest Signals	Movement, movement activity, movement rate, movement consistency, movement mirroring, movement mirroring lag, posture, posture activity, posture rate and posture mirroring

However, to explore whether the signals selected for feedback were acceptable, a more detailed analysis was conducted with a larger sample size. A further analysis was conducted included an even larger sample size pooled from all the data collected across all stages. Results can be seen in Appendix 4.2.

The second aim of this research was to investigate what the required sample size was for this type of training intervention:

What is the required sample size for this type of training intervention?

A power analysis revealed that a sample size of 22 is required and will be the target sample size for the experiment stage (Chapter 6). Next step in this research is presented in the next chapter which is to identify the easiest and understandable method of providing social signal feedback.

CHAPTER 5. USER CENTRED DESIGN OF SOCIAL SIGNALS FEEDBACK FOR COMMUNICATION SKILLS TRAINING

5.1. Introduction

The previous chapter showed that it is possible to detect social signals in a person-person interaction using COTS technologies. While these technologies all came with their own interfaces, except for sociometric badges, it is not clear whether these designs are optimised for providing user feedback. The literature suggests some guidelines but given the new context of use in media interviews, a user centred design is required. To do this, participants from the exploratory study were invited back to take part in this study to be given feedback based on earlier interviews using the default COTS interfaces as well as some customised designs.

The research question for this user-centred design stage is the following:

What is the best method of presenting social signal feedback in the context of media interview training that is actionable and understandable to trainees?

This chapter describes the second stage of research in an attempt to answer the overall research question of this thesis; can communication skills be improved using automated technology in a media interview context? This chapter investigates the most appropriate method of providing social signal feedback to trainees in a way that is actionable and understandable. It also presents the methods and discusses the results obtained from semi-structured interviews and system usability ratings. At the end of this chapter, how feedback will be employed in the following experiment stage will be discussed.

5.2. Data Collection

5.2.1. Participants

Participants from the exploratory stage who had expressed interest in returning to take part in further research were contacted via email. A total of five participants were recruited which included four females and one male, their age ranged from 18 – 35 years old. Nationalities of participants included Greek (1), Malaysian (1), Netherlands (1), South African (1) and Korean (1). A total of 4 participants were non-native English speakers and 1 participant was a native speaker. The roles that participants had within the university included research student (3), taught student (1) and research staff (1).

5.2.2. Feedback Display Methods

Several feedback display formats were explored which included those provided by the software providers and bespoke custom designs created for this project. The signals which were presented to participants were based on the preliminary data analysis conducted in the exploratory stage in

Chapter 4. Feedback styles and visualisations were presented to participants using summative (in summary form) and formative feedback (playback of video for reflection) techniques as described in Chapter 2 and is detailed in Hoque and colleagues (2013) and Fung et al (2015). A post-summary method for providing feedback has proven effective in previous research (Tanaka et al., 2015).

5.2.2.1. Facial Expression

There was a total of three options of visual displays of facial expressions which were presented to participants.

5.2.2.1.1. Method 1 – iMotions Emotion Dashboard with Video Playback

The first option was provided by iMotions which presents video playback with emotions displayed on an emotion dashboard below the video (juxtaposing facial expressions with the video) (Hoque, et al., 2013). This can be seen in Figure 5.1. Facial expression feedback included the seven basic emotions proposed by Ekman and Friesen (1972).

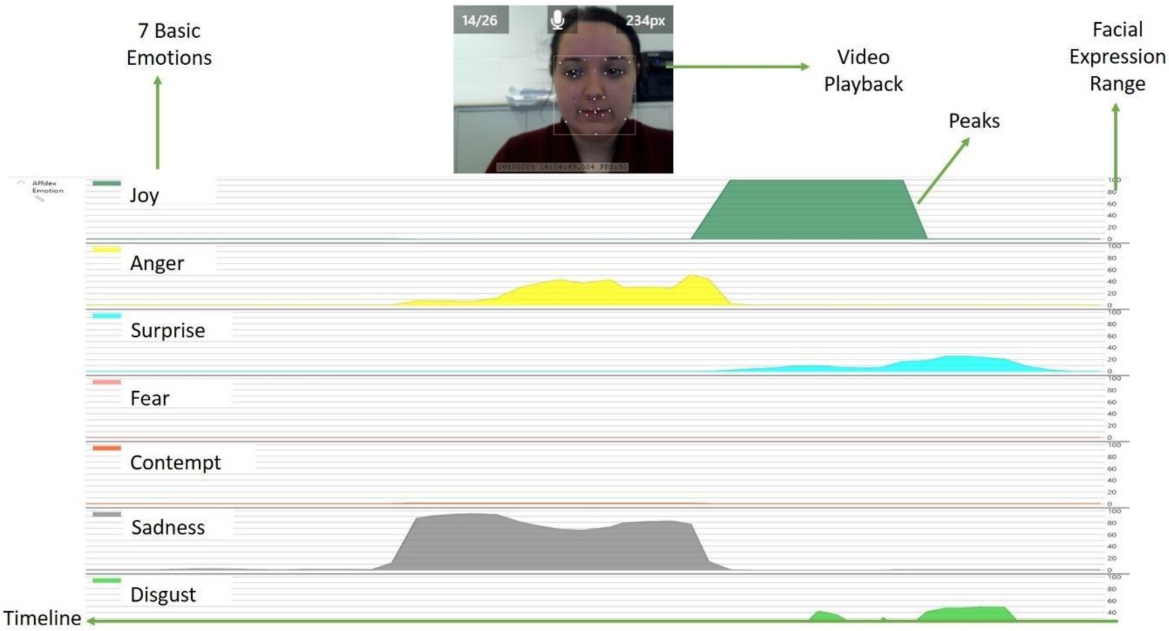


Figure 5.1. Method 1: iMotions video playback and emotion dashboard

5.2.2.1.2. Method 2 – Bespoke Emotional Dashboard

The second option was a software that was developed and implemented using MATLAB www.mathworks.com, (release R2015a8. 5.0. 197613, 64 bit) to process matrix-based data structures (Colecchia, Giacomin, & Hone, 2018). This can be seen in Figure 5.2 which included only 6 basic emotions.

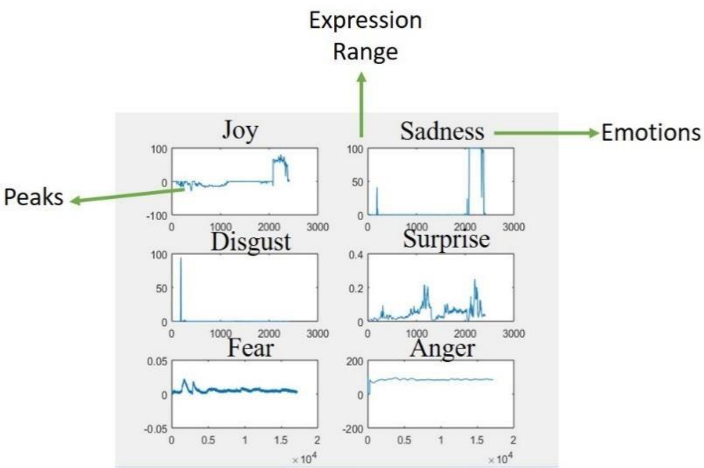


Figure 5.2. Method 2: Custom developed emotional dashboard

5.2.2.1.3. Method 3 – Bar Chart Template

The final method of displaying feedback for facial expression was presented in a bar chart style template. This can be seen in Figure 5.3 which included the 7 basic emotions and additional action units.

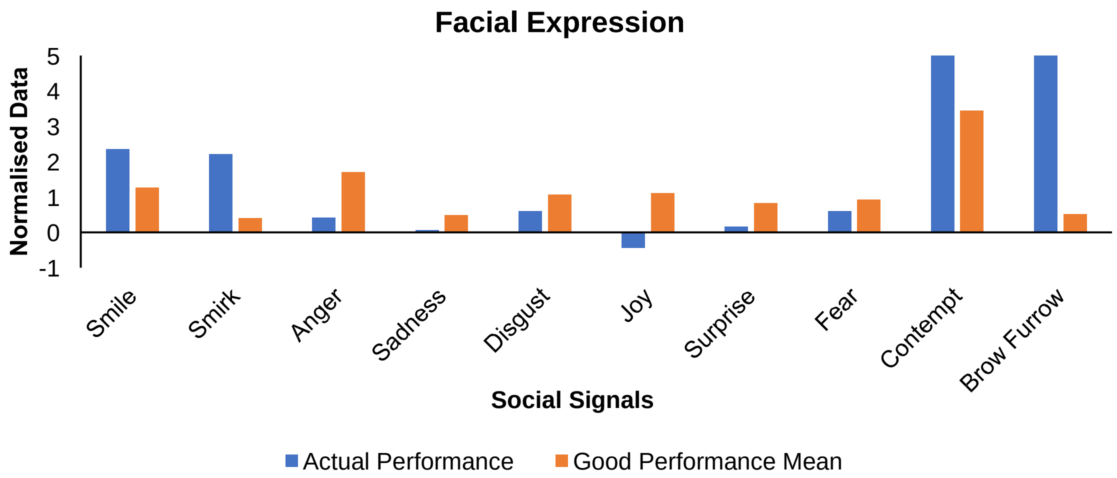


Figure 5.3. Method 3: Bar chart style template

The bar chart template adds information about what is considered a good performance, values for this was obtained by 'effective communicator' ratings by neutral observers and trainers which were collected in the exploratory stage. This information was added to allow trainees to infer behaviour changes that need to be made to improve. The data presented for each feature displayed in the bar chart was normalised using the mean and standard deviation of that feature to ensure that each feature is presented on the same scale. This would allow participants to observe their behaviour in relation to what is behaviour that is considered a good in media interviews.

5.2.2.2. Voice Emotion Recognition

Two methods of displaying voice emotion feedback were explored in this study.

5.2.2.2.1. Method – Emotional Diamond

The first method was an emotional diamond that is offered by Nemesysco[ltd]. This can be seen in Figure 5.4.

Figure 5.4. Method 1: Emotional Diamond provided by Nemesysco[ltd].

5.2.2.2.2. Method 2 – Bar chart template

The second method was the same approach as for facial expression method 3 (bar chart style template). This can be seen in Figure 5.5.

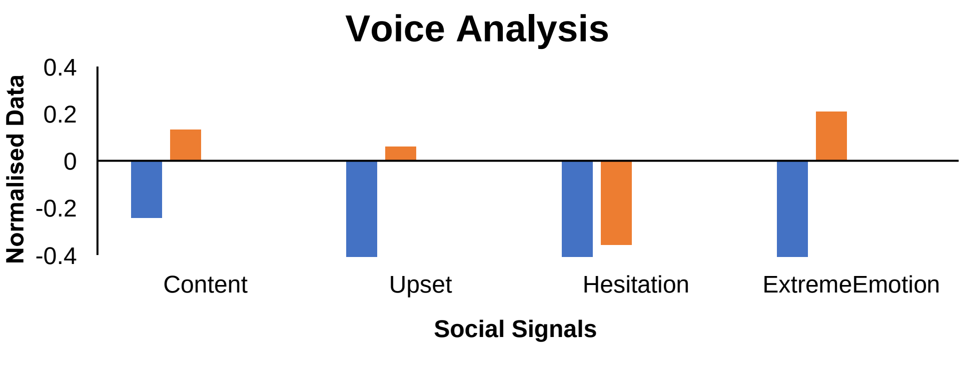


Figure 5.5.

Method 2: Bar chart style template for voice analysis feedback

5.2.2.3. Honest Signals

There was no visual display option provided by the sociometric badge general user interface (Sociometric Data Lab); therefore, it was decided to utilise the bar chart template in this instance to provide performance feedback to trainees. This was developed following the same procedure as facial expression feedback method 3 and voice emotion recognition method 2. This can be seen in Figure 5.6.

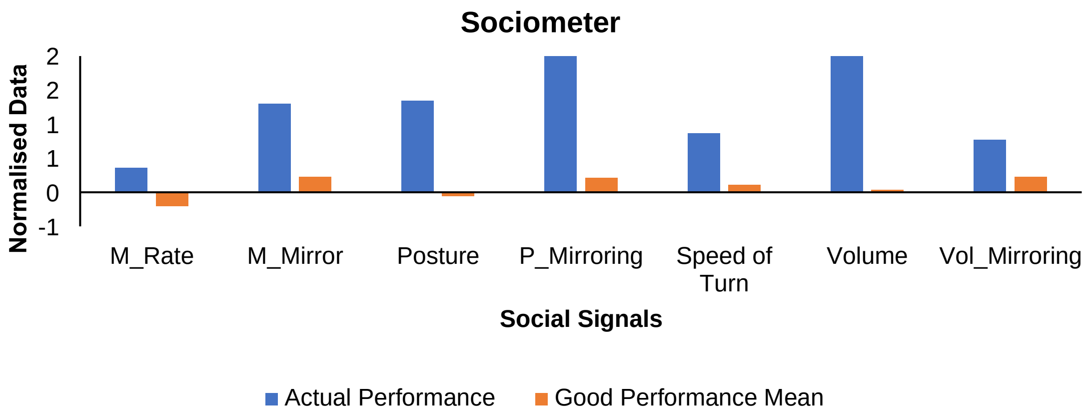


Figure 5.6. Honest signal feedback using bar chart feedback (Only Method)

5.2.2.4. Gestures

The only option that was provided for feedback was the video display offered by iMotions and a bar chart style template with an overall movement display. This follows the same procedure as method 3 for facial expression feedback, method 2 for voice emotion analysis and for method for honest signals. This can be seen in Figure 5.7.

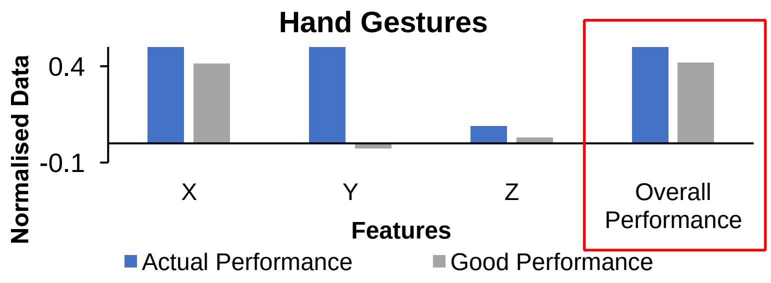


Figure 5.7. Bar chart style template with an overall movement element

While physiological recordings (e.g. GSR) were recorded in the exploratory stage, feedback was not developed for participants on these measures as this would be difficult to action.

5.2.3. Procedure

Participants were introduced to what the study would entail and were asked to give consent. Participants were provided feedback on their own interviews which were recorded in the exploratory stage with the different formats described above. Each individual participant was presented with each of the designs that were described above in the following order of presentation. A summary of the order of feedback can be seen in Table 5.1.

Table 5.1. Order of feedback method and feedback for each method

	Channel	Method of Feedback	Performance Feedback
1.	Facial feedback	Video playback, bespoke dashboard and bar chart template	Video was played back to trainees which would allow for trainees and trainers to pause and reflect on behaviours*
2.	Voice feedback	Emotional diamond and bar chart template	Talk more passionately about trainees' research*
3.	Honest signals	Bar chart template	Increase or decrease feature depending on comparison element *
4.	Gestures	iMotions and bar chart template	Increase or decrease gesture use based on comparison element. Video playback allows trainees to understand their use of gestures in context of discussion*

**Feedback from trainer to trainee would be to highlight key areas of improvement to not overload information which would restrict improvement (Baylor & Kim, 2009).*

Cognitive load theory proposes a human cognitive system that relates to the amount of information that working memory can retain at any one time (Sweller, 1988). Working memory holds information that is required to be processed (Baddeley & Hitch, 1974) which impacts development of learning materials so that information can be processed in a way that can be fully understand by the learner (Atkinson & Shiffrin, 1968). An example of this is when there is a 'problem space' between the learners' current ability and their desired ability. If the instructions provide the learner with too much information this will result in less effective learning and resulting a divided attention effect (Bennett & Flach, 1992).

Participants received feedback on their own behaviour in the exploratory study, this provides more realistic understanding of the different methods of feedback in context of their performance and how they can improve. The data presented to participants was only for the first 30 seconds of the entire duration of the media interview. After each presentation participants were asked to rate their understanding of the feedback display using a System Usability Scale (SUS) (Brooke, 1996). The SUS is a 10-item questionnaire with a total of 5 responses for each question which ranges from 'strongly agree' to 'strongly disagree' (Appendix 3.15). The SUS has been used to evaluate a number of systems and has been shown to be reliable (Bangor et al., 2008). Ratings from the SUS indicate the extent to which participants felt the method of feedback was usable (see Chapter 3).

In addition to the SUS, qualitative semi-structured interviews were conducted to obtain a more in-depth understanding of each participant's views on each method of presentation. This allowed more detailed discussion of the visual displays. Interview prompts were designed to probe participants on their views on each of the displays, including ones which they rated as preferred and not preferred. Interview questions are displayed in Table 5.2 below.

Table 5.2. Semi-Structured Interview Probing Questions for Social Signal Feedback Method

Semi-structured interview prompts
1. Please tell me which of the versions of the design you liked the best?
2. Why did you like this version better than the others you looked at?
3. Were there any elements of the less preferred version that you liked? If so, what?
4. Were there any elements of your preferred design that you don't like? If so, what?

Participants were remunerated £5 per hour for their participation. In total, the duration of the study was 1 hour per participant. A schema of the design procedure can be seen in Figure 5.8.

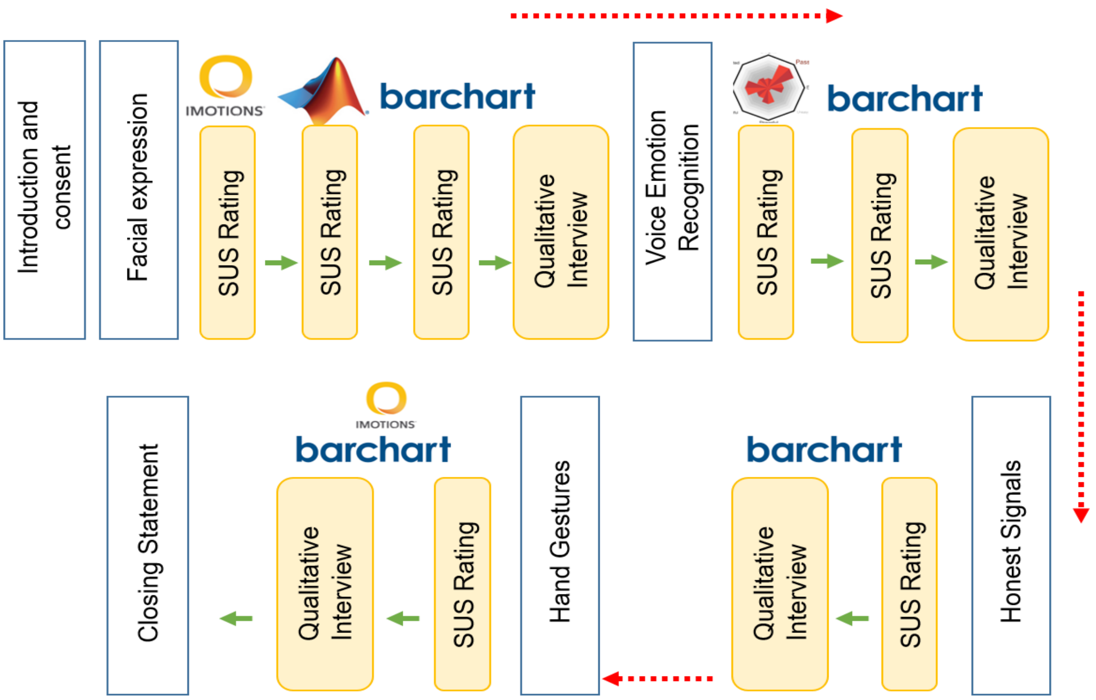


Figure 5.8. Procedure of user-centred design study

5.3. Data Analysis

5.3.1. System Usability Scale

The SUS yields a single number representing a composite measure of the overall usability of the system being studied. Each item's score contribution ranges from 0 to 4. For items 1,3,5,7 and 9 the score contribution is the scale position minus 1. For items 2,4,6,8 and 10, the contribution is 5 minus the scale position. The sum of these scores was multiplied by 2.5 to obtain the overall value of the systems usability (Brooke, 1996). This was calculated for each rating for each feedback option. These results were then summed and averaged for each option to illustrate how participants rated each system on average. A rating between 50 and 68 is indicative of moderate understanding and a score below 50 is concerning (Bangor et al., 2008; Kortum & Bangor, 2013). The benchmark for the usability of a method of feedback is a rating of 68.

5.3.2. Theme Identification

Interviews were transcribed verbatim. Semantic themes were identified within the explicit or surface meanings of the data, where analysis is not looking beyond what has been said or written. Furthermore, themes were identified in an inductive (bottom up) manner (Frith & Gleeson, 2004). Themes are discussed considering previous research and evidence from the qualitative interviews. Coding and theme analysis were completed according to Braun and Clarke's (2006) guidelines; 1) become familiar with the data (through conducting interviews and transcribing), 2) produce initial codes, 3) theme search, 4) review themes, 5) define generated themes and 6) write up.

5.4. Results and Discussion

The aim of the current research was to investigate the best method of providing social signal feedback that is actionable and understandable to trainees for improvement in communication skills.

5.4.1 System Usability Scale

A high SUS rating is indicative that participants found the method of feedback usable. The mean SUS scores for the emotional dashboard and iMotions for facial expressions were above the typical benchmark of usability at 74 and 70, respectively, while the bar chart was below benchmark with a SUS rating of 58. The results suggest that participants preferred the custom developed emotional dashboard as it is the easiest to understand of the three options. SUS descriptive statistics can be seen in Table 5.3.

Table 5.3. System Usability Scale Results for each method of feedback

Communication Channel	**Feedback Display**	**SUS (*SD*)**
Facial expression	iMotions	70 ↑* *(12.87)*
	Bespoke	74 ↑ *(10.09)*
	Bar chart	58 ↓* *(19.76)*
Voice	Diamond	84 ↑ *(9.12)*
	Bar chart	71 ↑ *(19.73)*
Honest Signals	Bar chart	72 ↑ *(16.45)*
Gesture	iMotions & bar chart	68 →* *(20.00)*

**Illustrating above benchmark is depicted in the table as↑, below benchmark as ↓ and on par as→*

Mean SUS ratings for voice analysis feedback suggest that participants preferred the Emotional Diamond method of providing voice emotion feedback with a rating of 84 compared to a value of 71 for the bar chart template. These results are consistent with the notion that humans are cognitively able to process information that presents multidimensional data in an octagon format or shapes that are symmetrical more effectively (Bennett & Flach, 1992; Coekin, 1970; Woods, 1991).

Participants rated the bar chart template for providing honest signal feedback from the sociometric badge data above the benchmark with a rating of 72 which is higher in this instance than for the bar chart in the context of facial expression feedback rated earlier in the study. This might suggest that participants had learned how to interpret the bar chart format to infer their performance upon repeated exposure.

The SUS ratings for the method of providing hand gesture feedback using iMotions together with the bar chart template was on par with a benchmark for understanding a system of 68.

The quantitative results suggest that the diamond (at 84) was the easiest to use format. However, the SUS ratings for bar chart on all but first exposure met or exceeded the benchmark. This could be a familiarity or learning effect; additionally, it could be that the bar chart is more suited to some types of social signals than others. To be sure, an experimental study would be required which is a possibility for future research. The next section provides more details about participants' understanding of each of the displayed methods.

5.4.2 Thematic Analysis of Interviews

5.4.2.1. Channel Specific Themes

Gestures and Voice Emotion Recognition: Scale considerations

Participants expressed concerns over the scale of the bar chart and how comprehensible it is. They did not understand what the numbers mean. However, each feature was normalised and presented to participants on the same scale so that they would understand each feature in relation to one another, i.e. their posture relative to their activity. This is important because it enables participants to understand their behaviour in a more cohesive way.

"*The scale is the only thing that I find hard to get my head around*." (P6, female)

Gestures and Facial Expression: Temporal Behaviour

Temporal behaviour is a timestamp of behaviour that is presented on a timeline so that trainees understand their use of hand gestures and facial expressions that has contextual information. This allows the trainee to grasp whether its use is appropriate.

Participants stated that they enjoyed observing the features in the bar chart template and compared it to the video of their interview.

> "*Well the fact that the first one is actually comparing behavioural of my gestures and then the temporal behaviour of the gestures is being translated into the graph. So, I can see that there is a behaviour of my temporal and a behavioural of my overall performance*." (P8, female)

Gestures and Sociometric Badges: Initial Guidance and Explanation

All participants stated that they would require an explanation for each of the social signals detected by the sociometric badges and their meaning. This is an important aspect of the feedback and would need to be integrated into feedback for participants to improve their performance.

> "*I need an assistant to tell me what's going on with these two, so if someone could tell me what this is and what is that and how is it, how is it, how this represents that one so that would be really helpful*." (P8 Female)

Facial Expression: Traditional feedback

Most participants preferred the iMotions video playback compared to the summative feedback methods (method 2 and 3). The primary reason given was because the playback of the video forms the primary basis for traditional feedback.

> "*This is how I did visually or in terms of video feedback and this is how I did statistically. That will be nice to see. So, have both of them side by side if possible*." (P4, male)

Voice Emotion Recognition: Value of Visual Display

Participants indicated that they preferred the colours of both feedback options (bar chart and diamond) as this would enable them to find the key elements for their feedback and focus on those points for improvement.

> "*It was the visual appeal of it, and it was quite simple to understand*" (P6, female)

The colour scheme of the bar-charts assists trainees in distinguishing between good and bad performances. This colour difference draws on previous research that found that colours draw a person's attention and as a result improves memory performance (Dzulkifli & Mustafar, 2013). Colours to draw attention to key features have also been utilised in previous research during real-time feedback highlighting incorrect behavioural cues in red and correct cues in green (Ali & Hoque, 2017).

5.4.2.2. Overarching Themes Identified

Comparison to good performance

Participants indicated that they valued the comparison element of the bar chart. One participant mentioned that they would like to see the maximum and minimum of what is considered a good performance.

> *"It will be good to know what kind of, it will be good to know where kind of the maximum amount of movement is. Where, if you were going above this threshold you'd be basically moving too much."* (P4, male)

Another participant stated that they would need to understand how much they moved their hands in comparison to other participants.

> *"I need to know how much or how little I moved my hands moved compared to others."* (P6, female)

This response suggests that participants were interested the threshold response that the bar chart / post summary feedback offers. This was utilised in Damian et al., (2015) training in presentation skills study who had implemented a threshold response; however, in this study the feedback provided was in real-time and participants did not often reach the threshold, suggesting inappropriate implementation of the threshold technique.

Combination of Displays

Participants recognised that there are pros and cons to the different displays and propose that these could be overcome by combining some displays. For gestures and facial expression participants prefer the conversion of the iMotions video and emotional dashboard into a numerical format which is compared to a 'good' performance. The bar chart was suggested by participants to present feedback in conjunction with the video playback function of iMotions. The primary reason for this was that the bar chart provides an overall performance that is compared to what is expected in a good performance.

> "*Combining option 3* [bar chart template] *with option 1*[video playback] *with the feature of the other people you know with the average of performance indicator, then that means I can see my face and I can see where I went wrong or where I went well compared to others*." (P6 female)

This is an important finding as video playback of interviews improves participants' self-awareness which has proven effective for improving communication skills (Argyle, 1988; Fung et al., 2015; Roter et al., 2004; Taylor, 2015; Zhao, Li, Barbosa, Ghoshal, & Hoque, 2017) and has been found in previous research for improving self-awareness.

Furthermore, previous research in training using automated feedback systems has found that a post summary feedback format is effective in improving communication skills (Tanaka et al., 2015).

Visualisation of social signal feedback was presented using a simple bar chart style template that contains a 'good performance' component in a summative format (Pereira, Colecchia, & Hone, 2018). This type of feedback is similar to the performance threshold developed in Damian and colleagues'

(2015) behavioural feedback loop system, Logue, that provides real-time feedback to participants about their use of social signals during a presentation. This threshold was implemented as a means of generating a level of appropriate or inappropriate behaviour, whereas the 'good performance' bars appropriateness level was generated by high and low thresholds of performance were obtained from the design study reported in Chapter 5. A bar chart style method of visualising feedback was also presented to participants in a study that included a model behaviour component (Tanaka et al., 2016).

5.5. Conclusion

The aim of this stage was to identify the most appropriate method of providing feedback:

What is the best method of presenting social signal feedback in the context of media interview training that is actionable and understandable to trainees?

The easiest display for the participants to use for facial expression was the bespoke customised design and for the voice emotion analysis displays participants preferred the Emotional Dashboard provided by Nemesysco Ltd QA 5. The SUS illustrated that while some systems are more easily usable, especially on first exposure, consistency aids in usability of display. Specifically, the bar chart was rated higher with repeated exposure. The consistent presentation of the bar chart display in each of the channels limits the number of display variables that are presented, thereby ensuring users do not have to learn new presentations for each channel.

The results obtained from the qualitative interviews identified two overarching themes across all communication channels. These included '*comparison to good performance'* and '*combination of displays'.* These themes were identified as participants repeatedly remarked on the performance threshold component presented by the bar chart which permitted participants to gain a clear understanding of what was expected. This may be because it reduces divided attention and will not tax cognitive resources by increasing cognitive load due to too many variables to consider in one display and will be overcome by a trainer highlighting key areas of improvement (Bennett & Flach, 1992). Even though participants had some concerns over understanding the scale of the bar chart template, this can be overcome by help from a trainer. This finding pertains to the *actionable* method of feedback highlighted in the research question.

Sub-themes were identified across communication channels which include scale considerations; value of visual display; guidance and explanations; and temporal behaviour. The themes identified suggest that joining feedback display methods where both temporal behaviour and traditional playback of interviews are discussed are important for feedback. The iMotions emotional dashboard has both capabilities. This gives participants the opportunity to stop / pause playback and discuss if there are any emotions that were displayed. Participants enjoyed the comparison element and consistency of the bar chart style template. Visual displays identified across communication channels pertains to the *understandable* method of feedback highlighted in the research question.

5.6. Affective Feedback Design

The findings from the exploratory stage and the current stage were collated to inform the design of the intervention of this research.

5.6.1. Feedback Selection Rationale

The following section describes the method of training feedback that will be delivered to trainees in Chapter 6. Table 5.4 summarises the design decisions made based on the work reported in both stages (Chapter 4 and 5) and shows the feedback presentation formats chosen for each signal type. Justification for each selection are detailed below.

Table 5.4. Experiment Feedback

Channel	Visual Display
Facial expression	Video Playback on iMotions Bar Chart Template
Hand Gestures	Video Playback on iMotions Bar Chart Template
Honest Signals	Bar Chart Template
Vocal Affect	Bar Chart Template

5.6.1.1. Video Playback – Formative Feedback

This will allow participants to watch their interview for self-reflection and discussion. This is effective for training as it raises self-awareness which has been found to improve communication skills (Schneider, Borner, Van Rosmalen, & Specht, 2016).

5.6.1.2. Behavioural Threshold - Post-summary feedback

The bar chart will provide a summary of the first 30 seconds of the interview for participants to understand their honest signals and voice emotions displayed. The trainer will observe the bar chart and note important changes that need to be made in the initial stages of the interview. Trainees can also infer behaviour changes, but the trainer will moderate the training feedback not to overload information which will impact information processing (Sweller, 1988).

In addition to having the good performance element in the bar chart, it was decided to add maximum and minimum values to the bar chart which provide more information about what is considered ‘good’. This is based on the *‘comparison to good performance’* theme and proposed by P4. However, participants would need a clear explanation about the scale used in the bar chart template.

5.6.1.3. Formative and Summative Feedback

Both playback of the video and the bar chart template that includes a performance threshold will benefit the trainee as the video will improve self-awareness while the bar-chart template will give the trainee details on how they performed in the initial stages of the interview compared to a performance threshold, allowing them to adjust their performance. Trainees will be able to playback their videos and relate it to the scale presented to them in the form of a summative feedback method.

The next step in this PhD research is to investigate whether presenting this method of feedback to trainees presenting them with the signals selected in the preliminary data analysis is effective in improving communication skills in the context of media interviews.

CHAPTER 6. ENHANCING COMMUNICATION SKILLS TRAINING THROUGH SOCIAL SIGNAL FEEDBACK

6.1. Introduction

Previous chapters identified the social signals necessary for effective communication in a media interview context, based on a preliminary analysis (Chapter 4). Chapter 5 recognised a method of feedback that is actionable and understandable to trainees. The method selected for feedback of social signals included playback of interviews and a comparative feature that allows trainees to understand how they need to improve. This chapter is an experimental evaluation of the social signal feedback technique developed in Chapter 4 and 5 which will be compared to standard media training feedback.

The research question for this experiment stage is the following:

Is the provision of social signal feedback more effective in enhancing communication skills during a person-person discourse compared to standard feedback provision?

The experimental hypotheses are:

H1: There will be significant improvements in performance (as measured subjectively) from pre-test to post-test interview in both training conditions (main effect of training);

H2: There will be significant changes in observed social signals detected between pre-test and post-test interview in both training conditions (main effect of training);

H3: Training gains (measured subjectively) from pre-test to post-test will be greater for the experimental condition (interaction effect);

H4: Greater changes in social signals will be detected between pre-test and post-test interview for the experimental condition (interaction effect).

6.2. Data Collection

6.2.1. Participants

An a priori test was carried out using G*Power to estimate the sample size for the design of this study (see Chapter 4). Briefly, G*Power was instructed based on a large effect size, a power of 0.95 and an alpha value of 0.01. This resulted in a total of 22 research staff and students recruited (age ranged from 18 – 55 years old; 6 male and 16 female) for this stage. Experience in public speaking ranged from no experience to extensive and from none to some experience in media interviews. The roles that participants had within the university included taught students (3), research staff (1) and research students (18).

Participants were from different cultural backgrounds including 6 native English speakers (participants who stated that English was their first language) and 16 non-native English speakers. Nationalities included were Lithuanian (3), British (6), Brazilian (2), Chinese (1), South African (1), Malaysian (1), Korean (1), Iranian (2), Taiwanese (1), Swedish (1), German (1), Nigerian (1) and not reported (1).

6.2.2. Research Design

6.2.2.1. Performance Evaluation

A 2x2 mixed ANOVA was used to analyse the difference in subjective ratings of communication performance. The between-participants independent variable was *feedback type* which has two levels: social signals feedback and traditional / standard feedback. The within-participants independent variable was *session* which also has two levels: pre-training and post-training interviews. The dependent variables were subjective judgements of communication skills provided by the journalist and three neutral observers in the pre-training and post-training interview.

6.2.2.2. Social Signal Use

A Principal Component Analysis (PCA) was used to obtain components of social signals which are correlated and linearly uncorrelated variables (Field, 2013). This was used to reduce, and merge similar variables included in the analysis (Cheng, Wang & Carrol, 2014). A multivariate analysis was conducted on the components selected for inclusion in the analysis where the between-participants independent variable was *feedback type*: social signals feedback and traditional feedback. The within-participants independent variable was *session*: pre-training and post-training interviews. The dependant variables were the components extracted from the PCA analysis.

6.2.2.3. Qualitative Interviews

Qualitative interviews were conducted to explore participants' views about the value of different types of training feedback. Semi-structured interviews were conducted to understand what participants found helpful / unhelpful about the feedback that they received during the training. Interviews were analysed using thematic analysis where themes extracted using Braun and Clarkes' (2006) guide and are based on verbally spoken words and not interpreted. A set of 5 key interview questions were developed for both groups and the experimental group were asked an additional five questions which can be seen in Table 6.1.

Table 6.1. Interview probing questions

Interview Questions for both conditions	Interview questions for experiment condition
1. How did you feel about the feedback you received on your performance today?	1. How did you feel about seeing the system feedback of your use of emotional signals and body language during the course of the training session today?
2. What, if anything, did you feel you were able to change about how you presented yourself today based on the feedback that you received?	2. Do you feel that the system feedback helped you appreciate aspects of your performance that you might not have noticed by just watching / hearing the playback?
3. To what extent do you feel that your performance in the areas identified in the feedback improved over the course of the day?	3. To what extent did you feel that the feedback you received was ‘actionable’ – that is to what extent did you think you could control or improve the behaviour which was highlighted in feedback?
4. To what extent do you feel the feedback you received today will help you present yourself better in the future?	4. Could the feedback have been presented in a clearer way? If so, what would you change about the way the feedback was presented?
5. Where there any aspects of your performance that you would have liked to have received feedback on which were not covered by the feedback you received? If so, what were these?	5. Were there additional aspects of your non-verbal behaviour that you think it would have been useful to receive feedback on? If so, which?

6.2.3. Materials and Measures

6.2.3.1. Subjective Assessment of Communication Skills

Like the exploratory stage, the CSRS was used to measure communication skills described in Chapter 3 (section 3.6). Ratings were obtained by three neutral observers who were not present on the day of the workshops, the journalist and the participants.

6.2.3.2. Emotion-recognition COTS Technology

A combination of media (cameras, tripods and voice recorders) and affective recognition technology was used to capture social signals during interviews like the exploratory stage (Chapter 4). However, the methods were different. Some data was collected in real time including voice affect, facial expression and hand gestures; the only data captured which was post-processed was sociometric badges due to the way the technology is designed (no visual display). A summary of how data was used can be seen in Table 6.2. Chapter 3 (section 3.8) contains more details about the software specificities.

Table 6.2. COTS technology, media technology and the methods used to capture signals

Social Signals	COTS Technology	Media Technology	Method
Voice Emotion Recognition	Vocal Affect Recognition	Zoom H4N Pro Handy Audio Recorder Laptop was positioned in front of participant to capture their voices	Real-time
Honest Signals	Sociometric Badges	■	Post-processed
Facial Expression	iMotions (Affdex + Shimmer 3)	Sony PJ 220 Handycam camera Adobe Photoshop (video trims)	Real-time
Hand Movement		■	Real-time

6.2.4 Journalist / Interviewer Experience

The content of this workshop included standard media skills training format was similar to the expert training observed in the exploratory stage (Chapter 4). However, for this stage the interviews were conducted by journalists who were still in training with some experience in field work. The difference between the journalists recruited in Chapter 4 and this workshop is that the exploratory stage' journalists were required to be experts whereas in this experiment they were to act solely as journalist and not as trainer.

6.2.5. Procedure

Training workshops were conducted at Brunel University London in Uxbridge (October 2017 - December 2017). Upon arrival, participants were introduced to the study and at this point demographics and consent were collected. Participants then engaged in a pre-test interview where subjective ratings and social signals were recorded. Participants were then split into pairs matched as closely as possible by gender, native language (English as first language), and based on average pre-test communication skill ratings. One member of each pair was allocated to the experimental or control condition at random (by the toss of a coin). Participants then engaged in a 30-minute pre-recorded lecture which introduced them to communication in media interviews.

Next, participants individually engaged in their training session which comprised of a radio, face-face and down-the line interview and lasted a total of 3 hours. Down-the-line interviews are where the interviewee cannot see the interviewer but can hear the questions asked and were instructed to look directly into the camera. The down-the-line interviews are found to be challenging, and generally people tend to find it challenging to master this interview technique (Taylor, 2015). For the same reasons discussed in the exploratory stage, difficulty of interview type and questions increased and received feedback following each practice interview. Following the practice session, participants took part in a post-test interview which was equally challenging as the baseline interview. This was done to ensure that any signals recorded were a result of the manipulation of the independent variable and not the nature of the interview. Furthermore, like the exploratory stage, practice interviews were conducted individually to reduce social pressure which is likely to affect performance. Both the control and the experimental conditions received feedback from the journalist after each of the three training interviews and video playback was presented following this for the experimental condition. However, the experiment condition received additional social signal feedback seen in social signal feedback visualisation section. All participants then engaged in a post-training interview where dependent variables were recorded like the pre-training. Participant interview recordings were later rated by three neutral observers using the CSRS. As in the exploratory stage, instructions for answering questions were explained as instructed by (Spitzberg & Adams, 2007). Both the journalist and neutral observers were blind to the participant feedback condition. Finally, participants were asked to take part in qualitative interviews to evaluate both methods of training feedback to understand what was helpful / unhelpful about the feedback that they received. At this point, participants were also asked to fill in a two-item questionnaire to evaluate participants perception of their confidence and skills post-training. Answers were rated on a scale ranging from no improvement to great improvement. Questions include the Table 6.3.

Table 6.3. Self -evaluation questionnaire post training

Self-evaluation Questionnaire
1. Please indicate on the scale below how much you estimate your **skill** in giving a media interview has improved over the course of today's training event
2. Please indicate on the scale below the extent to which your **confidence** in giving a media interview has improved over the course of today's training event

Participants were reimbursed £5 per hour in recognition of their time. The outline of the feedback for the experiment group can be seen in Figure 6.1. The traditional feedback group would receive the same method of feedback excluding the summative feedback.

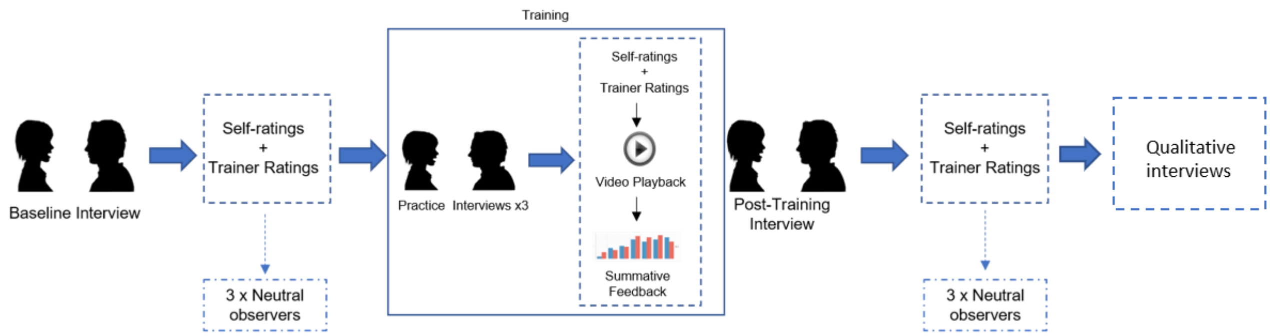


Figure 6.1. Experiment feedback outline for social signal feedback training

6.2.6. Social Signal Feedback Visualisation

6.2.6.1. Participant Feedback Procedure

The feedback design was selected to be delivered included a playback of video with the emotional dashboard presented below, a bar chart style template with a threshold for what is considered a good and bad performance. The method of feedback was based on participant responses in Chapter 5.

6.2.6.1.1. Traditional Feedback Group

After each interview participants were played back their videos for reflection and to improve self-awareness as this component as research has shown that those who are self-aware are known to be effective communicators (Hass & Eisenstadt, 1990; Schneider, Borner, et al., 2016; Wicklund, 1979). Videos were presented to participants without the facial point markers on the video. Participants were also given verbal feedback by the journalist about their performance.

6.2.6.1.2. Social Signal Feedback Group

The social signal feedback group also had their videos played back to them to improve self-awareness with the addition of the emotional dashboard below the video (Figure 6.2 below). Participants were given verbal feedback by the journalist and were presented with a summative feedback after each interview playback from the first 30 seconds of an interview.

Not all signals selected for feedback in the preliminary stage were fed back to participants (Chapter 4). Signals selected for feedback by the researcher based on whether participants' performance exceeded the minimum and maximum threshold (Good performance min [yellow] and max [grey]). Selective feedback was done to diminish an increase in cognitive overload as this has a negative impact on task performance as information held in our working memory that transfers into long-term memory is affected (Frein, Jones, & Gerow, 2013). Additionally, by feeding back additional information that is irrelevant, comprehension will be reduced (Sanchez & Wiley, 2006), otherwise known as the *seductive details effect* (Garner, Gillingham, & White, 1989). Selected reduced cue feedback to reduce cognitive load was also done in Ali (2017).

6.2.6.1.2.1. Facial Expression and Hand Gesture Feedback

Video playback demonstrating participants' facial expressions can be seen in Figure 6.2. This is a formative structure which allowed participants to view their overall performance for self-reflection. An emotional dashboard below the video display was used to observe expression and expression range (experimental condition only).

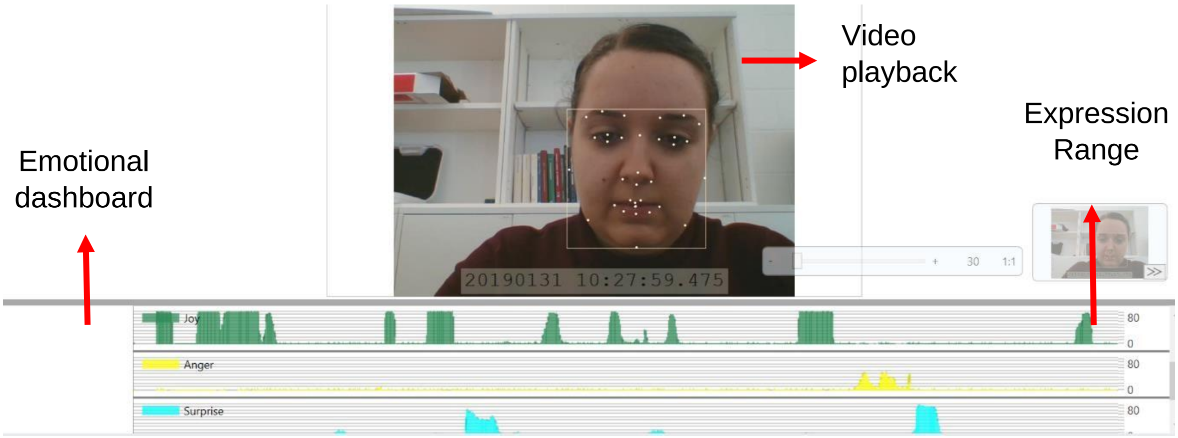


Figure 6.2. Formative exemplar feedback for facial expression and hand movement

An exemplar summative feedback of facial expressions presented to participants can be seen in Figure 6.3. Features fed back to participants included ‘smile’, ‘smirk’, ‘anger’, ‘sadness’, ‘disgust’, ‘joy’, ‘surprise’, ‘fear’, ‘contempt’ and ‘brow furrow’ (frown). Actual feedback given to participants were individualised for each participant based on their behaviours.

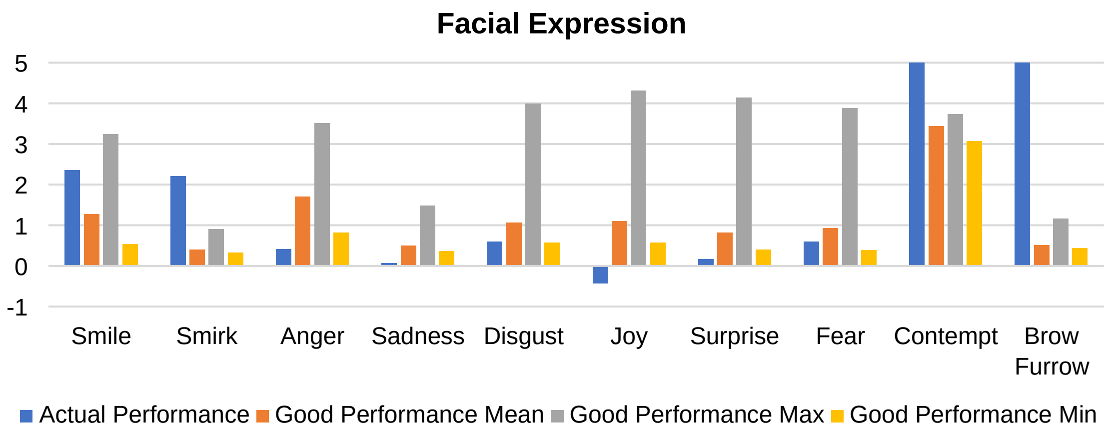


Figure 6.3. Exemplar summative feedback for facial expression

6.2.6.1.2.2. Honest Signal Feedback

The summative method of providing honest signal feedback can be seen in Figure 6.4. Features feedback includes ‘movement rate’ (m_rate), ‘movement mirroring’ (m_mirroring), ‘posture’, ‘posture mirroring’ (p_mirroring), ‘turn taking speed’ (speed of turn), ‘volume’ and ‘volume mirroring’ (vol_mirroring). Figure 6.4 demonstrates an exemplar feedback bar chart; participants actual feedback was individualised based on their displayed behaviours.

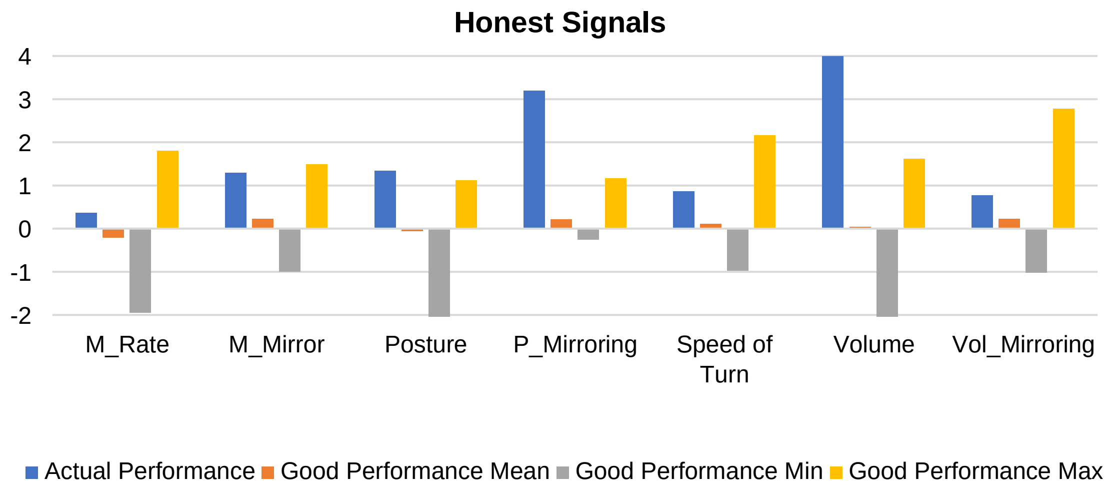


Figure 6.4. Exemplar summative feedback for honest signal displays

6.2.6.1.2.3. Vocal Affect

Summative feedback is presented by Figure 6.5 which demonstrates summative feedback of vocal emotion including ‘content’, ‘upset’, ‘hesitation’ and ‘extreme emotion’. This figure shows an example of how feedback was presented to participants. Their actual performances were individualised based on their behaviour and then fed back to them.

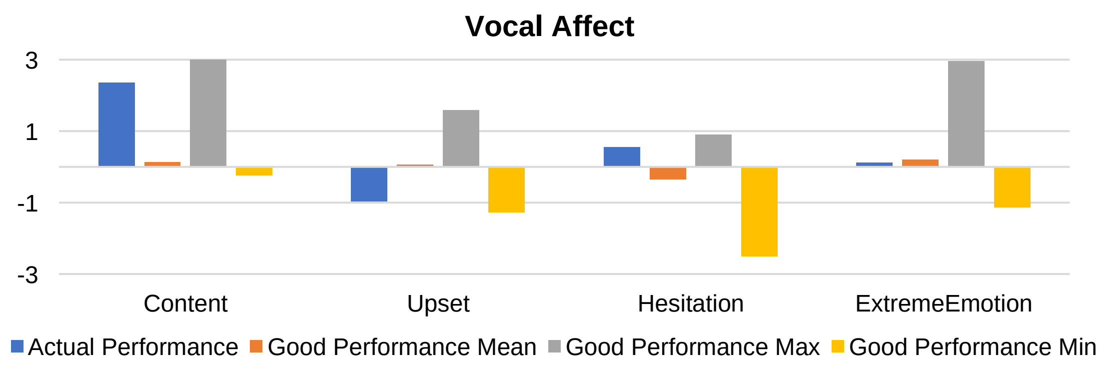


Figure 6.5. Exemplar summative feedback for emotional vocal behaviour detection

6.2.7. Subjective Ratings of Performance

Subjective ratings of performance were collected using three versions of the Conversation Skill Rating Scale (CSRS) (Spitzberg & Adams, 2007); self-report, observer and a ‘rating of partner’ version that includes 25 conversational feature scale ratings and a composite measure of 5-items on communication performance called molar scores. For this study, molar scores were used as a measurement of communication performance.

Two further questions were aimed at obtaining a rating of participants’ confidence and skills based on the training they received:

1. Please indicate on the scale below how much you estimate your ***skills*** in giving a media interview has improved over the course of today's training event.
2. Please indicate on the scale below the extent to which your ***confidence*** in giving a media interview has improved over the course of today's training event.

Participants in both conditions had rated their skills and confidence after training on a five-point scale of 'no improvement', 'slight improvement', 'some improvement', 'considerable improvement' and 'great improvement'.

6.3. Data Analysis

6.3.1. Subjective Ratings of Communication Skills Performance Evaluation

6.3.1.1. Data Pre-processing

Internal consistency of the CSRS was assessed using Cronbach's alpha and agreement among raters was assessed using intraclass correlation (Koo & Li, 2016; Mandrekar, 2011). The experimental hypotheses related to subjective judgements of participant performance were tested by means of 2x2 mixed factorial ANOVAs. Separate analyses were run on the journalist's judgements, neutral observer judgements and self-evaluations. A between-subjects t-test was run to investigate self-evaluation of skills and confidence across the two feedback conditions.

6.3.1.2. Self-report vs Journalist Ratings of Communication Skills

A 2x2 mixed ANOVA was used to investigate whether participants had perceived their performance as significantly different than the journalists in the pre-training and the post-training interviews. The independent variables were raters (self-report vs journalist) and session (pre-training vs post-training) and the dependent variable were ratings of self-reports of communication skills. The rationale for this investigation is that previous research has found that participants often misrepresent their abilities when self-reporting their capabilities in comparison to the competence measured by their conversational partner (Spitzberg & Adams, 2007).

6.3.2. Social signals

6.3.2.1. Data Pre-processing

Social signal analysis included facial expression were derived from the facial recognition software, honest signals captured using sociometric badges and emotions using voice analysis software. To meet the assumption of normal distribution the data was normalised using the minimum and maximum values of the dataset for each signal using the following formula:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Data was averaged 10s intervals which introduced 24 trials (4 minutes). In contrast to the exploratory stage (Chapter 4) that investigated the first 30 seconds of an interview, this stage investigated the whole interview. This was done because researchers wanted to assess whether participants would implement the feedback given to them during appraisals and assessing the first 30 seconds would be limiting as they would not be able to effectively demonstrate improvements. Data not collected by the technology were considered missing data. A tabulated pattern analyses of the missing data was produced for ‘pitch’, ‘volume mirroring’, ‘volume mirror lag’ and ‘speed of turn-taking’ because of missing data and produced no cohesive pattern in missing data and, as a result, was removed (Tabacknick, Fidell, & Ullman, 2007). Hand movements were also removed as the data collected was produced too many missing data. ‘Content’ detected by voice emotion analysis was also removed as all values produced by Nemesysco were mostly ‘0’ which significantly skewed the data. The resulting total number of valid cases produced were 646 cases.

6.3.2.2. Assumption Testing PCA

Significant Outliers. The method used to normalise the raw data that uses the minimum and maximum values of each feature / signals suppresses any significant outliers (Patro & Sahu, 2015). Upon further exploration of the data, there were no significant outliers when manually look at the data.

Sampling Adequacy. The current dataset includes an adequate sampling for a PCA and KMO produced a .632 which is above the threshold of .50 for adequate sample sizes (Field, 2013).

Bartlett’s Test of Sphericity. This tests the null hypothesis that the variables are unrelated and therefore suitable for structure detection, the value produced for this dataset was p = < .001 which suggests that PCA can be conducted. The data is suitable for data reduction as Bartlett’s test of Sphericity is less than p = < .001 (Field, 2013).

6.3.2.3. Assumption Testing for Multivariate Analysis

A multivariate analysis was conducted which included all the components extracted and produced by the PCA. Assumptions for multivariate test are discussed below.

Significant Outliers. Each component was analysed, and it was identified that there was an outlier for component 7. This was removed, resulting in 465 cases for this component.

Multivariate Normality. A Shapiro Wilks test was observed for normality and when the p value was less than .05 the researcher observed the distribution at the level of the eye. Histograms were normally distributed, and each signal / feature contained a pre-training interview histogram similar in its post training interview histogram demonstrating consistency in signals. This is acceptable to run an ANOVA (Schmider, Ziegler, Danay, Beyer, & Bühner, 2010). This enabled analysis using the PCA-

ANOVA method. Normality tests can be seen in Table 6.4. Similar skewness has been noted as appropriate for conducting comparative tests (Tabachnick & Fidell, 2013).

Table 6.4. Shapiro Wilk p-values for extracted PCA components

n	Pretraining	Post Training	Control Group	Experiment Group	Description
1	$p = .257$	$p = .527$	$p = .009$	$p =.227$	Normality assumed
2	$p = < .001$	$p = < .001$	$p = < .001$	$p = < .001$	Similar Skewness (eye observation)
3	$p = < .001$	$p = < .001$	$p = < .001$	$p = < .001$	Similar Skewness (eye observation)
4	$p = < .001$	$p = < .001$	$p = < .001$	$p = < .001$	Similar Skewness (eye observation)
5	$p = < .001$	$p = < .001$	$p = < .001$	$p = < .001$	Similar Skewness (eye observation)
6	$p = < .001$	$p = < .001$	$p = < .001$	$p = < .001$	Similar Skewness (eye observation)
7	$p = .001$	$p = .001$	$p = < .001$	$p = .125$	Similar Skewness (eye observation)

Homogeneity of Variance. A Levene's test revealed there is no homogeneity of variance assumed for components 1-5 and 7, $p = < .001$. There was assumed homogeneity of variance for component 6, $p = .469$. It has been suggested that if homogeneity of variance is not assumed then multiple univariate ANOVAs should be conducted, and interpretation of the alpha level should be stricter at $p = < .01$ (Allen & Bennett, 2007). A Welch's statistic will be reported to adjust for unequal variances in the datasets as it is a more powerful and conservative test compared to the Brown-Forsythe statistic for unequal variances in one-way ANOVAs (Field, 2013). The Box's M was significant ($p = < .001$); however, ANOVAs are robust when groups are larger than 30 (Field, 2013). The process of data analysis can be seen in Figure 6.6.

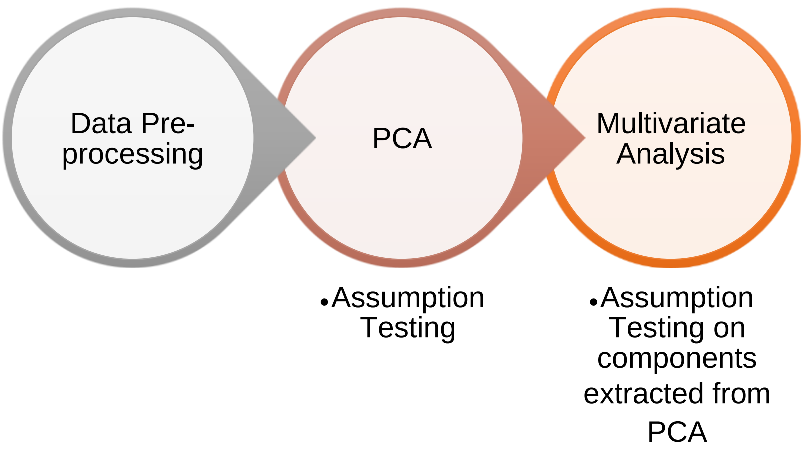


Figure 6.6. Process of data analysis

6.4. Results

A total of 11 trainees were randomly assigned to the social signal feedback group which included 3 males and 8 females whose age ranged from 18 to 45 years. Countries that trainees were from included Lithuania (1), Brazil (2), South Africa (1), Malaysia (1), Iran (1), United Kingdom (1), Germany (1), Nigeria (1), Bhutan (1) and Italy (1). Trainee roles within the university included research students (8), taught students (2) and research staff (1).

A total of 11 trainees were randomly assigned to the traditional feedback group which included 3 males and 8 females whose age ranged from 18 – 55 years. Countries that trainees in the control group were from included Belarus (1), China (1), South Korea (1), Lithuania (2), Sweden (1), Iran (1), and United Kingdom (1) and prefer not to say (1). Trainee roles within the university included research students (10) and taught students (1)

6.4.1. Subjective Ratings of Performance

6.4.1.1. Internal Consistency of Communication Skill Ratings

The CSRS communication skill ratings were assessed for internal consistency. A high internal consistency of α = .838 was obtained for *journalist molar ratings* for the pre-test interview. A moderate internal consistency was obtained for the post-test interview α = .677. A composite mean of all *three neutral observers'* molar ratings was obtained. A high internal consistency of α = .965 in the context of a media interview for the pre-test interview as well as for the post-test interview α = .965. As a *selfreport* measure the CSRS produced an internal consistency of α = .901 which is high for pre-test interview and for post-test interview α = .944.

The degree to which raters agreed on the participant's performance was a moderate to high. The average measure intraclass correlation was .696 with a 95% confidence interval from .394 to .863 ($F_{(21, 42)} = 3.79$, p = <.001) in the pre-training interview. Neutral observers were in moderate agreement in the post-training interview with an average measure of .603 with a 95% confidence interval from .146 to .827 ($F_{(21, 42)} = 3.868$, p = <.001).

6.4.1.2. Journalist Ratings of Performance

The normality assumption was met for pre-training data (p = > .05) and post training data (p = > .05). The assumption of equal variance was met (p = > .05). The mean and standard deviations of the journalist scores across the four conditions and the descriptive statistics can be seen in Table 6.5.

Table 6.5. Descriptive statistics of journalist ratings of participants' communication skills

Item	Traditional feedback Group		Social Signal Feedback Group	
	Pre-test	Post-test	Pre-test	Post-test
Communication Score	5.091 (.887)	5.873 (.734)	4.909 (.831)	6.145 (.614)

On average journalists rated the control group (traditional feedback) as improving by 15% on the subjective rating scale for overall communication skill and the experimental group (social signal feedback) as improving by 25% on the subjective rating scale. A 2 x 2 mixed ANOVA (feedback group [social signal feedback vs standard feedback group] vs time [pre-post training interview]) was conducted on communication skills rated by the journalist and results can be seen in Table 6.6.

Table 6.6. ANOVA results of journalist ratings of participants' overall communication skills

Conversation Score	*MS*	*F (df)*	*P value*	*Partial Eta Squared*
Time	11.201	18.725 $_{(1, 40)}$	< .001	.319
Group	.023	.038 $_{(1, 40)}$	.846	.001
Time x Group	.568	.950 $_{(1, 40)}$	.336	.023

Note: MS = Mean squares, effect size = η^2

A follow-up analysis of significant main effect illustrates that participants who received social signal feedback significantly improved from pre-test to post-test ($F_{(1, 20)}$ = 15.749, p = <.001, η^2 = 0.441, d = 1.692). Furthermore, journalist rated participants who received standard feedback significantly improved from pre-test to post-test ($F_{(1, 20)}$ = 5.074, p = .036, η^2 =0.202, d = 0.961).

6.4.1.3. Neutral Observer Ratings of Performance

The normality assumption was met for the pre-test interview (p = > .05) and for the post training interview (p = > .05). The homogeneity of variance assumption was also met (p = > .05). The mean and standard deviations of the neutral observer scores by feedback group and session can be seen in Table 6.7.

Table 6.7. Descriptive statistics of neutral observer ratings of participants' communication skills

Item	Traditional feedback Group		Social Signal Feedback Group	
	Pre-test	Post-test	Pre-test	Post-test
Communication Score	4.176 (.948)	4.787 (1.031)	4.436 (.891)	5.303 (.729)

On average neutral observers rated the control group (traditional feedback) as improving by 15% on the subjective rating scale for overall communication skill and the experimental group (social signal feedback) as improving by 20% on the subjective rating scale. A 2x2 mixed ANOVA (feedback group [social signal feedback vs standard feedback group] vs time [pre-post training interview]) was conducted on communication skills rated by neutral observers and results can be seen in Table 6.8.

Table 6.8. ANOVA results of neutral observers' ratings of participants' overall communication skills

Overall Conversation Score	*MS*	*F (df)*	*P value*	*Effect size*
Time	6.016	7.324 $_{(1, 40)}$	.010	.155
Group	1.654	2.013 $_{(1, 40)}$	.164	.048
Time x Group	.179	.218 $_{(1, 40)}$	.643	.005

Note: MS = Mean squares, effect size = η^2

A follow-up analysis of main effect that were significant illustrates that participants who received social signal feedback significantly improved in communication ratings from pre-test to post-test by neutral observers ($F_{(1, 20)}$ = 6.24, p = .021, η^2 = 0.238, d = 1.065). Furthermore, there was no significant difference between the pre and post-test for neutral observer ratings for traditional feedback group ($F_{(1, 20)}$ = 2.101, p = .163, η^2 = 0.095, d = 0.614). However, the difference in group effects did not translate into a significant interaction effect in ANOVA results.

6.4.1.4. Self-Report Ratings of Communication Skill Performance

The normality assumption was met for the baseline interview (p = > .05) and for the post-training interview (p = >.05) the normality assumption was also met. Homogeneity of variance was also met (p = >.05). The mean and standard deviations of the self-report scores can be seen in Table 6.9.

Table 6.9. Descriptive statistics of self-report ratings of communication skills

Item	Traditional feedback Group		Social Signal Feedback Group	
	Pre-test	**Post-test**	**Pre-test**	**Post-test**
Communication Score	5.055 (.759)	4.946 (1.096)	4.855 (.913)	4.982 (.969)

A 2x2 mixed ANOVA (feedback group [social signal feedback vs standard feedback group] vs time [pre-post training interview]) was conducted on self-report communication skills and results can be seen in Table 6.10. There was no significant main effect or interaction effects for self-report ratings of overall performance.

Table 6.10. ANOVA results of self-report ratings overall communication skills

Overall Conversation Score	***MS***	***F*** $_{(df)}$	***P value***	***Effect size***
Time	.001	.001 $_{(1, 40)}$	.975	.000
Group	.074	.083 $_{(1, 40)}$	.775	.002
Time x Group	.154	.173 $_{(1, 40)}$	.680	.004

Note: MS = Mean squares, effect size = η^2

6.4.1.5. Confidence and Skills Ratings

Ratings of confidence and skills were measured between groups to assess trainees' perceived confidence and skills following training. The mean and standard deviations of the self-evaluation scores between the two groups can be seen in Table 6.11.

Table 6.11. Descriptive statistics for self-evaluations of confidence and skill post-test

	Social signal Feedback Group	**Traditional Feedback Group**
Confidence	4.364 (.674)	3.364 (.809)
Skill	4.091 (.831)	3.182 (.874)

An independent t-test (social signal feedback vs traditional feedback group) was conducted which revealed that participants' who received social signal feedback reported higher *confidence* post-training (*M* = 4.36; *SD* = .67) than participants who did not (*M* = 3.364; *SD* = .809) ($t_{(20)} = 3.149$, $p = .005$, $d = 1.343$). This was also observed for skill ratings post-training where the social skills feedback report higher *skill* (*M* = 4.091; *SD* = .831) than the control group (*M* = 3.182; *SD* = .874) ($t_{(20)} = 2.500$, $p = .021$, $d = 1.066$).

6.4.1.6. Self- Report Scores vs Journalist Communication Scores

Homogeneity of variance was assumed for pre-training interview ($p = > .05$) and post-training interviews ($p = >.05$) which allowed for a mixed measures ANOVA (session [pre vs post-training], feedback [social signal feedback and traditional feedback group] and rater [self-report vs journalist communication scores]) to be conducted. Tests of normality were all met ($p = > .05$). Results revealed a significant main effect for session ($F_{(1, 40)} = 15.331$, $p = < .001$, $\eta^2 = .277$) and a significant interaction effect between session and rater ($F_{(1, 40)} = 14.789$, $p = < .001$, $\eta^2 = .270$). There was no significant interaction for session x feedback ($F_{(1, 40)} = 1.765$, $p = < .192$, $\eta^2 = .042$) and session x rater x feedback ($F_{(1, 40)} = .179$, $p = .677$, $\eta^2 = .004$).

A follow-up analysis was conducted to further explore the results obtained in the mixed ANOVA. This revealed no significant difference between the trainees (*M* = 4.955; *SD* = .826) and the journalist (*M* = 5.000; *SD* = .844) ratings of communication in the baseline interview ($F_{(1, 43)} = .033$, $p = .858$ $\eta^2 = .007$, $d = 0.052$). Results also revealed a significant difference between trainees (*M* = 4.964; *SD* = 1.010) and the journalist (*M* = 6.009; *SD* = .675) molar ratings in the post-training interview ($F_{(1, 43)} = 16.305$, $p = < .001$, $\eta^2 = 0.280$, $d = 1.217$)

6.4.2. Social Signal Detection

6.4.2.1. PCA Results

6.2.1.1. Preliminary Findings

A PCA was run on the data using a Varimax Rotation. Analysis of the scree plot revealed the second elbow would include 7 Components accounting for 42.85% of the total variance explained (Li & Wang, 2014). Variables retained in the rotation matrices were above .6 (Field, 2013). Component labels and items can be seen in Table 6.12*.*

Table 6.12. PCA components illustrating negative loadings

Component	Signals
Confidence	EmoCogRatio, Stressed (-), Excited, Energy, Imaginative Think (-), Upset (-), Brain Power, Imagination (-) and Uncertain (-)
Disgust	Disgust, Jaw Drop, Upper Lip Raise and Nose Wrinkle
Frowning	Brow Furrow, Lid Tighten, Anger and Sadness
Eagerness	Movement Activity, Volume, Movement Consistency (-), Unsuccessful Interruptions and Posture Activity
Expression Engagement	Engagement, Surprise and Brow Raise
Posed Expression	Dimpler, Lip Stretch, Lip Press and Lip Suck
Posture	Movement and Posture

6.4.2.2. Multivariate Analysis

A multivariate analysis of the 7 extracted Components revealed a significant main effect for session (pre-training vs post-training), $F_{(7, 635)} = 8.358$, $p = < .001$, $\eta^2 = .084$ and a significant main effect for feedback type (social signal feedback vs traditional feedback), $F_{(7, 635)} = 44.492$, $p = < .001$, $\eta^2 = .329$. There was also significant interaction effect for session x feedback type, $F_{(7, 635)} = 4.937$, $p = < .001$, $\eta^2 = .052$. Univariate analyses are conducted in the follow-up sections. Table 6.13 shows descriptive statistics for univariate analysis.

Table 6.13. Descriptives for univariate analysis for each component

Item	Traditional feedback Group		Social Signal Feedback Group	
	Pre-test	**Post-test**	**Pre-test**	**Post-test**
Confidence	.258 (.754)	-.068 (.966)	-.159 (.986)	-.025 (1.178)
Disgust	.567 (1.279)	.076 (.977)	-.168 (.776)	-.433 (.584)
Frowning	-.068 (.483)	-.164 (.389)	.559 (2.030)	-.161 (.305)
Eagerness	-.279 (.610)	-.291 (.676)	.353 (1.043)	.299 (1.315)
Engagement	-.259 (.903)	-.268 (.801)	.368 (1.192)	.247 (.987)
Posed expression	.026 (.721)	.249 (1.081)	-.125 (1.097)	-.206 (.981)
Posture	.036 (.891)	-.133 (.696)	.156 (1.093)	.077 (.821)

6.4.2.2.1. Confidence

A 2 way ANOVA (feedback type [social signal feedback vs traditional feedback] vs session [pre-training interview vs post-training interview]) revealed no significant main effect for session ($F_{(1, 642)} = 1.470$, $p = 226$, $\eta^2 = .002$) and for feedback type ($F_{(1, 642)} = 5.568$, $p = .019$, $\eta^2 = .009$). However, there was a significant interaction between session x feedback type ($F_{(1, 615)} = 8.433$, $p = .004$, $\eta^2 = .013$).

Follow-up between-group one-way ANOVA using a Welch's statistic was conducted to investigate difference between feedback types in the pretraining interview and the post-training interview. Results revealed that there was a significant difference in the pre-training interview between social signal feedback and traditional feedback types ($F_{(1, 231.961)} = 14.891$, $p = < .001$, $\eta^2 = 0.05$, $d = 0.475$) and no differences in groups in the post-test ($F_{(1, 342.956)} = .148$, $p = .701$, $\eta^2 = .00$, $d = .040$).

Follow-up repeated-measures ANOVA was conducted to investigate changes in confidence from pretraining to post-training interview. Results revealed a significant reduction in confidence from pre-training to post-training was observed the control group, ($F_{(1, 339.549)} = 12.265$, $p = .001$, $\eta^2 = 0.033$, $d = 0.376$). Whereas, those who were given social signal feedback had not significantly improved in confidence from the pre-training interview to the post-training interview, ($F_{(1, 293.599)} = 1.156$, $p = .283$, $\eta^2 = 0.004$, $d = 0.123$). Descriptive statistics can be seen in Figure 6.7.

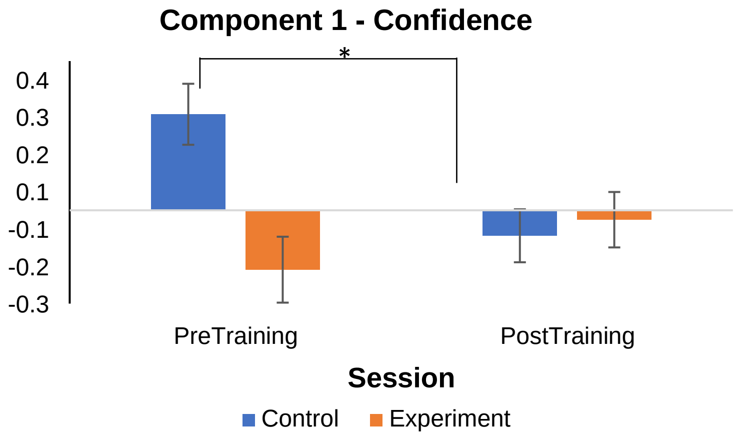


*Figure 6.7. Mean and Standard Error for Confidence. *significance*

6.4.2.2.2. Disgust

A 2 way ANOVA (feedback type [social signal feedback vs traditional feedback] vs session [pre-training interview vs post-training interview]) revealed a significant main effect for session ($F_{(1, 642)} = 25.737$, $p = < .001$, $\eta^2 = .039$) and feedback type ($F_{(1, 642)} = 69.691$, $p = < .001$, $\eta^2 = .098$). However, there was no significant interaction between session x feedback type ($F_{(1, 642)} = 2.333$, $p = .127$, $\eta^2 = .004$).

Follow-up between-group one-way ANOVA using a Welch's statistic was conducted to investigate difference between feedback types in the pretraining interview and the post-training interview. Results revealed a significant difference of performance between feedback types in the pre-training interview

($F_{(1, 245.401)}$ = 34.037, p = < .001, η^2 = 0.105, d = 0.695). In the post-training interview, there was also significant differences between groups ($F_{(1, 321.443)}$ = 37.963, p = < .001, η^2 = 0.089, d = 0.632).

Follow-up repeated-measures ANOVA was conducted to investigate changes in disgust from pretraining to post-training interview. Results revealed that the control group significantly reduced in demonstrations of disgust from the pre to post-training interview ($F_{(1, 264.137)}$ = 15.095, p = < .001, η^2 = 0.046, d = 0.431). Those who had received social signal feedback reduced displays of disgust from pre-training to post training ($F_{(1, 220.177)}$ = 10.438, p = .001, η^2 = 0.037, d = 0.386). Descriptive statistics can be seen in Figure 6.8.

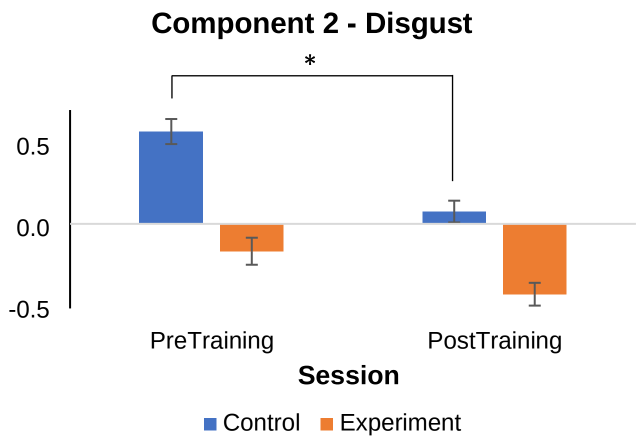


*Figure 6.8. Mean and Standard Errors for displays of disgust. *significance from pre-post training interview*

6.4.2.2.3. Frowning

A 2 way ANOVA (feedback type [social signal feedback vs traditional feedback] vs session [pre-training interview vs post-training interview]) revealed a significant main effect for session ($F_{(1, 642)}$ = 28.158, p = < .001, η^2 = .042) and feedback type ($F_{(1, 642)}$ = 16.780, p = < .001, η^2 = .025). There was also a significant interaction between session x feedback type ($F_{(1, 642)}$ = 16.460, p = <.001, η^2 = .025).

Follow-up between-group one-way ANOVA using a Welch's statistic was conducted to investigate difference between feedback types in the pretraining interview and the post-training interview. Results revealed a difference between feedback types in the pre-training interview ($F_{(1, 137.151)}$ = 11.449, p = .001, η^2 = 0.046, d = 0.425) but none in the post training interview between groups ($F_{(1, 362.912)}$ = .007, p = .933, η^2 = 0.000, d = 0.009).

Follow-up repeated-measures ANOVA was conducted to investigate changes in frowning from pretraining to post-training interview. Results revealed that the control group reduced displays of frowning / anger from the pre-training to the post-training interview ($F_{(1, 274.480)}$ = 3.899, p = .049, η^2 = 0.012, d = 0.219). Those who received social signal feedback displayed a significant reduction in displays of frowning / anger ($F_{(1,\ 129.012)}$ = 15.582, p = <.001, η^2 = 0.067, d = 0.496). Descriptive statistics can be seen in Figure 6.9.

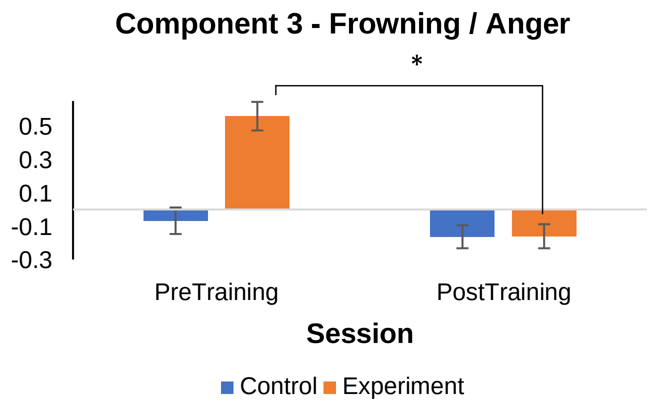


*Figure 6.9. Means and Standard Errors for Displays of frowning / anger. *significance from pre-post training interview*

6.4.2.2.4. Eagerness to speak

A 2 way ANOVA (feedback type [social signal feedback vs traditional feedback] vs session [pre-training interview vs post-training interview]) revealed no significant main effect for session ($F_{(1,\ 642)}$ = .186, p = .666, η^2 = .000), but a significant main effect for feedback type ($F_{(1,\ 642)}$ = 64.358, p = < .001, η^2 = .091). There was also no significant interaction between session x feedback type ($F_{(1,\ 642)}$ = .078, p = .780, η^2 = .000).

Follow-up between-group one-way ANOVA using a Welch's statistic was conducted to investigate difference between feedback types in the pretraining interview and the post-training interview. Results revealed a significant difference between social signal feedback and traditional feedback types in the pre-training interview ($F_{(1,\ 194.775)}$ = 35.837, p = < .001, η^2 = 0.125, d = 0.740) as well as in the post training interview ($F_{(1,\ 259.013)}$ = 28.872, p = < .001, η^2 = 0.076, d =0.564).

Follow-up repeated-measures ANOVA was conducted to investigate changes in eagerness to speak from pretraining to post-training interview. Results revealed that the control group did not significantly differ from the pre-training interview to the post-training interview ($F_{(1,\ 329.150)}$ = .028, p = .868, η^2 = .000, d = 0.019) and neither did those who had received social signal feedback ($F_{(1,\ 298.062)}$ = .160, p = .690, η^2 = .000, d = 0.046). Descriptive statistics can be seen in Figure 6.10.

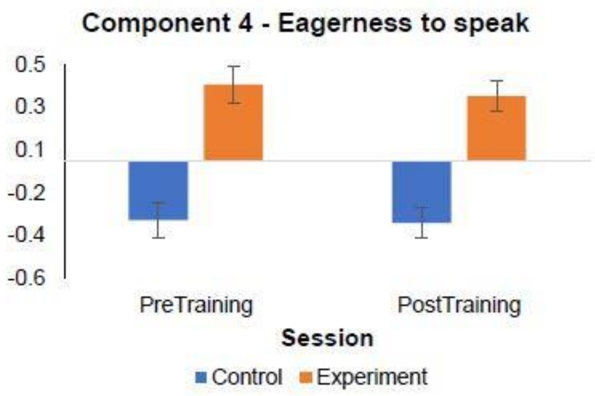


Figure 6.10. Means and standard error of displays of 'Eagerness to Speak'

6.4.2.2.5. Expression Engagement

A 2 way ANOVA (feedback type [social signal feedback vs traditional feedback] vs session [pre-training interview vs post-training interview]) revealed no significant main effect for session ($F_{(1, 642)}$ = .719, p = .397, η^2 = .001) but there was a main effect for feedback type ($F_{(1, 642)}$ = 55.381, p = < .001, η^2 = .079). There was no significant interaction between session x feedback type ($F_{(1, 642)}$ = .544, p = .461, η^2 = .001).

Follow-up between-group one-way ANOVA using a Welch's statistic was conducted to investigate difference between feedback types in the pretraining interview and the post-training interview. Results revealed that there was a significant difference between social signal feedback and traditional feedback types in the pre-training interview ($F_{(1, 230.506)}$ = 23.423, p = < .001, η^2 = 0.083, d = 0.593) as well as in the post training interview ($F_{(1, 341.316)}$ = 30.218, p = < .001, η^2 = 0.077, d = 0.573). Descriptive statistics can be seen in Figure 6.11.

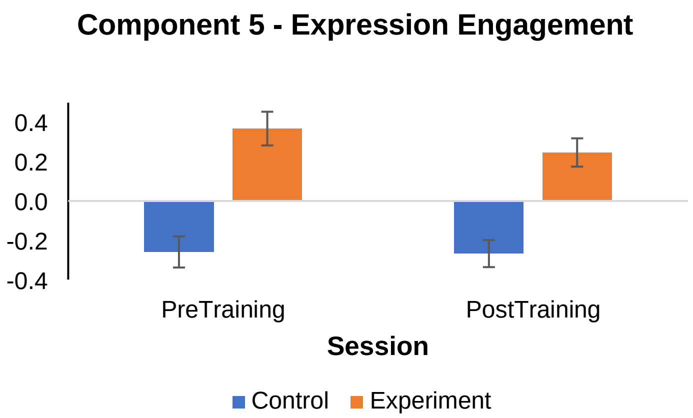


Figure 6.11. Means and Standard errors for 'Expression of Engagement'

6.4.2.2.6. Posed Expression

A 2 way ANOVA (feedback type [social signal feedback vs traditional feedback] vs session [pre-training interview vs post-training interview]) revealed no significant main effect for session ($F_{(1, 642)}$ = .819, p = .366, η^2 = .001), but a significant main effect for feedback type ($F_{(1, 642)}$ = 14.805, p = < .001, η^2 = .023). There was also no significant interaction between session x feedback type ($F_{(1, 642)}$ =3.732, p = .054, η^2 = .000).

Follow-up between-group one-way ANOVA using a Welch's statistic was conducted to investigate difference between feedback types in the pretraining interview and the post-training interview. Results revealed that there was no significant difference between social signal feedback and traditional feedback types in the pre-training interview ($F_{(1, 210.101)}$ = 1.735, p = .189, η^2 = 0.007, d = 0.163) but there was a significant difference observed in the post training interview ($F_{(1, 370.988)}$ = 18.130, p = < .001, η^2 = 0.046, d = 0.441). Descriptive statistics can be seen in Figure 6.12.

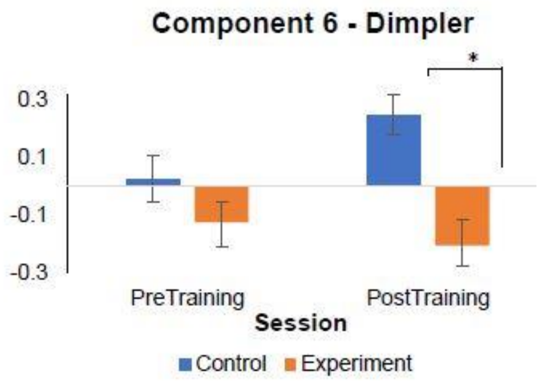


*Figure 6.12. Means and Standard errors for 'Dimpler' displays. *significance from pre-post training interview*

6.4.2.2.7 Posture

A 2 way ANOVA (feedback type [social signal feedback vs traditional feedback] vs session [pre-training interview vs post-training interview]) revealed a no significant main effect for session ($F_{(1, 641)}$ = 3.214, p = .073, η^2 = .005) but a significant main effect for feedback type ($F_{(1, 641)}$ = 5.703, p = .017, η^2 = .009). There was no significant interaction between session x feedback type ($F_{(1, 641)}$ = .415, p = .520, η^2 = .001).

Follow-up between-group one-way ANOVA using a Welch's statistic was conducted to investigate difference between feedback types in the pretraining interview and the post-training interview. Results revealed that there was no significant difference between social signal feedback and traditional feedback types in the pre-training interview ($F_{(1, 241.169)}$ = .970, p = .396, η^2 = 0.004, d = 0.120) but there was a significant difference observed in the post training interview ($F_{(1,\ 348.603)} = 6.976$, $p = .009$, $\eta^2 = 0.019$, $d = 0.276$). Descriptive statistics can be seen in Figure 6.13.

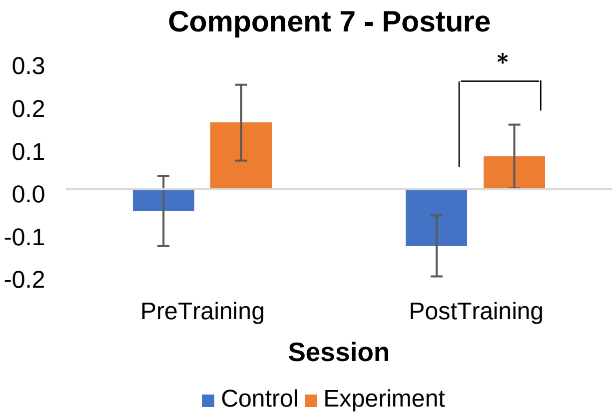


Figure 6.13. Means and standard errors of 'Posture'

6.4.3. Trainees' Thoughts about Feedback Types

This section presents the themes identified about both feedback groups' opinions about the method of feedback they received. This was collected using qualitative interviews conducted following training.

6.4.3.1. Traditional Feedback Group

A thematic map is presented in Figure 6.14 which is followed by a descriptive and interview quoted evidence.

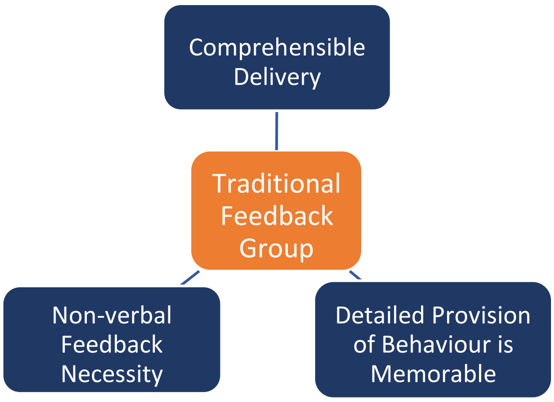


Figure 6.14. Thematic map of themes identified by interviews with those who received traditional feedback

Comprehensible Delivery. Trainees thought feedback was delivered comprehensively in a positive way. However, trainees felt that feedback should have been more detailed.

"Feedback, I felt – I didn't feel like it was delivered in a way that made me question myself" (P33, female)

"Feedback was good but it'd probably been more helpful to be more detailed" (P29, male)

Nonverbal Feedback Adjustable. Trainees were able to adjust their nonverbal signals with the feedback provided by the journalist.

"I change the feedback she gave me. Like my hand gestures, projection of voice." (P19, female)

Detailed Provision of Behaviours Are Memorable. Specific details from interview performance should highlighted for trainees to improve. Trainees thought that it is important for the following training session.

"Well it was good because she pointed out some specific points that I was doing in the interview and that was very helpful to improve for the future." (P26, female)

6.4.3.2. Social Signal Feedback Group

A thematic map is presented in *Figure 6.15* which is followed by a descriptive and interview quoted evidence.

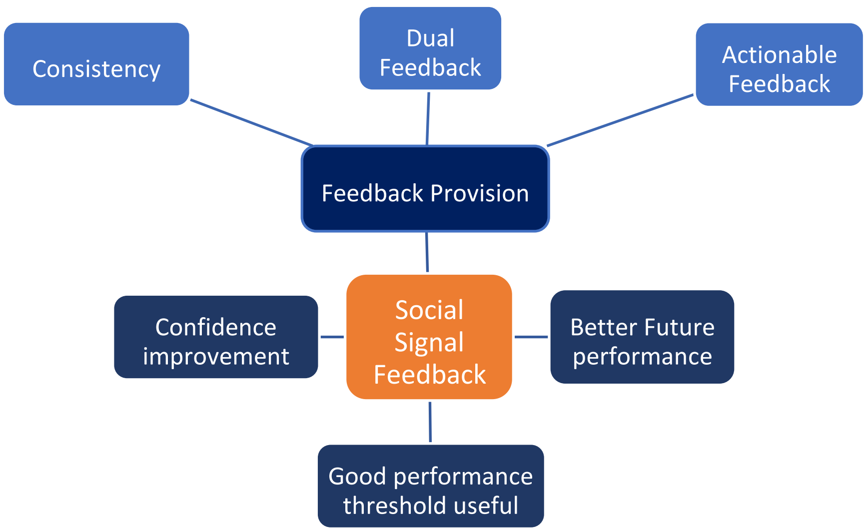


Figure 6.15. Thematic map of themes identified by interviews with those who received social signal feedback

Feedback Provision. Two subthemes were identified: consistency of feedback and dual-feedback helpful which will all be described.

Consistency of Feedback. This theme was identified as trainees stated that the consistency of social signal feedback was valuable for improvement.

> *"I think it was amazing, very fast I mean it takes time to put it but right away you can tell me how [I did]."* (P21, female)

Dual-feedback helpful. Trainees stated that receiving both traditional feedback about understanding of content as well as social signal feedback was useful for improvement. Trainees thought this was useful, as it teased out the specifics of performance by observing the playback of their interview.

> *"It was very consistent because you showed me with the technology and also the journalist gives good feedback, very impressive, very good points mentioned that I had to improve on and also that I did good with [social signal] feedback"* (P21, female)

Actionable feedback. This theme was identified as participants expressed the view that the social signal feedback was actionable. It was also expressed that any real improvements could only be observed over time or later, rather than on the same day. Furthermore, it was expressed that in order for feedback to be even more actionable it is beneficial to observe real physiological examples of the social signals that were fed back to participants.

> *"I definitely thought because the feedback was so specific; I definitely knew what I needed to improve. So, it was very actionable to me" (*P34, female)

> *"I think all the things that you mentioned or feedback to me, were things that I could change"* (P31, female)

> *"There has been some improvement I think, as I was aware of things that I had to change. This improvement you won't see needs some time to be implemented."* (P39, male)

> *"I'd like to know for an example, uh, the range of variability of posture. I would like to see an example; a good posture, a bad posture"* (P39, male)

Better Future Performance. Trainees felt that social signal feedback during training would contribute to their improvement in future interviews not limited to media interviews.

> *"I think it might help me to prepare myself better for the future for my presentation skills and interview skills."* (P25, male)

"I think I will use this information in even like the training videos of just not necessarily do a lot of interviews but can be applied to like presenting"
(P31, female)

Good Performance Threshold Useful. Trainees thought the good performance component of the summative visualisation method clarified their performance. They also felt that the minimum and maximum threshold was a good reference point for behaviour.

"It was useful to know what makes a good interview and then how well I was performing. It was good." (P31, female)

Confidence Improvement. Trainees felt the way social signal feedback was presented improved their confidence overall.

"I think I will actually improve more on confidence level" (P25, male)

"I think it was helpful. The one that is was changing was my confidence level"
(P28, male)

6.5. Discussion

The aim of this study was to investigate whether nonverbal feedback using commercial automated affect recognition technology is more effective in improving communication skills in media interview training than traditional methods of media interview training. The results obtained are in support of H1, H2 and H4. The first hypothesis proposed that there will be significant improvements in performance (as measured subjectively) from pre-test to post-test interview in both training conditions (main effect of training). The second hypothesis predicted that there will be significant changes in observed social signals detected between pre-test and post-test interview in both training conditions (main effect of training). Finally, the fourth hypothesis predicted that there will be greater changes in social signals detected between pre-test and post-test interview for the experimental condition (interaction effect). Subjective ratings of trainee performance and behavioural modifications by feedback are discussed considering previous research.

6.5.1. Subjective Ratings of Performance

Subjective judgements of communication skills provided by journalist and neutral observers (both blind to experimental condition) illustrate a clear improvement in both feedback groups, supporting hypothesis 1. According to journalist ratings, using the conversational skills rating scale, participants who received social signal feedback improved 25% from pre-test to post-test and those who received traditional feedback only improved 15%; however, the interaction effect was not statistically significant. Similarly, neutral observers rated the improvement of social signal feedback participants at 20%, compared to 15% for standard training, but again, the interaction effect was not statistically significant. The current data set does not provide statistical evidence for hypothesis 3. A larger effect size was found for improvements from pre-training to post-training interview for trainees who received social signal training (d = 1.064) than for those who received traditional feedback (d = 0.614).

In contrast to journalist and observer judgements, trainees did not rate themselves as improving from pre-test to post-test when measured by means of the conversation skills rating scale. However, when participants were asked to score how much they had improved directly, it was found that participants who received social signal feedback rated their improvement in skills and confidence as higher post training (on average citing ‘considerable improvement’) compared to participants that did not receive social signal feedback (on average citing ‘some improvement’ according to the scale labels). This provides partial support of hypothesis 3. Research has found that confidence encourages learning and improves performance (Costanzo, 1992; Skinner et al., 2013). These results suggest that standard training is not enough to provide trainees with the confidence to perform well and has been a suggested future work by Liu and colleagues (2007). In sum, these results suggest that social signal feedback improve confidence about performance which may be a result of understanding how performance improves over time by observing summative feedback (Damian et al., 2015). Previous research suggests that communication training increases confidence as well as perception of its effectiveness (Fukui, Ogawa, & Fukui, 2010).

6.5.1.1. Self-report vs Journalist Ratings of Communication Skills

The results revealed no different in ratings of performance between the journalist and trainee in the baseline interview; however, there was a significant difference between the two raters in the posttraining interview where journalist rated the trainees significantly better than trainees had rated themselves overall. The result obtained did not differ by group conditions suggesting that all trainees felt that they had not improved as considerably as the journalist reported. This result is contradictory to the skills and confidence ratings obtained at the end of the training day which was significantly different between groups where the experiment group had reported significantly higher confidence and perceived skill post-training in comparison to the control group. This discrepancy could be because of the nature of the questions in the CSRS as they measure communication effectiveness whereas the other questionnaire directly measures skill and confidence.

In general, the use of self-report measures has been questioned because participants often report the more socially acceptable answers than being truthful and they may lack the ability to be introspective as they are not familiar with this (Rosenman, Tennekoon, & Hill, 2011). As a result of this, researchers question the validity of the scores collected using self-report methods (Podsakoff & Organ, 1986). Individuals are often asked to rate their own performance and research has found that there is often a discrepancy between self-report and those generated by supervisors / employers (Smircich & Chesser, 1981). Where employees often make an attempt at providing scores that match up with the supervisors / employer's perspective (Smircich & Chesser, 1981). However, in contrast it has also been found that employees often report a lower score their job performance compared to scores gathered by their supervisors (Schoorman & Mayer, 2008). The latter was found in this research for the post-training interview.

6.5.1.2. Ceiling Effects

A ceiling effect is research is a measurement limitation which is the highest score on a measurement instrument. If this score is obtained in the pre-test then it decreases the likelihood that the instrument has accurately measured a training effect (Teshima, Xu, Sato, & Sugiyama, 2019). This is often observed with questionnaires. Ceiling effects can be described as when the independent variable does not have an effect the outcome variable or the independent variable is no longer measurable. When a maximum score obtained in the pre-test interview this effect will come into play in the current research.

It is possible that ceiling effects were present where trainees had displayed effective communication skills in the pre-training interview; thereby receiving a high score (such as scores 6 – 7 out of 7 for communication skills). Receiving a high score does not leave much room for improvement which impacts on the results obtained. Considering the relatively small sample size, this could have influenced the results. Research suggests that trainees who perform very poor in the pre-training interview / baseline are more likely to gain skills from training than those who scored high baseline scores (Aspegren, 1999). Future research could control for high scores obtained in the pre-training interview by excluding them.

6.5.2. Social Signals

A PCA was conducted to reduce and merge similar variables to conduct a multivariate analysis (Cheng Li & Wang, 2014). The reason for this was that the total number of variables included in the analysis was large and included multiple emotions across communication channels. Additionally, this was done to identify whether the signals which were fed back to participants were highlighted in the PCA analysis.

6.5.2.1. Confidence

The first component extracted by a PCA were derived from vocal emotion / affect analysis and included scores for 'rationality', 'stressed', 'excited', 'energy', 'logic' (loading positively) and 'imaginative thinking', 'upset', 'imagination' and 'uncertain' (loading negatively). These have been labelled as confidence based on the included signals. The results revealed that those who received traditional feedback had a significant reduction in confidence from pre-training interview to posttraining interview. This finding is partially supportive of hypothesis 4, the control group were not as confident as in the pre-training interview. The experiment group did not change in confidence score from pre to post-test resulting a relatively positive outcome compared to the control group. Trainees rated their skills and confidence higher in those who received social signal feedback than the control group. This finding is similar to what was found in research by Zhao and colleagues (2017).

6.5.2.2. Disgust

The second component extracted was disgust. The features included both EMFACS and facial expression classifiers (AU) resulting from Affdex expression recognition software. This included disgust, jaw drop (AU26 and AU25), upper lip raises (AU10 also shows slight AU25) and nose wrinkle (AU9 that also shows AU4 and AU10). Both groups displayed a reduction in disgust from pre-training to post-training interview which is partially in support of hypothesis 2. This could be a consequence of the AU involved in displays of disgust, which include nose wrinkler, lip corner depressor (AU15) and low lip depressor (AU16 with AU25). Jaw drop could have been captured due to its role in the act of verbal communication during conversation. Disgust is often confused with fear (AU1+AU2+AU4+AU5+AU7+AU20+AU26) (Klieger & Siejak, 1997), while fear could be presented in this context, it may be present in micro expressions. There was a reduction which could suggest that participants became less fearful of the situation from the pre-training interview to the post-training interview. Empirically, disgust has also been confused with anger (AU4+AU5+AU7+AU23) and contempt (AU12+AU14) (Calder & Young, 2005). It has been suggested that the confusion resulting from facial expression of disgust could be a consequence of culture-specific (Calder & Young, 2005).

Overall, the expression of disgust is revealing of a negative emotion suggesting that both groups were aware of their facial expressions which could be a result of watching videos of their interviews.

6.5.2.3. Anger / Frowning

The third component extracted was anger / frowning. The signals included in this component included both facial expressions classifiers (AU) and EMFACS captured by Affdex which included brow furrow (AU4), lid tighten (AU7), anger and sadness (AU1+AU4+AU15). Both groups had significantly reduced displays of frowning in post-training. A follow up analysis revealed that those in the experiment group displayed more frowning than the control group in the pre-training interview suggesting scope for improvement. Furthermore, trainees that received social signal feedback did not frown as much in the post training interview compared to the control groups. Media skills training guides suggest that frowning during an interview is negatively perceived by the audience (Taylor, 2010). From this it can be implied that social signal feedback helped trainees to improve their social skills in how they are perceived which may have been acquired from watching their interviews back for reflection and summative feedback that emphasises a threshold of good and bad performance. This was a key finding as previous research found no significance for the implementation of thresholds during feedback because trainees did not cross the threshold (Damian et al., 2015). However, in this instance feedback was not individualised.

6.5.2.4. Expression Engagement

The fourth component extracted using PCA was engagement. The features included are made up facial expression classifiers (AU) and EMFACS that include engagement, surprise (AU1+AU2+AU5+AU26) and brow raise (AU1). Trainees who received social signal feedback displayed more facial expression engagement in the pre-training interview and in the post-training interview compared to trainees that received standard feedback training. It suggests that social signal feedback provision encouraged trainees to use more facial expressions in their conversations with the journalist. Previous research has shown that engagement is required for effective communication (Naim et al., 2016; Rana el Kaliouby, Evan Kodra, Pankaj Jha, 2014; World Health Organization, 2005).

6.5.2.5. Posed Expressions

The fifth component extracted was '*posed expression'*. This component was derived from facial expression classifiers (AU) captured by Affdex. Included were dimpler (AU14), lip stretcher (AU20), lip press (AU24) and lip suck (AU28 and AU26). Those who received traditional training feedback had displayed more dimpler expression in the post-training interview. The presentation of dimpler has been found to be over exaggerated during posed expressions facial actions (Littlewort, Bartlett, & Lee, 2008), suggesting that expressions shown in the post-training interview were posed. This may signpost boredom or fatigue after training (Taylor, 2015). Similarly, the expression of dimpler is the main AU that makes up contempt which is often confused with disgust which was a PCA component. This may have been noticed by subjective raters and could be an explanation for the insignificant results between groups obtained using subjective ratings of performance.

6.5.2.6. Posture

Signals included in the *posture* component were honest signals resulting from sociometric badges detection of the interaction. Signals included were movement and posture. Those who received standard training feedback displayed a relaxed posture compared to trainees who received social signal feedback in the post-training interview. Research suggests that a relaxed posture is directly linked to attitude and signifies boredom or aloofness (Taylor, 2015; Vinciarelli et al., 2009). This could imply that those who received traditional feedback may have been bored after training session or that they were comfortable enough in the final interview. This is like the finding where those who received traditional feedback displayed more *posed expressions,* suggesting that they were bored. Research has also found that a straightened posture is associated with higher cognitive engagement and an increase in self-awareness (Kaakinen, Ballenghein, Tissier, & Baccino, 2018; Vinciarelli et al., 2009b), similar findings were found in (Muehlhan, Marxen, Landsiedel, Malberg, & Zaunseder, 2014). This suggests that trainees in the experiment group were more attentive and engaged after in the post training interview. This result is also consistent with results found in the exploratory study that those who were labelled as a poor communicator had displayed a significantly more relaxed posture than those who were labelled as effective communicators by three neutral observers and an expert trainer.

6.5.3 Mapping of Components to Summative Feedback

Social signals identified in the exploratory stage were fed back to trainees following their interviews, based on behaviours captured by the sensors. Detailed feedback given to each trainee was not recorded; however, a note was made about common feature feedback. Common feedback presented to participants included information about their postures, movements, the volume at which they spoke in comparison to the journalist, frowning and smiling. When disgust / surprise / sadness was identified, they were told to not use many facial expressions as previous research has shown that this is negatively perceived by the audience (Taylor, 2015).

Feedback given to participants included only the first 30 seconds. However, the PCA analysis included the duration of the interview which included smiling and brow furrow. These features were commonly fed back to participation. The identification of these features suggests that participants had retained the feedback provided during appraisal sessions following mock interviews and could suggest that the feedback was effective and no increase in cognitive load was prevalent.

Signals which were fed back to trainees' post interview were extracted from the PCA analysis. The facial expressions which were included in the PCA analysis that were also given as feedback to trainees included anger, sadness, disgust, surprise and brow furrow. Smiling was recognised as vital in the pre-interview lecture which may suggest why it was not selected for inclusion in the PCA analysis. These results suggest that the feedback provided to trainees was adopted and an attempt was made to utilise this feedback to better their communication skills during interviews.

The honest signals which were extracted from the PCA analysis included movement (related to movement rate), posture (also related to posture activity) and volume. It is possible that movement was extracted by the PCA analysis as a result of providing feedback to participants about their movement rate during performance appraisal. This could be because throughout the interview participants were conscious of their movement. The same could be explained for the extraction of posture where participants were made aware of their posture and recalled this feedback during the interview which could explain the inclusion of posture activity in the results. There was no extraction of any voice recognition labels which were feedback to trainees that were included in the PCA analysis.

6.5.4. Re-test Timing Limitation

In this study, pre-test and post-test interviews were conducted on the same day with training intervention taking place in between. This procedure was adopted to maximise retention of trainees within the University setting of this research as there was a concern that trainees might not return for a second testing session later. The study was therefore only able to assess short term gains due to training resulting in habituation (Hoque, Courgeon, & Martin, 2013; Tanaka, Negoro, Iwasaka, & Nakamura, 2017). There is some evidence that training effects often play out over a longer time and it may therefore be that differential effects of the social signal intervention might be observed if participants had been re-tested after more time had elapsed. Interestingly in the qualitative feedback, several trainees mentioned that they felt they might see the benefit of their social signal training experiences more in the future. This was therefore tested through the follow-up stage (see Chapter 7).

6.5.5. Qualitative Interviews

Themes identified from those who received traditional feedback included *comprehensible delivery; non-verbal feedback necessity and detailed provision of behaviours are memorable.* Overall, participants felt that the traditional feedback provided was comprehensible enough for development throughout their training session. Trainees in this study requested more information about their nonverbal behaviour and felt that it would enhance their performance as these details would help in remembering what to consider in behaviour during an interview. Research has found that the details surrounding an event helps in remembering (Kensinger, 2009). This finding suggests that incorporating trainees' nonverbal signals could be valuable to improve communication in media skills training programme development.

Themes established for trainees who received social signal feedback included a main theme of *feedback provision* with subthemes including *consistency of feedback; dual-feedback beneficial.* Other themes included were *perform better in future; comparison to good performance; actionable feedback;* and *confidence improvement*. Identification of these themes suggest several key findings; 1) that social signal feedback is valuable for boosting trainees' communication skills as it offers more detail; 2) it could have a widespread use for various contexts (i.e. presentations); 3) a combination of social signals feedback and video playback are useful for improvement; 4) it is applicable during training and improves trainees' confidence in their communication ability. The latter could be because trainees were able to track their progress using the summative feedback method as well as a result of the application of the 'good performance' threshold. Furthermore, consistent methods of feedback throughout training sessions helped trainees identify areas of improvement after repeated exposure and facilitated their understanding of their own behaviour in a coherent way. Together, these findings suggest that feedback of social signals is more effective than standard feedback alone as it offers detailed information that is memorable and, as result, makes it actionable.

6.6. Conclusion

The aim of this Chapter was to explore whether technology-enhanced feedback was more effective than standard feedback techniques:

Is the provision of social signal feedback more effective in enhancing communication skills during a person-person discourse compared to standard feedback provision?

Subjective ratings of performance revealed that the journalist (conversational partner) rated both the experiment group and the control group as improving in communication skills. However, a larger effect size was found for trainees in the experimental group. Moreover, an objective audience (neutral observers) rated that the experiment group and the control group has significantly improved in communication skills after training; however, a larger effect size was observed for the experiment group. Trainees in the experiment group rated their improvement in confidence more highly than those in the control group.

Social signals data showed a positive training effect for communication skill training in the specific case of frowning, with both groups showing a reduction. However, a medium effect size was produced for the experiment group (d = 0.496) compared to a small effect size produced for the traditional group (d = 0.219). Mixed methods approach used in this study allowed researchers to gain details surrounding the provision of social signal feedback using a combination of feedback techniques.

Together, these results suggest that the social signal feedback group are more effective communicators after training, as suggested by interpretation of effect sizes.

The results of this study highlight key points when implementing a technology enhanced method of communication skills training. The use of COTS technology in training is effective and is consistent with the literature. It proposes an alternative method to an already practical method, the behavioural feedback loop. Rather than providing real-time feedback, which could be distracting, the technique used in this study provides a summary of the behaviours displayed and improves self-awareness through formative feedback (video playback). This permits discussion of performance with the journalist and reflection of behaviour. Even though there were some limitations there were signs of habituation due to the differences between the two groups in social signal displays.

While these results are promising, interaction effects failed to reach significance. Qualitative comments on training suggested people thought it might take time for them to internalise and apply the feedback given during training. Furthermore, the literature suggests that training gains are often observed over time as this allows trainees to reflect of feedback (Aspegren, 1999). Therefore, the next chapter reports a follow up study conducted six months after initial training to assess training gains.

CHAPTER 7. 6 - MONTH FOLLOW-UP EVALUATION OF SOCIAL SKILLS TRAINING

7.1. Introduction

Findings from Chapter 6 suggest that there was a better improvement in communication skills for trainees who were provided with social signal feedback designed in Chapter 4 and 5 compared to standard training. Results from subjective ratings reported that both groups improved from pre-training interview (baseline skill evaluation) to post-training interviews. However, three neutral observers acting as an audience reported that the social signal feedback group had significantly improved in communication effectiveness. Additionally, those in the experiment group rated higher confidence and better perceived skills post training in comparison to the control group. Social signal feedback group and the traditional feedback group displayed improvement in frowning. However, the social signal feedback group produced a larger effect size. Maintenance of skills is a constituent of effective training so to further assess whether the proposed method was effective, the second part of this chapter investigates whether these skills were maintained by participants after 6 months and whether differences between training groups are more pronounced at six months.

Research Question:

Are there differential training effects for social signal feedback compared to standard feedback when tested after 6 months?

The experimental hypotheses for this research stage are:

1. Subjective ratings of observed interview performance will be higher for the experiment group (automated vs traditional training);
2. There will be significant differences in social signals between groups (automated vs traditional training)

7.2. Data Collection

7.2.1. Participants

Participants who took part in the experiment stage were recalled from the experiment stage using participant recruitment posters. Of the 22 participants trained in Chapter 6, a total of 16 participants (age ranged from 18 – 55 years old; 13 females and 3 males) were included in this follow-up study.

Participants included in this study were from Lithuania (2), Belarus (1), Brazil (2), South Africa (1), Malaysia (1), Iran (2), Sweden (1), Germany (1), Italy (1), UK (1) and other (2). A total of 14 participants were non-native English speakers and 2 participants were native English speakers. The roles that participants had within the university included research students (12), research staff (1) and taught students (3).

7.2.2. Materials and Measures

Subjective Communication Skills Ratings. Briefly, the CSRS was used to assess participants' communication skills performance. Ratings were collected from the participants, the journalist and, later, three neutral observers who were able to pause and play back each video serving as an independent rater.

7.2.3. Social Signal Detection

Non-verbal signals were captured on a PJ500 camera and a Zoom voice recorder. Video recordings were trimmed and imported into iMotions for post-processing of facial expressions and voice recordings were edited where the journalists voice was excluded resulting in voice analysis only including the participants' voice.

7.2.4. Experience since Training Questionnaire

Participants were given a questionnaire to gather information about their experience, confidence and the likelihood of accepting an invitation to give an interview. Information was gathered surrounding public speaking activities and media interviews. This can be seen in Appendix 7.1. Example questions include: 'How many media interviews did you take part in since training?', 'Do you feel capable of taking part in media interviews?' and 'How likely are you to accept an invitation to take part in media interview in as a result of training?'. Participants were also asked about whether they observed any changes in communication and to provide details about these changes.

7.2.5. Procedure

Once participants had arrived, they were be briefed on the study and the recording equipment, formal consent and demographics were collected. Permission was asked if researchers could access participants' data from the previous study and all participants had consented to this. Participants then filled in a questionnaire that assesses their experience since initial training.

Participants then took part in a 7 – 10-minutes media interview with a journalist. After the media interview was completed, participants interviews were played back to them and subsequently asked to fill in a CSRS to rate their own performance. Their performance was also rated by the journalist and later, three neutral observers.

7.3. Data Analysis

7.3.1. Subjective Ratings

7.3.1.1. Self-report of Communication Skills Rating

If the assumptions of a one-way analysis are met, it will be conducted to compare whether those who received social signal feedback had perceived their communication skills to be more effective than those who received traditional feedback group. The independent variable is group (experiment vs control group) and the dependent variable is the CSRS molar scores.

7.3.1.2. Self-awareness of Communication Skills at Different Training Points

Self-report ratings gathered in this study were compared to ratings gathered in pre-training interview and post-training interviews. This was done to get an understanding of how participants perceived their performance pre-training interview (T1), post training interview (T2) and after 6 months (T3). A mixed factorial design was conducted where the independent variables are time (T1, T2 and T3) and the dependent variables are self-report scores obtained using the molar scores of the CSRS.

7.3.1.3. Journalist Communication Skills Ratings of Trainee

Journalist ratings were used as a conversational partner in this context. The ratings gathered in this study (dependent variable) were used to compare communication skills between the control group and the experiment group (independent variables).

7.3.1.4. Experience since Training Questionnaire

The responses collected using the experience questionnaire designed for this research stage will be used to compare experience since training, perceived capabilities, likelihood of interview acceptance and any social signal changes observed since training in the last 6-months. A Mann-Whitney U test was conducted to compare experience, capabilities, likelihood of interview acceptance and social signal changes (dependent variable) between groups (independent variable).

7.3.2. Social Signal Data

Data analysis conducted mimicked that of the previous chapter. Social signal data collected was normalised using the min and max values of the dataset using the following formula:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Subsequently, a PCA was conducted as a means of grouping variables together to reduce the number of variables included in the analysis (Li & Wang, 2014). The number of components selected were based on an observation of the scree plot with a Varimax rotation (Li & Wang, 2014) and were included in a multivariate analysis.

7.4. Results

A total of 16 of the 22 trainees were successfully recruited from the experiment stage. The experiment group included a total of 8 trainees (6 females, 2 males), age ranged from 18 – 45 years old. A total of 7 trainees were non-native English speakers and 1 trainee was a native English speaker. Countries that trainees had originated from included Lithuania (1), Brazil (2), South Africa (1), Malaysia (1), Iran (1), Germany (1) and Italy (1). Trainee roles within the university included research students (5) taught students (2) and research staff (1).

The control group included a total of 8 participants (7 females and 1 male). Countries that trainees were from included Belarus (1), Lithuania (2), Sweden (1), Iran (1), British (1) and preferred not to say (1). A total of 7 participants were non-native English speakers and 1 was a native English speaker. Trainee roles within the university included research students (7) and a taught student (1).

7.4.1. Subjective Ratings of Communication Skills

7.4.1.1. Self-report Communication Skills Rating

According to a Shapiro-Wilke test the data collected and normalised was normally distributed for the traditional feedback group (p = > .05) and for the social signal feedback group (p = > .05) (Shapiro & Wilk, 1965). A Levene's test revealed that the assumption of homogeneity of variance was met (p = > .05) (Gastwirth et al., 2009; Lim & Loh, 1996). As a result, a one-way ANOVA revealed that there was no significant difference in trainee's own perception of communication skills in those who had received traditional feedback (M = 5.375; SD = .803) and those who had received social signal feedback (M = 4.800; SD = 1.031), $F_{(1, 14)}$ = 1.549, p = .23, η^2 = .100.

7.4.1.2. Self-awareness of Communication Skills at Different Training Points

A mixed factorial ANOVA for comparison in self-report rating scores collected at the pre-training interview (T1), post-training interview (T2) and the follow-up study (T3) sessions by treatment group (social signal feedback group vs traditional feedback group). Test of homogeneity of variance was met for T1, T2 and T3 (p = > .05) and for normality (p = > .05). Results revealed no significant main effect for session ($F_{(2, 28)}$ = .362, p = .700, η^2 = .025) and no significant main interaction effect for treatment group ($F_{(2, 28)}$ = .633, p = .538, η^2 = 0.043).

7.4.1.3. Journalist Communication Skills Ratings of Trainee

The assumption of normal distribution was also met for journalist ratings of the traditional feedback group (p = > .05) and for the social signal feedback group (p = > .05). Moreover, the assumption of homogeneity of variance was also met (p = > .05). As the normality and variance assumption was met a one-way ANOVA revealed that journalist rated participants who had received social signal feedback (*M* = 6.000, *SD* = .741) as significantly better communicators than those who had received traditional feedback (*M* = 4.575; *SD* = 1.452), $F_{(1, 14)} = 6.115$, $p = .027$, $\eta^2 = 0.304$.

7.4.1.4. Neutral Observer Communication Skills Rating

The interclass correlation was conducted to assess the amount of rating agreement between the trainer and three neutral observers. The average measure intraclass correlation was .972 with a 95% confidence interval from .936 to .989 ($F_{(15, 30)} = 34.148$, $p = <.001$, $\eta^2 = 0.245$) which suggests high level of agreement of trainee performance.

The data collected was normally distributed for the traditional feedback group ($p = > .05$) and for the social signal group ($p = >.05$). A Levene's test revealed a significance value of $p = > .05$ suggesting that homogeneity of variance is assumed. This allowed for a one-way ANOVA to be conducted which revealed that, on average, neutral observers rated participants who had received social signal feedback (*M* = 6.458; *SD* = .460) as significantly better communicators than those who had received traditional feedback (*M* = 5.333, *SD* = 1.419), $F_{(1,14)} = 4.553$, $p = .051$, $\eta^2 = .245$.

7.4.2. Experience, Awareness, Capabilities and Confidence

7.4.2.1. Experience since Training

Those in the social signal feedback did not take part in any media interviews (8), while those who received standard media skills training took part took part in none (5), only one (2) and three (1). However, this difference was not significantly different, $U = 20.00$, $p =.064$.

Those who received traditional feedback took part in public speaking: none (1), only one (1), two (1), three (2) and more than three (3). Those who received social signal feedback took part in public speaking: none (4), one (1), two (1), three (1) and more than three (1). However, this difference was not significantly different, $U = 16.50$, $p = .094$.

7.4.2.2. Awareness

Participants were asked to express some key elements that they noticed had changed during the 6-months from training. Those in the social signal group reported more awareness during communication settings relating to:

'Posture' (P32 and P26, Female)

'Rolling of the 'eyes' (P20, Female)

Whereas those in the control group expressed more conversational and capability notes which included:

'Confidence' (P32, Female),

‘Thinking about something before I say it’ (P27 and P34; Female and Male)

‘Articulation’ (P33 and P32, Female)

7.4.2.3. Capabilities

Capability responses in media interviews (A) and public speaking (B) are presented in Figure 7.1. A Mann-Whitney U test was conducted which revealed no significant difference between groups in perceived capabilities for media interviews U = 24.00, p = *.333* and for public speaking, U = 29.00, p = 643.

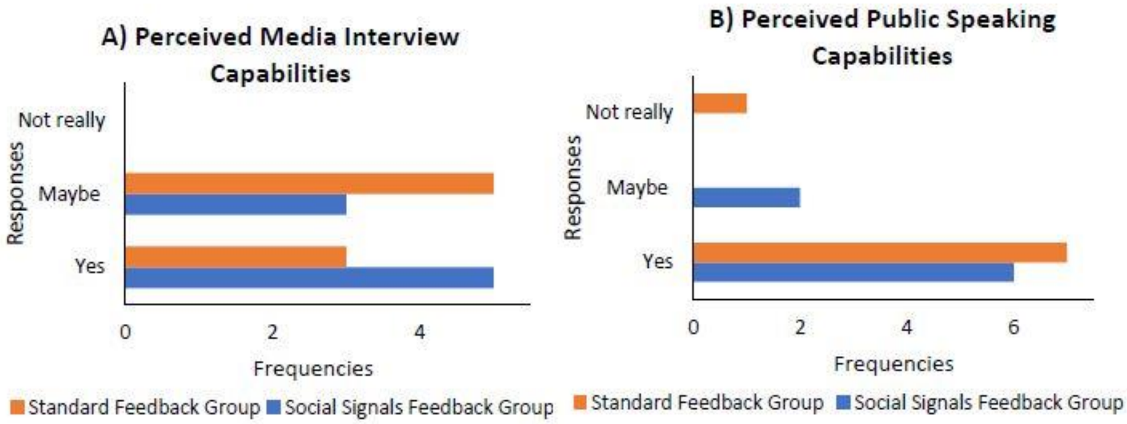


Figure 7.1. Trainees from both feedback groups perceived capabilities of taking part in media interviews (A) and public speaking (B)

7.4.2.4. Confidence

Participants confidence responses to take part in media interviews (A) and public speaking (B) can be seen in Figure 7.2. There was no significant difference in confidence for media interviews, U = 27.00, p = .559 or for public speaking, U = 26.00, p = .515.

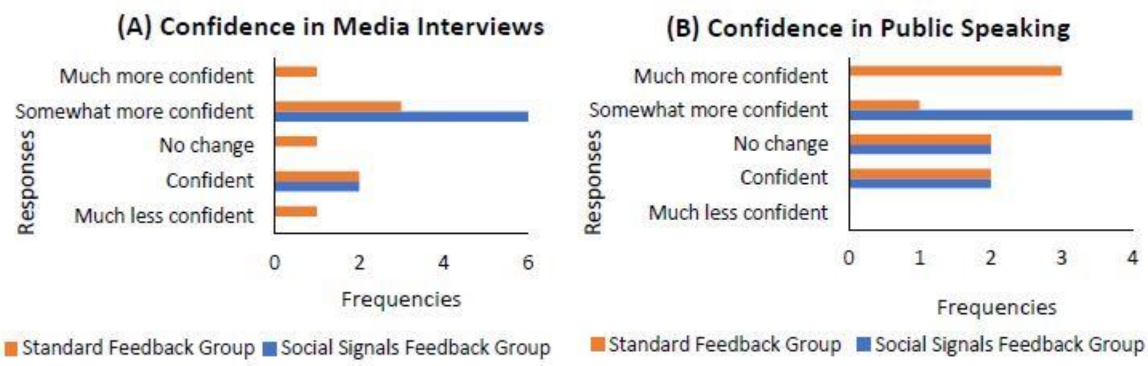


Figure 7.2. Trainees from both feedback group reports for confidence in taking part in media interviews (A) and public speaking (B)

7.4.2.5. Likelihood of Interview Acceptance

A frequency graph is presented in Figure 7.3 demonstrating participants responses for acceptance of an invitation for media interviews (A) and public speaking (B). There was no difference between groups responses for acceptance of invitations for media interviews, *U* = 28.50, *p* = .689 or for acceptance of invitations to public speaking, *U* = 20.00, *p* = .165.

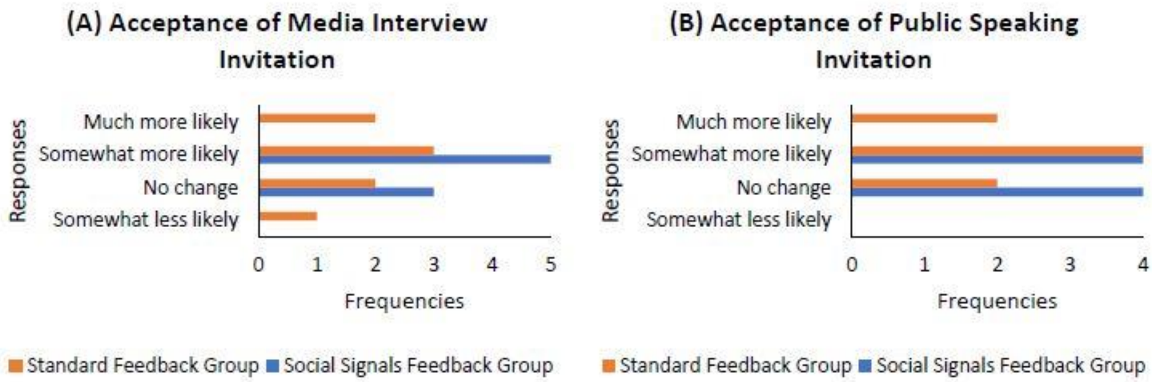


Figure 7.3. Trainees from both feedback group reports for likelihood of interview acceptance of taking part in media interviews (A) and public speaking (B)

7.4.2.6. Changes to Social Signal Since Training

Reports of any observed changes in performance did not differ between those who received social signal feedback (*Mean rank* = 9.75*)* and those who received traditional feedback (*Mean rank* = 7.25*)*, *U* = 22.00, *p* = .218.

7.4.3. Social Signal Detection

Data analysis was conducted in the same way as Chapter 6 for a clearer comparison between the two stages. Interviews ranged from 7 - 10 minutes and were averaged at 10 second intervals producing 30 trials for each participant. Interviews were longer for the follow-up stage than for the follow-up stage to account for potential missing data collected from sensors. The total number of cases produced were 480. Data which was not collected by the technology were considered missing data. Similarly, variables ‘pitch’, ‘voice mirroring’, ‘voice mirror lag’ and ‘speed of turn-taking’ were removed due to too many missing data (Tabacknick & Fidell, 2007). Similarly, ‘content’ and ‘upset’ were removed due to the software producing 0 values for these signals, indicating that participants were not content nor upset. These variables were produced by Nemesysco^{Ltd}. The remaining data (*n* = 387) normalised using the minimum and maximum to the interview [0, 1].

7.4.3.1. PCA Results

Significant Outliers. To reduce the number of outliers and standardise the values produced by each software, variables were normalised according to the mean and maximum of each variable. This would ensure that there were no significant outliers.

Sampling Adequacy. The number of valid instances in the analysis are 387 cases and it is suggested that a sample size of over 300 is adequate (Field, 2013). Additionally, the Kaiser-Meyer Olkin Measure of Sampling Adequacy value is above the suggested 0.5 threshold at a value of 0.595 suggesting the sample size is adequate for conducting a PCA (Field, 2013).

Bartlett's Test of Sphericity. The Bartlett's Test of Sphericity tests the null hypothesis that the variables are unrelated and therefore suitable for structure detection, the value produced for this dataset was p = < .001 which suggests that PCA can be conducted.

Preliminary Findings of PCA. Upon observation of the preliminary findings of a Varimax rotation, a scree plot suggested that a total of 6 components should be extracted which explains a total of 55% variance of the dataset. Variables retained in the rotation matrices were above 0.6 (Field, 2013). These can be seen in Table 7.1.

Table 7.1. PCA components including negative values

Component	Signals
Disgust	Upper Lip Raise, disgust, nose wrinkle, lid tighten, jaw drop and mouth open
Excited /Passionate	Lip press, dimpler, extreme emotion, excited, chin raise, and lip stretch
Eagerness to speak / Enthusiasm	Brow raise, arousal, body movement, posture (-), successful interruptions, surprise and unsuccessful interruptions
Positive Engagement	Smile, joy, intensive thinking, engagement and hesitation (-)
Anger	Volume, volume consistency (-) and brow furrow
Stressed	Stressed, energy (-), emotion / cognitive ratio (-) and brain power (-)

7.4.3.2. ANOVA Results

The number of valid cases for components were n = 387. Additionally, a multivariate analysis was conducted which included all the components extracted and produced by the PCA, but homogeneity of variance was not assumed, and it has been suggested that if this is the case then univariate analysis should be conducted with a stricter alpha level at p = 0.1 (Allen & Bennett, 2007; Tabacknick et al., 2007). A one-way analysis was conducted where the independent variable is feedback group (social signal feedback vs traditional feedback) and the dependent variables were components extracted from the PCA.

7.4.3.2.1. Disgust

While ANOVA is considered robust to violations of this assumptions, skewness can impact the data. The Shapiro-Wilk test was below p = <.001. When this occurs skewness, value should be > ± 3 or < ± -3 (Field, 2013). However, when running an ANOVA on PCA variables it has been suggested that a value of above 0.6 is acceptable (Fabrigar, Wegener, MacCallum and Strahan, 1999). Skewness for component 1, control group is 3.33 and the experiment group is .552. Eyeballing / observation of histograms and Q-Q plots were acceptable. Levene's test of homogeneity of variance was p = .09 suggesting this component meets this assumption. One-way ANVOA revealed that there was no significant difference in displays of disgust between those who had received traditional feedback (M = -.066, SD = 1.126) and those who had received social signal feedback (M = .037, SD = .650), $F_{(1, 384)} = 1.154$, p = .283, η^2 = .003.

7.4.3.2.2. Excited / Passionate

While ANOVA is considered robust to violations of the assumption of normality, skewness can impact the data. The Shapiro-Wilk test was below p = <.001. When this occurs skewness, value should be > ± 3 or < ± -3 (Field, 2013). However, when running an ANOVA on PCA variables it has been suggested that a value of above .6 is acceptable (Fabrigar, Wegener, MacCallum, & Strahan, 1999). Skewness for component 2, control group is 2.17 and the experiment group is 1.376. Eyeballing / observation of histograms and Q-Q plots were acceptable. Levene's test of homogeneity of variance was p = < .001. It has been suggested that if this is severe then the alpha level should be adjusted and interpreted as significant when below p = .01. Welshes test statistic is reported. The social signal feedback group (M = .207, SD = 1.068) had displayed significantly more excitement / passion than the standard feedback group (M = -.180, SD = .901), *Welch's* $F_{(1, 352.072)} = 14.643$, p = <.001, η^2 = 0.038

7.4.3.2.3. Eagerness to speak / Enthusiasm

Skewness for the control group was -.421 and for the experiment group were .717 which is within range. Levene's test of homogeneity of variance was p = < .001 and the significant alpha level will be interpreted as p = .01. Welshes test statistic is reported. The social signal feedback group (M = .468, SD = 1.163) were significantly more eager to speak than those who had received traditional feedback (M = -.407, SD = .583), *Welch's* $F_{(1, 255.069)} = 83.628$, p = <.001, η^2 = 0.191

7.4.3.2.4. Positive Engagement

Skewness for the control group was .439 and for the experiment group were .834 which is within range. Levene's test of homogeneity of variance was p = < .001 and the significant alpha level will be interpreted as p = .01. Welshes test statistic is reported. The results revealed that those who has received traditional (M = .220, SD = 1.071) were significantly more thoughtful and happier than those who had received social signal (M = - .253, SD = .847), *Welch's F* $_{(1,\ 381.\ 655)}$ = 23.497, p = <.001, η^2 = 0.056.

7.4.3.2.5. Anger

Skewness for the control group was .015 and for the experiment group were 1.269 which is within range. Levene's test of homogeneity of variance was p = < .001 and the significant alpha level will be interpreted as p = .01. Welshes test statistic is reported. The results revealed no significant difference between those who had received traditional feedback (M = .006; SD = .567) and those who had received social signal (M = -.007; SD = 1.337), *Welch's F* $_{(1,\ 234.332)}$ = .016, p = .898, η^2 = 0.000.

7.4.3.2.6. Stressed

Skewness for the control group was -.169 and for the experiment group were .287 which is within range. Levene's test of homogeneity of variance was p = .002 and the significant alpha level will be interpreted as p = .01. Welshes test statistic is reported. Results showed that the traditional feedback group (M = .189, SD = 1.091) were significantly more stressed than those who had received social signal (M = -.217, SD = .836), *Welch's F* $_{(1,\ 379.207)}$ = 17.149, p = < .001, η^2 = 0.041.

7.5. Discussion

A hallmark of effective training is the maintenance of skills after some time. The aim of this chapter was to observe if those who had received social signal feedback were more effective communicators compared to those who had received traditional feedback in media interview training conducted 6months prior. The results from this study were consistent with H1 in predicting that those who had received social signal feedback were rated as more effective communicators by the trainer and the average of three neutral observers. The results were partially also consistent with H2 in predicting that those who had received social signal feedback displayed significantly different social signals in media interviews compared to those who had received traditional media skills training.

7.5.1. Subjective Ratings of Performance

There was no significant difference in self-report ratings of communication performance. However, both the trainer and neutral observers rated those that received social signal feedback communicated better in interviews. This is an important finding as neutral observers acted as an objective audience. This is a more reliable finding than the journalist ratings who had also rated the experiment group as more effective communicators which is also a vital finding as there was high agreement among the ratings of neutral observers given to trainees which validates the findings.

Even though those who had received traditional feedback had substantial practice during the 6 months in contrast to the social signal group, the experiment group was rated more effective communicators by neutral observers and the journalist. This suggests that the skills gained in the original training session were long lasting. In addition to this, the findings suggest that participants had time to reflect on their feedback during the 6-months.

7.5.2. Evaluation of Experience, Capability and Confidence

The control group took part in more media and public speaking than the experiment group. However, the difference was not significant. Although this could be because they had received more invitations to take part in interviews, this was not recorded. Future research could account for this.

The results also revealed no difference in confidence to take part in media interviews and public speaking. The result is different to the result produced in the experiment stage as those in the experiment group had reported higher confidence after the post-training interview. As those in the control group had taken part in more media interviews than the experiment group this could have improved their confidence to take part in interviews resulting in better confidence due to practice of skills.

7.5.3. Social Signal Display

7.5.3.1. Disgust

Only facial expressions were included in this component and the AU included were upper lip raise (AU5), disgust, nose wrinkle (AU9, AU4 and AU10), lid tighten (AU7), jaw drop (AU26 and AU25) and mouth open (AU27). Results revealed that there was no difference between groups in displays of disgust. Those in the social signal feedback displayed less disgust; however, this was not significant. This is not consistent with either H1 or H2 hypotheses. Interestingly, this result was also found in the exploratory study suggesting the presence of ‘disgust’ is prevalent in media interviews. However, this could be a false positive due to the software. As mentioned previously, disgust is often confused with fear (Klieger & Siejak, 1997), anger or contempt (Calder & Young, 2005).

7.5.3.2. Excited / Passionate

The signals included in this component were both facial expression and voice affect: lip press (AU 24), dimpler (AU 14), extreme emotion (voice affect), excited (voice affect), chin raise (AU 17) and lip stretcher (AU 20). Results revealed that those who had received social signal feedback displayed significantly more excitement / passion. This is consistent with H1 and H2. Passion has been found to be a compelling feature in public speaking (Whitmarsh & O’Niell, 2010), media interviews (Taylor, 2015) and entrepreneurial pitches (Lucas, Kerrick, Haugen, & Crider, 2016). Research has also found that passion is associated with confidence (Hackman & Craig, 2009) which was identified in the first exploratory stage (Chapter 4) as somewhat better in the experiment group than those who received traditional feedback during training.

7.5.3.3. Eagerness to Speak / Enthusiasm

The signals included in the component labelled ‘eagerness to speak’ included facial expression, honest signals and vocal affect: brow raise, arousal, body movement, relaxed posture, successful interruptions, surprise and unsuccessful interruptions. The results were consistent with H1 and H2 and revealed that those who had received social signal feedback were significantly more eager to speak. ‘Eagerness to speak’ or enthusiasm in interviews is similar to ‘passionate’ in that it is the expression of interest. If an interviewee seems eager to talk about their research, it may be positively perceived by the receivers of the information. This is the case in job interview (Cuddy, Wilmuth, Yap, & Carney, 2015) and teaching (Thomson, Turner, & Nietfeld, 2012) and, in this instance, can be extended to media interviews.

7.5.3.4. Positive Engagement

The signals included in this component included those from facial expression and vocal affect detection: smile, joy, intensive thinking (vocal affect) engagement (facial expression) and less hesitation (vocal affect). A between-subjects ANOVA revealed that those who had received traditional / standard media skills training were significantly more thoughtful and happier than those who received social signal feedback. This is consistent with H2 in that there was a significant difference between the two groups but not higher for the experiment group (H1). Engagement was also captured during the first stage of evaluation; however, the addition of ‘smiling’ was included, hence the ‘positive’ engagement label. This is an important finding as this validates the training that the control group received as a good comparison measure.

7.5.3.5. Anger

The signals included in this analysis include facial expression and honest signals: volume, less volume consistency and brow furrow. The results found in this stage were neither consistent with H1 or H2 as there was no significant difference in displays of anger between groups. This could be because it may be possible that trainees in both groups had reflected on their performance after 6months. It is also possible that each group had more practice during the 6-months following training. In the first evaluation stage, those who received social signal feedback had exhibited significantly less anger / frowning during interviews. Again, the experience that the control group gained during the 6month gap between studies could be an explanation for this result.

7.5.3.6. Stressed

The social signals included in this component included only vocal affect detection: stressed, less energy, less emotion / cognitive ratio (logic) and less brain power. The results obtained were consistent with H1 and H2 which showed that those who received social signal training were significantly less stressed during interview. Stress has been found to be one of the biggest barriers to successful communication as this affects your use of nonverbal signals as a result of losing control of your emotions (Keeley-Dyreson, Bailey, & Burgoon, 1991).

7.6. Conclusion

The aim of this chapter was to investigate whether the skills obtained in the first training session (Chapter 6) was maintained after 6-months:

Are there differential training effects for social signal feedback compared to standard feedback when tested after 6 months?

The results suggest that providing social signal feedback during media interview training with the use of a comparison of 'good performance' was effective in skill maintenance after 6-months. This can be inferred because the social signal feedback group displayed better signals during interviews that are rated as important for media interviews in media training guides. This result was demonstrated over and above the control group even though that group had been exposed to more instances to practice the skills from the standard feedback framework. This positive result was further confirmed by the audience rating the experiment group as more effective communicators than the traditional feedback group. Together, these results suggest that social signal feedback is more effective than standard media skills training and could be a valuable addition in training interventions.

CHAPTER 8. LIMITATIONS AND FUTURE RESEARCH

8.1. Introduction

The results of this PhD research suggest provisional support for the value of social skills training using automated technology. This research also suggests that social skills training was effective in the long-term as suggestive from the results obtained from chapter 7. However, while the results suggest differences between the two groups (see chapter 6, section 6.4 and chapter 7, section 7.4), there are some limitations to take into consideration which suggest that the results should be interpreted with caution. This chapter discusses the limitations of this research and recommends future research in this domain.

8.2. Sample Size

8.2.1. *Exploratory Stage*

The sample size was small (n =17), particularly for the methods of data analysis. An increase in sample size in future work would be helpful to test the reliability of the findings described. In addition to this, a larger sample size would enable the investigation of gender differences and to assess whether there are any cultural differences in performance. Another implication of the results found are that the dataset was unbalanced and did not contain an equal number of poor and effective communicators. However, this was accounted for by conducting a bootstrapping analysis of the data with 100 iterations. This issue imbalance was accounted for by conducting a bootstrapping analysis of the data with 100 iterations. Real data should, wherever possible, be utilised in an analysis as this is more representative of the real world than bootstrapping with a small sample size, since bootstrapping will underestimate the variability in the data sample. As such, the strength of this finding is questionable with respect to its generalisability to a larger population.

To address this further, a post hoc analysis that included all the data was conducted and is reported in Appendix 4.2, this supported the relevance of a number of signals subsequently used in feedback.

Additionally, the sample was all early career researchers. This research could have included highly trained / exceptionally good communications in scope of this study to potentially get more differentiation between good and poor communication skills. This would have made the data more ecologically valid for training a wider audience.

8.2.2. Design Stage

This stage could include a larger sample size (n = 5). While this was the suggested sample size for usability studies, it would have been interesting to include more interviews and ratings. However, it is important to note that the findings of the interview results from the design stage concurred with those of the interviews conducted following the experimental evaluation. This suggests the results of the design stage collected with the small sample size suggested by Virzi (1992). However, a small sample size of only five should be interpreted with caution and it may be that the finding will be limited to the sample recruited in this user study

8.2.3. Experiment Stage

It is noted that the experiment study also includes a small sample size (*n* = 22). Although, this was supported by power calculations to demonstrate a training effect. It may have been better to use a Cohen's D effect size to undertake the power calculation rather than a nonparametric effect size. Given the approach taken, the sample size produced by the power calculation is open to question. However, some trainees showed a high degree of communication skill in their pre-test interviews (e.g. scores of 6 or 7 out of 7 for skill), which may have introduced ceiling effects into the study which did not have much scope for improvement, reducing the chances of demonstrating significant differences between conditions. Research has found that trainees with the lowest pre-course scores gain the most from communication skills training (Aspegren, 1999). It is therefore recommended that further studies are conducted to increase the sample size for this work to provide a stronger test of the interaction effect. Screening could also be introduced to select only those with clear scope to improve their communication skills.

Additionally, there may have been a possibility of type II errors as a result of using lower values of an alpha value in this study. A type II error is committed when researchers incorrectly accept the null hypothesis. However, this was suggested by Allen and Bennet (2007) as a method of accounting for violations of homogeneity of variance. As a result, there may be effects that were not found.

8.2.4. Follow-up Stage

The sample size included for this analysis was also small (*n* = 16). However, the number of cases included for the analysis (ten second intervals) was enough for analysis (see section 6.3.2.3). It would have been more representative of the larger population to include a larger sample size from the initial pool of participants which would in turn increase the return of participants in this stage (Prajapati et al., 2010). Claims made from the follow-up stage should be interpreted with caution, as the strength of these claims are reliant on the results from a small sample size limiting their confidence in them in terms of their generalisability.

8.3. Affective Feedback

The feedback design used in this study was based on the summative and formative method also used in (Damian, et al., 2015). A possible limitation for the emotional and social signals feedback design is that the data used to characterise 'good' performance was taken from a sample of early career trainees, so while the findings are representative of good performance within this cohort, there may be better exemplars on which to base the feedback method. Thus, future work could aim to collect emotional and social signals data from expert communicators and use this to formulate relevant feedback displays. It is also possible that the sandwich feedback method used in this study, might not have been the most beneficial as the negative behaviour that is observed in the interview is diluted. In a sense, this method detracts the reinforcement value of the positive comments and diminishes the corrective value of punishing consequences. An alternative to this method is to deliver it in a straightforward manner (Gigante, Dell, & Sharkey, 2011).

An understanding of communication skills training was obtained from the baseline to the post-training interview in the experiment study (section 6.4). However, an understanding of how skills had developed throughout the training was not assessed. To decouple the effects of training it would have been interesting to obtain neutral observer and trainer ratings to understand the skills gained throughout the training session.

Another limitation of this research is that the method of feedback provided in this research does not allow for testing whether you are required to provide feedback on all social signals or not. Future research could vary feedback modalities based on voice only, face only, face and voice only and so on.

The positive opinions gained from participants about the feedback obtained could have been a result of the novelty of the feedback. It would have been interesting to compare this to an additional technology enhanced method of training to more effectively evaluate the framework proposed in this thesis.

8.3.1. Hand Movements

Detection of gestures was collected from the non-dominant hand which may not be a true representation of hand gestures. However, it was important to obtain hand gestures in this manner so as not to compromise the quality of the physiological recordings as too much movement can impact the quality of the data recorded. Systems cannot recognise the difference between iconic gestures, emphasis gestures or waving hands without purpose or posed movements compared to a mistake. This issue can be seen in other hand gesture technology (Schneider et al., 2015). Future research could explore the use of non-contact detection of gestures to avoid this issue such as Microsoft Kinect which would have captured gestures in combination with other body movements. The results obtained from the Shimmer device only suggest hand movements and can therefore only be interpreted as movements.

8.3.2. Multichannel Communication

While this study included multiple emotional and social signals channels, the analysis reported here focusses on the data from each off-the-shelf technology separately. This allows combinations of signals to be examined, but only within the context of the signals detected by each tool. There is emerging evidence of the added value of combining signals across multiple modalities to improved classification accuracy (e.g. Turk, 2014; Pantic et al., 2005). The approach described in this paper facilitates understanding of the value and the insights that can be gained from each tool separately and as such is most relevant to providing feedback to interviewees, since they would need to know which signals from each tool to focus on improving performance. However, for potential applications where the focus is more on general classification of good and poor media performance, there could be value from an analysis which combines all signals, and future work could address this.

8.3.3. Cross Interview Evaluation

A possible limitation for this research is that the first 30 seconds were evaluated only in the exploratory study. Details surrounding trainee's communication performance could have improved or worsened over the course of the interview which was not included in this analysis. This could have affected the social signals captured for feedback in the experiment stage.

However, research has shown that judgements or impressions of performance are decided in the first stages of the observation and performance throughout the remainder of the interview are treated as confirmation of initial judgements made (Decoster, 2004; Sullivan, 2018). Future research could investigate the whole interaction instead of the first 30 seconds to draw out social signals to be fed back during social skills training.

8.4. Research Setting

8.4.1. User Population

The study looked only at one population composed of early career researchers within a university setting. While to some extent this population will be representative of the kind of professional role where employees may be called upon to engage in media interviews, it would be interesting to confirm the findings for trainees in other organisation types. In addition, none of the trainees were expert at media skills which could have restricted the range of performance. It would be interesting in future work to include expert as well as novice participants. However, the findings are relevant for the training context where trainees are usually not already experts and the findings are not generalisable beyond this context. Participants also received different questions from one another given their own research background. While this increased the ecological validity of the study, it reduced the degree of experimenter control over stimuli and may have led to differences in difficulty and/or emotional impact across different participants. Future work could potentially look to explore the use of more standardised question sets.

8.4.2. Study setting

It was necessary to use several different rooms to collect the data reported across the study in both the exploratory and experiment stage. While the researcher was careful to ensure that the room did not co-vary with condition (so as not to confound the results), the different lighting conditions could have influenced the detection of social signals that may have caused variation amongst participants' performance. In the real world the equipment will need to be robust to different conditions and it may be useful for future work to explore this issue directly through examining environmental effects.

Several trainees mentioned that they felt they would be able to use what they had gained from the social signal training intervention in other communication settings, for example delivering a presentation. It would be interesting to explore in future work whether transfer effects can be observed to other communication tasks.

8.4.3 Journalist

In this research the trainee and the trainer were not matched by culture. Researchers were unable to investigate this due to resource limitations. Different trainers / journalists across studies could have produced an array of different results. Whereas a single journalist / trainer would have produced ratings which are consistent across all studies. Future research could investigate whether matching the culture of the trainer with the culture of the trainee.

A limitation regarding this research is that the journalists recruited to conduct interviews in the experiment stage (Chapter 6) and the follow-up stage (Chapter 7) were not as skilled as those in the exploratory stage (Chapter 4). For consistency, future research could ensure that the same journalists are recruited are the same throughout the research stages. This would control for any extraneous variables associated with cultural, personality and skilled differences.

8.5. Summary of Recommended Research Based on Limitations

The present study was deliberately exploratory in nature as a necessary first step to validate the extent to which theories of emotion and social signals can be applied to training in communication skills (Chapter 4). The study looked only at one population composed of early career researchers within a university setting. While to some extent this population is representative of the kind of professional role where employees may be called upon to engage in media interviews, it would be interesting to confirm the findings for trainees in other organisation types.

In addition, none of the trainees were expert at media skills which could have restricted the range of performance. It would be interesting in future work to include expert as well as novice participants. However, the findings are relevant for the training context investigated in this research.

This research only investigated one population which was composed of early career and staff researchers within a university setting. Even though this population may be representative of a professional role, it would be interesting to confirm these findings with trainees in other types of organisations.

8.6. Future Research in Social Signal Processing Research

Future research could investigate whether the social signal feedback method of training can improve communication skills in the real world as research has found that there are clear differences between a controlled environment (lab setting) and in the wild where there is no control over extraneous variables (Dupré et al., 2018; Gunes et al., 2008). This would further validate the proposed training framework.

It would be interesting to investigate gender and cultural differences. Even though these were accounted for in the experiment by grouping participants together by gender, first language and experience, it would be interesting to explicitly investigate differences. Previous research has found that there is a difference in communication style observed in gender and culture (Gifford, 2012) and this is yet to be investigated in a media interview situation.

Future research could investigate whether real-time feedback would be more effective in improving communication skills in comparison to the summative and formative method presented in this research. While it has been shown that real-time feedback can be distracting, it would be interesting to include haptic feedback in the form of vibration as this has shown to be less distracting (Schneider et al., 2015). In addition to this, a post-summary feedback could also be generated which could include number of live feedback prompts, how long they performed well for and how long it took participants to adjust their behaviour as these are also suggestive of learning (Ali & Hoque, 2017). This is something which could be implemented in a person-agent interaction as this is not possible in a face-face interaction.

To eliminate the novelty of training that produces positive results during opinion gathering by participants, it would be interesting to implement this social signal training framework for a media skills training situation with a virtual conversational training avatar that provides real-time feedback. To avoid the novelty of this type of training technique, it could be interesting to compare this training technique with the previously proposed future research that includes real-time haptic feedback and compare both these training methods to a standard media skills training method as described in this thesis. Eye contact could also be captured as this channel of communication has been found to be important for effective conversation (Lapidot-Lefler & Barak, 2012). Additionally, trainees' performance would be evaluated by several neutral observers after training to assess their improvement. Researchers could recruit researchers in their first year with no previous experience in public speaking with a possible conference coming up. This would be beneficial as there would be a need for training.

Future work could improve this training by implementing a longer timeframe between practice interviews, excluding trainees that initially displayed good levels of communication skills performance and providing a concrete behavioural threshold for what is classified as a ‘good performance’.

Future experiments could decouple the effects of this automated feedback to gain new knowledge surrounding feedback generation beyond the scope of communication skills training. In other words, future research should better understand the development of the communication skills obtained to decide on a sufficient amount of training using the developed method of training. Furthermore, research has only considered the context of media skills, future work could look at other contexts where communication skills are important, e.g. medical training, negotiation skills training and teaching.

Furthermore, future research could conduct an experimental investigation of whether the bar chart is more effective to feedback specific signals than others.

This chapter considered limitations of the current research suggesting that the results and interpretation of the data should be treated with caution. The next chapter provides an overall conclusion for the research.

CHAPTER 9. CONCLUSION

9.1. Introduction

The aim of this PhD thesis was to explore whether training augmented with social skills detection could improve the impact of communication skills training. It included four research stages each with an aim to design an appropriate feedback that will allow trainees to benefit from a potential faster method of training that is more objective.

9.2. Research Questions

The research questions which this research attempted to address were:

1) *Can recently developed automated technology be used to evaluate communication effectiveness?*
2) *Can this be used to provide feedback that helps people improve their communication?*

Findings from the research stages positively addresses both research questions posed for this research; thereby suggesting that the use of commercial technology can be used to evaluate training effectiveness and could potentially be used to provide feedback that helps trainees improve their communication skills. However, there were a few concerns discussed in Chapter 8 which raised issues in relation to the impact of these findings and how they should be interpreted.

Table 9.1 demonstrates the different signals which were noted as important for media interviews with a small sample size. As the sample size were small, the literature was consulted to validate the signals captured in stage 1 and stages 3 and 4. Furthermore, the signals labelled in stages 3 and 4 were not validated but only noted based on the signals in each of the PCA components.

Table 9.1. Evidence for signals identified for media interviews

Stage	Common Social Signals Identified	Social Signal Change	Literature Evidence
1	Hesitation	--	(Skinner, et al., 2013)
	Extreme emotion		(Morgan, 2008)
	Posture		(Pentland & Heibeck, 2010)
	Movement rate / activity		(Gross & Levenson, 1997; Taylor, 2015)
	Smile		(Morgan, 2008; Taylor, 2015)
	Smirk		(Taylor, 2015)
	Sadness		(Morgan, 2008; Taylor, 2015)
	Joy		(Morgan, 2008; Taylor, 2015)
	Fear		(Morgan, 2008; Taylor, 2015)
	Contempt		(Morgan, 2008; Taylor, 2015)
	Brow furrow		(Morgan, 2008; Taylor, 2015)

3	• Confidence • Disgust • Frowning • Eagerness • Expression engagement • Dimpler • Posture	Frowning	(Taylor, 2015)
4	• Disgust • Excited / passionate • Eagerness to speak • Positive Engagement • Anger • Stressed	More excitement	(Taylor, 2015; Whitmarsh & O'Niell, 2010)
		More eager to speak	(Cuddy et al., 2015)
		Less stressed	(Keeley-Dyreson et al., 1991)

The aim of the second stage (Chapter 5) was to develop a method of visualising the social signal feedback. Based on mixed methods, this study suggested that trainees favoured colours, a comparison element or threshold for good and bad performances, a video playback for reflection and consistency in visualisation across communication channels (see section 5.6). However, future research could include a larger sample size and potentially include a real-time feedback option to evaluate both methods of providing feedback. These results are limited to the pool of participants recruited and it is questionable how generalisable these results are.

The aim of the third stage (Chapter 6) was to evaluate the training framework developed by comparing it to standard media skills training. Trainees' performance ratings suggested that the experiment group substantially improved in performance over the standard training method. However, this was based on a small sample size and the signals fed back to participants were based on a small sample size. Even though the signals were validated by searching the literature, they should still be interpreted with caution as they may not be specific to media skills training in this context.

Furthermore, an objective audience had also rated the social signal group as improving from pre-training to post-training whereas this was not the case for the standard feedback group. The results suggest that the social signal group had revealed more confidence in their signal displays (section 6.4.2.2.1), frowned less (section 6.4.2.2.3) and had a more natural facial expression after training (section 6.4.2.2.6). The results also suggested that the social signal feedback group also showed higher ratings of confidence post-training compared to the standard feedback group (section 6.4.2.2.1). However, the component labels were not validated and are therefore open to question.

Interview data suggest that trainees rated the method of enhanced training as potentially enhancing their future performances, they felt that the good vs bad performance threshold as helpful to change their cues in subsequent practice interviews, the consistency of feedback was important, and they felt that dual feedback of the post-summary feedback combined with video playback was effective for improvement (section 6.4.3.1 and 6.4.3.2). Even though there were signs of habituation, the results are consistent with previous literature on self-reflection and improved awareness in communication and conversation performance. Future research could enhance this training method by staggering out the practice interviews over a few weeks which will encourage more reflection about feedback.

After 6-months, subjective ratings obtained by the trainer and neutral observers suggest that trainees who had received social signal feedback were rated as better communicators by a journalist who was blind to their assigned feedback method and an objective audience (section 7.4.1). However, these results are based on a small sample size and should therefore be interpreted with caution.

In addition to this, the social signal training group had displayed more eagerness to talk and passion during their interview (section 7.4.3). In contrast, the results suggest that the standard training feedback had exhibited more positive engagement during interviews (section 7.4.3.2.4), this could have been a result of either the method of training demonstrating a good comparison of training methods as it was also effective in training or it could be because these trainees had been exposed to multiple public speaking and media interviews in the 6-months since training. Furthermore, the social signal labels of the PCA analysis were not validated and could suggest another feature. Future work should evaluate skill development at 3 months to get a better understanding of how skills develop.

Findings from all research stages positively addresses both research questions posed for this research; thereby suggesting that the use of commercial technology could be used to evaluate training effectiveness and could be used to provide feedback that helps trainees improve their communication skills. However, the results should be interpreted with caution.

9.3. Research Contributions

Overall, the contributions which were made for this research in Chapter 1 could be applied to practice and academic domains. Each sub-section addresses each contribution in academic and practice domains.

9.3.1. Empirical Contributions (Academic)

There is an improved understanding of how signals are detected by current COTS which map onto human judgements of communication skills

This was identified in Chapter 4 as the signals identified were matched with good and bad media interview performance labels as gathered by a weighted average of communication skill scores ratings collected by neutral observers and the trainer (section 4.2.5.2). The accuracies obtained were above chance and were high in the preliminary data (section 4.4.2) as well as a larger sample size (section 4.4.3).

An understanding of the short and longer-term impacts of training augmentation by social skills feedback

The short-term effects of training are observed in Chapter 6 as results suggest that neutral observer ratings had rated the social signal group and the control group as improving in communication skills from pre-training interview and the post training interview (section 6.4.1). The effect size was larger for those who received social signal feedback (section 6.4.1). The long-term impact of training augmentation by social skills feedback are detailed in Chapter 7. The results from Chapter 7 suggested that trainees who received social signal feedback were more effective communicators, as rated by neutral observers and journalist ratings (section 7.4.2). Furthermore, the social signal group could have displayed more positive signals during interviews as a result of social signal data obtained (section 7.4.3)

Understanding the potential effectiveness of social skills training interventions through experimental evaluation

The automated detection of social signals was identified as potentially effective using experimental methods (Chapter 7). The training intervention was compared to traditional methods of providing feedback at two different time points. This comparison resulted in identifying that augmenting training using automated recognition systems as a potential alternative to traditional methods of training when both groups have been exposed to the same training methods.

Automated detection and feedback of social signals in media interviews in this PhD research suggests improved communication effectiveness in the experiment stage (Chapter 6). It also suggests how a conversational partner (journalist) and an acting audience rate the interviewee (Chapter 7). Experimental methods in this instance enabled the researcher to control variables (gender, language and baseline scores).

9.3.2. Methodological Contribution (Academic and Practical)

A different approach to analysing data

This PhD research has suggested that rather than annotating video and audio files (which is time consuming and does not scale with large datasets) and using bespoke developed systems that use of commercial automated systems which produce prompt results and proof of concept for capturing signals in a media interview context could be useful (Chapter 5).

9.3.3. Artefact Contributions (Practical and Academic)

User-centred development of a training intervention based on detection of social signals through COTS.

This PhD research developed a method that could provide social signal training using COTS technology. The feedback provided to trainees / interviewees were based on the COTS data outputs. Feedback was produced in the form of a bar chart style template which contained threshold for effective communication. This threshold was defined by the data collected in the exploratory stage (Chapter 4). In other words, this threshold could be described as a *comparison to good performance.* Feedback provided to trainees / interviewers were selected based on the signal that required the most improvement.

9.3.4. Practical Contributions

To make recommendations for training practice

The results from this research suggest that augmenting communication skills training using automated recognition technology is possible. However, this method of training has its limitations of data collection that future research should be aware of; missing data, selective feedback so as not to increase cognitive load, training and practice interviews should be spread over a few days to enable trainees to absorb the feedback provided.

9.4. Summary

This research suggests that hybrid systems that combine the strengths of human judgements and computer feedback and can outperform either alone. The social signal training framework proposed in this thesis may encompass some of these strengths where feedback is more powerful than when generated from a machine or a human alone. However, a larger data set is needed to confidently state this which will allow for a more accurate account for frame-frame feedback to make every second count in media interviews as they are generally short in duration. As a result, humans alone could not evaluate communication skills, and neither could computers alone. The results from this research should be interpreted with caution owing to a number of considerations, noted throughout chapter 8 and 9. Therefore, this research proposes a possible solution to the barriers faced in communication skills training by companies in the UK to reduce the costs of training, improve the delivery of training, reduce the duration of training and improve access to training.

REFERENCES

Abernethy, M. (2010). Data mining with WEKA, Part 3: Nearest Neighbor and server-side library. *IBM DeveloperWorks*. http: // www.ibm.com/developerworkds/library/os-weka2/os-weka2pdf.pdf. Accessed 04 April 2017.

Adams, R. B., & Kveraga, K. (2015). Social Vision: Functional Forecasting and the Integration of Compound Social Cues. *Review of Philosophy and Psychology*, *6*(4), 591–610.

Adler, R., Rodman, G., & Kramer, E. M. (1991). *Understanding Human Communication*. Oxford University Press.

Ajmera, J., McCowan, I., & Bourlard, H. (2003). Speech/music segmentation using entropy and dynamism features in a HMM classification framework. *Speech Communication*, *40*(3), 351– 363.

Ali, M. R., & Hoque, E. (2017). Social skills training with virtual assistant and real-time feedback, 325–329.

Ali, M. R., Sen, T., Crasta, D., Nguyen, V. D., Rogge, R., & Hoque, M. E. (2018). The what, when, and why of facial expressions: An objective analysis of conversational skills in speed-dating videos. In *Proceedings - 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018* (pp. 203–209). IEEE.

Allen, M. (2017). Random Assignment of Participants. In *The SAGE Encyclopedia of Communication Research Methods*. 2455 Teller Road, Thousand Oaks California 91320: SAGE Publications, Inc.

Allen, P., & Bennett, K. (2007). *SPSS for the Health and Behavioural Sciences*. Cengage Learning Australia.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior. *Psychological Bulletin*, *111*(2), 256–274.

Amos, B., Ludwiczuk, B., & Satyanarayanan, M. (2016). OpenFace: A general-purpose face recognition library with mobile applications. *CMU School of Computer Science*, *6*.

Anderson, A., Hsiao, T., & Metsis, V. (2017). Classification of Emotional Arousal During Multimedia Exposure. *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments - PETRA ’17*, 181–184.

Argyle, M. (2013). *Bodily Communication* (2nd Editio). London: Routledge.

Argyle, Michael. (1988). *Bodily communication*. *Bodily communication*. Routledge.

Arnold, M. B. (1960). Emotion and personality. *New York Columbia University Press Aubergé V*. New York: Columbia University Press.

Aspegren, K. (1999). BEME Guide No. 2: Teaching and learning communication skills in medicine - A review with quality grading of articles. *Medical Teacher*, *21*(6), 564–570.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human Memory: A Proposed System and its Control Processes. *Psychology of Learning and Motivation*, *2*, 89–195.

Aylett, R., Lim, M., Ritter, C., Nazir, A., Hall, L., Paiva, A., … Kappas, A. (2014). *Werewolves, Cheats, and Cultural Sensitivity*. In *Proceedings of the 2014 International Conference on Autonomour Agents and Multi-agent Systems* (pp. 1085-1092). International Foundatiaon for Autonomous Agents and Multiagent Systems.

Aylett, R., Vala, M., Sequeira, P., & Paiva, A. (2007). FearNot! – an emergent narrativeapproach to virtual dramas for anti-bullying education. In *Interantional Conference on Virtual Storytelling* (pp. 2022-205). Springer, Berlin, Heidelberg.

Babad, E. (1999). Preferential Treatment in Television Interviewing: Evidence from Nonverbal Behavior. *Political Communication*, *16*(3), 337–358.

Baddeley, A. D., & Hitch, G. (1974). Working Memory. *Psychology of Learning and Motivation*, *8*, 47–89.

Bahreini, K., Nadolski, R., & Westera, W. (2017). Communication skills training exploiting multimodal emotion recognition. *Interactive Learning Environments*, *25*(8), 1065–1082.

Baldassarri, S., Hupont, I., Abadía, D., & Cerezo, E. (2015). Affective-aware tutoring platform for interactive digital television. *Multimedia Tools and Applications*, *74*(9), 3183–3206.

Bandura, A. (1986). The Explanatory and Predictive Scope of Self-Efficacy Theory. *Journal of Social and Clinical Psychology*, *4*(3), 359–373.

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. *Journal of Human–Computer Interaction Intl. Journal of Human–Computer Interaction*, *24*(6), 574–594.

Barmaki, R., & Hughes, C. (2018). Gesturing and Embodiment in Teaching : Investigating the Nonverbal Behavior of Teachers in a Virtual Rehearsal Environment. In *Thirty-Second AAAI Conference on Artificial Intelligence* (pp. 7893–7899).

Barrett, L. F. (2006, February 1). Valence is a basic building block of emotional life. *Journal of Research in Personality*. Academic Press.

Barth, J., & Lannen, P. (2011). Efficacy of communication skills training courses in oncology: a systematic review and meta-analysis. *Annals of Oncology*, *22*(5), 1030–1040.

Basu, S., & Weil, D. (1998). Appropriate Technology and Growth. *The Quarterly Journal of Economics*, *114*(4), 1025–2054.

Batrinca, L., Stratou, G., Shapiro, A., Morency, L. P., & Scherer, S. (2013). Cicero - Towards a multimodal virtual audience platform for public speaking training. In *International Workshop on Intelligent Virutal Agents* (pp. 116 - 128). Springer, Berlin, Heidelberg.

Baylor, A. L., & Kim, S. (2009). Designing Effective Nonverbal Communication for Pedagogical Agents. *Computers in Human Behavior*, *25*(2), 450–457.

Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, *190*(1), 80–91.

Bennett, K. B., & Flach, J. M. (1992). Graphical Displays: Implications for Divided Attention, Focused Attention, and Problem Solving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *34*(5), 513–533.

Bienvenu, M. J. (1971). An Interpersonal Communication Inventory. *Journal of Communication*, *21*(4), 381–388.

Bilakhia, S., Petridis, S., & Pantic, M. (2013). Audiovisual detection of behavioural mimicry. *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013*, 123–128.

Binali, H., & Potdar, V. (2012). Emotion detection state of the art. In *Proceedings of the CUBE International Information Technology Conference on - CUBE '12* (p. 501). New York, New York, USA: ACM Press.

Boersma, P., & van Heuven, V. (2001). Speak and unSpeak with Praat. *Glot International*, *5*(9–10), 341–347.

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, *49*(8), 1017–1034.

Brantley, C., & Miller, M. (2007). *Effective communication for colleges*. Cengage Learning.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140.

Brierley, J. A. (2017). The role of a pragmatist paradigm when adopting mixed methods in behavioural accounting research. *International Journal of Behavioural Accounting and Finance*, *6*(2), 140.

Brooke, J. (1996). SUS - A quick and dirty usability scale. *Usability Evaluation in Industry*, *189*(194), 4–7.

Brooke, J. (2013). SUS: a retrospecitve. *Journal fo Usability Studies, 8* (2), 29- 40.

Buchanan, M. (2009). Behavioural Science: Secret signals. *Nature*, *457*(7229), 528–530.

Buhrmester, D., Furman, W., Wittenberg, M. T., & Reis, H. T. (1988). *Five Domains of Interpersonal Competence in Peer Relationships*. *Journal of Personality and Social Psychology*, *55* (6), 991.

Bull, P. (1978). The interpretation of posture through an alternative methodology to role play. *British Journal of Clinical Psychology*, *17*(1), 1–6.

Burke, L. A., & Hutchins, H. M. (2007). Training transfer: An integrative literature review. *Human Resource Development Review*, *6* (3), 263- 296.

Burleson, B. R., Kunkel, A., Mortenson, S., Samter, W., & Xu, Y. (2003). *The experience and effects of emotional support: What the study of cultural and gender differences can tell us about close relationships, emotion, and interpersonal communication*. *Personal Relationships*, *10* (1), 1- 23.

Cafaro, A., Vilhjálmsson, H. H., Bickmore, T., Heylen, D., Jóhannsdóttir, K. R., & Valgarosson, G. S. (2012). First impressions: Users' judgments of virtual agents' personality and interpersonal attitude in first encounters. In *International Conference on Intelligent Virtual Agents* (pp. 6780). Springer, Berlin, Heidelberg.

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, *6*(8), 641–651.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and Quasi-Experimental Design for Research*. Ravenio Books.

Cannon, W. B. (1927). The James-Lange theory of emotions: A critical examination and an alternative theory. *The American Journal of Psychology*, *39*(1/4), 106–124.

Cannon, W. B. (1931). Again the James-Lange and the thalamic theories of emotion. *Psychological Review*, *38*(4), 281–295.

Cardon, P. W. (2008). A critique of Hall's contexting model: A meta-analysis of literature on intercultural business and technical communication. *Journal of Business and Technical Communication*, *22*(4), 399–428.

Carl, H. (1980). Nonverbal Communication during the Employment Interview. *ABCA Bulletin*, *43*(4), 14–19.

Cassell, J., & Tartaro, A. (2007). Intersubjectivity in human–agent interaction. *Interaction Studies*, *8*(3), 391–410.

Čereković, A., & Pandžić, I. S. (2011). Multimodal behaviour realization for embodied conversational agents. *Multimedia Tools and Applications*, *54*(1), 143-164

Cerrato, L. (2005). Linguistic functions of head nods. In *Proc. Conf. Multi--modal Communication* (pp. 137–152).

Chaby, L., Chetouani, M., Plaza, M., & Cohen, D. (2012). Exploring Multimodal Social-Emotional Behaviors in Autism Spectrum Disorders An interface between social signal processing and psychopathology. In *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing* (pp. 950-954). IEEE.

Chandler, D. (1994). *The Transmission Model of Communication.* University of Wales, Aberystwyth. Online: http://www.aber.ac.uk/media/documents/short/trans.html.

Chandrasegaran, S., Badam, S. K., Kisselburgh, L., Peppler, K., Elmqvist, N., & Ramani, K. (2016). *International Journal of Human-Computer Studies, 100,* 66-80.

Chang, C.-Y., Lange, B., Zhang, M., Koenig, S., Requejo, P., Somboon, N., … Rizzo, A. A. (2012). Towards pervasive physical rehabilitation using Microsoft Kinect. In *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2012 6th International Conference on* (pp. 159–162). IEEE.

Chang, K. Y., Liu, T. L., & Lai, S. H. (2009). Learning partially-observed hidden conditional random fields for faical expression recognition. In *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 533–540).

Chartrand, T. L., & Bargh, J. A. (2002). Nonconscious motivations: Their activation, operation and consequences. In A. Tesser, D. A. Stapel, & J. V. Wood (Eds.), *Self and Motivation: Emerging pscyhological perspectives* (pp. 13–41). Washington, DC, US.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357.

Cheng, D. S., Salamin, H., Salvagnini, P., Cristani, M., Vinciarelli, A., & Murino, V. (2014). Predicting online lecture ratings based on gesturing and vocal behavior. *Journal on Multimodal User Interfaces*, *8*(2), 151–160.

Chin, J. P., Diehl, V. A., & Norman, L. K. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems - CHI '88* (pp. 213–218). New York, New York, USA: ACM Press.

Chollet, M., Ghate, P., & Scherer, S. (2018). A generic platform for training social skills with adaptative virtual agents. In *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems,* (pp. 1800 - 1802). International Foundation for Autonomous Agents and Multiagent Systems.

Chollet, M., Wörtwein, T., Morency, L.-P., Shapiro, A., & Scherer, S. (2016). Exploring feedback strategies to improve public speaking: an interactive virtual audience framework. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing,* pp. 1143 - 1154. ACM 2015.

Choudhury, T., & Pentland, A. (2004). Characterizing social interactions using the sociometer. *Proceedings of NAACOS*, 1–4.

Chwalisz, K., Diener, E., & Gallagher, D. (1988). Autonomic arousal feedback and emotional experience: evidence from the spinal cord injured. *Journal of Personality and Social Psychology*, *54*(5), 820.

Cid, F., Manso, L. J., & Núñez, P. (2015). *A novel multimodal emotion recognition approach for affective human robot interaction*. In *Proceedings of FinE,* 1 - 9.

Coekin, J. A. (1970). An oscilloscope polar coordinate display for multi-dimensional data. *Radio and Electronic Engineer*, *40*(2), 97.

Cohen, D., & Crabtree, B. (2006). Qualitative Research Guidelines Project.

Cohen, I., Sebe, N., Garg, A., Chen, L. S., & Huang, T. S. (2003). Facial expression recognition from video sequences: temporal and static modeling. *Computer Vision and Image Understanding*, *91*(1–2), 160–187.

Cole, T. J. (2015). Too many digits: the presentation of numerical data. *Archives of Disease in Childhood*, *100*(7), 608–609.

Colecchia, F., Giacomin, J., & Hone, K. (2018). *A preliminary investigation towards the development of an emotion-aware partner agent for training control*. Contemporary Ergonomics and Human Factors.

Costanzo, M. (1992). Training Students to Decode Verbal and Nonverbal Cues: Effects on Confidence and Performance. *Journal of Educational Psychology*, *84*(3), 308–313.

Coulson, M. (2004). Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence. *Journal of Nonverbal Behavior*, *28*(2), 117–139.

Crawford, C., Crawford, R., & Jin, W. (2013). *The outlook for higher education spending by the department for business, innovation and skills* (No. R86). IFS Report.

Crystal, D., & Ardener, E. (1971). Prosodic and Paralinguistic Correlates of Social Categories. In E. Ardener (Ed.), *Social Anthropology and Language* (pp. 185-206). London: Tavistock.

Cuddy, A. J. C., Wilmuth, C. A., Yap, A. J., & Carney, D. R. (2015). Preparatory power posing affects nonverbal presence and job interview performance. *Journal of Applied Psychology*, *100*(4), 1286–1295.

Cunningham, D. W., Kleiner, M., Bulthoff, H. H., & Wallraven, C. (2004). The Components of Conversational Facial Expressions. In *Proceedings of the Symposium on Applied Perception in Graphics and Visualization* (pp. 143–150).

Cunningham, J. B., & McCrum-Gardner, E. (2007). Power, effect and sample size using GPower: practical issues for researchers and members of research ethics committees. *EvidenceBased Midwifery*, *5*(4), 132–137.

Curhan, J. R., & Pentland, A. (2007). Thin slices of negotiation: predicting outcomes from conversational dynamics within the first 5 minutes. *Journal of Applied Psychology*, *92*(3), 802.

Damian, I., Baur, T., & André, E. (2016). Measuring the impact of multimodal behavioural feedback loops on social interactions. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction - ICMI 2016* (pp. 201–208).

Damian, I., Baur, T., Lugrin, B., Gebhard, P., Mehlmann, G., & André, E. (2015). Games are better than books: In-situ comparison of an interactive job interview game with conventional training. In *International Conference on Artificial Intelligentce in Education* (pp. 84-94). Springer, Cham.

Damian, I., Dietz, M., & André, E. (2018). The SSJ Framework: Augmenting Social Interactions Using Mobile Signal Processing and Live Feedback. *Frontiers in ICT*, *5*, 13.

Damian, I., Dietz, M., Gaibler, F., & André, E. (2016). Social Signal Processing for Dummies. In *ICMI '16 Proceedings of the 18th ACM International Conference on Multimodal Interaction* (pp. 394–395).

Damian, I., Tan, C. S. (Sean), Baur, T., Schöning, J., Luyten, K., & André, E. (2015). Augmenting Social Interactions: Realtime Behavioural Feedback Using Social Signal Processing Techniques. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, (pp. 565–574).

Darwin, C., & Prodger, P. (1998). *The expression of the emotions in man and animals*. Oxford University Press.

Darwin, C. R. (1872). *The Expression of the Emotions in Man and Animals*. Oxford University Press.

Davis, D. A., Thomson, M. A., Oxman, A. D., & Haynes, R. B. (1995). Changing Physician Performance. *JAMA*, *274*(9), 700.

de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition & Emotion*, *14*(3), 289–311.

Decoster, J. (2004). A Meta-Analysis of Priming Effects on Impression Formation Supporting a General Model of Informational Biases. *Article in Personality and Social Psychology Review*.

DeGroot, T., & Motowidlo, S. J. (1999). Why visual and vocal interview cues can affect interviewers' judgments and predict job performance. *Journal of Applied Psychology*, *84*(6), 986–993.

Dupré, D., Morrison, G., & Mckeown, G. (2018). Accuracy of three commercial automatic emotion recognition systems across different individuals and their facial expressions. In *2018 IEEE International Conference on Pervasive Computing and Communications: Proceeedings* (pp. 746–751).

Dupré, D., Morrison, G., Mckeown, G., Andelic, N., Morrison, G., & Mckeown, G. (2017). Accuracy of three commercial automatic emotion recognition systems across different individuals and their facial expressions. In *2018 IEEE International Conference on Pervasive Computing and Communications: Proceeedings* (pp. 746–751).

Duran, R. L. (1983). Communicative adaptability: A measure of social communicative competence. *Communication Quarterly*, *31*(4), 320–326.

Dzulkifli, M. A., & Mustafar, M. F. (2013). The influence of colour on memory performance: a review. *The Malaysian Journal of Medical Sciences : MJMS*, *20*(2), 3–9.

Ekman, P, & Friesen, W. V. (1986). A New Pan-Cultural Facial Expression Of Emotion. *Motivation And Emotion*, *10*(2), 159–168.

Ekman, P, Levenson, R., & Friesen, W. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*, *221*(4616), 1208–1210.

Ekman, P, & Oster, H. (1979). Facial expressions of emotion. *Annual Review Psychology*, *30*, 527–554.

Ekman, Paul., & Davidson, R. J. (1994). The nature of emotions: Fundamental questions. *American Scientist*, *89*(4),

Ekman, Paul., Friesen, W. V., Hager, J. C., & A Human Face (Firm). (2002). *Facial action coding system*. Salt Lake City: A Human Face.

Ekman, Paul. (1992). An Argument for Basic Emotions. *Cognition and Emotion*.

Ekman, Paul. (1999). Facial Expression. In *Handbook of Cognition and Emotion* (Vol. 21, pp. 301–320).

Ekman, Paul. (2016). What Scientists Who Study Emotion Agree About. *Perspectives on Psychological Science*, *11*(1), 31–34.

Ekman, Paul, & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*(2), 124–129.

Ekman, Paul, & Friesen, W. V. (1978). *The Facial Action Coding System: a technique for the measurement of facial movement.* Consulting Psychologists: San Francisco.

Ekman, Paul, & Friesen, W. V. (1967). Head and body cues in the judgment of emotion: A reformulation. *Perceptual and Motor Skills*, *24*(3 PT 1), 711–724.

Ekman, Paul, Friesen, W. V, & Ellsworth, P. (1982). What emotion categories or dimensions can observers judge from facial behavior? In Paul Ekman (Ed.), *Emotion in the Face* (pp. 30–55). New York: Cambridge University Press.

Ekman, Paul, Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science*, *164*(3875), 86–88.

Ellyson, S. L., & Dovidio, J. F. (1985). Power, Dominance, and Nonverbal Behavior: Basic Concepts and Issues. In *Power, Dominance, and Nonverbal Behavior* (pp. 1–27). New York, NY: Springer New York.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, and Computers*, *28*(1), 1–11.

Fabrigar, L., Wegener, D., MacCallum, R., & Strahan, E. (1999). *Psychological methods. Psychological Methods* (Vol. 4). American Psychological Association.

Fallowfield, L., Jenkins, V., Farewell, V., & Solis-Trapala, I. (2003). Enduring impact of communication skills training: Results of a 12-month follow-up. *British Journal of Cancer*, *89*(8), 1445–1449.

Fang, X., van Kleef, G. A., & Sauter, D. A. (2018). Person perception from changing emotional expressions: primacy, recency, or averaging effect? *Cognition and Emotion*, pp. 1–14.

Fantini, A. E., & Smith, E. M. (1997). A survey of intercultural communication courses. International Journal of Intercultural Relations, 21, 125-148.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. In *Behavior Research Methods* (Vol. 39, pp. 175–191). Springer-Verlag.

Ferguson, N. (1977). Simultaneous speech, interruptions and dominance. *British Journal of Social and Clinical Psychology*, *16*(4), 295–302.

Ferreira, E., Lefevre, F., & Lefèvre, F. (2013). Social Signal and User Adaptation in Reinforcement Learning-based Dialogue Management. *Proceedings of the 2Nd Workshop on Machine Learning for Interactive Systems: Bridging the Gap Between Perception, Action and Communication*, 61–69.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics.* Sage Publications.

Fischbach, K., Gloor, P. A., Lassenius, C., Olguin, D. O., Pentland, A., Putzke, J., & Schoder, D. (2010). Analyzing the flow of knowledge with sociometric badges. *Procedia - Social and Behavioral Sciences*, *2*(4), 6389–6397.

Fiske, J. (2010). *Introduction to Communication Studies*. Routledge.

Forte, R., & Caan, J. (2015). *McDonald's Backing Soft Skills. A Plan for Recognising, Developing and Measuring Soft Skills at Every Stage of Education and Work*.

Frank, M. G., & Ekman, P. (2004). Appearing truthful generalizes across different deception situations. *Journal of Personality and Social Psychology*, *86*(3), 486–495.

Frauendorfer, D., Schmid Mast, M., Nguyen, L., & Gatica-Perez, D. (2014). Nonverbal Social Sensing in Action: Unobtrusive Recording and Extracting of Nonverbal Behavior in Social Interactions Illustrated with a Research Example. *Journal of Nonverbal Behavior*, *38*(2), 231–245.

Frein, S. T., Jones, S. L., & Gerow, J. E. (2013). When it comes to Facebook there may be more to bad memory than just multitasking. *Computers in Human Behavior*, *29*(6), 2179–2182.

Fridlund, A. J. (1994). *Human Facial Expression*. *Human Facial Expression*. Academic Press.

Friesen, W., & Ekman, P. (1983). *EMFACS-7: Emotional facial action coding system*. *Unpublished manuscript, University of California*.

Frith, H., & Gleeson, K. (2004). Clothing and Embodiment: Men Managing Body Image and Appearance. *Psychology of Men & Masculinity*, *5*(1), 40–48.

Fu, W. J., Carroll, R. J., & Wang, S. (2005). Estimating misclassification error with small samples via bootstrap cross-validation. *Bioinformatics Original Paper*, *21*(9), 1979–1986.

Fukui, S., Ogawa, K., & Fukui, N. (2010). Communication skills training on how to break bad news for Japanese nurses in oncology: Effects of training on nurses' confidence and perceived effectiveness. *Journal of Cancer Education*, *25*(1), 116–119.

Fung, M., Jin, Y., Zhao, R., & Hoque, M. (Ehsan). (2015). ROC speak: Semi-automated personalized feedback on nonverbal behaviour from recorded videos. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15* (pp. 1167–1178).

Futris, T. G., Campbell, K., Nielsen, R. B., & Burwell, S. R. (2010). The Communication Patterns Questionnaire-Short Form: A Review and Assessment. *The Family Journal*, *18*(3), 275–287.

Gao, Y., Bianchi-Berthouze, N., & Meng, H. (2012). What Does Touch Tell Us about Emotions in Touchscreen-Based Gameplay? *ACM Transactions on Computer-Human Interaction*, *19*(4), 1–30.

Garner, R., Gillingham, M. G., & White, C. S. (1989). Effects of "Seductive Details" on Macroprocessing and Microprocessing in Adults and Children. *Cognition and Instruction*, *6*(1), 41–57.

Gastwirth, J. L., Gel, Y. R., & Miao, W. (2009). The Impact of Levene's Test of Equality of Variances on Statistical Theory and Practice. *Statistical Science*, *24*(3), 343–360.

Gatica-Perez, D. (2009). Automatic nonverbal analysis of social interaction in small groups: A review. *Image and Vision Computing*, *27*(12), 1775–1787.

Gauvain, J. L., Lamel, L. F., & Adda, G. (1998). Partitioning and Transcription of Broadcast News Data. In *Fifth International Conference of Spoken Langauge Processing*.

Gebhard, P., Schneeberger, T., André, E., Baur, T., Damian, I., Mehlmann, G., … Langer, M. (2018). Serious Games for Training Social Skills in Job Interviews. *IEEE Transactions on Games*.

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for nonstatisticians. *International Journal of Endocrinology and Metabolism*, *10*(2), 486–489.

Gholami, B., Haddad, W. M., & Tannenbaum, A. R. (2009). Agitation and pain assessment using digital imaging. *Conference Proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, *2009*, 2176–2179.

Gifford, R. (2012). The Role of Nonverbal Communication in Interpersonal Relations. In *Handbook of Interpersonal Psychology: Theory, Research, Assessment, and Therapeutic Interventions* (pp. 171–190).

Gigante, J., Dell, M., & Sharkey, A. (2011). Getting Beyond "Good Job": How to Give Effective Feedback. *Pediatrics*, *127*(2), 205–207.

Glass, C. R., Merluzzi, T. V., Biever, J. L., & Larsen, K. H. (1982). Cognitive assessment of social anxiety: Development and validation of a self-statement questionnaire. *Cognitive Therapy and Research*, *6*(1), 37–55.

Goudbreek, M., Goldman, J. P., & Scherer, K. R. (2009). Emotion Dimoentions and formant position. In *Tenth Annual Conference of the International Speech Communication Association*.

Gould, S. J., & Grein, A. F. (2009). Think glocally, act glocally: A culture-centric comment on Leung, Bhagat, Buchan, Erez and Gibson (2005). *Journal of International Business Studies*, *40*(2), 237–254.

Gray, J. A. (1999). Cognition, emotion, conscious experience and the brain. In *Handbook of Cognition and Emotion.* (pp. 83–102). Chichester, UK, UK: John Wiley & Sons, Ltd.

Gross, J. J., & Levenson, R. W. (1993). Emotional suppression: Physiology, self-report, and expressive behavior. *Journal of Personality and Social Psychology*, *64*(6), 970–986.

Gross, J. J., & Levenson, R. W. (1997). Hiding feelings: The acute effects of inhibiting negative and positive emotion. *Journal of Abnormal Psychology*, *106*(1), 95–103.

Gudykunst, W. B., Matsumoto, Y., Ting-Toomey, S., Nishida, T., Kim, K., & Heyman, S. (1996). The influence of cultural individualism-collectivism, self construals, and individual values on communication styles across cultures. *Human Communication Research*, *22*(4), 510–543.

Gudykunst, W. B., & Nishida, T. (1986). Attributional confidence in low- and high-context cultures. Human Communication Research, 12, 525-549.

Guest, G., MacQueen, K. M., & Namey, E. E. (2012). *Applied thematic analysis*. Sage Publications.

Gul, A., Perperoglou, A., Khan, Z., Mahmoud, O., Miftahuddin, M., Adler, W., & Lausen, B. (2018). Ensemble of a subset of kNN classifiers. *Advances in Data Analysis and Classification*, *12*(4), 827–840.

Gulliksen, J., Göransson, B., Boivie, I., Blomkvist, S., Persson, J., & Cajander, Å. (2003). Key principles for user-centred systems design. In *Behaviour and Information Technology* (pp. 397–409).

Gunes, H., & Hung, H. (2015). Emotional and Social Signals: A Neglected Frontier in Multimedia Computing? *IEEE Multimedia*, *22*(2), 76–85.

Gunes, H., Piccardi, M., & Pantic, M. (2008). From the Lab to the Real World : Affect Recognition Using Multiple Cues and Modalities. *Affective Computing: Focus on Emotion Expression, Synthesis, and Recognition*, 185–218.

Hackman, J., & Craig, M. (2009). *Leaderhip: A Communication Perspective*. *Waveland Press*.

Halford, W. K., Lee, S., Hiew, D. N., & van de Vijver, F. J. R. (2018). Indirect couple communication and relationship satisfaction in Chinese, Western, and Chinese-Western Intercultural Couples. *Couple and Family Psychology: Research and Practice*, *7*(3–4), 183–200

Hall, E. T. (1976). *Beyond culture*. *Contemporary Sociology* (Vol. 7). New York: Anchor Books.

Hall, M. (1999). Correlation-based Feature Selection for Machine Learning. *Methodology*, *21i195i20*(April), 1–5.

Han, Y., & Nunes, J. (2010). Read the Signal But Don’t Mention It: How Consumption Embarrasses the Signaler. *Advances in Consumer Research*, *37*, 81–84.

Hart, Y., Czerniak, E., Karnieli-Miller, O., Mayo, A. E., Ziv, A., Biegon, A., … Alon, U. (2016). Automated video analysis of non-verbal communication in a medical setting. *Frontiers in Psychology*, *7*(AUG), 1130.

Hashimoto, M., Yamano, M., & Usui, T. (2009). Effects of emotional synchronization in human-robot KANSEI communications. In *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, 52–57.

Hass, R. G., & Eisenstadt, D. (1990). The effects of self-focused attention on perspective-taking and anxiety. *Anxiety Research*, *2*(3), 165–176.

Hirschberg, J. (1993). *Pitch accent in context: predicting intonational prominence from text*. *Artificial Intelligence* (Vol. 63).

Hirschberg, J., & Grosz, B. (1992). Intonational Features of Local and Global Discourse Structure. In *Proceedings of the Workshop on Speech and Natural Language* (pp. 441-446). Association for Computational Linguistics.

Ho, S., Foulsham, T., & Kingstone, A. (2015). Speaking and listening with the eyes: Gaze signaling during dyadic interactions. *PLoS ONE*, *10*(8), e0136905.

Ho, V. T., Wong, S. S., & Lee, C. H. (2011). A Tale of Passion: Linking Job Passion and Cognitive Engagement to Employee Work Performance. *Journal of Management Studies*, *48*(1), 26–47.

Hofstede, G. (2001). Culture's consequences: Comparing values, behaviors, institutions, and organizations across cultures (2nd ed.). Thousand Oaks, CA: Sage

Hofstede, G. (2011). Dimensionalizing Cultures: The Hofstede Model in Context. *Online Readings in Psychology and Culture*, *2*(1). https://doi.org/10.9707/2307-0919.1014

Holding, B. C., Sundelin, T., Lekander, M., & Axelsson, J. (2019). Sleep deprivation and its effects on communication during individual and collaborative tasks. *Scientific Reports*, *9*(1), 3131.

Hollandsworth, J. G., Hazelskis, R., Stevens, J., & Dressel, M. (1979). Relative contributions of verbal, articulative and non verbal communicaiton to emplyment decisions in the job interview. *Personnel Psychology*, *32*(2), 359–367.

Hoque, M., Courgeon, M., & Martin, J. (2013). Mach: My Automated Conversation coacH. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13)* (pp. 697–706).

Hoque, M., Courgeon, M., Martin, J., Mutlu, B., & Picard, R. W. (2013). Mach: My Automated Conversation coacH. *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13)*, 697–706.

Horvath, F., Mccloughan, J., Weatherman, D., & Slowik, S. (2013). The accuracy of auditors' and Layered Voice Analysis (LVA) operators' judgments of truth and deception during police questioning. *Journal of Forensic Sciences*, *58*(2), 385–392.

Huang, X., Acero, A., Chelba, C., Deng, L., Droppo, J., Duchene, D., … Wang, Y. (2001). MIPAD: A multimodal interaction prototype. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *1*, 9–12.

Izard, C. E. (1971). The Face of Emotion. In *Appleton-Centuary-Crofts*. New York.

Jain, A. K., & Ross, A. (2015). Bridging the gap: from biometrics to forensics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1674), 20140254.

James, W. (1884). What is an Emotion? *Mind*, *9*(34), 188–205.

Kaakinen, J. K., Ballenghein, U., Tissier, G., & Baccino, T. (2018). Fluctuation in cognitive engagement during reading: Evidence From Concurrent Recordings of Postural and Eye Movements. *Journal of Experimental Psychology: Learning Memory and Cognition*, *44*(10), 1671–1677.

Keeley-Dyreson, M. P., Bailey, W., & Burgoon, J. K. (1991). The Effects of Stress and Gender on Nonverbal Decoding Accuracy in Kinesic and Vocalic Channels. *Human Communication Research*, *17*(4), 584–605.

Keltner, D., & Haidt, J. (1999). Social Functions of Emotions at Four Levels of Analysis. *Cognition and Emotion*, *13*(5), 505–521.

Keltner, D., & Kring, A. M. (1998). *Emotion, Social Function, and Psychopathology*. *Review of General Psychology* (Vol. 2).

Kensinger, E. A. (2009). Remembering the details: Effects of emotion. *Emotion Review*. NIH Public Access.

Kim, M.-G. G., & Suzuki, K. (2014). Comparative Study of Human Behavior in Card Playing with a Humanoid Playmate. *International Journal of Social Robotics*, *6*(1), 5–15.

Kim, T., Chang, A., Holland, L., & Pentland, A. S. (Sandy). (2008). Meeting mediator. *Proceedings of the ACM 2008 Conference on Computer Supported Cooperative Work - CSCW '08*, 457.

Kim, D., Pan, Y., & Park, H. S. (1998). High- versus low-context culture: A comparison of Chinese, Korean, and American cultures. Psychology & Marketing, 15, 507-521

Kim, Y., Soyata, T., & Behnagh, R. F. (2018). Towards Emotionally Aware AI Smart Classroom: Current Issues and Directions for Engineering and Education. *IEEE Access*, *6*, 5308–5331.

Kleinke, C. L. (1986). Gaze and eye contact: a research review. *Psychological Bulletin*, *100*(1), 78–100.

Klieger, D. M., & Siejak, K. K. (1997). Disgust as the source of false positive effects in the measurement of ophidiophobia. *Journal of Psychology: Interdisciplinary and Applied*, *131*(4), 371–382.

Kluger, A. N., & Denisi, A. (1996). The effects of feedback internventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254–284.

Knapp, M., & Hall, J. (2009). *Nonverbal communication in human interaction*. Holt, Rinehart and Winston.

Knapp, M. L., Hall, J. A., & Horgan, T. G. (2013). *Nonverbal communication in human interaction*. Cengage Learning.

Knutson, B. (1996). Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior*, *20*(3), 165–182.

Kochanowicz, J., Tan, A.-H., & Thalmann, D. (2016). Modeling human-like non-rationality for social agents. *Proceedings of the 29th International Conference on Computer Animation and Social Agents - CASA '16*, 11–20.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2* (pp. 1137–1143). International Joint Conferences on Artificial Intelligence, Inc.

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Englewood Cliffs, NJ: Prentice-Hall.

Konopka, L., Duffecy, J., & Hur, K. (2010). Technology assisted PTSD screening (TAPS) for the Vietnam Veterans of America (VVA) study. In *Pilot Study Abstract, The Chicago School of Professional Psychology, Northwestern University, and The University of Illinois at Chicago.*

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163.

Kort, B., & Reilly, R. (2002). Analytical Models of Emotions, Learning and Relationships: Towards an Affect-sensitive Cognitive Machine. *Virtual Worlds Simulation Conference*, 1–15.

Kortum, P. T., & Bangor, A. (2013). Usability Ratings for Everyday Products Measured with the System Usability Scale. *International Journal of Human-Computer Interaction*, *29*(2), 67–76.

Kraut, R. E. (1978). Verbal and nonverbal cues in the perception of lying. *Journal of Personality and Social Psychology*, *36*(4), 380–391.

Kraut, R. E., & Johnston, R. E. (1979). Social and emotional messages of smiling: An ethological approach. *Journal of Personality and Social Psychology*, *37*(9), 1539–1553.

Krys, K., -Melanie Vauclair, C., Capaldi, C. A., Lun, V. M. C., Bond, M. H., Domínguez-Espinosa, A., … Yu, A. A. (2016). Be Careful Where You Smile: Culture Shapes Judgments of Intelligence and Honesty of Smiling Individuals. *Journal of Nonverbal Behavior*, *40*(2), 101–116.

Lacerda, F. (2009). LVA-technology-The illusion of “lie detection.” *Proceedings, FONETIK 2009*, 220–226.

Lacerda, F. (2012). Money Talks: The Power of Voice: A critical review of Mayew and Ventachalam’s The Power of Voice: Managerial Affective States and Future Firm Performance. *PERILUS*, 1–10.

Landman, B. A., Asman, A. J., Scoggins, A. G., Bogovic, J. A., Xing, F., & Prince, J. L. (2012). Robust statistical fusion of image labels. *IEEE Transactions on Medical Imaging*, *31*(2), 512–522.

Lapidot-Lefler, N., & Barak, A. (2012). Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in Human Behavior*, *28*(2), 434–443.

Lazarus, R. S. (1966). Psychological stress and the coping process. McGraw-Hill.

Lee, Y., Song, J., Min, C., Hwang, C., Lee, J., Hwang, I., … Lee, U. (2013). SocioPhone. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services - MobiSys ’13* (p. 375). New York, New York, USA: ACM Press.

Lepper, M. R., & Woolverton, M. (2002). The Wisdom of Practice: Lessons Learned from the Study of Highly Effective Tutors. In *Improving Academic Achievement* (pp. 135–158).

Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, *7*(1), 57–78.

Li, Cen. (2007). Classifying imbalanced data using a bagging ensemble variation (BEV). In *Proceedings of the 45th annual southeast regional conference on - ACM-SE 45* (p. 203). New York, New York, USA: ACM Press.

Cheng, L, Wang, B. & Carrol (2014). Selecting the number of principal components in Functional Data. *Journal of the Amercian Statistical Association, 108*(504), 1284-1294

Li, H. Z. (2001). Cooperative and intrusive interruptions in inter- and intracultural dyadic discourse. *Journal of Language and Social Psychology*, *20*(3), 259–284.

Li, H. Z., Yum, Y.-O., Yates, R., Aguilera, L., Mao, Y., & Zheng, &yue. (2005). Interruption and Involvement in Discourse : Can Intercultural Interlocutors be Trained ? *Journal of Lntercultural Communication Research*, *34*(4), 233–254.

Lifeng Shang, & Chan, K.-P. (2009). Nonparametric discriminant HMM and application to facial expression recognition. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2090–2096). IEEE.

Lim, T.-S., & Loh, W.-Y. (1996). A comparison of tests of equality of variances. *Computational Statistics & Data Analysis*, *22*(3), 287–301.

Littlewort, G., Bartlett, M. S., & Lee, K. (2008). *Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction*. *Lecture Notes in Computer Science* (Vol. 5042). Springer Verlag.

Liu, C., Calvo, R. A., & Lim, R. (2016). Improving Medical Students' Awareness of Their Non-Verbal Communication through Automated Non-Verbal Behavior Feedback. *Frontiers in ICT*, *3*, 11.

Liu, C., Scott, K. M., Lim, R. L., Taylor, S., & Calvo, R. A. (2016). EQClinic: a platform for learning communication skills in clinical consultations. *Medical Education Online*, *21*, 1–7.

Liu, J. E., Mok, E., Wong, T., Xue, L., & Xu, B. (2007). Evaluation of an integrated communication skills training program for nurses in cancer care in Beijing, China. *Nursing Research*, *56*(3), 202–209.]

Liu, Y., Huang, Y., Gao, H., & Cheng, X. (2017). Communication skills training: Adapting to the trends and moving forward. *BioScience Trends*, *11*(2), 142–147.

Logan-Terry, A., & Damari, R. R. (2015). Key Culture-general Interactional Skills for Military Personnel. *Procedia Manufacturing*, *3*, 3990–3997.

Lomas, J., Pickard, L., Bester, S., Elbard, H., Finlayson, A., & Zoghaib, C. (1989). The Communicative Effectiveness Index. *Journal of Speech and Hearing Disorders*, *54*(1), 113–124.

Lowe, M. R., & D'Ilio, V. (1985). Factor analysis of the social performance survey schedule. *Journal of Psychopathology and Behavioral Assessment*, *7*(1), 13–22.

Lowenthal, K., & Lewis, C. A. (2018). *An Introduction to Psychological Tests and Scales*. Psychology Press.

Lu, L., Zhang, H. J., & Jiang, H. (2002). Content Analysis for Audio Classification and Segmentation. *IEEE Transactions on Speech and Audio Processing*, *10*(7), 504–516.

Lucas, K., Kerrick, S., Haugen, J., & Crider, C. J. (2016). *Communicating Entrepreneurial Passion: Personal Passion vs. Perceived Passion in Venture Pitches*. *IEEE Transactions on Professional Communication*, *59*(4), 363-378

Lucey, P., Cohn, J. F., Prkachin, K. M., Solomon, P. E., & Matthews, I. (2011). Painful Data: The UNBC-McMaster shoulder pain expression archive database. *Face and Gesture*, 57–64.

MacKenzie, I. S. (2012). *Human-computer interaction.* Elsevier Science.

Mandrekar, J. N. (2011a). Measures of interrater agreement. *Journal of Thoracic Oncology*, *6*(1), 6–7.

Mandrekar, J. N. (2011b). Measures of Interrater Agreement. *Journal of Thoracic Oncology*, *6*(1), 6–7.

Mann, K., van der Vleuten, C., Eva, K., Armson, H., Chesluk, B., Dornan, T., … Sargeant, J. (2011). Tensions in Informed Self-Assessment: How the Desire for Feedback and Reticence to Collect and Use It Can Conflict. *Academic Medicine*, *86*(9), 1120–1127.

Martin, M. M., & Rubin, R. B. (1994). Development of a communication flexibility measure. *Southern Communication Journal*, *59*(2), 171–178.

Marsella, S. C., & Gratch, J. (2009). EMA: A process model of appraisal dynamics. *Cognitive Systems Research. 10*(1), 70-90.

McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review*, *90*(3), 215–238.

McDoughall, W. (1926). *An introduction to to social psychology*. Boston: Luce.

McGarry, J., & O'Leary, B. (1993). *The politics of ethnic conflict regulation : case studies of protracted ethnic conflicts*. Routledge.

McKnight, P. E., & Najab, J. (2010). Mann-Whitney U Test. In *The Corsini Encyclopedia of Psychology* (pp. 1–1). Hoboken, NJ, USA: John Wiley & Sons, Inc.

McNeill, D. (2008). *Gesture and thought*. Chicagot: University of Chicago Press.

Meissel, K., Meyer, F., Yao, E. S., & Rubie-Davies, C. M. (2017). Subjectivity of teacher judgments: Exploring student characteristics that influence teacher judgments of student ability. *Teaching and Teacher Education*, *65*, 48–60.

Mitchell, G. (2012). Revisiting Truth or Triviality. *Perspectives on Psychological Science*, *7*(2), 109–117.

Mohanta, A., & Mittal, V. K. (2016). Classifying emotional states using pitch and formants in vowel regions. In *2016 International Conference on Signal Processing and Communication (ICSC)* (pp. 258–463).

Montague, P. R., Berns, G. S., Cohen, J. D., McClure, S. M., Pagnoni, G., Dhamala, M., … Fisher, R. E. (2002). Hyperscanning: simultaneous fMRI during linked social interactions. *NeuroImage*, *16*(4), 1159–1164.

Moors, A. (2013). On the Causal Role of Appraisal in Emotion. *Emotion Review*, *5*(2), 132–140.

Morgan, N. (2008). How to Become an Authentic Speaker. *Harvard Business Review*, *86*(11), 115–119.

Mowrer, O. H. (1960). *Learning theory and behaviour*. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Muehlhan, M., Marxen, M., Landsiedel, J., Malberg, H., & Zaunseder, S. (2014). The effect of body posture on cognitive performance: a question of sleep quality. *Frontiers in Human Neuroscience*, *8*, 171.

Muralidhar, S. (2017). On Job Training: Automated Interpersonal Behavior Assessment & RealTime Feedback. In *Proceedings of the 25th ACM International Conference on Multimedia, 2017.*

Muralidhar, S., Nguyen, L. S., Frauendorfer, D., Odobez, J. M., Mast, M. S., & Gatica-Perez, D. (2016). Training on the job: Behavioral analysis of job interviews in hospitality. In *ICMI 2016 - Proceedings of the 18th ACM International Conference on Multimodal Interaction* (Vol. 16, pp. 84–91).

Naim, I., Tanveer, M. I., Gildea, D., Hoque, E., Mohammed, Hoque, & Hoque, E. (2016). Automated analysis and prediction of job interview performance. *IEEE Transactions on Affective Computing*, 1–14.

Navarro, J. (2003). A Four-Domain Model for Detecting Deception. *FBI Law Enforcement Bulletin*, 19–24.

Nguyen, L. S., & Gatica-Perez, D. (2015). I Would Hire You in a Minute. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction - ICMI '15* (pp. 51–58).

Nguyen, L. S., & Gatica-Perez, D. (2016). *Hirability in the Wild: Analysis of Online Conversational Video Resumes*. *IEEE Transactions on Multimedia* (Vol. 18).

Oatley, K., & Johnson-Laird, P. N. (1987). Towards a Cognitive Theory of Emotions. *Cognition and Emotion*, *1*(1), 29–50.

Ofek, E., Iqbal, S. T., & Strauss, K. (2013). Reducing disruption from subtle information delivery during a conversation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13* (p. 3111). New York, New York, USA: ACM Press.

Oguri, M., & Gudykunst, W. B. (2002, September 1). The influence of self construals and communication styles on sojourners' psychological and sociocultural adjustment. *International Journal of Intercultural Relations*. Pergamon.

Ohashi, R. (2000). High/low-context communication: Conceptualization and scale development. Dissertation Abstracts International, 61(05), 1680. (University Microfilms No. 9971973)

Olguín, D., Pentland, A., Olguin, D., & Pentland, A. (2007). Sociometric badges: State of the art and future applications. *IEEE 11th International Symposium on Wearable Computers. Boston, MA*, 1–3.

Opdenakker, R. (2006). Advantages and Disadvantages of the Four Interview Techniques. *Forum: Qualitative Social Research*, *7*(4), 1–10.

Ortony, A., & Turner, T. J. (1990). What's Basic About Basic Emotions ? *Psychological Review*, *97*(3), 315–331.

Otsuka, T., & Ohya, J. (1997). Recognizing multiple persons' facial expressions using HMM based on automatic extraction of significant frames from image sequences. In *IEEE International Conference on Image Processing* (Vol. 2, pp. 546–549). IEEE Comput. Soc.

Panksepp, J. (1982). Toward a general Psychobiological theory of emotions. *Behavioural and Brain Sciences*, *5*(3), 407–467.

Pantic, M., Sebe, N., Cohn, J. F., & Huang, T. (2005). Affective multimodal human-computer interaction. *Proceedings of the 13th Annual ACM International Conference on Multimedia,* 669.

Patel, V. R., & Mehta, R. G. (2011). Impact of Outlier Removal and Normalization Approach in Modified k-Means Clustering Algorithm. *IJCSI International Journal of Computer Science Issues*, *8*(2).

Patro, S. G. K., & Sahu, K. K. (2015). Normalization: A Preprocessing Stage. *IARJSET*, 20–22.

Patton, M. Q. (2002). Qualitative research & evaluation methods (3rd ed.). Thousand Oaks, CA: Sage

Paxton, A., Rodriguez, K., & Dale, R. (2015). PsyGlass: Capitalizing on Google Glass for naturalistic data collection. *Behavior Research Methods*, *47*(3), 608–619.

Pecchia, L., Chen, L. L., Nugent, C., & Bravo, J. (2014). Ambient Assisted Living and Daily Activities. In *6th international work-conference, IWAAL 2014, Belfast, UK* (Vol. 8868, p. 423).

Pekrun, R. (1992). The impact of emotions on learning and achievement: Towards a theory of cognitive/motivational mediators. *Applied Psychology*, *41*(4), 359–376.

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology*, *36*(1), 36–48.

Pentland, A., & Heibeck, T. (2010). *Honest signals: how they shape our world*. Cambridge, Massachusetts: MIT press.

Pereira, M., Colecchia, F., & Hone, K. (2018). User Centred Design of Social Signals Feedback for Communication Skills Training. *Proceedings of British HCI, 2018*, 1–5.

Picard, R. (1995). Affective Computing. *MIT Press*, (321), 1–16.

Picard, R. W., Vyzas, E., & Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(10), 1175–1191.

Piwek, L., Ellis, D. A., Andrews, S., & Joinson, A. (2016). The Rise of Consumer Health Wearables: Promises and Barriers. *PLOS Medicine*, *13*(2), e1001953.

Plutchik, R. (1994). *The psychology and biology of emotion.* HarperCollins College Publishers.

Plutchik, R. (2001). The Nature of Emotions Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, *89*(4), 344–350.

Podsakoff, P. M., & Organ, D. W. (1986). Self-Reports in Organizational Research: Problems and Prospects. *Journal of Management*, *12*(4), 531–544.

Poggi, I., & D 'errico, F. (2011). Social signals: A psychological perspective. In *Computer Analysis of Human Behaviour* (pp. 185–225). London: Springer.

Power, M., & Dalgleish, T. (2016). *Cognition and Emotion*. Psychology Press.

Prajapati, B., Dunne, M., & Armstrong, R. (2010). Sample size estimation and statistical power analyses. *OT Peer Reviewed*, 1–9.

Profita, J., & Bidder, T. G. (1988). Perfect pitch. *American Journal of Medical Genetics*, *29*(4), 763–771.

Psathas, G. (1995). *"Talk and social structure" and "studies of work." Human Studies* (Vol. 18).

Psychology, P. (2015). Transfer of Training : A Review and Directions for Future Research. *Personnel Psychology; Spring*, *41*(October), 63–105.

Pulvermüller, F., & Berthier, M. L. (2008). Aphasia therapy on a neuroscience basis. *Aphasiology*, *22*(6), 563–599.

Radbourne, J., Glow, H., & Johanson, K. (2013). *The audience experience: A critical analysis of audiences in the performing arts*. Intellect Books.

Rana el Kaliouby, Evan Kodra, Pankaj Jha, N. M. (2014). *Does Facial Coding Generalize Across Cultures? A Spotlight on Asia*. *Asia Pacific 2014 - Celebrating Asian Creativity*.

Rasipuram, S., & Sowmya. Prediction/Assessment of Communication Skill Using Multimodal Cues in Social Interactoin, *Proceedings of the 18th ACM International Conference on Multimodal Interaction - ICMI 2016* § (2016). New York, New York, USA: ACM Press.

Reason, J. (1990). *Human error*. Cambridge university press.

Reed, R. (2001). (Un-)Professional discourse? Journalists' and scientists' stories about science in the media. *Journalism*, *2*(3), 279–298.

Richardson, R. M., & Smith, S. W. (2007). The influence of high/low-context culture and power distance on choice of communication media: Students' media choice to communicate with professors in Japan and America. International Journal of Intercultural Relations, 31, 479501

Riggio, R. E. (2005). The Social Skills Inventory (SSI): Measuring nonverbal and social skills. *The Sourcebook of Nonverbal Measures: Going beyond Words*, 25–33.

Rosenman, R., Tennekoon, V., & Hill, L. G. (2011). Measuring bias in self-reported data. *International Journal of Behavioural & Healthcare Research*, *2*(4), 320–332.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–244). New York: Russell Sage Foundation.

Rosch, M., & Segler, K. G. (1987). Communication with the Japanese. Management International Review, 27, 56-67

Roter, D. L., Erby, L. H., Adams, A., Buckingham, C. D., Vail, L., Realpe, A., … Hall, J. A. (2014). Talking about depression: an analogue study of physician gender and communication style on patient disclosures. *Patient Education and Counseling*, *96*(3), 339–345.

Roter, D. L., & Hall, J. A. (2004). Physician Gender and Patient-Centered Communication: A Critical Review of Empirical Research. *Annual Review of Public Health*, *25*(1), 497–519.

Roter, D. L., Hall, J. A., Kern, D. E., Barker, L. R., Cole, K. A., & Roca, R. P. (1995). Improving physicians' interviewing skills and reducing patients' emotional distress: a randomized clinical trial. *Archives of Internal Medicine*, *155*(17), 1877–1884.

Roter, D. L., Larson, S., Shinitzky, H., Chernoff, R., Serwint, J. R., Adamo, G., & Wissow, L. (2004). Use of an innovative video feedback technique to enhance communication skills training. *Medical Education*, *38*(2), 145–157.

Roter D.L., L. S. (2004). Use of an innovative video feedback technique to. *Medical Education*, *38*, 145–157.

Rubin, A. M., & Windahl, S. (1986). The uses and dependency model of mass communication. *Critical Studies in Mass Communication*, *3*(2), 184–199.

Rubin, R. B., & Martin, M. M. (1994). Development of a measure of interpersonal communication competence. *Communication Research Reports*, *11*(1), 33–44.

Rudovic, O., Nicolaou, M. A., & Pavlovic, V. (2017). Machine learning methods for social signal processing. *Social Signal Processing*, 234–254.

Ruiz, N., Chen, F., & Oviatt, S. (2010). Multimodal Input. *Multimodal Signal Processing*, 231–255.

Russell, J. A., & Dolls, J. M. F. (1997). *The Psychology of Facial Expression*. Cambridge: Cambridge University Press.

Russell, James A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, *115*(1), 102–141.

Russell, James A, & Pratt, G. (1980). A description of the affective quality attributed to environments. *Journal of Personality and Social Psychology*, *38*(2), 311.

Rutkin, A. (2014). Off the clock, on the record. *New Scientist*, *224*(2991), 22–23.

Saket, B., Yang, S., Tan, H., Yatani, K., & Edge, D. (2014). TalkZones. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services - MobileHCI '14* (pp. 263–272). New York, New York, USA: ACM Press.

Saks, A. M. (2002). So what is a Good Transfer of Training Estimate? A Reply to Fitzpatrick. *The Industrial Organizational Psychologist*, *39*(3), 29–30.

Saldaña, J. (2014). *The Coding Manual for Qualitative Researchers*. *Sage*.

Sanchez, C. A., & Wiley, J. (2006). An examination of the seductive details effect in terms of working memory capacity. *Memory and Cognition*, *34*(2), 344–355.

Sapouna, M., Wolke, D., Vannini, N., Watson, S., Woods, S., Schneider, W., … Aylett, R. (2009). Virtual Learning Intervention to Reduce Bullying Victimization in Primary School: A Controlled Trial. *Journal of Child Psychology and Psychiatry, 51*(1), 104-112.

Scerbo, A. S., Freedman, L. W., Raine, A., Dawson, M. E., & Venables, P. H. (1992). A Major Effect of Recording Site on Measurement of Electrodermal Activity. *Psychophysiology*, *29*(2), 241– 246.

Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, *69*(5), 379–399.

Scheflen, A. E. (1964). The Significance of Posture in Communication Systems. *Psychiatry*, *27*(4), 316–331.

Scherer, K. R., (1984). On the nature of function of emotion: A component process approach. *Approaches to Emotion, 2293(*317), 31.

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, *44*(4), 695–729.

Scherer, K. R., & Giles, H. (1979). *Social markers in speech*. Cambridge University Press.

Scherer, K., & R., K. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40*(1–2), 227–256.

Scherer, K., R., Schorr, A. & Johnstone, T. (2001). *Appraisal processes in emotion: Theory, methods and research*. Oxford University Press.

Scherl, C. R., & Haley, J. (2000). Computer Monitor Supervision: A Clinical Note. *The American Journal of Family Therapy*, *28*(3), 275–282.

Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is It Really Robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology*, *6*(4), 147–151.

Schneeberger, T. (2018). Transfer of Social Human-Human Interaction to Social Human-Agent Interaction. In *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems* (pp. 1778 - 1980). International Foundation for Autonomous Agents and Multiagent Systems.

Schneider, J., Börner, D., Rosmalen, P. van, & Specht, M. (2015). Presentation Trainer, your Public Speaking Multimodal Coach. *ICMI '15: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, 539–546.

Schneider, J., Borner, D., Van Rosmalen, P., & Specht, M. (2016). Can You Help Me with My Pitch? Studying a Tool for Real-Time Automated Feedback. *IEEE Transactions on Learning Technologies*, *9*(4), 318–327.

Schneider, J., Börner, D., van Rosmalen, P., & Specht, M. (2016). Enhancing public speaking skills - An evaluation of the presentation trainer in the wild. In *European Conference on Technolgoy Enhanced Learning* (pp. 623-275). Springer, Cham.

Schoorman, F. D., & Mayer, R. C. (2008). The Value of Common Perspectives in Self-Reported Appraisals. *Organizational Research Methods*, *11*(1), 148–159.

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A Unified Embedding for Face Recognition and Clustering.

Seyama, J., & Nagayama, R. S. (2007). The Uncanny Valley: Effect of Realism on the Impression of Artificial Human Faces. *Presence: Teleoperators and Virtual Environments*, *16*(4), 337–351.

Shannon, C. E., & Weaver, W. (1963). The mathematical theory of communication. *The University of Illinois Press*, 1–131.

Shao, A. T., Bao,Y., & Gray, E. (2004). Comparative advertising effectiveness: A cross-cultural study. Journal of Current Issues and Research in Advertising, 26, 67-80.

Shapiro, S. S., & Wilk, ; M B. (1965). *An Analysis of Variance Test for Normality (Complete Samples)*. *Biometrika*, *52*(3/4), 591-611.

Shettleworth, S. J. (2009). *Cognition, evolution, and behaviour*. Oxford University Press.

Skinner, B. F. (1938). *The Behaviour of Organisms: An Experimental Analysis*. New York, NY:

Appleton-Century.

Skinner, B. R., Gordin, R. D., & Ed, D. (2013). The Relationship Between Confidence and Performance Throughout a Competitive Season, 2117.

Smircich, L., & Chesser, R. J. (1981). Superiors' and Subordinates' Perceptions of Performance: Beyond Disagreement. *Academy of Management Journal*, *24*(1), 198–205.

Smith, C.A., & Kirby, L. D. (2000). Consequences require antecedents: Toward a process model of emotion elicitation. In J. P. Forgas (Ed.), Studies in emotion and social interaction, second series. Feeling and thinking: The role of affect in social cognition (p. 83 – 106). Cambridge University Press.

Smith-Lovin, L., Lewis, M., & Haviland, J. M. (1995). Handbook of Emotions. *Contemporary Sociology*, *24*(3), 298.

Smith, R. C., Marshall-Dorsey, A. A., Osborn, G. G., Shebroe, V., Lyles, J. S., Stoffelmayr, B. E., … Gardiner, J. C. (2000). Evidence-based guidelines for teaching patient-centered interviewing. *Patient Education and Counseling*, *39*(1), 27–36.

Sondhi, S., Khan, M., Vijay, R., & K. Salhan, A. (2015). Vocal Indicators of Emotional Stress. *International Journal of Computer Applications*, *122*(15), 38–43.

Song, Y., Demirdjian, D., & Davis, R. (2015). Continuous body and hand gesture recognition for natural human-computer interaction. *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, *2*(Ijcai), 1–28.

Spitzberg, B. H., Canary, D. J., & Cupach, W. R. (1994). A competence-based approach to the study of interpersonal conflict. In D. D. Cahn (Ed.), *Conflict in Personal Relationships* (pp. 183–202). Hillsdale, NJ, US: Lawrence Erlbaum Associates Inc.

Spitzberg, Brian H. (1991). An examination of trait measures of interpersonal competence. *Communication Reports*, *4*(1), 22–29.

Spitzberg, Brian H, & Adams, T. W. (2007). CSRS, the Conversational Skills Rating Scale: An Instructional Assessment of Interpersonal Competence. *NCA, National Communication Association*.

Stöckli, S., Schulte-Mecklenbeck, M., Borer, S., & Samson, A. C. (2017). Facial expression analysis with AFFDEX and FACET: A validation study. *Behavior Research Methods*, *50*(4), 1–15.

Stratou, G., & Morency, L. P. (2017). MultiSense-Context-Aware Nonverbal Behavior Analysis Framework: A Psychological Distress Use Case. *IEEE Transactions on Affective Computing*, *8*(2), 190–203.

Sullivan, J. (2018). The Primacy Effect in Impression Formation. *Social Psychological and Personality Science*, *10*(4), 432 - 439.

Sung, M., & Pentland, A. (2005). PokerMetrics: Stress and lie detection through non-invasive physiological sensing. *Tech.Rep., MIT Media Lab*.

Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, *12*(2), 257–285.

Tabacknick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using Multivariate Statistics* (5th ed.). Boston, MA: Pearson.

Takahashi, M., Tanaka, K., & Miyaoka, H. (2006). Reliability and validity of communication skills questionnaire (CSQ). *Psychiatry and Clinical Neurosciences*, *60*(2), 211–218.

Tan, C. S. S., Schöning, J., Luyten, K., & Coninx, K. (2014). Investigating the effects of using biofeedback as visual stress indicator during video-mediated collaboration. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI ’14* (pp. 71–80). New York, New York, USA: ACM Press.

Tanaka, H., Negoro, H., Iwasaka, H., & Nakamura, S. (2017). Embodied conversational agents for multimodal automated social skills training in people with autism spectrum disorders. *PLoS ONE*, *12*(8), e0182151.

Tanaka, H., Sakti, S., Neubig, G., Negoro, H., Iwasaka, H., & Nakamura, S. (2016). Automated social skills training with audiovisual information. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* (Vol. 2016Octob, pp. 2262–2265). IEEE.

Tanaka, H., Sakti, S., Neubig, G., Toda, T., & Nakamura, S. (2015a). Multimodal Computer-Based Training for Social and Communication Skills. *IEICE Transactions on Information and Systems*, *E98D*(8), 1536–1544.

Tanaka, H., Sakti, S., Neubig, G., Toda, T., & Nakamura, S. (2015b). NOCOA+: Multimodal computerbased training for social and communication skills. *IEICE Transactions on Information and Systems*, *E98D*(8), 1536–1544.

Tanveer, M. I. (2016). Understanding and Intervening Communicational Behavior using Artificial Intelligence (pp. 150–153).

Tanveer, M. I., Lin, E., & Hoque, M. (2015). Rhema: A Real-Time In-Situ Intelligent Interface to Help People with Public Speaking. In *Proceedings of the 20th International Conference on Intelligent User Interfaces,* pp. 286- 295. ACM.

Tao, J., & Tan, T. (2009). *Affective information processing*. *Affective Information Processing*. London: Springer.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55.

Taylor, R. (2015). *Media Interview Techniques: A complete guide ot media training*. London: Kogan Page.

Terven, J. R., Raducanu, B., Meza-de-Luna, M. E., & Salas, J. (2016). Head-gestures mirroring detection in dyadic social interactions with computer vision-based wearable devices. *Neurocomputing*, *175*, 866–876.

Teshima, T., Xu, M., Sato, I., & Sugiyama, M. (2019). Clipped Matrix Completion: A Remedy for Ceiling Effects. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*, 5151– 5158.

Thomson, M. M., Turner, J. E., & Nietfeld, J. L. (2012). A typological approach to investigate the teaching career decision: Motivations and beliefs about teaching of prospective teacher candidates. *Teaching and Teacher Education*, *28*(3), 324–335.

Thornton, A., Sutton, R., Stanfield, C., & Leach, A. (2016). *The UK Commission's Employer Skills Survey 2015: UK Results*.

Tomarken, A. J., & Serlin, R. C. (1986). Comparison of ANOVA alternatives under variance heterogeneity and specific noncentrality structures. *Psychological Bulletin*, *99*(1), 90–99.

Tomkins, S. S. (1981). The role of facial response in the experience of emotion: A reply to Tourangeau and Ellsworth. *Journal of Personality and Social Psychology*, *40*(2), 355–357.

Tracy, J. L., Randles, D., Steckler, C. M., Crockett, M. J., & Cuddy, A. (2015). The nonverbal communication of emotions. *Current Opinion in Behavioral Sciences*, *3*, 25–30.

Trujillo Fernández, R. (2017). *Faculty of Arts Master in International Business Communication: The Effect of High/Low Context Style in CSR Communication on Spanish/Dutch*.

Tullis, T. S., & Stetson, J. N. (2004). A comparison of questionnaires for assessing website usability. *Usability Professional Association Conference*, 1–12.

Turk, M. (2014). Multimodal interaction: A review. *Pattern Recognition Letters*, *36*(1), 189–195.

Uddin, L. Q., Iacoboni, M., Lange, C., & Keenan, J. P. (2007). The self and social cognition: the role of cortical midline structures and mirror neurons. *Trends in Cognitive Sciences*, *11*(4), 153– 157.

Usui, T., Kume, K., Yamano, M., & Hashimoto, M. (2008). A robotic KANSEI communication system based on emotional synchronization. In *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS* (pp. 3344–3349). IEEE.

Utgoff, P. E., Cussens, J., Kramer, S., Jain, S., Stephan, F., Raedt, L. De, … Ng, A. Y. (2011). Instance-Based Learning. In *Encyclopedia of Machine Learning* (pp. 549–550). Boston, MA: Springer US.

Van den Stock, J., Righart, R., & de Gelder, B. (2007). Body Expressions Influence Recognition of Emotions in the Face and Voice. *Emotion*, *7*(3), 487–494.

van der Maaten, L., & Hendriks, E. (2012). Action unit classification using active appearance models and conditional random fields. *Cognitive Processing*, *13*(S2), 507–518.

van Lierde, K., van Heule, S., De Ley, S., Mertens, E., & Claeys, S. (2009). Effect of Psychological Stress on Female Vocal Quality. *Folia Phoniatrica et Logopaedica*, *61*(2), 105–111.

Vinciarelli, A., & Odobez, J.-M. (2006). Application of Information Retrieval Technologies to Presentation Slides. *IEEE Transactions on Multimedia*, *8*(5), 981–995.

Vinciarelli, A., Pantic, M., Heylen, D., Pelachaud, C., Poggi, I., D'Errico, F., & Schroeder, M. (2011). Bridging the Gap Between Social Animal and\nUnsocial Machine: A Survey of Social Signal\nProcessing. *IEEE Transactions on Affective Computing (to Appear)*, *3*(1), 1–20.

Vinciarelli, Alessandro, Pantic, M., & Bourlard, H. (2009a). Social signal processing: Survey of an emergin domain. *Image and Vision Computing*, *27*(November 2008), 1743–1759.

Vinciarelli, Alessandro, Pantic, M., & Bourlard, H. (2009b). Social signal processing: Survey of an emerging domain. *Image and Vision Computing*, *27*(12), 1743–1759.

Vinciarelli, Alessandro, Pantic, M., Heylen, D., Pelachaud, C., Poggi, I., D'Errico, F., & Schröder, M. (2012). Bridging the gap between social animal and unsocial machine: A survey of social signal processing. *IEEE Transactions on Affective Computing*, *3*(1), 69–87.

Vinciarelli, Alessandro, & Pentland, A. (Sandy). (2015). New Social Signals in a New Interaction World. *IEEE Systems, Man, and Cybernetics Magazine*, *1*(2), 10–17.

Virzi, R. A. (1992). Refining the Test Phase of Usability Evaluation: How Many Subjects Is Enough? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *34*(4), 457–468.

Vrij, A., Edward, K., & Roberts, K. P. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Ray Bull Journal of Nonverbal Behavior Winter*, *24*(4).

Warner-Søderholm, G. (2013). Beyond a Literature Review of Hall's Context Dimension: Scale Development, Validation & Empirical Findings within a Norwegian Study. *International Journal of Business and Management*, *8*(10).

Watson, J. B. (1930). *Behaviourism (Revised Edition)*. Chicago: University of Chicago Press.

Weiner, B., & Graham, S. (1984). An Approach to Emotional Development. *Emotions, Cognition, and Behaviour*, 167–191.

Weisfeld, G. E., & Beresford, J. M. (1982). Erectness of posture as an indicator of dominance or success in humans. *Motivation and Emotion*, *6*(2), 113–131.

Whitmarsh, L., & O'Niell, S. (2010). *Engaging the public with climate change: Communication and behaviour change and communication*. Routledge

Wicklund, R. A. (1979). The Influence of Self-Awareness on Human Behavior: The person who becomes self-aware is more likely to act consistently, be faithful to societal norms, and give accurate reports about himself. *American Scientists*, *67*(2), 187–193.

Wiemann, J. (1977). Explication and Test of a Model of Communicative Competence. *Human Communication Research*, *3*(3), 195–213.

Wiseman, J., Parry, E., & Baker, G. (2015). *Business Survey*. Luton.

Witten, I., & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques* (2nd ed.). Burlington, MA: Morgan Kaufmann: Elsevier.

Witten, I. H., Frank, E., & Hall, M. a. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. *Complementary literature None*.

Wobbrock, J., & Kientz, J. (2016). Research Contributions in Human-Computer Interaction. *Interactions*, 39–44.

Woods, D. D. (1991). The Cognitive Engineering of Problem Representations. In G. R. S. Weir & J. L. Alty (Eds.), *Human-Computer Interaction and Complex Systems* (pp. 169–188). London: Academic.

World Health Organization, & WHO. (2005). Effective media communication during public health emergencies - A WHO Handbook. *World Health*, (July).

World Medical Association. (2013). World Medical Association Declaration of Helsinki. *JAMA*, *310*(20), 2191.

Xia, C., & Maes, P. (2013). The design of artifacts for augmenting intellect. In *Proceedings of the 4th Augmented Human International Conference on - AH '13* (pp. 154–161). New York, New York, USA: ACM Press.

Xu, H., & Plataniotis, K. N. (2016). Affective states classification using EEG and semi-supervised deep learning approaches. *2016 IEEE 18th International Workshop on Multimedia Signal Processing (MMSP)*, 1–6.

Yang, Z., Metallinou, A., & Narayanan, S. (2014). Analysis and predictive modeling of body language behavior in dyadic interactions from multimodal interlocutor cues. *IEEE Transactions on Multimedia*, *16*(6), 1766–1778. Yelon, S., Sheppard, L., Sleight, D., & Ford, J. K. (2004). Intention to transfer: How do autonomous professionals become motivated to use new ideas? *Performance Improvement Quarterly*, *17*(2), 82–103.

Yule, G. (1996). *Pragmatics*. Oxford: Oxford University Press.

Zancanaro, M., Lepri, B., & Pianesi, F. (2006). Automatic detection of group functional roles in face to face interactions. In *Proceedings of the 8th international conference on Multimodal interfaces - ICMI '06* (p. 28). New York, New York, USA: ACM Press.

Zhang, Y., Olenick, J., Chang, C.-H., Kozlowski, S. W. J., & Hung, H. (2018). The I in Team: Mining Personal Social Interaction Routine with Topic Models from Long-Term Team Data.

Zhang, Y., Palo Alto Laboratory, F., Jeffrey Olenick, U., J Kozlowski, S. W., Olenick, J., Chang, C.H., & Hung, H. (2018). TeamSense: Assessing Personal Affect and Group Cohesion in Small Teams through Dyadic Interaction and Behavior Analysis with Wearable Sensors. *Proceedings of the ACM Interact. Mob. Wearable Ubiquitous Technology*, *2*(150), 150.

Zhang, Z., Luo, P., Loy, C. C., & Tang, X. (2018). From Facial Expression Recognition to Interpersonal Relation Prediction. *International Journal of Computer Vision*, *126*(5), 550–569.

Zhao, R., Li, V., Barbosa, H., Ghoshal, G., & Hoque, M. E. (2017). Semi-Automated 8 Collaborative Online Training Module for Improving Communication Skills. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(2), 1–20.

Žukauskas, P., Vveinhardt, J., & Andriukaitienė, R. (2018). Philosophy and Paradigm of Scientific Research. In *Management Culture and Corporate Social Responsibility*. InTech.

Appendix

Appendix 1

Research Contributions

Year	Title	Author	Type	Disseminated
1	Brunel Poster Conference	Monica Pereira	Conference: Poster	Brunel Postgraduate Event
1	3 Minute Thesis Finalist	Monica Pereira	Presentation	Brunel Postgraduate Event
2	Human Emotion Recognition Ethics Workshop	Helen Dudfield	Symposium: Workshop	Funder Event
2	User Centre Design of Social Signals Feedback for Communication Skills Training	Monica Pereira Federico Colecchia Kate Hone	Conference: Affective Computing Workshop (Double Blind Peer Reviewed Conference)	Proceedings of 32nd British HCI Annual Conference, Belfast, 2018
2	Media Skills Training Intervention Based on Automated Recognition of Human Emotion and Nonverbal Behaviour	Monica Pereira	Conference: Doctoral Consortium (Double Blind Peer Reviewed Conference)	Proceedings of 32nd British HCI Annual Conference, Belfast, 2018
3	Communication Skills Training Intervention Based on Automated Recognition of Human Emotion and Nonverbal Behaviour	Monica Pereira and Kate Hone	Computer Science Brunel PhD Symposium: Award for best Extended Abstract And Best PhD overall	Brunel Postgraduate Event
3	Prediction of Culture Based on Automated Detection of Multimodal Social Signals	Monica Pereira and Kate Hone	Computer Science Brunel PhD Symposium: Poster	Brunel Postgraduate Event
3	Communication Skills Training Intervention Based on Automated Recognition of Human Emotion and Nonverbal Behaviour	Monica Pereira and Kate Hone	Conference: Poster Award for best poster in college	Brunel Postgraduate Event
3	Communication Skills Training Intervention Based on Automated Recognition of Human Emotion and Nonverbal Behaviour	Monica Pereira and Kate Hone	Conference: Poster Award for best PhD research and presentation	British Computer Society
3	Detection of Social Signals During Communication in Media Skills Training using Commercial Automatic Emotion Recognition Systems	Monica Pereira, Hongying Meng and Kate Hone	Peer-reviewed Journal: Computers in Human Behaviour	Review

Appendix 2

Appendix 2.1. AU Descriptions and Associated Facial Muscles

Action Units Descriptions and Facial Muscles

Action Unit	Description	Facial Muscle
1	Inner Brow Raiser	Frontalis Pars Medialis
2	Outer Brow Raiser	Frontalis, Pars Lateralis
4	Brow Lowerer	Depressor Glabellae, Depressor Supercilli, Currugataor
5	Upper Lid Raiser	Levator Palpebrae Superioris
6	Cheek Raiser	Orbicularis Oculi, Pars Orbitalis
7	Lid Tightener	Orbicularis Oculi, Pars Palpebralis
9	Nose Wrinkler	Levator Labii Superioris Alaquae Nasi
10	Upper Lip Raiser	Levator Labii Superioris Caput Infraorbitalis
11	Nosolabial Deepender	Zygomatic Minor
12	Lip Corner Puller	Zygomatic Major
13	Cheek Puffer	Levator Anguliorís (Caninus)
14	Dimpler	Bucccinator
15	Lip Corner Depressor	Depressor Anguli Oris (Triangularis)
16	Lower Lip Depressor	Depressor Labii Inferioris
17	Chin Raiser	Mentalis
18	Lip Puckerer	Incisivii Labii Superioris and Incisivii Labii Inferioris
20	Lip Stretcher	Risorius
22 (with AU25)	Lip Funneler	Orbicularis Oris
23	Lid Tightener	Orbicularis Oris
24	Lip Pressor	Orbicularis Oris
25	Lips Part	Depressor Labii, Relationof Mentalis (AU17), Orbicularis Oris
26	Jaw Drop	Massetter; Temporal and Internal Pterygoid Relaxed
27	Mouth Stretch	Pterygoid, Digastric
28 (with AU26)	Lip Suck	Orbicularis Oris
41	Lid Droop	Relation of Flevator Palpebrae Superioris
42	Slit	Orbicularis Oculi
43	Eyes Closed	Relation of Levator Palpebrae Superioris
44	Squint	Orbicularis Oculi, Pars Palpebralis
45	Blink	Relation of Levator Palpebrae and Contraction of Orbicularis Oculi, Pars Palpebralis.
46	Wink	Levator Palpebrae Superioris; Orbicularis Oculi, Pars Palpebralis

Facial Expression and their corresponding action units

Emotion	Action Units	Description
Joy	6 + 12	Cheek raiser, Lip Corner Puller
Sadness	1 + 4 + 15	Inner Brow Raiser, Brow Lowerer, Lip Corner Depressor
Surprise	1 + 2 + 5 + 26	Inner Brow Raiser, Outer Brow Raiser, Upper lid Raiser, Jaw Drop
Fear	1 + 2 + 4 + 5 + 7 + 20 + 26	Inner Brow Raiser, Outer Brow Raiser, Brow Lowerer, Upper Lid Raiser, Lid Tightener, Lip Stretcher, Jaw Drop
Anger	4 + 5 + 7 + 23	Brow Lowerer, Upper Lid Raiser, Lid Tightener, Lip Tightener
Disgust	9 + 15 + 16	Nose Wrinkler, Lip Corner Depressor, Lower Lip Depressor
Contempt	12 + 14 (on one side of the face)	Lip Corner Puller, Dimpler

Appendix 2.2.

Author	Aim	Participant and context	Signals	Tech	Q's	Feedback Design	Feedback	Qualitative methods and raters	Procedure	Conclusions	Limitations and future work
Damian, Baur and Andre (2016)	To explore the concept of automatic behaviour feedback loops and compare between different feedback methods spanning multiple modalities	54 conversation	Speaking duration, speech rate, loudness and pitch information	Praat	Distraction questionnaire	Myo armband (tactile feedback), Aftershokz Bluez 2S bone conduction headphones (auditory feedback), Google glass (visual feedback), Microsoft surface 2 Pro (remote feedback). None vs automated	Real-time - behavioural feedback loop. Length of time left to speak	N/A	First discussion - control group BFL inactive. Feedback groups were active before second discussion group. P's instructed to click when noticed one peer receiving feedback.	BFL can improve behaviour during social interactions with some distraction	Disturbing. Explore different feedback strategies for each modality. Other dynamic methods of feedback.
Fung, Jin, Zhao and	To determine helpfulness and accuracy of machine	23 public speaking	Smiles, movement and volume	Shore Framework, Kinect, Praat,	Evaluation of own performance	Automated - graphs on smiles, body movement	Video and automated feedback only. Overall,	Interviews about experience with system	P's given joke to memorise. After narrating joke, one	Automated feedback was adding more value than solely watching	Features need refining, safe feedback on social media rather than Turkers.

Hoque (2015)	generated automated feedback.		modulation, word prosody	Google web speech and nuance speech recognition SDK		and volume modulation along with own videos but no subjective interpretation. Video only - only watched their video.	body gestures, volume modulation, friendliness. Ranked comments		group saw own video recording and the automated feedback after. Then interacted with feedback and completed surveys.	one's own video, context and personalisation added value to automated feedback	Support of Kinect, social media use for comments. Lab or school to unobtrusively impact mental health and behavioural assessment Features
	Explored whether non-expert Turkers would be viable source of helpful and accurate personalisation to the automated feedback. ROC Speak	20 Public speaking			Evaluation of system	After experiment recorded by system, p's received a feedback page via email with subjective ratings by Turkers and automated feedback. Helpfulness and accuracy	Overall body gestures, volume modulation, friendliness. Ranked comments	20 Turkers (10 watched videos; 10 rate helpfulness of the comments generated by the first)	P asked to memorise a joke and retell in front of a camera. Video recordings were sent to Turkers. Feedback sent to p's via email. Instructed to interact for 15 minutes with feedback only.	Graph are helpful for users, context specific feedback and aiding understanding of own nonverbal behaviour. Turks possible to generate helpful commentary and personalised feedback.	

	Researchers trained machine learning algorithms to automate classification. Also, evaluate improved user interface added value. ROC Speak	16 Public speaking			Questionnaire to evaluate public speaking experience. Post survey	Video group: observe video with general tips. ROC speak see interface with same general points	Overall body gestures, volume modulation, friendliness. Ranked comments	Turkers ratings	Pre-study survey. 2 min speech recorded. Rated by other Turks. Only interact with feedback for 10 -15 minutes. Post-study survey	Developed a framework for receiving personalised feedback on nonverbal behaviour.	
Hoque, Courgeon, Martin, Mutlu & Picard (2013)	Better understand how expert interviewers facilitate mock interview. MACH	90 Job interviews	Smile, head movement, voice prosody and speech recognition	Shore framework, Kawato and Ohya algorithm, Praat, Nuance	2 eval questions	Post-hoc.1) no feedback, 2) video only and 3) Summary feedback and, optional, focused feedback. 1) control group - watched educational video on interviewing for jobs.	**<u>Summary feedback:</u>** smile, pause, duration, speaking rate, weak language and pitch variation. **<u>Focused feedback:</u>** watch video (watch own video with	Open ended feedback on experience. Counsellors (blind to group condition)	All p's interview with professional counsellor. G2 and G3 brought into lab a few days after interview for an hour-long intervention. All p's brought back for	MACH enabled learning something new about behaviours and agreed to use in future.	None mentioned but could have implemented longer duration of study. Improve on the ubiquitous nature of MACH by extending implementation on to mobile platforms. **<u>Compare to past performance</u>**

						2) practiced interviews and watched themselves on video. 3) Practiced interviews with MACH and received feedback	informati on)		additional interview.		**<u>through progress charts.</u>**
Liu, Scott, Lim, Taylor and Calvo (2016)	1) provide medical students with opportuniti es to communica te with simulated patients via tele-consultatio n and an easy means of organising consultatio ns 2) provide video recordings of consultatio	8 Doctor-patient	Volume, pitch, turn taking patterns and speaking ratio, nodding, head shaking, smiling, frowning , head tilting and face-touch gestures.	EQClinic	Student-patient observed communica tion assessment form. Confidence , Reflection, System usability	Real time. Automated and simulated patient feedback. 1) two consultation s, 2) one consultation	Volume, pitch, turn talking patterns and speaking ratio, nodding, head shaking, smiling, frowning, head tilting and face-touch gestures.	Tutor, simulate d patient, self-assessme nt	Students asked to conduct two consultatio ns at least 3 days apart. 15-minute interview. SP assessed performanc e of student and Student evaluated own performanc e. System took 24 hours to generate nonverbal behaviour feedback	EQClinic provides an innovative solution for providing medical students with a means to practise and enhance their communica tion skills	No calibration procedure during SP training to ensure inter-rater agreement of assessments. Students were from different study years. System can only detect and evaluate non-verbal behaviour. Nurse education, help health professionals to reflect on and develop

	ns with feedback 3) automatically identify student's non-verbal behaviour								with SP comments. They then filled out system usability. Tutor then assessed student performance.		communication skills
Damian, Tan, Baur, Schoning, Luyten and Andre (2015)	Focus on collection of questionnaire data and measurement of social signals	15 Public speaking	Speech rate, energy and openness	Kinect, Vuzix HMD, Head worn microphone	Self-rating of previous experience and how skilful they think they are.	BFL. Logue feedback vs no feedback. Control condition - wore the whole system but feedback system deactivated. Experiment condition - wore system with feedback about nonverbal signals on.	Speech rate, energy and openness	Observers	Q filled in. 5-minute presentation to observers who were blind to condition. Two sessions two weeks apart and order of conditions randomised.	Logue provides users with behavioural feedback on two levels. 1) informs current state of speech rate, body energy and openness. 2) indicates quality of behavioural cues in relation to public speaking context.	Disturbing, weight and size. Conduct more in-depth studies to accurately determine which configuration of feedback classes and thresholds are representative for good and bad speaking behaviour
	Test Logue in a real presentation setting	3 Public speaking			N/A	Logue only (3 p's)		Semi-structured interview	After talk, an open discussion on the style		

								s with participants. 13 observers	of presentation occurred. Audience was asked questions about quality of talk or whether the system influenced the quality of presentation.		
Tanveer, Lin and Hoque (2015)	Maximise the usefulness of our system by minimising the level of distraction	30 Public speaking	Volume and speaking rate	Google glass, PRAAT	Speaking performance after speaking. Efficacy, learnability and future use.	Real-time. Google topics and feedback schemes were counter balanced to remove any ordering effects. 1) Continuous streams of information vs 2) sparse delivery of information	Volume and speaking rate	Post-study interview. 10 Amazon Turk Workers	Topic decided 2 days before. Presentation, survey after, then videos sent to Turkers to investigate distraction. Post-study interview - Effective reminder. Preferred word	Interface valuable to speaking performance. Word feedback more effective. No clear results because of low agreement among Turkers.	No post-speech feedback to user. Present information sporadically through secondary display. Explore haptic feedback schemes

						vs 3) no feedback. Words vs bars			feedback than bars		
Schneider, Borner, van Rosmalen and Specht (2015)	Explore Presentation trainer	40 Public speaking	body posture, use of gestures, voice volume, use of pauses, use of phonetic pauses and ability to stay grounded without shifting weight from one foot to another that resembles dancing.	Kinect	Evaluate user experience	Real-time. 1)Interruptive or 2) corruptive feedback is triggered whenever a mistake is considered severe. CC practiced using a version of PT whose interface only shows a mirrored image of user (no feedback). EC received both interruptive and immediate feedback.	Visual feedback on which signal is incorrect on PT interface and haptic feedback is through a wristwatch	N/A	5 min lecture about nonverbal communication for public speaking. 5 min lecture about elevator pitches. 5 min to create own elevator pitch. Practice in 5 successive training sessions.	PT supported users with development of public speaking skills by helping to improve performance	Performance measurement cannot be directly translated to the assessment that a human would make about quality of speech. Does not recognise differences in gestures. Expert study to extract rich set of nonverbal communication aspects and rules that influence quality of a presentation.
Batrinca, Stratou, Shapiro	Evaluate use of audience as means of	14 Public speaking	Voice (flow of speech, clear	Microsoft Kinect, webcams	Big five personality questionnaire, personal	N/A - Audience?	Audience displayed: posture, head	Toast makers	Presentation 5 - 15 minutes to audience.	Reveal several expert estimates of	Behavioural descriptors are crude and abstract.

, Morency and Scherer (2013)	feedback by presence		intonation, interrupted speech, speaks too quietly, vocal variety), Body (paces too much, gestures to emphasize, gestures too much) and gaze (gazes at audience, avoids audience)		report of confidence, self-statements during public speaking, Positive and negative affect schedule		orientation and eye gaze.		Experts watched videos once and rated performance.	nonverbal behaviours, identify automatic behaviour descriptors that correlate strongly with expert estimates of nonverbal behaviours, matching nonverbal signals with overall performance ratings	Expand on cicero to incorporate reactive audience
Chollet, Wortwein, Morency, Shapiro and	Explore feedback strategies for public speaking training based on an interactive	47 Public speaking	Eye contact and avoidance of pause filters	Microsoft Kinect, webcams	Evaluation of self-assessment, experiment assessment and two objectively	Real-time. Pre-post training paradigm. non-interactive (control), direct visual	avoiding pause fillers and gaze behaviour	Evaluation and Toast makers	P's completed questionnaires about public speaking anxiety and self-	Presenters enjoyed interacting with public speaking skills with virtual audience.	Not multimodal, not naturalistic, cultural error in interpretation. Compare

Scherer (2016)	virtual audience paradigm				annotated measures	feedback and nonverbal feedback			assessment. Each p gave 4 presentations	More engaging, captivating and challenging overall. Experts identified consistent improvement of skill from pre - post in both control and interactive audience conditions. Contact and avoid pause fillers improvement regardless of training. Virtual audience can act as an effective platform to improve public speaking and regulate anxiety.	different levels and types of feedback. Other audience behaviours such as yawning or falling asleep. Cultural error. Longitudinal. Natural characters.

Damian, Baur, Lugrin, Gebhard, Mehlmann and Andre (2015)	Evaluate a virtual job interview training game which has been adapted to special requirement so of young people with low chances on the job market.	20 Job interviews	Nodding, head tilting, eye contact	Kinect	Practitioners assessed: overall performance, recommend for employment, appropriate use of smiles, eye contact, gestures, nervousness, interested and focused. Pupils rated themselves on performance, nervous, filler words, focuses, aware of nonverbal, appropriate nonverbal.	Real-time. Random divided to EG or CG. 4 females and 6 males in EG and 5 females and four males in CG. Conventional vs game training	Nodding, head tilting, eye contact. Welcome, company presentation and strengths and weaknesses	Practitioner	Mock job interview - training over three days - mock interview	Benefits of computer-based job training systems for underprivileged pupils.	Not multimodal. Job interview training on large scale.
Schneider, Borner, van Rosmal	Explored the use of a multimodal learning application	9 Public speaking	Posture, gesture, volume, cadence, phonetic	Kinect	User experience questionnaire and presentatio	Real-time feedback. Session 1	body posture, use of gestures, voice	Tutor and peers and PT	Project presentation, feedback from peers and tutors,	Room for improvement	Kinect not owned by many students. Courses,

en and Specht (2016)	called the presentation trainer		pauses, dancing, blank face)		n evaluation	and two comparison	volume, phonetic pauses, filler sounds, use of pauses and facial expression.		homework prep. Pitch to audience-evaluation, feedback briefing, two practice sessions, second pitch to audience, evaluation and then user experience questionnaire		improving PT.
Barmaki and Hughes (2018)	Evaluate feedback application of real-time gesture recognition and application	30 Teaching	Gestures and postures	Kinect	Demographics, teaching experience, User experience and learning perception	Real-time. 2x 1x2 counterbalanced within subjects' study. Setting 1 (no visual feedback) setting 2 (visual feedback)	. Gestures and postures. Visual (traffic light model: green to red when closed gesture)	Peers and Tutors	Intro, pre-question, setting 1, post questionnaire-setting 2, post questionnaire. / setting 2, post questionnaire, setting 1, post questionnaire	Effective improving skills	Recruitment process - not able to do counterbalancing. Future work required to gain complete understanding of the embodiment in teaching using immediate

		3 Teaching	Gestures and postures			Real-time. 2x1x1 no counter balancing. Setting 1 (no visual feedback) setting 2 (visual and haptic l feedback)	Haptic (vibration) and visual (traffic light model: green to red when closed gesture)	N/A	All p's experienced both settings	Mere exposure suggests effectivenes s	feedback in training sessions. Could also be useful in police de-escalation training
Bahrein i, Nadolsk i and Webster (2017)	Investigates to what extent FILTWAM facial and vocal emotion recognition software can be used for improving serious game that delivers web-based training of communica tion skills	25 Gaming	Facial and vocal emotions	FILTWA M	Evaluations	Real-time. Within-subjects (repeated measure) with two experiment conditions. Direct feedback and no feedback	Correctne ss and incorrectn ess of expressio n	2 human raters	120 minutes. Breaks to avoid fatigue. Intro, face calibration. No voice needed. Game. Then fill out questionnai re	Suggest automated feedback improves learner's communica tion performanc es	Accuracy improved by combining sensory data. More real-time data. Improved feedback mechanism. More ecologically valid circumstance s. More detailed emotions
Ali & Hoque (2017)	Envision a virtual assistant that can	55 Speed dating	Smiles, eye contact, body	Praat and Shore, Visage.	CSRS	Real-time. Between. Treatment: practice	Smiles, eye contact, body	Research assistants .	After interactions (treatment) and video	System allows individual to practice	Eliminate flashing icon. Build more natural

	give individuals real-time feedback on their smiles, eye contact, body language and volume modulation that is available anytime, anywhere using a computer browser.		language, volume modulation			conversation. Control group read a pamphlet and watched YouTube videos	language, volume modulation. Green good and red bad		watch) the RA rated conversation skills.	conversation anytime and anywhere using a web browser.	system. Behaviour entrainment. Different population
Zhao and colleagues (2017)	Present a training module to aid in speaking skills for participants whenever they want on a web browser.	56 Public speaking	smile intensity, body movement, volume modulation, filler word usage, word cloud of spoke	ROC-Speak	User assessment, assessment by independent judges, self-assessment	Real-time. Between. Treatment (full feedback), control (written feedback by peers)	smile intensity, body movement, volume modulation, filler word usage, word cloud of spoken words and peer comments	Peers and Turkers raters	Introduce study, allowed to explore interface. 10 days with 5 prompts released every two days. Set up encouraged participants to prepare for presentatio	feedback improved more than control group	Sensitive to environmental factors, requires Kinect, google glass that people don't have access to. Incorporation of head nodding and hand waving. Longer study to assess skill maintenance

									ns and improve.		
Tanveer, Zhao, Chen, Tiet and Hoque (2016)	To present an intelligent interface that can automatically extract human gestures	27 Public speaking	Body pose	Kinect	User evaluation	Real-time. Exposed to all. Placebo vs treatment	Body movements	Interviews about experience with system. Self-rating and mechanical Turk	Speech 1 -> feedback (treatment vs placebo), Speech 2 -> feedback (treatment vs placebo), Speech 3 -> self-rating	Speakers liked interface as they became aware of body language	Participants unable to differentiate between placebo and treatment. Unsupervised approach with speaking experts. Longitudinal

Appendix 3

Appendix 3.1. Conversation Skills Rating Scale - Self-Report

Please rate how you felt you performed in the interview you have just given using the following scale: [Note the original Conversational Skills Rating Scale is shown here for illustration but will be adapted for use in the study by removing the name fields]

CONVERSATIONAL SKILLS RATING SCALE (Rating of Self Form)

Your Name:	Partner Name:
Your ID:	Partner ID:

Date:	Class:	Activity:

Rate how skillfully **YOU** used, or didn't use, the following communicative behaviors in the conversation, where:

1	=	**INADEQUATE**	(use is awkward, disruptive, or results in a negative impression of communicative skills)
2	=	**FAIR**	(occasionally awkward or disruptive, occasionally adequate)
3	=	**ADEQUATE**	(sufficient but neither noticeable nor excellent. Produces neither strong positive nor negative impression)
4	=	**GOOD**	(use was better than adequate but not outstanding)
5	=	**EXCELLENT**	(use is smooth, controlled, results in positive impression of communicative skills)

Circle the single most accurate response for each behavior:

1	2	3	4	5	=	(1)	Speaking rate (neither too slow nor too fast)
1	2	3	4	5	=	(2)	Speaking fluency (pauses, silences, "uh", etc.)
1	2	3	4	5	=	(3)	Vocal confidence (neither too tense/nervous nor overly confident sounding)
1	2	3	4	5	=	(4)	Articulation (clarity of pronunciation and linguistic expression)
1	2	3	4	5	=	(5)	Vocal variety (neither overly monotone nor dramatic voice)
1	2	3	4	5	=	(6)	Volume (neither too loud nor too soft)
1	2	3	4	5	=	(7)	Posture (neither too closed/formal nor too open/informal)
1	2	3	4	5	=	(8)	Lean toward partner (neither too forward nor too far back)
1	2	3	4	5	=	(9)	Shaking or nervous twitches (aren't noticeable or distracting)
1	2	3	4	5	=	(10)	Unmotivated movements (tapping feet, fingers, hair-twirling, etc.)
1	2	3	4	5	=	(11)	Facial expressiveness (neither blank nor exaggerated)
1	2	3	4	5	=	(12)	Nodding of head in response to partner statements
1	2	3	4	5	=	(13)	Use of gestures to emphasize what is being said
1	2	3	4	5	=	(14)	Use of humor and/or stories
1	2	3	4	5	=	(15)	Smiling and/or laughing
1	2	3	4	5	=	(16)	Use of eye contact
1	2	3	4	5	=	(17)	Asking of questions
1	2	3	4	5	=	(18)	Speaking about partner (involvement of partner as a topic of conversation)
1	2	3	4	5	=	(19)	Speaking about self (neither too much nor too little)
1	2	3	4	5	=	(20)	Encouragements or agreements (encouragement of partner to talk)
1	2	3	4	5	=	(21)	Personal opinion expression (neither too passive nor aggressive)
1	2	3	4	5	=	(22)	Initiation of new topics
1	2	3	4	5	=	(23)	Maintenance of topics and follow-up comments
1	2	3	4	5	=	(24)	Interruption of partner speaking turns
1	2	3	4	5	=	(25)	Use of time speaking relative to partner

For the next five items, rate your overall performance. I was a(n)...

POOR CONVERSATONALIST ::	1	2	3	4	5	6	7	: GOOD CONVERSATIONALIST
SOCIALLY UNSKILLED ::	1	2	3	4	5	6	7	: SOCIALLY SKILLED
INCOMPETENT COMMUNICATOR ::	1	2	3	4	5	6	7	: COMPETENT COMMUNICATOR
INAPPROPRIATE COMMUNICATOR ::	1	2	3	4	5	6	7	: APPROPRIATE COMMUNICATOR
INEFFECTIVE COMMUNICATOR ::	1	2	3	4	5	6	7	: EFFECTIVE COMMUNICATOR

Comments:

Appendix 3.2. Conversation Skills Rating Scale - Trainer and Neutral Observer

The version of the CSRS used by trainers and experimenters is shown here for illustration (again the fields requiring a name will be removed for the purposes of the study).

CONVERSATIONAL SKILLS RATING SCALE (Observer Rating of Conversant Form)

Your Name:		Partner Name:
Your ID:		Partner ID:
Date:	Class:	Activity:

Rate how skillfully **THIS INTERACTANT** used, or didn't use, the following communicative behaviors in the conversation, where:

1	=	**INADEQUATE**	(use is awkward, disruptive, or results in a negative impression of communicative skills)
2	=	**FAIR**	(occasionally awkward or disruptive, occasionally adequate)
3	=	**ADEQUATE**	(sufficient but neither noticeable nor excellent. Produces neither strong positive nor negative impression)
4	=	**GOOD**	(use was better than adequate but not outstanding)
5	=	**EXCELLENT**	(use is smooth, controlled, results in positive impression of communicative skills)

Circle the single most accurate response for each behavior:

1	2	3	4	5	=	(1)	Speaking rate (neither too slow nor too fast)
1	2	3	4	5	=	(2)	Speaking fluency (pauses, silences, "uh", etc.)
1	2	3	4	5	=	(3)	Vocal confidence (neither too tense/nervous nor overly confident sounding)
1	2	3	4	5	=	(4)	Articulation (clarity of pronunciation and linguistic expression)
1	2	3	4	5	=	(5)	Vocal variety (neither overly monotone nor dramatic voice)
1	2	3	4	5	=	(6)	Volume (neither too loud nor too soft)
1	2	3	4	5	=	(7)	Posture (neither too closed/formal nor too open/informal)
1	2	3	4	5	=	(8)	Lean toward partner (neither too forward nor too far back)
1	2	3	4	5	=	(9)	Shaking or nervous twitches (aren't noticeable or distracting)
1	2	3	4	5	=	(10)	Unmotivated movements (tapping feet, fingers, hair-twirling, etc.)
1	2	3	4	5	=	(11)	Facial expressiveness (neither blank nor exaggerated)
1	2	3	4	5	=	(12)	Nodding of head in response to partner statements
1	2	3	4	5	=	(13)	Use of gestures to emphasize what is being said
1	2	3	4	5	=	(14)	Use of humor and/or stories
1	2	3	4	5	=	(15)	Smiling and/or laughing
1	2	3	4	5	=	(16)	Use of eye contact
1	2	3	4	5	=	(17)	Asking of questions
1	2	3	4	5	=	(18)	Speaking about partner (involvement of partner as a topic of conversation)
1	2	3	4	5	=	(19)	Speaking about self (neither too much nor too little)
1	2	3	4	5	=	(20)	Encouragements or agreements (encouragement of partner to talk)
1	2	3	4	5	=	(21)	Personal opinion expression (neither too passive nor aggressive)
1	2	3	4	5	=	(22)	Initiation of new topics
1	2	3	4	5	=	(23)	Maintenance of topics and follow-up comments
1	2	3	4	5	=	(24)	Interruption of partner speaking turns
1	2	3	4	5	=	(25)	Use of time speaking relative to partner

For the next five items, rate this person's overall performance:

POOR CONVERSATONALIST ::	1	2	3	4	5	6	7	: GOOD CONVERSATIONALIST
SOCIALLY UNSKILLED ::	1	2	3	4	5	6	7	: SOCIALLY SKILLED
INCOMPETENT COMMUNICATOR ::	1	2	3	4	5	6	7	: COMPETENT COMMUNICATOR
INAPPROPRIATE COMMUNICATOR ::	1	2	3	4	5	6	7	: APPROPRIATE COMMUNICATOR
INEFFECTIVE COMMUNICATOR ::	1	2	3	4	5	6	7	: EFFECTIVE COMMUNICATOR

Comments:

Appendix 3.3. Systems Usability Scale

For the design you have just looked at please rate how usable you think it was using the following System Usability Scale:

	Strongly disagree				Strongly agree
1. I think that I would like to use this system frequently	1	2	3	4	5
2. I found the system unnecessarily complex	1	2	3	4	5
3. I thought the system was easy to use	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5
5. I found the various functions in this system were well integrated	1	2	3	4	5
6. I thought there was too much inconsistency in this system	1	2	3	4	5
7. I would imagine that most people would learn to use this system very quickly	1	2	3	4	5
8. I found the system very cumbersome to use	1	2	3	4	5
9. I felt very confident using the system	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this system	1	2	3	4	5

Appendix 3.4. Ministry of Defence Research Ethics Committee

Ministry of Defence
Research Ethics Committee

From the MODREC Secretariat
Building 5, GO2
Defence Science and Technology Laboratory
Porton Down, Salisbury, SP4 0JQ
Telephone: 01980 956351
E-mail: MODREC@dstl.gov.uk

Dr Kate Hone
Brunel University London,
Kingston Lane,
Uxbridge,
UB8 3PH

Telephone : 01895 265340

E-mail : kate.hone@brunel.ac.uk

Our Reference: 772/MODREC/16

Date: 24th January 2017

Dear Kate,

Empatic task 1b: Media skills training intervention based on automated recognition of human emotion and non-verbal behaviour

Thank you for submitting your revised Protocol 772/MODREC/16 with tracked changes, and the covering letter with detailed responses to the MODREC letter. I can confirm that the revised protocol has been given favourable opinion ex-Committee. I wish you and your colleagues a successful study.

In due course please send the Secretariat a final report containing a summary of the results so that these can be filed in accordance with the arrangements under which MODREC operates. Please would you also send a brief interim report in one year if the study is still on-going.

This approval is valid for three years and is conditional upon adherence to the protocol – please let the Secretariat know if any amendment becomes necessary.

Yours sincerely

Dr Simon Kolstoe,

Appendix 3.5. Brunel Research Ethics Committee Approval for Exploratory Stage, User Centered Design Stage and Experiment Stage

Brunel University London

University Research Ethics Committee
Brunel University London
Kingston Lane
Uxbridge
UB8 3PH
United Kingdom

www.brunel.ac.uk

16 November 2016

LETTER OF APPROVAL

Applicant: Dr Kate Hone

Project Title: Empatic task 1b

Reference: 3795-SS-Nov/2016- 4310-1

Dear Dr Kate Hone,

The Research Ethics Committee has considered the above application recently submitted by you.

The Chair, acting under delegated authority has agreed that there is no objection on ethical grounds to the proposed study. Approval is given on the understanding that the conditions of approval set out below are followed:

- The agreed protocol must be followed. Any changes to the protocol will require prior approval from the Committee by way of an application for an amendment.

Please note that:

- Research Participant Information Sheets and (where relevant) flyers, posters, and consent forms should include a clear statement that research ethics approval has been obtained from the relevant Research Ethics Committee.
- The Research Participant Information Sheets should include a clear statement that queries should be directed, in the first instance, to the researcher. Complaints, on the other hand, should be directed, in the first instance, to the Chair of the relevant Research Ethics Committee.
- Approval to proceed with the study is granted subject to receipt by the Committee of satisfactory responses to any conditions that may appear above, in addition to any subsequent changes to the protocol.
- The Research Ethics Committee reserves the right to sample and review documentation, including raw data, relevant to the study.

Kind regards,

Peter C Hobson

Professor Peter Hobson

Chair, University Research Ethics Committee

Brunel University London

Page 1 of 1

Appendix 3.6. Brunel Research Ethics Committee Approval for Follow-up Stage

Brunel University London

College of Engineering, Design and Physical Sciences Research Ethics Committee
Brunel University London
Kingston Lane
Uxbridge
UB8 3PH
United Kingdom

www.brunel.ac.uk

24 April 2018

LETTER OF APPROVAL

Applicant: Miss Monica Pereira

Project Title: Media Skills Training Using Automated Technology

Reference: 11294-LR-Apr/2018- 12578-1

Dear Miss Monica Pereira

The Research Ethics Committee has considered the above application recently submitted by you.

The Chair, acting under delegated authority has agreed that there is no objection on ethical grounds to the proposed study. Approval is given on the understanding that the conditions of approval set out below are followed:

- The agreed protocol must be followed. Any changes to the protocol will require prior approval from the Committee by way of an application for an amendment.

Please note that:

- Research Participant Information Sheets and (where relevant) flyers, posters, and consent forms should include a clear statement that research ethics approval has been obtained from the relevant Research Ethics Committee.
- The Research Participant Information Sheets should include a clear statement that queries should be directed, in the first instance, to the Supervisor (where relevant), or the researcher. Complaints, on the other hand, should be directed, in the first instance, to the Chair of the relevant Research Ethics Committee.
- Approval to proceed with the study is granted subject to receipt by the Committee of satisfactory responses to any conditions that may appear above, in addition to any subsequent changes to the protocol.
- The Research Ethics Committee reserves the right to sample and review documentation, including raw data, relevant to the study.
- You may not undertake any research activity if you are not a registered student of Brunel University or if you cease to become registered, including abeyance or temporary withdrawal. As a deregistered student you would not be insured to undertake research activity. Research activity includes the recruitment of participants, undertaking consent procedures and collection of data. Breach of this requirement constitutes research misconduct and is a disciplinary offence.

Zhao Hua

Professor Hua Zhao

Chair

College of Engineering, Design and Physical Sciences Research Ethics Committee
Brunel University London

Appendix 3.7. Recruitment Email

Title of Study: Media Skills Training using automated recognition of human emotion and non-verbal behaviour
MoDREC Reference: 772/MoDREC/2019

Subject: Media Skills Training **Email**

body:

Circular e-mail for use for recruitment of volunteers for study ref. 772/MoDREC/2019, approved by the Ministry of Defence Research Ethics Committee. You are under no obligation to reply to this email, however if you choose to, participation in this research is voluntary and you may withdraw at any time.

We wish to invite you to participate in a research programme run by the Human Centred Design Institute (HCDI) at Brunel University London. The study provides the opportunity to take part in a oneday media skills training course at the Brunel campus delivered by [name of training provider] on [date]. The course aims to help develop research dissemination skills and the purpose of the research is to investigate whether such training can be augmented by the use of automated detection of emotion and non-verbal behaviour. Should you be interested in participating, please don't hesitate to contact us.

Contact information: Monica.Pereira@brunel.ac.uk

Best Regards,

The Brunel HCDI Team

Targets: Brunel students and staff members; internal Brunel mailing lists; online groups relevant to the research carried out at the Brunel Human Centred Design Institute (HCDI).

Appendix 3.8. Participant Information Sheet for Chapter 4

Participant Information Sheet

Study Title

Media Skills Training using automated recognition of human emotion and non-verbal behaviour

MoDREC Application No: 772/MoDREC/2019

Invitation to take part

We would like to invite you to take part in a research study. Before you decide, you need to understand why the research is being done and what it would involve for you. Please read the following information carefully and talk to others about the study if you wish.

Please feel free to contact us if anything is unclear or you need additional information: Email: kate.hone@brunel.ac.uk, Phone: 01895266009.

Take at least 24 hours to decide whether or not you want to take part. Thank you.

What is the purpose of the research?

We aim to improve media skills training by collecting and analysing information about an interviewee's displayed emotion and use of non-verbal signals (body language) in an automated fashion. One of our primary objectives is to develop ways of providing feedback for trainees that will help them to improve how they are perceived in the context of a media interview, and we need your help to design novel solutions.

Who is doing this research?

We are a team of researchers at Brunel University London in Uxbridge funded by the Ministry of Defence.

Why have I been invited to take part?

Because you are a student or member of staff at Brunel University London.

Do I have to take part?

No, your participation is entirely voluntary. Please get in contact with us if you wish to be involved. Please be aware that you also have the right to withdraw from the study at any time, without any penalty and without having to explain your decision.

What will I be asked to do?

You will take part in a media skills training course where you will get the chance to practice giving an interview about your work to a journalist, including speaking to camera. During the training we will use a camera-based system to detect your facial expression, tone of voice and body movements during each interview you practice. The study also involves the use of two sensors which are worn, one a badge that detects movement, tone of voice and alignment and the other a wearable wristband that measures skin conductivity and heart rate. The overall training will be a day-long event, but the total amount of time during which you'll be wearing sensors or being recorded will not exceed 2 hours in total. 

A wearable 'badge' that detects tone of voice and movement.

Acceptance criteria

You must be over 18 to take part and you must be able to take part in a spoken conversation. You must not have previously attended media skills training at Brunel University London within the past 12 months.

What is the device or procedure that is being tested?

We are testing a method of improving the interpersonal communication skills performance by detecting emotional state and non-verbal signal (body language) while taking part in a media interview. One of our primary objectives is to develop ways of providing feedback for trainees that will help them to improve how they come across in a media interview. After the training event we will therefore compare the data obtained from the automated recognition of emotional and non-verbal signals to the ratings provided by you, your trainer and the experimenters to identify the most important non-verbal signals to provide feedback on.

What are the benefits of taking part?

By taking part in the study you will receive media skills training which can aid your own personal and professional development. The information we get from the study will help to develop new solutions which will help will enhance future training in interpersonal skills.

What are the possible disadvantages and risks of taking part?

Some people find giving a media interview somewhat stressful. Some of the questions that the interviewer might ask you during training could be probing or intrusive to some degree which is reflective of real journalistic methods. However, practising the skills involved within the context of a training course should be less stressful that doing a real media interview and could help you feel more prepared to give real media interviews in future.

Will my taking part in this study be kept confidential?

Yes, your decision as to whether you decide to take part, not to take part, or withdraw from this study will be treated in confidence.

Can I withdraw from the research and what will happen if I do?

Yes, you have the right to withdraw at any time during the course of the study, without having to explain why, and without any consequence. This will not impact your eligibility to remain in the training session (not as part of the study) or access future training in media skills. You also have the right to withdraw any data already provided by you in the study; any such request must be made by the end of the training day.

Are there any expenses and payments which I will get?

You will receive a £30 voucher to thank you for participating in our study.

Will my taking part or not taking part affect my career?

No.

Whom do I contact if I have any questions?

Should you have any questions, please don't hesitate to contact the Brunel research team: Kate.Hone@brunel.ac.uk, 01895266009.

Whom do I contact if I have a complaint?

Please direct any concerns or complaints to the Chair of the Brunel University Research Ethics Committee, Prof. Peter Hobson, Peter.Hobson@brunel.ac.uk.

What happens if I suffer any harm?

If you suffer any harm as a direct result of taking part in this study, you can apply for compensation under the MoD's 'No-Fault Compensation Scheme'.

Will my records be kept confidential?

Your records will be anonymised and kept confidential at all times. Hard copies of the consent forms will be stored in a locked cabinet, and all relevant electronic documents will be assigned an identifier for the purpose of anonymisation and will be stored exclusively on a Brunel file system. Only the research team will have access to your records. Your personal data will be treated in accordance with the provisions of the Data Protections Act 1998 at Brunel University London. Following completion of the study your signed consent form will be forwarded to the MoDREC Secretariat for retention in accordance with UK legislation and MoD Policy.

Who is organising and funding the research?

The training will be conducted on the Brunel campus. They are part of a research project led by QinetiQ and carried out in collaboration with Brunel University London. The project is funded by the MoD (Dstl).

Who has reviewed the study?

This study has been reviewed and given favourable opinion by the Ministry of Defence Research Ethics Committee (MoDREC), and by the Brunel University Research Ethics Committee.

Further information and contact details

Name: Dr. Kate Hone

Address: Brunel University London

Kingston Lane

Uxbridge UB8 3PH

Tel No: 01895 266009

E-mail: Kate.hone@brunel.ac.uk

Compliance with the Declaration of Helsinki

This study complies, and at all times will comply, with the Declaration of Helsinki[1] as adopted at the 64th WMA General Assembly at Fortaleza, Brazil in October 2013.

All data will be collected and stored in accordance with the Data Protection Act 1998

Brunel University London is committed to the UK Concordat on Research Integrity, http://www.rcuk.ac.uk/funding/researchintegrity

[1] World Medical Association Declaration of Helsinki [revised October 2013]. Recommendations Guiding Medical Doctors in Biomedical Research Involving Human Subjects. 64th WMA General Assembly, Fortaleza (Brazil).

Appendix 3.9. Participant Information Sheet for Chapter 5

Media Skills Training : ***Participant Information Sheet***

Study Title Design of feedback intervention for media skills training

MoDREC Application No: 772/MoDREC/2019

Invitation to take part

As a participant in our Media Skills Training study (held on date xx/xx/xx) you kindly volunteered to be contacted again to help use with some follow-up work. We would therefore like to invite you to take part in the next phase of our research. Before you decide, you need to understand why the research is being done and what it would involve for you. Please read the following information carefully and talk to others about the study if you wish.

Please feel free to contact us if anything is unclear or you need additional information: Email: kate.hone@brunel.ac.uk, Phone: 01895266009.

Take at least 24 hours to decide whether or not you want to take part. Thank you.

What is the purpose of the research?

We aim to improve media skills training by collecting and analysing information about an interviewee's displayed emotion and use of non-verbal signals (body language) in an automated fashion. One of our primary objectives is to develop ways of providing feedback for trainees that will help them to improve how they are perceived in the context of a media interview, and we need your help to design novel solutions. This study is being conducted over a two-year period.

Who is doing this research?

We are a team of researchers at Brunel University London in Uxbridge funded by the Ministry of Defence.

Why have I been invited to take part?

You are being invited to take part as you kindly provided your contact details and indicated a willingness to be contacted about follow-up work after the Media Skills Training event held on (xx/xx/xx). The follow-up studies are expected to involve a total of 5-6 participants.

Do I have to take part?

No, your participation is entirely voluntary. Please get in contact with us if you wish to be involved. Please be aware that you also have the right to withdraw from the study at any time, without any penalty and without having to explain your decision.

What will I be asked to do?

You will be given the chance to see recordings of the interviews you gave at the Media Skills Training course, supplemented with visual displays based on the emotion and non-verbal signals (body language) data which was collected in real time during your interviews. We have devised several different ways of displaying this information and we'd like your help to choose the most effective format to help trainees improve their performance. The overall session will not exceed 2 hours in total.

Acceptance criteria

You must be over 18 to take part and have taken part in the Media Skills Study on date (xx/xx/xx).

What is the device or procedure that is being tested?

We are evaluating different ways of providing feedback to trainees on their use of emotional and non-verbal signals (body language). We'd like to get the views of potential users to ensure that the feedback is meaningful and usable.

What are the benefits of taking part?

There are no direct benefits to you of taking part but the information we get from the study will help to develop new solutions which will help will enhance future training in interpersonal skills.

What are the possible disadvantages and risks of taking part?

We do not anticipate any disadvantages or risks of taking part.

Will my taking part in this study be kept confidential?

Yes, your decision as to whether you decide to take part, not to take part, or withdraw from this study will be treated in confidence.

Can I withdraw from the research and what will happen if I do?

Yes, you have the right to withdraw at any time during the course of the study, without having to explain why, and without any consequence. You also have the right to withdraw any data already provided by you in the study; any such request must be made by the end of the day of the study.

Are there any expenses and payments which I will get?

We will give you a £15 voucher as a modest token of appreciation to thank you for participating in our studies.

Will my taking part or not taking part affect my career?

No.

Whom do I contact if I have any questions?

Should you have any questions, please don't hesitate to contact the Brunel research team: Kate.Hone@brunel.ac.uk, 01895266009.

Whom do I contact if I have a complaint?

Please direct any concerns or complaints to the Chair of the Brunel University Research Ethics

Committee, Prof. Peter Hobson, Peter.Hobson@brunel.ac.uk.

What happens if I suffer any harm?

If you suffer any harm as a direct result of taking part in this study, you can apply for compensation under the MoD's 'No-Fault Compensation Scheme'.

Will my records be kept confidential?

Your records will be anonymised and kept confidential at all times. Hard copies of the consent forms will be stored in a locked cabinet, and all relevant electronic documents will be assigned an identifier for the purpose of anonymisation and will be stored exclusively on a Brunel file system. Only the research team will have access to your records. Your personal data will be treated in accordance with the provisions of the Data Protections Act 1998 at Brunel University London. Following completion of the study your signed consent form will be forwarded to the MoDREC Secretariat for retention in accordance with UK legislation and MoD Policy.

Who is organising and funding the research?

The training will be conducted on the Brunel campus. They are part of a research project led by QinetiQ and carried out in collaboration with Brunel University London. The project is funded by MoD (Dstl).

Who has reviewed the study?

This study has been reviewed and given favourable opinion by the Ministry of Defence Research Ethics Committee (MoDREC), and by the Brunel University Research Ethics Committee.

Further information and contact details

Name: Dr. Kate Hone

Address: Brunel University London

Kingston Lane

Uxbridge UB8 3PH

Tel No: 01895 266009

E-mail: Kate.hone@brunel.ac.uk

Compliance with the Declaration of Helsinki

This study complies, and at all times will comply, with the Declaration of Helsinki[2] as adopted at the 64th WMA General Assembly at Fortaleza, Brazil in October 2013.

All data will be collected and stored in accordance with the Data Protection Act 1998

Brunel University London is committed to the UK Concordat on Research Integrity, http://www.rcuk.ac.uk/funding/researchintegrity

[2] World Medical Association Declaration of Helsinki [revised October 2013]. Recommendations Guiding Medical Doctors in Biomedical Research Involving Human Subjects. 64th WMA General Assembly, Fortaleza (Brazil).

Appendix 3.10. Participant Information Sheet for Chapter 6

Participant Information Sheet

Study Title

Media Skills Training using automated recognition of human emotion and non-verbal behaviour

MoDREC Application No: 772/MoDREC/2019

Invitation to take part

We would like to invite you to take part in a research study. Before you decide, you need to understand why the research is being done and what it would involve for you. Please read the following information carefully, and talk to others about the study if you wish.

Please feel free to contact us if anything is unclear or you need additional information: Email: kate.hone@brunel.ac.uk, Phone: 01895266009.

Take at least 24 hours to decide whether or not you want to take part. Thank you.

What is the purpose of the research?

We aim to improve media skills training by collecting and analysing information about an interviewee's displayed emotion and use of non-verbal signals (body language) in an automated fashion. One of our primary objectives is to develop ways of providing feedback for trainees that will help them to improve how they are perceived in the context of a media interview, and we need your help to evaluate potential solutions.

Who is doing this research?

We are a team of researchers at Brunel University London in Uxbridge funded by the Ministry of Defence.

Why have I been invited to take part?

Because you are a student or member of staff at Brunel University London.

Do I have to take part?

No, your participation is entirely voluntary. Please get in contact with us if you wish to be involved. Please be aware that you also have the right to withdraw from the study at any time, without any penalty and without having to explain your decision.

What will I be asked to do?

You will take part in a media skills training course where you will get the chance to practice giving interviews about your work to a journalist, including speaking to camera, and receive feedback on your performance. During the training we will use a camera-based system to detect your facial expression, tone of voice and body movements during each interview you practice. The study also involves the use of two sensors which are worn, one a badge that detects movement, tone of voice and alignment and the other a wearable wristband that measures skin conductivity and heart rate.

The overall training will be a day-long event, but the total amount of time during which you'll be wearing sensors or being recorded will not exceed 2 hours in total.

A wearable 'badge' that detects tone of voice and movement:

Acceptance criteria

You must be over 18 to take part and you must be able to take part in a spoken conversation. You must not have previously attended media skills training at Brunel University London within the past 12 months.

What is the device or procedure that is being tested?

We are testing a method of improving the interpersonal communication skills performance by detecting emotional state and non-verbal signal (body language) while taking part in a media interview. One of our primary objectives is to develop ways of providing feedback for trainees that will help them to improve how they come across in a media interview. During the training we will vary the type of feedback that participants receive and after the training we will test the impact of these different feedback approaches on training effectiveness.

What are the benefits of taking part?

By taking part in the study you will receive media skills training which can aid your own personal and professional development. The information we get from the study will help to evaluate new solutions which aim to enhance future training in interpersonal skills.

What are the possible disadvantages and risks of taking part?

Some people find giving a media interview somewhat stressful. Some of the questions that the interviewer might ask you during training could be probing or intrusive to some degree which is reflective of real journalistic methods. However, practising the skills involved within the context of a training course should be less stressful that doing a real media interview and could help you feel more prepared to give real media interviews in future.

Will my taking part in this study be kept confidential?

Yes, your decision as to whether you decide to take part, not to take part, or withdraw from this study will be treated in confidence.

Can I withdraw from the research and what will happen if I do?

Yes, you have the right to withdraw at any time during the course of the study, without having to explain why, and without any consequence. This will not impact your eligibility to remain in the training session (not as part of the study) or access future training in media skills. You also have the right to withdraw any data already provided by you in the study; any such request must be made by the end of the training day.

Are there any expenses and payments which I will get?

You will receive a £30 voucher to thank you for participating in our study.

Will my taking part or not taking part affect my career?

No.

Whom do I contact if I have any questions?

Should you have any questions, please don't hesitate to contact the Brunel research team: Kate.Hone@brunel.ac.uk, 01895266009.

Whom do I contact if I have a complaint?

Please direct any concerns or complaints to the Chair of the Brunel University Research Ethics Committee, Prof. Peter Hobson, Peter.Hobson@brunel.ac.uk.

What happens if I suffer any harm?

If you suffer any harm as a direct result of taking part in this study, you can apply for compensation under the MoD's 'No-Fault Compensation Scheme'.

Will my records be kept confidential?

Your records will be anonymised and kept confidential at all times. Hard copies of the consent forms will be stored in a locked cabinet, and all relevant electronic documents will be assigned an identifier for the purpose of anonymisation and will be stored exclusively on a Brunel file system. Only the research team will have access to your records. Your personal data will be treated in accordance with the provisions of the Data Protections Act 1998 at Brunel University London. Following completion of the study your signed consent form will be forwarded to the MoDREC Secretariat for retention in accordance with UK legislation and MoD Policy.

Who is organising and funding the research?

The training will be conducted on the Brunel campus. They are part of a research project led by QinetiQ and carried out in collaboration with Brunel University London. The project is funded by the MoD (Dstl).

Who has reviewed the study?

This study has been reviewed and given favourable opinion by the Ministry of Defence Research Ethics Committee (MoDREC), and by the Brunel University Research Ethics Committee.

Further information and contact details

Name: Dr. Kate Hone

Address: Brunel University London

Kingston Lane

Uxbridge UB8 3PH

Tel No: 01895 266009

E-mail: Kate.hone@brunel.ac.uk

Compliance with the Declaration of Helsinki

This study complies, and at all times will comply, with the Declaration of Helsinki[3] as adopted at the 64th WMA General Assembly at Fortaleza, Brazil in October 2013.

All data will be collected and stored in accordance with the Data Protection Act 1998

Brunel University London is committed to the UK Concordat on Research Integrity, http://www.rcuk.ac.uk/funding/researchintegrity

[3] World Medical Association Declaration of Helsinki [revised October 2013]. Recommendations Guiding Medical Doctors in Biomedical Research Involving Human Subjects. 64th WMA General Assembly, Fortaleza (Brazil).

Appendix 3.11. Participant Information Sheet for Chapter 7 Media Skills Training Using Automated Technology: A Follow-up Study

Invitation to take part

We would like to invite you to take part in a research study. Before you decide, you need to understand why the research is being done and what it would involve for you. Please read the following information carefully and talk to others about the study if you wish.

Please feel free to contact us if anything is unclear to you or you need additional information:

Email: monica.pereira@brunel.ac.uk

What is the purpose of the research?

The aim of the current study is to examine the longer-term impact of the media skills training that you received in the earlier study.

Who is doing this research?

A team at Brunel.

Do I have to take part?

No, your participation is entirely voluntary. Please get in touch with us if you wish to be involved. Please be aware that you also have the right to withdraw from the study at any time, without any penalty and without having to explain your decision.

What will I be asked to do?

You will take part in a single, face-to-face media interview with a journalist. This interview will include speaking to a camera. You will receive feedback about your performance. During the interview we will use a camerabased system to detect your facial expression, tone of voice, body movements. The study involves the use of two sensors which are worn, one a badge that detects movement, tone of voice and alignment and the other, a wearable wristband that measures skin conductivity and heart rate. The overall time taken for this study will not exceed 30 minutes.

Acceptance criteria

You must be over the age of 18 to take part. You will have to have taken part in our previous study to assess the effects of training on your performance.

What is the device or procedure that is being tested?

We are evaluating the potential of automated technology within a communication skills context.

What are the benefits of taking part?

By taking part in the study you will receive feedback on your interviewing skills which can aid your own personal and professional development. The information we get form the study will help to develop new solutions which will help enhance future training in interpersonal skills.

What are the possible disadvantages and risks of taking part?

Some people find giving a media interview somewhat stressful. Some of the questions that the interviewer might ask you during training could be probing or intrusive to some degree which is reflective of real journalistic methods. However, practising the skills involved within the context of a training session should be less stressful than doing a real media interview and could help you feel more prepared to give real media interviews in future.

Will my taking part in this study be kept confidential?

Yes, you decision as to whether you decide to take part, not to take part, or withdraw from this study will be treated in confidence.

Can I withdraw from the research and what will happen if I do?

Yes, you have the right to withdraw at any time during the course of the study, without having to explain why, and without any consequence. You also have the right to withdraw any data already provided by you in the study, any such request must be made by the end of the session.

Are there any expenses and payments which I will get?

No.

Will my taking part or not taking part affect my career?

No.

Who do I contact if I have any questions?

Should you have any questions, please don't hesitate to contact the Brunel research team:

Kate.hone@brunel.ac.uk, 01895266009 (lead) and Monica.pereira@brunel.ac.uk (researcher) **Whom**

do I contact if I have a complaint?

Please direct any concerns or complaints to the Chair of the Brunel University Research Ethics Committee, Prof. Peter Hobson, Peter.Hobsin@brunel.ac.uk **Will my records be kept confidential?**

Your records will be anonymised and kept confidential at all times. Hard copies of the consent forms will be stored in a locked cabinet, and all electronic documents will be assigned an identifier for the purpose of anonymization and will be stored exclusively on a Brunel file system. Only the research team will have access to your records. Your personal data will be treated in accordance with the provision of the Data Protections Act 1998 at Brunel University London.

Who is organising and funding the research?

The training will be conducted on the Brunel Campus. The project is funded by the EPSRC.

Further information and contact details:

Name: Prof Kate Hone
Address: Brunel University London
Kingston Late
Uxbridge UB8 3PH
Tel No: 01895 266009
Email: Kate.hone@brunel.ac.uk

Name: Monica Pereira
Address: Brunel University London
Kingston Lane
Uxbridge UB8 3PH
Email: monica.pereira@brunel.ac.uk

Compliance with the Declaration of Helsinki

The study complies, and at all times will comply, with the Declaration of Helsinki[4]

[4] World Medical Association Declaration of Helsinki [revised October 2013]. Recommendations Guiding Medial Doctors in Biomedical Research Involving Human Subjects. 64th WMA General Assembly, Fortaleza (Brazil).

Appendix 3.12. Consent Form for All Research Stages

Title of Study: Media Skills Training

MoDREC Reference: 772/MoDREC/2019

Please Initial or Tick Boxes

- The nature, aims and risks of the research have been explained to me. I have read and understood the Information for Participants and understand what is expected of me. All my questions have been answered fully to my satisfaction.
- I understand that if I decide at any time during the research that I no longer wish to participate in this project, I can notify the researchers involved and be withdrawn from it immediately without having to give a reason. I also understand that I may be withdrawn from it at any time, and that in neither case will this be held against me in subsequent dealings with the Ministry of Defence.
- I understand that video and audio recordings of me will be made and that these will be reviewed by experimenters as part of the analysis of the data from this study.
- I consent to the processing of my personal information for the purposes of this research study. I understand that such information will be treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998.
- I agree to volunteer as a participant for the study described in the information sheet and give full consent.
- This consent is specific to the particular study described in the Information for Participants attached and shall not be taken to imply my consent to participate in any subsequent study or deviation from that detailed here.
- I understand that in the event that I reveal, during the course of taking part, that I have been involved in activity which represents professional misconduct, this will be reported in line with Brunel University London's Regulations and Codes of Practice (e.g. Senate Regulation 6 for student participants; Research Integrity Code for staff participants).
- I understand that in the event of my sustaining injury, illness or death as a direct result of participating as a volunteer in Ministry of Defence research, I or my dependants may enter a claim with the Ministry of Defence for compensation under the provisions of the no-fault compensation scheme, details of which are attached.
- I understand the compensation arrangements that have been provided.

Participant's Statement:

I ……………………………………………………agree that the research project named above has been explained to me to my satisfaction and I agree to take part in the study. I have read both the notes written above and the Information for Participants about the project and understand what the research study involves.

Signed: Date:

Witness Name:

Signature:

Date:

Investigator's Statement:

I ……………………………………………………… confirm that I have carefully explained the nature, demands and any foreseeable risks (where applicable) of the proposed research to the Participant.

Signed: **Date:**

Authorising Signatures

The information supplied above is to the best of my knowledge and belief accurate. I clearly understand my obligations and the rights of research participants, particularly concerning recruitment of participants and obtaining valid consent.

Signature of Chief Investigator

…………………………………………………… **Date:**

Name and Contact Details of Independent Medical Officer (*if appropriate*):

N/A

Name and Contact Details of Chief Investigator:

Dr Kate Hone, kate.hone@brunel.ac.uk

Appendix 3.13. Demographics Questionnaire

Please complete the following questions about yourself which may help us in interpreting the results we obtain from this study:

My experience of public speaking / presenting is: none / a little / some / extensive

My experience of giving media interviews is: none / a little / some / extensive

I am: a taught programme student / a research student / a member of research staff / a member of academic staff / a professional or administrative member of staff / prefer not to say

I am: Male / Female / Prefer not to say / Other (please specify)

I am: 18-25 years old / 26-35 years old / 36-45 years old / 46-55 years old / 56-65 years old / 66-75 years old / 75 years old or order / prefer not to say

I am: White or White British / Black or Black British / Asian or Asian British / Chinese / Mixed background / Arab / Any other ethnic background / prefer not to say

My place of birth: UK / Other (please specify) / prefer not to say

My nationality is: [open text] / prefer not to say

My first language is: English / Other (please specify) / prefer not to say

Do you consider yourself to have a social/communication impairment (such as Asperger's syndrome/other autistic spectrum disorder)?: No / Yes / prefer not to say

Appendix 3.14. Closing statement for exploratory stage

Thank you for taking part in this study. The purpose of our work is to improve media skills training by collecting and analysing information about an interviewee's displayed emotion and use of non-verbal signals (body language) in an automated fashion. After the training event we will therefore compare the data obtained from the automated recognition of emotional and non-verbal signals to the ratings provided by you, your trainer and the experimenters to identify the most important non-verbal signals to provide feedback on. If you have any concerns about this proposed use of your data, you can still decide to withdraw from the study without consequence at this time. You are also welcome to discuss any concerns you have with the study team before deciding.

Following today's training you are reminded that your records will be anonymised and kept confidential at all times. Hard copies of the consent forms will be stored in a locked cabinet, and all relevant electronic documents will be assigned an identifier for the purpose of anonymisation and will be stored exclusively on a Brunel file system. Only the research team will have access to your records. Your personal data will be treated in accordance with the provisions of the Data Protection Act 1998 at Brunel University London. Following completion of the study your signed consent form will be forwarded to the MoDREC Secretariat for retention in accordance with UK legislation and MoD Policy.

If you would like to find out about the study results you are welcome to request a copy of the final project findings by emailing the study lead, Dr Kate Hone (kate.hone@brunel.ac.uk).

We are also looking to recruit volunteers who participated in this study to join us for a follow-up study to look at prototype designs for giving trainees feedback on their use of non-verbal signals. This would involve watching recordings of the interviews you gave today, supplemented with visual displays based on the emotion and non-verbal signals (body language) data which was collected in real time during your interviews. We will demonstrate different ways of displaying this information and we'd like your help to choose the most effective format. The overall session will not exceed 2 hours in total and a small token of our appreciation will be provided (a voucher of £15). There will be a chance to read further information about this study before deciding whether to participate, but we'd be grateful if you could give an indication of whether or not you'd be willing to be contacted to take part in this follow up. If you do consent to be contacted, please fill in the section below.

I consent to be contacted by the study team with details of how to participate in a follow-up study, my contact details are ……………………………………………………

If you have any concerns or queries please contact the Chief Investigator, Dr Kate Hone (kate.hone@brunel.ac.uk or 0189566009). Alternatively, if you have a complaint you can contact the Chair of the Brunel University Research Ethics Committee, Prof. Peter Hobson, Peter

Appendix 3.15. Closing statemen for user centered study

Thank you for taking part in this study. We will use the feedback you have provided on the designs to help us decide on the most appropriate way to present this kind of information to trainees in future.

Following today's session, you are reminded that your records will be anonymised and kept confidential at all times. Hard copies of the consent forms will be stored in a locked cabinet, and all relevant electronic documents will be assigned an identifier for the purpose of anonymisation and will be stored exclusively on a Brunel file system. Only the research team will have access to your records. Your personal data will be treated in accordance with the provisions of the Data Protection Act 1998 at Brunel University London. Following completion of the study your signed consent form will be forwarded to the MoDREC Secretariat for retention in accordance with UK legislation and MoD Policy.

If you would like to find out about the study results you are welcome to request a copy of the final project findings by emailing the study lead, Dr Kate Hone (kate.hone@brunel.ac.uk).

If you have any concerns or queries please contact the Chief Investigator, Dr Kate Hone (kate.hone@brunel.ac.uk or 0189566009). Alternatively, if you have a complaint you can contact the Chair of the Brunel University Research Ethics Committee, Prof. Peter Hobson, Peter.Hobson@brunel.ac.uk

Appendix 3.16. Closing statement for experiment stage

Thank you for taking part in this study. The purpose of our work is to improve media skills training by collecting and analysing information about an interviewee's displayed emotion and use of non-verbal signals (body language) in an automated fashion. After the training event we will compare the training of those who received specific feedback on their use of emotional and non-verbal signals with those who received standard training feedback. If you have any concerns about this proposed use of your data, you can still decide to withdraw from the study without consequence at this time. You are also welcome to discuss any concerns you have with the study team before deciding.

Following today's training you are reminded that your records will be anonymised and kept confidential at all times. Hard copies of the consent forms will be stored in a locked cabinet, and all relevant electronic documents will be assigned an identifier for the purpose of anonymisation and will be stored exclusively on a Brunel file system. Only the research team will have access to your records. Your personal data will be treated in accordance with the provisions of the Data Protection Act 1998 at Brunel University London. Following completion of the study your signed consent form will be forwarded to the MoDREC Secretariat for retention in accordance with UK legislation and MoD Policy.

If you would like to find out about the study results you are welcome to request a copy of the final project findings by emailing the study lead, Dr Kate Hone (kate.hone@brunel.ac.uk).

If you have any concerns or queries please contact the Chief Investigator, Dr Kate Hone (kate.hone@brunel.ac.uk or 0189566009). Alternatively, if you have a complaint you can contact the Chair of the Brunel University Research Ethics Committee, Prof. Peter Hobson, Peter.Hobson@brunel.ac.uk

Appendix 3.17. Closing statement for follow-up stage

Thank you for taking part in this study. The purpose of this study is to assess whether the skills you had gained in your initial training were maintained after 6-months. If you have any concerns about this proposed use of your data, you can still decide to withdraw from the study without consequence at this time. You are also welcome to discuss any concerns you have with the study team before deciding. Following today's training you are reminded that your records will be anonymised and kept confidential at all times. Hard copies of the consent forms will be stored in a locked cabinet, and all relevant electronic documents will be assigned an identifier for the purpose of anonymisation and will be stored exclusively on a Brunel file system. Only the research team will have access to your records. Your personal data will be treated in accordance with the provisions of the Data Protection Act 1998 at Brunel University London.

If you would like to find out about the study results you are welcome to request a copy of the final project findings by emailing the study lead, Dr Kate Hone (kate.hone@brunel.ac.uk).

If you have any concerns or queries please contact the Chief Investigator, Dr Kate Hone (kate.hone@brunel.ac.uk or 0189566009). Alternatively, if you have a complaint you can contact the Chair of the Brunel University Research Ethics Committee, Prof. Peter Hobson, Peter.Hobson@brunel.ac.uk

Appendix 4

Appendix 4.1. Descriptive scores for individual participants

	Radio Interview - cut off = 23.70	
Participant	**Molar Score (classification)**	**Weighted Average (classification)**
1	30.50 (good)	30.50 (good)
2	27.25 (good)	28.40 (good)
3	24.75 (good)	24.00 (good)
4	27.25 (good)	26.20 (good)
5	16.25 (bad)	16.20 (bad)
6	30.75 (good)	28.80 (good)
7	21.25 (bad)	20.00 (bad)
8	23.75 (good)	23.00 (bad)
9	31.50 (good)	30.20 (good)
10	25.00 (good)	24.00 (good)
11	23.50 (bad)	22.20 (bad)
12	31.75 (good)	30.80 (good)
13	29.25 (good)	28.40 (good)
14	25.75 (good)	26.60 (good)
15	20.25 (bad)	20.20 (bad)
16	19.50 (bad)	20.60 (bad)
17	17.00 (bad)	16.60 (bad)
	On-camera Interview – cut off = 25.20	
Participant	**Molar Score (classification)**	**Weighted Average (classification)**
1	31.00 (good)	30.00 (good)
2	29.00 (good)	28.25 (good)
3	25.60 (good)	26.50 (good)
4	27.60 (good)	29.00 (good)
5	16.60 (bad)	16.50 (bad)
6	30.60 (good)	32.00 (good)
7	20.00 (bad)	21.25 (bad)
8	27.00 (good)	26.00 (good)
9	32.60 (good)	33.50 (good)
10	23.60 (bad)	24.50 (bad)
11	25.80 (good)	27.75 (good)
12	33.80 (good)	34.00 (good)
13	29.20 (good)	30.25 (good)
14	29.40 (good)	29.25 (good)
15	23.60 (bad)	24.50 (bad)
16	23.80 (bad)	23.50 (bad)
17	23.20 (bad)	24.50 (bad)

Appendix 4.2.

Multimodal Fusion Analysis of Social Signals in Media Interviews

Introduction

Previous chapters in this thesis have found that social signal feedback is more effective in improving communication skills than the traditional method of providing feedback during training. However, the signals selected to feedback to participants were based on a preliminary data analysis. A more detailed analysis was later conducted; though, the setting was different as interviews included were both radio and on-camera interviews. This chapter investigates which combination of social signals are necessary for the context of media interviews with a larger sample size in the context of an oncamera media interview only. A larger sample size would provide deeper insight into media interview settings and would enable better inference from the data.

Data Handling

Dataset and Pre-processing

To create the dataset for this analysis data from all CSRS-rated on-camera interviews from across the three reported social signals data collection stages (exploratory study, experimental evaluation, follow up evaluation) were combined. This resulted in a dataset of 77 interview cases. The first 30 seconds were extracted from the data recording to investigate a primacy effect in impression formation and were enough to obtain participants' responses to the first question in each interview. The data was normalised so that all feature values are in the range [0, 1]. This reduces outliers and normalises the data. Data was removed because if missing data were replaced using the mean of the dataset, this would not be a good representation of the data as the classifier used is an instancebased classifier. Missing data was removed resulting in 67 data instances removed. See Table 6.1 for sample size demonstration.

Table 4.2.1. Sample of Data set included

Research Stage	**Sample Included**	**Removed**
Exploratory	17	5 (*n* = 12)
Intervention Evaluation	44	1 (*n* = 43)
Intervention Evaluation Follow-up	16	4 (*n* = 12)
Total	77	10 (*n* = 67)

Ground Truth and Feature Extraction

The midpoint of the dataset was obtained by observing the centre of the histogram which was used as a cut-off point. There was a total of 36 effective communicator cases and 31 poor communicator cases. Feature selection was applied using the correlation-based feature selection (CFS) method which selects features correlated with the class variable and uncorrelated with each other (Hall, 1999). Highly ranked features were selected (with a cut-off of 0.2). Features removed were content, anger, pitch, volume mirroring and volume mirroring lag due to missing data or data produced as 0.

Prediction Analysis and Group Differences

K-NN was conducted on all the features selected for inclusion in the analysis. As the sample size is unequal, a bootstrap aggregation (bagging) will be used as described in previous chapter (Chapter 4, section 1) with 100 iterations. Furthermore, a unimodal analysis using Mann-Whitney U analysis was conducted to assess if there were any differences in individual signals (from those selected using CSF) between effective and poor communicators.

Results

The results are shown in this section. The features extracted using the CFS method included arousal, energy, posture mirroring, sadness, extreme emotion and hesitation. See Figure 6.1 for the correlation values of each feature and its corresponding communication channel to be included in the classification analysis.

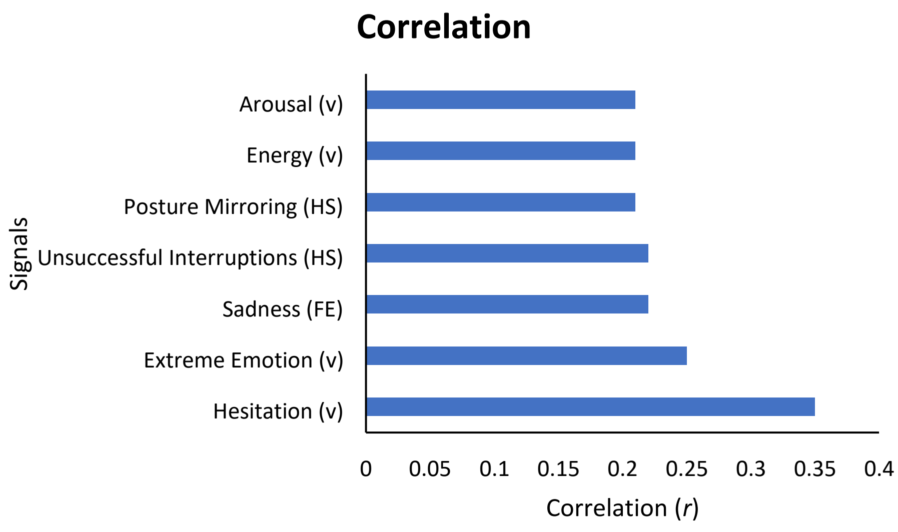


Figure 4.2.1. Features included in analysis. V = vocal analysis, FE = facial expression and HS = honest signals

Based on the extracted analysis, a k-NN analysis was conducted using k = 1, 2, 3 and 4 with a leaveone-out cross validation. The data was not partitioned in to 34 training and 34 testing as in

Chapter 4, section 1 as the data presented in this case is all on camera interviews whereas the data in Chapter 4 included radio interview and on-camera interviews. An additional analysis was conducted where half the data was partitioned into training and the remainder was left for testing. Table 6.2 demonstrates k-NN parameters and bagging accuracy.

Table 4.2.2. k-NN analysis results on extracted features including confusion matrix

k	Cross validation	Accuracy	Effective; Poor communicators	Bagging Accuracy	Effective; Poor communicators
1	Leave-one-out	55%	19 of 36; 18 of 31	57%	21 of 36; 17 of 31
2	Leave-one-out	62%	27 of 36; 15 of 31	68%	22 of 36; 22 of 31
3	Leave-one-out	66%	23 of 36; 21 of 31	66%	23 of 36; 21 of 31
4	Leave-one-out	63%	25 of 26; 17 of 31	61%	19 of 36; 22 of 31

A Mann-Whitney-U test (McKnight and Najab, 2010) was conducted to formally assess significant differences between effective communicators and poor communicators. The results revealed that poor communicators hesitated more than effective communicators (U = 320, p = .003). Poor communicators also demonstrated more sadness than effective communicators (U = 353.50, p = .01). There was no significant difference for vocal energy (U = 445.50, p = .16), extreme emotion (U = 410.50, p = .06), arousal (U = 468.50 p = .26), unsuccessful interruptions (U = 530, p = .706) and posture mirroring (U = 457, p = .204). Descriptive statistics can be seen in Figure 6.2.

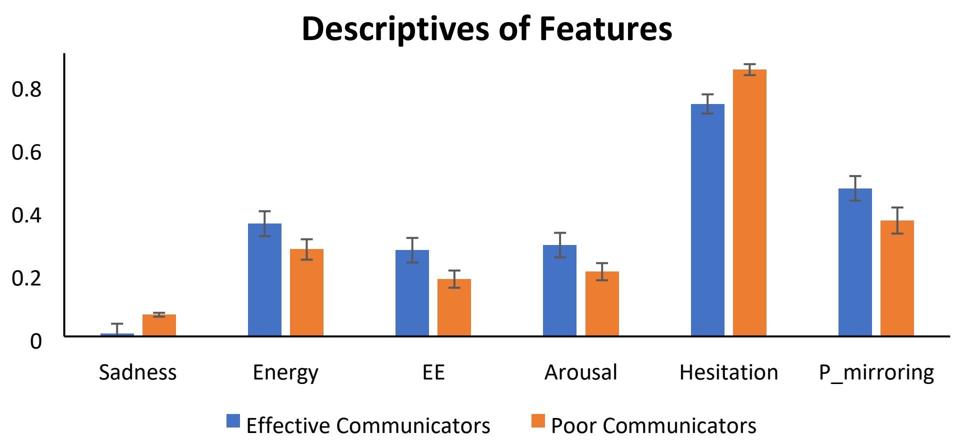


Figure 4.2.2. Mean and standard error for each. EE = extreme emotion, P_mirroring = posture mirroring.

Discussion

The aim of this chapter was to investigate which combination of social signals were identified based on subjective ratings. Signals were included across all communication channels suggesting that communication is a multimodal event. Additionally, similar signals were identified when using a smaller sample size in Chapter 4, this is discussed in this section to highlight the importance of these signals.

‘Energy’, ‘extreme emotion’ and ‘hesitation’ were included in the vocal affect channel. Results confirm relevance of ‘extreme emotion’ and ‘hesitation’ signals included in feedback design. This is because these signals were included in the preliminary analysis in Chapter 4. This suggests that feedback provided to participants were on par and accurate as identified with a larger sample size. In other words, the inclusion of these signals has been found in a small sample size and in the larger sample size analysis suggesting that these signals are relevant for appraising communication skills in media interviews. Both energy and extreme emotion signals feed into their description of passion and confidence which was found in both the experimental and follow-up research stages.

‘Posture mirroring’ was identified for the honest signal channel confirming the result from the preliminary data analysis and for providing feedback to trainees with a larger sample size. Suggesting that mirroring the interviewers’ posture is important in how an interviewee is perceived by the audience. This was also suggested by a media training guide (Taylor, 2015).

The facial expression ‘Sadness’ was included. Interestingly, sadness was also found in preliminary and detailed analysis suggesting relevance for media interviews. However, identification of this facial expression could suggest this signal is often confused with fear and anger as described in Chapter 6 discussion.

Hand gestures were identified as important for media interviews for on-camera interviews in Chapter 4. However, in this analysis, hand gestures were not included using correlational methods. This could be a result of the shimmer itself. It could also be because movement is not well perceived in interviews; however, this is less likely as empirical research has found that hand gestures are important for effective communication (Argyle, 1988; Čereković & Pandžić, 2011; Gunes et al., 2008; Kim, Soyata, & Behnagh, 2018; Morgan, 2008; Poggi & D ’errico, 2011).

Conclusion

The results from this analysis reveal that facial expression (sadness), vocal affect (extreme emotion, arousal and hesitation) and honest signals (posture mirroring) are important for communication in on camera media interviews. This chapter is consistent with previous literature stating that communication is a multimodal event. Some of the signals identified were also included in the preliminary data analysis and fed back to participants in the evaluation stage. The consistency in signal identification suggests that the signals fed back to participants in the evaluation stage were appropriate for appraising performance in media interviews.

Data output analysis

Correlation Selection Feature Method

```
Attribute Evaluator (supervised, Class (nominal): 45 Category):
        Correlation Ranking Filter
Ranked attributes:
 0.34815   18 Hesitation
 0.24577   23 ExtremeEmotion
 0.21996    4 Sadness
 0.2159    37 Unsuccessful interruptions
 0.20916   34 P_Mirroring
 0.20914   11 Energy
 0.20781   24 Arousal
 0.17966   10 BrowFurrow
 0.17226   20 Embar
 0.1516    42 Volume Consistency
 0.14876   28 M_Consistency
 0.14696    3 Anger
 0.14187   36 Successful interruptions
 0.14054   40 Total Speaking
```

K1- Leave-one-out

```
Relation:     Multimodal for Weka-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribu
Instances:    67
Attributes:   8
              Sadness
              Energy
              Hesitation
              ExtremeEmotion
              Arousal
              P_Mirroring
              Unsuccessful interruptions
              Category
Test mode:    67-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          37               55.2239 %
Incorrectly Classified Instances        30               44.7761 %
Kappa statistic                          0.1075
Mean absolute error                      0.4493
Root mean squared error                  0.6594
Relative absolute error                 89.0387 %
Root relative squared error            130.3318 %
Total Number of Instances               67

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.528    0.419    0.594      0.528    0.559      0.108    0.554     0.567     Effective
                 0.581    0.472    0.514      0.581    0.545      0.108    0.554     0.493     Poor
Weighted Avg.    0.552    0.444    0.557      0.552    0.553      0.108    0.554     0.533

=== Confusion Matrix ===

  a  b   <-- classified as
 19 17 |  a = Effective
 13 18 |  b = Poor
```

K1 - Leave-one-out – Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk --
Relation:     Multimodal for Weka-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.a
Instances:    67
Attributes:   8
              Sadness
              Energy
              Hesitation
              ExtremeEmotion
              Arousal
              P_Mirroring
              Unsuccessful interruptions
              Category
Test mode:    67-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDi

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          38               56.7164 %
Incorrectly Classified Instances        29               43.2836 %
Kappa statistic                          0.1314
Mean absolute error                      0.4478
Root mean squared error                  0.5609
Relative absolute error                 88.7503 %
Root relative squared error            110.8618 %
Total Number of Instances               67

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.583    0.452    0.600      0.583    0.592      0.131    0.586     0.568     Effective
                 0.548    0.417    0.531      0.548    0.540      0.131    0.586     0.572     Poor
Weighted Avg.    0.567    0.435    0.568      0.567    0.568      0.131    0.586     0.570

=== Confusion Matrix ===

  a  b   <-- classified as
 21 15 |  a = Effective
 14 17 |  b = Poor
```

K2 - Leave one out

```
Scheme:       weka.classifiers.lazy.IBk -K 2 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.
Relation:     Multimodal for Weka-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised
Instances:    67
Attributes:   8
              Sadness
              Energy
              Hesitation
              ExtremeEmotion
              Arousal
              P_Mirroring
              Unsuccessful interruptions
              Category
Test mode:    67-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 2 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          42               62.6866 %
Incorrectly Classified Instances        25               37.3134 %
Kappa statistic                          0.2376
Mean absolute error                      0.4485
Root mean squared error                  0.5891
Relative absolute error                 88.8887 %
Root relative squared error            116.4292 %
Total Number of Instances               67

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.750    0.516    0.628      0.750    0.684      0.243    0.585     0.581     Effective
                 0.484    0.250    0.625      0.484    0.545      0.243    0.585     0.541     Poor
Weighted Avg.    0.627    0.393    0.627      0.627    0.620      0.243    0.585     0.562

=== Confusion Matrix ===

  a  b   <-- classified as
 27  9 |  a = Effective
 16 15 |  b = Poor
```

K2 - Leave one out – Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk -
Relation:     Multimodal for Weka-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.
Instances:    67
Attributes:   8
              Sadness
              Energy
              Hesitation
              ExtremeEmotion
              Arousal
              P_Mirroring
              Unsuccessful interruptions
              Category
Test mode:    67-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 2 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanD

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          44               65.6716 %
Incorrectly Classified Instances        23               34.3284 %
Kappa statistic                          0.3172
Mean absolute error                      0.4449
Root mean squared error                  0.5238
Relative absolute error                 88.1666 %
Root relative squared error            103.5249 %
Total Number of Instances               67

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.611    0.290    0.710      0.611    0.657      0.321    0.628     0.600     Effective
                 0.710    0.389    0.611      0.710    0.657      0.321    0.628     0.633     Poor
Weighted Avg.    0.657    0.336    0.664      0.657    0.657      0.321    0.628     0.615

=== Confusion Matrix ===

  a  b   <-- classified as
 22 14 |  a = Effective
  9 22 |  b = Poor
```

K3 - Leave one out

```
Scheme:       weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.c
Relation:     Multimodal for Weka-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.
Instances:    67
Attributes:   8
              Sadness
              Energy
              Hesitation
              ExtremeEmotion
              Arousal
              P_Mirroring
              Unsuccessful interruptions
              Category
Test mode:    67-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 3 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          44               65.6716 %
Incorrectly Classified Instances        23               34.3284 %
Kappa statistic                          0.3142
Mean absolute error                      0.4384
Root mean squared error                  0.5298
Relative absolute error                 86.8856 %
Root relative squared error            104.7133 %
Total Number of Instances               67

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.639    0.323    0.697      0.639    0.667      0.315    0.630     0.625     Effective
                 0.677    0.361    0.618      0.677    0.646      0.315    0.630     0.565     Poor
Weighted Avg.    0.657    0.340    0.660      0.657    0.657      0.315    0.630     0.597

=== Confusion Matrix ===

  a  b   <-- classified as
 23 13 |  a = Effective
 10 21 |  b = Poor
```

K 4 - Leave one out

```
Scheme:       weka.classifiers.lazy.IBk -K 4 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.c
Relation:     Multimodal for Weka-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.
Instances:    67
Attributes:   8
              Sadness
              Energy
              Hesitation
              ExtremeEmotion
              Arousal
              P_Mirroring
              Unsuccessful interruptions
              Category
Test mode:    67-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 4 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          42               62.6866 %
Incorrectly Classified Instances        25               37.3134 %
Kappa statistic                          0.2445
Mean absolute error                      0.4741
Root mean squared error                  0.5413
Relative absolute error                 93.9493 %
Root relative squared error            106.9865 %
Total Number of Instances               67

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.694    0.452    0.641      0.694    0.667      0.245    0.581     0.595     Effective
                 0.548    0.306    0.607      0.548    0.576      0.245    0.581     0.527     Poor
Weighted Avg.    0.627    0.384    0.625      0.627    0.625      0.245    0.581     0.563

=== Confusion Matrix ===

  a  b   <-- classified as
 25 11 |  a = Effective
 14 17 |  b = Poor
```

K4 Leave one out - bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk -- -K 4 -W 0 -A "weka.core.ne:
Relation:     Multimodal for Weka-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-R1-3,5-10,1:
Instances:    67
Attributes:   8
              Sadness
              Energy
              Hesitation
              ExtremeEmotion
              Arousal
              P_Mirroring
              Unsuccessful interruptions
              Category
Test mode:    67-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 4 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          41               61.194  %
Incorrectly Classified Instances        26               38.806  %
Kappa statistic                          0.2333
Mean absolute error                      0.4556
Root mean squared error                  0.5022
Relative absolute error                 90.288  %
Root relative squared error             99.2644 %
Total Number of Instances               67

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.528    0.290    0.679      0.528    0.594      0.240    0.641     0.617     Effective
                 0.710    0.472    0.564      0.710    0.629      0.240    0.641     0.639     Poor
Weighted Avg.    0.612    0.374    0.626      0.612    0.610      0.240    0.641     0.627

=== Confusion Matrix ===

  a  b   <-- classified as
 19 17 |  a = Effective
  9 22 |  b = Poor
```

Mann Whitney U

Ranks

	Category	N	Mean Rank	Sum of Ranks
Sadness	Effective	36	28.32	1019.50
	Poor	31	40.60	1258.50
	Total	67		
Energy	Effective	36	37.13	1336.50
	Poor	31	30.37	941.50
	Total	67		
Hesitation	Effective	36	27.39	986.00
	Poor	31	41.68	1292.00
	Total	67		
Unsuccessfulinterruptions	Effective	36	33.22	1196.00
	Poor	31	34.90	1082.00
	Total	67		
P_Mirroring	Effective	36	36.81	1325.00
	Poor	31	30.74	953.00
	Total	67		
ExtremeEmotion	Effective	36	38.10	1371.50
	Poor	31	29.24	906.50
	Total	67		
Arousal	Effective	36	36.49	1313.50
	Poor	31	31.11	964.50
	Total	67		

Test Statistics[a]

	Sadness	Energy	Hesitation	Unsuccessfulinterruptions	P_Mirroring	ExtremeEmotion	Arousal
Mann-Whitney U	353.500	445.500	320.000	530.000	457.000	410.500	468.500
Wilcoxon W	1019.500	941.500	986.000	1196.000	953.000	906.500	964.500
Z	-2.572	-1.415	-2.993	-.378	-1.270	-1.855	-1.126
Asymp. Sig. (2tailed)	.010	.157	.003	.706	.204	.064	.260

a. Grouping Variable: Category

Appendix 4.3. Results

4.4.1. Subjective Ratings of Communication Skills

4.4.1.1. Radio Interview

Trainer Ratings

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.960	.961	5

Neutral Observers

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.980	.982	5

Self-report

Reliability Statistics

Cronbach's Alpha	N of Items
.961	5

Inter-rater Agreement (trainer and neutral observers)

Intraclass Correlation Coefficient

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.362[a]	.128	.637	3.324	16	48	.001
Average Measures	.694[c]	.370	.875	3.324	16	48	.001

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A intraclass correlation coefficients using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

4.4.1.2 On-camera Interview

Trainer

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.973	.974	5

Neutral Observers

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.975	.976	5

Self-report

Reliability Statistics

Cronbach's Alpha	N of Items
.974	5

Inter-rater agreement (trainer and neutral observers)

Intraclass Correlation Coefficient

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.308[a]	.087	.590	2.939	16	48	.002
Average Measures	.640[c]	.275	.852	2.939	16	48	.002

Two-way mixed effects model where people effects are random and measures effects are fixed. a. The estimator is the same, whether the interaction effect is present or not.

b. Type A intraclass correlation coefficients using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

4.4.2. Preliminary Analysis

4.4.2.1. Vocal Affect Recognition

4.4.2.1.1. Radio Interview

Vocal Affect Recognition– Radio Interview - Energy

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"w
Relation:     LVA primacy voice_WEKA-weka.filters.unsupervised.attribute.Standardize-weka.filters.uns
Instances:    17
Attributes:   2
              Energy
              Overall performance
Test mode:    17-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances           9               52.9412 %
Incorrectly Classified Instances         8               47.0588 %
Kappa statistic                         -0.0149
Mean absolute error                      0.5
Root mean squared error                  0.6602
Relative absolute error                 97.4522 %
Root relative squared error            127.0693 %
Total Number of Instances               17

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.700    0.714    0.583      0.700    0.636      -0.015   0.457     0.562     Good
                 0.286    0.300    0.400      0.286    0.333      -0.015   0.457     0.408     Bad
Weighted Avg.    0.529    0.544    0.508      0.529    0.512      -0.015   0.457     0.499

=== Confusion Matrix ===

 a b   <-- classified as
 7 3 | a = Good
 5 2 | b = Bad
```

Vocal affect recognition – radio interview - upset

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"we
Relation:     LVA primacy voice_WEKA-weka.filters.unsupervised.attribute.Standardize-weka.filters.unsu
Instances:    17
Attributes:   2
              upset
              Overall performance
Test mode:    17-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          11               64.7059 %
Incorrectly Classified Instances         6               35.2941 %
Kappa statistic                          0.2031
Mean absolute error                      0.457
Root mean squared error                  0.5569
Relative absolute error                 89.0683 %
Root relative squared error            107.1749 %
Total Number of Instances               17

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.900    0.714    0.643      0.900    0.750      0.240    0.443     0.605     Good
                 0.286    0.100    0.667      0.286    0.400      0.240    0.443     0.487     Bad
Weighted Avg.    0.647    0.461    0.653      0.647    0.606      0.240    0.443     0.556

=== Confusion Matrix ===

 a b   <-- classified as
 9 1 | a = Good
 5 2 | b = Bad
```

Vocal Affect Recognition – Radio Interview - hesitation

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"we
Relation:     LVA primacy voice_WEKA-weka.filters.unsupervised.attribute.Standardize-weka.filters.unsu
Instances:    17
Attributes:   2
              Hesitation
              Overall performance
Test mode:    17-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances           8               47.0588 %
Incorrectly Classified Instances         9               52.9412 %
Kappa statistic                         -0.0699
Mean absolute error                      0.5538
Root mean squared error                  0.7008
Relative absolute error                107.943  %
Root relative squared error            134.8795 %
Total Number of Instances               17

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.500    0.571    0.556      0.500    0.526      -0.070   0.414     0.539     Good
                 0.429    0.500    0.375      0.429    0.400      -0.070   0.414     0.396     Bad
Weighted Avg.    0.471    0.542    0.481      0.471    0.474      -0.070   0.414     0.480

=== Confusion Matrix ===

 a b   <-- classified as
 5 5 | a = Good
 4 3 | b = Bad
```

Vocal Affect Recognition – Radio Interview – All signals

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"w
Relation:     LVA primacy voice_WEKA-weka.filters.unsupervised.attribute.Standardize-weka.filters.uns
Instances:    17
Attributes:   4
               Energy
               upset
               Hesitation
              Overall performance
Test mode:    17-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          14               82.3529 %
Incorrectly Classified Instances         3               17.6471 %
Kappa statistic                          0.6277
Mean absolute error                      0.2124
Root mean squared error                  0.3999
Relative absolute error                 41.4013 %
Root relative squared error             76.9719 %
Total Number of Instances               17

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.900    0.286    0.818      0.900    0.857      0.633    0.807     0.795     Good
                 0.714    0.100    0.833      0.714    0.769      0.633    0.807     0.713     Bad
Weighted Avg.    0.824    0.209    0.824      0.824    0.821      0.633    0.807     0.761

=== Confusion Matrix ===

 a b   <-- classified as
 9 1 | a = Good
 2 5 | b = Bad
```

4.4.2.1.2. On-camera Interview

Vocal Affect Recognition – On-camera - Content

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"
Relation:     LVA video primacy-weka.filters.unsupervised.attribute.Standardize-weka.filters.unsuper
Instances:    17
Attributes:   2
              Content
              Performance
Test mode:    17-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          11               64.7059 %
Incorrectly Classified Instances         6               35.2941 %
Kappa statistic                          0
Mean absolute error                      0.4927
Root mean squared error                  0.509
Relative absolute error                101.1862 %
Root relative squared error            100.8438 %
Total Number of Instances               17

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.647      1.000    0.786      ?        0.000     0.601     Good
                 0.000    0.000    ?          0.000    ?          ?        0.000     0.353     Bad
Weighted Avg.    0.647    0.647    ?          0.647    ?          ?        0.000     0.514

=== Confusion Matrix ===

  a  b   <-- classified as
 11  0 |  a = Good
  6  0 |  b = Bad
```

Vocal Affect Recognition – On-camera - Upset

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"w
Relation:     LVA video primacy-weka.filters.unsupervised.attribute.Standardize-weka.filters.unsuperv:
Instances:    17
Attributes:   2
              upset
              Performance
Test mode:    17-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances           8               47.0588 %
Incorrectly Classified Instances         9               52.9412 %
Kappa statistic                         -0.2047
Mean absolute error                      0.5903
Root mean squared error                  0.708
Relative absolute error                121.2325 %
Root relative squared error            140.2835 %
Total Number of Instances               17

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.636    0.833    0.583      0.636    0.609      -0.207   0.227     0.537     Good
                 0.167    0.364    0.200      0.167    0.182      -0.207   0.227     0.300     Bad
Weighted Avg.    0.471    0.668    0.448      0.471    0.458      -0.207   0.227     0.454

=== Confusion Matrix ===

 a b   <-- classified as
 7 4 | a = Good
 5 1 | b = Bad
```

Vocal Affect Recognition– On-camera - Hesitation

```
=== Run information ===

Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"we
Relation:     LVA video primacy-weka.filters.unsupervised.attribute.Standardize-weka.filters.unsupervi
Instances:    17
Attributes:   2
              Hesitation
              Performance
Test mode:    17-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          10               58.8235 %
Incorrectly Classified Instances         7               41.1765 %
Kappa statistic                          0.063
Mean absolute error                      0.4216
Root mean squared error                  0.6075
Relative absolute error                 86.5772 %
Root relative squared error            120.376  %
Total Number of Instances               17

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.727    0.667    0.667      0.727    0.696      0.064    0.530     0.661     Good
                 0.333    0.273    0.400      0.333    0.364      0.064    0.530     0.369     Bad
Weighted Avg.    0.588    0.528    0.573      0.588    0.578      0.064    0.530     0.558

=== Confusion Matrix ===

 a b   <-- classified as
 8 3 | a = Good
 4 2 | b = Bad
```

Vocal Affect Recognition– On-camera- Extreme Emotion

```
=== Run information ===

Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"we
Relation:     LVA video primacy-weka.filters.unsupervised.attribute.Standardize-weka.filters.unsupervi
Instances:    17
Attributes:   2
              ExtremeEmotion
              Performance
Test mode:    17-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          11               64.7059 %
Incorrectly Classified Instances         6               35.2941 %
Kappa statistic                          0.2273
Mean absolute error                      0.3693
Root mean squared error                  0.5654
Relative absolute error                 75.8389 %
Root relative squared error            112.0184 %
Total Number of Instances               17

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.727    0.500    0.727      0.727    0.727      0.227    0.576     0.685     Good
                 0.500    0.273    0.500      0.500    0.500      0.227    0.576     0.420     Bad
Weighted Avg.    0.647    0.420    0.647      0.647    0.647      0.227    0.576     0.591

=== Confusion Matrix ===

 a b   <-- classified as
 8 3 | a = Good
 3 3 | b = Bad
```

Vocal Affect Recognition – on-camera – all signals

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"we
Relation:     LVA video primacy-weka.filters.unsupervised.attribute.Standardize-weka.filters.unsupervi
Instances:    17
Attributes:   5
               Content
               upset
               Hesitation
               ExtremeEmotion
              Performance
Test mode:    17-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          14               82.3529 %
Incorrectly Classified Instances         3               17.6471 %
Kappa statistic                          0.6277
Mean absolute error                      0.2124
Root mean squared error                  0.3999
Relative absolute error                 43.6242 %
Root relative squared error             79.2427 %
Total Number of Instances               17

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.818    0.167    0.900      0.818    0.857      0.633    0.826     0.854     Good
                 0.833    0.182    0.714      0.833    0.769      0.633    0.826     0.654     Bad
Weighted Avg.    0.824    0.172    0.834      0.824    0.826      0.633    0.826     0.783

=== Confusion Matrix ===

 a b   <-- classified as
 9 2 | a = Good
 1 5 | b = Bad
```

4.4.2.2. Honest Signals

4.4.2.2.1. Radio Interview

Honest Signals – Radio Interview - Movement Rate

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"w
Relation:     Weka_SB_Voice Interview-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsu
Instances:    16
Attributes:   2
              M Rate
              Performance
Test mode:    16-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances           9               56.25   %
Incorrectly Classified Instances         7               43.75   %
Kappa statistic                          0.0968
Mean absolute error                      0.4449
Root mean squared error                  0.6241
Relative absolute error                 88.9706 %
Root relative squared error            121.6999 %
Total Number of Instances               16

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.600    0.500    0.667      0.600    0.632      0.098    0.550     0.650     Good
                 0.500    0.400    0.429      0.500    0.462      0.098    0.550     0.402     Bad
Weighted Avg.    0.563    0.463    0.577      0.563    0.568      0.098    0.550     0.557

=== Confusion Matrix ===

 a b   <-- classified as
 6 4 | a = Good
 3 3 | b = Bad
```

Honest Signals – Radio Interview - Movement Mirror

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"w
Relation:     Weka_SB_Voice Interview-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsu
Instances:    16
Attributes:   2
              M_Mirror
              Performance
Test mode:    16-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          12               75      %
Incorrectly Classified Instances         4               25      %
Kappa statistic                          0.4667
Mean absolute error                      0.2794
Root mean squared error                  0.4733
Relative absolute error                 55.8824 %
Root relative squared error             92.3024 %
Total Number of Instances               16

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.800    0.333    0.800      0.800    0.800      0.467    0.733     0.765     Good
                 0.667    0.200    0.667      0.667    0.667      0.467    0.733     0.569     Bad
Weighted Avg.    0.750    0.283    0.750      0.750    0.750      0.467    0.733     0.692

=== Confusion Matrix ===

 a b   <-- classified as
 8 2 | a = Good
 2 4 | b = Bad
```

Honest Signals – Radio Interview - Posture

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"w
Relation:     Weka_SB_Voice Interview-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsu
Instances:    16
Attributes:   2
              Posture
              Performance
Test mode:    16-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          10               62.5    %
Incorrectly Classified Instances         6               37.5    %
Kappa statistic                          0.1429
Mean absolute error                      0.3897
Root mean squared error                  0.5782
Relative absolute error                 77.9412 %
Root relative squared error            112.7556 %
Total Number of Instances               16

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.800    0.667    0.667      0.800    0.727      0.149    0.567     0.658     Good
                 0.333    0.200    0.500      0.333    0.400      0.149    0.567     0.417     Bad
Weighted Avg.    0.625    0.492    0.604      0.625    0.605      0.149    0.567     0.568

=== Confusion Matrix ===

 a b   <-- classified as
 8 2 | a = Good
 4 2 | b = Bad
```

-

Honest Signals - Radio Interview - Posture Mirroring

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"we
Relation:     Weka_SB_Voice Interview-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsup
Instances:    16
Attributes:   2
              P_Mirroring
              Performance
Test mode:    16-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances           9               56.25   %
Incorrectly Classified Instances         7               43.75   %
Kappa statistic                          0.0968
Mean absolute error                      0.4449
Root mean squared error                  0.6241
Relative absolute error                 88.9706 %
Root relative squared error            121.6999 %
Total Number of Instances               16

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.600    0.500    0.667      0.600    0.632      0.098    0.550     0.650     Good
                 0.500    0.400    0.429      0.500    0.462      0.098    0.550     0.402     Bad
Weighted Avg.    0.563    0.463    0.577      0.563    0.568      0.098    0.550     0.557

=== Confusion Matrix ===

 a b   <-- classified as
 6 4 | a = Good
 3 3 | b = Bad
```

Honest Signals – Radio Interview - Speed of Turn-taking

```
=== Run information ===

Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"w
Relation:     Weka_SB_Voice Interview-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsu
Instances:    16
Attributes:   2
              Speed of Turn
              Performance
Test mode:    16-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances           8               50      %
Incorrectly Classified Instances         8               50      %
Kappa statistic                         -0.1429
Mean absolute error                      0.5
Root mean squared error                  0.6668
Relative absolute error                100      %
Root relative squared error            130.0304 %
Total Number of Instances               16

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.700    0.833    0.583      0.700    0.636      -0.149   0.433     0.596     Good
                 0.167    0.300    0.250      0.167    0.200      -0.149   0.433     0.354     Bad
Weighted Avg.    0.500    0.633    0.458      0.500    0.473      -0.149   0.433     0.505

=== Confusion Matrix ===

 a b   <-- classified as
 7 3 | a = Good
 5 1 | b = Bad
```

Honest Signals – Radio Interview - Volume

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"wek
Relation:     Weka_SB_Voice Interview-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupe
Instances:    16
Attributes:   2
              Volume
              Performance
Test mode:    16-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          10               62.5    %
Incorrectly Classified Instances         6               37.5    %
Kappa statistic                          0.2
Mean absolute error                      0.3897
Root mean squared error                  0.5782
Relative absolute error                 77.9412 %
Root relative squared error            112.7556 %
Total Number of Instances               16

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.700    0.500    0.700      0.700    0.700      0.200    0.600     0.678     Good
                 0.500    0.300    0.500      0.500    0.500      0.200    0.600     0.438     Bad
Weighted Avg.    0.625    0.425    0.625      0.625    0.625      0.200    0.600     0.588

=== Confusion Matrix ===

 a b   <-- classified as
 7 3 | a = Good
 3 3 | b = Bad
```

Honest Signals – Radio Interview - Volume Mirror

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"wel
Relation:     Weka_SB_Voice Interview-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupe
Instances:    16
Attributes:   2
              Vol_Mirroring
              Performance
Test mode:    16-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          10               62.5    %
Incorrectly Classified Instances         6               37.5    %
Kappa statistic                          0.25
Mean absolute error                      0.3897
Root mean squared error                  0.5782
Relative absolute error                 77.9412 %
Root relative squared error            112.7556 %
Total Number of Instances               16

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.600    0.333    0.750      0.600    0.667      0.258    0.633     0.700     Good
                 0.667    0.400    0.500      0.667    0.571      0.258    0.633     0.458     Bad
Weighted Avg.    0.625    0.358    0.656      0.625    0.631      0.258    0.633     0.609

=== Confusion Matrix ===

 a b   <-- classified as
 6 4 | a = Good
 2 4 | b = Bad
```

Honest Signals – Radio Interview - All Signals

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"we
Relation:     Weka_SB_Video Interview-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-weka.fil
Instances:    16
Attributes:   8
              M_Rate
              M_Mirror
              Posture
              P_Mirroring
              Speed of Turn
              Volume
              Vol_Mirroring
              Performance
Test mode:    16-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          11               68.75   %
Incorrectly Classified Instances         5               31.25   %
Kappa statistic                          0.3103
Mean absolute error                      0.3346
Root mean squared error                  0.5284
Relative absolute error                 66.9118 %
Root relative squared error            103.0377 %
Total Number of Instances               16

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.800    0.500    0.727      0.800    0.762      0.313    0.650     0.707     Good
                 0.500    0.200    0.600      0.500    0.545      0.313    0.650     0.488     Bad
Weighted Avg.    0.688    0.388    0.680      0.688    0.681      0.313    0.650     0.625

=== Confusion Matrix ===

 a b   <-- classified as
 8 2 | a = Good
 3 3 | b = Bad
```

4.4.2.2.2. On-camera Interview

Honest Signals – On-camera Interview - Movement Rate

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"we
Relation:     Weka_SB_Video Interview-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-weka.fil
Instances:    16
Attributes:   2
              M_Rate
              Performance
Test mode:    16-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          10               62.5    %
Incorrectly Classified Instances         6               37.5    %
Kappa statistic                          0.25
Mean absolute error                      0.3897
Root mean squared error                  0.5782
Relative absolute error                 77.9412 %
Root relative squared error            112.7556 %
Total Number of Instances               16

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.600    0.333    0.750      0.600    0.667      0.258    0.633     0.700     Good
                 0.667    0.400    0.500      0.667    0.571      0.258    0.633     0.458     Bad
Weighted Avg.    0.625    0.358    0.656      0.625    0.631      0.258    0.633     0.609

=== Confusion Matrix ===

 a b   <-- classified as
 6 4 | a = Good
 2 4 | b = Bad
```

Honest Signals – On-camera Interview - Movement Mirror

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"w
Relation:     Weka_SB_Video Interview-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-weka.fi
Instances:    16
Attributes:   2
              M_Mirror
              Performance
Test mode:    16-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances           9               56.25   %
Incorrectly Classified Instances         7               43.75   %
Kappa statistic                          0.0968
Mean absolute error                      0.4449
Root mean squared error                  0.6241
Relative absolute error                 88.9706 %
Root relative squared error            121.6999 %
Total Number of Instances               16

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.600    0.500    0.667      0.600    0.632      0.098    0.550     0.650     Good
                 0.500    0.400    0.429      0.500    0.462      0.098    0.550     0.402     Bad
Weighted Avg.    0.563    0.463    0.577      0.563    0.568      0.098    0.550     0.557

=== Confusion Matrix ===

 a b   <-- classified as
 6 4 | a = Good
 3 3 | b = Bad
```

Honest Signals – On-camera Interview -Posture

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"we
Relation:     Weka_SB_Video Interview-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-weka.fil
Instances:    16
Attributes:   2
              Posture
              Performance
Test mode:    16-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          15               93.75   %
Incorrectly Classified Instances         1                6.25   %
Kappa statistic                          0.8621
Mean absolute error                      0.114
Root mean squared error                  0.2421
Relative absolute error                 22.7941 %
Root relative squared error             47.2082 %
Total Number of Instances               16

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.167    0.909      1.000    0.952      0.870    0.917     0.909     Good
                 0.833    0.000    1.000      0.833    0.909      0.870    0.917     0.896     Bad
Weighted Avg.    0.938    0.104    0.943      0.938    0.936      0.870    0.917     0.904

=== Confusion Matrix ===

  a  b   <-- classified as
 10  0 |  a = Good
  1  5 |  b = Bad
```

Honest Signals – On-camera Interview - Posture Mirroring

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"we
Relation:     Weka_SB_Video Interview-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-weka.fil
Instances:    16
Attributes:   2
              P_Mirroring
              Performance
Test mode:    16-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances           8               50      %
Incorrectly Classified Instances         8               50      %
Kappa statistic                         -0.1429
Mean absolute error                      0.5
Root mean squared error                  0.6668
Relative absolute error                100      %
Root relative squared error            130.0304 %
Total Number of Instances               16

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.700    0.833    0.583      0.700    0.636      -0.149   0.433     0.596     Good
                 0.167    0.300    0.250      0.167    0.200      -0.149   0.433     0.354     Bad
Weighted Avg.    0.500    0.633    0.458      0.500    0.473      -0.149   0.433     0.505

=== Confusion Matrix ===

 a b   <-- classified as
 7 3 | a = Good
 5 1 | b = Bad
```

Honest Signals – On-camera - Speed of Turn-taking

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"wek
Relation:     Weka_SB_Video Interview-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-weka.filt
Instances:    16
Attributes:   2
              Speed of Turn
              Performance
Test mode:    16-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          12               75      %
Incorrectly Classified Instances         4               25      %
Kappa statistic                          0.4667
Mean absolute error                      0.2811
Root mean squared error                  0.4813
Relative absolute error                 56.227  %
Root relative squared error             93.8567 %
Total Number of Instances               16

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.800    0.333    0.800      0.800    0.800      0.467    0.658     0.783     Good
                 0.667    0.200    0.667      0.667    0.667      0.467    0.658     0.571     Bad
Weighted Avg.    0.750    0.283    0.750      0.750    0.750      0.467    0.658     0.703

=== Confusion Matrix ===

 a b   <-- classified as
 8 2 | a = Good
 2 4 | b = Bad
```

Honest Signals – on-camera - volume

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"wel
Relation:     Weka_SB_Video Interview-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-weka.filt
Instances:    16
Attributes:   2
              Volume
              Performance
Test mode:    16-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances           6               37.5    %
Incorrectly Classified Instances        10               62.5    %
Kappa statistic                         -0.3333
Mean absolute error                      0.6103
Root mean squared error                  0.7449
Relative absolute error                122.0588 %
Root relative squared error            145.2652 %
Total Number of Instances               16

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.500    0.833    0.500      0.500    0.500      -0.333   0.333     0.563     Good
                 0.167    0.500    0.167      0.167    0.167      -0.333   0.333     0.340     Bad
Weighted Avg.    0.375    0.708    0.375      0.375    0.375      -0.333   0.333     0.479

=== Confusion Matrix ===

 a b   <-- classified as
 5 5 | a = Good
 5 1 | b = Bad
```

Honest signals – on-camera – interview - Volume Mirroring

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"we
Relation:     Weka_SB_Video Interview-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-weka.fil
Instances:    16
Attributes:   2
              Vol_Mirroring
              Performance
Test mode:    16-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances           4               25      %
Incorrectly Classified Instances        12               75      %
Kappa statistic                         -0.6
Mean absolute error                      0.6672
Root mean squared error                  0.7675
Relative absolute error                133.4329 %
Root relative squared error            149.663  %
Total Number of Instances               16

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.400    1.000    0.400      0.400    0.400      -0.600   0.267     0.563     Good
                 0.000    0.600    0.000      0.000    0.000      -0.600   0.267     0.333     Bad
Weighted Avg.    0.250    0.850    0.250      0.250    0.250      -0.600   0.267     0.477

=== Confusion Matrix ===

 a b   <-- classified as
 4 6 | a = Good
 6 0 | b = Bad
```

Honest Signals- On-camera Interview - All Signals

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"we
Relation:     Weka_SB_Video Interview-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-weka.fil
Instances:    16
Attributes:   8
              M_Rate
              M_Mirror
              Posture
              P_Mirroring
              Speed of Turn
              Volume
              Vol_Mirroring
              Performance
Test mode:    16-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          11               68.75   %
Incorrectly Classified Instances         5               31.25   %
Kappa statistic                          0.3103
Mean absolute error                      0.3346
Root mean squared error                  0.5284
Relative absolute error                 66.9118 %
Root relative squared error            103.0377 %
Total Number of Instances               16

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0.800    0.500    0.727      0.800    0.762      0.313    0.650     0.707     Good
                0.500    0.200    0.600      0.500    0.545      0.313    0.650     0.488     Bad
Weighted Avg.   0.688    0.388    0.680      0.688    0.681      0.313    0.650     0.625

=== Confusion Matrix ===

 a b   <-- classified as
 8 2 | a = Good
 3 3 | b = Bad
```

4.4.2.2.3. Facial Expression (on camera interview only)

Facial Expression - Anger

```
=== Run information ===

Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"w
Relation:     Facial Expression-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-weka.filters.
Instances:    14
Attributes:   2
              Anger
              Performance_Cat_Overall
Test mode:    14-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          10               71.4286 %
Incorrectly Classified Instances         4               28.5714 %
Kappa statistic                          0
Mean absolute error                      0.4735
Root mean squared error                  0.57
Relative absolute error                105.7751 %
Root relative squared error            118.0848 %
Total Number of Instances               14

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.714      1.000    0.833      ?        0.163     0.641     Good
                 0.000    0.000    ?          0.000    ?          ?        0.163     0.220     Bad
Weighted Avg.    0.714    0.714    ?          0.714    ?          ?        0.163     0.521

=== Confusion Matrix ===

  a  b   <-- classified as
 10  0 |  a = Good
  4  0 |  b = Bad
```

Facial Expression - Joy

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"we
Relation:     Facial Expression-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-weka.filters.u
Instances:    14
Attributes:   2
              Joy
              Performance_Cat_Overall
Test mode:    14-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances           9               64.2857 %
Incorrectly Classified Instances         5               35.7143 %
Kappa statistic                         -0.129
Mean absolute error                      0.423
Root mean squared error                  0.5705
Relative absolute error                 94.4923 %
Root relative squared error            118.1774 %
Total Number of Instances               14

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.900    1.000    0.692      0.900    0.783      -0.175   0.175     0.622     Good
                 0.000    0.100    0.000      0.000    0.000      -0.175   0.175     0.210     Bad
Weighted Avg.    0.643    0.743    0.495      0.643    0.559      -0.175   0.175     0.505

=== Confusion Matrix ===

 a b   <-- classified as
 9 1 | a = Good
 4 0 | b = Bad
```

Facial Expression - Contempt

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka
Relation:     Facial Expression-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-weka.filters.uns
Instances:    14
Attributes:   2
              Contempt
              Performance_Cat_Overall
Test mode:    14-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          11               78.5714 %
Incorrectly Classified Instances         3               21.4286 %
Kappa statistic                          0.4324
Mean absolute error                      0.321
Root mean squared error                  0.4663
Relative absolute error                 71.715  %
Root relative squared error             96.6088 %
Total Number of Instances               14

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.900    0.500    0.818      0.900    0.857      0.440    0.525     0.737     Good
                 0.500    0.100    0.667      0.500    0.571      0.440    0.525     0.462     Bad
Weighted Avg.    0.786    0.386    0.775      0.786    0.776      0.440    0.525     0.658

=== Confusion Matrix ===

 a b   <-- classified as
 9 1 | a = Good
 2 2 | b = Bad
```

Facial Expression - Brow Furrow

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"wel
Relation:     Facial Expression-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-weka.filters.ui
Instances:    14
Attributes:   2
              BrowFurrow
              Performance_Cat_Overall
Test mode:    14-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          10               71.4286 %
Incorrectly Classified Instances         4               28.5714 %
Kappa statistic                          0.1765
Mean absolute error                      0.3624
Root mean squared error                  0.5225
Relative absolute error                 80.9675 %
Root relative squared error            108.2493 %
Total Number of Instances               14

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.900    0.750    0.750      0.900    0.818      0.194    0.263     0.676     Good
                 0.250    0.100    0.500      0.250    0.333      0.194    0.263     0.296     Bad
Weighted Avg.    0.714    0.564    0.679      0.714    0.680      0.194    0.263     0.568

=== Confusion Matrix ===

 a b   <-- classified as
 9 1 | a = Good
 3 1 | b = Bad
```

Facial Expression - All Signals

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"wel
Relation:     Facial Expression-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0-weka.filters.ur
Instances:    14
Attributes:   11
              Smile
              Smirk
              Anger
              Sadness
              Disgust
              Joy
              Surprise
              Fear
              Contempt
              BrowFurrow
              Performance_Cat_Overall
Test mode:    14-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          10               71.4286 %
Incorrectly Classified Instances         4               28.5714 %
Kappa statistic                          0.1765
Mean absolute error                      0.3143
Root mean squared error                  0.5021
Relative absolute error                 70.2128 %
Root relative squared error            104.0069 %
Total Number of Instances               14

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.900    0.750    0.750      0.900    0.818      0.194    0.575     0.746     Good
                 0.250    0.100    0.500      0.250    0.333      0.194    0.575     0.339     Bad
Weighted Avg.    0.714    0.564    0.679      0.714    0.680      0.194    0.575     0.630

=== Confusion Matrix ===

 a b   <-- classified as
 9 1 | a = Good
 3 1 | b = Bad
```

4.4.3. Detailed Analysis

4.4.3.1. Social Signals Displayed During Interviews

Group Statistics

	Performance	N	Mean	Std. Deviation	Std. Error Mean
WA separate	Good	21	28.5810	2.65247	.57882
	Bad	13	20.7385	2.81826	.78164

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
WA separate	Equal variances assumed	.124	.727	8.183	32	.000	7.84249	.95843	5.89023	9.79475
	Equal variances not assumed			8.063	24.372	.000	7.84249	.97262	5.83671	9.84827

4.4.3.2. Vocal Affect Recognition

Correlation Feature Selection (0.2 cut off)

```
Search Method:
        Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 15 Range):
        Correlation Ranking Filter
Ranked attributes:
 0.3236    3 Stressed
 0.3233   13 ExtremeEmotion
 0.302     8 Hesitation
 0.2926    6 Concentrated
 0.233    14 Arousal
 0.151    10 Embar
 0.1463    2 Upset
 0.1387    5 Excited
 0.076     4 Uncertain
 0.0681   12 Imagin
 0.0557    9 BrainPower
 0.0537    7 EmoCogRatio
 0.0225   11 IThink
 0.017     1 Energy

Selected attributes: 3,13,8,6,14,10,2,5,4,12,9,7,11,1 : 14
```

K=1, 17-Fold Cross Validation

```
=== Run information ===

Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance
Relation:     LVA_test1-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-R16-weka.filte
Instances:    34
Attributes:   6
              Stressed
              Concentrated
              Hesitation
              ExtremeEmotion
              Arousal
              Range
Test mode:    17-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          23               67.6471 %
Incorrectly Classified Instances        11               32.3529 %
Kappa statistic                          0.3249
Mean absolute error                      0.3339
Root mean squared error                  0.5526
Relative absolute error                 69.6751 %
Root relative squared error            112.4115 %
Total Number of Instances               34

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.714    0.385    0.750      0.714    0.732      0.326    0.665     0.712     Good
                 0.615    0.286    0.571      0.615    0.593      0.326    0.665     0.499     Bad
Weighted Avg.    0.676    0.347    0.682      0.676    0.679      0.326    0.665     0.631

=== Confusion Matrix ===

  a  b   <-- classified as
 15  6 |  a = Good
  5  8 |  b = Bad
```

K=1, 17-Fold Cross Validation – Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk -- -K 1 -W 0 -A "weka.core
Relation:     LVA_test1-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-R16-weka.filters.u
Instances:    34
Attributes:   6
              Stressed
              Concentrated
              Hesitation
              ExtremeEmotion
              Arousal
              Range
Test mode:    17-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          23               67.6471 %
Incorrectly Classified Instances        11               32.3529 %
Kappa statistic                          0.3249
Mean absolute error                      0.374
Root mean squared error                  0.4903
Relative absolute error                 78.0323 %
Root relative squared error             99.7392 %
Total Number of Instances               34

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.714    0.385    0.750      0.714    0.732      0.326    0.681     0.777     Good
                 0.615    0.286    0.571      0.615    0.593      0.326    0.681     0.579     Bad
Weighted Avg.    0.676    0.347    0.682      0.676    0.679      0.326    0.681     0.701

=== Confusion Matrix ===

  a  b   <-- classified as
 15  6 |  a = Good
  5  8 |  b = Bad
```

K = 1, Leave-one-out Fold Cross Validation

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"wek
Relation:     LVA_test1-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribu
Instances:    34
Attributes:   6
              Stressed
              Concentrated
              Hesitation
              ExtremeEmotion
              Arousal
              Range
Test mode:    34-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          23               67.6471 %
Incorrectly Classified Instances        11               32.3529 %
Kappa statistic                          0.3249
Mean absolute error                      0.3336
Root mean squared error                  0.553
Relative absolute error                 68.4483 %
Root relative squared error            110.6328 %
Total Number of Instances               34

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.714    0.385    0.750      0.714    0.732      0.326    0.665     0.712     Good
                 0.615    0.286    0.571      0.615    0.593      0.326    0.665     0.499     Bad
Weighted Avg.    0.676    0.347    0.682      0.676    0.679      0.326    0.665     0.631

=== Confusion Matrix ===

  a  b   <-- classified as
 15  6 |  a = Good
  5  8 |  b = Bad
```

K=1, Leave-one-out Fold Cross Validation

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk -- -K 1 -W 0 -A "weka.core
Relation:     LVA_test1-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-R16-weka.filters.u
Instances:    34
Attributes:   6
              Stressed
              Concentrated
              Hesitation
              ExtremeEmotion
              Arousal
              Range
Test mode:    34-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          23               67.6471 %
Incorrectly Classified Instances        11               32.3529 %
Kappa statistic                          0.3249
Mean absolute error                      0.3749
Root mean squared error                  0.4953
Relative absolute error                 76.9264 %
Root relative squared error             99.0801 %
Total Number of Instances               34

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.714    0.385    0.750      0.714    0.732      0.326    0.685     0.795     Good
                 0.615    0.286    0.571      0.615    0.593      0.326    0.685     0.602     Bad
Weighted Avg.    0.676    0.347    0.682      0.676    0.679      0.326    0.685     0.722

=== Confusion Matrix ===

  a  b   <-- classified as
 15  6 |  a = Good
  5  8 |  b = Bad
```

K=2, Leave-one-out Fold Cross Validation

```
Scheme:       weka.classifiers.lazy.IBk -K 2 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"we
Relation:     LVA_test1-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attrib
Instances:    34
Attributes:   6
              Stressed
              Concentrated
              Hesitation
              ExtremeEmotion
              Arousal
              Range
Test mode:    34-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 2 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          22               64.7059 %
Incorrectly Classified Instances        12               35.2941 %
Kappa statistic                          0.1535
Mean absolute error                      0.3858
Root mean squared error                  0.5238
Relative absolute error                 79.1582 %
Root relative squared error            104.7793 %
Total Number of Instances               34

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.905    0.769    0.655      0.905    0.760      0.186    0.630     0.693     Good
                 0.231    0.095    0.600      0.231    0.333      0.186    0.630     0.475     Bad
Weighted Avg.    0.647    0.512    0.634      0.647    0.597      0.186    0.630     0.609

=== Confusion Matrix ===

  a  b   <-- classified as
 19  2 |  a = Good
 10  3 |  b = Bad
```

K=2, Leave-one-out Fold Cross Validation - Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk -- -K 2 -W 0 -A "weka.core.
Relation:     LVA_test1-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-R16-weka.filters.un
Instances:    34
Attributes:   6
              Stressed
              Concentrated
              Hesitation
              ExtremeEmotion
              Arousal
              Range
Test mode:    34-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 2 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          22               64.7059 %
Incorrectly Classified Instances        12               35.2941 %
Kappa statistic                          0.2302
Mean absolute error                      0.4055
Root mean squared error                  0.4923
Relative absolute error                 83.1882 %
Root relative squared error             98.4852 %
Total Number of Instances               34

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.762    0.538    0.696      0.762    0.727      0.232    0.656     0.777     Good
                 0.462    0.238    0.545      0.462    0.500      0.232    0.656     0.595     Bad
Weighted Avg.    0.647    0.424    0.638      0.647    0.640      0.232    0.656     0.707

=== Confusion Matrix ===

  a  b   <-- classified as
 16  5 |  a = Good
  7  6 |  b = Bad
```

K=3, Leave-one-out Fold Cross Validation

```
Scheme:       weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"we
Relation:     LVA_test1-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attrib
Instances:    34
Attributes:   6
              Stressed
              Concentrated
              Hesitation
              ExtremeEmotion
              Arousal
              Range
Test mode:    34-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 3 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          20               58.8235 %
Incorrectly Classified Instances        14               41.1765 %
Kappa statistic                          0.1282
Mean absolute error                      0.4423
Root mean squared error                  0.5368
Relative absolute error                 90.7562 %
Root relative squared error            107.3742 %
Total Number of Instances               34

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.667    0.538    0.667      0.667    0.667      0.128    0.566     0.652     Good
                 0.462    0.333    0.462      0.462    0.462      0.128    0.566     0.427     Bad
Weighted Avg.    0.588    0.460    0.588      0.588    0.588      0.128    0.566     0.566

=== Confusion Matrix ===

  a  b   <-- classified as
 14  7 |  a = Good
  7  6 |  b = Bad
```

K=3, Leave-one-out Fold Cross Validation- Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk -- -K 3 -W 0 -A "weka.core.
Relation:     LVA_test1-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-R16-weka.filters.un
Instances:    34
Attributes:   6
              Stressed
              Concentrated
              Hesitation
              ExtremeEmotion
              Arousal
              Range
Test mode:    34-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          20               58.8235 %
Incorrectly Classified Instances        14               41.1765 %
Kappa statistic                          0.1282
Mean absolute error                      0.4234
Root mean squared error                  0.4936
Relative absolute error                 86.8751 %
Root relative squared error             98.7432 %
Total Number of Instances               34

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.667    0.538    0.667      0.667    0.667      0.128    0.630     0.768     Good
                 0.462    0.333    0.462      0.462    0.462      0.128    0.630     0.569     Bad
Weighted Avg.    0.588    0.460    0.588      0.588    0.588      0.128    0.630     0.692

=== Confusion Matrix ===

  a  b   <-- classified as
 14  7 |  a = Good
  7  6 |  b = Bad
```

K=4, Leave-one-out Fold Cross Validation

```
Scheme:       weka.classifiers.lazy.IBk -K 4 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"we
Relation:     LVA_test1-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attrik
Instances:    34
Attributes:   6
              Stressed
              Concentrated
              Hesitation
              ExtremeEmotion
              Arousal
              Range
Test mode:    34-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 4 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          23               67.6471 %
Incorrectly Classified Instances        11               32.3529 %
Kappa statistic                          0.2609
Mean absolute error                      0.4276
Root mean squared error                  0.5025
Relative absolute error                 87.7252 %
Root relative squared error            100.5178 %
Total Number of Instances               34

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.857    0.615    0.692      0.857    0.766      0.277    0.606     0.699     Good
                 0.385    0.143    0.625      0.385    0.476      0.277    0.606     0.494     Bad
Weighted Avg.    0.676    0.435    0.667      0.676    0.655      0.277    0.606     0.621

=== Confusion Matrix ===

  a  b   <-- classified as
 18  3 |  a = Good
  8  5 |  b = Bad
```

K=4, Leave-one-out Fold Cross Validation - Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk -- -K 4 -W 0 -A "weka.core.
Relation:     LVA_test1-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-R16-weka.filters.ur
Instances:    34
Attributes:   6
              Stressed
              Concentrated
              Hesitation
              ExtremeEmotion
              Arousal
              Range
Test mode:    34-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 4 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          19               55.8824 %
Incorrectly Classified Instances        15               44.1176 %
Kappa statistic                          0.0794
Mean absolute error                      0.4224
Root mean squared error                  0.4804
Relative absolute error                 86.6742 %
Root relative squared error             96.1068 %
Total Number of Instances               34

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.619    0.538    0.650      0.619    0.634      0.080    0.623     0.761     Good
                 0.462    0.381    0.429      0.462    0.444      0.080    0.623     0.597     Bad
Weighted Avg.    0.559    0.478    0.565      0.559    0.562      0.080    0.623     0.698

=== Confusion Matrix ===

  a  b   <-- classified as
 13  8 |  a = Good
  7  6 |  b = Bad
```

Mann Whitney U Test for individual signals

Ranks

	Performance	N	Mean Rank	Sum of Ranks
Stressed	Good	21	19.98	419.50
	Bad	13	13.50	175.50
	Total	34		
Concentrated	Good	21	19.71	414.00
	Bad	13	13.92	181.00
	Total	34		
Hesitation	Good	21	15.24	320.00
	Bad	13	21.15	275.00
	Total	34		
ExtremEmo	Good	21	19.38	407.00
	Bad	13	14.46	188.00
	Total	34		
Arousal	Good	21	18.93	397.50
	Bad	13	15.19	197.50
	Total	34		

Test Statistics[a]

	Stressed	Concentrated	Hesitation	ExtremEmo	Arousal
Mann-Whitney U	84.500	90.000	89.000	97.000	106.500
Wilcoxon W	175.500	181.000	320.000	188.000	197.500
Z	-1.843	-1.648	-1.684	-1.400	-1.063
Asymp. Sig. (2-tailed)	.065	.099	.092	.161	.288
Exact Sig. [2*(1-tailed Sig.)]	.065[b]	.104[b]	.096[b]	.169[b]	.292[b]

a. Grouping Variable: Performance

b. Not corrected for ties.

Difference in selected signals by interview type

Test Statistics[a]

	Stressed	Concentrated	Hesitation	ExtremEmo	Arousal
Mann-Whitney U	139.000	106.000	98.000	130.000	131.000
Wilcoxon W	292.000	259.000	251.000	283.000	284.000
Z	-.189	-1.326	-1.602	-.500	-.465
Asymp. Sig. (2-tailed)	.850	.185	.109	.617	.642
Exact Sig. [2*(1-tailed Sig.)]	.865[b]	.193[b]	.114[b]	.634[b]	.658[b]

a. Grouping Variable: Interview

b. Not corrected for ties.

4.4.3.3. Honest Signals

Correlation Feature Selection

```
=== Attribute Selection on all input data ===

Search Method:
        Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 22 Category):
        Correlation Ranking Filter
Ranked attributes:
 0.4654    8 P_Activity
 0.4031    4 M_Consistency
 0.3793    2 M_Activity
 0.3311    7 Posture
 0.2602    3 M_Rate
 0.2279   13 Unsuccessful interruptions
 0.1742   19 Pitch
 0.1682    6 M_MirrorLag
 0.1435   20 Vol_Mirroring
 0.1416   17 Volume
 0.1305   14 Speed of Turn
 0.122     9 P_Rate
 0.0971   12 Successful interruptions
 0.0795   11 P_MirrorLag
 0.0685   15 Overlap
 0.0667   21 Vol_MirrorLag
 0.0639    1 Movement
 0.0636   16 Total Speaking
 0.0319   10 P_Mirroring
 0.0168    5 M_Mirror
 0.0132   18 Volume Consistency

Selected attributes: 8,4,2,7,3,13,19,6,20,17,14,9,12,11,15,21,1,16,10,5,18 : 21
```

K = 1, 17- Fold Cross Validation

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"wek
Relation:     Book1-weka.filters.unsupervised.attribute.Remove-R1,23,25-weka.filters.unsupervised.attri
Instances:    32
Attributes:   7
              M_Activity
              M_Rate
              M_Consistency
              Posture
              P_Activity
              Unsuccessful interruptions
              Category
Test mode:    17-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          25               78.125  %
Incorrectly Classified Instances         7               21.875  %
Kappa statistic                          0.541
Mean absolute error                      0.2363
Root mean squared error                  0.4539
Relative absolute error                 49.6579 %
Root relative squared error             92.6842 %
Total Number of Instances               32

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.800    0.250    0.842      0.800    0.821      0.542    0.788     0.814     Good
                 0.750    0.200    0.692      0.750    0.720      0.542    0.788     0.619     Bad
Weighted Avg.    0.781    0.231    0.786      0.781    0.783      0.542    0.788     0.741

=== Confusion Matrix ===

  a  b   <-- classified as
 16  4 |  a = Good
  3  9 |  b = Bad
```

K = 1, 17-Fold Cross Validation – Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk -- -K 1 -W
Relation:     Book1-weka.filters.unsupervised.attribute.Remove-R1,23,25-weka.filters.unsupervised.attribute.Remove-
Instances:    32
Attributes:   7
              M_Activity
              M_Rate
              M_Consistency
              Posture
              P_Activity
              Unsuccessful interruptions
              Category
Test mode:    17-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          25               78.125  %
Incorrectly Classified Instances         7               21.875  %
Kappa statistic                          0.541
Mean absolute error                      0.2947
Root mean squared error                  0.4099
Relative absolute error                 61.9342 %
Root relative squared error             83.6886 %
Total Number of Instances               32

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.800    0.250    0.842      0.800    0.821      0.542    0.838     0.882     Good
                 0.750    0.200    0.692      0.750    0.720      0.542    0.838     0.843     Bad
Weighted Avg.    0.781    0.231    0.786      0.781    0.783      0.542    0.838     0.867

=== Confusion Matrix ===

  a  b   <-- classified as
 16  4 |  a = Good
  3  9 |  b = Bad
```

K = 1, Leave-one-out Fold Cross Validation

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"we
Relation:     Book1-weka.filters.unsupervised.attribute.Remove-R1,23,25-weka.filters.unsupervised.attr
Instances:    32
Attributes:   7
              M_Activity
              M_Rate
              M_Consistency
              Posture
              P_Activity
              Unsuccessful interruptions
              Category
Test mode:    32-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          25               78.125  %
Incorrectly Classified Instances         7               21.875  %
Kappa statistic                          0.541
Mean absolute error                      0.2358
Root mean squared error                  0.4543
Relative absolute error                 48.6328 %
Root relative squared error             91.0742 %
Total Number of Instances               32

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.800    0.250    0.842      0.800    0.821      0.542    0.775     0.799     Good
                 0.750    0.200    0.692      0.750    0.720      0.542    0.775     0.613     Bad
Weighted Avg.    0.781    0.231    0.786      0.781    0.783      0.542    0.775     0.729

=== Confusion Matrix ===

  a  b   <-- classified as
 16  4 |  a = Good
  3  9 |  b = Bad
```

K = 1, Leave-one-out Fold Cross Validation - bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk
Relation:     Book1-weka.filters.unsupervised.attribute.Remove-R1,23,25-weka.filters.unsupervised.attribu
Instances:    32
Attributes:   7
              M_Activity
              M_Rate
              M_Consistency
              Posture
              P_Activity
              Unsuccessful interruptions
              Category
Test mode:    32-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.Euclidea

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          25               78.125  %
Incorrectly Classified Instances         7               21.875  %
Kappa statistic                          0.541
Mean absolute error                      0.2943
Root mean squared error                  0.4046
Relative absolute error                 60.7023 %
Root relative squared error             81.1125 %
Total Number of Instances               32

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.800    0.250    0.842      0.800    0.821      0.542    0.829     0.876     Good
                 0.750    0.200    0.692      0.750    0.720      0.542    0.829     0.817     Bad
Weighted Avg.    0.781    0.231    0.786      0.781    0.783      0.542    0.829     0.854

=== Confusion Matrix ===

  a  b   <-- classified as
 16  4 |  a = Good
  3  9 |  b = Bad
```

K = 2, Leave-one-out Fold Cross Validation

```
Scheme:       weka.classifiers.lazy.IBk -K 2 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka
Relation:     Book1-weka.filters.unsupervised.attribute.Remove-R1,23,25-weka.filters.unsupervised.attrib
Instances:    32
Attributes:   7
              M_Activity
              M_Rate
              M_Consistency
              Posture
              P_Activity
              Unsuccessful interruptions
              Category
Test mode:    32-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 2 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          26               81.25   %
Incorrectly Classified Instances         6               18.75   %
Kappa statistic                          0.5714
Mean absolute error                      0.3184
Root mean squared error                  0.4476
Relative absolute error                 65.6616 %
Root relative squared error             89.7233 %
Total Number of Instances               32

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.950    0.417    0.792      0.950    0.864      0.596    0.771     0.787     Good
                 0.583    0.050    0.875      0.583    0.700      0.596    0.771     0.687     Bad
Weighted Avg.    0.813    0.279    0.823      0.813    0.802      0.596    0.771     0.749

=== Confusion Matrix ===

  a  b   <-- classified as
 19  1 |  a = Good
  5  7 |  b = Bad
```

K = 2, Leave-one-out Fold Cross Validation – Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk
Relation:     Book1-weka.filters.unsupervised.attribute.Remove-R1,23,25-weka.filters.unsupervised.attrib
Instances:    32
Attributes:   7
              M_Activity
              M_Rate
              M_Consistency
              Posture
              P_Activity
              Unsuccessful interruptions
              Category
Test mode:    32-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 2 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.Euclidea

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          25               78.125  %
Incorrectly Classified Instances         7               21.875  %
Kappa statistic                          0.541
Mean absolute error                      0.3226
Root mean squared error                  0.4052
Relative absolute error                 66.5268 %
Root relative squared error             81.2219 %
Total Number of Instances               32

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.800    0.250    0.842      0.800    0.821      0.542    0.821     0.868     Good
                 0.750    0.200    0.692      0.750    0.720      0.542    0.821     0.830     Bad
Weighted Avg.    0.781    0.231    0.786      0.781    0.783      0.542    0.821     0.854

=== Confusion Matrix ===

  a  b   <-- classified as
 16  4 |  a = Good
  3  9 |  b = Bad
```

K = 3, Leave-one-out Fold Cross Validation

```
Scheme:       weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"we
Relation:     Book1-weka.filters.unsupervised.attribute.Remove-R1,23,25-weka.filters.unsupervised.attr
Instances:    32
Attributes:   7
              M_Activity
              M_Rate
              M_Consistency
              Posture
              P_Activity
              Unsuccessful interruptions
              Category
Test mode:    32-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 3 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          24               75      %
Incorrectly Classified Instances         8               25      %
Kappa statistic                          0.4839
Mean absolute error                      0.3266
Root mean squared error                  0.4039
Relative absolute error                 67.3705 %
Root relative squared error             80.9726 %
Total Number of Instances               32

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.750    0.250    0.833      0.750    0.789      0.488    0.819     0.847     Good
                 0.750    0.250    0.643      0.750    0.692      0.488    0.819     0.687     Bad
Weighted Avg.    0.750    0.250    0.762      0.750    0.753      0.488    0.819     0.787

=== Confusion Matrix ===

  a  b   <-- classified as
 15  5 |  a = Good
  3  9 |  b = Bad
```

K = 3, Leave-one-out Fold Cross Validation - Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IB
Relation:     Book1-weka.filters.unsupervised.attribute.Remove-R1,23,25-weka.filters.unsupervised.attr
Instances:    32
Attributes:   7
              M_Activity
              M_Rate
              M_Consistency
              Posture
              P_Activity
              Unsuccessful interruptions
              Category
Test mode:    32-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.Euclide

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          25               78.125  %
Incorrectly Classified Instances         7               21.875  %
Kappa statistic                          0.541
Mean absolute error                      0.3463
Root mean squared error                  0.4175
Relative absolute error                 71.4224 %
Root relative squared error             83.6862 %
Total Number of Instances               32

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.800    0.250    0.842      0.800    0.821      0.542    0.821     0.866     Good
                 0.750    0.200    0.692      0.750    0.720      0.542    0.821     0.842     Bad
Weighted Avg.    0.781    0.231    0.786      0.781    0.783      0.542    0.821     0.857

=== Confusion Matrix ===

  a  b   <-- classified as
 16  4 |  a = Good
  3  9 |  b = Bad
```

K = 4, Leave-one-out Fold Cross Validation

```
Scheme:       weka.classifiers.lazy.IBk -K 4 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"w
Relation:     Book1-weka.filters.unsupervised.attribute.Remove-R1,23,25-weka.filters.unsupervised.att
Instances:    32
Attributes:   7
              M_Activity
              M_Rate
              M_Consistency
              Posture
              P_Activity
              Unsuccessful interruptions
              Category
Test mode:    32-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 4 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          24               75      %
Incorrectly Classified Instances         8               25      %
Kappa statistic                          0.4483
Mean absolute error                      0.3385
Root mean squared error                  0.4257
Relative absolute error                 69.8242 %
Root relative squared error             85.3385 %
Total Number of Instances               32

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.850    0.417    0.773      0.850    0.810      0.453    0.785     0.827     Good
                 0.583    0.150    0.700      0.583    0.636      0.453    0.785     0.683     Bad
Weighted Avg.    0.750    0.317    0.745      0.750    0.745      0.453    0.785     0.773

=== Confusion Matrix ===

  a  b   <-- classified as
 17  3 |  a = Good
  5  7 |  b = Bad
```

K = 4, Leave-one-out Fold Cross Validation – Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk -- -K 4 -W 0 -A "weka.core.
Relation:     Book1-weka.filters.unsupervised.attribute.Remove-R1,23,25-weka.filters.unsupervised.attribute.Remove-R1,5-6,9-12,14-21
Instances:    32
Attributes:   7
              M_Activity
              M_Rate
              M_Consistency
              Posture
              P_Activity
              Unsuccessful interruptions
              Category
Test mode:    32-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 4 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          26               81.25   %
Incorrectly Classified Instances         6               18.75   %
Kappa statistic                          0.6
Mean absolute error                      0.3597
Root mean squared error                  0.4243
Relative absolute error                 74.1945 %
Root relative squared error             85.055  %
Total Number of Instances               32

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.850    0.250    0.850      0.850    0.850      0.600    0.792     0.848     Good
                 0.750    0.150    0.750      0.750    0.750      0.600    0.792     0.780     Bad
Weighted Avg.    0.813    0.213    0.813      0.813    0.813      0.600    0.792     0.822

=== Confusion Matrix ===

  a  b   <-- classified as
 17  3 |  a = Good
  3  9 |  b = Bad
```

Mann Whitney U Test for individual signals

Ranks

	Performance	N	Mean Rank	Sum of Ranks
M_activity	Good	20	14.90	298.00
	Bad	12	19.17	230.00
M_rate	Good	20	13.78	275.50
	Bad	12	21.04	252.50
M_consistency	Good	20	18.30	366.00
	Bad	12	13.50	162.00
Posture	Good	20	15.25	305.00
	Bad	12	18.58	223.00
P_activity	Good	20	13.73	274.50
	Bad	12	21.13	253.50
UnsucInterrupt	Good	20	17.43	348.50
	Bad	12	14.96	179.50

Test Statistics[a]

	M_activity	M_rate	M_consistency	Posture	P_activity	UnsucInterrupt
Mann-Whitney U	88.000	65.500	84.000	95.000	64.500	101.500
Wilcoxon W	298.000	275.500	162.000	305.000	274.500	179.500
Z	-1.246	-2.122	-1.402	-.973	-2.161	-.735
Asymp. Sig. (2-tailed)	.213	.034	.161	.330	.031	.462
Exact Sig. [2*(1-tailed Sig.)]	.224[b]	.032[b]	.170[b]	.346[b]	.029[b]	.477[b]

a. Grouping Variable: Performance

b. Not corrected for ties.

Difference in selected signals by interview type

Test Statistics[a]

	M_activity	M_rate	M_consistency	Posture	P_activity	UnsucInterrupt
Mann-Whitney U	121.000	94.000	75.000	98.000	110.000	113.500
Wilcoxon W	257.000	230.000	211.000	234.000	246.000	249.500
Z	-.264	-1.282	-1.998	-1.131	-.679	-.558
Asymp. Sig. (2-tailed)	.792	.200	.046	.258	.497	.577
Exact Sig. [2*(1-tailed Sig.)]	.809[b]	.210[b]	.047[b]	.270[b]	.515[b]	.590[b]

a. Grouping Variable: Interview

b. Not corrected for ties.

4.4.3.4. Facial Expression

Correlation Feature Selection

```
=== Attribute Selection on all input data ===

Search Method:
        Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 11 Category):
        Correlation Ranking Filter
Ranked attributes:
 0.5974   10 BrowFurrow
 0.4721    9 Contempt
 0.4061    2 Smirk
 0.3728    4 Sadness
 0.3317    1 Smile
 0.2477    8 Fear
 0.2215    6 Joy
 0.1606    3 Anger
 0.0874    5 Disgust
 0.0451    7 Surprise

Selected attributes: 10,9,2,4,1,8,6,3,5,7 : 10
```

Facial Expression – K = 1, 17-Fold Cross Validation

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.cor
Relation:     Facial expression_normalised-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupe
Instances:    14
Attributes:   8
              Smile
              Smirk
              Sadness
              Joy
              Fear
              Contempt
              BrowFurrow
              Category
Test mode:    7-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          11               78.5714 %
Incorrectly Classified Instances         3               21.4286 %
Kappa statistic                          0.3226
Mean absolute error                      0.2551
Root mean squared error                  0.4345
Relative absolute error                 58.1395 %
Root relative squared error             91.8503 %
Total Number of Instances               14

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.750    0.769      1.000    0.870      0.439    0.625     0.769     Good
                 0.250    0.000    1.000      0.250    0.400      0.439    0.625     0.464     Bad
Weighted Avg.    0.786    0.536    0.835      0.786    0.735      0.439    0.625     0.682

=== Confusion Matrix ===

  a  b   <-- classified as
 10  0 |  a = Good
  3  1 |  b = Bad
```

.

Facial Expression – K = 1, 17-Fold Cross Validation: Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk -- -K 1 -W 0 -A "weka.core
Relation:     Facial expression_normalised-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-
Instances:    14
Attributes:   8
              Smile
              Smirk
              Sadness
              Joy
              Fear
              Contempt
              BrowFurrow
              Category
Test mode:    7-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          11               78.5714 %
Incorrectly Classified Instances         3               21.4286 %
Kappa statistic                          0.3226
Mean absolute error                      0.3121
Root mean squared error                  0.4312
Relative absolute error                 71.1266 %
Root relative squared error             91.1529 %
Total Number of Instances               14

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.750    0.769      1.000    0.870      0.439    0.550     0.746     Good
                 0.250    0.000    1.000      0.250    0.400      0.439    0.550     0.639     Bad
Weighted Avg.    0.786    0.536    0.835      0.786    0.735      0.439    0.550     0.716

=== Confusion Matrix ===

  a  b   <-- classified as
 10  0 |  a = Good
  3  1 |  b = Bad
```

Facial Expression – Leave-one-out Fold Cross Validation -K = 1

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""
Relation:     Facial expression_normalised-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-R3,5,7
Instances:    14
Attributes:   8
              Smile
              Smirk
              Sadness
              Joy
              Fear
              Contempt
              BrowFurrow
              Category
Test mode:    14-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          11               78.5714 %
Incorrectly Classified Instances         3               21.4286 %
Kappa statistic                          0.3226
Mean absolute error                      0.2524
Root mean squared error                  0.4361
Relative absolute error                 56.383  %
Root relative squared error             90.3369 %
Total Number of Instances               14

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.750    0.769      1.000    0.870      0.439    0.625     0.769     Good
                 0.250    0.000    1.000      0.250    0.400      0.439    0.625     0.464     Bad
Weighted Avg.    0.786    0.536    0.835      0.786    0.735      0.439    0.625     0.682

=== Confusion Matrix ===

  a  b   <-- classified as
 10  0 |  a = Good
  3  1 |  b = Bad
```

Facial expression - Leave-one-out Fold Cross Validation, K = 1: Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk -- -K 1 -W 0 -A "weka.core
Relation:     Facial expression_normalised-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove
Instances:    14
Attributes:   8
              Smile
              Smirk
              Sadness
              Joy
              Fear
              Contempt
              BrowFurrow
              Category
Test mode:    14-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          11               78.5714 %
Incorrectly Classified Instances         3               21.4286 %
Kappa statistic                          0.3226
Mean absolute error                      0.3304
Root mean squared error                  0.4504
Relative absolute error                 73.8019 %
Root relative squared error             93.2997 %
Total Number of Instances               14

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.750    0.769      1.000    0.870      0.439    0.625     0.830     Good
                 0.250    0.000    1.000      0.250    0.400      0.439    0.625     0.583     Bad
Weighted Avg.    0.786    0.536    0.835      0.786    0.735      0.439    0.625     0.759

=== Confusion Matrix ===

  a  b   <-- classified as
 10  0 |  a = Good
  3  1 |  b = Bad
```

Facial Expression, *K = 2, Leave one out Fold Cross Validation*

```
Scheme:       weka.classifiers.lazy.IBk -K 2 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka
Relation:     Facial expression_normalised-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.ur
Instances:    14
Attributes:   8
              Smile
              Smirk
              Sadness
              Joy
              Fear
              Contempt
              BrowFurrow
              Category
Test mode:    14-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 2 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          10               71.4286 %
Incorrectly Classified Instances         4               28.5714 %
Kappa statistic                          0
Mean absolute error                      0.3342
Root mean squared error                  0.472
Relative absolute error                 74.6581 %
Root relative squared error             97.7743 %
Total Number of Instances               14

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.714      1.000    0.833      ?        0.600     0.759     Good
                 0.000    0.000    ?          0.000    ?          ?        0.600     0.343     Bad
Weighted Avg.    0.714    0.714    ?          0.714    ?          ?        0.600     0.640

=== Confusion Matrix ===

  a  b   <-- classified as
 10  0 |  a = Good
  4  0 |  b = Bad
```

Facial expression - Leave-one-out Fold Cross Validation - K = 2, Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk -- -K 2 -W 0 -A "weka.core
Relation:     Facial expression_normalised-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove
Instances:    14
Attributes:   8
              Smile
              Smirk
              Sadness
              Joy
              Fear
              Contempt
              BrowFurrow
              Category
Test mode:    14-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 2 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          10               71.4286 %
Incorrectly Classified Instances         4               28.5714 %
Kappa statistic                          0
Mean absolute error                      0.3514
Root mean squared error                  0.4683
Relative absolute error                 78.4953 %
Root relative squared error             97.0192 %
Total Number of Instances               14

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.714      1.000    0.833      ?        0.575     0.790     Good
                 0.000    0.000    ?          0.000    ?          ?        0.575     0.568     Bad
Weighted Avg.    0.714    0.714    ?          0.714    ?          ?        0.575     0.726

=== Confusion Matrix ===

  a  b   <-- classified as
 10  0 |  a = Good
  4  0 |  b = Bad
```

Facial expression - K = 3 - Leave-one-out Fold Cross Validation

```
Scheme:       weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.c
Relation:     Facial expression_normalised-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsu
Instances:    14
Attributes:   8
              Smile
              Smirk
              Sadness
              Joy
              Fear
              Contempt
              BrowFurrow
              Category
Test mode:    14-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 3 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          10               71.4286 %
Incorrectly Classified Instances         4               28.5714 %
Kappa statistic                          0
Mean absolute error                      0.3868
Root mean squared error                  0.4981
Relative absolute error                 86.4037 %
Root relative squared error            103.179  %
Total Number of Instances               14

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.714      1.000    0.833      ?        0.450     0.695     Good
                 0.000    0.000    ?          0.000    ?          ?        0.450     0.268     Bad
Weighted Avg.    0.714    0.714    ?          0.714    ?          ?        0.450     0.573

=== Confusion Matrix ===

  a  b   <-- classified as
 10  0 |  a = Good
  4  0 |  b = Bad
```

Facial Expression - K = 3 - Leave-one-out Fold Cross Validation - Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk -- -K 3 -W 0 -A "weka.core.
Relation:     Facial expression_normalised-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-
Instances:    14
Attributes:   8
              Smile
              Smirk
              Sadness
              Joy
              Fear
              Contempt
              BrowFurrow
              Category
Test mode:    14-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          10               71.4286 %
Incorrectly Classified Instances         4               28.5714 %
Kappa statistic                          0
Mean absolute error                      0.3581
Root mean squared error                  0.4682
Relative absolute error                 79.998  %
Root relative squared error             96.9923 %
Total Number of Instances               14

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.714      1.000    0.833      ?        0.625     0.830     Good
                 0.000    0.000    ?          0.000    ?          ?        0.625     0.583     Bad
Weighted Avg.    0.714    0.714    ?          0.714    ?          ?        0.625     0.759

=== Confusion Matrix ===

  a  b   <-- classified as
 10  0 |  a = Good
  4  0 |  b = Bad
```

Facial Expression - K = 4 - Leave-one-out Fold Cross Validation

```
Scheme:       weka.classifiers.lazy.IBk -K 4 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"wel
Relation:     Facial expression_normalised-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.
Instances:    14
Attributes:   8
              Smile
              Smirk
              Sadness
              Joy
              Fear
              Contempt
              BrowFurrow
              Category
Test mode:    14-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 4 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          10               71.4286 %
Incorrectly Classified Instances         4               28.5714 %
Kappa statistic                          0
Mean absolute error                      0.3796
Root mean squared error                  0.4997
Relative absolute error                 84.8109 %
Root relative squared error            103.5143 %
Total Number of Instances               14

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                1.000    1.000    0.714      1.000    0.833      ?        0.400     0.680     Good
                0.000    0.000    ?          0.000    ?          ?        0.400     0.254     Bad
Weighted Avg.   0.714    0.714    ?          0.714    ?          ?        0.400     0.558

=== Confusion Matrix ===

  a  b   <-- classified as
 10  0 |  a = Good
  4  0 |  b = Bad
```

Facial expression - K = 4 - Leave-one-out Fold Cross Validation - Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk -- -K 4 -W 0 -A "weka.core
Relation:     Facial expression_normalised-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove
Instances:    14
Attributes:   8
              Smile
              Smirk
              Sadness
              Joy
              Fear
              Contempt
              BrowFurrow
              Category
Test mode:    14-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 4 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          10               71.4286 %
Incorrectly Classified Instances         4               28.5714 %
Kappa statistic                          0
Mean absolute error                      0.362
Root mean squared error                  0.4649
Relative absolute error                 80.8727 %
Root relative squared error             96.3169 %
Total Number of Instances               14

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.714      1.000    0.833      ?        0.625     0.830     Good
                 0.000    0.000    ?          0.000    ?          ?        0.625     0.583     Bad
Weighted Avg.    0.714    0.714    ?          0.714    ?          ?        0.625     0.759

=== Confusion Matrix ===

  a  b   <-- classified as
 10  0 |  a = Good
  4  0 |  b = Bad
```

Facial Expression - Mann Whitney U Test for individual signals

Ranks

	RANGE_CAT	N	Mean Rank	Sum of Ranks
Smile	Good	10	8.65	86.50
	Bad	4	4.63	18.50
Smirk	Good	10	7.00	70.00
	Bad	4	8.75	35.00
Sadness	Good	10	7.00	70.00
	Bad	4	8.75	35.00
Joy	Good	10	8.20	82.00
	Bad	4	5.75	23.00
Fear	Good	10	7.45	74.50
	Bad	4	7.63	30.50
Contempt	Good	10	6.25	62.50
	Bad	4	10.63	42.50
BrowFurrow	Good	10	6.40	64.00
	Bad	4	10.25	41.00

Test Statistics[a]

	Smile	Smirk	Sadness	Joy	Fear	Contempt	BrowFurrow
Mann-Whitney U	8.500	15.000	15.000	13.000	19.500	7.500	9.000
Wilcoxon W	18.500	70.000	70.000	23.000	74.500	62.500	64.000
Z	-1.635	-.710	-.707	-.990	-.071	-1.770	-1.564
Asymp. Sig. (2-tailed)	.102	.478	.480	.322	.944	.077	.118
Exact Sig. [2*(1-tailed Sig.)]	.106[b]	.539[b]	.539[b]	.374[b]	.945[b]	.076[b]	.142[b]

a. Grouping Variable: RANGE_CAT

b. Not corrected for ties.

4.4.3.5. Hand Movement

Hand gestures (radio interview) – K = 1, 9-fold cross validation

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-R3
Instances:    17
Attributes:   2
              Movement
              Performance
Test mode:    9-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances           6               35.2941 %
Incorrectly Classified Instances        11               64.7059 %
Kappa statistic                         -0.3077
Mean absolute error                      0.63
Root mean squared error                  0.7581
Relative absolute error                127.7085 %
Root relative squared error            151.6356 %
Total Number of Instances               17

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.400    0.714    0.444      0.400    0.421      -0.310   0.329     0.518     Good
                 0.286    0.600    0.250      0.286    0.267      -0.310   0.329     0.366     Bad
Weighted Avg.    0.353    0.667    0.364      0.353    0.357      -0.310   0.329     0.455

=== Confusion Matrix ===

 a b   <-- classified as
 4 6 | a = Good
 5 2 | b = Bad
```

Hand gestures (radio interview) – K = 1, 9-fold cross validation: Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IE
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsuper
Instances:    17
Attributes:   2
              Movement
              Performance
Test mode:    9-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.Euclide

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances           6               35.2941 %
Incorrectly Classified Instances        11               64.7059 %
Kappa statistic                         -0.3077
Mean absolute error                      0.5941
Root mean squared error                  0.6462
Relative absolute error                120.4421 %
Root relative squared error            129.2388 %
Total Number of Instances               17

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.400    0.714    0.444      0.400    0.421      -0.310   0.271     0.539     Good
                 0.286    0.600    0.250      0.286    0.267      -0.310   0.271     0.429     Bad
Weighted Avg.    0.353    0.667    0.364      0.353    0.357      -0.310   0.271     0.494

=== Confusion Matrix ===

 a b   <-- classified as
 4 6 | a = Good
 5 2 | b = Bad
```

Hand Movements (radio interview) - K = 1 - Leave-one-out Fold Cross Validation

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"we
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsuper
Instances:    17
Attributes:   2
              Movement
              Performance
Test mode:    17-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances           5               29.4118 %
Incorrectly Classified Instances        12               70.5882 %
Kappa statistic                         -0.4571
Mean absolute error                      0.683
Root mean squared error                  0.7941
Relative absolute error                133.121  %
Root relative squared error            152.8259 %
Total Number of Instances               17

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.400    0.857    0.400      0.400    0.400      -0.457   0.271     0.513     Good
                 0.143    0.600    0.143      0.143    0.143      -0.457   0.271     0.373     Bad
Weighted Avg.    0.294    0.751    0.294      0.294    0.294      -0.457   0.271     0.455

=== Confusion Matrix ===

 a b   <-- classified as
 4 6 | a = Good
 6 1 | b = Bad
```

Hand movement (radio interview) - K = 1 - Leave-one-out Fold Cross Validation: Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk -
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervise
Instances:    17
Attributes:   2
              Movement
              Performance
Test mode:    17-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanD

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances           5               29.4118 %
Incorrectly Classified Instances        12               70.5882 %
Kappa statistic                         -0.4571
Mean absolute error                      0.6489
Root mean squared error                  0.6798
Relative absolute error                126.4765 %
Root relative squared error            130.8366 %
Total Number of Instances               17

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.400    0.857    0.400      0.400    0.400      -0.457   0.057     0.419     Good
                 0.143    0.600    0.143      0.143    0.143      -0.457   0.057     0.278     Bad
Weighted Avg.    0.294    0.751    0.294      0.294    0.294      -0.457   0.057     0.361

=== Confusion Matrix ===

a b   <-- classified as
4 6 | a = Good
6 1 | b = Bad
```

Hand movements (radio interview) - K = 2- Leave-one-out Fold Cross Validation

```
Scheme:       weka.classifiers.lazy.IBk -K 2 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"wek
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsuperv
Instances:    17
Attributes:   2
              Movement
              Performance
Test mode:    17-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 2 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          10               58.8235 %
Incorrectly Classified Instances         7               41.1765 %
Kappa statistic                          0
Mean absolute error                      0.6384
Root mean squared error                  0.6735
Relative absolute error                124.4286 %
Root relative squared error            129.6134 %
Total Number of Instances               17

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.588      1.000    0.741      ?        0.143     0.588     Good
                 0.000    0.000    ?          0.000    ?          ?        0.143     0.342     Bad
Weighted Avg.    0.588    0.588    ?          0.588    ?          ?        0.143     0.487

=== Confusion Matrix ===

  a  b   <-- classified as
 10  0 |  a = Good
  7  0 |  b = Bad
```

Hand movements (radio interview) - K = 2 - Leave-one-out Fold Cross Validation: Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk -- -K 2
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attri
Instances:    17
Attributes:   2
              Movement
              Performance
Test mode:    17-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 2 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances           3               17.6471 %
Incorrectly Classified Instances        14               82.3529 %
Kappa statistic                         -0.7
Mean absolute error                      0.6285
Root mean squared error                  0.6435
Relative absolute error                122.4909 %
Root relative squared error            123.8569 %
Total Number of Instances               17

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.300    1.000    0.300      0.300    0.300      -0.700   0.014     0.409     Good
                 0.000    0.700    0.000      0.000    0.000      -0.700   0.014     0.272     Bad
Weighted Avg.    0.176    0.876    0.176      0.176    0.176      -0.700   0.014     0.353

=== Confusion Matrix ===

a b   <-- classified as
3 7 | a = Good
7 0 | b = Bad
```

Hand movement (radio interview) - K = 3 - Leave-one-out Fold Cross Validation

```
Scheme:       weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"w
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupe
Instances:    17
Attributes:   2
              Movement
              Performance
Test mode:    17-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 3 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances           1                5.8824 %
Incorrectly Classified Instances        16               94.1176 %
Kappa statistic                         -0.7895
Mean absolute error                      0.66
Root mean squared error                  0.6691
Relative absolute error                128.6369 %
Root relative squared error            128.7684 %
Total Number of Instances               17

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.000    0.857    0.000      0.000    0.000      -0.883   0.071     0.588     Good
                 0.143    1.000    0.091      0.143    0.111      -0.883   0.071     0.340     Bad
Weighted Avg.    0.059    0.916    0.037      0.059    0.046      -0.883   0.071     0.486

=== Confusion Matrix ===

  a  b   <-- classified as
  0 10 |  a = Good
  6  1 |  b = Bad
```

Hand movements (radio interview)- K = 3 - Leave-one-out Fold Cross Validation: Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk -- -K 3
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attr
Instances:    17
Attributes:   2
              Movement
              Performance
Test mode:    17-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistanc

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances           4               23.5294 %
Incorrectly Classified Instances        13               76.4706 %
Kappa statistic                         -0.6131
Mean absolute error                      0.6059
Root mean squared error                  0.6175
Relative absolute error                118.0838 %
Root relative squared error            118.8503 %
Total Number of Instances               17

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.400    1.000    0.364      0.400    0.381      -0.618   0.000     0.407     Good
                 0.000    0.600    0.000      0.000    0.000      -0.618   0.000     0.271     Bad
Weighted Avg.    0.235    0.835    0.214      0.235    0.224      -0.618   0.000     0.351

=== Confusion Matrix ===

 a b   <-- classified as
 4 6 | a = Good
 7 0 | b = Bad
```

Hand movements (radio interview) - K = 4 - Leave-one-out Fold Cross Validation

```
Scheme:       weka.classifiers.lazy.IBk -K 4 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"we
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsuper
Instances:    17
Attributes:   2
              Movement
              Performance
Test mode:    17-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 4 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances           9               52.9412 %
Incorrectly Classified Instances         8               47.0588 %
Kappa statistic                         -0.1148
Mean absolute error                      0.5856
Root mean squared error                  0.5969
Relative absolute error                114.1285 %
Root relative squared error            114.8819 %
Total Number of Instances               17

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.900    1.000    0.563      0.900    0.692      -0.209   0.129     0.565     Good
                 0.000    0.100    0.000      0.000    0.000      -0.209   0.129     0.342     Bad
Weighted Avg.    0.529    0.629    0.331      0.529    0.407      -0.209   0.129     0.473

=== Confusion Matrix ===

 a b   <-- classified as
 9 1 | a = Good
 7 0 | b = Bad
```

Hand movements (radio interview) - K = 4 - Leave-one-out Fold Cross Validation: Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk -- -K 4 -W 0 -A "weka.core.
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-R3
Instances:    17
Attributes:   2
              Movement
              Performance
Test mode:    17-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 4 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances           4               23.5294 %
Incorrectly Classified Instances        13               76.4706 %
Kappa statistic                         -0.6131
Mean absolute error                      0.5815
Root mean squared error                  0.5922
Relative absolute error                113.3414 %
Root relative squared error            113.9758 %
Total Number of Instances               17

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.400    1.000    0.364      0.400    0.381      -0.618   0.029     0.412     Good
                 0.000    0.600    0.000      0.000    0.000      -0.618   0.029     0.273     Bad
Weighted Avg.    0.235    0.835    0.214      0.235    0.224      -0.618   0.029     0.355

=== Confusion Matrix ===

a b   <-- classified as
4 6 | a = Good
7 0 | b = Bad
```

Hand movements (on-camera interview) - K = 1, 8-fold cross validation

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"w
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupe
Instances:    16
Attributes:   2
              Movement
              Performance
Test mode:    8-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          12               75      %
Incorrectly Classified Instances         4               25      %
Kappa statistic                          0.4667
Mean absolute error                      0.2813
Root mean squared error                  0.4719
Relative absolute error                 58.0645 %
Root relative squared error             94.9308 %
Total Number of Instances               16

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.800    0.333    0.800      0.800    0.800      0.467    0.733     0.765     Good
                 0.667    0.200    0.667      0.667    0.667      0.467    0.733     0.569     Bad
Weighted Avg.    0.750    0.283    0.750      0.750    0.750      0.467    0.733     0.692

=== Confusion Matrix ===

a b   <-- classified as
8 2 | a = Good
2 4 | b = Bad
```

Hand movement (on-camera interview) - K = 1, 8-fold cross validation: Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IB
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsuperv
Instances:    16
Attributes:   2
              Movement
              Performance
Test mode:    8-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.Euclide

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          12               75      %
Incorrectly Classified Instances         4               25      %
Kappa statistic                          0.4667
Mean absolute error                      0.363
Root mean squared error                  0.4606
Relative absolute error                 74.933  %
Root relative squared error             92.6663 %
Total Number of Instances               16

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.800    0.333    0.800      0.800    0.800      0.467    0.683     0.775     Good
                 0.667    0.200    0.667      0.667    0.667      0.467    0.683     0.581     Bad
Weighted Avg.    0.750    0.283    0.750      0.750    0.750      0.467    0.683     0.702

=== Confusion Matrix ===

a b   <-- classified as
8 2 | a = Good
2 4 | b = Bad
```

Hand gestures (on-camera interview) - K = 1 - Leave-one-out Fold Cross Validation

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.c
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervise
Instances:    16
Attributes:   2
              Movement
              Performance
Test mode:    16-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          12               75      %
Incorrectly Classified Instances         4               25      %
Kappa statistic                          0.4667
Mean absolute error                      0.2794
Root mean squared error                  0.4733
Relative absolute error                 55.8824 %
Root relative squared error             92.3024 %
Total Number of Instances               16

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.800    0.333    0.800      0.800    0.800      0.467    0.733     0.765     Good
                 0.667    0.200    0.667      0.667    0.667      0.467    0.733     0.569     Bad
Weighted Avg.    0.750    0.283    0.750      0.750    0.750      0.467    0.733     0.692

=== Confusion Matrix ===

 a b   <-- classified as
 8 2 | a = Good
 2 4 | b = Bad
```

Hand movement (on-camera interview) - K = 1 - Leave-one-out cross validation: Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IB
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsuperv
Instances:    16
Attributes:   2
              Movement
              Performance
Test mode:    16-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.Euclide

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          12               75      %
Incorrectly Classified Instances         4               25      %
Kappa statistic                          0.4667
Mean absolute error                      0.34
Root mean squared error                  0.4398
Relative absolute error                 67.996  %
Root relative squared error             85.7576 %
Total Number of Instances               16

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.800    0.333    0.800      0.800    0.800      0.467    0.725     0.753     Good
                 0.667    0.200    0.667      0.667    0.667      0.467    0.725     0.619     Bad
Weighted Avg.    0.750    0.283    0.750      0.750    0.750      0.467    0.725     0.703

=== Confusion Matrix ===

 a b   <-- classified as
 8 2 | a = Good
 2 4 | b = Bad
```

Hand movements (on-camera interview) - K = 2 - Leave one out fold cross validation

```
Scheme:       weka.classifiers.lazy.IBk -K 2 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"we
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsuper
Instances:    16
Attributes:   2
              Movement
              Performance
Test mode:    16-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 2 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          11               68.75   %
Incorrectly Classified Instances         5               31.25   %
Kappa statistic                          0.2593
Mean absolute error                      0.3535
Root mean squared error                  0.4766
Relative absolute error                 70.7031 %
Root relative squared error             92.9313 %
Total Number of Instances               16

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.900    0.667    0.692      0.900    0.783      0.289    0.708     0.756     Good
                 0.333    0.100    0.667      0.333    0.444      0.289    0.708     0.535     Bad
Weighted Avg.    0.688    0.454    0.683      0.688    0.656      0.289    0.708     0.673

=== Confusion Matrix ===

 a b   <-- classified as
 9 1 | a = Good
 4 2 | b = Bad
```

Hand movements (on-camera interview) – K = 2 - Leave-one-out Fold Cross Validation: Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk -- -K 2
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attri
Instances:    16
Attributes:   2
              Movement
              Performance
Test mode:    16-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 2 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          11               68.75   %
Incorrectly Classified Instances         5               31.25   %
Kappa statistic                          0.3103
Mean absolute error                      0.3795
Root mean squared error                  0.4594
Relative absolute error                 75.9076 %
Root relative squared error             89.5882 %
Total Number of Instances               16

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.800    0.500    0.727      0.800    0.762      0.313    0.683     0.727     Good
                 0.500    0.200    0.600      0.500    0.545      0.313    0.683     0.699     Bad
Weighted Avg.    0.688    0.388    0.680      0.688    0.681      0.313    0.683     0.717

=== Confusion Matrix ===

a b   <-- classified as
8 2 | a = Good
3 3 | b = Bad
```

Hand movements (on-camera interview) – K = 3 - Leave-one-out Fold Cross Validation

```
Scheme:       weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.c
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervise
Instances:    16
Attributes:   2
              Movement
              Performance
Test mode:    16-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 3 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          10               62.5    %
Incorrectly Classified Instances         6               37.5    %
Kappa statistic                          0.2
Mean absolute error                      0.4003
Root mean squared error                  0.4893
Relative absolute error                 80.0532 %
Root relative squared error             95.4103 %
Total Number of Instances               16

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.700    0.500    0.700      0.700    0.700      0.200    0.642     0.718     Good
                 0.500    0.300    0.500      0.500    0.500      0.200    0.642     0.464     Bad
Weighted Avg.    0.625    0.425    0.625      0.625    0.625      0.200    0.642     0.622

=== Confusion Matrix ===

 a b   <-- classified as
 7 3 | a = Good
 3 3 | b = Bad
```

Hand movements (on-camera interview) – K = 3 - Leave-one-out Fold Cross Validation: Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk --
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.
Instances:    16
Attributes:   2
              Movement
              Performance
Test mode:    16-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDis

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          11               68.75   %
Incorrectly Classified Instances         5               31.25   %
Kappa statistic                          0.3103
Mean absolute error                      0.4129
Root mean squared error                  0.4818
Relative absolute error                 82.5803 %
Root relative squared error             93.9454 %
Total Number of Instances               16

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.800    0.500    0.727      0.800    0.762      0.313    0.650     0.719     Good
                 0.500    0.200    0.600      0.500    0.545      0.313    0.650     0.616     Bad
Weighted Avg.    0.688    0.388    0.680      0.688    0.681      0.313    0.650     0.681

=== Confusion Matrix ===

 a b   <-- classified as
 8 2 | a = Good
 3 3 | b = Bad
```

Hand movements (on-camera interview) – K = 4, Leave-one-out-fold cross validation

```
Scheme:       weka.classifiers.lazy.IBk -K 4 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"we
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsuper
Instances:    16
Attributes:   2
              Movement
              Performance
Test mode:    16-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 4 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances           9               56.25   %
Incorrectly Classified Instances         7               43.75   %
Kappa statistic                         -0.12
Mean absolute error                      0.4395
Root mean squared error                  0.5196
Relative absolute error                 87.9032 %
Root relative squared error            101.3265 %
Total Number of Instances               16

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.900    1.000    0.600      0.900    0.720      -0.200   0.550     0.668     Good
                 0.000    0.100    0.000      0.000    0.000      -0.200   0.550     0.416     Bad
Weighted Avg.    0.563    0.663    0.375      0.563    0.450      -0.200   0.550     0.573

=== Confusion Matrix ===

a b   <-- classified as
9 1 | a = Good
6 0 | b = Bad
```

Hand Movements (on-camera interview) – K = 4, Leave-one-out-fold cross validation: Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk -- -K 4 -W 0 -A "weka.core.ne
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-R3
Instances:    16
Attributes:   2
              Movement
              Performance
Test mode:    16-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 4 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          11               68.75   %
Incorrectly Classified Instances         5               31.25   %
Kappa statistic                          0.3103
Mean absolute error                      0.4246
Root mean squared error                  0.4886
Relative absolute error                 84.9278 %
Root relative squared error             95.2739 %
Total Number of Instances               16

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.800    0.500    0.727      0.800    0.762      0.313    0.617     0.702     Good
                 0.500    0.200    0.600      0.500    0.545      0.313    0.617     0.539     Bad
Weighted Avg.    0.688    0.388    0.680      0.688    0.681      0.313    0.617     0.641

=== Confusion Matrix ===

 a b   <-- classified as
 8 2 | a = Good
 3 3 | b = Bad
```

Hand movements (merged dataset) – K = 1, 17-fold cross validation

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"we
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1,4
Instances:    33
Attributes:   2
              Movement
              Performance
Test mode:    17-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          18               54.5455 %
Incorrectly Classified Instances        15               45.4545 %
Kappa statistic                          0.0351
Mean absolute error                      0.4573
Root mean squared error                  0.6542
Relative absolute error                 94.5655 %
Root relative squared error            132.5121 %
Total Number of Instances               33

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.650    0.615    0.619      0.650    0.634      0.035    0.533     0.634     Good
                 0.385    0.350    0.417      0.385    0.400      0.035    0.533     0.410     Bad
Weighted Avg.    0.545    0.511    0.539      0.545    0.542      0.035    0.533     0.546

=== Confusion Matrix ===

  a  b   <-- classified as
 13  7 |  a = Good
  8  5 |  b = Bad
```

Hand movements (merged dataset) – K = 1, 17-fold cross validation: Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1,4
Instances:    33
Attributes:   2
              Movement
              Performance
Test mode:    17-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.Euclidea

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          18               54.5455 %
Incorrectly Classified Instances        15               45.4545 %
Kappa statistic                          0.0351
Mean absolute error                      0.4879
Root mean squared error                  0.5836
Relative absolute error                100.9056 %
Root relative squared error            118.22   %
Total Number of Instances               33

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.650    0.615    0.619      0.650    0.634      0.035    0.481     0.615     Good
                 0.385    0.350    0.417      0.385    0.400      0.035    0.481     0.393     Bad
Weighted Avg.    0.545    0.511    0.539      0.545    0.542      0.035    0.481     0.528

=== Confusion Matrix ===

  a  b   <-- classified as
 13  7 |  a = Good
  8  5 |  b = Bad
```

Hand movements (merged dataset) – K = 1, Leave-one-out-fold cross validation

```
Scheme:       weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"we
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1,4
Instances:    33
Attributes:   2
              Movement
              Performance
Test mode:    33-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          18               54.5455 %
Incorrectly Classified Instances        15               45.4545 %
Kappa statistic                          0.0351
Mean absolute error                      0.4572
Root mean squared error                  0.6547
Relative absolute error                 92.7667 %
Root relative squared error            130.157  %
Total Number of Instances               33

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.650    0.615    0.619      0.650    0.634      0.035    0.517     0.615     Good
                 0.385    0.350    0.417      0.385    0.400      0.035    0.517     0.403     Bad
Weighted Avg.    0.545    0.511    0.539      0.545    0.542      0.035    0.517     0.531

=== Confusion Matrix ===

  a  b   <-- classified as
 13  7 |  a = Good
  8  5 |  b = Bad
```

Hand movement (merged dataset) – K = 1, Leave-one-out-fold cross validation: Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk -- -K 1 -W 0 -A "weka.core.i
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1,4
Instances:    33
Attributes:   2
              Movement
              Performance
Test mode:    33-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          18               54.5455 %
Incorrectly Classified Instances        15               45.4545 %
Kappa statistic                          0.0351
Mean absolute error                      0.485
Root mean squared error                  0.5867
Relative absolute error                 98.4002 %
Root relative squared error            116.6294 %
Total Number of Instances               33

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.650    0.615    0.619      0.650    0.634      0.035    0.442     0.559     Good
                 0.385    0.350    0.417      0.385    0.400      0.035    0.442     0.398     Bad
Weighted Avg.    0.545    0.511    0.539      0.545    0.542      0.035    0.442     0.496

=== Confusion Matrix ===

  a  b   <-- classified as
 13  7 |  a = Good
  8  5 |  b = Bad
```

Hand movements (merged dataset) – K = 2, Leave-one-out-fold cross validation

```
Scheme:       weka.classifiers.lazy.IBk -K 2 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"wek
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1,4
Instances:    33
Attributes:   2
              Movement
              Performance
Test mode:    33-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 2 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          17               51.5152 %
Incorrectly Classified Instances        16               48.4848 %
Kappa statistic                         -0.1733
Mean absolute error                      0.4903
Root mean squared error                  0.6018
Relative absolute error                 99.4695 %
Root relative squared error            119.6347 %
Total Number of Instances               33

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.850    1.000    0.567      0.850    0.680      -0.255   0.454     0.586     Good
                 0.000    0.150    0.000      0.000    0.000      -0.255   0.454     0.391     Bad
Weighted Avg.    0.515    0.665    0.343      0.515    0.412      -0.255   0.454     0.509

=== Confusion Matrix ===

  a  b   <-- classified as
 17  3 |  a = Good
 13  0 |  b = Bad
```

Hand movements (merged dataset) – K = 2, Leave-one-out-fold cross validation: Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk -- -K 2 -W 0 -A "weka.core
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1,4
Instances:    33
Attributes:   2
              Movement
              Performance
Test mode:    33-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 2 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          17               51.5152 %
Incorrectly Classified Instances        16               48.4848 %
Kappa statistic                         -0.0435
Mean absolute error                      0.5052
Root mean squared error                  0.5701
Relative absolute error                102.5035 %
Root relative squared error            113.3365 %
Total Number of Instances               33

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.650    0.692    0.591      0.650    0.619      -0.044   0.415     0.553     Good
                 0.308    0.350    0.364      0.308    0.333      -0.044   0.415     0.367     Bad
Weighted Avg.    0.515    0.557    0.501      0.515    0.506      -0.044   0.415     0.480

=== Confusion Matrix ===

  a  b   <-- classified as
 13  7 |  a = Good
  9  4 |  b = Bad
```

Hand movements (merged dataset) – K = 3, Leave-one-out-fold cross validation

```
Scheme:       weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"wek
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1,4
Instances:    33
Attributes:   2
              Movement
              Performance
Test mode:    33-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 3 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          12               36.3636 %
Incorrectly Classified Instances        21               63.6364 %
Kappa statistic                         -0.4289
Mean absolute error                      0.5469
Root mean squared error                  0.5991
Relative absolute error                110.9596 %
Root relative squared error            119.1006 %
Total Number of Instances               33

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.600    1.000    0.480      0.600    0.533      -0.456   0.338     0.544     Good
                 0.000    0.400    0.000      0.000    0.000      -0.456   0.338     0.399     Bad
Weighted Avg.    0.364    0.764    0.291      0.364    0.323      -0.456   0.338     0.487

=== Confusion Matrix ===

  a  b   <-- classified as
 12  8 |  a = Good
 13  0 |  b = Bad
```

Hand gestures (merged dataset) – K = 3, Leave-one-out-fold cross validation: Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk -- -K 3 -W 0 -A "weka.core.
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1,4
Instances:    33
Attributes:   2
              Movement
              Performance
Test mode:    33-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          16               48.4848 %
Incorrectly Classified Instances        17               51.5152 %
Kappa statistic                         -0.1567
Mean absolute error                      0.5079
Root mean squared error                  0.5562
Relative absolute error                103.0431 %
Root relative squared error            110.573  %
Total Number of Instances               33

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.700    0.846    0.560      0.700    0.622      -0.167   0.396     0.543     Good
                 0.154    0.300    0.250      0.154    0.190      -0.167   0.396     0.352     Bad
Weighted Avg.    0.485    0.631    0.438      0.485    0.452      -0.167   0.396     0.468

=== Confusion Matrix ===

  a  b   <-- classified as
 14  6 |  a = Good
 11  2 |  b = Bad
```

Hand movement (merged dataset) – K = 4, Leave-one-out-fold cross validation

```
Scheme:       weka.classifiers.lazy.IBk -K 4 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"w
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1,4
Instances:    33
Attributes:   2
              Movement
              Performance
Test mode:    33-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 4 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          16               48.4848 %
Incorrectly Classified Instances        17               51.5152 %
Kappa statistic                         -0.2276
Mean absolute error                      0.5015
Root mean squared error                  0.5511
Relative absolute error                101.7549 %
Root relative squared error            109.5636 %
Total Number of Instances               33

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.800    1.000    0.552      0.800    0.653      -0.299   0.446     0.632     Good
                 0.000    0.200    0.000      0.000    0.000      -0.299   0.446     0.416     Bad
Weighted Avg.    0.485    0.685    0.334      0.485    0.396      -0.299   0.446     0.547

=== Confusion Matrix ===

  a  b   <-- classified as
 16  4 |  a = Good
 13  0 |  b = Bad
```

Hand movements (merged dataset) – K = 4, Leave-one-out-fold cross validation: Bagging

```
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.lazy.IBk -- -K 4 -W 0 -A "weka.core.
Relation:     Weka_accel_non normal-weka.filters.unsupervised.attribute.Remove-R1,4
Instances:    33
Attributes:   2
              Movement
              Performance
Test mode:    33-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 100 iterations and base learner

weka.classifiers.lazy.IBk -K 4 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          19               57.5758 %
Incorrectly Classified Instances        14               42.4242 %
Kappa statistic                          0.061
Mean absolute error                      0.5041
Root mean squared error                  0.5444
Relative absolute error                102.2762 %
Root relative squared error            108.2323 %
Total Number of Instances               33

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.750    0.692    0.625      0.750    0.682      0.063    0.415     0.547     Good
                 0.308    0.250    0.444      0.308    0.364      0.063    0.415     0.372     Bad
Weighted Avg.    0.576    0.518    0.554      0.576    0.556      0.063    0.415     0.478

=== Confusion Matrix ===

  a  b   <-- classified as
 15  5 |  a = Good
  9  4 |  b = Bad
```

Hand gestures (merged dataset) – ANOVA

Descriptive Statistics

Dependent Variable: Gestures

Performance	Interview	Mean	Std. Deviation	N
Good	Radio	10.5200	4.60120	10
	On-camera	12.2530	3.05404	10
Bad	Radio	10.1486	4.26651	7
	On-camera	6.4567	4.25172	6
Total	Radio	10.3671	4.33248	17
	On-camera	10.0794	4.47452	16

Levene's Test of Equality of Error Variances[a,b]

		Levene Statistic	df1	df2	Sig.
Gestures	Based on Mean	.879	3	29	.463
	Based on Median	.756	3	29	.528
	Based on Median and with adjusted df	.756	3	28.557	.528
	Based on trimmed mean	.872	3	29	.467

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Dependent variable: Gestures

b. Design: Intercept + Performance + Interview + Performance * Interview

Tests of Between-Subjects Effects

Dependent Variable: Gestures

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	127.241[a]	3	42.414	2.594	.072	.212
Intercept	3043.323	1	3043.323	186.161	.000	.865
Performance	74.660	1	74.660	4.567	.041	.136
Interview	7.531	1	7.531	.461	.503	.016
Performance * Interview	57.759	1	57.759	3.533	.070	.109
Error	474.087	29	16.348			
Total	4053.237	33				
Corrected Total	601.328	32				

a. R Squared = .212 (Adjusted R Squared = .130)

Post-hoc Analysis

Radio Interview

Group Statistics

	Performance	N	Mean	Std. Deviation	Std. Error Mean
Gestures	Good	10	10.5200	4.60120	1.45503
	Bad	7	10.1486	4.26651	1.61259

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Gestures	Equal variances assumed	.019	.893	.169	15	.868	.37143	2.20300	-4.32416	5.06702
	Equal variances not assumed			.171	13.695	.867	.37143	2.17199	-4.29678	5.03964

On-camera Interview

Group Statistics

	Performance	N	Mean	Std. Deviation	Std. Error Mean
Gestures	Good	10	12.2530	3.05404	.96577
	Bad	6	6.4567	4.25172	1.73576

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Gestures	Equal variances assumed	1.107	.310	3.181	14	.007	5.79633	1.82224	1.88801	9.70465
	Equal variances not assumed			2.918	8.141	.019	5.79633	1.98634	1.22963	10.36303

Heart rate 2x2 ANOVA

Descriptive Statistics

Dependent Variable: Beats.min

Performance	Interview	Mean	Std. Deviation	N
Good	Radio	104.6985	36.36012	10
	On-camera	119.5498	43.84559	11
	Total	112.4778	40.17351	21
Bad	Radio	126.6644	40.64325	7
	On-camera	138.6704	39.22177	6
	Total	132.2056	38.80364	13
Total	Radio	113.7433	38.56532	17
	On-camera	126.2983	42.08285	17
	Total	120.0208	40.25368	34

Levene's Test of Equality of Error Variances[a,b]

		Levene Statistic	df1	df2	Sig.
Beats.min	Based on Mean	.319	3	30	.812
	Based on Median	.285	3	30	.836
	Based on Median and with adjusted df	.285	3	28.315	.836
	Based on trimmed mean	.373	3	30	.773

Tests the null hypothesis that the error variance of the dependent variable is equal across groups. a. Dependent variable: Beats.min

b. Design: Intercept + Performance + Interview + Performance * Interview

Tests of Between-Subjects Effects

Dependent Variable: Beats.min

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	4745.979[a]	3	1581.993	.974	.418	.089
Intercept	478968.583	1	478968.583	294.896	.000	.908
Performance	3373.281	1	3373.281	2.077	.160	.065
Interview	1441.379	1	1441.379	.887	.354	.029
Performance * Interview	16.179	1	16.179	.010	.921	.000
Error	48725.862	30	1624.195			
Total	543241.360	34				
Corrected Total	53471.841	33				

a. R Squared = .089 (Adjusted R Squared = -.002)

Radio Interview

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Radio	17	7.4840	6.38414	-.41	25.21
Performance_cat_overall_voice	17	1.4118	.50730	1.00	2.00

Ranks

	Performance_cat_overall_voice	N	Mean Rank
Radio	Good	10	8.30
	Bad	7	10.00
	Total	17	

Test Statistics[a,b]

	Radio
Kruskal-Wallis H	.467
df	1
Asymp. Sig.	.495

a. Kruskal Wallis Test

b. Grouping Variable: Performance_cat_overall_voice

On-camera Interview

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
SMEAN(Oncamera)	17	9.8840	7.04090	-2.12	28.83
Performance_cat_overall_video	17	1.3529	.49259	1.00	2.00

Ranks

	Performance_cat_overall_video	N	Mean Rank
SMEAN(Oncamera)	Good	11	8.09

	Bad	6	10.67
	Total	17	

Test Statistics[a,b]

	SMEAN(Oncamera)
Kruskal-Wallis H	1.010
df	1
Asymp. Sig.	.315

a. Kruskal Wallis Test

b. Grouping Variable: Performance_cat_overall_video

Difference in selected signals by interview type

Test Statistics[a]

	Gestures
Mann-Whitney U	128.500
Wilcoxon W	264.500
Z	-.270
Asymp. Sig. (2-tailed)	.787
Exact Sig. [2*(1-tailed Sig.)]	.790[b]

a. Grouping Variable: Interview

b. Not corrected for ties.

4.5. Sample Estimation for Experiment Stage

Ranks

		N	Mean Rank	Sum of Ranks
WA_Video_overall - WA_Voice_overall	Negative Ranks	0[a]	.00	.00
	Positive Ranks	14[b]	7.50	105.00
	Ties	3[c]		
	Total	17		

a. WA_Video_overall < WA_Voice_overall

b. WA_Video_overall > WA_Voice_overall

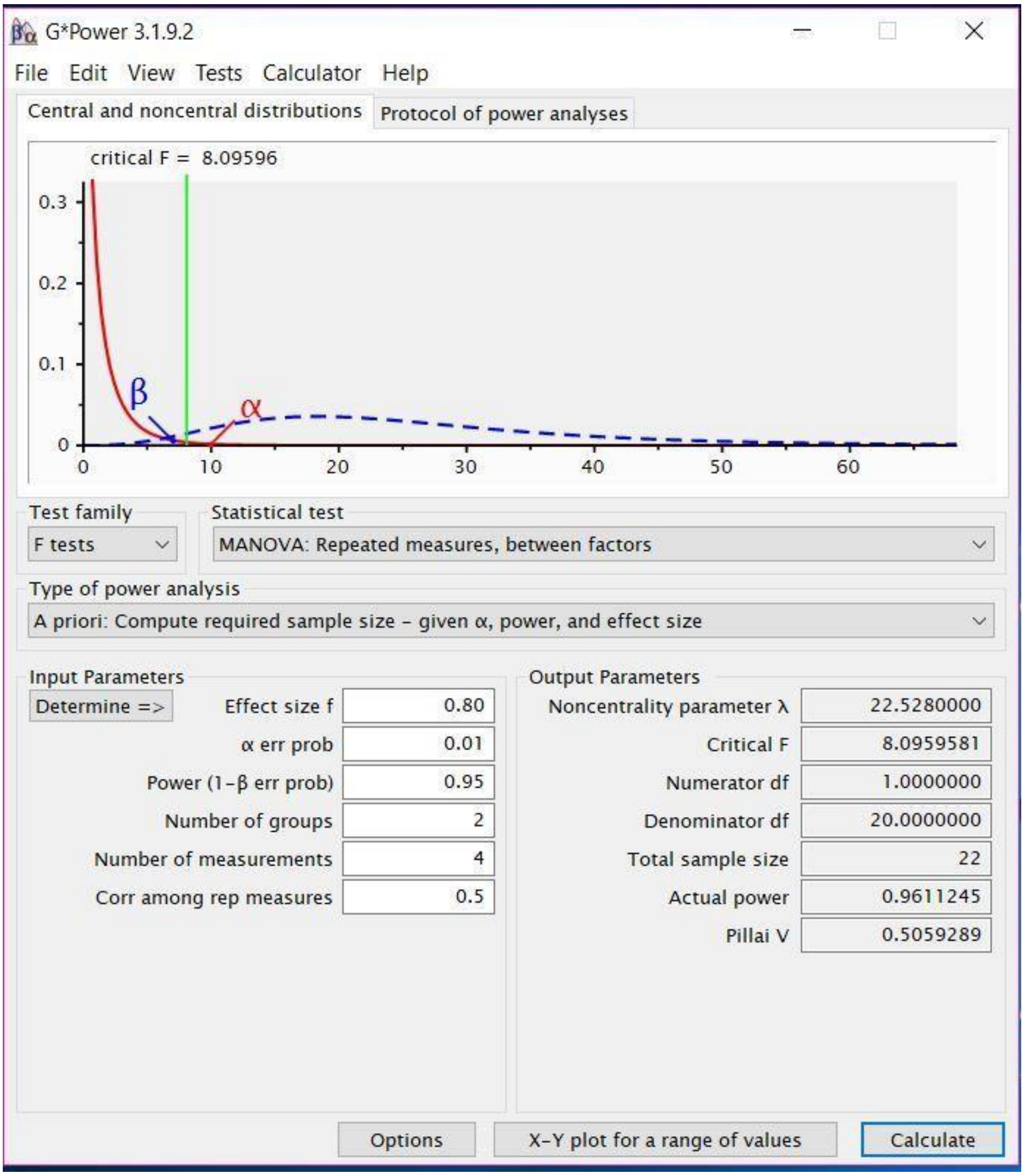
c. WA_Video_overall = WA_Voice_overall

Test Statistics[a]

	WA_Video_overall - WA_Voice_overall
Z	-3.314[b]
Asymp. Sig. (2-tailed)	.001

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.



Appendix 5

Appendix 5.1. System Usability Scale

	Facial Expression					Sociometers
	Bar Chart	Putty File	iMotions			Bar Chart
P6	37.5	77.5	67.5		P6	57.5
P8	40	65	50		P8	72.5
P4	72.5	90	80		P4	85
P12	82.5	67.5	82.5		P12	90
P7	55	70	70		P7	52.5
Mean	57.50	74.00	70.00		Mean	71.50
SD	19.76	10.09	12.87		SD	16.45
	LVA					Accelerometer
	Diamond	BarChart				Bar Chart - Imotions
P6	87.5	60			P6	72.5
P8	75	60			P8	35
P4	90	82.5			P4	70
P12	92.5	100			P12	90
P7	72.5	52.5			P7	70
Mean	83.50	71.00			Mean	67.50
SD	9.12	19.73			SD	20.00

Appendix 5.2. Thematic Analysis

Accelerometer

Key Codes

P4	P6	P7	P8	P12
Comparison element, actual data and actual movements. - Would be good to see max and minimum of performance. - Let's you know whether you are moving at all (comparison chart) -	iMotions does not give me much feedback, just my data. - Comparison with others and temporal is good, bar chart puts into perspective. - The scale for the bar chart is confusing. Need an assistant to tell me what each mean. -	iMotions doesn't give comparison but joining of the bar chart and iMotions gives me a more complete picture. - Video play back important for understanding gestures. -	I like that one is measuring temporal gestures and then that is translated into a graph. - Need an assistant to help me understand. -	The graph and iMotions are useful and straight to the point. - I would need an initial explanation of what each channel needs. -

Themes

Comparison to good performance

Temporal behaviour

Scale considerations

Initial Guidance and Explanation

Facial Expression Channel

Coding

P4	P6	P7	P8	P12
Prefer version 2 (putty tool) as it is clear to understand. - Prefer visual presentation of data. - Doesn't like video playback as doesn't like to view themselves on camera.	Prefers version 2 (putty tool) it is easy to understand when compared to graph. - Doesn't like bar chart as it is not clear. - Liked the colours of the bar chart and	Preferred video playback (iMotions) as prefers to video feedback in context of conversation. It is very intuitive. - Not good with graphs. -	Preferred the putty tool because it is intuitive but also likes the video playback because you would act naturally. - Liked bar chart because of comparison element. -	Preferred the bar chart because of the comparison element but also liked the video playback because you can see your emotions in context. - Preferred the bar chart because video

- Bar chart and putty tool suggestion. -	comparison element. - Suggestion of including a baseline - Preferred option 3 (over 2). See results in relation to own face and in context of what I am saying. - Suggestion of combining bar chart and video playback		The tool had real-time feedback that considers time. - The video playback likes the time frame too. - The bar chart scale is confusing. -	is translated nicely. - Suggests more emotions -

Themes

iMotions

Comparison to Good Performance

Temporal Behaviour

Honest Signals

Coding

P4	P6	P7	P8	P12
The bar chart is easy to interpret. - Suggest the diamond as an alternative -	Could be easier to understand. - Would need assistance. - Suggest definitions on screen.	Don't like the visual of the graph. - Like comparison element. -	Need assistance to explain each feature and what's going on. -	Simple to understand. - Nothing confusing. - Comparison good. -

Themes

Initial Guidance and Assistance

Comparison to Good Performance Essential

Voice Analysis

P4	P6	P7	P8	P12
Prefer the diamond. - It is simple	Likes both. The diamond gives a snapshot and the bar chart	Prefers diamond -	Prefers diamond due to simple emotions.	Preferred bar chart -
- Found the bar chart interesting and would choose the bar chart over diamond if min and max was included. -	compares to another person's performance - Overall prefers option 2. - Likes option 2 because of more emotions. - Understands option 2, scale needs improvement.	Helps understand the scale of emotions. - Like bar chart due to more emotions. - Comparison is preferred for improvement -	- Already explained the bar chart so understanding is fine. - Prefers comparison element of bar chart - Prefers diamond -	Did not like diamond because there was no comparison to a good performance. -

Themes

Visual display important

Comparison of Good performance

Scale of Measurement Important

Appendix 5.3. Raw Interviews

M = Monica

P= Participant

Participant 4 - Hand Gestures

M: This is the recording for the accelerometer with the bar chart together with the iMotions feedback so that's the temporal aspect of it. So, if you could. Okay so, do you like the presentations of the designs together or?

P: Yeah, the presentation is nice, definitely.

M: okay, alright. Uhm, okay so work together or yeah?

P: They do work together yes. They do assess each other.

M: Okay so why do you like it?

P: Uhm, well from this, like I said you can compare kind of a good persons or a good set of movements to the actual movements itself and you can see the difference between the two, whether you are moving too much or too little. Uhm and then combine that with the actual X y and Z it shows how you are moving as well.

M: okay and so what are the elements of the design that you find confusing?

P: Uhm, the one thing I would say is in terms of where you said that the orange bars represent a good person's movements and the blue bar represents.

M: yeah well if you remember for this one you wouldn't see the blue.

P: yes, you would only see the only, so yeah it will be good to know what kind of, it will be good to know where kind of the maximum amount of movement is. Where, if you were going above this threshold you'd be basically moving too much. If that makes sense.

M: Yeah it does. Uhm.

P: So yeah, a minimum and maximum type thing. Well, because I can see that I have moved quite decently but it will get to a stage where that movement is a bit too much.

M: Yeah

P: So, it will be nice to see where that ends.

M: Okay so the actual performance would need to be in there then, so because this would be the maximum.

P: okay so then I am above the maximum at this point.

M: well, you're below it.

P: okay, no, sorry yeah yeah yeah.

M: so, you didn't move your hands enough, let's say.

P: okay cool, yeah that will be fine.

M: okay, and then what are the elements of the design that you like?

P: Uh, I like the combination of the X Y Z type thing, and it showed that I might be moving too much in a forward or horizonal vertical way, or whether I am moving in spatial, combined with the fact of whether I am moving at all or enough.

M: Okay, yeah that.

P: So, you can have making small gestures but not a lot of it which it does show that I am doing something.

M: Yeah that makes sense. Okay, thank you very much.

P6 - Hand Gestures

M: Alright, so do you like the presentation of the design for iMotions and the comparison chart? Which would be better?

P: Uhm, I do like, I like the iMotions interface, but it doesn't really give me much feedback as in it is easy to understand but it doesn't it doesn't really tell me much about how I performed, so it doesn't really give me feedback about my performance it is just my data. I would like that it would show me how much hand movement I had so the accelerometer data so in relation to others because I think that's more feedback than just presentation of the results. So, I would learn something from that and I don't need to know exactly at which point I moved my hand, but I need to know how much or how little I moved my hands moved compared to others.

M: Okay so that's why you don't like? why do you like the bar chart.

P: Well it puts things into perspective by showing how the others performed and what is the average performance. Uhm so I know where I should improve and where I kind of over the or underperforming.

M: So, it will tell you whether to use your hands more in comparison to, or not. Okay is there anything you find confusing? Is it still the scale?

P: Yeah it is still the scale.

M: Scales confusing. Okay so there are some parts that you don't like which is the scale.

P: the scale is the only thing that I find hard to get my head around, but I am not very like.

M: yes, but it needs to be understandable. So, do you have anything else you would like to add about the accelerometer other than you want to…

P: I think it depends on the level, sorry not the level of feedback but the nature of the feedback that you are looking to give to people because if its, for example if you are showing, if you are doing this for a politician it will be very interesting for them to see when they move their hands because it might not be interesting to see overall hand movement. But they might have moved their hands at a particular moment that was really bad. So, in terms of overall feedback I think it is official to show this as how much you moved or how much you didn't move your hands compared to how you compared with the others. But if you wanted to drill deeper and find exactly whether you moved your hands in a appropriate time then you would need a timeline, that the iMotions would offer and not just the chart.

M: It is, when you move your hands it would be in conjunction with your words. Yeah, it is temporal, so what do you think about playing that back, because like you say, this is not really understandable, but to play it back and have those peaks and there would be an average in the bar chart, would that be an option. Played in conjunction.

P: aha. Is this feedback, will this feedback be given to participants as it is, or will it be given to participants in conjunction with some verbal feedback from the trainers.

M: Uhm, no it will just be this. We want to see which one is better. Okay.

P7 - Hand Gestures

M; okay so this this the recording for the accelerometer option, with the option being the combined bar chart and the video being played back on iMotions with participant number 7. Uhm so, do you like the presentation of the design with it being combined?

P: yeah, I like that because if you do not merge both of them then it could be many things to be improved because this particular system doesn't show you the level of the other comparison with the other participants. But yeah anyway the combining of those two could be more give some complete picture of the data.

M: and then so, you have explained why you like it. So, anything visually to explain why you like it or?

P: Uhm, so do you have any plan to how to combine the two?

M: oh no, just like this.

P: okay, yeah, I prefer like this layout. Yeah. What is the difference…

M: yeah so if you have a look, X Y and then Z so that's why I say it is just an overall movement and that you can tell in the beginning that you weren't moving your hands which is the baseline, which is what I was looking for.

P: oh no I mean the yellow and red.

M: yeah so x is yellow

P: ah, okay yeah, I like those layouts than the traditional bar chart.

M: So, is there anything that you find confusing or that you don't understand specifically in either or together?

P: Well if the system has kind of features integrated with the, if you can see the video with this then.

M: yes, it will be

played back. P:

okay,

M: because I recorded the video post hoc. So, with this one, you have that video played back.

P: okay then that's good.

M: okay so which of the elements of the design that you do like? Oh yes, we have discussed this earlier. So, anything that you don't like?

P: Nothing to say really.

M: Okay then, thank you very much.

P8 - Hand Gestures

M: So, this is the recording for feedback of the accelerometer for participant 8. So, do you like the presentation of the combined designs?

P: Well the, of the designs well certainly represent my gestures. So, yeah, I do like the presentation of each two together.

M: okay then so why do you like it?

P: well the fact that the first one is actually comparing behavioural of my gestures and then the temporal behaviour of the gestures is being translated into the graph. So, I can see that there is a behaviour of my temporal and a behavioural of my overall performance.

M: Right, okay I see what you're saying. Right, so are there any elements of the design, in conjunction with each other that you find confusing or that you don't understand?

P: well again if I could just, I need an assistant to tell me what's going on with these two, so if someone could tell me what this is and what is that and

M: Right, okay I see what you're saying. Right, so are there any elements of the design, in conjunction with each other that you find confusing or that you don't understand?

P: well again if I could just, I need an assistant to tell me what's going on with these two, so if someone could tell me what this is and what is that and how is it, how is it, how this represents that one so that would be really helpful.

M: okay okay, yeah well you would have the voice playing back there so you're alright with the temporal aspect of the… and then the bar chart in comparison to the others. Then are there elements of the design that you really like, I know you mentioned.

P: Well, the fact that it in the first device and then it is translated into the other format on the second one of the bar chart is very interesting for me.

M: okay okay, and then visually as well.

P: well visually, yes, I do like the visualisation of the.

M: what about the bar chart?

P: It is really good, it represents everything that happened in the first device.

M: well in a different way but yeah. Thank you so much. Do you have anything else that you would like to add?

P: no.

M: Okay thank you very much?

P12 - Hand Gestures

M: This the recording for the accelerometer with presentation of iMotions and bar chart in combination for participant 12. So, do you the presentation of the design of the combination of the two?

P: I do really it is really easy to understand, I like it. It is straight to the point.

M: okay you have just answered the second one. So, what are the elements that you might find confusing or that you don't understand.

P: Uhm, nothing in particular.

M: So, if you had a look at iMotions you'd understand that.

P: I thought, that was hand motions.

M: No, it's not eye motions it's I- Motions

P: oh haha, that's my bad. Uhm, yes, I would understand it, well the only thing I would need an explanation of the Z Y and X and what they were but obviously once I knew that. Also, you can tell in the X you know, which is where you're moving your hands is when you weren't, that I combination with that too which shows you how you move and how much you should move I think it's alright.

M: alright, so what are the elements that you like that you have just answered. Okay thank you very much.

P4 – Facial Expression

M: okay so this is the recording for the facial expression and feedback on feedback. So, the, okay so of the three options; that is option 1, option 2 was the feedback, the live feedback and then option three was the video played back together with your performance of facial expressions. So, the one I just showed you now. Okay. So, the first question is, so which of the versions do you like best?

P: I prefer version 2.

M: Version 2. Right, so why?

P: So, because I don't like to see myself visually represented again like. It's probably because I am a statistical person and I like to see numbers.

M: okay so yeah okay.

P: statistical person does like to see numbers and graphs and visual representation but when I actually see myself doing something I don't ahh.

M: So then how would you compare that, so then what is the difference between Options 2 and options 1.

P; There isn't much difference because I learned the same thing from both of them. It's just that I prefer seeing numbers over… M: video playback

P: a video of myself that's just a personal character.

M: so, then options one and two, you'd still pick 2?

P: still pick two.

M: why?

P: Ah, because I prefer just the way that the graphs are displayed and the way that things looked.

M: okay that's perfect. Uh, okay so and then uh which of the less preferred uh so the less preferred being the first one and the option 3, what don't you like about it.

P: sorry which more so between the first one?

M: No sorry so if you were given this, what don't you like about it.

P: uhm, what I don't like about it? M: yeah P Uhm, it might be difficult because there isn't anything that I dislike like about it uhm, but it just when I saw the second system it was just I prefer that one a lot better than this. It is obviously still not … and I still enjoy seeing. That's why I said I don't really dislike it. Uhm if I think of a reason why I dislike it I don't think I have any. Can I say that I don't have any?

M: That's fine. And then you said for the third option which is the playback for the video you don't like that one?

P: The system itself is amazing, uhm it's very good but it's just personally I don't like to see myself being played back like the visual play back.

M: That's fine, so you prefer option 2. So, of the design that you preferred, which is the second one uhm which of the elements did you not like?

P: which of the elements did I not like? Uhm, again I don't have a not like here. If I can say that. Can I say that?

M: Yeah you can I mean.

P: And that's the reason why I chose the second one because there wasn't anything really that I didn't like.

M: okay so you loved it, okay.

P: It was very straight forward.

M: okay that's wonderful, uhm so is there anything else that you would like to add with regards to the three designs.

P: add in terms of?

M: Feedback about how we are giving feedback to you as the participant. Which one is easier to understand for you, I mean visually ah, anything really.

P: The second one for me was very easy to understand the last one is also very easy for me to understand uhm it might be good If you just show them side by side and then at that point you can get preference from people. If that makes sense. So, for example if I saw both systems alongside, I would say okay, this is how I did visually or in terms of video feedback and this is how I did statistically.
That will be nice to see. So, have both of them side by side if possible. If not, I prefer the second one.

M: Okay, alright okay, thank you.

P6 – Facial Expression

M: So first one question, please tell me which of the versions of the design you like the best. Is it the option 1 with the bar chart and the normalised data or is it the uhm the tool that records your facial expressions live?

P: I like this option, option b (putty tool) because it is easy to understand. Uhm I don't, I am not really good with graphs. Uhm so it would be confusing for me to figure out what why some of the bars are going downwards and some are going upwards, so if I glanced at it first time I wouldn't understand that as well with the first option, while with the second option I could kind of understand that you know that I showed a bit of fear, a bit of disgust, and there was one instance where I felt sadness, so it is easy for me to understand without any training, to just see it without any help. Option b is better for me.

M: Okay that's perfect, thank you very much. So why did you like this version better than the others?

P: It's more visual I think and its more, it makes more, that timeline of emotions more it's easier to understand than just the graph.

M: okay that's fine. Okay. Alright and then uhm so you're saying the scale is the issue with the bar chart?

P: yeah.

M: Right wonderful. Okay and were there any elements of the less preferred version that you liked, If so what?

P: I like the colours, I like knowing that the orange its kind of like a performance of other people and the blue is mine, so I can see how mine is compared, but still wouldn't understand, for example, in surprise mine is higher than the orange one, so the other people, but I won't really know what that means. Does it mean that I felt more surprised?

M: yeah so it means that you felt more surprised than the average person that performed good.

P: So, there were more instances of surprise or was my surprise was longer?

M: Well no, more instances because this is presented as an average a normalised average.

P: In fear, my graph goes down, and the other people's graph goes, so that means I showed.

M: So, the average of people is 0 so you showed less fear, so you just showed less fear. The zero is the mean and then, so on a bell curve you'd be on the lower end.

P: I am not very good with graphs.

M: But this is why we are doing this. This is fantastic, thank you very much. Uhm okay so you liked the colours and the visual comparison. Okay and were there any elements of the preferred version that you liked? If so, what?

P: Uhm, I liked that it was shown in little icons in little, in separate little sections for each emotion. I don't like that it doesn't show that for other people's results, I just see mine. I obviously cannot judge from that whether it is good or bad. I can see my results, but I cannot see whether a lot of fear is good. What is necessarily good or not. Or average even. But I like how they are all split into emotions and I like the squares. I would have preferred if they, I know you didn't ask me that, but I would have preferred if, you know where they have the numbers. That doesn't really mean anything to me so perhaps if there was a baseline. Instead of showing, so the y axis shows the what does it show now? Is that the measurement?

M: That's the maximum measurement

P: and the x axis is …

M: is time, well relative to that interview. So, if the interview was 20 minutes then that would still be 200 but that would be the number of people.

P: So maybe if there was another colour there that showed the average of what is considered good and perhaps a baseline.

M: Okay thank you.

P6 – Facial Expression

M: this is option 3 now so please tell me which of the versions, well

that you like P: I like option 3.

M: oh, you like option 3, so why?

P: shall I tell you why?

M: why did you like this version (option 3) than the others?

P: Um because it shows my face so that I can see the results in context if that makes sense, so rather than seeing a summary or a graph of the results, I see the results in relation to my face. And then, which I think is very useful for feedback because I can see where I smiled, and I can see exactly what happened when I smiled. So, if the feedback says I smiled too much or frown too much then I can go back and review where I frowned. So, I can tell myself if I frowned too much, cause maybe I frowned while I was waiting for the question. Do you see what I mean cause maybe I need to see exactly what, cause if someone says you're frowning too much because you don't know what your face does, so maybe if you go back and see it then it helps with feedback. It will help me at least. On the other hand, I find it a little bit too complicated because it does not show a snapshot of how I done. It just shows the raw data.

M: so, what about using option 1 and option 3 together?

P: I was thinking more option 2 and option 3.

M: Because I was just thinking for comparison wise. To see how the participants performed.

P: I see yeah.

M: okay so of option 2/3, you prefer option?

P: I prefer option 3,

M: okay so yes, I would not illustrate the actual performance but just good performance and say okay, so we can look at you performed here so this would be the comparison graph. Then have a look and say on average people showed less fear and here is recorded a lot of fear.

P: Yeah because you have to combine this with what other people have done. Because this might say that, I don't know I smiled I know don't know, a few times, but this doesn't necessarily say that this is good or not. Do you see what I mean?

M: Absolutely, okay.

P: it just tells you what I have done. That I was engaged, but still that doesn't mean that engagement, it looks like it's a lot of engagement but that's not compared to what the others have done so I cannot judge if it is good or bad. So, something between combining option 3 with option 1 with the feature of the other people you know with the average of performance indicator, then that mean I can see my face and I can see where I went wrong or where I went well compared to others, so It is a bit more complete.

M: Okay that's wonderful, thank you.

P7 – Facial Expression

M: So, this is the recording for participant number 7 for the design options for facial expressions feedback. So please tell me, of the 3 options, option 1 being the bar chart, option is the tool that provides the emotion peaks in the, and option 3 is the iMotions where you are able to view your video.
So, which of the 3 options do you like?

P: I prefer the iMotions.

M: so, the one where you can see your video?

P: because I can see all the different emotions in the video and at the same time the video helps me to understand why and at what point each specific emotion and what exactly happened in that situation, so I can get more information about the emotions. But why there are specific emotions.

M: Alright, thank you. Okay, of the less preferred version which is the tool that provides you the peaks or the bar chart, uhm what are the elements of each that you liked? So, of the tool, what are the features that you liked?

P: I liked well, the graphs, the peaks that the graphs gives me some information about the level of each emotion. So, it just looks intuitive. So, it is very clear and easy to recognise which one is the level of certain emotions. So, that I liked.

M: okay and then what about the bar chart?

P: uhm, bar chart, well I am not good at graphs. So well personally I don't like bar charts. So, for me it's quite difficult to understand what it means.

M: okay so even if you had someone sitting here explaining that the zero is the mean?

P: yes, even with that.

M: okay, so of the preferred version, which is the iMotions playback of the video, what were the elements that you didn't like?

P: uhm, well, nothing special that I don't like. But I can hear the voice of the video?

M: yes?

P: There's nothing special that I don't like, okay.

M: okay thank you.

P8 – Facial Expression

M: Okay so this is the recording or the interview for the feedback for facial expressions options for providing feedback to participants about their performance for which you were a participant. So, of the three options for facial expression please tell me which of the versions of the design that you liked the best? Was it the bar chart? Was it the, you know, the one that recorded in real time? Or was it iMotions which allows you to view your video together with the peaks in iMotions. Which one would you prefer?

P: For me I certainly prefer the second one.

M: So, the tool? Okay that's wonderful. Okay, alright. Why would you do you prefer this version compared to the other ones?

P: For me I think it is very hard to understand the first one. And then for the third one because you are talking with human to human then when I realise that for the third option and then my hand tends to cover my face because when you talk --- then my hand is moving. I believe that when the hand moves to the face I think it will make a false positive in the graph that is presented. I think I don't know I think so for this one machine-machine interaction then there will be the camera and me. The camera will only look at me and detect my face so that it will have the results very clear and accurate compared to human to human.

M: Okay so which one is clearer in how you performed compared to other participants? So if you, which one is better for you to understand how you performed and how you can improve?

P: How I perform, if I were alone at that time and given the three options. Okay maybe I would choose the third one.

M: Okay so why would you select the third one?

P: Uhm, well because if I were alone at that time and someone was watching me or a camera sort of. Uhm, because of I would tend to act naturally. When I am interacting with the machine I don't think there is…

M: So, the output of the tool. Is it easy for you to understand? So, is it the graph where zero is the mean? So that would be excessive or but bear in mind you have to compare yourself to other participants because how would you know how you done?

P: If that is the case, can I change my mind?

M: yes of course, I mean I want you to give a full answer.

P: Well it maybe that one.

M: Why is it?

P: Well if I look at my results it clearly said that I am happy.

M: So that is the bar chart.

P: yes, the bar chart. It clearly says that I am joyous enough during the interview.

M: Okay so that's fine, so you can clearly see your emotions. Okays so next is what are the elements of the less preferred version that you liked. So, of options 2 and 3. So 2 is the tool and the third one is the one you viewed here. So, of the tool what did you like about it?

P: What do I like about the tool? It is a real time performance.

M: Okay so what that would help you improve your performance then?

P: because I can see that my results clearly in real time so that would people there are actually learning by time line in the first minute. I didn't get my results very good. But maybe in the second minute so that I would improve so much.

M: Oh, I see what you're saying.

P: because I learn from my previous experience the first time, one minute but then I would improve. So that is what I like about this one.

M: Oh so, so this one view here so. Alright so what about this one then?

P: This option, uhm.

M: So, I'll just run that option again so that you can see it. Oh yeah because it hasn't recorded. Uhm but yeah, the valence would be feedback.

P: Okay, so one question, what is the difference between.

M: These figures are those little charts and then this is the option 3 which is this one where you can view the time line.

P: Oh, so this is the time line so the second one, Well I can see that in each time what is the result of my valence what I did fear the most and did I disgust the most. What I was surprised the most during time frame.

M: okay I see what you're saying. And this one? What do you think about this one.

P: well for this one I would well actually because my face is in there and then that you would catch the temporal behaviour on my face which is very good which in terms of facial expression that point is relevant if you want to make regression problem or classification problem.

M: Okay so of the 3 you prefer the bar chart?

P: I think of the 3 I will prefer the bar chart.

M: okay so were there elements of the bar chart that you don't like? If so what?

P: Uhm none, how should I say. How do you produce this graph?

M: so, what I do is I normalise the data food good and then I normalise your data and then I compare all the participants that performed good and then I use your actual data and that's your actual data and that's the average performance for performance

P: or maybe should I say that the way your produce this result doesn't really reflects on me.

M: but let me explain why I did it. Each feature produces in its real form or in its raw form and produces so, one would produce a maximum and another would produce another of 70 so I have normalised them all so that I could have a real value for each so that it could be each across all features. Yeah, that is the only way, well not only, there might be another way but that is why I am doing this so that I can get feedback. Uhm but I mean if I normalise the zero is the mean, so you must not show anger, because the majority of participants show no anger. Well because it is the mean, any other ones would be a false positive. Uhm, but do you see what I am saying.

P: Well the scale does not reflect on me.

M: Okay so everyone has an issue with the scale. Okay so that is, is there anything else you would like to add about nay of the options?

P: Maybe again, reflects on the scale.

M: reflects, could you elaborate on that please.

P: The values, for this one, for the blue is my actual performance so the good performance is the orange.

M: is the good performance.

P: My one is very high on joy but actually it is on the negative scale on the good performance.

M: I mean the other way to present it is on a bell curve.

P: Ah yeah,

M: do you see what I mean? Just for a person that knows that has seen graphs before, a person of the public has not seen a bell curve. So, this was the other option. Maybe a bell curve would work?

P: Maybe.

M: but I have tried to think for people that have never been exposed to research or statics or a graph. Well this is great thank you very much. Okay I am going to end then and show you another one.

P12 – Facial Expression

M: okay so this the recording for facial expression for participant number 12. Uhm so I have shown you the three options for the design and we would like your feedback, so I am going to ask you some questions about the three options. Yeah okay, so please tell me which of the versions of the design that you like the best, of the three. So, the first one was the bar chart, the second one was the tool and the third one was the iMotions where you were able to have seen or watch your video.

P: So, when you say best you mean to say which one is going to help me perform better?

M: yes.

P: If I am being honest then I would say probably the first one, the first one is the bar chart.

M: Okay so why do you prefer this version compared to the others?

P: Because it says what a good performance is and what you performed as, so you are able to adjust your performance based on what the bar chart tells you is a good performance. But I also like the last one.

M: Okay. So why do you like the last one?

P: because it is in real-time, it is recording your face and you can change your emotions or facial expressions like instantly, so I like that one.

M: Yeah but you got to think more like an interview setting. So, you would be focused on the interview and then watch yourself afterwards.

P: So, if that is the case then I would say the first one.

M: So, of the less preferred version, so the iMotions, which is not less preferred but for improvement as you say so of the iMotions and the tool that provided the real time feedback, which of the elements of that you liked? So, let's start with iMotions being the third option where you are able to see your video being played back to you.

P: Which elements I liked of it? I liked the fact that you can see it being translated so from the video straight into the peaks.

M: So, peaks being the actions.

P: yeah.

M: what about the tool? The second one, the one where the graphs would be showed.

P: I liked that, because you could see the variance from the video, same as the first one. I like the scale, and I liked. It just made sense, the way it was done. You could see when you needed more fear and when you needed less joy for example. Or something like that, that's the reason I preferred still the first one, that's because the first one told you what the performance was, and the rest didn't.

M: Okay thank you so that was a good option. So, of the one that you liked, so the bar chart, the first option. What were the features that you didn't like or some of the elements that you don't like?

P: The colours.

M: you don't like the colours?

P: I am joking. Ya no, perhaps maybe more like, cause that's just 1,2,3,4,5,6 obviously there's a lot more to an interview. So maybe more.

M: oh yes of course but this is limited to facial expressions and there will be other channels that I will show you. There is a theorist that proposed that this is there are 6 basic emotions that are observed across cultures. So, these are for facial expressions that he proposed which are based on facial expressions. So that is why we have selected these. There are other ones, but they are not the 6 basic emotions.

P: There is not a lot that I don't like about this if I am being honest.

M: okay but you would say to add more to that.

P: Yeah, I would say add more.

M: Okay that is wonderful. Thank you.

P4 - Sociometric Badges

M: So, this is the recording for the sociometric badges uhm ah. So, do you like the presentation of the design?

P: Yes, simple.

M: okay why?

P: Yeah, because you can see the bars in terms of how high the bars are compared to the other bar you know you are doing better or worse. It's quite easy to interpret.

M: okay and then uhm, are there elements of the design that you find confusing?

P: not really no. Like I said, it's quite easy to understand or interpret.

M: okay, so then what are the elements of the design that you do like?

P: well, like I said earlier the comparison of the bars to each other you can see which one is higher and lower, you either know that you are doing better or worse. And also, you can see kind of the scale going on and you can see which area you are lacking in. I think in this case my speed I might have been speaking a bit too much. Which says volume speed is quite above what the regular… I might be speaking a bit too fast.

M: yeah, well compared to the mean, yes. I mean the other option that could do is a bell curve here. Do you see what I mean cause that is the mean and it illustrates how much you deviate from it compared to the person, but I thought this would be the best way of doing it.

P: I mean this is a good way because you can always. Cause you know like this thing, you know the other system that you had. The one with the diamond thing? If you had that for a similar thing in the spreadsheet, that would have been much easier to read. Because with this thing, the minute you said, the minute you saw that frequency thing.

M: Right okay, I will have a look at that, but the pie chart.

P: I don't' think it's called a pie chart, it's called something else. The reason why is because when we used to play football or FIFA quite a lot and we had the same thing and it tells straight away which players are better than the other players. I know it's called something else.

M: Okay. That's interesting.

P: I'll message you if I find it.

M: okay thank you. Okay so that is everything for the sociometric badges.

P6 - Sociometric Badges

M: This is the sociometric badge interview, uh, for feedback for the sociometric badges. So, uhm, do you like the presentation of this design for the sociometric badges?

P: I don't particularly like it. But then again, I don't dislike it. I think it could be a bit easier. It could have been designed so that it is a bit easier to understand. But because I have seen this particular style of showing the data for the voice analysis and the emotion analysis it becomes a bit clearer in my head how I would read this graph. So, it looks a bit easier now than it did when you first showed me. Uhm, again I like that it shows in a difference in a difference colour, the other people performance. Uhm, and I would need more uhm explanation about what some of those values of those uhm metrics, like what does overlap mean. But apart from that I think.

M: Well during we would explain what they mean.

P: but I think of overall uhm it's okay. It wouldn't be my absolute favourite, but I think you could read it and you could be useful.

M: Okay so that's why you like it, is that right? Okay, so then uhm are there elements that you find confusing or that you don't understand.

P: Uhm, no apart from that I would need explanation of what those values mean, I am assuming that these would happen if this was a real system. So, you could perhaps roll over our mouse and it would tell you what overlap means. Uhm, the thing I like the most is the colour.

M: Okay you like the colour. Uhm, okay that's great, thank you. And then uhm so the elements that you like are the ones that you have mentioned are the colour.

P: Perhaps it would be easier for first time users for people like, it would be easy if I could have access to this and I could, if I drag my mouse, let's say, my results for overlap and if I left my mouse there I would get a pop up to say an explanation of that result. So, for this instance say for example let's say this means that you dah dah dah dah and you explain the result. I am not sure how easy it will be to create that system but uhm or you could have some more general information like this is the result of how many times you spoke over someone else.

M: Right okay so verbal ones. Yeah okay so you want verbal popups. Okay, I'll have a look at the pop-up, that way the participant can come back all the time and have a little look over it. Okay that's wonderful thank you. Okay so we are done with that, do you have anything else to add?

P: No.

P7 - Sociometric Badges

M: This is the recording for sociometric badges for participant number 7 for one option of the feedback. Uhm, so, do you like the way this data is presented?

P: Uhm well, I really don't say I hate it, but I don't like. I don't prefer.

M: Okay, so why don't you prefer it?

P: well to some degree, the bar graph explains different information in one place but well this personally I don't like the visual of the explaining the data in a graph.

M: Okay, and then are there elements of the design that you find confusing or you don't understand?

P: other features?

M: well, anything that you find confusing or you don't understand?

P: uhm, well, what I particularly don't like is the normally bar chart or graphs have two different levels, so the average is the mean level and then, how can I say, the uhm,

M: So, there would be two sections, I just thought that it would be right next to each other and you would see how much you need to improve or reduce. That's the only reason why I have done it that way. I see what you're saying that it is not a typical graph that you would understand.

P: yeah.

M: Okay, so what are the elements of the design that you do like?

P: Two bar graphs compare the level of the data each other quite clearly. So that's good part. So I can see the, you know, okay this one you know, this one more than this one, so I can see exactly and directly compare those two graphs between the data. That the only thing. M: okay thank you

P8 - Sociometric Badges

M: This is the recording for the sociometric badges so how you felt the feedback was presented. So, the first question is, do you like the presentation of the design?

P: Overall?

M: Yes, overall.

P: yes, I do.

M: okay what do you like about it?

P: that there is a more classification on there, and there is a more an activity inside the sociometer. And then then the representation of it is really good.

M: why?

P: What else? I can see clearly how I am highly active, and my volume is very fast and then I can compare myself with the majority of the people in the interview. I can see that I am really good at one point but maybe I am below average on another point probably where I can improve more during the interview.

M: yes, that is wonderful, thank you very much for that. Uhm okay so are there elements of the design that you find confusing or that you don't understand?

P: Well, first if I see it I don't understand it but thanks to you, if someone can assist me to understand what the whole thing because I didn't want to successful interruption so how interruption is sometimes successful how interruption can be unsuccessful. If someone can explain to me what is going on so that would be good.

M: Well for the explanation, yeah there does need to be explanation. Okay well you have explained the elements of the design that you do like. That is wonderful thank you. So is there anything else that you would like to add to this?

P: Uhm, no.

M: Okay thank you, that is wonderful. Alright, thank you.

P12 - Sociometric Badges

M: So, this is the recording for sociometric badges for feedback for the design for feedback for participant interview number 12. So uhm, do you like the presentation of the design?

P: yes, I do.

M: Great, can you tell me why do you like it?

P: because it is simple, very simple.

M: can you elaborate?

P: So, since GCSE we have been trained to look at bar charts, it is the simplest form of displaying data.

M: what about the scale, is that understandable?

P: 100 percent. That would be a yes, it is very understandable.

M: So, are there elements of the design that you might find confusing or that you don't understand?

P: No, no there isn't.

M: so, if I wasn't here would you be able to.

P: yes. Just that one though

M: okay that is fine thank you very much and then uhm what are the elements of the design that you do like for feedback?

P: Well basically that it directly compares it to what a good performance is. Cause you can tell that this person spoke too softly, you know. This person spoke too much. You see what I am saying, yes, it is simple, it is very simple to read.

M: Okay thank you very much

P4 – Vocal Affect Recognition

M: so, this is the voice recording for voice emotion recognition software and which option the participant likes. SO, please tell me which of the version you like?

P: I like the version, is this version 2?

M: that's version one. The diamond.

P: The diamond, alright I like version 1.

M: wonderful can you tell me why you prefer this one compared to that one?

P: I compared kind of the diamond thing to kind of filling up the pie. Uhm and you can see like what areas of the pie chart and you need to kind of like and is easy to read and very quick without really scanning too much.

M: okay

P: so not much cognitive reasoning is required so I can see oh I need to be more confident or I need to be more uhm concentrated it the way that I am speaking essentially.

M: okay and then of the less preferred versions not really less as you

say but… P: they are quite similar to be honest uhm.

M: so, what were the elements of the less preferred version that you liked?

P: well the thing that I didn't like with this?

M: The things that you didn't like.

P: I like that it's got all the uhm what do we call them? Emotions I'd call them or...

M: yeah, emotions or features

P: The features at the bottom, I can understand all the features at the bottom I can see areas that I am lacking and areas that I need to improve, similar to the pie example but because it is laid out flat it is just a bit more, not annoying but you need to think about it a bit more.

M: It requires an increase in cognitive load.

P: yes Exactly.

M: okay so of the preferred design which elements of the design don't you like?

P: some people might not understand some of these frequencies of energy at the bottom uhm so maybe that can be represented in some other way. Maybe similar to like another pie graph or maybe even similar to these like bars and so more of uhm, because we don't know what the minimum or the maximum here is so something that can show that this is the minimum amount of speaking you have and here is the maximum.

M: Well that's why I decided to do this design. Because it gives you somewhat of a comparison being the maximum or the minimum. Do you see what I mean?

P: I think within this system or the first one you could also like design this pie chart or this diamond type thing and that point I would probably take this one (motioned to bar chart) over this one (motioned to diamond).

M: Pie chart?

P: I don't know, would you call this a pie chart, or it's called something else.

M: Oh, that's called the uhm, the actual shape.

P: Yes, but I can understand both options, I understand these frequencies here ….

M: Okay I will look into it. Thank you so much. So okay. We are done with that. Thank you.

P6 – Vocal Affect Recognition

M: This is the voice recording for voice analysis feedback uhm option 1 and 2. So please tell me, which of the versions did you like best from the voice analysis?

P: Uhm, I like both, I like certain features of one and I like certain features of 2. I like the feature of one that was like a snapshot of uhm your voice analysis like, so it showed you the video of the person and then it showed you the emotions that it could detect from your voice. Well the dominant ones. I like how it changed the light. But then I like the option of the second one that shows what the others have done, and you faired against the others. I feel that the second one might be a bit more useful for feedback because it just gives you an overall snapshot of what you have done. Cause you might not have necessarily want to watch yourself again. Uhm, so it's a quicker way, the option B (option 2) give you a quicker way of just overall seeing how you've done and how you've done compared to others.

M: Okay, so would you say it would be a good option to join the two?

P: It would be a good option to join the two but if I had to choose, I'd probably choose option B. (Option 2). Because it is less time consuming to get a quick overall view of how you have done.

M: Uhm, what do you mean in the sense of time consuming?

P: because you can just see, in a minute, how you have done when compared to the others. You don't have to watch the entire.

M: well, it would give you a summary in the end, so you press online recording and it records, you press end and it gives you a summary.

P: Okay, but still I would prefer this version because it is more emotions.

M: So why did you pick this version (option 2)?

P: Because it shows what others have done.

M: Okay so there is a comparison?

P: yeah, because if you had not shown me that the others have not exhibited much energy I would definitely not have thought that I have shown more energy than others. So if you did show me the orange bit and it just showed me that I have shown loads of energy, like option 1 would that would still be out of context for me. I would still need to see what the others have done.

M: Yeah of course. Okay, and were there any elements of the less preferred version, so the diamond, that you liked, if so, what?

P: it was the visual appeal of it and it was quite simple to understand, you didn't have a graph, you just had that diamond shape.

M: okay so the diamond, you preferred that.

P: yeah because it didn't have any number or, so it wasn't hard to understand. The bigger the diamond that is leaning towards an emotion is stronger bit. Or the more exhibited it was. So it was easy to understand.

M: Okay. That's great, thank you. And then uhm were there any elements of the preferred version that you don't like? If so, what? So, of this one, what don't you like?

P: Uhm, I think it's, the thing I don't like is that some go below 0 and some go above zero.

M: Okay that's the scale issue for you. But if it was explained to you before?

P: Yeah if it was explained to me before that would have been fine.

M: Because I mean as a participant…

P: Yeah yeah yeah. Because I can kind of understand when both graphs are below 0 or above 0. But I would need to figure out what it means when one of them are below and one of them are above in upset. But mine is below and other peoples are above.

M: So, it's literally that they showed more that they were upset, and you showed less. I mean the whole point is to not show upset because…

P: but then wouldn't that be the same if the blue bit was above 0 and the orange bit was below?

M: Yes, that's exactly right.

P: That's what I can't get my head around.

M: Well, okay maybe the scale needs to be worked out. Because at the moment the normalised data is the only way of presenting the data that way. And the quickest way, because you know I have to do a lot of exporting files and calculating and then put it in the table and show you.

P: No, no I understand that there was a reason for you to do this. I am just saying that.

M: No but that's good, thank you. We need to work out the scale.

P: I just keep thinking to myself okay so that one that means they did more, and I showed less. So, it's not just easy to look at it and get it straight away.

M: So, the diamond was, okay so you do prefer the diamond, but you like this one because of the comparison. So, would you have shown that the option, having, like we did with the facial expression having the diamond presented and then just the good performance.

P: Okay yes, that would be the easiest.

M: okay, okay we are done with the voice. Is there anything else you would like to add?

P: No that's fine.

P7 – Vocal Affect Recognition

M: This is the recording for the voice analysis for emotion recognition software for two options for participant number 7. So, can you please tell me which of the versions of the design that you like the best? Option 1 is the diamond shaped feedback and the option 2 is the bar chart.

P: Uhm, yeah, I prefer option 1 that shows the infographic the emotions and scales so that time and that graphic provides very visualised the information so that helps me understand you know the scale of the specific emotions. Also, the emotional profile that the features are quite good because they are quiet. I think that those are looks pretty systematic to see the different kind of views of the data. So, I prefer this one.

M: Okay so then of this version, which is the bar chart what do you like?

P: Well, it is very hard to find what I like. Uhm, it's as you already explained what each level means and what, you know, each specific bar means so I now I get what they are. So, it's not really difficult to understand the graphs bar data. So that's good part. Well when I see it the first time it quite makes many emotions on the axis, but you explained each by each so now I get it. So, it's has its quite complex emotion. So that's another good part.

M: Okay and then of the preferred design what don't you like?

P: Uhm, if I don't know whether they provide those features or not, if that bar or the scale provides the exact number of the scale.

M: uhm no.

P: Then yeah, it's much more accurate and yeah.

M: so, you're saying the comparison would be good? Well with regards to the maximum you could go?

P: yeah.

M: okay, that's wonderful thank you.

P8 – Vocal Affect Recognition

M: SO, this is the recording interview for voice emotion recognition options for the design stage. Uhm, okay so, of the two options for the voice emotion recognition software for feedback, which of the versions did you like the best?

P: for me I would choose the second.

M: so that was the first one.

P: Yeah, the first one.

M: okay and why did you like this version?

P: Well first it's the fact that the interface is wonderful because of the diamond ring over there. Well if I am pretty confident and I am pretty concentrated or energetic then it will show that and give me the energetic graphs and the confident graph. So that reflects on my emotions during that time. I like that because of the interface that I like the most. I can relate the most.

M: Okay so why do you like this version, oh you explained. So, were there any elements of the less preferred version that you liked?

P: The last thing that I liked was well the one that I liked. Because it shows for all, it summarises my interview in one graph. So, I can see my interview, during my half an hour interview what is my emotions during that time. But even that it is not in real time, it summarises the whole things emotions during the interview.

M: and you can also tell, I mean, okay. Sure, and then were there any elements of this design that you didn't like for option 1?

P: For option 1 what I didn't like uhm, well, none. I wish I could learn for something from that.

M: So, to sum up which one would you use to better perform uhm well which ever one you understand better to improve compared to others.

P: Well I would choose the first one. If I want to improve?

M: Well yes, that's why we are doing this is because we want feedback because we want to improve. Do you think we could use these?

P: To compare?

M: Yes, so to look at this and say, this is what you did.

P: okay so this is what I did, how you want to improve and how you want to produce that one.

M: so, you prefer the diamond if you want to improve?

P: yeah yeah. I prefer the diamond.

M: Okay so how would you improve then? Okay so if the consensus is that you seemed quite uneasy, now how do you know that that is something that is important for an interview? Do you see what I mean? So, you are in an interview and you illustrate on the diamond that you are uneasy, how would you know that being uneasy in an interview. Do you see what I mean?

P: If I am being uneasy, so there must be an uneasy scale over here. If I am uneasy during that time. Here all of them, they are very positive, such as stressful.

M: Oh yeah there is a scale.

P: all of this is quite positive maybe you should add negative emotions like uneasy.

M: oh, I couldn't add, but I see what you're saying. Okay so, I mean, you'd use that. I mean they would say, overall, they would say – while this gives you a comparison. This gives you a comparison for what is good. This one, I mean before the interview, they would say you need to perform confidently, energetically and these types of things would be produced in the previous study. So, they will pick out what is important for a good interview and tell you on that basis. Yeah, and then we would record it.
So, you would you would pick the diamond, not perfect because it is beautiful and easy to understand.
Okay, right. Is there anything else that you would like to add?

P: Uhm so in here. There is no scale for in video you have a video for valence, but here.

M: There is arousal over here. I mean there is calculated. Arousal does feed into passion and energy.

P: Okay because from my experience arousal can be predicted more in audio.

M: yes, it is illustrated here. Which it does give you more options than that option.

P: yeah, this does give you more options, this one is only 1,2,3,4,5,6 so only small. So maybe put more emotions.

M: I can't change that.

P: Oh yeah, I can't change that.

M: That is fixed because we got that software, I mean the whole idea is that we take this technology and this person in a company would just get it commercially and then train their employees. This is just an example. So that is the best way. Okay. So, you still pick the diamond yeah?

P: Yeah, I still pick the diamond.

P12 – Vocal Affect Recognition

M: this is the recording for p12 for voice analysis. The two options have been presented to the participant. So please tell me which of the options you liked best, option 1 being the diamond and option 2 being the bar chart.

P: I liked option 2.

M: okay can you tell me why?

P: for the same reasons, there is a direct comparison to what your performance was to what was a good performance.

M: okay so that's fine, so what were, of the less preferred options, which was the diamond, what did you not like about it?

P: just that there was no comparison. Going in you wouldn't know what would be a good interview if you didn't have any prior training. You would need to know what a good interview for you would be to sit here and go I need less fear, I need more of this and I need more of that.

M: well you would have your trainer with you because this is a tool that would aid the trainer.

P: That is the only reason that I prefer the bar chart. Other than that, I would prefer the diamond.

M: Okay so if you had a trainer here. Which one would you prefer?

P: the first option because then as the trainer is saying 'be less fearful' for example, you could instantly change that. Whereas with the bar chart that is fixed and already done. It is not real time.

M: Okay that is wonderful thank you. So, of the less preferred version being the diamond, what are elements that you don't like?

P: I just said, that there is not a direct comparison to the good performance.

M: yes, you did, so what about the bar chart what don't you like?

P: That is already an output, that you can't change anything so if you had a trainer present, the trainer wouldn't be able to say be less fearful or be more joyful, you'd have to go redo the bar chart to see what your actual performance was. Does that make sense?

M: for every interview

P: for every single interview, whereas this gives you something right now.

M: okay thank you.

Appendix 6

Chapter 6. Enhancing Communication Skills Training by Providing Social Signal Feedback

6.3.2. Social Signals

6.3.2.1. Data pre-processing

Statistics

		REGR factor score 1 for analysis 1	REGR factor score 2 for analysis 1	REGR factor score 3 for analysis 1	REGR factor score 4 for analysis 1	REGR factor score 5 for analysis 1	REGR factor score 6 for analysis 1	REGR factor score 7 for analysis 1
N	Valid	646	646	646	646	646	646	646
	Missing	410	410	410	410	410	410	410

6.3.2.2. Assumption Testing of Principal Component Analysis

Sampling adequacy and KMO and Bartlett's Test of Sphericity **KMO and Bartlett's Test**

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.632
Bartlett's Test of Sphericity	Approx. Chi-Square	25962.157
	df	1711
	Sig.	.000

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Condition_num	Statistic	df	Sig.	Statistic	df	Sig.
Browfurrow	PreTraining	.446	67	.000	.261	67	.000
	PostTraining	.394	100	.000	.283	100	.000
Browraise	PreTraining	.220	67	.000	.752	67	.000
	PostTraining	.148	100	.000	.887	100	.000
Engagement	PreTraining	.112	67	.035	.969	67	.088
	PostTraining	.086	100	.065	.949	100	.001

Lip_corner_depres	PreTraining	.419	67	.000	.476	67	.000
	PostTraining	.337	100	.000	.482	100	.000
Smile	PreTraining	.339	67	.000	.542	67	.000
	PostTraining	.355	100	.000	.408	100	.000
InnerBrowRaise	PreTraining	.271	67	.000	.757	67	.000
	PostTraining	.308	100	.000	.625	100	.000
NoseWrinkle	PreTraining	.354	67	.000	.612	67	.000
	PostTraining	.367	100	.000	.443	100	.000
UpperLipRaise	PreTraining	.390	67	.000	.361	67	.000
	PostTraining	.359	100	.000	.376	100	.000
LipSuck	PreTraining	.363	67	.000	.379	67	.000
	PostTraining	.329	100	.000	.482	100	.000
LipPress	PreTraining	.262	67	.000	.656	67	.000
	PostTraining	.275	100	.000	.604	100	.000
MouthOpen	PreTraining	.131	67	.006	.913	67	.000
	PostTraining	.179	100	.000	.820	100	.000
ChinRaise	PreTraining	.403	67	.000	.479	67	.000
	PostTraining	.362	100	.000	.437	100	.000
Smirk	PreTraining	.430	67	.000	.311	67	.000
	PostTraining	.431	100	.000	.157	100	.000
LipPucker	PreTraining	.352	67	.000	.440	67	.000
	PostTraining	.273	100	.000	.633	100	.000
Anger	PreTraining	.507	67	.000	.194	67	.000
	PostTraining	.379	100	.000	.284	100	.000
Sadness	PreTraining	.484	67	.000	.204	67	.000
	PostTraining	.422	100	.000	.222	100	.000
Disgust	PreTraining	.357	67	.000	.632	67	.000
	PostTraining	.407	100	.000	.329	100	.000

Joy	PreTraining	.391	67	.000	.485	67	.000
	PostTraining	.403	100	.000	.293	100	.000
Surprise	PreTraining	.210	67	.000	.762	67	.000
	PostTraining	.156	100	.000	.878	100	.000
Fear	PreTraining	.381	67	.000	.627	67	.000
	PostTraining	.331	100	.000	.547	100	.000
Contempt	PreTraining	.452	67	.000	.329	67	.000
	PostTraining	.467	100	.000	.114	100	.000
CheekRaise	PreTraining	.391	67	.000	.472	67	.000
	PostTraining	.289	100	.000	.623	100	.000
Dimpler	PreTraining	.281	67	.000	.613	67	.000
	PostTraining	.351	100	.000	.382	100	.000
EyeWiden	PreTraining	.295	67	.000	.723	67	.000
	PostTraining	.217	100	.000	.794	100	.000
LidTighten	PreTraining	.363	67	.000	.390	67	.000
	PostTraining	.390	100	.000	.440	100	.000
LipStretch	PreTraining	.372	67	.000	.370	67	.000
	PostTraining	.391	100	.000	.297	100	.000
JawDrop	PreTraining	.171	67	.000	.851	67	.000
	PostTraining	.153	100	.000	.850	100	.000
Energy	PreTraining	.132	67	.006	.926	67	.001
	PostTraining	.124	100	.001	.945	100	.000
Upset	PreTraining	.213	67	.000	.800	67	.000
	PostTraining	.222	100	.000	.767	100	.000
Angry	PreTraining	.534	67	.000	.101	67	.000
	PostTraining	.536	100	.000	.124	100	.000
Stressed	PreTraining	.096	67	.200*	.975	67	.190

	PostTraining	.046	100	.200*	.976	100	.060
Uncertain	PreTraining	.100	67	.093	.983	67	.477
	PostTraining	.046	100	.200*	.986	100	.385
Excited	PreTraining	.098	67	.178	.971	67	.120
	PostTraining	.116	100	.002	.955	100	.002
Concentrated	PreTraining	.058	67	.200*	.992	67	.950
	PostTraining	.048	100	.200*	.987	100	.460
EmoCogRatio	PreTraining	.041	67	.200*	.992	67	.946
	PostTraining	.088	100	.055	.942	100	.000
Hesitation	PreTraining	.103	67	.077	.977	67	.260
	PostTraining	.072	100	.200*	.960	100	.004
BrainPower	PreTraining	.077	67	.200*	.974	67	.181
	PostTraining	.055	100	.200*	.987	100	.432
Embar	PreTraining	.387	67	.000	.562	67	.000
	PostTraining	.288	100	.000	.659	100	.000
I_think	PreTraining	.132	67	.006	.943	67	.004
	PostTraining	.186	100	.000	.751	100	.000
Imagine	PreTraining	.067	67	.200*	.966	67	.063
	PostTraining	.140	100	.000	.820	100	.000
ExtremeEmo	PreTraining	.097	67	.190	.948	67	.007
	PostTraining	.131	100	.000	.909	100	.000
Arousal	PreTraining	.089	67	.200*	.951	67	.010
	PostTraining	.141	100	.000	.896	100	.000
Movement	PreTraining	.226	67	.000	.812	67	.000
	PostTraining	.139	100	.000	.872	100	.000
M_Activity	PreTraining	.153	67	.001	.885	67	.000
	PostTraining	.187	100	.000	.817	100	.000
M_Rate	PreTraining	.142	67	.002	.934	67	.001

	PostTraining	.077	100	.158	.990	100	.630
M_Consistency	PreTraining	.100	67	.094	.969	67	.094
	PostTraining	.178	100	.000	.722	100	.000
M_Mirror	PreTraining	.123	67	.014	.959	67	.027
	PostTraining	.076	100	.167	.956	100	.002
M_MirrorLag	PreTraining	.068	67	.200*	.972	67	.133
	PostTraining	.080	100	.113	.978	100	.093
Posture	PreTraining	.241	67	.000	.817	67	.000
	PostTraining	.155	100	.000	.885	100	.000
P_Activity	PreTraining	.162	67	.000	.879	67	.000
	PostTraining	.141	100	.000	.861	100	.000
P_Rate	PreTraining	.206	67	.000	.796	67	.000
	PostTraining	.143	100	.000	.896	100	.000
P_Mirroring	PreTraining	.110	67	.044	.964	67	.048
	PostTraining	.081	100	.100	.966	100	.012
P_MirrorLag	PreTraining	.061	67	.200*	.983	67	.482
	PostTraining	.094	100	.028	.983	100	.208
Successfulinterruptions	PreTraining	.223	67	.000	.867	67	.000
	PostTraining	.311	100	.000	.682	100	.000
Unsuccessfulinterruptions	PreTraining	.292	67	.000	.757	67	.000
	PostTraining	.433	100	.000	.573	100	.000
Speed_Turn	PreTraining	.172	67	.000	.904	67	.000
	PostTraining	.156	100	.000	.883	100	.000
Overlap	PreTraining	.086	67	.200*	.982	67	.431
	PostTraining	.104	100	.009	.937	100	.000
Total_speaking	PreTraining	.082	67	.200*	.943	67	.004
	PostTraining	.150	100	.000	.847	100	.000
Volume				.029	.969	67	.093

	PreTraining	.115	67				
	PostTraining	.195	100	.000	.799	100	.000
V_consistency	PreTraining	.520	67	.000	.227	67	.000
	PostTraining	.184	100	.000	.791	100	.000
Pitch	PreTraining	.145	67	.001	.922	67	.000
	PostTraining	.248	100	.000	.756	100	.000
Volume_mirror	PreTraining	.105	67	.063	.957	67	.022
	PostTraining	.129	100	.000	.900	100	.000
Vol_mirrorlag	PreTraining	.088	67	.200*	.955	67	.017
	PostTraining	.151	100	.000	.897	100	.000

*. This is a lower bound of the true significance. a. Lilliefors Significance Correction

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
Browfurrow	Based on Mean	16.461	1	165	.000
	Based on Median	4.697	1	165	.032
	Based on Median and with adjusted df	4.697	1	67.708	.034
	Based on trimmed mean	5.763	1	165	.017
Browraise	Based on Mean	.239	1	165	.626
	Based on Median	.845	1	165	.359
	Based on Median and with adjusted df	.845	1	147.319	.359
	Based on trimmed mean	.395	1	165	.531
Engagement	Based on Mean	.410	1	165	.523
	Based on Median	.323	1	165	.570

	Based on Median and with adjusted df	.323	1	164.872	.570
	Based on trimmed mean	.350	1	165	.555
Lip_corner_depres	Based on Mean	.126	1	165	.723
		.002	1	165	.963
	Based on Median Based on Median and with adjusted df	.002	1	164.978	.963
	Based on trimmed mean	.028	1	165	.868
Smile	Based on Mean	4.858	1	165	.029
		1.694	1	165	.195
	Based on Median Based on Median and with adjusted df	1.694	1	155.746	.195
	Based on trimmed mean	3.928	1	165	.049
InnerBrowRaise	Based on Mean	18.443	1	165	.000
	Based on Median	6.951	1	165	.009
	Based on Median and with adjusted df	6.951	1	145.582	.009
	Based on trimmed mean	16.151	1	165	.000
NoseWrinkle	Based on Mean	35.358	1	165	.000
	Based on Median	9.506	1	165	.002
	Based on Median and with adjusted df	9.506	1	108.793	.003
	Based on trimmed mean	26.251	1	165	.000
UpperLipRaise	Based on Mean	1.916	1	165	.168
	Based on Median	1.105	1	165	.295

	Based on Median and with adjusted df	1.105	1	144.880	.295
	Based on trimmed mean	1.303	1	165	.255
LipSuck	Based on Mean	14.525	1	165	.000
	Based on Median	4.968	1	165	.027
	Based on Median and with adjusted df	4.968	1	79.695	.029
	Based on trimmed mean	7.616	1	165	.006
LipPress	Based on Mean	4.330	1	165	.039
	Based on Median	1.799	1	165	.182
	Based on Median and with adjusted df	1.799	1	152.867	.182
	Based on trimmed mean	3.240	1	165	.074
MouthOpen		1.967	1	165	.163
	Based on Mean	2.628	1	165	.107
	Based on Median Based on Median and with adjusted df	2.628	1	156.832	.107
	Based on trimmed mean	2.418	1	165	.122
ChinRaise	Based on Mean	2.464	1	165	.118
	Based on Median	1.097	1	165	.297
	Based on Median and with adjusted df	1.097	1	140.923	.297
	Based on trimmed mean	.853	1	165	.357
Smirk	Based on Mean	1.405	1	165	.238
	Based on Median	.467	1	165	.496
	Based on Median and with adjusted df	.467	1	164.901	.496

	Based on trimmed mean	.701	1	165	.404
LipPucker	Based on Mean	23.266	1	165	.000
	Based on Median	12.120	1	165	.001
	Based on Median and with adjusted df	12.120	1	118.981	.001
	Based on trimmed mean	18.435	1	165	.000

Anger	Based on Mean	14.222	1	165	.000
	Based on Median	3.577	1	165	.060
	Based on Median and with adjusted df	3.577	1	66.016	.063
	Based on trimmed mean	3.604	1	165	.059
Sadness	Based on Mean	9.748	1	165	.002
	Based on Median	2.435	1	165	.121
	Based on Median and with adjusted df	2.435	1	74.139	.123
	Based on trimmed mean	2.611	1	165	.108
Disgust	Based on Mean	.065	1	165	.799
	Based on Median	.026	1	165	.873
	Based on Median and with adjusted df	.026	1	137.608	.873
	Based on trimmed mean	.254	1	165	.615
Joy	Based on Mean	12.740	1	165	.000
	Based on Median	4.214	1	165	.042
		4.214	1	129.717	.042

	Based on Median and with adjusted df				
	Based on trimmed mean	9.580	1	165	.002
Surprise	Based on Mean	5.308	1	165	.022
	Based on Median	1.470	1	165	.227
	Based on Median and with adjusted df	1.470	1	121.515	.228
	Based on trimmed mean	3.248	1	165	.073
Fear	Based on Mean	.405	1	165	.526
	Based on Median	.116	1	165	.734
	Based on Median and with adjusted df	.116	1	155.238	.734
	Based on trimmed mean	.034	1	165	.853
Contempt	Based on Mean	2.162	1	165	.143
	Based on Median	.748	1	165	.388
	Based on Median and with adjusted df	.748	1	164.219	.388
	Based on trimmed mean	.852	1	165	.357
CheekRaise	Based on Mean	6.171	1	165	.014
	Based on Median	.801	1	165	.372
	Based on Median and with adjusted df	.801	1	120.165	.373
	Based on trimmed mean	2.232	1	165	.137
Dimpler	Based on Mean	.192	1	165	.662
	Based on Median	.374	1	165	.542
	Based on Median and with adjusted df	.374	1	157.067	.542

	Based on trimmed mean	.386	1	165	.535
EyeWiden	Based on Mean	2.094	1	165	.150
	Based on Median	1.595	1	165	.208
	Based on Median and with adjusted df	1.595	1	163.918	.208
	Based on trimmed mean	1.986	1	165	.161
LidTighten	Based on Mean	7.607	1	165	.006
	Based on Median	2.810	1	165	.096
	Based on Median and with adjusted df	2.810	1	100.465	.097
	Based on trimmed mean	3.814	1	165	.053
LipStretch	Based on Mean	8.401	1	165	.004
	Based on Median	2.901	1	165	.090
	Based on Median and with adjusted df	2.901	1	111.540	.091
	Based on trimmed mean	5.211	1	165	.024
JawDrop	Based on Mean	3.243	1	165	.074
	Based on Median	2.357	1	165	.127
	Based on Median and with adjusted df	2.357	1	161.079	.127
	Based on trimmed mean	2.928	1	165	.089
Energy	Based on Mean	.719	1	165	.398
	Based on Median	.501	1	165	.480
	Based on Median and with adjusted df	.501	1	164.983	.480

	Based on trimmed mean	.676	1	165	.412
Upset	Based on Mean	11.711	1	165	.001
	Based on Median	8.528	1	165	.004
	Based on Median and with adjusted df	8.528	1	132.139	.004
	Based on trimmed mean	10.010	1	165	.002
Angry	Based on Mean	2.964	1	165	.087
	Based on Median	.715	1	165	.399
	Based on Median and with adjusted df	.715	1	77.747	.400
	Based on trimmed mean	.715	1	165	.399
Stressed	Based on Mean	.580	1	165	.447
	Based on Median	.384	1	165	.536
	Based on Median and with adjusted df	.384	1	163.217	.536
	Based on trimmed mean	.511	1	165	.476
Uncertain	Based on Mean	.858	1	165	.356
	Based on Median	.827	1	165	.365
	Based on Median and with adjusted df	.827	1	164.924	.365
	Based on trimmed mean	.840	1	165	.361
Excited	Based on Mean	1.633	1	165	.203
	Based on Median	1.039	1	165	.310
	Based on Median and with adjusted df	1.039	1	163.028	.310
	Based on trimmed mean	1.471	1	165	.227

Concentrated	Based on Mean	.503	1	165	.479
	Based on Median	.526	1	165	.469
	Based on Median and with adjusted df	.526	1	163.447	.469
	Based on trimmed mean	.518	1	165	.473
EmoCogRatio	Based on Mean	.029	1	165	.864
	Based on Median	.011	1	165	.915
	Based on Median and with adjusted df	.011	1	154.126	.915
	Based on trimmed mean	.023	1	165	.881
Hesitation	Based on Mean	.546	1	165	.461
	Based on Median	.548	1	165	.460
	Based on Median and with adjusted df	.548	1	156.573	.460
	Based on trimmed mean	.527	1	165	.469
BrainPower	Based on Mean	1.007	1	165	.317
	Based on Median	1.046	1	165	.308
	Based on Median and with adjusted df	1.046	1	161.126	.308
	Based on trimmed mean	1.028	1	165	.312
Embar		.027	1	165	.871
	Based on Mean	.519	1	165	.472
	Based on Median Based on Median and with adjusted df	.519	1	164.947	.472
	Based on trimmed mean	.113	1	165	.737
I_think		2.097	1	165	.149

	Based on Mean				
	Based on Median	.764	1	165	.383
	Based on Median and with adjusted df	.764	1	124.049	.384
	Based on trimmed mean	1.241	1	165	.267
Imagine	Based on Mean	1.895	1	165	.171
	Based on Median	1.636	1	165	.203
	Based on Median and with adjusted df	1.636	1	141.531	.203
	Based on trimmed mean	1.645	1	165	.201
ExtremeEmo	Based on Mean	.585	1	165	.445
	Based on Median	.415	1	165	.520
	Based on Median and with adjusted df	.415	1	162.363	.520
	Based on trimmed mean	.515	1	165	.474
Arousal	Based on Mean	3.141	1	165	.078
	Based on Median	3.025	1	165	.084
	Based on Median and with adjusted df	3.025	1	162.043	.084
	Based on trimmed mean	3.131	1	165	.079
Movement	Based on Mean	13.090	1	165	.000
	Based on Median	3.751	1	165	.054
	Based on Median and with adjusted df	3.751	1	145.885	.055
	Based on trimmed mean	12.159	1	165	.001
M_Activity	Based on Mean	7.787	1	165	.006

	Based on Median	4.455	1	165	.036
	Based on Median and with adjusted df	4.455	1	137.680	.037
	Based on trimmed mean	5.864	1	165	.017
M_Rate	Based on Mean	.839	1	165	.361
	Based on Median	.791	1	165	.375
	Based on Median and with adjusted df	.791	1	148.042	.375
	Based on trimmed mean	.834	1	165	.363
M_Consistency	Based on Mean	3.712	1	165	.056
	Based on Median	2.342	1	165	.128
	Based on Median and with adjusted df	2.342	1	116.447	.129
	Based on trimmed mean	2.680	1	165	.104
M_Mirror	Based on Mean	.290	1	165	.591
	Based on Median	.309	1	165	.579
	Based on Median and with adjusted df	.309	1	163.632	.579
	Based on trimmed mean	.310	1	165	.579
M_MirrorLag	Based on Mean	1.782	1	165	.184
	Based on Median	1.517	1	165	.220
	Based on Median and with adjusted df	1.517	1	162.293	.220
	Based on trimmed mean	1.704	1	165	.194
Posture	Based on Mean	7.399	1	165	.007
	Based on Median	1.642	1	165	.202

	Based on Median and with adjusted df	1.642	1	152.279	.202
	Based on trimmed mean	7.056	1	165	.009
P_Activity	Based on Mean	1.844	1	165	.176
	Based on Median	.970	1	165	.326
	Based on Median and with adjusted df	.970	1	161.315	.326
	Based on trimmed mean	1.394	1	165	.239
P_Rate	Based on Mean	3.341	1	165	.069
		3.638	1	165	.058
	Based on Median Based on Median and with adjusted df	3.638	1	121.026	.059
	Based on trimmed mean	3.497	1	165	.063
P_Mirroring	Based on Mean	.029	1	165	.865
		.062	1	165	.804
	Based on Median Based on Median and with adjusted df	.062	1	164.436	.804
	Based on trimmed mean	.055	1	165	.815
P_MirrorLag	Based on Mean	.835	1	165	.362
	Based on Median	.805	1	165	.371
	Based on Median and with adjusted df	.805	1	164.983	.371
	Based on trimmed mean	.806	1	165	.371
Successfulinterruptions	Based on Mean	9.091	1	165	.003
	Based on Median	9.537	1	165	.002
	Based on Median and with adjusted df	9.537	1	148.710	.002

	Based on trimmed mean	10.611	1	165	.001
Unsuccessfulinterruptions	Based on Mean	29.183	1	165	.000
	Based on Median	3.875	1	165	.051
	Based on Median and with adjusted df	3.875	1	108.265	.052
	Based on trimmed mean	17.918	1	165	.000
Speed_Turn	Based on Mean	21.056	1	165	.000
	Based on Median	20.854	1	165	.000
	Based on Median and with adjusted df	20.854	1	139.806	.000
	Based on trimmed mean	20.875	1	165	.000
Overlap	Based on Mean	16.743	1	165	.000
	Based on Median	14.167	1	165	.000
	Based on Median and with adjusted df	14.167	1	158.447	.000
	Based on trimmed mean	16.924	1	165	.000
Total_speaking	Based on Mean	.433	1	165	.511
		.170	1	165	.680
	Based on Median	.170	1	141.052	.680
	Based on Median and with adjusted df Based on trimmed mean	.245	1	165	.621
Volume	Based on Mean	3.660	1	165	.057
	Based on Median	1.863	1	165	.174
	Based on Median and with adjusted df	1.863	1	123.515	.175

	Based on trimmed mean	2.324	1	165	.129
V_consistency	Based on Mean	18.049	1	165	.000
	Based on Median	4.275	1	165	.040
	Based on Median and with adjusted df	4.275	1	66.145	.043
	Based on trimmed mean	4.269	1	165	.040
Pitch	Based on Mean	29.949	1	165	.000
	Based on Median	14.674	1	165	.000
	Based on Median and with adjusted df	14.674	1	113.296	.000
	Based on trimmed mean	24.787	1	165	.000
Volume_mirror	Based on Mean	11.049	1	165	.001
	Based on Median	10.174	1	165	.002
	Based on Median and with adjusted df	10.174	1	142.906	.002
	Based on trimmed mean	10.336	1	165	.002
Vol_mirrorlag		6.096	1	165	.015
	Based on Mean	5.153	1	165	.024
	Based on Median				
	Based on Median and with adjusted df	5.153	1	164.056	.025
	Based on trimmed mean	5.961	1	165	.016

6.3.2.2. Assumption Testing for Multivariate Analysis

Multivariate Normality

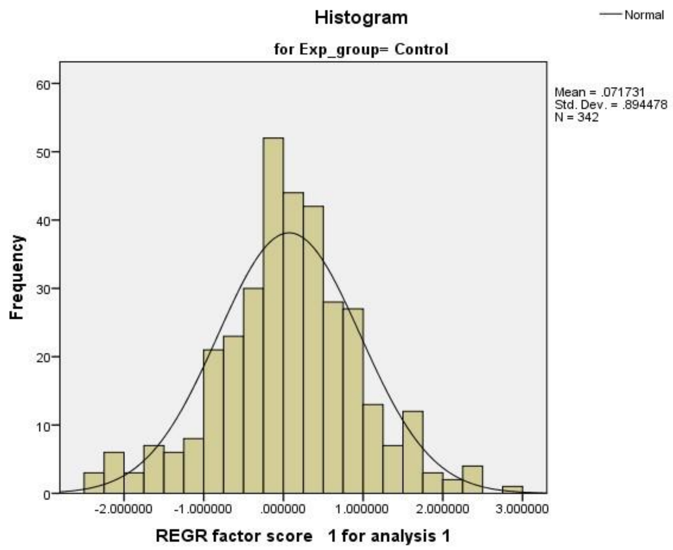
Component 1 – Confidence

Tests of Normality

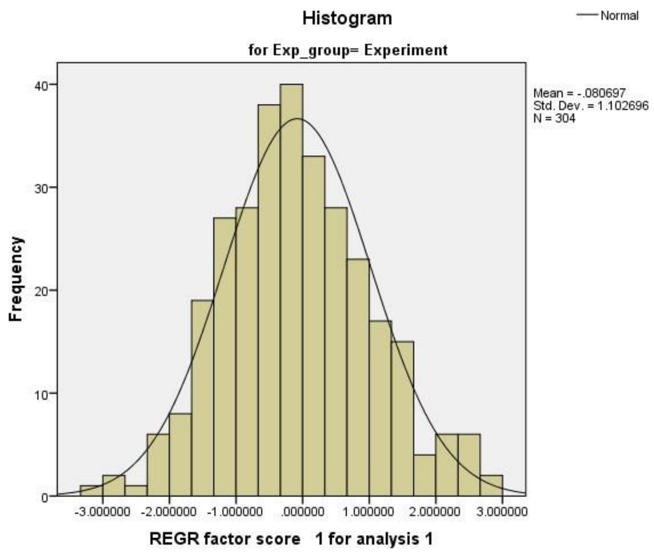
	Exp_group	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
REGR factor score 1 for analysis 1	Control	.053	342	.022	.989	342	.009
	Experiment	.042	304	.200*	.994	304	.227

*. This is a lower bound of the true significance. a. Lilliefors Significance Correction

Control Group



Experiment Group



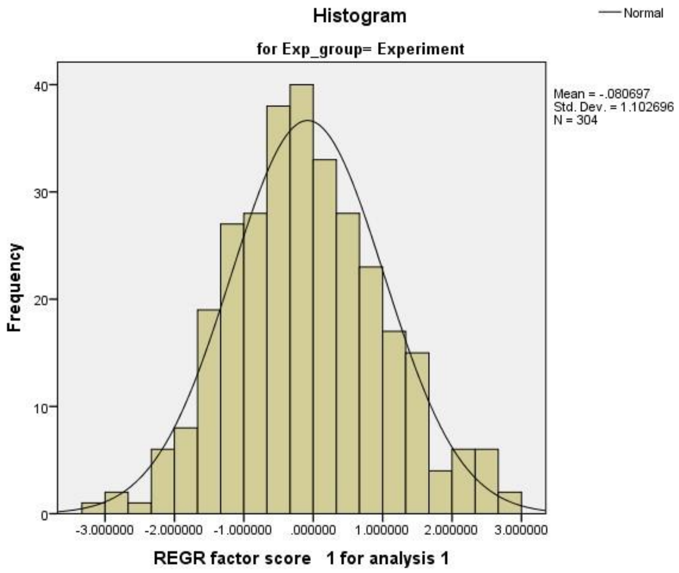
Interview Type

Tests of Normality

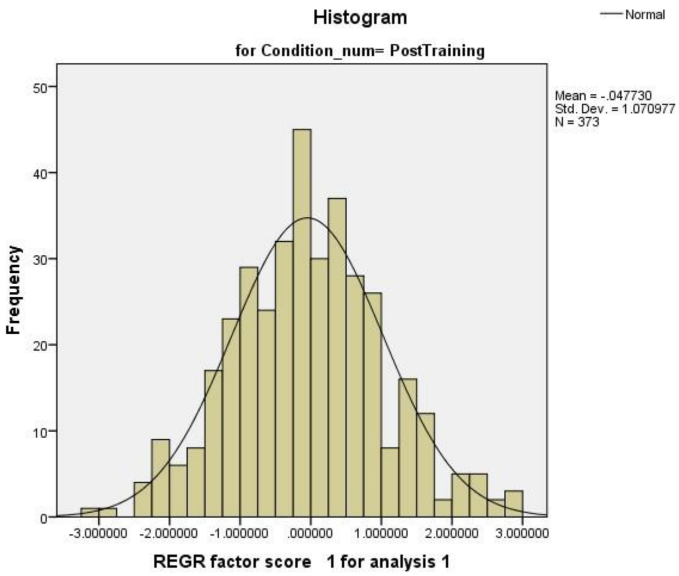
		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Condition_num	Statistic	df	Sig.	Statistic	df	Sig.
REGR factor score 1 for analysis 1	PreTraining	.043	273	.200*	.993	273	.275
	PostTraining	.028	373	.200*	.996	373	.527

*. This is a lower bound of the true significance. a. Lilliefors Significance Correction

Pretraining Interview



Post-training Interview



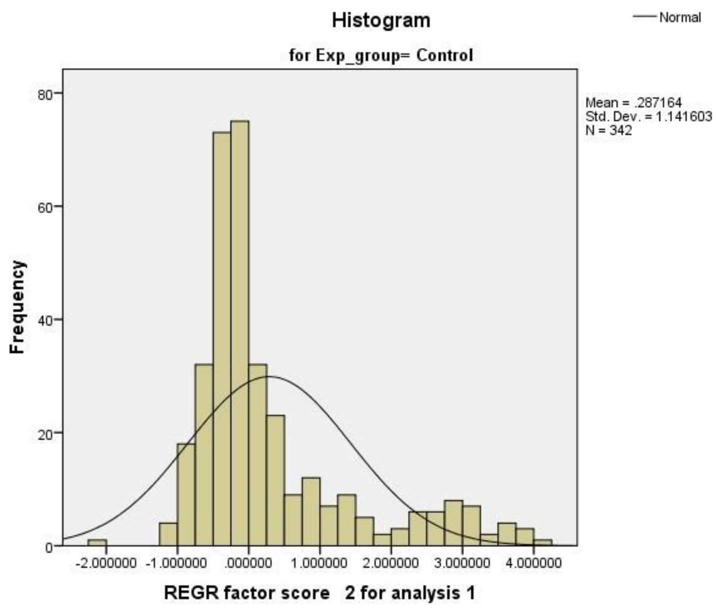
Component 2 - Disgust

Tests of Normality

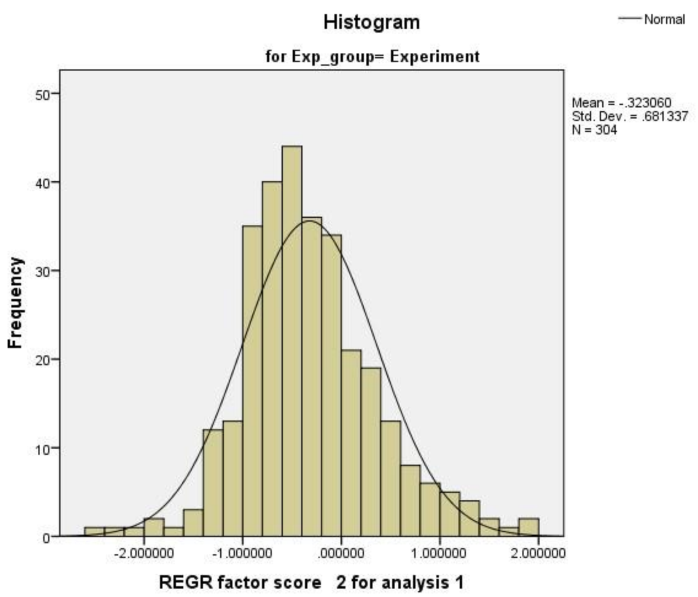
		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Exp_group	Statistic	df	Sig.	Statistic	df	Sig.
REGR factor score 2 for analysis 1	Control	.215	342	.000	.801	342	.000
	Experiment	.071	304	.001	.978	304	.000

a. Lilliefors Significance Correction

Control Group



Experiment Group



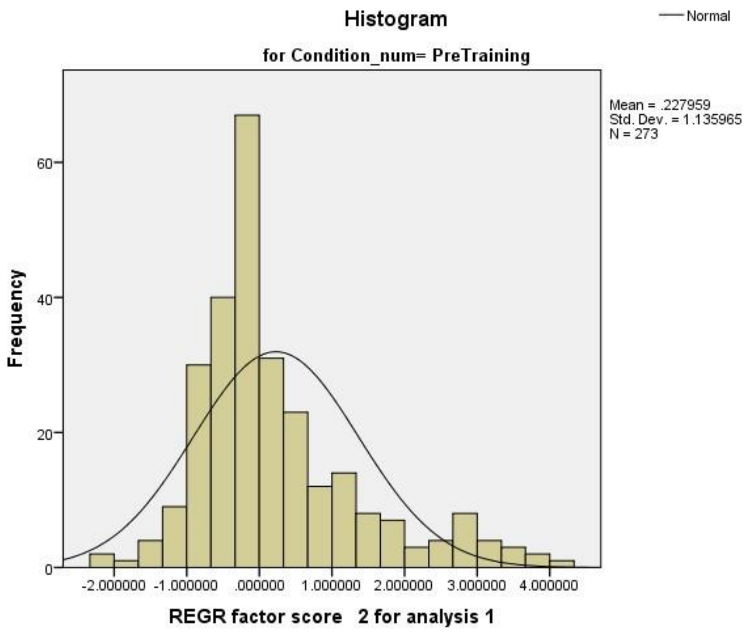
Interview Type

Tests of Normality

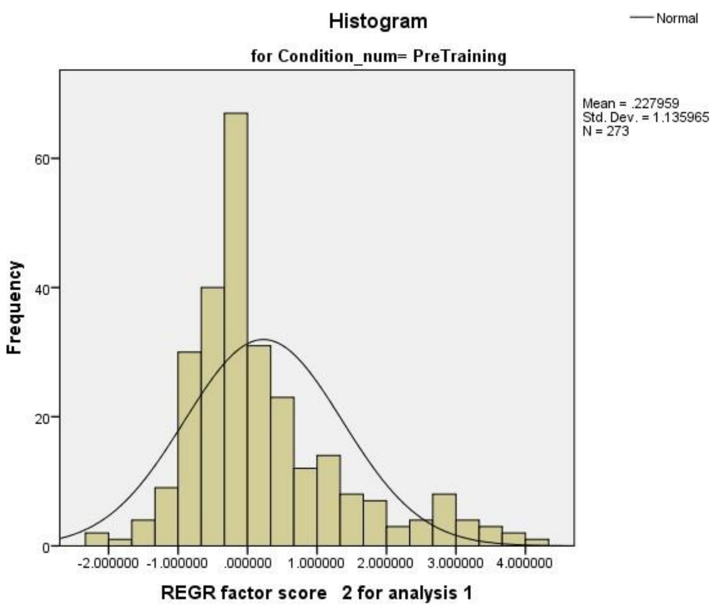
		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Condition_num	Statistic	df	Sig.	Statistic	df	Sig.
REGR factor score 2 for analysis 1	PreTraining	.156	273	.000	.902	273	.000
	PostTraining	.164	373	.000	.793	373	.000

a. Lilliefors Significance Correction

Pretraining Interview



Post-training Interview

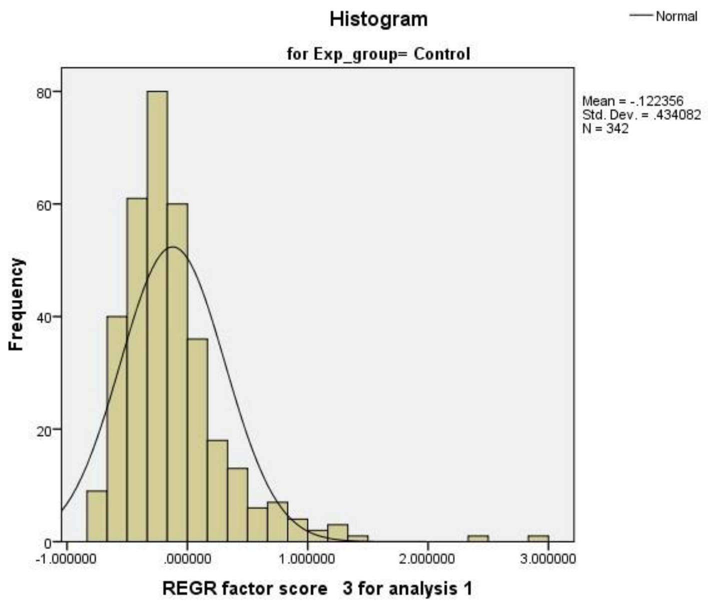


Component 3 - Frowning

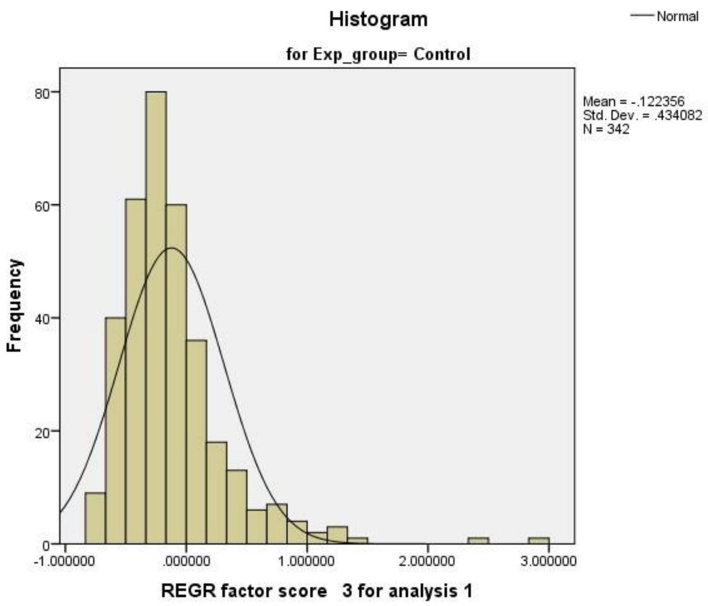
	Exp_group	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
REGR factor score 3 for analysis 1	Control	.129	342	.000	.833	342	.000
	Experiment	.334	304	.000	.368	304	.000

a. Lilliefors Significance Correction

Control Group



Experiment Group



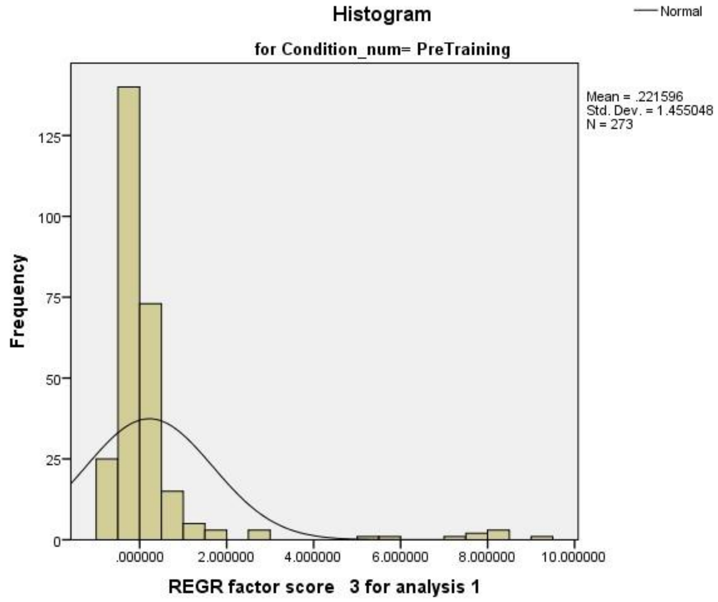
Interview Type

Tests of Normality

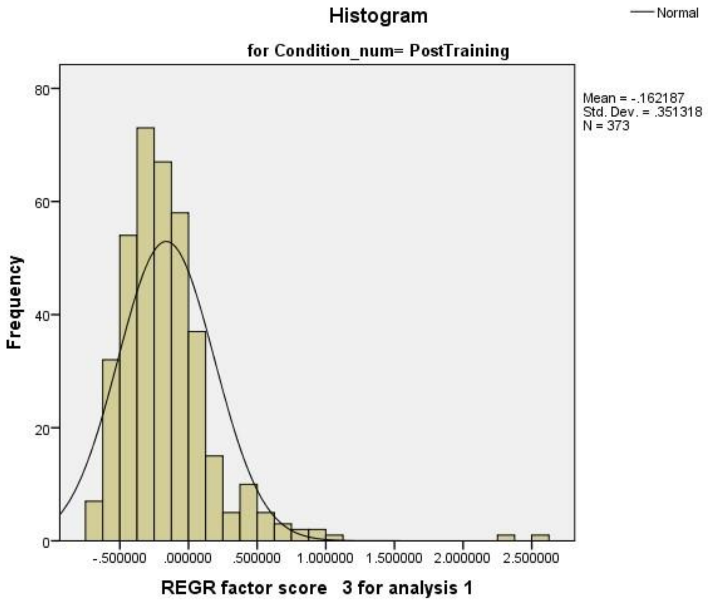
		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Condition_num	Statistic	df	Sig.	Statistic	df	Sig.
REGR factor score 3 for analysis 1	PreTraining	.306	273	.000	.434	273	.000
	PostTraining	.114	373	.000	.811	373	.000

a. Lilliefors Significance Correction

Pre-training Interview



Post-training Interview

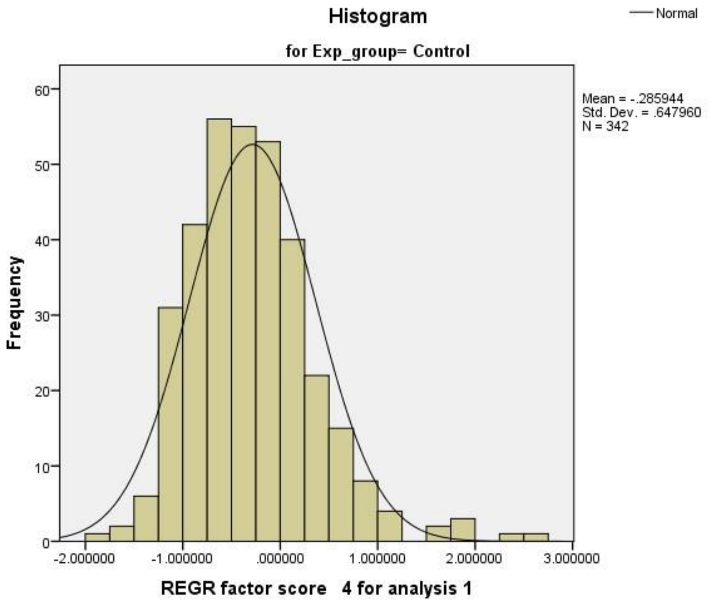


Component 4 – Eagerness to speak

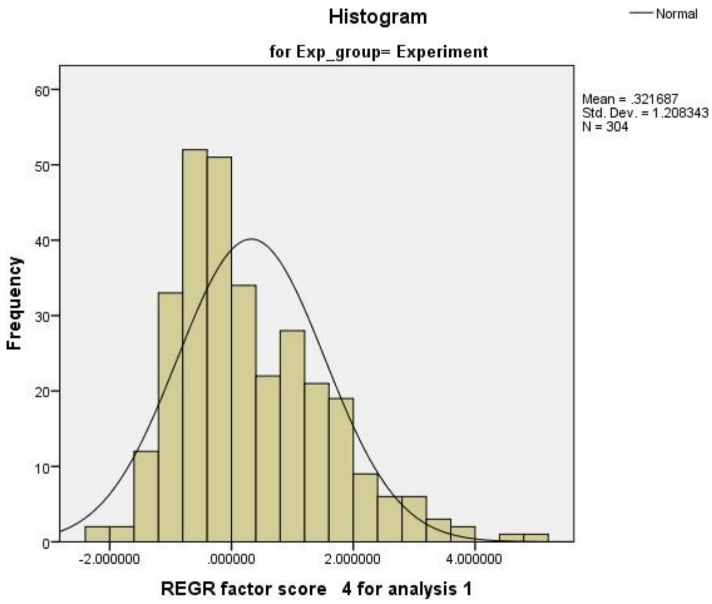
		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Exp_group	Statistic	df	Sig.	Statistic	df	Sig.
REGR factor score 4 for analysis 1	Control	.051	342	.033	.955	342	.000
	Experiment	.113	304	.000	.947	304	.000

a. Lilliefors Significance Correction

Control Group



Experiment Group



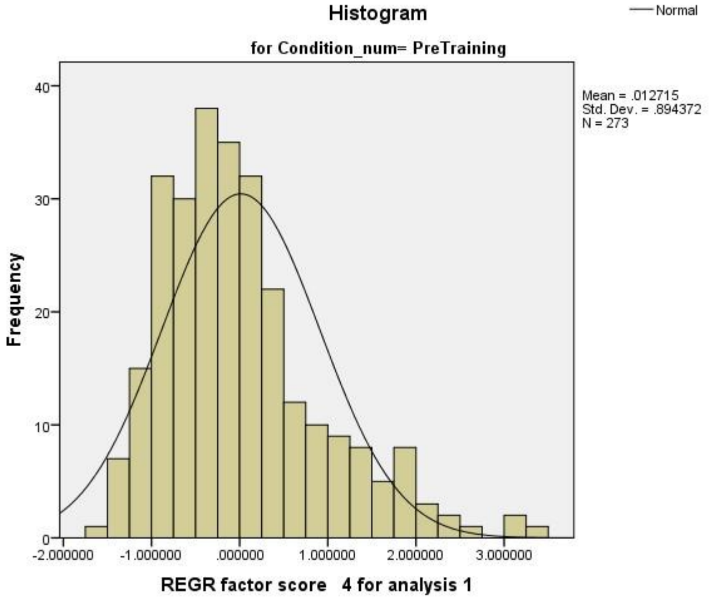
Interview Type

Tests of Normality

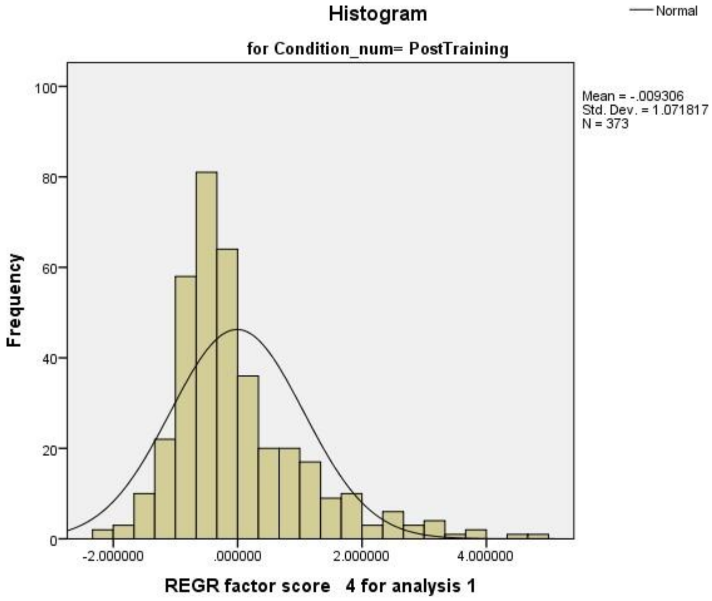
	Condition_num	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
REGR factor score 4 for analysis 1	PreTraining	.098	273	.000	.935	273	.000
	PostTraining	.141	373	.000	.891	373	.000

a. Lilliefors Significance Correction

Pretraining Interview



Post-training Interview

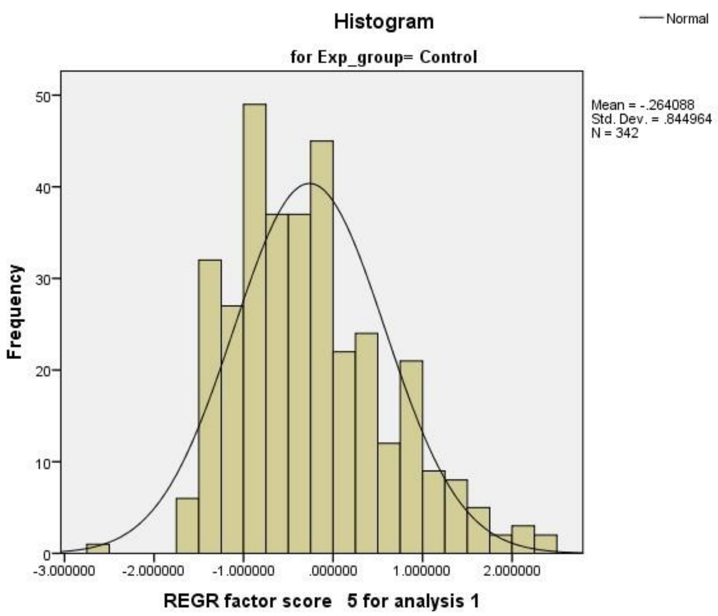


Component 5 - Expression engagement

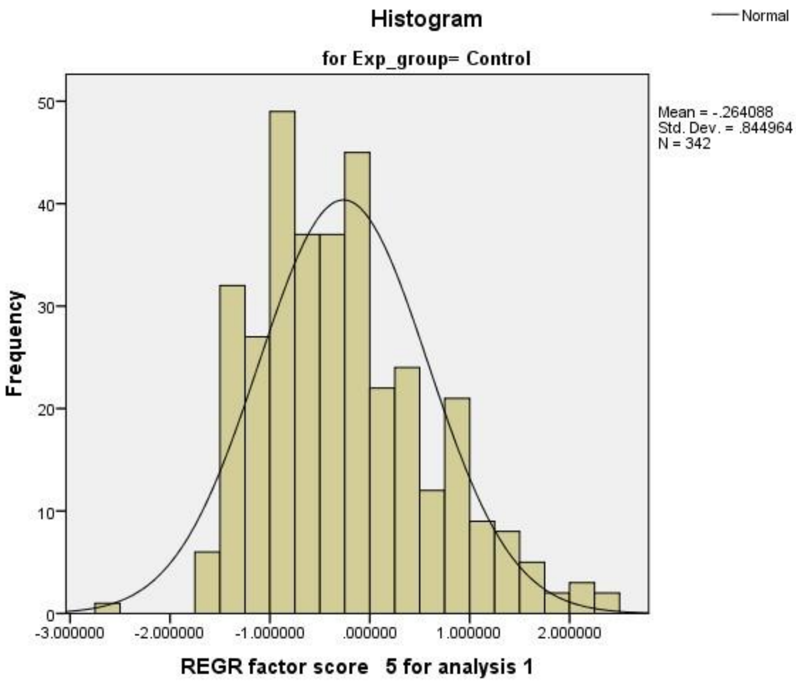
	Exp_group	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
REGR factor score 5 for analysis 1	Control	.067	342	.001	.967	342	.000
	Experiment	.089	304	.000	.935	304	.000

a. Lilliefors Significance Correction

Control Group



Experiment Group



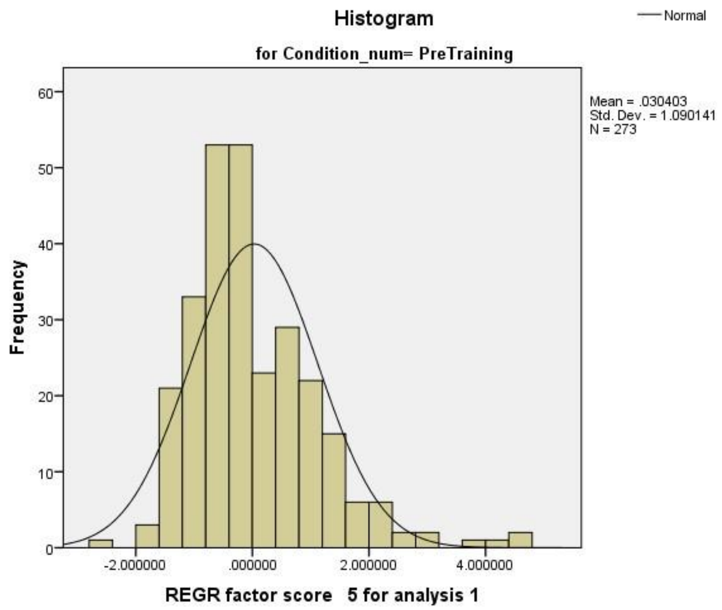
Interview Type

Tests of Normality

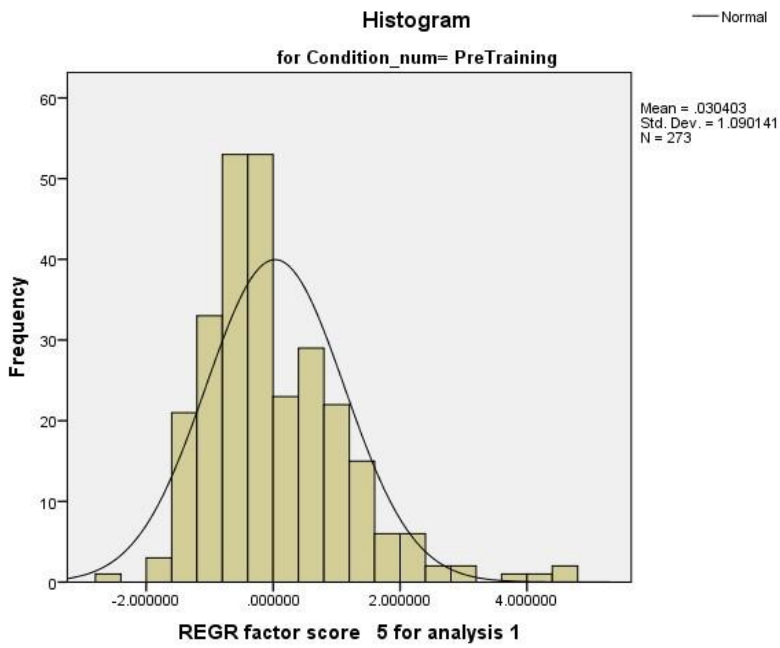
	Condition_num	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
REGR factor score 5 for analysis 1	PreTraining	.114	273	.000	.924	273	.000
	PostTraining	.058	373	.004	.956	373	.000

a. Lilliefors Significance Correction

Pretraining Interview



Post training Interview



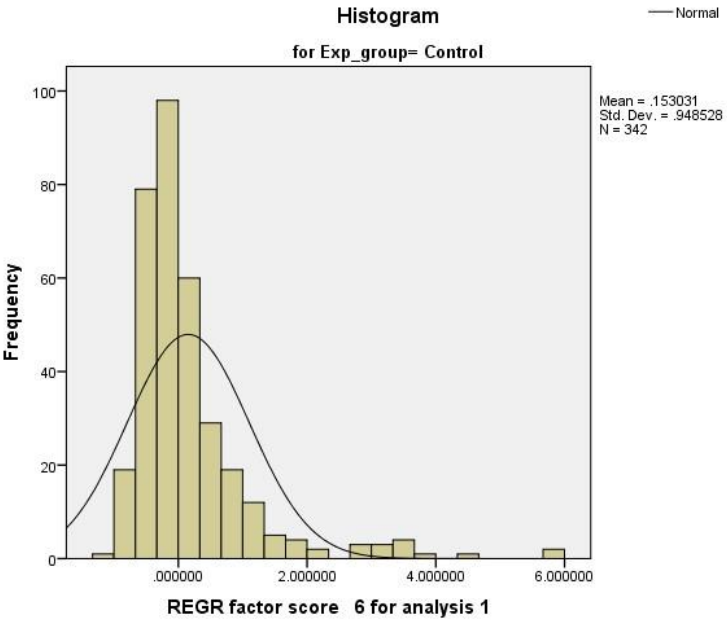
Component 6 – Posed Expression

Tests of Normality

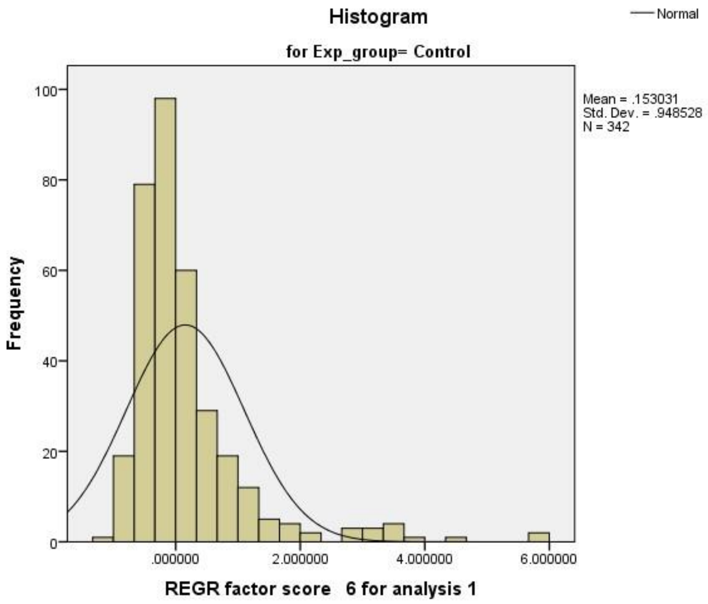
	Exp_group	Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
REGR factor score 6 for analysis 1	Control	.182	342	.000	.732	342	.000
	Experiment	.189	304	.000	.709	304	.000

a. Lilliefors Significance Correction

Control Group



Experiment Group



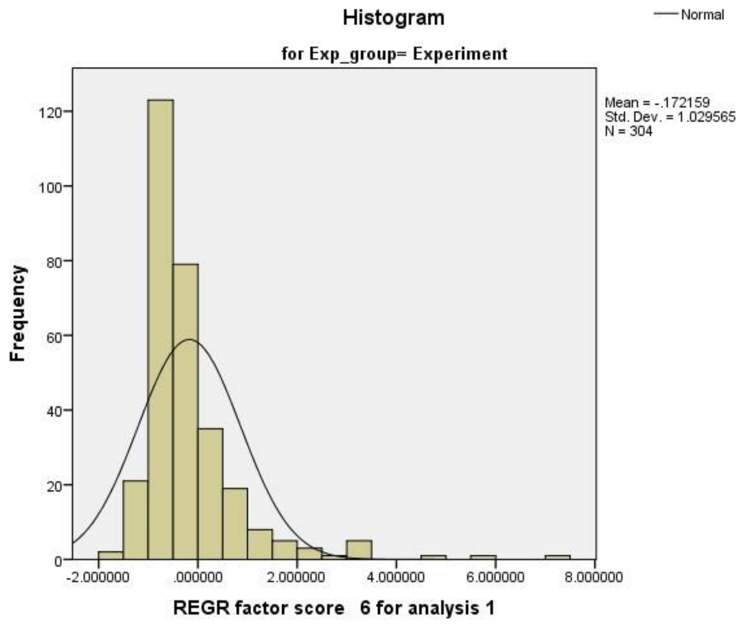
Interview Type

Tests of Normality

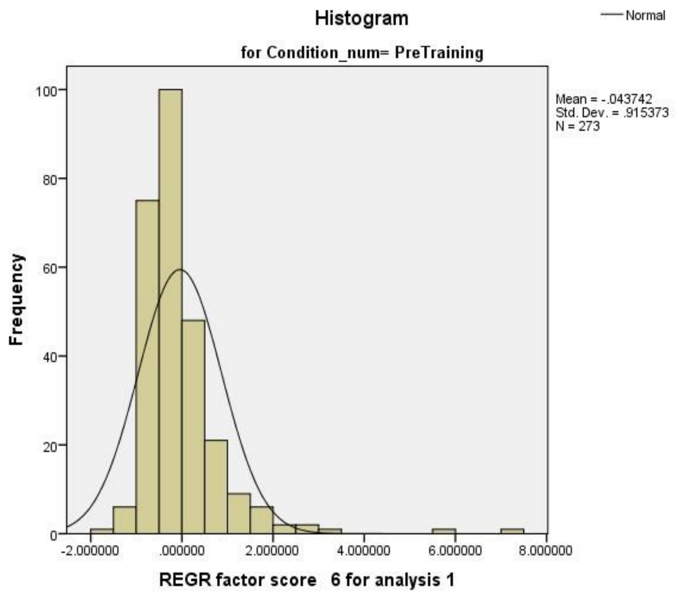
	Condition_num	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
REGR factor score 6 for analysis 1	PreTraining	.167	273	.000	.714	273	.000
	PostTraining	.170	373	.000	.764	373	.000

a. Lilliefors Significance Correction

Pretraining Interview



Post-training Interview



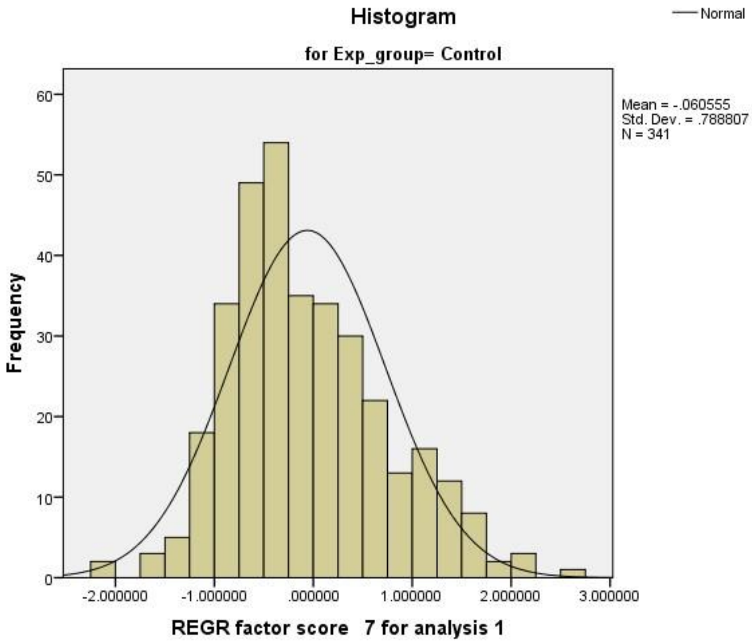
Component 7 - Posture

Tests of Normality

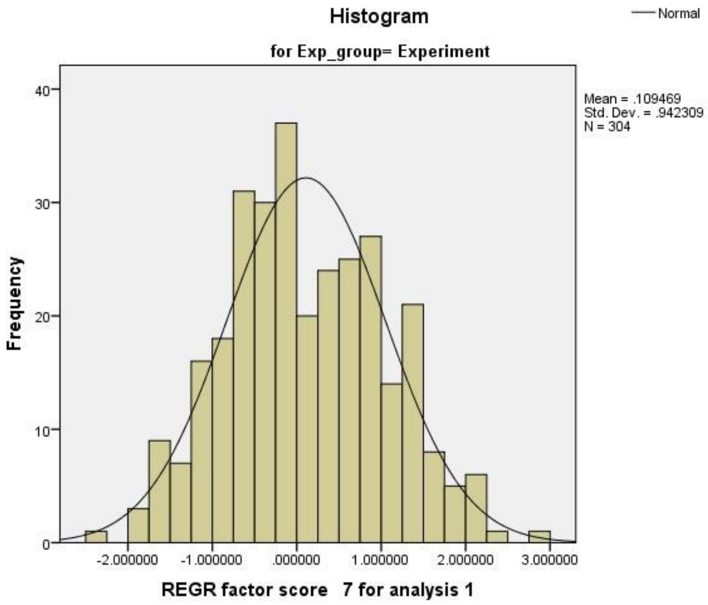
		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Exp_group	Statistic	df	Sig.	Statistic	df	Sig.
REGR factor score 7 for analysis 1	Control	.088	341	.000	.972	341	.000
	Experiment	.058	304	.016	.992	304	.125

a. Lilliefors Significance Correction

Control Group



Experiment Group



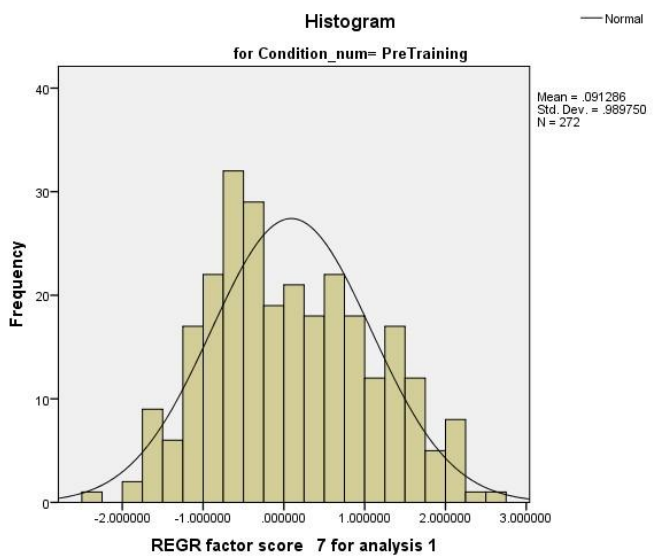
Interview Type

Tests of Normality

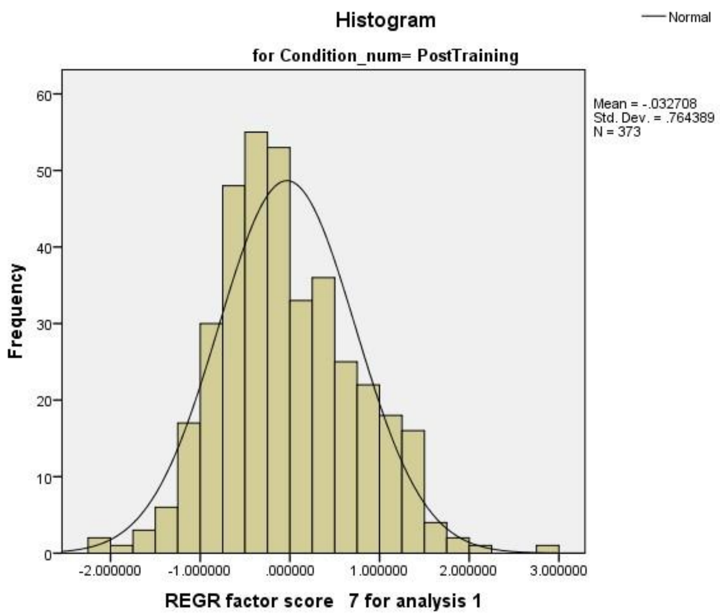
	Condition_num	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
REGR factor score 7 for analysis 1	PreTraining	.074	272	.001	.981	272	.001
	PostTraining	.082	373	.000	.985	373	.001

a. Lilliefors Significance Correction

Pretraining Interview



Post-training Interview



Test of homogeneity of variance

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
REGR factor score 1 for analysis 1	Based on Mean	16.629	1	644	.000
	Based on Median	15.787	1	644	.000
	Based on Median and with adjusted df	15.787	1	630.194	.000
	Based on trimmed mean	16.374	1	644	.000

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
REGR factor score 2 for analysis 1	Based on Mean	40.869	1	644	.000
	Based on Median	13.725	1	644	.000
	Based on Median and with adjusted df	13.725	1	466.188	.000
	Based on trimmed mean	29.159	1	644	.000

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
REGR factor score 3 for analysis 1	Based on Mean	41.866	1	644	.000
	Based on Median	22.082	1	644	.000
	Based on Median and with adjusted df	22.082	1	300.349	.000
	Based on trimmed mean	23.791	1	644	.000

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
REGR factor score 4 for analysis 1	Based on Mean	109.421	1	644	.000
	Based on Median	81.660	1	644	.000
	Based on Median and with adjusted df	81.660	1	479.147	.000
	Based on trimmed mean	102.380	1	644	.000

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
REGR factor score 5 for analysis 1	Based on Mean	7.984	1	644	.005
	Based on Median	6.738	1	644	.010
	Based on Median and with adjusted df	6.738	1	583.116	.010
	Based on trimmed mean	7.208	1	644	.007

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
REGR factor score 6 for analysis 1	Based on Mean	.524	1	644	.469
	Based on Median	.219	1	644	.640
	Based on Median and with adjusted df	.219	1	636.036	.640
	Based on trimmed mean	.393	1	644	.531

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
REGR factor score 7 for analysis 1	Based on Mean	29.949	1	643	.000
	Based on Median	28.352	1	643	.000
	Based on Median and with adjusted df	28.352	1	632.154	.000
	Based on trimmed mean	29.768	1	643	.000

Box's Test of Equality of Covariance Matrices[a]

Box's M	1513.294
F	17.657
df1	84
df2	771134.755
Sig.	.000

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + Condition_num + Exp_group + Condition_num * Exp_group

6.4. Results

6.4.1. Subjective Ratings of Performance

6.4.1.1. Internal Consistency of Communication Skill Ratings

Trainer molar ratings for pre-test interview - Baseline Interview

Case Processing Summary

		N	%
Cases	Valid	22	100.0
	Excluded[a]	0	.0
	Total	22	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.838	.831	5

Item Statistics

	Mean	Std. Deviation	N
BL_TR_conversation	5.0909	1.19160	22
BL_TR__sociallyskilled	5.0455	.84387	22
BL_TR_competent	4.9091	1.10880	22
BL_TR_appropriate	5.0455	1.04550	22
BL_TR_effective	4.9091	1.19160	22

Trainer molar ratings for post-test interview

Case Processing Summary

		N	%
Cases	Valid	22	100.0
	Excluded[a]	0	.0
	Total	22	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.677	.654	5

Item Statistics

	Mean	Std. Deviation	N
PT_TR_conversation	5.9091	1.01929	22
PT_TR_sociallyskilled	6.0455	.72225	22
PT_TR_competent	6.1818	1.05272	22
PT_TR_appropriate	6.0455	1.13294	22
PT_TR_effective	5.8636	1.12527	22

Neutral observer molar ratings for pre-test interview

Case Processing Summary

		N	%
Cases	Valid	22	100.0
	Excluded[a]	0	.0
	Total	22	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.965	.967	5

Item Statistics

	Mean	Std. Deviation	N
BL_conversation	4.1068	.87566	22
BL_sociallyskilled	4.0150	.91726	22
BL_competent	4.3182	1.10064	22
BL_appropriate	4.8477	.85316	22
BL_effective	4.2423	1.06909	22

Neutral observer molar ratings for post-test interview

Case Processing Summary

		N	%
Cases	Valid	22	100.0
	Excluded[a]	0	.0
	Total	22	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.965	.972	5

Item Statistics

	Mean	Std. Deviation	N
PT_conversation	4.7427	1.15400	22
PT_sociallyskilled	4.8036	1.02200	22
PT_competent	5.1064	1.05600	22
PT_appropriate	5.5455	.67890	22
PT_effective	5.0314	.87803	22

Self-report molar ratings for pre-test interview

Case Processing Summary

		N	%
Cases	Valid	22	100.0
	Excluded[a]	0	.0
	Total	22	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.901	.903	5

Item Statistics

	Mean	Std. Deviation	N
BL_SR_conversation	4.7273	.93513	22
BL_SR_sociallyskilled	4.8636	1.12527	22
BL_SR_competent	5.0909	.92113	22
BL_SR_appropriate	5.1818	.95799	22
BL_SR_effective	4.9091	.92113	22

Self-report molar ratings for post-test interview

Case Processing Summary

		N	%
Cases	Valid	22	100.0
	Excluded[a]	0	.0
	Total	22	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.944	.946	5

Item Statistics

	Mean	Std. Deviation	N
PT_SR_conversation	5.0455	.95005	22
PT_SR_sociallyskilled	5.0000	1.06904	22
PT_SR_competent	5.0455	1.13294	22
PT_SR_appropriate	4.8182	1.18065	22
PT_SR_effective	4.9091	1.23091	22

Neutral observer agreement scores for pretraining interview

Case Processing Summary

		N	%
Cases	Valid	22	100.0
	Excluded[a]	0	.0
	Total	22	100.0

a. Listwise deletion based on all variables in the procedure.

Intraclass Correlation Coefficient

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.433[a]	.178	.677	3.790	21	42	.000
Average Measures	.696[c]	.394	.863	3.790	21	42	.000

Two-way mixed effects model where people effects are random and measures effects are fixed. a.

The estimator is the same, whether the interaction effect is present or not.

b. Type A intraclass correlation coefficients using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Neutral observer agreement scores for post-test interview

Case Processing Summary

		N	%
Cases	Valid	22	100.0
	Excluded[a]	0	.0
	Total	22	100.0

a. Listwise deletion based on all variables in the procedure.

Intraclass Correlation Coefficient

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.336[a]	.054	.615	3.868	21	42	.000
Average Measures	.603[c]	.146	.827	3.868	21	42	.000

Two-way mixed effects model where people effects are random and measures effects are fixed. a.

The estimator is the same, whether the interaction effect is present or not.

b. Type A intraclass correlation coefficients using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

6.4.1.2. Journalist scores of communication skills

Descriptive Statistics

Dependent Variable: Trainer_molar

Time	Group	Mean	Std. Deviation	N
BL	SS Feedback	4.9091	.83121	11
	Traditional Feedback	5.0909	.88708	11
	Total	5.0000	.84403	22
PT	SS Feedback	6.1455	.61378	11
	Traditional Feedback	5.8727	.73361	11
	Total	6.0091	.67465	22
Total	SS Feedback	5.5273	.95328	22
	Traditional Feedback	5.4818	.88943	22
	Total	5.5045	.91141	44

ANOVA Results

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	11.792[a]	3	3.931	6.571	.001	.330
Intercept	1333.201	1	1333.201	2228.755	.000	.982
Time	11.201	1	11.201	18.725	.000	.319
Group	.023	1	.023	.038	.846	.001
Time * Group	.568	1	.568	.950	.336	.023
Error	23.927	40	.598			
Total	1368.920	44				
Corrected Total	35.719	43				

Follow-up analysis

Social signal group improvement

ANOVA

Trainer_molar

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	8.407	1	8.407	15.749	.001
Within Groups	10.676	20	.534		
Total	19.084	21			

Traditional feedback group

ANOVA

Trainer_molar

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3.362	1	3.362	5.074	.036
Within Groups	13.251	20	.663		
Total	16.613	21			

6.4.1.3. Neutral observer scores of communication skills

Descriptive data

Descriptive Statistics

Dependent Variable: No_molar

Time	Group	Mean	Std. Deviation	N
BL	SS Feedback	4.4355	.89132	11
	Traditional Feedback	4.1755	.94750	11
	Total	4.3055	.90748	22
PT	SS Feedback	5.3027	.72911	11
	Traditional Feedback	4.7873	1.03052	11
	Total	5.0450	.91018	22
Total	SS Feedback	4.8691	.91019	22
	Traditional Feedback	4.4814	1.01550	22
	Total	4.6752	.97297	44

ANOVA Results

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	7.849[a]	3	2.616	3.185	.034	.193
Intercept	961.741	1	961.741	1170.799	.000	.967
Time	6.016	1	6.016	7.324	.010	.155
Group	1.654	1	1.654	2.013	.164	.048
Time * Group	.179	1	.179	.218	.643	.005
Error	32.858	40	.821			
Total	1002.448	44				
Corrected Total	40.707	43				

Follow-up analysis

Social Signal Feedback

ANOVA

No_molar

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4.137	1	4.137	6.239	.021
Within Groups	13.260	20	.663		
Total	17.397	21			

Traditional

ANOVA

No_molar

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.059	1	2.059	2.101	.163
Within Groups	19.597	20	.980		
Total	21.656	21			

6.4.1.4. Self-rater scores of communication skills Descriptive data

Descriptive Statistics

Dependent Variable: SR_molar

Time	Group	Mean	Std. Deviation	N
BL	SS Feedback	4.8545	.91254	11
	Traditional Feedback	5.0545	.75943	11
	Total	4.9545	.82562	22
PT	SS Feedback	4.9818	.96935	11
	Traditional Feedback	4.9455	1.09578	11

ANOVA Results

Tests of Between-Subjects Effects

Dependent Variable: SR_molar

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	.228[a]	3	.076	.086	.967	.006
Intercept	1082.074	1	1082.074	1219.300	.000	.968
Time	.001	1	.001	.001	.975	.000
Group	.074	1	.074	.083	.775	.002
Time * Group	.154	1	.154	.173	.680	.004
Error	35.498	40	.887			
Total	1117.800	44				
Corrected Total	35.726	43				

a. R Squared = .006 (Adjusted R Squared = -.068)

Confidence and Skills Ratings

Group Statistics

	Feedback or none	N	Mean	Std. Deviation	Std. Error Mean
SelfEval_skill	Feedback	11	4.0909	.83121	.25062
	No feedback	11	3.1818	.87386	.26348
SelfEval_confidence	Feedback	11	4.3636	.67420	.20328
	No feedback	11	3.3636	.80904	.24393

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2tailed)	Mean Difference	Std. Error Difference	Lower	Upper
SelfEval_skill	Equal variances assumed	.206	.655	2.500	20	.021	.90909	.36364	.15056	1.66762
SelfEval_confidence	Equal variances assumed	.699	.413	3.149	20	.005	1.00000	.31753	.33764	1.66236

6.4.1.6. Self-report scores vs journalist communication scores

Levene's Test of Equality of Error Variances[a]

		Levene Statistic	df1	df2	Sig.
RatingsBL	Based on Mean	.220	3	40	.882
	Based on Median	.164	3	40	.920
	Based on Median and with adjusted df	.164	3	37.694	.920
	Based on trimmed mean	.216	3	40	.885
RatingsPT	Based on Mean	1.161	3	40	.337
	Based on Median	.841	3	40	.479
	Based on Median and with adjusted df	.841	3	34.145	.481
	Based on trimmed mean	1.093	3	40	.363

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Rater + Feedback + Rater * Feedback

Within Subjects Design: session

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Rater	Statistic	df	Sig.	Statistic	df	Sig.
RatingsBL	self-report	.152	22	.200*	.972	22	.763
	trainer	.182	22	.057	.920	22	.076
RatingsPT	self-report	.163	22	.133	.947	22	.270
	trainer	.152	22	.200*	.941	22	.209

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
session	Sphericity Assumed	5.702	1	5.702	15.331	.000	.277
	Greenhouse-Geisser	5.702	1.000	5.702	15.331	.000	.277
	Huynh-Feldt	5.702	1.000	5.702	15.331	.000	.277
	Lower-bound	5.702	1.000	5.702	15.331	.000	.277
session * Rater	Sphericity Assumed	5.500	1	5.500	14.789	.000	.270
	Greenhouse-Geisser	5.500	1.000	5.500	14.789	.000	.270

	Huynh-Feldt	5.500	1.000	5.500	14.789	.000	.270
	Lower-bound	5.500	1.000	5.500	14.789	.000	.270
session * Feedback	Sphericity Assumed	.656	1	.656	1.765	.192	.042
	Greenhouse-Geisser	.656	1.000	.656	1.765	.192	.042
	Huynh-Feldt	.656	1.000	.656	1.765	.192	.042
	Lower-bound	.656	1.000	.656	1.765	.192	.042
session * Rater * Feedback	Sphericity Assumed	.065	1	.065	.176	.677	.004
	Greenhouse-Geisser	.065	1.000	.065	.176	.677	.004
	Huynh-Feldt	.065	1.000	.065	.176	.677	.004
	Lower-bound	.065	1.000	.065	.176	.677	.004
Error(session)	Sphericity Assumed	14.876	40	.372			
	Greenhouse-Geisser	14.876	40.000	.372			
	Huynh-Feldt	14.876	40.000	.372			
	Lower-bound	14.876	40.000	.372			

Follow-up analysis

Descriptives

RatingsBL

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
self-report	22	4.9545	.82562	.17602	4.5885	5.3206	3.00	6.40
trainer	22	5.0000	.84403	.17995	4.6258	5.3742	3.80	6.60
Total	44	4.9773	.82543	.12444	4.7263	5.2282	3.00	6.60

ANOVA

RatingsBL

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.023	1	.023	.033	.858
Within Groups	29.275	42	.697		
Total	29.297	43			

Descriptives

RatingsPT

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
self-report	22	4.9636	1.00974	.21528	4.5159	5.4113	2.80	6.60
trainer	22	6.0091	.67465	.14384	5.7100	6.3082	4.80	7.00
Total	44	5.4864	.99990	.15074	5.1824	5.7904	2.80	7.00

ANOVA

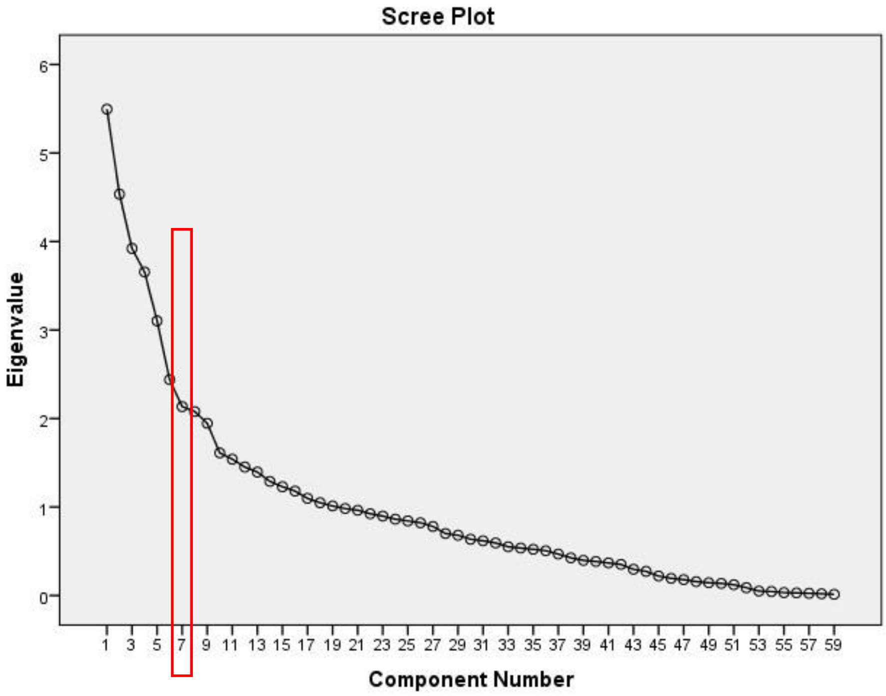
RatingsPT

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	12.023	1	12.023	16.305	.000
Within Groups	30.969	42	.737		
Total	42.992	43			

6.4.2. Social Signal Detection

6.4.2.1. Principal Component Analysis

Scree Plot



Total variance explained

Total Variance Explained

Component	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.495	9.314	9.314	5.187	8.791	8.791
2	4.535	7.686	17.000	4.162	7.054	15.845
3	3.922	6.647	23.646	3.783	6.412	22.257
4	3.656	6.197	29.843	3.605	6.110	28.367
5	3.103	5.259	35.102	3.119	5.287	33.654
6	2.439	4.134	39.236	3.051	5.172	38.826
7	2.134	3.617	42.853	2.376	4.028	42.853
8	2.079	3.524	46.377			
9	1.945	3.297	49.674			
10	1.611	2.731	52.405			
11	1.540	2.611	55.016			
12	1.452	2.461	57.477			
13	1.394	2.363	59.840			
14	1.290	2.186	62.026			

Extraction Method: Principal Component Analysis.

PCA Components Extracted

Rotated Component Matrix[a]

	Component						
	1	2	3	4	5	6	7
EmoCogRatio	.907						
Stressed	-.757						
Excited	.751						
Energy	.741						
I_think	-.733						
Upset	-.683						
BrainPower	.662						
Imagine	-.631						
Uncertain	-.604						
Disgust		.880					
JawDrop		.752					
UpperLipRaise		.733					
NoseWrinkle		.622					
Browfurrow			.931				
LidTighten			.893				
Anger			.869				
Sadness			.744				
M_Activity				.701			
Volume				.692			
M_Consistency				-.665			
Unsuccessfulinterruptions				.643			
P_Activity				.632			
Engagement					.737		
Surprise					.702		
Browraise					.655		
Dimpler						.735	
LipStretch						.705	
LipPress						.658	
LipSuck						.627	
Movement							.670
Posture							-.660

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.[a]

a. Rotation converged in 11 iterations.

6.4.2.2.1. Multivariate analysis

Descriptive Statistics

	Condition_num	Exp_group	Mean	Std. Deviation	N
REGR factor score 1 for analysis 1	PreTraining	Control	.25117271	.753086599	146
		Experiment	-.15908465	.985696402	126
		Total	.06112702	.890847615	272
	PostTraining	Control	-.06828726	.965515341	195
		Experiment	-.02520933	1.178090300	178
		Total	-.04772996	1.070977178	373
	Total	Control	.06849032	.893779796	341
		Experiment	-.08069713	1.102696461	304
		Total	-.00182438	.999699620	645
REGR factor score 2 for analysis 1	PreTraining	Control	.58636395	1.263443640	146
		Experiment	-.16832973	.775909320	126
		Total	.23676320	1.128688770	272
	PostTraining	Control	.07573235	.976897427	195
		Experiment	-.43258743	.583446105	178
		Total	-.16684385	.851053942	373
	Total	Control	.29436054	1.135485476	341
		Experiment	-.32305957	.681337316	304
		Total	.00335944	.997121277	645
REGR factor score 3 for analysis 1	PreTraining	Control	-.08794367	.416835594	146
		Experiment	.55900101	2.030093535	126
		Total	.21174394	1.448578216	272
	PostTraining	Control	-.16362726	.389362695	195
		Experiment	-.16060953	.305339204	178
		Total	-.16218716	.351318402	373
	Total	Control	-.13122314	.402504288	341
		Experiment	.13765010	1.371403267	304
		Total	-.00449839	.994213439	645
REGR factor score 4 for analysis 1	PreTraining	Control	-.28227002	.611429914	146
		Experiment	.35340282	1.043135212	126
		Total	.01219607	.895979149	272
	PostTraining	Control	-.29095067	.676412814	195
		Experiment	.29923694	1.315220320	178

		Total	-.00930618	1.071817057	373
	Total	Control	-.28723403	.648472623	341
		Experiment	.32168727	1.208343156	304
		Total	-.00023856	1.000757699	645
REGR factor score 5 for analysis 1	PreTraining	Control	-.24325381	.884857685	146
		Experiment	.36833767	1.191631024	126
		Total	.04005695	1.080396692	272
	PostTraining	Control	-.26773156	.800867031	195
		Experiment	.24667194	.986527658	178
		Total	-.02225214	.929382827	373
	Total	Control	-.25725135	.836678304	341
		Experiment	.29709918	1.076071968	304
		Total	.00402394	.995528237	645
REGR factor score 6 for analysis 1	PreTraining	Control	.03130552	.720647762	146
		Experiment	-.12488511	1.096973404	126
		Total	-.04104749	.915975617	272
	PostTraining	Control	.24893538	1.080939661	195
		Experiment	-.20562338	.980836701	178
		Total	.03201458	1.057721675	373
	Total	Control	.15575662	.948579658	341
		Experiment	-.17215949	1.029564531	304
		Total	.00120391	1.000307469	645
REGR factor score 7 for analysis 1	PreTraining	Control	.03555756	.891260619	146
		Experiment	.15586003	1.092937693	126
		Total	.09128591	.989750086	272
	PostTraining	Control	-.13251585	.696142217	195
		Experiment	.07663118	.820881212	178
		Total	-.03270842	.764389493	373
	Total	Control	-.06055480	.788807148	341
		Experiment	.10946945	.942309405	304
		Total	.01958066	.868038182	645

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.005	.448[b]	7.000	635.000	.872	.005
	Wilks' Lambda	.995	.448[b]	7.000	635.000	.872	.005
	Hotelling's Trace	.005	.448[b]	7.000	635.000	.872	.005
	Roy's Largest Root	.005	.448[b]	7.000	635.000	.872	.005
Interview type (session)	Pillai's Trace	.084	8.358[b]	7.000	635.000	.000	.084
	Wilks' Lambda	.916	8.358[b]	7.000	635.000	.000	.084
	Hotelling's Trace	.092	8.358[b]	7.000	635.000	.000	.084
	Roy's Largest Root	.092	8.358[b]	7.000	635.000	.000	.084
Group	Pillai's Trace	.329	44.492[b]	7.000	635.000	.000	.329
	Wilks' Lambda	.671	44.492[b]	7.000	635.000	.000	.329
	Hotelling's Trace	.490	44.492[b]	7.000	635.000	.000	.329
	Roy's Largest Root	.490	44.492[b]	7.000	635.000	.000	.329
Interview*Group	Pillai's Trace	.052	4.937[b]	7.000	635.000	.000	.052

	Wilks' Lambda	.948	4.937[b]	7.000	635.000	.000	.052
	Hotelling's Trace	.054	4.937[b]	7.000	635.000	.000	.052
	Roy's Largest Root	.054	4.937[b]	7.000	635.000	.000	.052

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	REGR factor score 1 for analysis 1	13.420[a]	3	4.473	4.550	.004	.021
	REGR factor score 2 for analysis 1	88.189[b]	3	29.396	34.129	.000	.138
	REGR factor score 3 for analysis 1	50.301[c]	3	16.767	18.332	.000	.079
	REGR factor score 4 for analysis 1	59.815[d]	3	19.938	21.841	.000	.093
	REGR factor score 5 for analysis 1	50.532[e]	3	16.844	18.371	.000	.079
	REGR factor score 6 for analysis 1	21.717[f]	3	7.239	7.452	.000	.034
	REGR factor score 7 for analysis 1	7.468[g]	3	2.489	3.340	.019	.015
Intercept	REGR factor score 1 for analysis 1	7.770E-5	1	7.770E-5	.000	.993	.000
	REGR factor score 2 for analysis 1	.147	1	.147	.170	.680	.000
	REGR factor score 3 for analysis 1	.844	1	.844	.923	.337	.001
	REGR factor score 4 for analysis 1	.247	1	.247	.271	.603	.000

	REGR factor score 5 for analysis 1	.424	1	.424	.462	.497	.001
	REGR factor score 6 for analysis 1	.099	1	.099	.102	.750	.000
	REGR factor score 7 for analysis 1	.719	1	.719	.965	.326	.002
Condition_num	REGR factor score 1 for analysis 1	1.349	1	1.349	1.372	.242	.002
	REGR factor score 2 for analysis 1	23.518	1	23.518	27.304	.000	.041
	REGR factor score 3 for analysis 1	24.773	1	24.773	27.085	.000	.041
	REGR factor score 4 for analysis 1	.155	1	.155	.169	.681	.000
	REGR factor score 5 for analysis 1	.837	1	.837	.912	.340	.001
	REGR factor score 6 for analysis 1	.734	1	.734	.756	.385	.001
	REGR factor score 7 for analysis 1	2.395	1	2.395	3.214	.073	.005
Exp_group	REGR factor score 1 for analysis 1	5.280	1	5.280	5.371	.021	.008
	REGR factor score 2 for analysis 1	62.479	1	62.479	72.538	.000	.102
	REGR factor score 3 for analysis 1	16.546	1	16.546	18.091	.000	.027
	REGR factor score 4 for analysis 1	58.857	1	58.857	64.473	.000	.091
	REGR factor score 5 for analysis 1	49.658	1	49.658	54.159	.000	.078
	REGR factor score 6 for analysis 1	14.610	1	14.610	15.040	.000	.023

	REGR factor score for analysis 1 7	4.251	1	4.251	5.703	.017	.009
Condition_num * Exp_group	REGR factor score for analysis 1 1	8.049	1	8.049	8.187	.004	.013
	REGR factor score for analysis 1 2	2.377	1	2.377	2.760	.097	.004
	REGR factor score for analysis 1 3	16.240	1	16.240	17.756	.000	.027
	REGR factor score for analysis 1 4	.081	1	.081	.089	.766	.000
	REGR factor score for analysis 1 5	.370	1	.370	.403	.526	.001
	REGR factor score for analysis 1 6	3.487	1	3.487	3.589	.059	.006
	REGR factor score for analysis 1 7	.309	1	.309	.415	.520	.001
Error	REGR factor score for analysis 1 1	630.193	641	.983			
	REGR factor score for analysis 1 2	552.109	641	.861			
	REGR factor score for analysis 1 3	586.267	641	.915			
	REGR factor score for analysis 1 4	585.161	641	.913			
	REGR factor score for analysis 1 5	587.721	641	.917			
	REGR factor score for analysis 1 6	622.679	641	.971			
	REGR factor score for analysis 1 7	477.780	641	.745			
Total	REGR factor score for analysis 1 1	643.615	645				
	REGR factor score for analysis 1 2	640.305	645				

	REGR factor score 3 for analysis 1	636.582	645				
	REGR factor score 4 for analysis 1	644.976	645				
	REGR factor score 5 for analysis 1	638.264	645				
	REGR factor score 6 for analysis 1	644.397	645				
	REGR factor score 7 for analysis 1	485.495	645				
Corrected Total	REGR factor score 1 for analysis 1	643.613	644				
	REGR factor score 2 for analysis 1	640.298	644				
	REGR factor score 3 for analysis 1	636.568	644				
	REGR factor score 4 for analysis 1	644.976	644				
	REGR factor score 5 for analysis 1	638.253	644				
	REGR factor score 6 for analysis 1	644.396	644				
	REGR factor score 7 for analysis 1	485.248	644				

a. R Squared = .021 (Adjusted R Squared = .016)

b. R Squared = .138 (Adjusted R Squared = .134)

c. R Squared = .079 (Adjusted R Squared = .075)

d. R Squared = .093 (Adjusted R Squared = .088)

e. R Squared = .079 (Adjusted R Squared = .075)

f. R Squared = .034 (Adjusted R Squared = .029)

g. R Squared = .015 (Adjusted R Squared = .011)

6.4.2.2.1. Component 1 – confidence

Descriptive Statistics

Dependent Variable: REGR factor score 1 for analysis 1

Condition_num	Exp_group	Mean	Std. Deviation	N
PreTraining	Control	.25746900	.754375541	147
	Experiment	-.15908465	.985696402	126
	Total	.06521347	.891768257	273
PostTraining	Control	-.06828726	.965515341	195
	Experiment	-.02520933	1.178090300	178
	Total	-.04772996	1.070977178	373
Total	Control	.07173078	.894477986	342
	Experiment	-.08069713	1.102696461	304
	Total	.00000000	1.000000000	646

Tests of Between-Subjects Effects Dependent Variable: REGR factor score 1 for analysis 1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	13.956[a]	3	4.652	4.733	.003	.022
Intercept	.001	1	.001	.001	.975	.000
Condition_num	1.445	1	1.445	1.470	.226	.002
Exp_group	5.473	1	5.473	5.568	.019	.009
Condition_num * Exp_group	8.289	1	8.289	8.433	.004	.013
Error	631.044	642	.983			
Total	645.000	646				
Corrected Total	645.000	645				

a. R Squared = .022 (Adjusted R Squared = .017)

One-way analysis follow-up pre-training

ANOVA

REGR factor score 1 for analysis 1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	11.772	1	11.772	15.598	.000
Within Groups	204.536	271	.755		
Total	216.308	272			

Robust Tests of Equality of Means REGR factor score 1 for analysis 1

	Statistic[a]	df1	df2	Sig.
Welch	14.981	1	231.961	.000

a. Asymptotically F distributed.

One-way analysis for post-training

ANOVA

REGR factor score 1 for analysis 1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.173	1	.173	.150	.699
Within Groups	426.508	371	1.150		
Total	426.681	372			

Robust Tests of Equality of Means REGR factor score 1 for analysis 1

	Statistic[a]	df1	df2	Sig.
Welch	.148	1	342.956	.701

a. Asymptotically F distributed.

One-way anova analysis for control group

ANOVA

REGR factor score 1 for analysis 1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	8.894	1	8.894	11.458	.001
Within Groups	263.937	340	.776		
Total	272.831	341			

Robust Tests of Equality of Means

REGR factor score 1 for analysis 1

	Statistic[a]	df1	df2	Sig.
Welch	12.265	1	339.549	.001

a. Asymptotically F distributed.

One-way analysis follow-up for experiment group

ANOVA

REGR factor score 1 for analysis 1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.322	1	1.322	1.088	.298
Within Groups	367.107	302	1.216		
Total	368.430	303			

Robust Tests of Equality of Means REGR factor score 1 for analysis 1

	Statistic[a]	df1	df2	Sig.
Welch	1.156	1	293.599	.283

a. Asymptotically F distributed.

6.4.2.2.2. Component 2 - Disgust

Descriptive Statistics

Dependent Variable: REGR factor score 2 for analysis 1

Condition_num	Exp_group	Mean	Std. Deviation	N
PreTraining	Control	.56763470	1.279422415	147
	Experiment	-.16832973	.775909320	126
	Total	.22795881	1.135965218	273
PostTraining	Control	.07573235	.976897427	195
	Experiment	-.43258743	.583446105	178
	Total	-.16684385	.851053942	373
Total	Control	.28716406	1.141603348	342
	Experiment	-.32305957	.681337316	304
	Total	.00000000	1.000000000	646

Tests of Between-Subjects Effects Dependent Variable: REGR factor score 2 for analysis 1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	85.363[a]	3	28.454	32.642	.000	.132
Intercept	.071	1	.071	.081	.776	.000
Condition_num	22.436	1	22.436	25.737	.000	.039
Exp_group	60.750	1	60.750	69.691	.000	.098
Condition_num * Exp_group	2.033	1	2.033	2.333	.127	.004
Error	559.637	642	.872			
Total	645.000	646				
Corrected Total	645.000	645				

a. R Squared = .132 (Adjusted R Squared = .128)

One-way ANOVA follow-up for pretraining Interview

ANOVA

REGR factor score 2 for analysis 1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	36.748	1	36.748	31.691	.000
Within Groups	314.245	271	1.160		
Total	350.993	272			

Robust Tests of Equality of Means

REGR factor score 2 for analysis 1

	Statistic[a]	df1	df2	Sig.
Welch	34.037	1	245.401	.000

a. Asymptotically F distributed.

One-way ANOVA follow-up for post-training Interview

ANOVA

REGR factor score 2 for analysis 1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	24.045	1	24.045	36.352	.000
Within Groups	245.392	371	.661		
Total	269.437	372			

Robust Tests of Equality of Means

REGR factor score 2 for analysis 1

	Statistic[a]	df1	df2	Sig.
Welch	37.963	1	321.443	.000

a. Asymptotically F distributed.

One-way ANOVA for control group

ANOVA

REGR factor score 2 for analysis 1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	20.281	1	20.281	16.258	.000
Within Groups	424.130	340	1.247		
Total	444.411	341			

Robust Tests of Equality of Means

REGR factor score 2 for analysis 1

	Statistic[a]	df1	df2	Sig.
Welch	15.095	1	264.137	.000

a. Asymptotically F distributed.

One-way ANOVA of experiment group

ANOVA

REGR factor score 2 for analysis 1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	5.152	1	5.152	11.482	.001
Within Groups	135.507	302	.449		
Total	140.659	303			

Robust Tests of Equality of Means

REGR factor score 2 for analysis 1

	Statistic[a]	df1	df2	Sig.
Welch	10.438	1	220.177	.001

a. Asymptotically F distributed.

6.4.2.2.3. Component 3 - frowning

Descriptive Statistics

Dependent Variable: REGR factor score 3 for analysis 1

Condition_num	Exp_group	Mean	Std. Deviation	N
PreTraining	Control	-.06760759	.483067918	147
	Experiment	.55900101	2.030093535	126
	Total	.22159638	1.455047938	273
PostTraining	Control	-.16362726	.389362695	195
	Experiment	-.16060953	.305339204	178
	Total	-.16218716	.351318402	373
Total	Control	-.12235565	.434081692	342
	Experiment	.13765010	1.371403267	304
	Total	.00000000	1.000000000	646

Tests of Between-Subjects Effects Dependent Variable: REGR factor score 3 for analysis 1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	49.857[a]	3	16.619	17.928	.000	.077
Intercept	1.096	1	1.096	1.183	.277	.002
Condition_num	26.103	1	26.103	28.158	.000	.042
Exp_group	15.555	1	15.555	16.780	.000	.025
Condition_num * Exp_group	15.258	1	15.258	16.460	.000	.025
Error	595.143	642	.927			
Total	645.000	646				
Corrected Total	645.000	645				

a. R Squared = .077 (Adjusted R Squared = .073)

Pretraining

ANOVA

REGR factor score 3 for analysis 1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	26.639	1	26.639	13.144	.000
Within Groups	549.230	271	2.027		
Total	575.869	272			

Robust Tests of Equality of Means REGR factor score 3 for analysis 1

	Statistic[a]	df1	df2	Sig.
Welch	11.449	1	137.151	.001

a. Asymptotically F distributed.

Post training

ANOVA

REGR factor score 3 for analysis 1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.001	1	.001	.007	.934
Within Groups	45.913	371	.124		
Total	45.914	372			

Robust Tests of Equality of Means

REGR factor score 3 for analysis 1

	Statistic[a]	df1	df2	Sig.
Welch	.007	1	362.912	.933

a. Asymptotically F distributed.

Control- one-way ANOVA

ANOVA

REGR factor score 3 for analysis 1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.773	1	.773	4.139	.043
Within Groups	63.481	340	.187		
Total	64.254	341			

Robust Tests of Equality of Means REGR factor score 3 for analysis 1

	Statistic[a]	df1	df2	Sig.
Welch	3.899	1	274.480	.049

a. Asymptotically F distributed.

Experiment Group

ANOVA

REGR factor score 3 for analysis 1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	38.204	1	38.204	21.701	.000
Within Groups	531.662	302	1.760		
Total	569.866	303			

Robust Tests of Equality of Means REGR factor score 3 for analysis 1

	Statistic[a]	df1	df2	Sig.
Welch	15.582	1	129.012	.000

a. Asymptotically F distributed.

6.4.2.2.4. Component 4 – Eagerness to speak

Descriptive Statistics

Dependent Variable: REGR factor score 4 for analysis 1

Condition_num	Exp_group	Mean	Std. Deviation	N
PreTraining	Control	-.27930306	.610393290	147
	Experiment	.35340282	1.04313521 2	126
	Total	.01271504	.894371717	273
PostTraining	Control	-.29095067	.676412814	195
	Experiment	.29923694	1.31522032 0	178
	Total	-.00930618	1.07181705 7	373
Total	Control	-.28594424	.647960255	342
	Experiment	.32168727	1.20834315 6	304

Total	.00000000	1.00000000 0	646

Tests of Between-Subjects Effects Dependent Variable: REGR factor score 4 for analysis 1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	59.650[a]	3	19.883	21.808	.000	.092
Intercept	.266	1	.266	.292	.589	.000
Condition_num	.170	1	.170	.186	.666	.000
Exp_group	58.679	1	58.679	64.358	.000	.091
Condition_num * Exp_group	.071	1	.071	.078	.780	.000
Error	585.350	642	.912			
Total	645.000	646				
Corrected Total	645.000	645				

a. R Squared = .092 (Adjusted R Squared = .088)

Pretraining Group

ANOVA

REGR factor score 4 for analysis 1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	27.160	1	27.160	38.655	.000
Within Groups	190.413	271	.703		
Total	217.573	272			

Robust Tests of Equality of Means REGR factor score 4 for analysis 1

	Statistic[a]	df1	df2	Sig.
Welch	35.837	1	194.775	.000

a. Asymptotically F distributed.

Post training Group

ANOVA

REGR factor score 4 for analysis 1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	32.414	1	32.414	30.449	.000
Within Groups	394.937	371	1.065		
Total	427.351	372			

Robust Tests of Equality of Means REGR factor score 4 for analysis 1

	Statistic[a]	df1	df2	Sig.
Welch	28.872	1	259.013	.000

a. Asymptotically F distributed.

One way ANOVA of Control group

ANOVA

REGR factor score 4 for analysis 1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.011	1	.011	.027	.870
Within Groups	143.158	340	.421		
Total	143.170	341			

Robust Tests of Equality of Means REGR factor score 4 for analysis 1

	Statistic[a]	df1	df2	Sig.
Welch	.028	1	329.150	.868

a. Asymptotically F distributed.

One-way ANOVA of Experiment group

ANOVA

REGR factor score 4 for analysis 1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.216	1	.216	.148	.701
Within Groups	442.192	302	1.464		
Total	442.408	303			

Robust Tests of Equality of Means REGR factor score 4 for analysis 1

	Statistic[a]	df1	df2	Sig.
Welch	.160	1	298.062	.690

a. Asymptotically F distributed.

6.4.2.2.5. Component 5 – expression engagement

Descriptive Statistics

Dependent Variable: REGR factor score 5 for analysis 1

Condition_num	Exp_group	Mean	Std. Deviation	N
PreTraining	Control	-.25925508	.902910988	147
	Experiment	.36833767	1.191631024	126
	Total	.03040311	1.090141403	273
PostTraining	Control	-.26773156	.800867031	195
	Experiment	.24667194	.986527658	178
	Total	-.02225214	.929382827	373
Total	Control	-.26408816	.844963584	342
	Experiment	.29709918	1.076071968	304
	Total	.00000000	1.000000000	646

Tests of Between-Subjects Effects Dependent Variable: REGR factor score 5 for analysis 1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	51.784[a]	3	17.261	18.681	.000	.080
Intercept	.304	1	.304	.329	.566	.001
Condition_num	.665	1	.665	.719	.397	.001
Exp_group	51.173	1	51.173	55.381	.000	.079
Condition_num * Exp_group	.503	1	.503	.544	.461	.001
Error	593.216	642	.924			
Total	645.000	646				
Corrected Total	645.000	645				

a. R Squared = .080 (Adjusted R Squared = .076)

Pretraining Group

ANOVA

REGR factor score 5 for analysis 1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	26.723	1	26.723	24.422	.000
Within Groups	296.524	271	1.094		
Total	323.247	272			

Robust Tests of Equality of Means REGR factor score 5 for analysis 1

	Statistic[a]	df1	df2	Sig.
Welch	23.423	1	230.506	.000

a. Asymptotically F distributed.

Post training Group

ANOVA

REGR factor score 5 for analysis 1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	24.624	1	24.624	30.791	.000
Within Groups	296.692	371	.800		
Total	321.316	372			

Robust Tests of Equality of Means

REGR factor score 5 for analysis 1

	Statistic[a]	df1	df2	Sig.
Welch	30.218	1	341.316	.000

a. Asymptotically F distributed.

6.4.2.2.6. Component 6 – posed expression

Descriptive Statistics

Dependent Variable: REGR factor score 6 for analysis 1

Condition_num	Exp_group	Mean	Std. Deviation	N
PreTraining	Control	.02581010	.721259634	147
	Experiment	-.12488511	1.096973404	126
	Total	-.04374153	.915373216	273
PostTraining	Control	.24893538	1.080939661	195
	Experiment	-.20562338	.980836701	178
	Total	.03201458	1.057721675	373

Tests of Between-Subjects Effects Dependent Variable: REGR factor score 6 for analysis 1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	21.673[a]	3	7.224	7.441	.000	.034
Intercept	.122	1	.122	.126	.723	.000
Condition_num	.796	1	.796	.819	.366	.001
Exp_group	14.374	1	14.374	14.805	.000	.023
Condition_num * Exp_group	3.623	1	3.623	3.732	.054	.006
Error	623.327	642	.971			
Total	645.000	646				
Corrected Total	645.000	645				

a. R Squared = .034 (Adjusted R Squared = .029)

Pretraining

ANOVA

REGR factor score 6 for analysis 1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.541	1	1.541	1.844	.176
Within Groups	226.370	271	.835		
Total	227.911	272			

Robust Tests of Equality of Means REGR factor score 6 for analysis 1

	Statistic[a]	df1	df2	Sig.
Welch	1.735	1	210.101	.189

a. Asymptotically F distributed.

Post training Interview

ANOVA

REGR factor score 6 for analysis 1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	19.228	1	19.228	17.970	.000
Within Groups	396.957	371	1.070		
Total	416.184	372			

Robust Tests of Equality of Means REGR factor score 6 for analysis 1

	Statistic[a]	df1	df2	Sig.
Welch	18.130	1	370.988	.000

a. Asymptotically F distributed.

6.4.2.2.7. Component 7 - Posture

Descriptive Statistics

Dependent Variable: REGR factor score 7 for analysis 1

Condition_num	Exp_group	Mean	Std. Deviation	N
PreTraining	Control	.03555756	.891260619	146
	Experiment	.15586003	1.092937693	126
	Total	.09128591	.989750086	272
PostTraining	Control	-.13251585	.696142217	195
	Experiment	.07663118	.820881212	178
	Total	-.03270842	.764389493	373

Tests of Between-Subjects Effects Dependent Variable: REGR factor score 7 for analysis 1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	7.468[a]	3	2.489	3.340	.019	.015
Intercept	.719	1	.719	.965	.326	.002
Condition_num	2.395	1	2.395	3.214	.073	.005
Exp_group	4.251	1	4.251	5.703	.017	.009
Condition_num * Exp_group	.309	1	.309	.415	.520	.001
Error	477.780	641	.745			
Total	485.495	645				
Corrected Total	485.248	644				

a. R Squared = .015 (Adjusted R Squared = .011)

Pretraining Group

ANOVA

REGR factor score 7 for analysis 1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.979	1	.979	.999	.318
Within Groups	264.494	270	.980		
Total	265.473	271			

Robust Tests of Equality of Means REGR factor score 7 for analysis 1

	Statistic[a]	df1	df2	Sig.
Welch	.970	1	241.169	.326

a. Asymptotically F distributed.

Post training Group

ANOVA

REGR factor score 7 for analysis 1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4.071	1	4.071	7.080	.008
Within Groups	213.286	371	.575		
Total	217.356	372			

Robust Tests of Equality of Means REGR factor score 7 for analysis 1

	Statistic[a]	df1	df2	Sig.
Welch	6.976	1	348.603	.009

a. Asymptotically F distributed.

Qualitative Interview transcriptions

P 18 did not record but here are the notes
P18_ liked the layout of the design and couldn't find anything more that she would like feedback on. She did feel that it was very rushed, however. Perhaps to make things quicker is to feed into the graph itself or get some help.

1. I felt good about it, it was very clear
2. Yes, as I said because the feedback was very clear
3. I felt that I improved based on the feedback
4. I feel that I will be more aware of the issues surrounding my performance today 5. I can't think of any as everything was feedback to me well

1. It was very clear and informative
2. Yes, it was very informative about my body language
3. I was able to respond to the feedback but maybe there was too many things to be aware of. It was hard to focus on which of the elements that I should improve
4. I can't think of any
5. Just pay attend to time and it felt very rushed.

P19

M: so the first question is how do you feel about the feedback you received on your performance today?

P: I believe it was good, very honest, if you understand that.

M: Okay is that. So thank you. So what, if anything, did you feel you were able to change based on the feedback that you received?

P: Yeah I would change like the feedback she gave me. Like my hand gestures, projection of voice, I feel it should be better. The way of explanation as well. You need to sometimes you need to make a sentence in your head before speaking, so it should be in your head what you are speaking. So instead of saying, ah, uhm, you know. She speaks like that.

M: Okay, so you felt you were able to change that, so the last… P: Confident.

M: Oh okay, more confident, that's good. So to what extent do you feel your performance in the areas identified in the feedback improved over the course of the training?

P: Yes, it improves, for the training ones, face to face is a different experience in the camera that is different. Yeah. So the last, I was the best I could give, I think.

M: Yeah, yeah? So you felt more confident in face to face. Okay that's good. So to what extent do you feel the feedback you received today will help you present will help you present yourself better in future?

P: It will definitely help me, yeah. Through the feedback because uhm without the training and today I came here, I do not understand what are the drawbacks and the or how I have to present myself in front of people or I have to give an interview to someone. I didn't know that. So today was a great experience. I gave interview to someone and she gave me feedback. That was, I like it.

M: Okay, thank you. That's good. Okay. So you think you'll be able to present yourself better in the future? That's always good to hear. Okay so were there any element of the feedback that you would have liked to receive feedback on that were not covered in the feedback?

P: Sorry?

M: So do you feel that there were some things that you would have liked to receive feedback on that you did not receive feedback on?

P: No it's not like that. I received what I ever I was thinking she was gonna say and she said that.

M: Okay so you felt that it was useful?

P: Yeah?

M: So do you have anything else you would like to add? About the feedback specifically?

P: Everything was great and everything was nice. As a research everything was fine. But for time management your program you have to switch, but maybe there could be things in a better way? Maybe, I am not sure because I am not in your field, I don't know, you know. It is really difficult. I really don't know. Maybe that is a time consuming.

M: It is time consuming and that is the problem with this but it is great that you mention that. It is possible that on the 3rd or 4th day that I run this experiment that it will be better. Uhm, but regards to feedback? Anything?

P: No, it's great.

M: Okay that's great. Thank you.

P20

M: Ok hi, thank you very much for taking part in this and you know you can always withdraw your participation within the study until and if we publish. Ok so I have a few questions for you, and the first one is, how do you feel about the feed back you received on your performance today?

P: I think it was good, it had a good direction for what I needed to sort of maybe do next, maybe a bit more details sort of like examples, but I think it was great.

M: Ok good. So what if anything did you feel you were able to change about how you presented yourself today based on the feedback you received?

P: Yes it was sort of like the move in the arms and being sort of less monitoring, so I tried to incorporate that in the next interview.

M: Ok thank you and just three more, ok so to what extent do you feel your performance in the areas identified in the feedback improved over the course of the day?

P: I think in general it improved just because you sort of practice it, but it was good to hear the feedback, just because then you know from obviously the persons sitting opposite you exactly what you can improve on or in general what you are doing right so you can improve on the things they telling you and then you can just carry on doing the things you are doing right. So I think it improved in general and became better overall.

M: Ok good. Ok to what extent do you feel the feedback you received will help you better yourself in the future?

P: I think this is quite useful because I am not great at any size of presentation or interview, so this was great to sort if practice and build the confidence up, especially about speaking about research because we do a lot of that at sort of conference, different networking events, so this was great you know to learn how to come across better and how to present your research better to other people.

M: Great. Ok were there any aspects of your performance that you would like to have liked to have received back on that weren't covered?

P: Umm

M: And if so what were they?

P: I, ooh god

M: I know it's a bit of a tough one

P: I think everything was covered maybe the body language because well..

M: Yeah

P: But I think that would have been better to see sort of whether I could of done better or how I.. It would have been interesting to see how I've dealt with that in general and if there were areas for improvement or if there was areas I was actually alright in I think that's the main one.

M: Ok, that's interesting. Ok then that is everything, thank you. P: no worries

P21

M: So, thank you -put her name here I don't know it- for coming today, you should know that you can withdraw your participation at any point that you would like, up until we publish that's if we publish. Ok so there is 10 questions that I have to ask you, that I would like to ask you today, the first one is how do you feel about the feedback you received on your performance today?

P: I think it was very consistent because you showed me like with the technology and also the journalist she gives like good feedback and I mean very impressive because she didn't write anything on the paper and she told me very good points that I probably have to improve and also that I did good and with your feedback showed me that it was very good also. I think very important for me to you know try and be better in the interviews.

M: Definitely. Ok so what if, what did you feel you were able to change how you presented yourself today based on the feedback you received?

P: I think based on the feedback I think I can be like paying more attention with the people I am talking to, you know and also I speak very loud -laughing- for example the volume so it's better to pay attention where as the volume where as speaking to be, so that was like a good feedback because sometimes you don't realise you can be a little bit more equal when you are talking and I think this is kind of polite thing to do because sometimes if they are speaking very low and you are very loud. So I think, yeah I think that, things like that the posture also like how to be more relax not to be too much front, be more like middle and I think it was like great feedback and I will take with me you know like next interviews.

M: Well then essentially umm, ok so how do you feel you performed over the areas, within the feedback that's been given do you think you have improved throughout the day?

P: I think so because as you show me in the results from the technology, I think I did try improve a little bit.

M: Well yeah, ok so using this feedback do you think yo u will be able to present better in the future?

P: For sure yeah of course. I will pay more attention, I think I have seen present in conference you know because I hate it because I do a lot of expressions, so I don’t like a ot I have to remember to not make too much face but yeah because I think like when you show upset, maybe its because of that also because im talking a lot doing faces.

M: ok so were there any elements of the performance that you would have liked to receive feedback from that you weren’t received that weren’t given to you?

P: I think that it was no, I think we talk about everything about the posture about the volume about the hands, I think was like a great feedback

M: ok is there anything else that you would like to know about? With egarding to feedback?

P: No I don’t think so, I think it was very good, was enough you know about like presenting and in like the media. I think you cover everything

M: OK great that’s good. Ok and just five more. How do you feel about seeing the systems feedback on your use of emotional signals and body language during the course of the day?

P: I think its amazing because you can show me right , the first time I didn’t understand that I am gonna see t and do it again to be honest, I thought that maybe you gonna give me the feedback and we gonna do like other kinds of tests not like interview again to see the same results. I think it was amazing, very fast I mean it takes time to put it but right away you can tell me how the things
M: I did

P: You did so its amazing I like it.

M: ok good . so do you think that the system heloed you appreciate aspects of your performance that you might not have noticed by watching the playback?

P: well yes because as you told me for example the smile thing because I did not realise that I am making a happy face

M: yeah yeah

P: because I am talking and thinking about what I am saying but yeah and then the devices gave me that response you know so

M: Yeah definaely, ok so what extent did you feel the feedback received was actionable so you could act on the feedback given to you?

P: for example like the volume to be a little bit less

M: yeah so like that kinda stuff and then did you think you could control or improve that feedback?

P: I think it can improve if I pay more attention with my posture and volume maybe

M: and during the sessions? So when the feedback was given to you? Did you feel you could act on it? Or was it quite tough?

P: no it was not tough I tried to

P22

M: Okay so. Thank you for taking part in this study. You should know that you can withdraw your participation at any point up until, because we are aiming to publish this, but up until we publish the paper you can withdraw.

P: Okay so when you publish, you can give me a copy.

M: I will of course. Uhm okay, so I have several questions to ask you about everything or how you felt today went. The first question is how did you feel about the feedback that you received on your performance today?

P: Yeah, I think it is really helpful for my first year actually. Because I have less experience of interview. I don't know. When I have interview there is eye reflection. So this has given me a good experience. Yeah.

M: Okay that is good. So the second one is what did you feel you were able to change based on the feedback that you received?

P: Uhm, how I changed?

M: Yeah, so in the beginning how you performed and then feedback given by the journalist based on how you changed?

P: Yeah, Uhm, actually in the last interview. I control my eyes because the journalist in the beginning.
Also it is in the nature of the body behaviour, it is too much. I haven't touched my mouth like the last. And I try not to speak too much with 'kisses teeth'. I see I improved.

M: So you felt you were able to change that throughout?

P: Yeah, yeah sure.

M: Okay so to what extent do you think your performance improved over the day? So do you think you improved over the day?

P: Yeah sure.

M: Yeah? Okay. So do you think the feedback you received today will help you present yourself better in the future?

P: Yes.

M: Anything else? Just yes?

P: Yes it was helpful.

M: haha okay that's fine. So were there any aspects of the performance that you would have liked to receive feedback on that you didn't receive feedback on?

P: Uhm, sorry?

M: So is there any point in your performance that you would have liked to receive feedback but didn't get any feedback for it?

P: Uhm, actually I wanted feedback about my… M: Content?

P: No, the uhm, what I am talking.

M: Yeah.

P: Okay cause some questions I think I answered more academic and not very simple to understand. I mean like MATLAB or something like that.

M: I noticed that, that's why I kept telling her to tell you about your content. Cause that's important.

P: Yeah, last time she asked me about the code I tried ot use a simple word to explain but it was very hard.

M: Yes it is hard. Okay so contact you would have liked more feedback on?

P: yeah, for others, year I guess I would have liked feedback about my minor movements, eyes and uhm loud or something. I think that's enough.

M: Okay then, thank you very much.

P23

M: Thank you for taking part in the study today. You should know that you can withdraw up until we publish the paper on the research. So, I have a few questions for you today and that is up to ten. The first question is how did you feel about the feedback you received on your performance today?

P: It was great. Educational it was great.

M: ok that's good. So, what did you feel you were able to change about how you presented yourself today based on the feedback you received?

P: I realised I moved my eyebrows too much when I talk, which is very, very true.

M: Yeah that's interesting. Ok so just your eyebrows?

P: And uh whatever you are feeling these software's can detect -laughing- M: So, tell me then what do you mean? Can you elaborate that?

P: ok so with for example the content issue, I was getting asked the same questions over and over again, which is obviously I understand it is part of it, but even when I thought 'oh here we go again' it came across in the…. M: That you were upset, yeah

P: even though it wasn't as I was like upset I was slightly annoyed and it came across, and I am known to show everything on my face and it's amazing that it detected that.

M: Great that's amazing, ok so to what extent did you feel the performance in areas identified in feedback improved over the course of the day?

P: Very well. I dunno I didn't get my feedback from my last interview M: Oh, right the post test.

P: Yeah, the post test is what I meant, because that's when I found out about the eyebrow situation

M: With regards to down the line, from voice, face

P: The one thing I must say is, especially with the voice interview because of my whole situation of saying 'you know what I mean?' that went away M: Yeah?

P: because of the feedback with my performance it got better. I didn't once say or have the urge, I didn't have to tell myself consciously don't say it. I just knew, it just went away. Which is good M: yeah that's very good, ok, so to what extent do you think the feedback you received will help you present yourself better in the future?

P: Massively it's made me a lot more, it's made me less fearful of going through an interview process like it's made me realise, it's all about control like if you like to control situations I now know what a good performance in an interview is. So, thank you

M: Of course. Ok so were there any aspects of your performance that you would of liked to receive feedback on that was not covered in the feedback you received? Do you understand the question?

P: yeah sorry I just need to think about it

M: Ok sure, sure

P: No, no I don't think so because every time I thought I wonder if this is ok I got feedback on it M:

From what?

P: Like the voice, the volume

M: so, with the system feedback?

P: Yeah, the feedback you gave me. So, every time I thought 'ooh am I too loud?' I got feedback. Am I hesitating too much? Ooh and that's another thing I learnt, not to hesitate just going back to your previous question... Every time I thought 'ooh I wonder if this is ok I got feedback on it, I didn't have any uncoveredness

P: ok that's great. So were there any aspects of your performance that you.. I just read that to you I'm so sorry -laughing- how do you feel about seeing the system feedback on your use of emotional signals and body language during the course of the training session today?

P: it was great, very useful because it is immediate, not immediate it happens straight after, I like the fact that you do the interview and five minutes later you get your feedback and then you get to do another interview and then five minutes later.. its great, perfect training because you do, you get feedback, you do you get feedback

M: And specifically how the feedback was presented?

P: yeah very easy to understand, because there is a direct comparison of what's good and what's yours. Its all very good,

P: ok so do you feel the system feedback helped you appreciate your performance that you might not have noticed by watching or?

M: definitely yeah,

M: ok

P: made me realise in=m not as loud as I think I am -laughing-

M: ok so to what extent did you feel the feedback you received was actionable?

P: because, definitely I agree it was actionable, it wasn't things I didn't understand 'you're hesitating too much' so I didn't hesitate, you raise your eyebrows too much so I didn't raise them. I realised I got feedback on saying you know what I mean when I shouldn't have and I was able to rectify that. And was also given another opportunity to not do it, does that make sense?

M: absolutely

P: given repeat exposure

M: ok so could the feedback be presented in a clear way? If so… P:

in a clearer way?

M: yeah

P: no

P24

M: Ok so first off thank you for taking part in the study today and you should know that you can withdraw at any point up until when and if we publish. Right so I have a few questions for you and they are about 10 questions, so the first one is how did you feel about the feedback you received on your performance today?

P: I felt good about the feedback that you gave me but honestly, I felt like the feedback from the journalists was not very specific was too general. Ok its good but I felt like you know I wasn't able to understand where I needed to improve whereas you specifically said this this and that I got it so I knew where I had to improve next time, that's how I felt.

M: Ok that's good, ok thank you. Ok so what did you feel you were able to change about how you presented yourself based on the feedback you received?

P: you mean your feedback or in general?

M: Both feedback? P: well as I said based on the journalist's feedback, I wouldn't even remember afterwards what I needed to improve. Like uh the only thing I remembered was emotion that she said I had to improve because she was specific.

M: Oh, like the monotone?

P: Like she said maybe something to improve is show more emotion, so I remembered that because it was specific but apart from that, when you mentioned the points you know, responding quicker, mirroring her tone and try not to …. -sneezing- oh bless you

M: Thank you

P: try not to do that too much, try to smile more, a lot of things that you mentioned were in my head when I was doing the next session.

M: Ok, that's good. That's amazing ok. So how do you think the feedback helped you improve throughout the day? Through the training?

P: I think well it made 100% of the difference because I could, you know one session after the other I was able to focus on what we just discussed. So, after listening to it I wanted to do it differently, so I was like excited to do it better.

M: yeah that's fair enough. Ok that's good. So, do you think that this will help you present yourself better in the future?

P: I do think yeah, I love doing presentations, take it as a challenge and sometimes I try to be careful not to smile because I smile a lot I mean I see you and I'm like hi, how are you? And professionally I try not to but and now its good to know that I can relax more and be more friendly even though I'm talking about serious problems

M: Yeah that's fair enough. Ok so what are the aspects of the performance that you would of liked to receive feedback on that you didn't receive feedback on?

P: Ummm I can't think about anything now like you know you mentioned pasture, expression, um hesitation. I don't think I will be able to say.

M: ok o how did you feel about the system feedback, how did you feel about the body language and the emotional signals that was feedback to you?

P: I feel like it did correspond to the reality like I can see myself doing the things that you mentioned when I was doing well and what I could improve. And I liked seeing the charts.

M: Yeah

P: Yeah but I found it a little bit difficult to differentiate the orange from the blue and like what those meant but then you explained, and I got it but if I were to see it on my own I would probably struggle a little bit, but not a big deal.

M: Ok, right that's wonderful thank you. So, did you feel that the system feedback helped you appreciate the aspects of the performance that you might not have noticed by just watching or hearing the playback? P: Yeah as I said I'm very hard on myself, very difficult for me to see the good points. And I normally know that I try to speak well like in general. Never spoke to the media apart from my YouTube channel and a few interviews and small things but I didn't know how I would do. Yeah based on what the journalist said it looks like I do well so it's good for me to feel more confident about it.

M: That's what you need. So, do you feel like the system heled you appreciate aspects of your performance that you might not have not by watching or hearing the video?

P: umm sorry what do you mean like which video?

M: so, the performance, you know watching your interview again, so did the bar chart and the features. Did you think you got more from this than watching the video, if you were to watch the video?

P: right so I think if just watch the video I would not look at it the same way you guys would because I'm more used to seeing myself speak like I may find things ok but actually I know what I know what I wanna say, where as you have to understand what I wanna say without knowing what I wanna say if that makes sense? I think it's a positive thing to watch the video but I think its different, its another thing to have someone tell you how you've done. For me those are two positive things but you are always bias when you watching yourself, because you may not see things because you are used to it because you know what you meant so its easier for you. Where if its another person it more unbiased feedback I think. Its good, but maybe I think one compliments the other I would hear your feedback first and then watch my video because if I watch it first.. I don't know if that made sense?

M: No it does absolutely just thinking about what you're saying. Ok so did you think that the feedback you received was actionable, were you able to act on it and improve?

P: Yeah sure

M: Ok could the feedback be presented in a clearer way do you think?

P: well maybe just in terms of the colours of the chart and the columns I don't know its just an impression that maybe I don't know how but it could be clearer but at the same time because you explained it I got it, I think just the speed of response something like that, these terms im not familiar with it I think that's maybe why I wasn't able to look at it and understand straight away seeing what you meant, but of course after you explained it I think it got it that's a big deal.

M: no, it is that makes a lot of sense thank you. So, were there any additional aspects that the nonverbal behaviour that you think you would have been useful to receive feedback on?

P: no verbal I don't think so

M: Ok do you have any comments or experiences that you would like to share about how you presented?

P: no, I think it was a nice experience, I mean I like participating because I know how difficult it is for researches to get participants but I also like the idea of learning something new. Like you always coming thinking its not gonna be that good but I do honestly think it was good, especially doing a PhD you don't really receive a lot of feedback so when you receive it, it makes you more motivated well I feel that way so I think it was a good experience and listen to some points that I can work on to make it better to listen and hear from you and you guys that its good even though there are some points of being prepped so I feel more motivated so in general it was a good experience M: ok good thank you so much

P25

M: Okay, thank you so much for taking part in this study and you should know that you can withdraw your participation at any point, uhm, up until we publish.

P: Okay.

M: If and when.

P: Okay.

M: Okay, so I have a few questions for you, P: Okay.

M: Uhm, there are ten questions, uhm, as you were in the experimental condition, so uhm, the first question is; how did you feel about the feedback you received on your performance today?

P: Uhm, I think it's very good feedback because I got from the first experiment until the end of [unclear] so I can see how actually [unclear] to myself, M: Mhmm.

P: From the [unclear] I [unclear] the first thing I think I lack more confident but then through the experiment, uh, seems like quite improved.

M: Mhmm, you did.

P: The – all the contact, the recognition, the [unclear] volume, I think it's quite good, yes.

M: Okay. Good, wonderful. Okay so what did you feel you were able to change about how you presented yourself today based on the feedback that you received?

P: Uhm, I think I will actually improve more on confident level of the voice recognition, uh, and the volume as well and especially for the face to [unclear] uh, interview, yes.

M: Oh, okay.

P: I'd like to –

M: Oh, face-to-face, yeah, P: Yeah.

M: We just – yeah, okay. Uhm, so you feel that you'd need to improve on that?

P: Yes.

M: Yeah?

P: I believe that.

M: Okay. So would you – do you think although – but do you think the feedback given to you today will help you present yourself better in the future?

P: Uh, yes. I think, uh, might help me, uh, to prepare myself better in future for my presentations skill and you know, for my interviews skill, because it help me a lot. It can be like, based on the feedback, can boost up my confidence level, yes.

M: Oh, okay. That's wonderful. Okay, uhm, were there any other aspects of the performance that you would have liked to receive feedback on but were not given to you?

P: Uhm, no I don't think so. I think it's been briefly described and the feed base [unclear] or, M:

Okay.

P: For evidence, yes.

M: Okay, no that's wonderful. Uhm, okay so now looking at the feedback given from the bar-charts,

P: Yeah.

M: And how we give that feedback to you, uhm, how did you feel about seeing the feedback with the use of emotional signals and body language during the course of training today?

P: Uh, I think it's good.

M: Okay.

P: Because I mean, uh, the interpretation of, uh, body signal and the emotion, because, uh, previously I had no idea being, uh, I mean like [unclear] the emotion especially when interview so it doesn't know, uh, I mean like the image [unclear] doesn't know what is your emotion level, what the [unclear] research or this feedback, so I know either I'm at the [unclear] so.

M: Mhmm.

P: Uh –

M: Sorry, what was that? You were [unclear] P: I mean the emotion, the level of emotion.

M: Okay.

P: So it can reflect, yeah.

M: Okay.

P: I think it's good.

M: Yeah?

P: Yeah.

M: Did you agree with…

P: Yes, I'm agree, yes.

M: Okay. Wonderful, uhm, so do you feel that the system feedback helped you appreciate aspects of your performance that you might not have noticed just by watching your video?

P: Uh, I believe yes, but uhm, the system, uhm, but it's not 100%.

M: Okay. P: Because uhm, I'm not sure how it works but then by giving a reflect which is the thing that I'm not sure, I think is wonderful because the, you know how your emotion, your body gesture when you present or you speak, M: Okay.

P: For an interview, yes.

M: Okay, but uhm, do you feel that what you've – what we gave you feedback on, P: Yes?

M: You would have – do you think you would be able to get more from watching the video yourself? Or uh, from the… P: Just from the graph?

M: Yeah. P: Uhm, I'm not sure how to say this [unclear] uhm, it might be at a value if you can watch the video as well because at least you know how you see yourself, then you know which – or in which area to improve. But then, uh – by then by, uh, seeing the feedback only, uhm, it's only help you al lots to improve.

M: Oh, with the feedback, P: Yeah.

M: With the system feedback. Okay. That's good. Okay and then, uh, to what extent did you feel the feedback you received was actionable? So the feedback that you were given; how did you think – did you think that you were able to improve on it or was it sometimes feedback that you were just like, "I can't improve on that."? P: Uh, it's very – actually it's very subjective the things that you need, you can or you cannot improve but then in this case, uhm, how you speak, your body language, your emotion, your tone or your eye contact, which is you can improve from time to time.

M: Mhmm.

P: But then, uhm, based on the feedback, uh, I believe myself, I can improve by receiving of feedback. So uhm, yeah, I think it's a good feedback which you can improve. Yes.

M: Yeah?

P: [unclear]

M: On based on – yeah, because yeah, this one was verbal – uhm, non-verbal feedback.

P: Yeah.

M: Okay, uhm, do you think that this could be presented in a clearer way; the non-verbal feedback that was presented today, using the bar-chart?

P: Uh, mmm, I'm not sure – I'm not have any other sort of idea at the moment but then for me, the bar-chart, the one you showed me, uh, it's quite clear.

M: Mhmm.

P: Uhm, because you're comparing yourself at the min value and the maximum value, so you know at what part you are [unclear] I think it's good, yes.

M: Okay, thank you. Okay, uhm, all right, so were there any additional aspects of your non-verbal behaviour that you would like to have received feedback on that you weren't received - given feedback on the system?

P: Uh…

M: With the bar-chart.

P: I think – no, I think all cover, yes.

M: Okay wonderful and then the last question; is uh, do you have any further comments about your experiences or the way the feedback was presented today?

P: Mmm, [unclear] comments for me would be wonderful to be part of this, uh, research. But then, uhm, there is some cases, uh, which interview is not only by sitting.

M: Of course, yes.

P: So sometimes by standing by then so I think that aspect,

M: By…

P: Should be covered because sometime when people during interview in the standing, M: Yeah.

P: So the movement of the whole body language, how they walk, how they stop, how they move around.

M: Mhmm.

P: Uh, it kind of reflects how the confident level, the emotion,

M: Mmm, absolutely. You're very right.

P: Yeah because…

M: You are very right.

P: Sometimes when you sit down, your body gesture or posture, most of might have a perfect but then when you stand, some people might like stand very straight, some like not over confidence,

M: Yeah, you're very right.

P: Not having less confidence so this - bit different way by standing.

M: Yeah, you are very, very right. Thank you so much,

P: Yeah, I think that -

M: For that, that's very right.

P: Yeah.

M: Yeah, okay.

P: So that's all?

M: That's – okay, thank you. All right, thank you very much.

P: Yeah? Thank you. You're welcome, if you need –

P26

M: Thank you for taking part in this study today, and you should know that you can withdraw your participation up until/if we publish. So, I have about 5 questions to ask you today, the first one is how did you feel about the feedback you received on your performance today?

P: well is was good because she pointed out some specific points that I was doing in the interview and that was very helpful, to be improve for the futures.

M: what did you feel you were able to change about how you presented yourself today based on the feedback you received?

P: Uhm I think that I feel, once I listened, heard feedback and I unconsciously tried to improve what I am talking, so that was uh kind of improvement from the feedback M: so, you were able to change it?

P: yeah

M: ok good

M: so, to what extent did you feel that your performance in the areas that Lina identified improved over your interviews?

P: uhm, how do I say….

M: did you feel that what she mentioned you improved on?

P: yeah yeah, I think so but I need more, to practice more slightly I feel that I was better than the first interview

M: ok that's good

P: yeah

M: where there any aspects of your performance that you would have liked feedback on that you did not receive feedback on?

P: no

M: ok then that's all

P: whats the last question again?

M: that one was just, is there anything you wanted extra feedback on

P: no

M: so that was it thank you…

P27

P: So this is not an interview anymore?

M: No.

P: [laughs] Okay. It's just a question… [laughs]

M: That's okay. So thank you for taking part today. And you should know that you can withdraw at any point up until when and if we publish this work.

P: Mhmm.

M: Uhm, okay so I have five questions for you. Please answer as honestly and give me as much information as possible, please.

P: Okay.

M: Okay, so the first question is; how did you feel about the feedback you received on your performance today?

P: Uhm, it was a – what I liked in that feedback is that there are some points that I, let's say that; that terminology part… M: Mhmm.

P: Uhm, I found it important for me to have feedback from a person that is not from my field. So I didn't – I noticed with that feedback I could see what amount of information is filtered and how much they really understand… M: Mhmm. P: And how much wasn't really there. So I found this feedback very interesting actually, and important. Interesting because I mean, I looked at my research a little bit differently; so how am I supposed to do things? How am I supposed to tell things to people so it's understandable?

M: Mmm.

P: That was a very valuable point.

M: Okay. Okay, thank you. Uhm, okay so the next question is; what did you feel you were able to change about how you presented yourself today based on the feedback that you received?

P: Uhm, again well comparing the very first time I was interviewed and the very last time; I finally changed terminology into just a simple language and I found this extremely challenging. Because I felt the need to use that terminology because it – we usually use terminology to save time and to just tell things quicker… M: Absolutely.

P: And then instead of saying that quicker; I actually need to go a long route and explain things; which it was challenging to find those routes first of all. Secondly I thought that uhm, like we were talking too long and you have that little stress going on… but then again within those interviews and training that we had; I gained a little bit more confidence and I knew answers, I had the feeling of; "Oh, I know how to answer this question. I know how to say that."

M: Yeah?

P: So that twist from using terminology to just a proper language; it helped - it wasn't that sharp.

M: Okay. That's good. Okay, so thank you. And the next question is; to what extent do you feel that your performance in the areas identified in the feedback improved over the course of the day?

P: Sorry, can you repeat that?

M: Ja, sure. So basically do you feel that you improved over the day – throughout the day?

P: Definitely.

M: Yeah? Okay.

P: Definitely. First of all just thinking about research from different angles and to be honest; I – there were many little as I call them; candies – in my – in me; that I thought; "Oh, this would be interesting. Oh, this would be useful… This one I want to include…" M: Yeah?

P: But I just had those – a lot of information that isn't interesting to me but I never really referred to those details within the research or through the research. And today I had the chance to – "Oh, this is what I would see in the long run. This is how I would develop something after the research…" So it felt that within this experience, within these interviews and with this training; I got a bit of space of other thinking about research. So yeah, that was a very interesting training.

M: Okay good. So you felt you improved? Yeah?

P: Absolutely.

M: Alright. Wonderful. Uhm, so do you think that based on the feedback you received today; you would be able to uhm, present better in the future?

P: Yes. I believe that I do have still some feedback – my own feedback from me, something that I would still consider changing; but that is probably a very good outcome of the training and all those interviews is that it kept me thinking about things. The interview about how I present my research…
M: Mhmm?

P: To an audience.

M: Okay good. So you feel that you would be able to – it did help you for the future?

P: Yes.

M: Okay. Uhm, so were there any aspects of the performance that you would like to have been given feedback on that you were not given feedback on?

P: Uhm, to be honest well; yes and no. My no part is because the feedback that I received was quiet informative and I found it enough to change my view on certain details that I had. But we also had – I had my own looking at my own performance and I had a chance to give my own feedback on my performance. And all those little gaps that I thought oh it would be interesting to know how things went there, I actually gave my own – I looked at things and evaluated things myself. Which was another – I think very important experience and that was to some extent I valued that part of me giving my own feedback and evaluating my own recordings, a little bit more valuable than the information that was given.

M: Okay. So what do you think that you gained from the playback then – the playback of the video; that you didn't get from the trainers or the…

P: I also looked at the points that I did well. So I think that giving feedback; positive points are also important because when I am doing something which is, let's say a final interview or something; I know that just in case if something; I feel thrilled, I know where to backup, what is okay, what did I do well. So it feels a little – it just builds confidence.

M: Mhmm. P: Uhm, so that was the bit that I thought extremely useful from the recordings. Uhm, so yeah, that's probably all that I could say by now. And every other [unclear] is a little bit more personal research, feedback, which probably wouldn't be – the person who gave me the feedback wouldn't be able to give it because it's my area and I am doing it on my own.

M: Absolutely. Okay so uhm, okay – so you are saying – let me just sum up what you are saying; basically you are saying that what you would like to have received feedback on – a journalist couldn't give you because it's basically about your research?

P: Yes.

M: Yes? Okay. Uhm, and with regards to how you preformed communication wise? Uhm, do you think that was covered? With feedback? You know?

P: Yes.

M: With regards to [unclear] and things like that?

P: Yes.

M: Okay.

P: That was the important bit. And yes. I valued that.

M: Okay.

P: Thank you very much for that.

M: No, thank you. Alright, okay. So that is everything. Uhm, let me just turn this off…

P28

M: So, thank you for taking part in the study today, uhm

P: you are welcome

M: of course, you should know that uhm, you can withdraw at any point, uhm up until and if we publish you can withdraw your participation, ok, uhm, so there just a few questions I have for you today, 10 in total. So, the first one is uh, how did you feel about the feed back you received on your performance today?

P: It was much better than what I thought

M: ok, good

P: Yeah

M: Ok, uhm so what did you feel you were able to change about how you presented yourself based on the feedback you received?

P: MMM, uhm, I think I could improve it, uh

M: so, you thought you did, you improved it

P: Yes

M: ok that's good

P: I tried, I tried because yeah based on what you told me I tried just to quirk something, so it helped me

M: ok, ok and uhm so what did you feel, did you, to what extant did you feel your performance in the areas identified in the feedback improved over the interviews?

P: Sorry what do you mean?

M: So, the feedback you received, do you think in the first interview, do you think you improved in the second and third interview?

P: HMM, I think so

M: Yeah

P: Yeah

M: Yeah based on the feedback cos I mean you had your voice interview, then you had your face interview, video interview, when you looked at the improvements

P: Yeah

M: Where you able to improve, you felt. Did you feel you were able to improve?

P: I think, yeah

M: Yeah? Ok, Ok and uh do you think that this feedback you received would help you present yourself better in the future?

P: Of course

M: Yeah

P: of course, yes

M: That's great, uhm and then were there any aspects of the performance that you would have liked to receive feedback on that you didn't receive feedback on?

P: Not yet

M: Not yet, alright, ok. Ok so you saw the bar chart feedback system uhm, so how did you feel about uhm seeing the syst... seeing that?

P: it is good, yeah, yeah, yeah

M: yeah

P: yeah bec... uhh, I think it was so helpful. The one that it was rating confident and it was changing, yeah?

M: Uh No the bar chart

P: Oh, the bar chart? Yeah it is good, at first it was just make me confused

M: Of course

P: But then after I found out what's happening

M: Ok so possibly, yeah? So eventually you got

P: Yeah

M: Yeah, eventually you understood it and you felt comfortable with it?

P: Of course

M: Ok, ok. Uhm, so did that bar chart help you appreciate aspects of your performance that you might not have noticed by simply watching your video?

P: Hmmmm

M: Do you understand the question?

P: not (I don't understand what participant is saying)

M: Ok that's fine, so the different elements of the bar charts, so you know the facial expressions, the posture, things like that, where you able to get more from the bar chart then watching the video on how to improve?

P: Watching the video that I was talking

M: agreeing

P: My video, no bar chart was less helpful. Is that I've seen myself how I talk or but I don't know which part is a strong or it compare for me, it definitely less helpful

M: yeah? Ok, ok that's great, uhm so

P: Sorry for example, for posture maybe I can't see, that how I am, is it good or not, but then it compares with the (I don't understand participant) so I can just inform that ok it's not correct so

M: yeah definitely, ok great. Ok uh, do you think, uh were there any additional aspects... oops, so sorry. Do you think that the feedback you received was actionable in that the feedback you received you were able to improve on mentally, so you...?

P: Yes

M: yeah, do you see what I mean

P: yeah

M: Ok

P: Yeah, uhm especially when you told me about the smiling and the emotion, even, even I could feel that ok in this second part my smile was not that much firm since that uh comparing to the second part, then you tell me definitely, yeah, I can understand ok what, what I did and so definitely helps to improve it in the future.

M: Ok, that's wonderful, thank you. Uhm do you have any, uh so do you think that the uhm feedback could be presented in a clearer way if you understood that with the actionable, do you think it was fine?

P: It was fine

M: Ok and then uhm, do you have any further comments about your experiences or the way that feedback was presented today?

M: Ooh so sorry

M: Uh, so do you think there are any aspects of the non-verbal behaviour uh that you think would've been useful to receive feedback on that you weren't given feedback on

P: No

M: No

P: No

M: With regards to non-verbal behaviour

P: No, no for non-verbal behaviour it was all fine

M: ok

P: Just for verbal, I the problem that I had I you told me that you have not been passionate hmmm, but the problem that I have in my mind is about hmm I cannot interpretate it to my English language or hmm the, the research that I am doing it, you know what I mean?

M: Yeah, absolutely but there is more you can seem pleased and joyful about, joyful, you don't wanna be like someone's got pain in their legs! Do you see what I mean? There's a different way of saying it, you could come across as uh, a smile could come across as compassionate, do you see what I mean? Do you understand?

P: Yeah M: Yeah? So, there's different ways of doing that, and I mean as well, uhm the way I'm talking now to you, I am expressing some emotion to you but that only because I want you to improve, do you understand? So, I'm feeling passionate about you improving, do you see what I mean? So, there's the difference, I'm not saying excited passion about, you know uhm,

P: No, No, No, No, uh, uh, I just, I don't know why I am I've been upset in the first interview, I'm just talking about first interview

M: Oh

P: when it was about voice but uh

M: I think it could've been related to that question with the diabetes

P: 'Agrees'

M: Yeah, do you, Yeah?

P: No, no, the first interview showed upset for me yeah

M: yes

P: it was about not being passionate in my hmm speaking

M: agrees

P: Ok, uh my problem is not related to maybe, it's not related to hmmm research, but I think that I couldn't show patience because I was thinking behind in my mind, I was thinking about English language

M: Uh, right. Yeah, Yeah of course, no of course, of course, yeah so, its concentration

M: yeah, ok thank you

M: So, Yeah

P: So, I just ask if its like that, it could be like that or there's no relation between English or patience in my speaking

M: I'll speak to you a little bit about it, So I will speak to you a little about this afterwards, ok so I have one more question, do you have any further comments about your experiences or the way feedback was presented today?

P: No, it was so interesting for me

M: ok, ok

P: thank you so much

M: No thank you

P29

M: Thank you for taking part in this study, and you should know that you can with draw you participation at any point up until and if we publish, uhm, if you wish to withdraw your participant I.D is Participant 29. Ok so like said I have a few uh 5 question for you

P: Yup

M: Uh answer them as honestly as you can please. So the first question is, how did you feel about the feedback you received on your performance today?

P:Uh, the feedback was good, uhm, in some areas it'd probably been more helpful to be more detailed M: ok?

P: Yeah

M: Detailed in what sense?

P: uh just there was small things, maybe it would be quite handy to get more feedback, but then generally where, where it said I was quite good

M: ok

P: its hard to sort of be like right M: what was good?

P: yeah

M: yeah

P: uhm

M: What to focus on and keep consistent through?

P: yeah

M: right, ok. Alright so what did you feel you were able to change about how you presented yourself today based on the feedback that you received?

P: I slowed my speech down a bit, I took more time to think, uh and then some of that I don't know whether re- watching and listening and hearing myself say uhm, I don't know whether I'd then sort of think ah shit I need to work on that, but I did try to use less technical terms as well

M: Yeah you did

P: With feedback

M: Yeah

P: Give more examples

M: ok so you felt you were able to change that

P: Yeah, yeah

M: ok, that’s great. Ok so then to what extant do you feel your performance in the areas identified in the feedback improved over the course of the day?

P: I think they improved M: Yeah?

P: Yeah

M: Yeah?

P: in those areas, definitely

M: Yeah? Definitely, yeah. Uhm, so do you feel that based on the feedback you received today you’d present yourself better in the future?

P: Yes

M: Yeah?

P: I think so, yeah

M: Ok, and then where there any aspects of the performance that you would of liked to have received feedback on that were not covered?

P: MMM, maybe my body language

M: Ok

P: ‘Cos I know body language is really more important then actually what you say M: Yeah

P: ‘Laughs’

M: Yeah

P: Yeah

M: ok so that is al my questions, uh Thank you.

P30

M: Thank you for taking part today, you should know you can withdraw your participation at any point up until and when – if and when we publish. Uhm, so I have five questions for you, please answer them as honestly as you can. Uhm, so how did you feel about the feedback you received on your performance today?

P: I was more than [unclear] appreciate to get some any sort of feedback you know, because I have a chance to improve.

M: Mhmm, definitely. Okay, so what did you feel you were able to change about how you presented yourself based on the feedback that you received?

P: I need to go straight to the point when I'm talking.

M: Okay.

P: Uhm, less pauses in the middle of the sentence together with the pauses after saying three words because I do the chop, chop… M: Okay.

P: Yeah.

M: Okay, so that was one of the feedback that you received.

P: I cannot – I should answer the questions more straight.

M: Yeah. Definitely answer the questions, yeah. Uhm, okay so to what extent do you feel your performance in the areas that were identified in the feedback, improved over the course of the day?

P: I would rate myself that it improved just a little.

M: Just a little, okay.

P: That's how I feel about myself.

M: Okay. All right, uh, all right. Uhm, just a little?

P: Yeah.

M: Yeah?

P: I'm still feeling a bit critical about myself.

M: No, of course you would but based on the feedback that was given to you, yeah. P: Yeah, based on the feedback, M: Okay.

P: I improved more,

M: Oh, okay. That's fine.

P: Than I rate myself.

M: Of course. Uhm, okay so what do you feel based on the feedback you received today, you were able to present yourself better in the future?

P: Yes, definitely.

M: Yeah? Okay, good. All right, were there any aspects of the performance that you would like to receive feedback on that you didn't received feedback on?

P: My body movement.

M: Okay, wonderful. Thank you very much and that's all the questions.

P: Cheers.

P31

M: Okay, so thank you for taking part in the study today,

P: That's fine.

M: You can withdraw at any point, P: Yeah.

M: Uhm, so this is just, uhm, an interview asking you for some questions – uhm, asking you some questions to get - you know how – what you thought about today and how we can improve. Okay? So uh, how did you feel about the feedback you've received on the performance today?

P: Uhm, yeah I thought it was quite useful. Uhm, I don't really consider all of those things when I'm doing an interview or well, I don't interview quite often but, M: Yeah.

P: At – when we did the first one, I didn't consider like, how I was really sitting, uhm, how low I was speaking and everything like that, so it definitely was useful to know what makes a good interview and then how well I was performing and it was good, yeah.

M: Yeah, okay good, thank you. Okay, so what did you feel you were able to change based on how you presented yourself today?

P: Uhm,

M: Based on the feedback you received.

P: Okay, so I think all the things that you mentioned or that you fed back to me, were things that I could change.

M: Mhmm.

P: Uhm, some of them I didn't really know I was doing, like the frowning for an example, like I didn't really think about changing it but then when you mentioned like smiling or changing how you're speaking, I think that – directly then changes the frowning itself.

M: Yes.

P: Uhm, so I felt like in the last one, when I kind of – I smiled at one point, then I thought, "Okay, I'm probably looking less grumpy this time." Uhm, yeah so…

M: Okay, thank you. Okay so to what extent did you feel the performance in the areas identified in the feedback improved over the course of the day?

P: Uhm, yeah so I think by probably articulating a little bit better or I would say is so as from – I can't remember which number interview it was but when I – when it was said that I was waffling a little bit, I thought, "Yeah, that's true." And then when we watched it back, I thought that was very unnecessary to say, so I kind of tried to be a bit more concise with my answers.

M: Mhmm.

P: Uhm, and maybe, yeah like change how I was sitting a little bit, maybe try to be a little bit more elaborative on certain points with my voice, for an example.

M: Mhmm, definitely. Okay, thank you. Uhm, so do you think this feedback will help you present yourself better in the future?

P: Uhm, yeah. Probably, I think so. I think I will use this kind of information in even like the training videos of just not necessarily do a lot of interviews but even can be applied to like presenting and stuff like that.

M: Definitely. P: Uhm, so yeah, that's something that I will think about and then in a more [unclear] early on in my PHD's so I'm probably not looking to go to interviews for like jobs or anything at this stage but in maybe a few years' time when I think of going into that stage, I think this kind of thing would have been quite useful for that.

M: Absolutely, yes. Uhm, okay so were there any aspects of the performance that you'd like to receive feedback on that you weren't received feedback on?

P: Uhm, I'm not sure. I don't think so, I think you covered how I was sitting, so maybe like body language, you've covered my facial expressions, my voice, my answers which are the main aspects I think. Uhm, I can't think of anything else.

M: Okay.

P: Yeah?

M: Yeah.

P: Sorry.

M: No, that's okay. Just five more questions to go.

P: Okay.

M: Okay, so how did you feel about seeing the system feedback on your emotional signals and body language during the course of training today?

P: Uhm, it's quite interesting to see, so it's not something that I'm used to seeing, so it's quite interesting to see how I feel like I've spoken or how I feel like I've been during the interview and then to see how exactly seeing how you're frowning. I didn't really realise but, M: Yeah.

P: Yeah, it's quite interesting to see the objective versus my subjective feeling about it.

M: Absolutely, yes. Okay, uh, do you feel the system helped you appreciate aspects of your performance that you might not have noticed whilst watching or hearing the playback?

P: Yeah, definitely. So I think having the system itself, uh, is because it's an objective feedback, it makes me think it's real. So like if you said to me, "Are you frowning?" I would be thinking I'm not but like because the system's telling me I'm frowning, I'm thinking I must be actually frowning, if you know what I mean? Like I'm – not that I'm questioning your judgment,

M: Of course.

P: But that is more subjective from your point of view but because I've got subjective feedback, it's made me think about it for the next one, basically the next interview,

M: Mhmm, yes. That's true. Okay, uh, so did you feel that, uh, the feedback you received was actionable?

P: Uh, yeah. So I think, uh, the – everything that you mentioned was something I could improve on in the next interview, yeah.

M: Okay, so what about the bar-charts? Do you feel that the feedback could be presented in a better way or clearer way?

P: Uhm, I don't know, I think your bar-chart's quite easy to understand so, like for me, I'm obviously used to this kind of environment in a way because I'm doing this kind of research or whatever but someone that's not used to seeing the chart, if you did a different kind of chart, they'd maybe be a bit confused but the bar-chart's quite an easy one to read so I think that's quite a good way of doing it.

M: Okay.

P: And it's quick to get the feedback as well, so it generates it quite quickly which was quite good.

M: Yeah? So okay that's good, thank you for that.

P: Yeah.

M: Okay, uhm, and would you change anything about it? Or…

P: About the feedback or just the…

M: About the – how uhm, the bar-chart was?

P: Uhm, no, I don't think so. I thought it was quite good.

M: Okay. Uhm, so were there any additional aspects of the non-verbal behaviour that you'd like to receive feedback that you didn't receive feedback on? Specifically non-verbal.

P: Uhm, I don't think so.

M: No? Okay.

P: No, I think, uh, you covered – because of any – the few main facial expressions that you need to focus on, there's no point in adding. I thought it was quite good because there's quite a lot of variables but you focused on the main one, so you could have gone through every single one.

M: Yeah.

P: And by the end of it, I would have been like, "Oh, I have no idea what she said." M: Yeah.

P: But you focused on the main ones then I knew exactly what I need to improve on for the next one. M: Mhmm.

P: And then it was more useful going forward whereas if you'd gone, "Okay, so on this [unclear] you'd done this, this…" I would have been like, "I can't remember which one I was supposed to change." M: Yeah, it's just overload.

P: Yeah, exactly.

M: Absolutely. Yeah, okay, uh, so do you have any further comments about your experiences or the way that the feedback was presented today?

P: Uhm, I don't think so. The only thing that was – well, not bad but being at the lecture, M: Yeah?

P: Was the voice over was quite like,

M: Yeah, I needed to keep it without emotion, P: Yeah.

M: Because if I had emotion, then it was – uhm, made you remember more, P: Okay.

M: Things than others.

P: Yeah.

M: But that's good, thank you very much for that feedback, uhm, I'll have a [unclear] –

P: But I think the good thing about the lecture, is that it was only half an hour which isn't too long and there's videos there to break it up which I thought was quite good because if you had just had someone speaking for half an hour, it would have probably got a bit towards the end, M: Yeah.

P: Not boring but like…

M: Well especially the way it was spoken.

P: Yeah, definitely.

M: Definitely, yeah.

P: Uhm, but no, other than that it was good, yeah.

M: Okay, good. All right, wonderful. Thank you.

P32

M: Okay, so thank you for taking part today in this study. Uhm, you should know that you can withdraw your participation if you feel uncomfortable, up until when and if we publish. Okay?

P: Mhmm.

M: And if you wish to withdraw; you participant id is number thirty two. Okay?

P: Okay.

M: Alright, so I just have about five questions for you; the first question is; how did you feel about the feedback you received on your performance today?

P: It was very helpful.

M: Okay, that's good. Okay, so what did you feel you were able to change about how you presented yourself based on the feedback you received?

P: I think I should answer questions – I should think before answer the questions.

M: Okay that's good. Okay, uhm, so is that what you were able to change from the beginning to now?

P: Yeah.

M: Yeah? Okay. So to what extent do you feel that the performance in your areas that were given to you in feedback; improved over the course of the day?

P: Yeah.

M: It did improve?

P: Yes.

M: Okay. So to what extent do you feel that the feedback you received today will help you present yourself better in the future?

P: Uh, sorry…

M: That's okay. So to what extent do you feel that the feedback that you received today, will help you perform better in the future?

P: Uhm, yes, so I think I will think of everything that I heard from this session and then I will like, I will think about everything about my research as well and how I can describe my research next time – introduce my research to like people. Not academic people; like or academic people but in different subjects.

M: Absolutely. Yes.

P: Ja.

M: Okay, so do you feel like the feedback will help you do that? Yes?

P: Yes.

M: Okay good. Okay, so where there any elements of the performance that you would of liked to receive feedback on that you didn't receive feedback on?

P: No.

M: No? You think we covered everything? Do you think our feedback covered everything?

P: Ja.

M: Ja? Alright then.

P: [laughs] M: Okay.

P33

P: Okay.

M: Thank you for taking part in this study, uhm, you should know that at any point, you can withdraw your, uh, participation and, uh, up until if and when we publish, uhm, and that your – if you wish to withdraw, your participant ID is number 33.

P: Okay.

M: Okay? Uhm, so I just have five questions for you,

P: Mhmm.

M: Relating to how you, uhm, performed today. Okay?

P: Mhmm.

M: So the first question is; how did you feel about the feedback you received on your performance today?

P: Uhm, I thought it was really good, M: Okay.

P: Feedback, I felt – I didn't feel like it was delivered in a way that made me question myself, so that was good.

M: Okay, that's good. Okay, uhm, what did you feel you were able to change about how you presented yourself based on the feedback you received today?

P: I think I – I'd like to think that I took the comments on board, it's specifically about the message of my research, I think that and I tried to not do the hair twirling thing.

M: Amazing, okay good. Uhm, okay so do you feel that your performance in the areas are identified in the feedback improved over the course of the day?

P: Uhm, I think I did, I feel like I did. I feel more comfortable towards the end.

M: Yeah?

P: Yeah.

M: Okay, that's good.

P: Yeah.

M: Uhm, and then do you feel that this feedback will help you present yourself better in the future?

P: Absolutely.

M: Yeah?

P: Yeah.

M: Okay, good. Uhm, and then were there any aspects of your performance that you would have liked to have received feedback that were not covered?

P: No.

M: No?

P: No, I don't think, uhm – maybe just on the content.

M: Okay. Anything specific?

P: No, but maybe more like, uhm, if – but it's difficult because it's my research so whoever's going to ask questions, is going to base the questions on what they know of my research.

M: Mhmm.

P: So content – but I guess content wise, in a way it was already addressed but yeah, maybe uhm, more like talk more about that part because that's more important than that part. If you know what I mean. Does that make sense?

M: Okay, no.

P: So like uhm, yeah, like don't talk about the things – I guess, uhm, so maybe she could have said it's better for you to focus on only the things that trigger the anxiety, M: Okay.

P: Rather than uhm, the overall findings so – M: Oh, I see.

P: Yeah. So maybe something like that, not to say that that was what it should have but I think content wise, that would be good to kind of know which is more important.

M: Okay, thank you Jess.

P: Okay.

P34

M: Sorry, thank you for taking part in this study and if you feel that you want to withdraw; your participant number is thirty four. And if you want to withdraw you can uhm, withdraw up until when and if we publish. Uhm, okay, so I have ten questions for you today. So how did you feel about the feedback you received on your performance today?

P: I thought it was very uhm, specific, which was very useful. I knew exactly in terms of content and also kind of the specific kind of competencies and a way on what to improve and how to improve them as well.

M: Okay.

P: Yeah. So it was very useful.

M: So competencies regarding…?

P: Things like I guess uhm, maybe perhaps factors that will influence how I come across; like posture and…

M: Oh okay.

P: Ja.

M: Okay, wonderful. Okay, so what did you feel you were able to change about how you presented yourself today?

P: Definitely posture. Definitely things like uhm, I think I am much more aware of my tone of voice and…

M: Okay.

P: …kind of variety of it as well which I know I still struggle with.

M: Okay.

P: But it's something I bear in mind moving forward which is useful. Uhm, also not too often to get to the point… M: Okay.

P: Ja. That is quiet useful.

M: Okay good. That's good news, thank you for that. Okay, so to what extent did you feel your performance was identified – that was identified in the feedback improved over the training?

P: I remember that initially uhm, when answering a question; I couldn't even remember the question anymore – I kind of went off track. I didn't talk much about the project itself. And towards the last interview I felt like I am answering the question, I am giving precise answers and I am talking about what I actually should be talking about.

M: Okay, that's good. Alright, good. Uhm, okay so how did you feel this feedback will help you present yourself better in the future?

P: I think I give much more thought to how I am being interviewed and take things accordingly to that.

M: Okay.

P: I have been much more conscious about what I need to say. So my key messages and what are my key messages in terms of the project as well. And I think this will be quiet helpful to answer questions and to bring them back to what I wanted to say and convey, kind of the relevant information as well about the study itself.

M: Okay. Okay, and with regards to the non-verbal feedback that was given to you?

P: That was very useful because you never get feedback on that at all.

M: Yeah.

P: Uhm, so things like – you quiet unconscious about them as well; like tone of voice – I never really thought about that.

M: Yeah.

P: Uhm, or even things like how much I move and my hands or my body. So that was quiet useful to know that this will also be taken into account.

M: Ja, of course. That's good. Uhm, okay so were there any elements that you would of liked to receive feedback on that you didn't receive feedback on?

P: Uhm, I can't think of any.

M: No?

P: No. Uhm… I don't know. I guess, but this is kind of tricky – maybe just to know what a good presentation would look like. I know I saw some examples but uhm, I don't know, maybe just knowing how to move your hands or in what way…

M: There was the range and focusing for the average… P: Mhmm.

M: I mean, that was – with in regards to what way; you have got to use it within what you are saying, in context to you are saying. That's it.

P: Ja. But no, I can't really think of anything.

M: That was if you had to say something… P: Ja.

M: Okay, wonderful. Aright thank you. And then, so how did you feel – you know the bar chart… P: Mhmm.

M: Okay, how did you feel about the system feedback on – about the uhm, the emotions that you – and the body language? Did you feel that the bar chart was useful?

P: Yeah.

M: Yeah?

P: Definitely.

M: Okay, so uhm, did it help you throughout the course of the day?

P: Mhmm.

M: Yeah? Okay good. Do you feel the system helped you appreciate aspects of the performance that you might not have noticed just by watching or hearing playback?

P: Definitely.

M: Yeah? I know you did say that a bit earlier. Okay, and to what extent did you feel that the feedback you received was actionable?

P: Uhm, yeah I definitely thought because the feedback was so specific; I definitely knew what I needed to improve. So it was very actionable to me. I think it's a useful thing about those criteria's that were things that I would [unclear] my control.

M: Yeah.

P: Uhm, ja.

M: Ja, that's the good point of the – okay. And then uhm, okay so to what extent – okay yes. Uhm, so okay, do you think that anything could be presented to you clearer? In clearer ways?

P: How do you mean? In terms of the feedback?

M: Yeah.

P: No. [unclear]

M: Okay, so you understood what they were?

P: Yeah.

M: Okay good. Okay, and then were there any additional aspects of the non-verbal behaviour that would have been useful to receive feedback that wasn't?

P: No, I thought the feedback was really useful because she would give me feedback on the content as well like if in questions you diverted. So that was very good.

M: Okay good. Thank you. Uhm, I have one more… P: Mhmm?

M: Do you have any further comments or experiences that you would like to share about today?

P: No, I thought it was really useful.

M: Okay good.

P: Where we can use them moving forward as well. Ja.

M: Ja? Okay good. Thank you.

P35

M: Okay, so thank you very much for taking part today.

P: It's a pleasure.

M: I really appreciate it. Uhm, and if you want to withdraw your participation at any point; your participation id number is thirty five.

P: Mhmm.

M: Uhm, you can withdraw up until when and if we publish. Okay? Fingers crossed. [laughs] So I have five questions for you now, so the first question is how did you feel about the feedback you received on your performance today?

P: Ja, I feel that it is useful.

M: Okay.

P: Ja, I can use it and try to improve.

M: Yeah?

P: Yeah.

M: Okay good.

P: To change a little bit. Ja.

M: Okay, alright. And then what did you feel you were able to change about how you presented yourself today based on the feedback that you received?

P: First the gesture and how I move my hands and you know, exactly I'm going to the point that the interviewer asked me. Not you know, just confusing – so ja.

M: What do you mean? I'm so sorry.

P: So I mean when they ask me some questions; so I try to you know, give an example to make it simplified and not trying to complicate it or just exactly answer the question that she wants. You know? Sometimes I say more or something… [laughs]

M: Oh ja, okay. [laughs] Okay. And then, so do you think that your performance improved in the areas that were identified?

P: Yes.

M: Throughout the training? Yeah?

P: Yes, of course. Ja.

M: Okay good. And do you think you would be able to use this information to present yourself better in the future?

P: Of Couse. I will.

M: Ja? Okay good. And then were there any aspects of the performance that you would like to have received feedback on that you didn't receive feedback on?

P: Uhm, no. But I would be happy that I will have all the - you know, just [unclear] about if it's possible… M: What do you mean?

P: About my weakness.

M: Oh, so you want me to summaries it for you? Or sorry, I don't understand.

P: Not now, later. Maybe in the – ja.

M: Okay, I see what you mean. Definitely. Absolutely.

P: Ja.

M: So you don't think anything else? Okay.

P: Ja, I think everything is okay.

M: Ja? Okay. And then uhm, so alright, so you received everything you – all the feedback. Okay. Alright, that's everything. Thank you so much.

P: Thank you.

P36

M: So, thank you so much for taking the study today, I really, really appreciate it, uhm, ok so uh if you want to withdraw your participation in this study, for any chance uhm but for any reason uhm your participant number is number 36 ok, and uh you can withdraw up until when and if we publish, ok? Ok, uhm so I have just f. about 10 questions for you today, uhm if you can answer them as honest as you can please. Ok so the first question is, how did you feel about the feedback you received t. on your performance today?

P: They were quite constructive because they helped me improved

M: Yeah

P: Yes, Yes

M: Yeah

M: Ok, good

P: Yeah, it was because they were like uhm, they were reflective of my performance M: Yeah

P: so uhm all I had to do was that I had (unsure) the key things to watch out for and I don't even forget, so (unsure) to the next stage

M: Ok, fantastic. That's good, thank you

P: Yeah

M: Ok, so what did you feel you were able to change about how you presented yourself based on the feedback you received?

P: Ok so for, I sat my posture, I sat up like, like I like (unsure)

M: Yeah

P: and then secondly, I relaxed cos I smiled more uhm like on that (unsure) so I relaxed, smiled more you know, and then finally I was able to use like my hands to demonstrate more than I did before M: Absolutely.

P: Yeah, so that, that

M: Yeah, ok good

P: and of course, my speech, I tried wish like I wasn't watching my speech

M: uh ha, yes definitely

P: Yes

M: Yeah definitely, good

M: ok uh so to what extent did you feel you pref... your performance improved over the course of the day?

P: oh um, that like uhm compared to the first one uh I'll give it uhm a 95 P: Laughs

M: Laughs, ok, ok, that's good. Ok good, thank you, Yeah? Ok good

M: Uhm so would you think this feedback will help you present yourself better in the future?

P: Yes

M: yeah

P: Definitely, yes.

M: alright

P: Because you know the stage fright is taken away, uhm the, like I said when the attention is no longer there

M: Yeah

P: The anxiety, uhm I'm not, I feel more confident

M: Well good

P: Yeah

M: Yeah

P: and I know the key things to watch out for, you know? Like my posture in a real-life situation, yeah and no, I need to like mush, I'm not mushing my words, you know I didn't tell you I've done 3 minutes presentation before

M: OK

M: (not sure)

P: Yes, I've read.. what I've said, what I had to say was possibility for 3 minutes M: Yes

P: I say its about 2 minutes

M: 2 minutes, oh

P: woah, woah, woah, woah

M: oh

P: as soon as I finished they took, they uhm (unsure) called and said do you talk fast naturally as it is? It's only if I'm under.. when I'm tense. Laughs out loud

M: Aaw

P: So, you can see what this did to me today M: Yeah but maybe you can do it again?

P: Yes precisely

M: Based on this thing

P: Exactly, yes

M: Yeah

P: Yes

M: Yeah, Yeah. Good

P: I enjoyed it actually

M: Yeah, yeah, good, good, Uhm ok so, were the any aspects of the performance that you would have liked to received feedback on that did you didn't receive feedback on?

P: I think I received feedback on, on majority because data was, was similar to what they have given me the first feedback

M: Agrees

P: And of course because I have improved we now, obviously we understand that actually I have improved. Yes, so I actually (unsure) received feedback on yeah, (unsure) most of that you know they say the gonna send it me

M: Oh yes, of course

P: and it makes me really happy.

M: Ok good, alright. So you can watch it and reflect

P: Exactly

M: Of course, of course. Ok, so uh you know the bar chart that you saw your feedback, with your (unsure recording skipped) and the how maximum and minimum

P: agrees

M: Yeah, so how did you feel about seeing the feedback on your emotional signals uhm, how did you feel about that system

P: ok, yeah, you mean the first time you should (recording skipped and not sure of what participant said)

M: ok but more, more the feedback, so h..how did you understand it?

P: yes, so uh uhm, actually its really un-effective for me, I mean it's a measure, whatever you know platform, because its actually caption

M: Ok

P: And uhm like those seem be asking (unsure) yes actually I don't really like in uhm anxiety, you know, yes so

M: Yeah

P: its actually true, its real what it's saying what I mean, you know

M: Yeah

P: and the I now seem to improve, so when you gave me the second feedback , I was so happy because I could see improvement in me

M: ok, that's good, ok but did you understand the bar chart

P: Yes

M: how I did

P: Yes, yes

M: how it presented?

P: you took, like what, what each of the columns means

M: ok good

P: and made like this, you know like a range unexpected to like perform of 2 M: yes

P: what's acceptable yeah (unsure) so I totally understood

M: Ok good

P: I understood what the chart said

M: Ok. Oh that's good, thank you. Uh So, how did you feel about that?

P: Oh, great. Good, good

M: Good, good

P: Unsure of what participants is saying

M: ok good, Uhm do you feel the syst.. the bar chart feedback helped you appreciate aspects of the performance that you might not have noticed by watching only

P: Yes, lots, yes a good deal

M: ok

P: A good deal cos watching, you know, is notice enough

M: Agrees

P: but this one told me that you are not too happy, so that actually made me work on myself M:

Absolutely.

P: Yes

M: Yeah, yeah, yeah

P: and that I.. I improved more, so yes, it actually said a lot, a lot that just watching myself actually told me. So, I really like the system, I like the system where you know you've got bar charts that tells you lots, it has to watch out for and to see if you've improved this because it will tell you if you have improved or not (unsure of what participant said) M: Ok, good, thank you.

P: You're welcome

M: Uhm… Do you feel the feedback that was given to you was actionable?

P: Yes

M: Yeah

P: It was

M: Ok. Do you think anything uh, could've been made clearer?

P: I'm not too sure, because in the first place I didn't even know how this ones existed, laughs out loud

M: Laughs, that's fair enough

P: Honestly

M: that's fair enough P: So honestly I.. I know, honestly its really (unsure) I don't even know how (unsure), Honestly I learnt a lot, this (unsure of what participant is saying) I learnt (unsure of what participant is saying), it teaches you (unsure of what participant is saying) I learnt a lot of things (unsure of what participant is saying) and it has true you've helped me

M: agrees

P: Yes, because you know, its one thing to try and fake a smile even when you are un.. anxious or you are under tension

M: Agrees

P: (unsure of what participant is saying)

M: Yeah massively, yeah

P: so.. so this one checks a lot of things inside (unsure) which is a lot of things

M: Yeah, yeah.. it.. yeah definitely, I see exactly what you are saying

P: and that, you know even uhm, just looking at the interview can tell you, you can see me smile. Imagine if im faking it M: Absolutely.

P: even if im struggling or this one (unsure) exactly what is happening

M: Yes

P: (Unsure of what participant is saying) anxiety level (unsure) Flow naturally

M: Uh ha exactly, definitely

P: Yeah, now talking about first year I was like, how did you know my first year (unsure of what participant is saying)

M: Pay attention

P: (Unsure of what participant is saying) M: Really?

P: Cos you know if I sit and write, you are apply too much from your body then (unsure) uncomfortable but now I had to sit up right (unsure)

M: It made you change your view on it

P: Exactly

M: Fantastic

P: yes

M: Oh that's amazing

P: And I feel really relaxed

M: Good, that's what.. yes, that's why you performed better, you know

P: Yeah

M: Uhm, ok, so where there any asp.. additional aspects of the non-verbal behaviour that you would like to hear feedback on that you didn't receive feedback on?

P: I don't think so, like said (unsure) I don't even know what there was (unsure) or there should be, all that you've told me is perfect

M: and based on the lectures measuring all the

P: oh yes, yes M: is fine?

P: Yes

M: no im mean because in the lecture we measure.. we told you some non verbal behaviour

P: Yes

M: uh so it mimicked that

P: Yes

M: is that right?

P: Yes

M: ok

M: ok, do you have any uh other further comments or experiences or anything mentioned about how things were.. feedback was presented today?

P: No, I think everything was nice M: Yeah?

P: it was nice

M: Perfect

P: it was nice

M: that's good

P: it was nice, as I can see

M: Yeah

P: Yeah

M: Good

P: I feel (unsure) jumping out there (unsure)

M: oh that's amazing, ok. Thank you

P: You're welcome

P37

M: Okay, so I have uhm, thank you for taking part today. Thank you very much for uhm, your full engagement…

P: Sorry for leaving you behind today.

M: Bless you. [laughs]

P: [laughs]

M: You should know that you can withdraw your participation at any point up until when and if we publish, and if you do wish to withdraw; your participant id is number 37. Okay?

P: When you say publish, you're not going to use my name or anything like that?

M: Absolutely not. No.

P: So in the videos I reckon its uhm, it's linked to analysing itself?

M: Yes, absolutely.

P: You know, because I don't want to be you know, to be based in it. I don't want to see myself in YouTube.

M: No, definitely not. [laughs] Okay, so I have ten questions for you.

P: Aright.

M: The first one is; how did you feel about the feedback you received on your performance today?

P: Uhm, overall I am quite satisfied. Uhm, I could understand some of the feedback and uhm, some improvements I had to make but otherwise still like, uhm, I am not saying that you have to have a leading question but if it's something you expect then those are the questions that should be addressed. Uhm, so [unclear]. So its aims and objectives – well we did focus on it twice and it's like sometimes you know, if it was asked ahead of time, or during the beginning of the interview than definitely talk on it and once we have learned something about [unclear]… M: Mhmm?

P: I'm sure like that if that aim and objective had been asked earlier and it had been repeated a few times during the interview.

M: Okay.

P: Even as an interviewee, you know, I will find it more comfortable speaking about it again and again.

M: Mhmm.

P: Especially after going through the lecture.

M: Mhmm. Okay.

P: So, organising the questions was – I think could have been a little bit… M: Better?

P: …better. Yes.

M: Definitely. Okay, so what about your performance – the feedback that you received?

P: It was good.

M: Yeah?

P: I'm quiet satisfied.

M: Okay good.

P: And it did makes sense of uhm, you know the energy level and frowning and you know… M: Mhmm.

P: Because of the other factors that was affecting it.

M: Of course. No of course. Definitely. Okay, thank you. So what did you feel you were able to change about how you presented yourself based on the feedback that you received?

P: Uhm, I think one of the improvements I made today would be talking a bit slow.

M: Mhmm?

P: Because usually I will just go on my own speed and because [unclear] someone is now interviewing me and often I don't take a lot of feedbacks on board either because you don't learn the things you don't want to learn. So… [laughs] M: Yeah.

P: I would have ignored a lot of things but I think talking slow; that was a factor I had addressed and posture as well and smiling. And often what I have felt was that uhm, because it's more of a formal interview I will try and minimize the use of my hands – expressing with my hands and stuff… M: Okay.

P: So to begin with I thought that's [unclear] but then again when I realised like I should be normal and use the hands as often as I did… M: Yes.

P: And I guess the hand movement and stuff had to [unclear]…

M: Absolutely. And as a [unclear] having a [unclear] effect. Okay that's good. Thank you very much. Uhm, so did you feel your performance in the areas identified in the feedback improved over the course of the training?

P: Confidence did definitely go up.

M: Okay.

P: And with the feedback I think I was able to be more informative and you know, comply more with the activity that is going on here; that's being a part of the interview… M: Okay.

P: And uhm…

M: Okay that's really good. Thank you.

P: That was it? Two things…

M: Ja? Okay wonderful. Thank you. Okay, so do you think this will help you present – the feedback you received today; will help you present yourself better in the future?

P: Now that's a bit of a – the word future is quiet you know… [unclear].

M: Oh I mean for other interviews perhaps?

P: If I had to do another interview in a short space of time, say if I was involved in the same, next one week or two weeks – all this information would be fresh and it's like practice makes it more convenient and more I would say "perfect". Especially in this field of interview, getting an interview, it would make me more confident and comfortable. So in the next two or three weeks I will be able to give a good interview in this selected topic again… M: Mhmm?

P: Not if someone talks to me about some financial world, I would get lost again and… M: [laughs]

P: But ja, in the next two or three weeks, talking about the same topic and the same line of research; I will be more comfortable talking about it.

M: Ja, so you would use the feedback that you got today?

P: Absolutely.

M: Okay great. Thank you. So were there any aspects of your performance that you would have liked to receive feedback on that weren't covered?

P: I believe we went through you know, using technology to review your feedback is not something I would of thought about but from going through the video tape and you know, looking at myself and the posture; which I felt is quiet important; mentioning eye contact, uhm, because of the glasses; I can't do it very often.

M: Of Couse ja.

P: And uhm, but uhm, overall I was very satisfied. I would say the use of the different aspects we looked at [unclear], expressions, the use of hands, you know, if you can actually number it then it's like tactically analysing after a game and seeing where we went wrong and how we should improve it. So, I don't think at the moment with the technology that we have, that we could have done it better. So I am quite happy with that.

M: Okay good. Thank you. Uhm so how do you feel about, you know the bar charts that was given feedback with the actual performance, the range and the average… Uhm, so how do you feel about seeing that system on your use of emotional signals and body language over the course of the training session?

P: Well with bar charts; I think it was quiet confusing.

M: Okay.

P: Uhm, it was confusing in a sense that I wasn't used to – I didn't expect that to begin with. And when I saw it maybe I didn't process it well either.

M: Okay.

P: And uhm, coming to think about it, I guess that's how I would of – if it was – and it's a field that I wasn't aware of, about saying you know; conveying your interviews into a bar chart… M: Mhmm?

P: [unclear] emotions and stuff; I would of expected that on financial field or something to do with numbers. So I got confused with that to begin with.

M: Okay.

P: And even at the end of when we looked at it again; I wasn't sure exactly what it was. But then again you know, you can't know everything so I was like; "Okay, she probably know better than me."

M: No, but I mean the whole point is me explaining it to you and then eventually you would be able to – I mean you always will have someone using this. You will always have someone explaining it.

P: Uh, with the bar chart I would say, I will be honest with you… M: No, please do.

P: That's one thing I wasn't very clear about.

M: Okay. That's fine. Okay so, alright. Do you feel though that the bar chart would have given you something that you appreciated or that you would gain more rather than simply watching the video? P: Uhm, that's a mixed area and that's a bit of a grey area. The reason I would say is like I am so mixed and I didn't understand what exactly you were reflecting on asking that. Like this is the average, this is where I stood and…

M: The minimum and the maximum… ja.

P: And maybe it's the frequency as well; that we didn't use it that didn't allow me to completely understand compared to the other analysis you did use today.

M: Mhmm.

P: I would think – I would say if I was to you know, rank them; then bar chart would come at the bottom of the list.

M: Absolutely.

P: This is something uhm, I didn't kind of understand or perhaps I didn't make enough effort to understand it again. Mind you, the [unclear] level was low when we started the interview.

M: Ja. [laughs] But do you think that when I played back your video and we had a look at the emotions as you were watching…?

P: Oh yeah.

M: That was a better one?

P: That was a good one.

M: Ja?

P: That was a good one. Uhm, even though the – it was a good one. That kind of helped me understand and you know… M: Okay, good.

P: Because it's a [unclear] learning, it's not graphic and you know. But overall I would say it was good.

M: Okay. Alright, thank you very much. So to what extent did you feel though that the feedback you received was actionable?

P: Uhm…

M: So if I said to you just to improve your posture or something like that; were you able to act on those?

P: I was. Yeah. And I believe that I did take them aboard and I did make the changes necessary.

M: Mhmm?

P: And uhm, yeah, posture, smiling, use of hand gestures and stuff… I think they were quiet important tips.

M: Okay, alright, that's good. Uhm, so do you think the feedback could have been presented in a clearer way?

P: The feedback process was good. It was like, you go through a chart; what has been you know, what has been analysed by the machine and you interpreted it well. So it's on a one to one bases. I don't think it could have been any better. I think it's uh, I would say it was good.

M: Okay. Thank you. So would you change anything about the bar chart?

P: About the…?

M: About the bar chart or the way that it's been given feedback to you.

P: Uhm, I think with the initial ones; perhaps there is a need to spend that one extra minute to go through that again.

M: Okay.

P: And uhm, to make sure that the interviewee you know, the person who is being interviewed; he actually does get it.

M: Okay.

P: You know, lead him to point certain things out from the chart.

M: Okay.

P: One of the examples I will give you is like; I am very bad with names and stuff.

M: Okay.

P: So what I like to do is like, if I meet somebody again I say like; "Hang on…there you go…" and I will open my contacts, the address book and I will say; "Won't you pop in your name and number…"
M: Oh that's smart. I see.

P: "I will get in touch with you…" And stuff like that.

M: Okay.

P: Or if I just want to get someone's number then I will be like; "Yeah and your mobile number is 07…?"

M: Yeah.

P: [unclear] even with that one, you know, you would explain maybe one or two and on the third you would expect them to [unclear] that they do understand.

M: Absolutely. Yeah okay.

P: [laughs]

M: No, that's good. Thank you very much for that. Okay, so where there any additional aspects of nonverbal behaviour that you would of found useful in your feedback?

P: Uh, nonverbal? No. I think like to the extent needed for an interview, I think I got all the feedback needed. And of that [unclear].

M: Okay good. Thank you. Okay so is there any other comments or experiences that you want to share with me?

P: On what?

M: Based on the feedback?

P: Feedbacks were pretty good.

M: Okay.

P: I'm not going to lie about that. The only thing; like I said; you know – a few of the things were probably, you didn't encounter before – like someone with glasses and the distance… M: Limitations. Of course. Ja.

M: Mhmm. Definitely. Okay, thank you so much for that. Anything else?

P: Uh, that's it.

M: Okay.

P: I just felt [unclear] distraction. But I don't know, for me – I don't know. I didn't want to get distracted away from the topic, and taking the glasses off. I was like…

M: No definitely. But that's limitations of the technology.

P: Yeah.

M: Thank you.

P38

M: Okay. Thank you for taking part in this study. Uhm, you should know that you can withdraw your participation. You can withdraw up until when and if we publish. And if you do wish to withdraw your participant id is number thirty eight. Okay? Uhm, so I have five questions for you. Okay, the first one is; how did you feel about the feedback you received on your performance today?

P: Ja, fine.

M: Ja? Okay. So what did you feel you were able to change about how you presented yourself based on the feedback you received today?

P: Uhm, it was the smiling… M: Yeah?

P: Uhm, and then trying to not go off on tangents but keep – keep it tight.

M: Okay alright. So to what extent do you feel your performance in the areas identified in the feedback improved over the training?

P: Yes I do feel it improved. Sorry, was the question – what's the…

M: No its fine. I was just wondering; you know, the feedback that you were given… P: Ja?

M: Do you feel that it improved throughout the training? As the training went on?

P: Ja.

M: Yeah?

P: Ja.

M: Okay. Uhm, okay so to what extent do you feel that the feedback you received will help you present yourself better in the future?

P: Well it will be for example; knowing that if I am not having a camera interview; to look directly at the camera and not worry about looking at the interviewer.

M: Ja.

P: Uhm, and then feeling a bit more confident about my research because the feedback that I had is really positive; about the content.

M: Okay, that's good. Okay, and how you present yourself?

P: Put makeup on… [laughs] M: Oh no! [laughs] No?

P: Mhmm.

M: Okay. It's because you had to watch it back… [laughs] P: [laughs] Ja.

M: [laughs] Okay, uh, so were there any aspects of the performance that you would of liked to have received feedback on that were not covered?

P: No. I think she was quiet positive. It could have been more negative.

M: Okay. Alright then. Okay, thank you very much.

P39

M: So, thank you very much for taking part in this study. Uhm… P: This is a video recording or audio?

M: Just record – audio recording.

P: [unclear]

M: Uh, thank you very much for taking part in this study, uhm, the main idea of this is, uhm, just to let you know that if you wish to withdraw your participation, you can withdraw up until when and if we publish,

P: [unclear]

M: And, uhm, the – and if you wish to withdraw your participant ID [unclear] okay?

P: Okay.

M: Okay. Uhm, one moment while I get the… okay, so I have ten questions for you: uh, the first question is; how do you feel about the feedback you received on your performance today?

P: I found it interesting, uhm, because it made me aware of things I didn't think about, especially while speaking I'm still finding it hard to be aware of, uhm, my posture, the volume of my voice, M: Yeah.

P: When I speak, there's a concentrate on the answer or the question and the answer, I don't think too much about the, uh, body language or the, uh, the features of my, M: Mhmm.

P: Voice for an example, my facial expression. So it was an interesting to see stuff I don't automatically recognise myself.

M: Mmm, that's good. Okay, all right. So then, uhm, what did you feel you were able to change how you presented yourself today, based on the feedback that you received?

P: Well, I only know what I have, uh, changed every individual element that were supposed to but I'm definitely more aware of, uh, having to control the volume of my voice more carefully. Right now as I'm speaking, I am trying to control it consciously for once and, uhm, I hope – I think this over time may actually improve my performance. You know, this – once single exposure may not be enough but, uh, over time, I think I could make an effort and build on these things to have some [unclear] effect.

M: Mhmm, okay that's good. So that leads me nicely into my next question; so do you – based on the feedback that you received today, to what extent do you feel it will help you present yourself better, P: Well,

M: In the future?

P: There has been some improvement I think, as I was aware of things that had to change.

M: Mhmm.

P: But I wouldn't say a dramatic improvement.

M: Okay.

P: Uhm, I think it would be interesting for someone to try this system out over an extended period of time.

M: Okay.

P: Because maybe that, uhm – the improvement you won't see needs some time to be implemented if you know what I mean.

M: Mhmm.

P: It may now be enough to have once in the session to actually see significant improvement, maybe [unclear]

M: Yeah, well okay. But do you think the feedback you received today will help you present yourself better in the future?

P: I think it's very – M: Yeah.

P: Can be very useful if sustained over time.

M: Okay, thank you. That's good, thank you. Uhm, were there any aspects of the performance that you would have liked to receive feedback on that you didn't receive feedback on? P: No, they already covered already everything, M: Okay.

P: Uh, they covered my, uh, voice, [unclear] the, uh, facial expression. Uh, I was a bit confused about the feedback given by the badges.

M: Okay.

P: Because I couldn't really, uhm – well, I'd like to know for an example, uh, the range of variability for the posture.

M: Mhmm.

P: I would like to see an example; a good posture, a bad posture, M: Oh, like physical pictures?

P: [unclear] posture – exactly.

M: Right.

P: Because it's hard for me to – I mean, all I have in mind is a certain, uh, well straight,

M: Mmm.

P: As opposed to be a slouch, I can't really, uhm, picture anything else and I think it's my problem because I feel the signals are detected by the [unclear] than I actually think about, M: Self-aware, mmm.

P: I may be losing something, I may be missing something that is in the data, M: Mmm.

P: But I'm just not now able to mark that to a physical [unclear] M: Yeah.

P: My body.

M: Yeah, well we saw that as you improved, there was some improvement in your posture, so as you straightened up, there was some improvement in your posture.

P: [unclear]

M: But then it was not natural what you were doing.

P: Right, exactly.

M: Yeah.

P: You see, not only that because it's just – it may not be what I naturally do,

M: Mhmm.

P: But if you improve things, so be it.

M: Yeah.

P: The thing is that if you take it to the extreme, then the person won't be comfortable anymore. They will sit rigidly and [unclear] be counterproductive,

M: Exactly because that will have a domino effect on other things, absolutely.

P: So – exactly. So uhm…

M: Well that's where you've got to look at what's most important. P: Yes.

M: Yeah.

P: That's [unclear]

M: And work on individual things at a time, definitely.

P: But I found the voice and the facial expression more useful.

M: Yeah?

P: [unclear] yes.

M: Interesting. Okay, thank you. Okay so, uh, how did you feel about the bar-chart or the system feedback on the use of social signals and, uh, body language during the course of the training session?

P: Well I found it interesting. Uh, it was a bit, uhm – it wasn't straight forward to understand, M: Of course.

P: All of it, but with your explanations, it was absolutely clear again.

M: Oh, okay. That's great, thank you. Do you feel the system helped you appreciate aspects of your performance that you might not have known, just by watching and hearing your playback?

P: Absolutely.

M: Yeah?

P: Because I'm never aware of these things when I speak and, uh, this has been a way for me to actually see, uh, some objective data about the way I speak than I'm completely unaware of otherwise.

M: Okay, good. Uhm, and so to what extent did you feel the feedback was actionable?

P: Actionable; well that's time dependant question and time dependant answer. M: Okay.

P: There's – I don't think it's very much actionable in the short [unclear] I wouldn't be – I don't think I would be able to implement all the changes within one session. But if you asked me to use this system over a month for example, one hour a day, M: Mmm.

P: And then try to improve after a month, I think I would be able to do something useful there.

M: Okay but, uhm, based in – on the feedback that was highlighted in the bar-chart, were you – do you think you were able to improve your performance?

P: I hope, I tried to – I hope it was improved a bit but I think the potential this technology goes well beyond what you can do within one single session. See, if you extend the exposure and the user will [unclear] time, I think you may be able to see a bigger improvement.

M: Mhmm, and then do you feel that the feedback could be presented in a clearer way or, uhm – yeah?

P: It's a difficult one because there aren't many, M: Options. P: Ways I can think of to make it clearer because you are, uhm, you need, as you've mentioned, you need to know normalise to be able to show, uh, the variability across participants in a meaningful way. And of course when you normalise that, then you have to explain, M: Yeah.

P: How you've done it and it may not be the simplest thing of all to understand this for everyone, M: Yeah.

P: Uh, so there are limitations of course. Uhm, but I think the, uh – it's good to have a comparison between, uhm, what the good interview looks like in terms of your distant features and what mine looks like, M: Right.

P: Within the range acceptable, M: Right.

P: Acceptable ranges. One thing I would like – would be interesting to see, would be – you see, this gives you an integrated view over the whole interview – across the whole interview.

M: Uh, with the first 30 seconds P: Uh, pardon? Okay.

M: Yeah, the first 30 seconds.

P: All right, but it would be nice to see the first 30 seconds, "Oh, okay. It's good." But it would be nice to see some – how those signals changed over time, for an example.

M: Oh absolutely, yeah.

P: For instance, uh, if I have, uhm, a high posture signal; is that because I kept a certain posture throughout those 30 seconds or did I change suddenly and – you see, is that signal an artefact of this [unclear] event? Was that something sustained consistently over those 30 seconds?

M: Mmm. I'll have to have a look at that, that's very interesting. Thank you. P: Okay.

M: And, uh, so were there any additional aspects of the non-verbal behaviour that you think you would have found useful?

P: I wouldn't know honestly because I can't say no because I'm sure there will be something that's useful, I'm just simply not thinking about now. Uh, in terms of the sense though, were used, uh, tonight, the only thing I'm still confused about is the way the, uh, [unclear] badges [unclear] the way they allow you to abstract information of the [unclear] about how often I mirrored the, uh,

M: Ah, you can change the…

P: [unclear]

M: You can change it to 20 seconds, 10 seconds, you can change how it – basically it uses a [unclear] P: It escapes me a bit because you can face the other person but then mirroring is also the other things, it's also about, uhm, how I use my hands in a similar way as you and that will be captured by another sensor, which is the, uh, accelerometer

M: But the accelerometer doesn't account for the other – it bounce off, uhm, the [unclear] P: I know, sorry to end,

M: It's okay.

P: The correlation can only be done with the analysis [unclear]

M: Absolutely yeah, which the badges do as it's extracts the data or exporting the data… P: You see, mirroring also means that I smile when you smile.

M: Mhmm.

P: And this is not captured,

M: No, of course it isn't.

P: Unless you do it [unclear] analysis.

M: Of course it isn't but that's why it's the non-verbal and, P: Sure.

M: That's why we're looking at the, uh, facial expressions but thank you. P: No problem.

M: No good, thank you. Uhm, so – okay, I see what you're saying; so there were no – so there were none of that, okay so do you have any further comments or experiences that you want to share with me?

P: About what?

M: About today? About the feedback.

P: Well you know, apart from this being a very different experience, M: Okay.

P: From anything else, uhm…

M: And how the feedback was presented specifically?

P: Yeah, I think that's helpful. M: Okay.

P: And uh, but it's up to the essential that the expert at hand explains…

M: [unclear] definitely.

P: [unclear] if you only look at he [unclear] there's no way I can understand it. M: Oh, of course not.

P: So there has to be that explanation, yes.

M: Okay, good. Thank you.

P: No problem, thank you.

Appendix 7

7.1 Experience Questionnaire

Please answer questions below. The answers provided should be based on the media skills training research that you took part in 6 months ago.

1. **How many media interview have you given since the media training?**

   0 / 1 / 2 / 3 / more than 3

2. **If so, what interviews have you taken part in?**

---

3. **How many times have you engaged in public speaking since the media training:**
   0 / 1 / 2 / 3 / more than 3
4. **To what extent do you feel your confidence in giving a media interview has changed as a result of the training you completed in the previous study?**

   Much less confident / confident / somewhat less confident / no change / somewhat more confident / much more confident

5. **To what extent do you feel your confidence in giving a public talk has changed as a result of the training you completed in the previous study?**

   Much less confident / confident / somewhat less confident / no change / somewhat more confident / much more confident

6. **Do you feel you are capable of taking part in media interviews after the media training?** Yes / maybe / not really
7. **Do you feel you are capable of taking part in public speaking after the media training?** Yes / maybe / not really
8. **Do you think you are more or less likely to accept an invitation to take part in a media interview in the future as a result of the training you received in the previous study?**
   Much less likely / somewhat less likely / no change / somewhat more likely / much more likely

9. **Do you think you are more or less likely to take part in public speaking in the future as a result of the training you received in the previous study?**

   Much less likely / somewhat less likely / no change / somewhat more likely / much more likely

10. **Have you noticed any other changes in your communication skills following the training you received in the previous study?**
    Yes / No / not sure
11. **If you answered yes, please briefly describe the changes you've noticed.**

7.4 Data Analysis

7.4.1. Subjective ratings

7.4.1.1. Self-report of Communication Skills Ratings

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Feedback or none	Statistic	df	Sig.	Statistic	df	Sig.
SR_molar	No Feedback	.202	8	.200*	.933	8	.540
	Feedback	.220	8	.200*	.921	8	.436

*. This is a lower bound of the true significance. a.

Lilliefors Significance Correction

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
SR_molar	Based on Mean	.678	1	14	.424
	Based on Median	.196	1	14	.665
	Based on Median and with adjusted df	.196	1	10.213	.668
	Based on trimmed mean	.586	1	14	.457

Descriptive Statistics

Dependent Variable: SR_molar

Feedback or none	Mean	Std. Deviation	N
No Feedback	5.3750	.80312	8
Feedback	4.8000	1.03095	8
Total	5.0875	.94083	16

Tests of Between-Subjects Effects

Dependent Variable: SR_molar

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	1.322[a]	1	1.322	1.549	.234	.100
Intercept	414.122	1	414.122	484.962	.000	.972
Group	1.323	1	1.323	1.549	.234	.100
Error	11.955	14	.854			
Total	427.400	16				
Corrected Total	13.278	15				

a. R Squared = .100 (Adjusted R Squared = .035)

7.4.1.2. Self-awareness of communication skills at different training Points

Levene's Test of Equality of Error Variances[a]

		Levene Statistic	df1	df2	Sig.
SR_Followup	Based on Mean	.678	1	14	.424
	Based on Median	.196	1	14	.665
	Based on Median and with adjusted df	.196	1	10.213	.668
	Based on trimmed mean	.586	1	14	.457
SR_molar_T1	Based on Mean	.131	1	14	.722
	Based on Median	.060	1	14	.810
	Based on Median and with adjusted df	.060	1	13.585	.810
	Based on trimmed mean	.101	1	14	.755
SR_molar_T2	Based on Mean	.228	1	14	.641
	Based on Median	.244	1	14	.629
	Based on Median and with adjusted df	.244	1	13.934	.629
	Based on trimmed mean	.237	1	14	.634

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Group Within Subjects Design: Time

Tests of Normality

	Group	Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
SR_Followup	SS Feedback	.220	8	.200*	.921	8	.436
	Traditional Feedback	.202	8	.200*	.933	8	.540
SR_molar_T1	SS Feedback	.217	8	.200*	.940	8	.609
	Traditional Feedback	.273	8	.082	.897	8	.269
SR_molar_T2	SS Feedback	.229	8	.200*	.881	8	.193
	Traditional Feedback	.203	8	.200*	.924	8	.463

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Time	Sphericity Assumed	.347	2	.173	.362	.700	.025
	Greenhouse-Geisser	.347	1.678	.207	.362	.664	.025
	Huynh-Feldt	.347	2.000	.173	.362	.700	.025
	Lower-bound	.347	1.000	.347	.362	.557	.025
Time * Group	Sphericity Assumed	.607	2	.303	.633	.538	.043
	Greenhouse-Geisser	.607	1.678	.362	.633	.513	.043
	Huynh-Feldt	.607	2.000	.303	.633	.538	.043
	Lower-bound	.607	1.000	.607	.633	.440	.043
Error(Time)	Sphericity Assumed	13.420	28	.479			
	Greenhouse-Geisser	13.420	23.494	.571			
	Huynh-Feldt	13.420	28.000	.479			
	Lower-bound	13.420	14.000	.959			

7.4.1.3. Journalist Communication Skills Ratings of Trainee

Tests of Normality

	Feedback or none	Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
Trainer_molar	No Feedback	.166	8	.200*	.949	8	.701
	Feedback	.250	8	.150	.913	8	.374

*. This is a lower bound of the true significance. a.

Lilliefors Significance Correction

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
Trainer_molar	Based on Mean	4.784	1	14	.046
	Based on Median	4.285	1	14	.057
	Based on Median and with adjusted df	4.285	1	12.217	.060
	Based on trimmed mean	4.766	1	14	.047

Descriptive Statistics

Dependent Variable: Trainer_molar

Feedback or none	Mean	Std. Deviation	N
No Feedback	4.5750	1.45185	8
Feedback	6.0000	.74066	8
Total	5.2875	1.33460	16

Tests of Between-Subjects Effects Dependent Variable: Trainer_molar

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	8.123[a]	1	8.123	6.115	.027	.304
Intercept	447.322	1	447.322	336.785	.000	.960
Group	8.122	1	8.122	6.115	.027	.304
Error	18.595	14	1.328			
Total	474.040	16				
Corrected Total	26.718	15				

a. R Squared = .304 (Adjusted R Squared = .254)

7.4.1.4. Neutral Observer Communication Skills Rating

Intraclass Correlation Coefficient

	Intraclass Correlation[b]	95% Confidence Interval Lower Bound	Upper Bound	F Test with True Value 0 Value	df1	df2	Sig
Single Measures	.920[a]	.829	.969	34.148	15	30	.000
Average Measures	.972	.936	.989	34.148	15	30	.000

Two-way random effects model where both people effects and measures effects are random. a.

The estimator is the same, whether the interaction effect is present or not.

b. Type A intraclass correlation coefficients using an absolute agreement definition.

Tests of Normality

	Feedback or none	Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
NOs_molar	No Feedback	.250	8	.150	.898	8	.275
	Feedback	.300	8	.033	.822	8	.049

a. Lilliefors Significance Correction

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
NOs_molar	Based on Mean	4.096	1	14	.063
	Based on Median	3.702	1	14	.075
	Based on Median and with adjusted df	3.702	1	9.144	.086
	Based on trimmed mean	3.825	1	14	.071

Descriptive Statistics

Dependent Variable: NOs_molar

Feedback or none	Mean	Std. Deviation	N
No Feedback	5.3333	1.41870	8
Feedback	6.4583	.45973	8
Total	5.8958	1.17277	16

Tests of Between-Subjects Effects

Dependent Variable: NOs_molar

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	5.063[a]	1	5.063	4.553	.051	.245
Intercept	556.174	1	556.174	500.145	.000	.973
Group	5.062	1	5.062	4.553	.051	.245
Error	15.568	14	1.112			
Total	576.804	16				
Corrected Total	20.631	15				

a. R Squared = .245 (Adjusted R Squared = .191)

7.4.2. Experience, Awareness, Capabilities and Confidence

Ranks

	group	N	Mean Rank	Sum of Ranks
engagedmedia	SS Feedback	8	7.00	56.00
	Traditional Feedback	8	10.00	80.00
	Total	16		
engagedpublic	SS Feedback	8	6.56	52.50
	Traditional Feedback	8	10.44	83.50
	Total	16		
confidencemedia	SS Feedback	8	9.13	73.00
	Traditional Feedback	8	7.88	63.00
	Total	16		
confidencepublic	SS Feedback	8	7.75	62.00
	Traditional Feedback	8	9.25	74.00
	Total	16		
capablemedia	SS Feedback	8	7.50	60.00
	Traditional Feedback	8	9.50	76.00
	Total	16		
capablepublic	SS Feedback	8	8.88	71.00

	Traditional Feedback	8	8.13	65.00
	Total	16		
Likelymedia	SS Feedback	8	8.06	64.50
	Traditional Feedback	8	8.94	71.50
	Total	16		
Likelypublic	SS Feedback	8	7.00	56.00
	Traditional Feedback	8	10.00	80.00
	Total	16		
Changes	SS Feedback	8	9.75	78.00
	Traditional Feedback	8	7.25	58.00
	Total	16		

Test Statistics[a]

	Engaged media	Engaged public	Confidence media	Confidence public	Capable media	Capable public	Likely media	Likely public	Change
U	20.000	16.500	27.000	26.000	24.000	29.000	28.500	20.000	22.00
Sig. (2-tailed)	.064	.094	.559	.515	.333	.643	.689	.165	.21

a. Grouping Variable: group

b. Not corrected for ties.

7.4.1. Social Signal Detection

7.4.2.1. Principal Component Analysis

Sampling Adequacy

Case Processing Summary

	Exp_group	Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
BrowFurrow	Control	207	86.3%	33	13.8%	240	100.0%
	Experiment	180	75.0%	60	25.0%	240	100.0%

KMO and Bartlett's Test

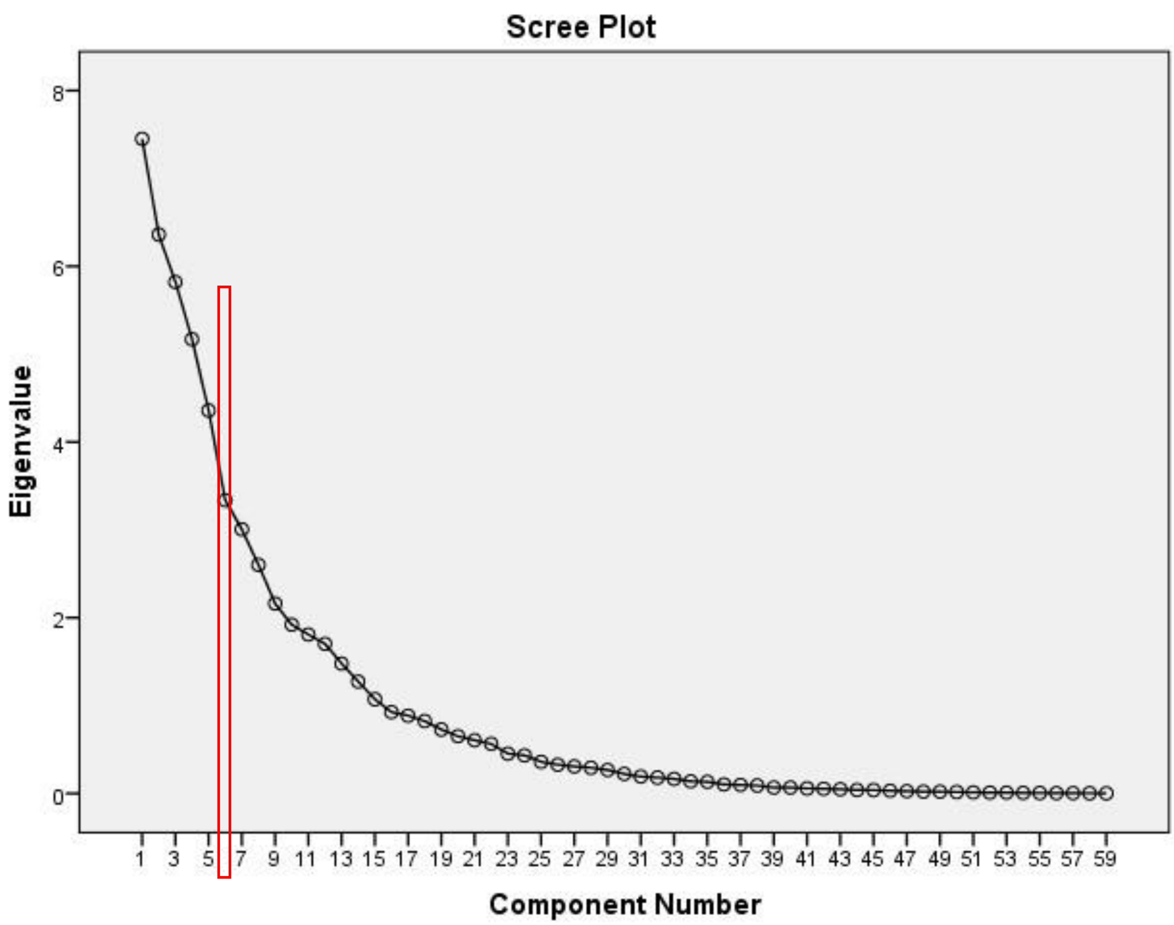
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.595
Bartlett's Test of Sphericity	Approx. Chi-Square	36567.345
	df	1711
	Sig.	.000

Preliminary Analysis

Total Variance Explained

Component	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	7.450	12.628	12.628	6.121	10.375	10.375
2	6.361	10.781	23.409	5.977	10.130	20.505
3	5.820	9.864	33.274	5.803	9.835	30.340
4	5.170	8.762	42.036	5.592	9.477	39.818
5	4.359	7.389	49.424	4.536	7.688	47.506
6	3.337	5.656	55.080	4.469	7.574	55.080

Extraction Method: Principal Component Analysis.



7.4.2.2. ANOVA Results

Disgusted

Descriptives

	Exp_group			Statistic	Std. Error
REGR factor score 1 for analysis 1	Control	Mean		-.0317336	.08524071
		95% Confidence Interval for Mean	Lower Bound	-.1997897	
			Upper Bound	.1363224	
		5% Trimmed Mean		-.2153551	
		Median		-.2807598	
		Variance		1.504	
		Std. Deviation		1.22640029	
		Minimum		-1.33693	
		Maximum		7.01989	
		Range		8.35683	
		Interquartile Range		.76434	
		Skewness		3.333	.169
		Kurtosis		13.247	.337
	Experiment	Mean		.0364937	.04847646
		95% Confidence Interval for Mean	Lower Bound	-.0591652	
			Upper Bound	.1321525	
		5% Trimmed Mean		.0128100	
		Median		-.1493088	
		Variance		.423	
		Std. Deviation		.65037997	
		Minimum		-1.05388	
		Maximum		1.68973	
		Range		2.74362	
		Interquartile Range		.98921	
		Skewness		.552	.181
		Kurtosis		-.521	.360

Excited / Passionate

Descriptives

	Exp_group			Statistic	Std. Error
REGR factor score 2 for analysis 1	Control	Mean		-.1802955	.06265235
		95% Confidence Interval for Mean	Lower Bound	-.3038176	
			Upper Bound	-.0567735	
		5% Trimmed Mean		-.2927313	
		Median		-.2878753	
		Variance		.813	
		Std. Deviation		.90141038	
		Minimum		-1.20888	
		Maximum		3.43102	
		Range		4.63990	
		Interquartile Range		.67838	
		Skewness		2.170	.169
		Kurtosis		5.235	.337
	Experiment	Mean		.2073398	.07959972
		95% Confidence Interval for Mean	Lower Bound	.0502653	
			Upper Bound	.3644144	
		5% Trimmed Mean		.1209633	
		Median		-.0297885	
		Variance		1.141	
		Std. Deviation		1.06794224	
		Minimum		-1.26801	
		Maximum		4.68628	
		Range		5.95429	
		Interquartile Range		1.39548	
		Skewness		1.376	.181
		Kurtosis		2.675	.360

Eagerness to Speak / Enthusiasm

	Exp_group			Statistic	Std. Error
REGR factor score 3 for analysis 1	Control	Mean		-.4070941	.04050271
		95% Confidence Interval for Mean	Lower Bound	-.4869471	
			Upper Bound	-.3272411	
		5% Trimmed Mean		-.3907825	
		Median		-.4127589	
		Variance		.340	
		Std. Deviation		.58273247	
		Minimum		-2.20179	
		Maximum		1.40737	
		Range		3.60916	
		Interquartile Range		.58174	
		Skewness		-.421	.169
		Kurtosis		2.086	.337
	Experiment	Mean		.4681582	.08671731
		95% Confidence Interval for Mean	Lower Bound	.2970384	
			Upper Bound	.6392779	
		5% Trimmed Mean		.4225091	
		Median		.2753857	
		Variance		1.354	
		Std. Deviation		1.16343485	
		Minimum		-1.19737	
		Maximum		2.89053	
		Range		4.08790	
		Interquartile Range		1.64243	
		Skewness		.717	.181
		Kurtosis		-.501	.360

Positive Engagement

	Exp_group			Statistic	Std. Error
REGR factor score 4 for analysis 1	Control	Mean		.2199856	.07442066
		95% Confidence Interval for Mean	Lower Bound	.0732618	
			Upper Bound	.3667094	
		5% Trimmed Mean		.1938040	
		Median		.1893527	
		Variance		1.146	
		Std. Deviation		1.07072678	
		Minimum		-2.13745	
		Maximum		2.74829	
		Range		4.88574	
		Interquartile Range		1.15273	
		Skewness		.439	.169
		Kurtosis		-.153	.337
	Experiment	Mean		-.2529835	.06310123
		95% Confidence Interval for Mean	Lower Bound	-.3775015	
			Upper Bound	-.1284654	
		5% Trimmed Mean		-.2990183	
		Median		-.2963173	
		Variance		.717	
		Std. Deviation		.84659190	
		Minimum		-1.84013	
		Maximum		2.62430	
		Range		4.46444	
		Interquartile Range		.70576	
		Skewness		.834	.181
		Kurtosis		1.391	.360

Anger

	Exp_group			Statistic	Std. Error
REGR factor score 5 for analysis 1	Control	Mean		.0063739	.03938228
		95% Confidence Interval for Mean	Lower Bound	-.0712701	
			Upper Bound	.0840179	
		5% Trimmed Mean		.0061827	
		Median		-.0257219	
		Variance		.321	
		Std. Deviation		.56661234	
		Minimum		-1.39216	
		Maximum		1.45716	
		Range		2.84931	
		Interquartile Range		.52589	
		Skewness		.015	.169
		Kurtosis		.531	.337
	Experiment	Mean		-.0073300	.09963398
		95% Confidence Interval for Mean	Lower Bound	-.2039383	
			Upper Bound	.1892783	
		5% Trimmed Mean		-.1054076	
		Median		-.4670201	
		Variance		1.787	
		Std. Deviation		1.33673017	
		Minimum		-1.90123	
		Maximum		4.10893	
		Range		6.01016	
		Interquartile Range		1.11060	
		Skewness		1.269	.181
		Kurtosis		1.156	.360

Stressed

	Exp_group			Statistic	Std. Error
REGR factor score 6 for analysis 1	Control	Mean		.1889965	.07579794
		95% Confidence Interval for Mean	Lower Bound	.0395573	
			Upper Bound	.3384357	
		5% Trimmed Mean		.2091816	
		Median		.0064026	
		Variance		1.189	
		Std. Deviation		1.09054245	
		Minimum		-2.21508	
		Maximum		2.32698	
		Range		4.54206	
		Interquartile Range		1.41682	
		Skewness		-.169	.169
		Kurtosis		-.436	.337
	Experiment	Mean		-.2173460	.06231383
		95% Confidence Interval for Mean	Lower Bound	-.3403102	
			Upper Bound	-.0943818	
		5% Trimmed Mean		-.2245827	
		Median		-.3904196	
		Variance		.699	
		Std. Deviation		.83602777	
		Minimum		-1.86891	
		Maximum		1.35836	
		Range		3.22727	
		Interquartile Range		1.35485	
		Skewness		.287	.181
		Kurtosis		-.940	.360

Test of Homogeneity of Variances

	Levene Statistic	df1	df2	Sig.
REGR factor score 1 for analysis 1	2.909	1	384	.089
REGR factor score 2 for analysis 1	14.825	1	385	.000
REGR factor score 3 for analysis 1	85.072	1	385	.000
REGR factor score 4 for analysis 1	13.183	1	385	.000
REGR factor score 5 for analysis 1	83.100	1	385	.000
REGR factor score 6 for analysis 1	9.471	1	385	.002

Multivariate Analysis

		Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Disgust	Control	-.0659648	1.12590137	.07844527	-.2206278	.0886982	-1.33693	5.59399
	Experiment	.0364937	.65037997	.04847646	-.0591652	.1321525	-1.05388	1.68973
	Total	-.0181863	.93502334	.04759144	-.1117579	.0753854	-1.33693	5.59399
Excited	Control	-.1802955	.90141038	.06265235	-.3038176	-.0567735	-1.20888	3.43102
	Experiment	.2073398	1.06794224	.07959972	.0502653	.3644144	-1.26801	4.68628
	Total	.0000000	1.00000000	.05083286	-.0999439	.0999439	-1.26801	4.68628
Eager	Control	-.4070941	.58273247	.04050271	-.4869471	-.3272411	-2.20179	1.40737
	Experiment	.4681582	1.16343485	.08671731	.2970384	.6392779	-1.19737	2.89053
	Total	.0000000	1.00000000	.05083286	-.0999439	.0999439	-2.20179	2.89053
Positive	Control	.2199856	1.07072678	.07442066	.0732618	.3667094	-2.13745	2.74829
	Experiment	-.2529835	.84659190	.06310123	-.3775015	-.1284654	-1.84013	2.62430

	Total	.0000000	1.00000000	.05083286	-.0999439	.0999439	-2.13745	2.74829
Anger	Control	.0063739	.56661234	.03938228	-.0712701	.0840179	-1.39216	1.45716
	Experiment	-.0073300	1.33673017	.09963398	-.2039383	.1892783	-1.90123	4.10893
	Total	.0000000	1.00000000	.05083286	-.0999439	.0999439	-1.90123	4.10893
Stress	Control	.1889965	1.09054245	.07579794	.0395573	.3384357	-2.21508	2.32698
	Experiment	-.2173460	.83602777	.06231383	-.3403102	-.0943818	-1.86891	1.35836
	Total	.0000000	1.00000000	.05083286	-.0999439	.0999439	-2.21508	2.32698

Tests of Between-Subjects Effects Dependent Variable: REGR factor score 1 for analysis 1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	1.008[a]	1	1.008	1.154	.283	.003
Intercept	.083	1	.083	.095	.758	.000
Exp_group	1.008	1	1.008	1.154	.283	.003
Error	335.585	384	.874			
Total	336.721	386				
Corrected Total	336.593	385				

a. R Squared = .003 (Adjusted R Squared = .000)

Robust Tests of Equality of Means

		Statistic[a]	df1	df2	Sig.
Excited	Welch	14.643	1	352.072	.000
Eagerness	Welch	83.628	1	255.069	.000
Positive	Welch	23.497	1	381.655	.000
Anger	Welch	.016	1	234.332	.898
Stressed	Welch	17.149	1	379.207	.000

a. Asymptotically F distributed.