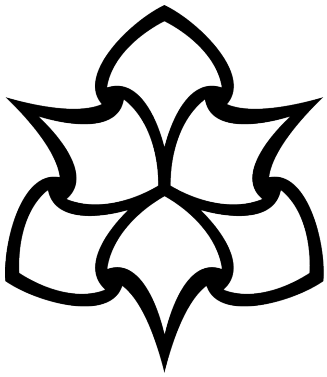


Manchester Metropolitan University

Buckingham, Fiona Jane (2016)*Detecting human comprehension from non-verbal behaviour using artificial neural networks.* Doctoral thesis (PhD), Manchester Metropolitan University.

Downloaded from: http://e-space.mmu.ac.uk/617426/

Usage rights: Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Please cite the published version

https://e-space.mmu.ac.uk

Detecting Human Comprehension From Nonverbal Behaviour Using Artificial Neural Networks

Fiona Jane Buckingham

A thesis submitted in partial fulfilment of the requirements of the Manchester Metropolitan University for the degree of Doctor of Philosophy

School of Computing, Mathematics and Digital Technology

Manchester Metropolitan University

May 2016

Abstract

Every day, communication between humans is abundant with an array of nonverbal behaviours. Nonverbal behaviours are signals emitted without using words such as facial expressions, eye gaze and body movement. Nonverbal behaviours have been used to identify a person's emotional state in previous research. With nonverbal behaviour being continuously available and almost unconscious, it provides a potentially rich source of knowledge once decoded. Humans are weak decoders of nonverbal behaviour due to being error prone, susceptible to fatigue and poor at simultaneously monitoring numerous nonverbal behaviours.

Human comprehension is primarily assessed from written and spoken language. Existing comprehension assessments tools are inhibited by inconsistencies and are often time-consuming with feedback delay. Therefore, there is a niche for attempting to detect human comprehension from nonverbal behaviour using artificially intelligent computational models such as Artificial Neural Networks (ANN), which are inspired by the structure and behaviour of biological neural networks such as those found within the human brain.

This Thesis presents a novel adaptable system known as FATHOM, which has been developed to detect human comprehension and non-comprehension from monitoring multiple nonverbal behaviours using ANNs. FATHOM's Comprehension Classifier ANN was trained and validated on human comprehension detection using the error-backpropagation learning algorithm and cross-validation in a series of experiments with nonverbal datasets extracted from two independent comprehension studies where each participant was digitally video recorded: (1) during a mock informed consent field study and (2) in a learning environment. The Comprehension Classifier ANN repeatedly achieved averaged testing classification accuracies (CA) above 84% in the first phase of the mock informed consent field study. In the learning environment study, the optimised Comprehension Classifier ANN achieved a 91.385% averaged testing CA. Overall, the findings revealed that human comprehension and non-comprehension patterns can be automatically detected from multiple nonverbal behaviours using ANNs.

Acknowledgements

First and foremost, I would like to express a huge thanks to my Director of Studies, Dr. Keeley Crockett, for her continuous support and encouragement throughout my doctoral research.

I would also like to thank the Manchester Metropolitan University and Family Health International 360 for funding the PhD studentship. Without the studentship it would have not been possible for me to embrace a doctoral research opportunity.

Last but not least, I would like to thank my family and friends for their enduring support throughout this journey.

Table of Contents

Table of Figures

Table of Tables

Abbreviations

AAM	Active Appearance Model
AI	Artificial Intelligence
AIDS	Acquired Immune Deficiency Syndrome
ALVINN	Autonomous Land Vehicle in a Neural Network
ANN	Artificial Neural Network
API	Application Programming Interface
ASL	Average Sentence Length
ASW	Average Number of Syllables Per Word
AU	Action Unit
AVI	Audio-Video Interleaved
BAP	Body Action and Posture
CA	Classification Accuracy
CF	Confidence Factor
CRF	Circulating Recombinant Form
DFD	Data Flow Diagram
DICCT	Deaconess Informed Consent Comprehension Test
ELM	Extreme Learning Machine
FACS	Facial Action Coding System
FHI 360	Family Health International 360
FKGL	Flesch-Kincaid Grade Level
FN	False Negative
FP	False Positive
FPS	Frames Per Second
FRES	Flesch Reading Ease Score
GUI	Graphical User Interface
HIV	Human Immunodeficiency Virus
HPTN	HIV Prevention Trials Network
IC-C	Close-ended Measure
IC-NV	Nonverbal Measure
IC-O	Open-ended Measure
IC-SP	Self-perception Measure

ID3	Iterative Dichotomiser 3
IEC	Institutional Ethics Committees
IRB	Institutional Review Boards
ISG	Intelligent Systems Group
MF	Media Foundation
MME	Micromomentary Expressions
MMU	Manchester Metropolitan University
MNC	Minimum Number of Cases
MNO	Minimum Number of Objects
MPEG	Moving Picture Experts Group
MSE	Mean Square Error
NIMR	National Institute of Medical Research
No-Prop	No-Propagation
OED	Oxford English Dictionary
PrEP	Pre-Exposure Prophylaxis
Q&A	Question and Answer
QuIC	Quality of Informed Consent
RMS	Root Mean Square
SSE	Sum-of-Squares Error
TN	True Negative
TP	True Positive
UK	United Kingdom
UNAIDS	Joint United Nations Programme on HIV and AIDS
XML	Extensible Markup Language
XOR	Exclusive-or
ZPD	Zone of Proximal Development

Chapter 1 Introduction

1.1 Introduction

This Thesis shall investigate whether an adaptable system can detect human comprehension and non-comprehension from monitoring multiple nonverbal behaviours using Artificial Neural Networks (ANN). Therefore, the adaptable system known as FATHOM was constructed and validated in order to pursue an answer to the investigation proposed. Most importantly, this Chapter establishes the general foci of this Thesis. The Chapter begins by providing research background and highlighting key motivations. The Thesis aim and objectives are stated. Lastly, a brief overview of the Thesis structure is provided.

1.2 Background and Motivation

The research presented in this Thesis takes a multidisciplinary approach. The creation of the automated adaptable system for human comprehension detection required knowledge to be combined from three discrete disciplines: human comprehension, nonverbal behaviour and ANNs.

Reliably detecting whether a person comprehends is difficult because at present we are unable to directly access their mind to discover their mental constructions (Newton, 2000). A person can reveal whether they comprehend through spoken language but it cannot always be relied upon due to their spoken responses being susceptible to concealment. Assessment tools are frequently employed to probe, capture and evaluate human understanding on a particular subject area e.g. asking closed/open questions, problem solving tasks and formal examinations etc. However, there is not a single generic assessment tool that is capable of being applied across all assessment contexts to detect human comprehension.

Human nonverbal behaviour provides a potential solution to the problem of reliably detecting human comprehension for numerous reasons. Nonverbal behaviour is a communication format where messages are transmitted via visual and auditory aspects such as facial expressions, gaze, posture, body movement, gestures, touch, non-linguistic vocal sounds and vocal acoustic parameters (Babad, 2009; Hall, 2007). Unlike spoken language, nonverbal behaviour is constantly available and can be produced subconsciously. Previous researchers have been able to identify universally

recognised emotions from nonverbal behaviours (Ekman and Friesen, 1971). Trained human observers can manually extract observed nonverbal behaviours by hand for analysis (Frauendorfer et al., 2014). However, human observers are prone to fatigue and other known weaknesses, so an automated computational approach to extracting the observed nonverbal behaviours is preferred.

ANNs are 'parallel computational models comprised of densely interconnected adaptative processing units' (Hassoun, 1995:1). ANNs are inspired by biological neural networks found in the human nervous system. Experiential knowledge is stored in the ANNs interconnections. ANNs have been used to solve a range of complex real world problems such as image and pattern recognition (Rothwell et al. 2006; LeCun, 1990b; Pomerleau, 1989). Hence, the latter properties make ANNs a suitable computational system for automatically identifying nonverbal behaviours and for finding possible classification patterns associated with human comprehension and non-comprehension.

At present, no attempts have been made to detect human comprehension from multiple nonverbal behaviours using ANNs. If an adaptable system is able to automatically detect human comprehension and non-comprehension from multiple nonverbal behaviours reliably then it could provide a new, flexible comprehension assessment tool that overcomes the weaknesses of human observers and allows comprehension states to be captured earlier e.g. during a learning event rather than post learning. The motivation for this research stems from the niche for an automated adaptable system for detecting human comprehension detection from nonverbal behaviour. Thus, the primary research question that this Thesis shall investigate is: can ANNs detect distinct patterns of human comprehension and non-comprehension from multiple nonverbal behaviours?

1.3 Thesis Aim and Objectives

The primary aim of the research in this Thesis is to develop an adaptable system that can detect human comprehension and non-comprehension from monitoring multiple nonverbal behaviours using ANNs.

The objectives of the research in this Thesis are:

- Research and review nonverbal behaviour, ANNs and human comprehension.

- Identify nonverbal behaviours known to be associated with human comprehension and non-comprehension from previous research that could be built into the adaptable system.
- Build an adaptable system that monitors multiple human nonverbal behaviours using ANNs.
- Design a study to enable the capturing of nonverbal behaviours associated with human comprehension and non-comprehension.
- Extract a dataset containing multiple nonverbal behaviours from a human comprehension study using the adaptable system.
- Optimise an ANN for classifying human comprehension and non-comprehension through training and validation with a nonverbal dataset extracted from a human comprehension study.
- Evaluate the performance and accuracy of the optimised ANN at classifying human comprehension and non-comprehension.
- Disseminate research findings in one or more publications.

1.4 Thesis Outline

A structural overview of this Thesis is shown in Figure 1.1. There are nine chapters in this Thesis. Each chapter falls under one heading: (1) literature review; (2) adaptable system design and implementation; (3) adaptable system exploratory research methodology and results; or (4) conclusions. The literature review is comprised of three distinct fields of research: nonverbal behaviour, human comprehension and ANNs. Chapter 2 introduces nonverbal behaviour and reviews how nonverbal behaviours have been isolated and extracted in previous research. In Chapter 3, human comprehension is defined and the existing methods used for human comprehension detection within informed consent and learning environments are discussed. Existing approaches on human comprehension detection using nonverbal behaviour(s) are also highlighted in Chapter 3. Chapter 4 describes the architecture of ANNs and the error-backpropagation learning algorithm, a popular learning algorithm for training ANNs. The ANN review also identifies common variations in the application of the error-backpropagation learning algorithm and provides examples of problems, which ANNs have been successfully applied to.

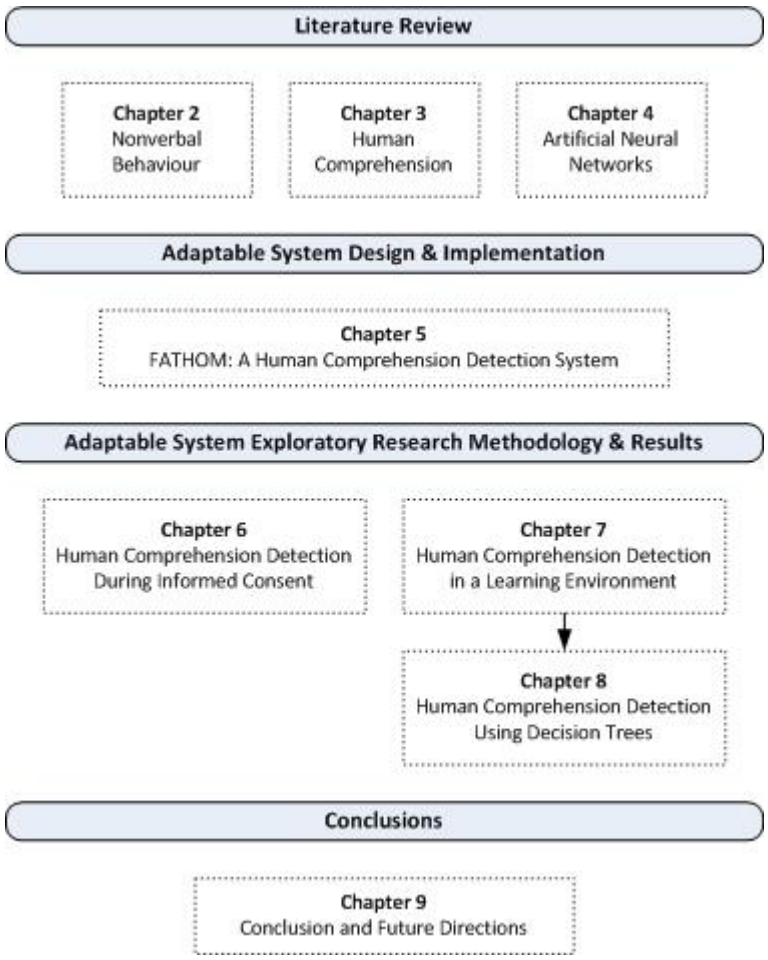


Figure 1.1 Thesis Outline

Following the literature review is the adaptable system implementation, where Chapter 5 describes how the novel adaptable system for human comprehension detection was constructed and how it works. Afterwards, investigative exploratory research using the adaptable system is introduced in three distinct Chapters, with each Chapter outlining the methodology, stating the results and providing empirical discussions. Chapter 6 explains how an ANN within the adaptable system has been trained and validated on human comprehension detection with a dataset from an informed consent field study. Chapter 7 explains how an ANN in the adaptable system has been trained and validated on human comprehension detection with a dataset from a learning environment study. In Chapter 8, decision trees are described and a well known decision tree induction algorithm is introduced. More importantly, Chapter 8 examines whether an optimised decision tree in the adaptable system can detect human comprehension from the dataset collated from the study outlined in Chapter 7. Lastly, Chapter 9 concludes the Thesis by summarising the main contributions of this research and provides detailed suggestions on potential future research directions.

Chapter 2 Nonverbal Behaviour

2.1 Introduction

Every day, people communicate with one another by exchanging messages through speaking, writing or using some other recognised medium such as nonverbal communication. Nonverbal communication is the process of sending, receiving and interpreting wordless cues known as nonverbal behaviours. This chapter starts by defining nonverbal behaviour and then identifies how nonverbal behaviours have been isolated, extracted and measured in previous works. A comparison between verbal and nonverbal behaviour is also included to further define the distinct properties of nonverbal behaviour and to highlight similarities.

2.2 Nonverbal Behaviour

Communication is a process of sending encoded messages and decoding the meaning of the received messages by speaking, writing, or using some other medium. Humans communicate with one another using verbal communication and/or nonverbal communication. Verbal communication encompasses communication involving words that may by spoken, written or signed. Nonverbal communication can be simply defined as any form of communication without using words e.g. nonverbal behaviour. Nonverbal behaviour encompasses visual and auditory aspects such as facial expressions, gaze, posture, body movement, gestures, touch, non-linguistic vocal sounds and vocal acoustic parameters (Babad, 2009; Hall, 2007).

Throughout everyday social interactions, humans are exposed to an abundance of nonverbal behaviours. Humans often communicate face-to-face. During face-to-face communication, people predominantly focus their attention primarily upon the face as it contains a rich source of information for nonverbal feedback and indication of a person's emotional state (Knapp and Hall, 1992). The face is one of the most important areas of the nonverbal production systems (Scherer and Wallbott, 1985) because of its physically complex structure and the diversity of nonverbal behaviours that it can emit e.g. facial expressions, eye gaze behaviour and physiological changes such as blushing. Cohn and Ekman (2008) argue that the face commands so much visual attention because it is always visible, continually delivers information and it houses four active sensory organs (smell, taste, sight and hearing), which play a critical role in human

survival. Even when the face is resting, Cohn and Ekman (2008) suggest that it may still deliver information about an individual's emotional state. Mehrabian (1968) identified that the facial component of a message delivered by an individual accounts for approximately 55 percent of the message's total impact. Birdwhistell (1970:158) suggested that 30-35% 'of the social meaning of a conversation or interaction is carried by the words'. Therefore, it is not surprising that the face naturally attracts more attention than other nonverbal behaviours in nonverbal research. Furthermore, when a person allocates more weighting to facial behaviour than any other communication channel, it is regarded as "facial primacy" (Knapp and Hall, 1992).

The field of nonverbal behavioural research is broad, which is reflected in the variety of the theoretical and empirical research papers published in Springer's *Journal of Nonverbal Research*, which span many different perspectives, approaches, applications and disciplines. As a result, only a selection of prominent nonverbal behavioural research has been highlighted throughout this review. Charles Darwin (1872) was amongst the earliest researchers to lay the foundations of nonverbal communication through his research ideas on emotional expression and nonverbal behaviour published in *The Expression of the Emotions in Man and Animals*. In his book, Darwin discusses the nature of how man expresses complex emotional states such as surprise, fear and anger from photographic illustrations predominantly focused upon facial expressions. One of the reasons for the 'lack of influence from Darwin's book was his reliance on anecdotal rather than systematic data' (Ekman, 2006:3). Many years later, Haggard and Isaacs (1966) discovered Micromomentary Expressions (MME) whilst looking for nonverbal communication between therapist and patient in photographic films at a slower frame rate. MMEs are subtle, subliminal, short-lived facial expressions, which lasts for a fraction of a second i.e. so rapid that you can miss a MME with the blink of an eye. Nowadays, MMEs are more commonly referred to as microexpressions. After much hypothesising and debate amongst researchers regarding the existence of universals in facial behaviour, in 1971 Ekman and Friesen (1971) found evidence supporting the hypothesis that there are some universally recognised patterns of muscular facial behaviour related to specific emotions. They uncovered six universal facial expressions of emotion: happiness, sadness, anger, fear, surprise and disgust.

During social interactions, individuals produce bodily movements, often using the hands, the head or other body parts (Argyle, 2013). These nonverbal body movements are referred to as gestures. The distinct types of gestures are: emblems, illustrators and adaptors. Emblems are distinct nonverbal bodily movements or configurations that have a precise meaning, which is directly translatable in to words (Ekman and Friesen, 1969a) e.g. hand gestures such as the okay sign (👌), the hitchhiking sign (👍) and the victory sign (✌). Therefore, emblems can be used in the absence of speech. Emblems are learnt in a social context so vary between cultures (Ekman, 2004) and are usually produced voluntarily (Ekman and Friesen, 1969a). Illustrators are bodily movements that accompany speech and are interconnected with the verbal content (Ekman and Friesen, 1969a) e.g. pointing finger to indicate direction whilst verbally giving directions to a lost traveller. Adaptors are bodily movements that manipulate the self or an object within the environment (Krauss et al., 1996) e.g. scratching, tapping and fidgeting. Typically, individuals have the lowest conscious awareness of adaptor gesture production (Ekman and Friesen, 1969a).

2.3 Nonverbal Channel and Multichannels

In nonverbal research, the term nonverbal channel is used to describe a single isolatable unit of nonverbal behaviour e.g. left eye blink. The isolated nonverbal channels are then measured and/or monitored for research purposes. Nonverbal channels can also be referred to as nonverbal cues or signals. The division of nonverbal behaviours in to nonverbal channels can differ between researchers (Hall et al., 2008). For example, one researcher may define the raising of the eyebrows as one nonverbal channel but another researcher may categorise the same nonverbal behaviour down into three nonverbal channels: the inner brow raise, the outer brow raise and the concatenation of an inner and an outer brow raise (Cohn and Ekman, 2008). Section 2.5 on the measurement of nonverbal behaviour identifies a selection of nonverbal coding systems, which provides one explanation as to why researchers using different nonverbal coding systems have different definitions of nonverbal channels.

Patterson (2014) highlighted that in early nonverbal research, researchers focused upon the examination of a single nonverbal channel because of the expense and limitations of video technology for capturing nonverbal behaviour and the limited physical mental processing capabilities of human judges for the observational

recording of multiple nonverbal channels (multichannels). Nowadays, nonverbal researchers are less restricted when it comes to capturing nonverbal behaviours because digital recording technology is cheaper, more sensitive, lightweight and abundant. Patterson (2014) also argues that nonverbal researchers should focus upon patterns of behaviour because during natural social interactions individuals simultaneously communicate through multichannels of behaviour, which are then judged by others in a holistic manner. Furthermore, App et al. (2011) suggest that individuals do not use all nonverbal channels at once to communicate emotional messages because of the quick delivery speed of messages and the energy cost for using all nonverbal channels. Thus, each nonverbal channel constantly fluctuates between an active and inactive (redundant) state to produce nonverbal multichannel patterns that are associated with specific emotions. On the other hand, Birdwhistell (1970:70) argues that 'while no single channel is in constant use, one or more channels are always in operation'.

Birdwhistell (1970) proposed that when the human communicational process is broken down in to discontinuous isolated channels that it simplifies and provides order to the process, which when reassembled become a continuous process with a multilevel description. Therefore, the benefits of adopting the approach to breakdown the human nonverbal communication process in to discrete nonverbal channels far outweigh the disadvantages.

2.4 Nonverbal Behaviour Versus Verbal Behaviour

There are key differences between nonverbal and verbal communication. Guerrero and Farinell (2009) enumerated the distinguishing features of nonverbal and verbal communication, which are as follows. For nonverbal communication, the distinctive characteristics are that it is analog, multimodal and variable, more spontaneous, only occurs in the present and is non-reflexive. Whereas, for verbal communication the distinctive characteristics are that it is digital, unimodal and constant, more intentional, can refer back in time and is reflexive. Hence, the properties of verbal and nonverbal communication contrast and complement one another. From a structural perspective, verbal communication is constructed of spoken words, sentences and individual letters whereas nonverbal communication is composed of a variety of nonverbal behaviours e.g. facial expressions, body movement and posture etc. Due to

the abundance and continuous availability of nonverbal channels, they provide a potentially wealthier source of information to work with in comparison to the solitary, discontinuous verbal channel.

Ekman and Friesen (1969a) developed the Display Rule theory whereby emotions experienced by individuals in social contexts are expressed according to one of the four socially learnt display rules, which vary culturally i.e. de-intensify, over-intensify, affectless (neutral) or mask. Similarly, Snyder (1974) developed the Self-Monitoring approach, where individuals monitor and control their nonverbal and verbal behavioural channels according to the social context and their position on the validated Self-Monitoring Scale i.e. high self-monitors are highly sensitive in social contexts and adapt their behaviour based upon the observed social interactional cues whereas low self-monitors are less sensitive in social contexts, so focus less upon social interaction cues and adapt their behaviour less. Because behavioural expressions can be regulated to a degree, it leads to the question of whether the true, underlying emotional expression can be identified from communicated channels of verbal and/or nonverbal behaviour.

Some channels of verbal and nonverbal communication are regarded as being more controllable than others. Thus, a controlled channel often inherits the properties of being voluntary and intentional. Ekman and Friesen (1969b) proposed the nonverbal leakage hierarchy where the nonverbal channels are ranked according to their degree of controllability and leakage. By leakage, they meant nonverbal cues that are inadvertently emitted by the sender when controlling their emotional expressions, which reveals information about what the sender is attempting to conceal or deceive. Within the nonverbal leakage hierarchy, the most controllable nonverbal channels to the leakiest are: the face, hands, legs and then the feet. Later, Zuckerman et al. (1982) found that the tone of voice is leakier and less controllable than facial expressions in a study on the communication of honest and deceptive messages. On the other hand, verbal content is considered to be highly controllable but also susceptible to verbal leakage. McQuaid et al. (2015) revealed linguistic differences between honest and deceitful pleaders from manually transcribed pleas using an automated linguistic tool. Furthermore, Mehrabian and Wiener (1967) investigated the interpretation of inconsistent (contradictory) messages communicated via two channels (single spoken words and vocal tone) and found that the total message impact was predominantly

judged on the attitude (feeling) delivered in the vocal tone. Therefore, underlying emotional expressions can be detected from the leakage of true emotions in the nonverbal and verbal channels of communication.

2.5 Measurement of Nonverbal Behaviour

The measurement of nonverbal behaviour(s) can be achieved by using manual, automatic or semiautomatic extraction techniques. In nonverbal research, the measurement of nonverbal behaviour is typically done via manual extraction (Frauendorfer et al., 2014) where a human observer is employed to hand code observed instances of each nonverbal channel under analysis either live, in real-time or post study from a recorded medium. Manual coding of nonverbal behaviour is time-consuming and expensive. The human observer needs to be trained on how to code each nonverbal channel, which is mentally (attentionally) demanding and tiring. When there are numerous nonverbal channels to be coded, human observers are unable to cope with simultaneously processing all nonverbal channels in memory. Miller (1956) argues that humans cannot process more than 7 ± 2 pieces of information at the same time in immediate memory. To counteract the memory capacity limitation, human observers have to adopt a serial approach to coding multichannels of nonverbal behaviour by repeatedly viewing video recorded observations, which is laborious and delays findings. Although, extra human observers can be hired to help overcome the latter problems, it introduces the additional issue of inter-observer agreement and reliability. There are multimedia annotation software tools such as ANVIL (Kipp, 2015; Kipp, 2001) and ELAN (The Language Archive, 2015; Wittenburg et al., 2006), which human observers can utilise to aid nonverbal coding of audiovisual streams but the task still remains manually time-consuming and the human observers have to be trained to use the software. In a review of 75 studies, Hall (1978) found that females were significantly better decoders of nonverbal communication than males. Regardless of sex, Gulabovska and Leeson (2014) also found that individuals with higher levels of emotional intelligence were better decoders of nonverbal behaviour. Hence, the disadvantages of the manual extraction and coding technique predominantly stem from the human observer weaknesses.

At present, there is no universally recognised coding system, which encompasses the coding of all human nonverbal behaviour. However, there are well known

independent coding systems, which exist for human facial expression analysis and body movement classification. Ekman and Friesen (1978) developed the Facial Action Coding System (FACS), which is a standard tool for the measurement of observed human facial expressions. Within FACS, observable facial behaviour is decomposed in to coded Action Units (AU), which anatomically represent distinct contractions of facial muscle(s) e.g. AU1 is an Inner Brow Raiser that uses the frontalis pars medialis facial muscle. The observed AUs and their combinations are then used to describe the emitted human facial behaviour. Bartlett et al. (1999) report that it takes over 100 hours of training to achieve minimal competency on FACS, and that each minute of video footage takes approximately 1 hour to score. Although FACS provides foundations for a consistent approach to coding human facial behaviour, it still remains a manual task when using human observers and incurs additional training costs.

Dael et al. (2012) developed the Body Action and Posture (BAP) coding system, which is a research tool for the measurement of observed human body movement. In BAP, there are 141 behaviour variables to be manually coded by human observers, spanning from articulation of the head, neck, torso, arms and the lower limbs. Observable human skeletal behaviour is categorised as a coded Action Unit or Posture Unit. Action Units are articulations of the body part(s) e.g. head shake. Posture Units are when the body part(s) are positioned in resting configurations e.g. arms crossed. 'Like FACS, BAP is anchored in human anatomy and descriptions are thus closely related to the mechanisms of movement production' (Dael et al., 2012:117). Weaknesses of BAP are that it lacks detailed movement of body parts such as the hands, it requires the manual setup of the XML based BAP coding scheme within the ANVIL software and it relies upon complex manual coding. The present lack of a universally recognised coding system may be inhibited by the issue of differences between researchers on the definition of an isolatable unit of nonverbal behaviour, which was discussed in Section 2.3.

Nowadays, video recordings are preferred over still photographs as a storage medium for capturing observed nonverbal behaviours for analysis. An advantage of digital and analogue videos is that it provides the researcher with a real-time sequential medium, which can be repeatedly viewed. Furthermore, videos can be modified so that only the audio or video data stream is played back to the observer for

analysis. Interestingly, Haggard and Isaacs (1966:154) found that 'a more vivid picture of the nonverbal aspects of the therapist-patient interchange' was obtained when repeatedly running their study of psychotherapy film footage in different modes i.e. at normal speed, backwards or forwards, fast or slow, frame-by-frame and silently. Therefore, video recordings can collate large volumes of rich data for analysis, which can be both advantageous and disadvantageous. On the other hand, photographs are static images, which are frozen at a fixed point in time without sound so cannot capture the natural, continuous flow of spontaneous human nonverbal behaviour as easily as video recordings.

Automated techniques for extracting nonverbal behaviour have been attempted but with varying degrees of success. Velloso et al. (2013) developed AutoBAP, a prototype system that automatically extracts and codes continuous human body movement according to the BAP coding system using input from a full body motion tracking suit and eye tracking glasses. Although the latter wearable sensing devices have the advantage of overcoming manual extraction weaknesses, their major pitfalls are that they are: expensive, cumbersome, sometimes uncomfortable, may require calibration, can inhibit/distract natural behaviour and are often restricted to use in a laboratory environment. Mukhopadhyay (2015) provides a comprehensive review on wearable sensors. Won et al. (2014a; 2014b) used Microsoft Kinect sensors to automatically capture basic human body movement, which is less obtrusive than wearable sensors but is confined to operational distances (1.2-4m), depth ranges and viewing angles. Bartlett et al. (1999) developed an automated hybrid system based upon three computer image analysis methods to classify six upper facial actions (AU's 1-6) from FACS, which achieved 91% accuracy and in comparisons, performed as well as the three expert human certified FACS coders. As a consequence of advances in the quality of digital technology and the computer vision field, it has provided nonverbal researchers with a selection of automated techniques at their disposal and the opportunity to produce finer grained analyses. However, the advancements are limited, often utilising expensive specialist sensors.

2.6 Summary

This chapter has described nonverbal behaviour, how it can be isolated in to discrete channels and measured using different types of extraction techniques.

Nonverbal communication is complex and multifaceted but it has the ability to provide insight on true human emotional state. Unlike verbal behaviour, nonverbal behaviour has the strength of being continuously available under less conscious control, thus potentially providing a rich dataset to work with. Furthermore, the face has been found to be the main focal point during human nonverbal communication. Advances in technology have made extraction and analysis of nonverbal behaviour easier, richer, faster and less inhibited by human decoder weaknesses. However, specialist technologies such as wearable sensors can be expensive and affect the expression of naturally occurring nonverbal behaviours. The next chapter introduces human comprehension and reviews previous work on human comprehension detection.

Chapter 3 Human Comprehension

3.1 Introduction

Throughout the human lifespan, a person regularly encounters tasks where understanding is required. Some tasks are easier to understand than others but comprehending them is dependent upon the individual's knowledge and ability. This chapter begins by defining human comprehension and then focuses upon how human comprehension is detected within two independent contexts: informed consent and learning environments. Because learning environments cultivate human comprehension, they are abundant with moments of comprehension and non-comprehension. The previous chapter introduced human nonverbal behaviour. This raises the question of whether human comprehension and non-comprehension can be detected from nonverbal behaviour and how. Therefore, the remainder of the chapter reviews previous work on detecting human comprehension and non-comprehension from nonverbal behaviour(s) using manual and automatic methods.

3.2 Human Comprehension

The Oxford English Dictionary (OED) (2016: online) definition of the word *comprehension* states that it is when a person has a 'mental grasping, understanding'. The OED definition is somewhat inadequate, vague and fuzzy, which reflects the complexity and broadness of the nontrivial term. Academic researchers have attempted to define comprehension. For example, Nickerson (1985:234) provides a more comprehensive comprehension definition:

> It requires the connecting of facts, the relating of newly acquired information to what is already known, the weaving of bits of knowledge into an integrated and cohesive whole. In short, it requires not only having knowledge but also doing something with it.

In other words, comprehension is a process whereby to comprehend one must be able to apply meaning obtained from knowledge through the synthesis of digested facts. The word *understanding* is synonymous with the word *comprehension*; therefore they are used interchangeably throughout this Thesis. On the other hand, non-comprehension is the opposite of comprehension, where knowledge state 'ranges

from uncertainty to complete lack of understanding' (Waring, 2002:1712). Human comprehension is not always black and white; it lies on a relatively unexplored continuum with comprehension and non-comprehension at the extremities. Therefore, this Thesis focuses upon the extremities.

Reliably detecting whether someone understands is difficult because at present we cannot directly access an individual's mind to discover their mental constructions (Newton, 2000). Furthermore, we cannot fully rely upon an individual's subjective verbal opinion on whether they comprehended because the verbal channel is susceptible to Self-Monitoring (Snyder, 1974) as highlighted in Section 2.4. For example, an individual may verbally conceal their lack of understanding due to embarrassment. There are a wide range of assessment tools that can be used to probe and capture evidence of human understanding for evaluation. Some examples of assessment tools are: closed questions, open questions, concept mapping, think-aloud protocols, observation and eavesdropping and problem solving tasks (Newton, 2000). The assessment tools enable the capturing of expressed human understanding to be articulated via different formats i.e. reading-writing (written), listening-speaking (oral), motor skills (practical) etc. Demonstrating how to tie a shoe lace is an example of a motor skill based activity. Each assessment tool has it strengths and weaknesses. Therefore, it is the responsibility of the assessor to decide, which assessment tool is suitable for evaluating understanding in a given context. The duration and frequency of the application of assessment tools varies according to the assessors evaluative requirements. Furthermore, there are different functions of assessment: formative (e.g. verification test), diagnostic (e.g. entrance test) and summative (e.g. General Certificate of Secondary Education exam). Even though there are a variety of assessment tools for obtaining evidence of human understanding, the techniques are generally not comparable in accuracy or reliability. Moreover, there is still not a single established assessment tool that can be applied to all assessment contexts to detect or predict human comprehension.

Some distinguishing properties of human comprehension are that: it is idiosyncratic; it can develop over time; it is a mental process that occurs within the individuals mind; it can be spontaneous; it varies in degree; it is context dependent and it is predominantly an unconscious activity. Miyake (1986) proposed that human understanding develops by progressing through a set of steps applied across levels

that contain gradually more complex concepts when observing how participants understood the function-mechanism hierarchy of a complex physical device. One reason why understanding varies between individuals is because of the approach that the individual chooses to adopt during learning. Marton and Säljö (2005) identified two distinct approaches to learning, which university students commonly adopted when tackling the task of learning a text-based article: the surface approach and the deep approach. Learners who used the surface approach, just attempted to memorise text, whereas learners that utilised the deep approach tried to understand the text by searching for meaningful links within the text and relating the text to real world applications. The latter finding highlights the importance of the assessor applying a suitable assessment tool in order to accurately identify evidence of comprehension so that corrective action can be applied as soon as non-comprehension is identified. Hence, successful recalling of a memorised fact does not demonstrate a deeper understanding of the fact.

3.3 Comprehension in Informed Consent

Put simply, informed consent is 'an individual's *autonomous authorisation* of a medical intervention or of participation in research' (Beauchamp and Childress, 2001:78). By autonomous authorisation, it is meant that the individual has the freedom and respect to self-determine whether to permit or refuse their willingness to participate. Informed consent stems from globally recognised ethical guidelines for medical research involving human subjects outlined in the Nuremburg Code (1949), the Declaration of Helsinki (World Medical Association, 2013) and the Belmont Report (1979). Faden et al. (1986) provide a comprehensive historical and theoretical review on informed consent.

The five core components of informed consent are: competence, disclosure, understanding, voluntariness and consent (Beauchamp and Childress, 2001). Competence refers to the individual's capacity to make a rationally informed decision. Disclosure entails the dissemination of adequate, accurate and relevant information about the research study/trial to all participants i.e. transmission of facts such as the nature, purpose, risks and benefits of the research including the right to withdraw. Understanding requires that the individual comprehends the disclosed research information and how the research will personally and potentially affect them.

Voluntariness means that the individual has the independent right to decide whether to participate without coercion, deception, influence, bribes or reprisal. Consent is when the individual actively decides upon whether to participate by authorising their decision through the formal handwritten signing of a consent form in the presence of a witness. An individual's participation in a research study/trial should not commence until valid informed consent has been received. Overall, informed consent is an interactive communicative process that occurs between research staff and their participants via verbal and written channels (Moodley et al., 2005).

The purpose of informed consent is to protect the individual's right to autonomous authorisation and to protect them from harm (Jefford and Moore, 2008). Institutional Review Boards (IRB)/Institutional Ethics Committees (IEC) are independent committees that exist within institutions that are responsible for reviewing and approving medical research involving human subjects by ensuring that it adheres to ethical guidelines. Understanding is a critical element of the informed consent process. The Declaration of Helsinki (World Medical Association, 2013) even states that the physician is responsible for checking for participant understanding. Although universally accepted ethical guidelines and IRB/IEC's exist to ensure that research is ethical and informed, accurately capturing, identifying and evaluating an individual's understanding during the informed consent process still remains a difficult and complex task. Thus, signing a consent form does not automatically imply comprehension (Jefford and Moore, 2008).

Researchers have an array of assessment tools at their disposal for capturing human comprehension during the informed consent process e.g. interviews and questionnaires etc. However, their use and success varies. Lindegger et al. (2006) compared four comprehension assessment tools: self-reports, checklists, vignettes and narratives and found significant differences between the measures of understanding. Moreover, the findings also suggested that that the degree of understanding was dependent upon the assessment tool used and that the closed-ended measures may overestimate levels of participant understanding. Thus, assessment of understanding in the informed consent process is typically manually tailored on a study-by-study basis using written and/or verbal assessment methods. Only a few attempts have been made to produce a standardised comprehension measurement tool for the informed consent process e.g. the Quality of Informed Consent (QuIC) (Joffe et al., 2001) questionnaire but its application is limited to cancer clinical trials and the Deaconess

Informed Consent Comprehension Test (DICCT) (Miller et al., 1996). At present, there is no universally accepted informed consent comprehension assessment tool with reliable benchmarks.

In 2015, Tam et al. (2015) systematically reviewed 103 studies on participants understanding of informed consent in clinical trials over a thirty year period (1983-2013) and found that there was no significant change in the understanding of informed consent elements but that understanding was significantly affected by covariates (controlled variables): age, educational level, critical illness, the study phase and location. Mandava et al. (2012) compared the quality of informed consent in studies between developing and developed countries, published from 1966 to 2010 and confirmed that it was poorer in developing countries. However, participant understanding of study information fluctuated across both developing and developed countries. Common findings across informed consent reviews are that individuals often have difficulty understanding experimental aspects such as randomisation (Hereu et al., 2010; Mandava et al., 2012; Afolabi et al., 2014; Tam et al., 2015) and placebo (Mandava et al., 2012; Afolabi et al., 2014; Tam et al., 2015). Therefore, the latter systematic reviews provide evidence of widespread comprehension difficulties in the informed consent process and emphasise the need for improvement.

Sand et al. (2010) reviewed how understanding has been defined and measured in the informed consent process and discovered that studies often lacked a common definition of understanding and the measurement tools varied considerably in length, depth and timing of the assessment(s). Berger et al. (2009) examined the length of informed consent documents from Norwegian oncology studies and found that within a twenty year period (1987-2007) that the overall length and quantity of informed consent components discussed had increased significantly. Therefore, the length and complexity of informed consent documents may hinder individual's understanding because humans cannot process more than 7 ± 2 pieces of information at the same time in immediate memory (Miller, 1956). However, Paris et al. (2010) attempted to improve individuals understanding of an informed consent document by enhancing the lexicosyntactic readability (i.e. long sentences shortened and words with 3+ syllables were replaced with shorter synonyms where possible) and by using working groups but there was no significant improvement. Using less technical language, larger fonts, adding images and decreasing the reading level are some examples of other

interventions, which have been used in an attempt to improve understanding of informed consent forms but with limited success (Flory and Emanuel, 2004). Language can be a barrier to informed consent comprehension. Lema et al. (2009) highlight that in parts of sub-Saharan Africa some words used in informed consent for clinical trials do not have a direct translation in to local languages e.g. placebo and randomisation. Therefore, improving the comprehension component in the informed consent process remains an ongoing challenge.

Failure to understand during informed consent means that the individual's action is not autonomous (Faden et al., 1986), therefore does not qualify as a truly informed decision. Furthermore, misunderstanding during informed consent has potential consequences relating to the individual such as noncompliance, discontent feelings, unrealistic expectations and therapeutic misconceptions. Checking for an individual's understanding can be time-consuming (West and Baile, 2010). Therefore, there is a niche for developing new techniques to improve the capture of understanding during informed consent. For that reason, this Thesis has focused solely upon the understanding component within informed consent.

3.4 Comprehension in Learning Environments

By learning environment, this Thesis is referring to physical spaces, which students reside in when engaging in learning. For example, it may be a classroom within a school, college or university.

Within the classroom, informal techniques traditionally used by teachers for monitoring student comprehension during lessons are: question asking and observations via room circulation (Schumm et al., 1997). Asking a student a question relating to subject matter enables the teacher to immediately obtain evidence of comprehension or non-comprehension and then react accordingly. Observations enable teachers to view student's approaches and progress on subject related tasks. Both informal techniques are restricted to focusing on one student at a time i.e. serial processing. The latter techniques are also utilised within adult education. Some other examples of informal comprehension monitoring techniques, which teachers have at their disposal, are: student summaries of main points, lesson reaction sheets, learning logs, collaborative open-note tests and fake pop quizzes etc (Schumm et al., 1997). Each informal technique has it strengths and weaknesses. Therefore, it is the

responsibility of the teacher to decide, which technique is suitable for monitoring student understanding during each lesson.

More formal techniques can be used to evaluate a learners understanding e.g. written assignments, in-class tests and examinations. Although teachers can assess an entire class's understanding from student's written work, it has several weaknesses. Marking written work is time-consuming resulting in feedback delays; the incorrect answers don't always reveal how the answer was obtained and recapping problematic topics at a later stage in the course can confuse students (Neill, 1991). Thus, highlighting the need for a human comprehension monitoring system that can detect non-comprehension in near real-time so that educators can provide immediate feedback and apply corrective actions to continually scaffold learning and maximise teaching time.

During learning, learners may ask their instructor or a peer for help when experiencing difficulty understanding a task. For example, within the classroom students usually request help from their teacher by intentionally raising their hand (✋) to attract their attention. Therefore, the hand-raising gesture could be regarded as an indicant of non-comprehension. However, help-seeking behaviour cannot always be relied upon because it is influenced by the social environment, thus making it susceptible to Self-Monitoring (Snyder, 1974) (Section 2.4). Help-seeking is a degrading activity that can lower an individual's self-esteem (Gall, 1985). In elementary classroom observation studies, frequencies of children's help-seeking behaviour was found to be higher in mathematics than reading lessons and varied across ability levels i.e. high, average or low (Gall and Glor-Scheib, 1985). Without appropriate help, a student may not reach their Zone of Proximal Development (ZPD). ZPD 'is the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers' (Vygotsky, 1980:86). Production of help-seeking behaviour is also dependent upon the learner's metacomprehension monitoring ability. Metacomprehension 'refers to a person's ability to judge his or her own learning and/or comprehension of text materials' (Dunlosky and Lipko, 2007:228). Therefore, a learner that has a poor metacomprehension monitoring ability struggles to learn well and is less aware of the need for help-seeking during learning difficulties. As a result, learner's subjective

evaluations of their comprehension level during learning cannot always be relied upon, which emphasises the need for a more reliable human comprehension detection method.

In education, teacher/lecturers are regularly exposed to teaching an array of children/students that do and do not experience difficulty comprehending elements of subject matter at different points in time. As a result, this makes the educator a primary observer and judge of their cohorts learning behaviour(s). Jecker et al. (1964) argues that as the audience becomes larger that the teacher becomes more reliant upon nonverbal feedback from the class to make communication judgements as verbal feedback is less viable. Therefore, it is not surprising that some researchers have approached educators in order to obtain their perspective on the existence of nonverbal behavioural indicators of human comprehension and non-comprehension in Section 3.5.1.

The consequences of failing to comprehend in learning environments are numerous and can have both short-term and long-term effects. For instance, it can affect attainment of expertise and academic performance i.e. qualification grade(s). When a student is unable to scaffold their learning to the ZPD, they may fall behind as the material becomes progressively more difficult on the same concept. As a result, failing to comprehend may hinder an individual's future pathways e.g. subsequent qualifications and career choices. Failing to understanding also presents personal challenges such as frustration, embarrassment, disappointment and can be demotivating.

3.5 Human Comprehension and Nonverbal Behaviour

The following subsections identify significant previous work on the detection of human comprehension and non-comprehension from nonverbal behaviour(s) via two approaches. The first approach in subsection 3.5.1 covers current manual methods of detecting human comprehension from nonverbal behaviour(s). The second approach in subsection 3.5.2 encompasses a selection of automated techniques for detecting human comprehension from nonverbal behaviour(s). Specific nonverbal behaviours associated as being indicants of human comprehension are also highlighted and some are used to guide the experimental design contained in Chapters 5-8.

3.5.1 Manual Detection of Human Comprehension

Webb et al. (1997) examined expert and novice teacher's ability to judge student understanding from the visual nonverbal behaviours contained in soundless videos, where each student completed a multiple-choice test that was presented verbally by the teacher from a script. Webb et al.'s (1997) investigation uncovered several findings: (1) all teachers regarded excessive physical movements, eye movements and facial expressions as indicators of comprehension and non-comprehension; (2) all teachers regarded secure/relaxed behaviour as an indicant of comprehension and nervous/insecure behaviour as an indicant of non-comprehension or confusion; (3) response latency, response confidence, attention and deliberate/impulsive behaviour were nonverbal cues commonly used by the teachers to judge student understanding; and (4) a slow response can be an indicant of comprehension or non-comprehension. Thus, the latter findings emphasize the complexity and difficulty of accurately decoding comprehension level from nonverbal behaviour(s).

Van Amelsvoort and Krahmer (2009) identified that adults can perceive task difficulty by watching the facial expressions of children solving easy and difficult math problems in soundless videos. Further analysis, revealed that out of the five facial expressions loosely coded using FACS (Webb et al., 1997) (smiling, eye gaze, frowning, funny face and visual response delay) that smiling occurred most frequently during easy math questions whilst the remaining facial expressions occurred more frequently during the difficult math questions. In a separate set of studies using the same videos with sound, Van Amelsvoort et al. (2013) asked 51 adults to indicate their reason(s) for each perceived task difficulty rating decision made for each video clip. Interestingly, 94% of adults reported looking at the children's facial expressions and/or listening to the voice, 86% reported paying attention to the children's pauses and delays in answering and 39% paid attention to the children's hesitations or (un)certainty in the voice. Furthermore, when delays (pauses) in children's answers to arithmetic problems were removed from the video clips, 52 newly recruited adults were still able to successfully perceive task difficulty and the categories of cues indicated were the same as before. To determine whether lecturers use student's nonverbal behaviour as an aid for detecting student comprehension level in virtual learning environments, Sathik and Jonathan (2013) distributed a questionnaire to 100 lecturers, each with over 10 years

lecturing experience and found from rankings that they most frequently used facial expressions followed by body language, gestures and then hands. Moreover, facial AUs associated as sources of comprehension level by the lecturers were: eye enlarge/shrink, eyebrow raised/lower, mouth widen, forehead wrinkle and lips curled. Thus, human comprehension has been manually detected in different learning environments with participants of different ages and from different perspectives i.e. teacher/adult/researcher observations.

Goldin-Meadow and Alibali (2013:257) argue that 'gesture reflects speakers' thoughts, often their unspoken thoughts, and thus can serve as a window onto cognition'. Hand gestures have been found to provide additional knowledge of a child's understanding when they verbally explain a concept. "Gesture-speech matches" occur when gestures and verbal responses deliver interconnected information whereas "gesture-speech mismatches" occur when gestures and verbal response deliver disconnected information (Church and Goldin-Meadow, 1986). Goldin-Meadow (2004) observed school aged children producing gesture-speech mismatches and matches when solving mathematical equivalence problems. However, the findings are limited to children and the identification of gesture-speech mismatches have a constant reliance upon the presence and synchronous processing of the verbal channel with spontaneous hand gestures.

Few cultural investigations have been executed to identify whether universal patterns of nonverbal human comprehension and non-comprehension exist. Machida (1986) found that Anglo- and Mexican-American teachers were able to distinguish whether Anglo- and Mexican-American children understood an easy and difficult lesson on animal habitats, even though there were slight cultural differences in their nonverbal behavioural displays. The Anglo children produced more head movements than the Mexican-Americans. The Anglo girls and Mexican-American boys produced more hand movements than the Anglo boys and Mexican-American girls. Furthermore, the children produced more head tilting, hand and body movements and less eye contact when participating in the difficult lesson than in the easy lesson. Therefore, the finding suggests that the prospect of a universally recognised set of human comprehension nonverbal behavioural cues existing still remains a possibility.

Because behavioural expressions can be regulated to a degree as highlighted in Section 2.4, it poses the question of whether the true, underlying emotional

expression of human comprehension can be identified from communicated channels of nonverbal behaviour. Hrubes and Feldman (2001) found that college students do emit facial and upper body nonverbal behavioural displays of cognitive difficulty when independently engaging in easy and hard problem solving tasks. Furthermore, they also found that there were fewer nonverbal displays of cognitive difficulty during the easy problem solving tasks than the hard and that the nonverbal behavioural displays of cognitive difficulty were more easily distinguishable from the low self-monitoring students than the high self-monitors. Allen and Atkinson (1978) found that adult observers were able to successfully estimate the level of understanding from spontaneous and deliberate nonverbal behaviours emitted by high-and low-achieving children (9-10 year olds) viewing an easy and difficult electricity lesson on a television. Hindmarsh et al. (2011) analysed audio-visual recordings of naturally occurring dental training sessions in a UK dental school clinic and found that student dentists may make verbal claims of understanding but that it is the emitted nonverbal bodily behaviour that supports or contradicts the verbal statement, which the supervisors used to assess the student dentists understanding in real-time. Therefore, the underlying emotional expressions of human comprehension and non-comprehension can be detected from the leakage of true emotions within the nonverbal behavioural channels.

3.5.2 Automatic Detection of Human Comprehension

After discovering that adults can perceive task difficulty from nonverbal behaviours contained in videos of children solving arithmetic problems, Van Amelsvoort et al. (2013) wanted to see if task difficulty could be automatically predicted by tracking head movements in the first second (25 frames, adopted Thin Slice (Ambady and Rosenthal, 1992) analysis approach) of the child's response using an Active Appearance Model (AAM) (Cootes et al., 2001) validated with leave-one-out cross-validation. An AAM (Cootes et al., 1998) contains a statistical model of the shape and appearance of an object (e.g. a face), which has been trained to recognise presentations of previously unseen images of the object. The MATLAB AAM method applied a set of landmarks to each child's face, which enabled head motion to be tracked from subsequent landmark coordinate calculations. Results revealed a 71% difficulty prediction accuracy and that the easy tasks predominantly contained vertical

head motion whereas difficult tasks predominantly contained diagonal head motion. However, 66 landmarks had to be manually specified for frames in the training set, which is a time-consuming process.

From the perspective of reading comprehension and spoken language comprehension, various experimental studies have shown that eye behaviour can provide insight on human cognition. For example, Graesser et al. (2005) recruited thirty college students to read illustrated texts about common devices (e.g. cylinder lock) and then invoked cognitive disequilibrium by providing a device breakdown scenario for the participant to comprehend by verbally asking questions. During the device breakdown scenario, each participant was seated, audiovisual recorded and had an eye tracker attached to their head. The eye tracking correlations revealed that the low comprehenders were randomly looking at the device displayed on the computer screen and only fixating on device fault regions at chance levels whereas the high comprehenders were not randomly looking the device, they were fixating on device fault regions at above chance levels. Furthermore, the high comprehenders were also found to spend more time looking at device fault regions before and during question generation. Using a head mounted eye tracker, Tannenhaus et al. (1995) tracked the eye movements of participants that were verbally instructed to interact with a set of common objects in order to monitor comprehension of spoken language e.g. 'Put the apple that's on the towel in the box'. Whilst comprehending the instructions the participants incrementally, visually focused upon the objects involved after hearing the objects name. Moreover, comprehenders fixated upon the target object before more complex instructions had even finished being delivered. Rayner et al. (2006) tracked the eye movements of sixteen participants whilst they silently read a collection of easy and difficult text-based passages that had been subjectively rated by an independent group. After every passage the participant was presented with a multiple choice question in order to assess their comprehension of the given passage. Total reading time and the number of eye fixations increased when reading difficult passages. Thus, results indicated that the global passage difficulty level (easy or difficult) was reflected in eye movements. Although, the eye trackers in the previous studies were able to detect human comprehension difficulties, they all inherited the following weaknesses: required calibration for each participant, only tracked one eye, expensive technology and bulky head mounted hardware that could hinder/distract natural behaviour.

3.6 Summary

This chapter began by defining human comprehension and outlining how human comprehension is currently detected within informed consent and learning environments. The reason why informed consent and learning environments have been selected is because human understanding plays a critical role in those types of real world environments and they have been used as settings within studies contained within Chapters 6-8. The chapter concluded by reviewing previous work on manual and automatic methods of detecting human comprehension and non-comprehension from nonverbal behaviour(s). The review in Section 3.5 revealed that human comprehension detection has had a predominant focus upon facial nonverbal channels, which highlights the importance of facial nonverbal behaviours in the role of human comprehension detection and may be interlinked with the facial primacy theory (Knapp and Hall, 1992). Previous studies on human comprehension detection have focused on few fine-grained nonverbal channels so there is a niche for further investigations using an automated multichannel approach. So far, none of the automated human comprehension detection approaches have attempted to use ANNs. Therefore, the following chapter introduces ANNs.

Chapter 4 Artificial Neural Networks

4.1 Introduction

Artificial Neural Networks (ANNs) are parallel computational models comprised of interconnected artificial neurons. ANNs are inspired by the structure and behaviour of biological neural networks found within the human brain, which are composed of interconnected biological neurons. Experiential knowledge is contained within the neural networks interconnections, which is obtained from the environment through the application of a learning algorithm. Neural networks have the strength of being able to learn from experience (examples) and generalise rather than being hard programmed (Peterson and Rögnvaldsson, 1991). This Thesis is primarily interested in using ANNs to classify human comprehension by detecting and continually monitoring multiple channels of nonverbal behaviour. This chapter begins by describing the architecture of ANNs and outlining error-backpropagation, a popular neural network learning algorithm. Common variations in the application of the error-backpropagation learning algorithm are reviewed. Existing applications of ANNs are also highlighted.

4.2 Artificial Neuron

An artificial neuron is a simple processing unit, which emulates the basic structure and behaviour of a biological neuron. The biological neuron (Figure 4.1) is composed of a cell body, dendrites, synapses and an axon. The biological neuron receives electrochemical signals from many other neurons through its synaptic connections in the dendrites. The cell body combines the incoming signals and fires a signal if the product exceeds the threshold, which is transmitted down the axon to other neurons. The strength of the signal is controlled by the synaptic connections, which releases neurotransmitter chemicals that cause an inhibition or an excitatory effect.

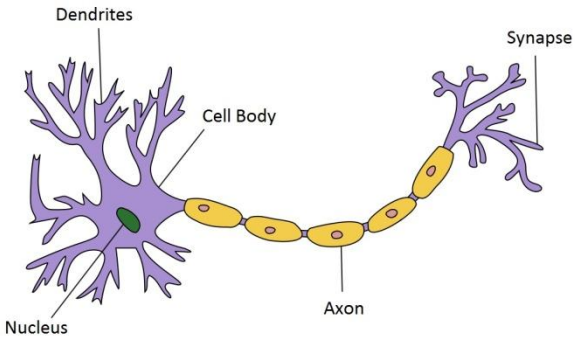


Figure 4.1 Biological Neuron

In 1943, McCulloch and Pitts (1943) developed the artificial neuron, a mathematical representation of a single biological neuron. The artificial neuron (Figure 4.2) is composed of a node, weights, and an output.

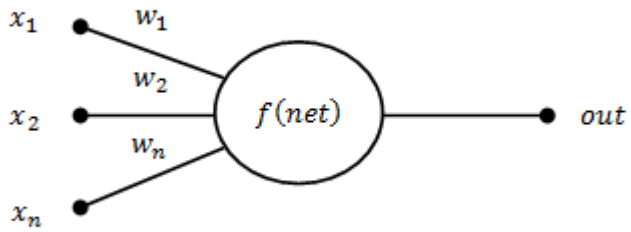


Figure 4.2 Artificial Neuron

The artificial neuron receives a number of input signals (x_1, x_2...x_n), which all have an associated weight (w_1, w_2...w_n) that controls the strength of the signal, causing an inhibition or an excitatory effect. The node combines the incoming input signals and weights in a product summation (*net*).

$$net = \sum_{i=1}^{n} w_i x_i \tag{4.1}$$

The node then applies a transfer (activation) function on the inner product, *f(net)*, which fires an output signal (*out*) to other neurons. There are different types of transfer functions (Duch and Jankowski, 1999) but the sigmoid function (logistic function) is the most widely used in artificial neural networks because of its monotonicity, simple form and the derivatives (Minai and Williams, 1993), which make them suitable for use with the error-backpropagation learning algorithm (Section 4.4). The sigmoid is a continuous "squashing" function that ensures that the artificial neurons output is within a limited range e.g. *out* ≤ 1 and *out* ≥ 0. The bipolar sigmoid has been adopted,

$$f(net) = \frac{2}{1 + e^{-\lambda net}} - 1 \tag{4.2}$$

where $e \approx 2.7183$ is a mathematical constant and the lambda (λ) controls the steepness of the slope in the transfer function. Using the bipolar sigmoid limits the artificial neurons output range to out ≤ +1 and out ≥ -1. The derivative of the bipolar sigmoid is

$$f'(net) = \frac{1}{2}(1 - f(net)^2). \tag{4.3}$$

An artificial neuron with two inputs is capable of solving linearly separable logic functions (Widrow and Lehr, 1998) e.g. the logic OR function illustrated in Figure 4.3(a) where the two classes (C_1 when *out* = 0 and C_2 when *out* = 1) are separated by a single hyperplane (decision boundary). A fundamental problem of the artificial neuron is its inability to solve nonlinear problems such as the exclusive-or (XOR) logic function, which was identified by Minsky and Papert (1988). An additional hyperplane is required in order to solve the separation of the two classes in the logic XOR function, as illustrated in Figure 4.3(b). Nonlinear problems were later solved by connecting multiple artificial neurons together in a layered architecture known as a multilayer artificial neural network (Section 4.3).

Figure 4.3 Logic Functions

4.3 Multilayer Artificial Neural Networks

Multilayer ANNs are computational models that replicate the behaviour of biological neural networks found within the human nervous system. Multilayer ANNs are comprised of multiple interconnected artificial neurons, arranged in a layered architecture. By connecting simple artificial neurons into neural networks it enables the harnessing of parallel computational power for solving complex nonlinear problems. Figure 4.4 shows a three-layer ANN that has an input layer, two hidden layers each with two artificial neurons and an output layer with one artificial neuron. Weighted interconnections connect all artificial neurons to the preceding layers, thus making it a fully connected feedforward neural network. The constant bias (-1) is an extra input that connects to all artificial neurons in the multilayer ANN and controls the threshold (θ) position. The neural network topology in Figure 4.4 can be expressed as (*n* + *bias*):2:2:1.

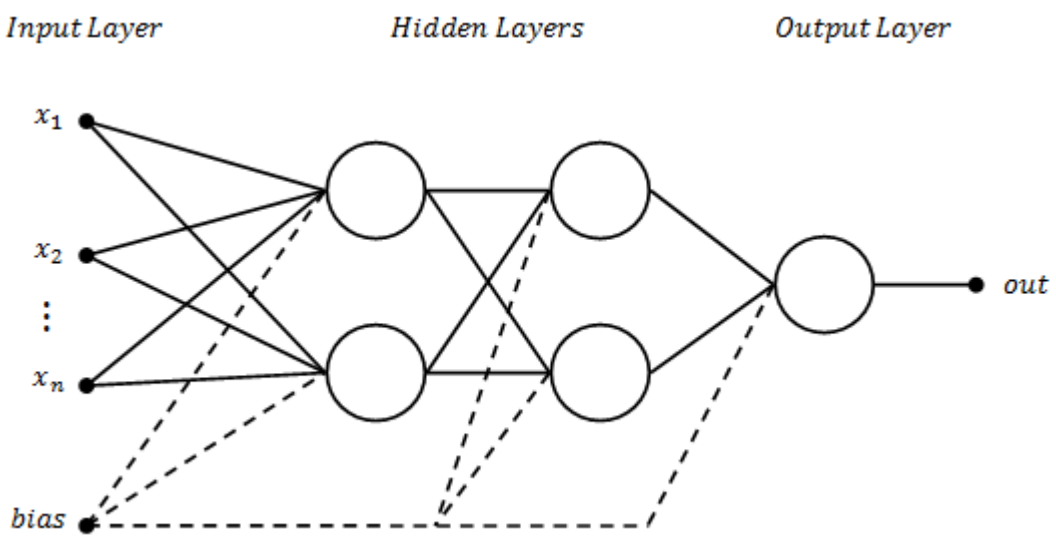


Figure 4.4 Multilayer Artificial Neural Network

A multilayer ANN has one or more hidden layers between the input and output layer (Fausett, 1994). A network with an input layer, no hidden layers and more than one artificial neuron on the output layer is called a single-layer ANN. Cybenko (1988) demonstrated that only two-layer ANNs with continuous transfer functions are required to approximate any arbitrary nonlinear continuous function, provided that there is a sufficient number of artificial neurons in the hidden layer. Therefore, this research has used multilayer ANNs with no more than two hidden layers.

4.4 Error-backpropagation Learning Algorithm

Neural networks acquire experiential knowledge in their interconnections (synaptic weights) by learning from and dynamically adapting to examples presented during training through the application of a suitable learning algorithm. The adaption of an ANN during training replicates the plasticity found within biological neural networks such as the developing human brain adapting to its surrounding environment (Haykin, 1999). The distribution of experiential knowledge in the neural networks interconnections is regarded as a connectionist model in the field of neurocomputing. Learning algorithms operate by searching the weight space for a set of weights that enable the ANN to generalise output on a given set of examples.

Error-backpropagation is a learning algorithm that is often referred to as backpropagation. Werbos (1974) originally founded the backpropagation algorithm in 1974 but it was later made popular by Rumelhart et al. (1986a). The purpose of the supervised learning algorithm is to find a set of weights during training that for every

set of inputs the ANN is able to produce output(s) close to the desired output(s). The advantages of the backpropagation algorithm are: it is a relatively straightforward learning algorithm to implement; only a few parameters to tune, its ability to solve complex nonlinear problems and that it has been widely used on feedforward neural networks (Priddy and Keller, 2005). Therefore, in this research the backpropagation algorithm has been adopted to train the feedforward neural networks.

The backpropagation algorithm has two major phases: a forward pass and a backward pass. In the forward pass the neural networks weights are "frozen". The training set contains a set of patterns known as the input-output pairs ($\mathbf{p} = [(\mathbf{x}^1,\mathbf{d}^1),(\mathbf{x}^2,\mathbf{d}^2) \ldots (\mathbf{x}^p,\mathbf{d}^p)]$), where $\mathbf{x}^p = [x_1, x_2 \ldots x_n]^T$ is the input vector and $\mathbf{d}^p = [d_1, d_2 \ldots d_n]^T$ is the desired response (target output) vector. An input vector from the training set is presented to the ANNs input layer. The activation signal of each artificial neuron is computed using Equations 4.1-4.2 by traversing from the hidden layer(s) to the output layer. The activation signals of the output layers artificial neurons form the output vector ($\mathbf{o} = [out_1, out_2 \ldots out_n]^T$).

During the backward pass the neural networks weights are "unfrozen" so that neuronal learning can occur. For neuronal learning to take place a learning rule has to be applied in the backward pass. The most commonly used neuronal learning rule for multilayer feedforward neural networks is the delta (δ) rule (Widrow and Hoff, 1960; Rumelhart et al., 1986b). The purpose of the delta rule is to locate a set of weights during training that for every set of inputs the ANN is able to produce output(s) close to the desired response(s). This is supervised learning because the neural network is provided with the correct answer in the desired response vector, which is used as a "teacher" for every input vector during the learning process. For each input vector, the delta rule computes the difference between the output and the desired response for each output neuron to determine the error signal using

$$\delta_{pj} = (d_{pj} - o_{pj})\, f'_j(net_{pj}) \tag{4.4}$$

where *p* is the input-output pattern and *j* is the j^{th} value of the input-output vector. After the hidden delta rule

$$\delta_{pj} = f'_j(net_{pj}) \sum_k \delta_{pk} w_{kj} \tag{4.5}$$

is used to compute the error signal for each neuron in the hidden layer(s) so that the error can be propagated from the output layer to the input layer, where *k* is the

number of the neuron in the succeeding layer. The weight adjustments (Δw) for each neuron is computed by using

$$\Delta_p w_{ji} = \eta \delta_{pj} \tag{4.6}$$

where η is the learning rate, which determines the step size of the gradient descent and w_{ji} is the weight from the i^{th} to the j^{th} neuron. Lastly, the error signal is minimised by updating the weights with the weight adjustments

$$w_{ji}(t+1) = w_{ji}(t) + \Delta_p w_{ji} \tag{4.7}$$

where t is time. 'The adjustment made to a synaptic weight of a neuron is proportional to the product of the error signal and the input signal of the synapse in question' (Haykin, 1999:53). The delta rule is an error-correction learning process. If the error signal is zero, no neuronal learning takes place; otherwise the weights are adjusted by attributing proportional "blame" to reduce the error signal.

Immediately after the backward pass the measure of error (E) for the presentation of the single input-output pattern can be determined using the sum-of-squares error (SSE) function

$$E = \frac{1}{2} \sum_j (d_{pj} - o_{pj})^2. \tag{4.8}$$

The forward and backward passes are repeated until all patterns in the training set have been presented once to the ANN (1 epoch). If E is below the preset maximum error (E_{max}) or the epoch counter (*epoch*) is at the preset maximum epochs ($epoch_{\max}$) then the backpropagation learning algorithm terminates. Otherwise, the backpropagation learning algorithm continues to process further epochs until E_{max} or $\text{epoch}_{\max}$ is encountered. Therefore, the SSE function can be extended to compute E on an epoch-by-epoch basis using

$$E(t+1) = E(t) + E \tag{4.9}$$

where t is time. The shape of the error surface in weight space (Figure 4.5) is characterised by hills, valleys and plateaus. As a gradient descent method, the delta rule minimizes E recursively in an attempt to converge at the global minimum (Figure 4.5), which is the optimal solution to the problem. Gradient descent is not guaranteed to converge at the global minimum and can get trapped in a local minimum (Figure 4.5). The learning pace is dependent on the shape of the error surface in weight space (Minsky, 1961). There are techniques in Section 4.5 that can be applied to help

accelerate gradient descent and increase the likelihood of locating an optimal solution when using the backpropagation learning algorithm with the delta rule.

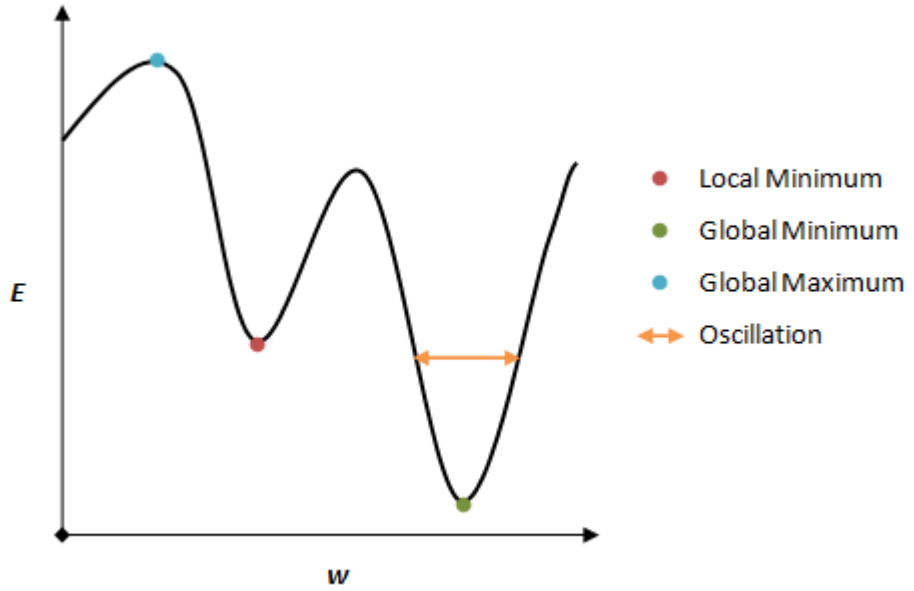


Figure 4.5 Minima in Weight Space

The frequency of weight updating can be performed by using incremental learning or batch learning. Incremental learning is when the weights are updated after every individual input-output pattern presentation. In batch learning the weights are only updated after all of the input-output patterns in the training set have been presented once i.e. every epoch. Batch learning moves the weight search in the direction of true gradient descent but is computationally demanding (Plaut et al., 1986). On the other hand, incremental learning moves the weight search in the direction of an estimate of true gradient descent (Plaut et al., 1986), requires less memory space and is stochastic (random), which permits wider search space exploration and can potentially produce better solutions (Gurney, 1997). Therefore, incremental learning has been used in this Thesis.

A summary of the backpropagation learning algorithm is shown in Figure 4.6. 'The backward pass has the same computational complexity as the forward pass, and so it is not unduly expensive' (Rumelhart et al., 1986b:327). The following section reviews approaches to improving the performance of the backpropagation learning algorithm from the literature.

i. Initialise weight vector ($\mathbf{w} = [w_{11}, w_{12}, \ldots w_{ij}]^T$) with random values
ii. Set η, *bias*, $epoch_{max}$ and E_{max}
iii. Randomly pick $\mathbf{x}^p$ and $\mathbf{d}^p$ from $\mathbf{p}$
iv. Compute forward pass
    a. Calculate *out* for each neuron from first hidden layer to output layer using Eq. 4.1-4.2
v. Compute backward pass
    a. Calculate δ's in output layer neurons using Eq. 4.4
    b. Calculate δ's in hidden layer(s) neurons using Eq. 4.5
    c. Calculate Δw using Eq.4.6
    d. Update w_{ji} with Δw using Eq. 4.7
vi. Calculate *E* using Eq. 4.8-4.9
vii. Repeat steps iii-vi until $epoch_{max}$ is encountered or *E* is below E_{max}

Figure 4.6 Error-backpropagation Learning Algorithm

4.5 Backpropagation Variations

When developing a neural network with a dataset using the backpropagation learning algorithm there are a number of variations on how it is applied and parameter values, which need to be empirically determined. These choices can be critical in determining the success of the search for an optimal solution. Therefore, the main problem is the number of parameters that can be used and establishing their optimal values. Another difficulty is that the optimal choices are usually unknown prior to investigation due to the problem being dependent upon the dataset (Thimm and Fiesler, 1995). Therefore, the following subsections have reviewed literature on the backpropagation learning algorithm variations and parameter values in order to guide the experimental design for the optimisation of the neural networks used in this research.

4.5.1 Data Preparation

The choice of encoding method on the input vectors representation has been found to have an effect upon training time and ANN generalisation performance. For example, Ahmad and Tesauro (1989) found that by changing the input representation range from [0, 1] to [-1, 1] that training time decreased and generalisation improved by 5-10%. Lawrence (1991) suggests that the raw inputs should be normalised to fall within a fixed interval, which the neurons understand i.e. usually unipolar [0, 1] or the bipolar [-1, 1] interval, which are the operational ranges used by neurons with sigmoidal squashing functions. Normalisation equalizes and rescales the inputs so that they are of the same order of magnitude. Without normalisation of the inputs, the

neurons in the neural network will operate in the saturation zone, which is detrimental to learning. If an input has a larger magnitude than the others then it can take the ANN longer to learn patterns because the learning algorithm has to compensate for the order-of-magnitude differences (Peterson and Rögnvaldsson, 1991). Scaling speeds up learning because it helps to balance out the rate at which the weights connected to the input neurons learn (LeCun et al., 1998). More complex normalisations procedures have the weakness of increasing the computational cost. Therefore, this research has normalised the dataset's representation range to [-1, 1] and utilised the bipolar sigmoid transfer function.

4.5.2 Topology Size

An ongoing 'question in neural network research is the size of the neural network needed to solve a particular problem' (Sietsma and Dow, 1988:325) i.e. the number of layers, hidden neurons and weights contained within the ANNs architecture (topology). Unfortunately, at present there is no "golden rule" for determining the optimal neural network architecture for any given problem, therefore the optimal solution can only be determined through experimentation. Smaller neural networks that generalise well are preferred over large networks because they have the advantage of being: cheaper to construct, computationally faster at processing input-output pairs and are easier to interpret their operational behaviour. The trial and error approach has been used to search for the optimal network architecture, which is a simple to implement but time-consuming because many networks of different fixed sizes have to be trained before an acceptable solution is discovered. There are theorem and approaches that can be applied to assist in the discovery of the optimum topology for a given problem.

The number of hidden layers required in a neural network is dependent upon the complexity of the given problem. Neural networks with no hidden layers are only capable of solving linear separable problems whereas multilayer ANNs are capable of solving more complex nonlinear problems. Cybenko (1988) demonstrated that only two hidden layer ANNs with continuous transfer functions are required to approximate any arbitrary nonlinear continuous function, provided that there is a sufficient number of artificial neurons in the hidden layers. Later, Cybenko (1989) went on to show that continuous feedforward single hidden layer neural networks could approximate any

arbitrary decision region well. Furthermore, Hornik et al. (1989) established multilayer feedforward neural networks as “universal approximators”, capable of approximating any arbitrary function to any degree of accuracy, provided that there are a sufficient amount of hidden neurons. From experimental investigations, de Villiers and Barnard (1993) concluded that feedforward neural networks with two hidden layers were more susceptible to local minima than single hidden layer feedforward neural networks. Therefore, this research has only experimented with network topologies with a maximum of two hidden layers. Although the latter publications have successfully determined the capabilities of hidden layers in neural networks, none of them have identified how many hidden neurons should reside in the hidden layer(s).

The number of artificial neurons in a neural network is crucial. When a neural network has too many neurons, it becomes prone to over-fitting the training dataset by memorising the input-output patterns, which causes poor generalisation on the unseen (test) dataset. On the other hand, when a neural network has too few neurons, it can experience difficulty learning and is susceptible to under-fitting, which causes poor generalisation on the training and test datasets. To help determine the optimum size of the ANN architecture there are two different incremental approaches that can be applied: growing (constructive) or pruning (destructive) methods (Bebis and Georgiopoulos, 1994). Growing algorithms start with a small neural network and gradually add hidden neurons one by one during the training process until performance plateaus e.g. Cascade Correlation (Fahlman and Lebiere, 1990). Pruning algorithms start with a large neural network and gradually discards unnecessary weights and/or artificial neurons whilst maintaining an acceptable performance. Therefore, pruning is based on the assumption that there is a large amount of redundant information stored in a fully connected feedforward neural network (Hush and Horne, 1993). Reed (1993) provides a detailed survey of pruning algorithms for feedforward neural networks. In general, there are two main subsets of pruning algorithms: (1) sensitivity methods that prune the trained neural network using a sensitivity measure e.g. Optimal Brain Damage (LeCun et al. 1990a) and (2) penalty-term methods that prune the neural network during training by modifying the error function, E in Equation 4.8 e.g. iterative pruning algorithm (Castellano et al., 1997).

Although incremental algorithms have the advantage of automatically determining optimal network size they have some major disadvantages. With pruning algorithms

there are the issues of when to stop pruning (Reed, 1993) and determining what initial neural network size is regarded as large for the given problem (Castellano et al., 1997). Growing methods and penalty-term methods have the weakness of interfering with the learning algorithm. Therefore, in this research the optimum topology search has used the trial and error approach with a maximum of two hidden layers.

4.5.3 Weight Initialisation

Several weight initialisation approaches have been proposed for multilayer artificial neural networks. Rumelhart et al. (1986b) suggest initialising ANN weights with small random values in order to counteract the problem of symmetry breaking, which is when the ANN starts with equal weight values and is unable to learn due to having identical error signals on each layer of neurons, thus remaining at a constant local maximum i.e. permanently sat at the top of a hill in weight space (Figure 4.5). Typically, weights are initialised to small random values in the range [-1, 1] or [-0.5, 0.5] (Fausett, 1994). Random weight initialisation has the advantage of being simple to implement. Kolen and Pollack (1990) demonstrated the sensitivity of the backpropagation learning algorithm from the initial weight vector through empirical studies and found that the optimum initial weight range for breaking symmetry was [≥ -0.5, ≤ 0.5].

Initial weights that are too large or too small should be avoided. If the initial weights are too large then it can saturate the neurons sigmoid i.e. the neurons output is in the upper or lower plateau of the "S" shaped sigmoid function. When the neurons are saturated they become insensitive to the learning process (Wessels and Barnard, 1992), taking small descents in gradient causing extremely slow learning (LeCun et al., 1998). On the other hand, if the initial weights are too small then the neurons output will be close to zero, which also causes small descents in gradient resulting in slow learning (Fausett, 1994; LeCun et al., 1998). Therefore, identifying intermediate initial weights that reside in between weight values that are too large and too small should overcome the latter issues.

Wessels and Barnard (1992) proposed a weight initialisation method that generates random weights on the order of $[-1/\sqrt{f},\ 1/\sqrt{f}]$ by using the range $[-3/\sqrt{f}, 3/\sqrt{f}]$ where *f* is fan-in of the neuron i.e. the number of weights entering the neuron. Thimm and Fiesler (1995) evaluated the performance of weight initialisation

methods on single hidden layer ANNs and found that Wessels and Barnard's (1992) weight initialisation method performed the best on average. Furthermore, they found that the weight range [-0.77, 0.77] produced the best average performance for all eight datasets used in their empirical study. Because the initial weights can have such an impact upon the success of the backpropagation learning algorithm, care should be taken when selecting the weight initialisation method in relation to the choice of sigmoid and data representation (LeCun et al., 1998) to avoid neuronal saturation and to thwart weight symmetry.

4.5.4 Learning Rate

The convergence of the gradient descent method in the backpropagation learning algorithm is dependent upon the learning rate (η) value used in the delta (δ) rule (Equation 4.6). The η is a small positive number, typically between zero and one, which determines the gradient descent step size i.e. the size of the weight changes. If the learning rate is too small then the step size is small resulting in a smoother descent on E in weight space but at the cost of a slower rate of learning. On the other hand, if the learning rate is too large then the step size is large resulting in a faster descent on E in weight space but tends to lead to an unstable neural network that overshoots the global minimum or oscillates (Figure 4.5). Oscillation is when the neural network repeatedly moves between the same two points in weight space.

Choosing an optimal learning rate for any given problem is not trivial because of the abundance of variations in the properties of the error surfaces. Therefore, the optimal learning rate is usually determined through experimentation using trial and error. The learning rate is often selected to be as large as possible without resulting in oscillations (Battiti, 1989). Zurada (1992) reports that η values ranging from 0.005 to 10 having been used successfully in the literature on experiments utilising the backpropagation learning algorithm. LeCun et al. (1998:22) argue that learning rates in the earlier layers should be larger than those in the later layers to accommodate correction of the 'second derivative of the cost function with respect to weights in the lower layers being generally smaller than that of the higher layers'. McLean et al. (1998) also suggest that larger learning rates should be applied to earlier layers than later layers. Through experimentation, Plaut and colleagues (Plaut et al. 1986; Plaut

and Hinton, 1987) found that using learning rates that were inversely proportional to neuronal fan-in speeded up learning by almost two factors.

There are alternative approaches that can be applied to accelerate the convergence speed of the learning algorithm by adapting the learning rate automatically. Rumelhart et al. (1986b) modified the delta (δ) rule (Equation 4.7) to include the momentum (α) term,

$$w_{ji}(t+1) = w_{ji}(t) + \alpha\Delta_p w_{ji} \tag{4.10}$$

where α is a constant positive number between 0 and 1 (Plaut and Hinton, 1987) that determines the update of the current weights using a fraction of the previous weight adjustment. The momentum term introduces the following effects: (1) causes acceleration when the error surface is continually sloping in a downhill direction by increasing the step size and (2) causes deceleration when the error surface is bumpy by decreasing the step size to dampen oscillations. An undesirable limitation of the momentum term is that it can cause the weights to be adjusted so that the slope(s) in the error surface are ascended (Jacobs, 1988). Furthermore, the user has the additional problem of finding the optimal combination of the η and α parameters, which absorbs computational resources in a meta-optimization phase (Battiti, 1989). Therefore, in this research only the learning rate has been empirically investigated.

4.5.5 Stopping Criteria

When training a multilayer ANN with the backpropagation learning algorithm it is generally difficult to determine when it is appropriate to stop the learning algorithm as there is no guarantee that it will converge. If the neural network is trained for too long i.e. too many epochs then the network is prone to over-fitting (overtraining) the training dataset causing poor generalisation performance on the test dataset. On the other hand, if training is stopped too early then the neural network is susceptible to under-fitting (undertraining) causing poor generalisation on the training and test datasets. Therefore, the choice of stopping criteria is important because the 'stopping criterion predominately involves a trade-off between training time (epochs) and generalisation error' (Prechelt, 2012:55).

The simplest stopping criterion is to halt the learning algorithm after a preset maximum number of epochs ($\text{epoch}_{\text{max}}$) have elapsed but this is most susceptible to over-fitting/under-fitting when used independently. Another stopping criterion is to

use a fixed threshold (Hush and Horne, 1993). A fixed threshold can be used on the measure of error so that the learning algorithm terminates when $E < E_{max}$. Commonly used error measures are the Mean Square Error (MSE), which is defined as

$$E_{mse} = \frac{1}{2}\sum_{p}\sum_{j}(d_{pj} - o_{pj})^2 \tag{4.11}$$

and the Root Mean Square (RMS) error, an extension of MSE, which is defined as

$$E_{rms} = 1/\sqrt{E_{mse}}. \tag{4.12}$$

Using a fixed threshold on E as the only stopping criterion is not adequate enough because the preselected E_{max} may be unattainable, causing an infinite loop on the learning algorithm (Kramer and Sangiovanni-Vincentelli, 1989). Therefore, to overcome the latter issue, the $epoch_{max}$ should be integrated in to the stopping criteria as shown in Figure 4.6, step vii. Kaastra and Boyd (1996) suggest determining the maximum number of epochs by plotting the error measure every epoch or at preset intervals (every n^{th} epoch) so that the $epoch_{max}$ can be set at the time where the error measure begins to plateau.

The most basic method for counteracting over-fitting is to stop training early by applying hold-out validation. In hold-out validation, the entire dataset is randomly split in to two independent subsets (two-way data split): (1) the training set, which is used to train the neural network in the learning algorithm and (2) the validation (test) set, which is unseen data used to measure the generalisation performance of the trained neural network. The learning algorithm continues to train the neural network with the training set until the error/performance measure deteriorates i.e. the MSE/RMS begins to increase on the test set or the classification accuracy decreases on the test set. It is common practice to have a two-way data split where 2/3 of the dataset is designated as the training set and 1/3 of the dataset as the test set (Stone, 1974). Alternatively, the entire dataset could be randomly split in to three independent subsets (three-way data split): (1) training set - used to train the neural network in the learning algorithm (2) validation set – stops the learning algorithm when the error measure increases on the validation set and (3) test set – used to evaluate the performance of the trained neural network. A limitation of hold-out validation is that large variance (Refaeilzadeh et al., 2009) can occur between the generalisation performance measure on the training and test sets.

A more widely employed technique for counterattacking over-fitting and evaluating or comparing neural network performance is cross-validation. There are various types of cross-validation: *k*-fold cross-validation, leave-one-out cross-validation and repeated *k*-fold cross-validation. In *k*-fold cross-validation (Stone, 1974) the entire dataset is randomly divided into *k* approximately equally sized subsets (folds). $k - 1$ subsets are used as the training set and the remaining subset is used as the test set. The cross-validation process is repeated *k* times, with each *k* subset being used once as the test set. $k = 10$ is typically used in *k*-fold cross-validation. A true estimate of error can be obtained by averaging the error measure results from the *k* folds. The estimate of accuracy, known as classification accuracy (CA) is calculated as the percentage of correct classifications (Kohavi, 1995), which can also be averaged from the *k* folds. Repeated *k*-fold cross-validation is when *k*-fold cross-validation is run a preset number of times e.g. 10 x 10-fold cross-validation = 100 ANNs trained and tested. The advantage of repeated *k*-fold cross-validation is that it can provide a more accurate measure of performance but at the cost of computational expense (Witten et al., 2011). To enhance *k*-fold cross-validation and repeated *k*-fold cross-validation, stratification can be applied, which is when the dataset is randomly split so that the contents of the training and test sets mirror the proportions of the classes in the entire dataset. Leave-one-out cross-validation is essentially *k*-fold cross-validation with *k* equal to the total number of instances in the dataset. The leave-one-out cross-validation technique is beneficial when the dataset is small as it yields the largest possible training dataset but it is computationally expensive and cannot be stratified (Witten et al., 2011). Kohavi (1995) compared common cross-validation methods and found stratified 10-fold cross-validation to be the best technique for model selection. In this research, cross-validation has been adopted because it overcomes the limitations of hold-out validation.

4.6 Neural Network Applications

Artificial neural networks have been utilised to tackle a broad range of problems such as pattern recognition, data compression, function approximation and process control (Hammerstrom, 1993; Widrow et al., 1994; Zhang, 2000). Hence, industrial and research-based applications of neural networks are numerous, spanning many disciplines. Widrow et al. (1994) provide a review of successful applications of neural

networks in industry, business and science, see references therein. Hammerstrom (1993) details how neural networks have been applied to optical character recognition, financial forecasting and process control in the workplace. The remainder of this section briefly examines four examples of backpropagation neural network applications, which emphasises their strong problem solving ability for providing solutions to a diverse range of challenging real-world problems. Some of the examples are directly related to the research conducted in this Thesis.

One of the earliest backpropagation ANN pattern recognition applications was NETtalk (Sejnowski and Rosenberg, 1987; Sejnowski and Rosenberg, 1988), a multilayer ANN trained to convert written English text to audible speech using only 203 inputs, 80 hidden neurons, 26 output neurons and 18,629 weights. NETtalk achieved a 78% testing classification accuracy on a small continuous informal speech word corpus containing 1024 words in the training set and 439 words in the test set. A well known perception system is ALVINN (Autonomous Land Vehicle in a Neural Network) (Pomerleau, 1989), which is a single hidden layer backpropagation ANN trained to autonomously keep the vehicle on the road using input from an attached video camera and a laser range finder. ALVINN's neural network, which consists of 1232 inputs (30x32 video input plus 8x32 laser range input), 29 hidden neurons and 45 output neurons, was trained on a dataset produced by a simulated artificial road generator. LeCun et al. (1990b) trained a four hidden layer backpropagation ANN to distinguish digits in a dataset of 9298 digitised normalised handwritten zip codes provided by the United States Postal Service and achieved 92% recognition accuracy. These image and pattern recognition examples demonstrate how versatile and powerful backpropagation ANNs can be.

A recent advance in the application of backpropagation ANNs is Silent Talker (Rothwell et al., 2006), which is a psychological profiling system that concurrently monitors multiple channels of human facial nonverbal behaviour to detect truthful and deceptive behavioural patterns using a bank of interconnected backpropagation neural networks. Through training and testing on digital recordings of participant interview responses to a simulated theft scenario where the participant was asked to steal/not steal money out of a box, Silent Talker was able to detect patterns of truthful and deceptive nonverbal behaviour with 80% of interviews correctly classified. A key feature of Silent Talker's interconnected neural network architecture is that there are

three distinct types of backpropagation ANNs that have been developed each with their own purpose:

(1) Object Locators: identify the location of facial nonverbal features e.g. left eye position.
(2) Pattern Detectors: categorise the state of the object identified by the object locator e.g. the left eye is gazing to the left.
(3) Classifier: decides whether the participant's response was truthful or deceptive based upon the states of the facial nonverbal behaviours collated over a predefined period of time.

Thus, the Silent Talker's backpropagation ANNs are used for image compression, image recognition and pattern recognition. Advantages of the Silent Talker system are that it is non-invasive, automatic and able to profile truthful/deceptive behaviour in near real-time. Disadvantages of the Silent Talker system is that the technology is offline, the interrelationships between the nonverbal channels used have not been analysed in depth and it has only been applied to a single laboratory based study in the field of deception detection. Stemming from the invention of Silent Talker is a patent published internationally on the method used for the "analysis of a human subject" (Bandar et al., 2002). As a result, there is the potential for adapting the method outlined in the patent to profile human behaviour in different fields of research. Therefore, there exists an unexplored niche for empirical investigations on the detection of human comprehension from multichannels of nonverbal behaviour using backpropagation neural networks, which this Thesis shall investigate.

4.7 Summary

The beginning of this chapter introduced the fundamentals of artificial neural networks and how they are able to replicate the naturally occurring behaviour of biological neural networks from the application of a learning algorithm, which harvests experiential knowledge in the ANNs interconnections. Multilayer neural networks with two sufficiently sized hidden layers are capable of approximating any arbitrary nonlinear continuous functions so have been used in this Thesis. The error-backpropagation learning algorithm, which is a well known supervised learning algorithm that uses gradient descent to search the weight space for an optimal solution, has been outlined. The key choices on the backpropagation learning

algorithm variations and parameter values have been reviewed in detail to ensure that appropriate training configurations are selected to optimise the feedforward multilayer neural networks to the problems presented in Chapters 6 and 7. Lastly, a plethora of industrial and research-based applications of backpropagation neural networks were identified, which revealed the uncharted opportunity of adapting Silent Talker's patented methodology for use within the human comprehension detection domain. In the following chapter a human comprehension detection system that uses a bank of interconnected backpropagation neural networks known as FATHOM is introduced.

Chapter 5 FATHOM: A Human Comprehension Detection System

5.1 Introduction

This chapter introduces FATHOM, a novel human comprehension detection software application that has been built to automatically detect human comprehension and non-comprehension from monitoring multichannels of nonverbal behaviour using ANNs. FATHOM builds upon an existing methodology on the analysis of the behaviour of a human subject using ANNs, which is disclosed in international patent number WO02087443 (Bandar et al., 2002) and was discussed in Section 4.6. A high-level overview of FATHOM's architecture is provided before introducing low-level descriptions of the distinct internal components within FATHOM and how they operate. Lastly, all of the channels monitored by FATHOM are defined and described within their respective categories.

Chapter 3 discovered that detection of human understanding is predominantly captured through written assessment and spoken language. Chapter 2 revealed that nonverbal behaviours are highly abundant but humans are poor manual decoders and judges of nonverbal behaviour. Furthermore, at present there is no automatic human comprehension detection system that detects human comprehension and non-comprehension by monitoring multiple nonverbal behaviours using ANNs. Therefore, the implementation of FATHOM attempts to address the latter niche. FATHOM overcomes the weaknesses of manual human coders by using ANNs. FATHOM investigates the unexplored niche by taking a nonverbal approach in human comprehension detection. Because the face has been found to be a prime source for indicators of human comprehension and non-comprehension in Section 3.5, FATHOM has a dominant focus upon facial nonverbal behaviours.

5.2 FATHOM Overview

FATHOM is a computer-based human comprehension detection system that uses a collection of interconnected backpropagation neural networks to simultaneously monitor multiple channels of human nonverbal behaviour from digital recordings. Figure 5.1 shows a screenshot of the FATHOM software application, which contains the

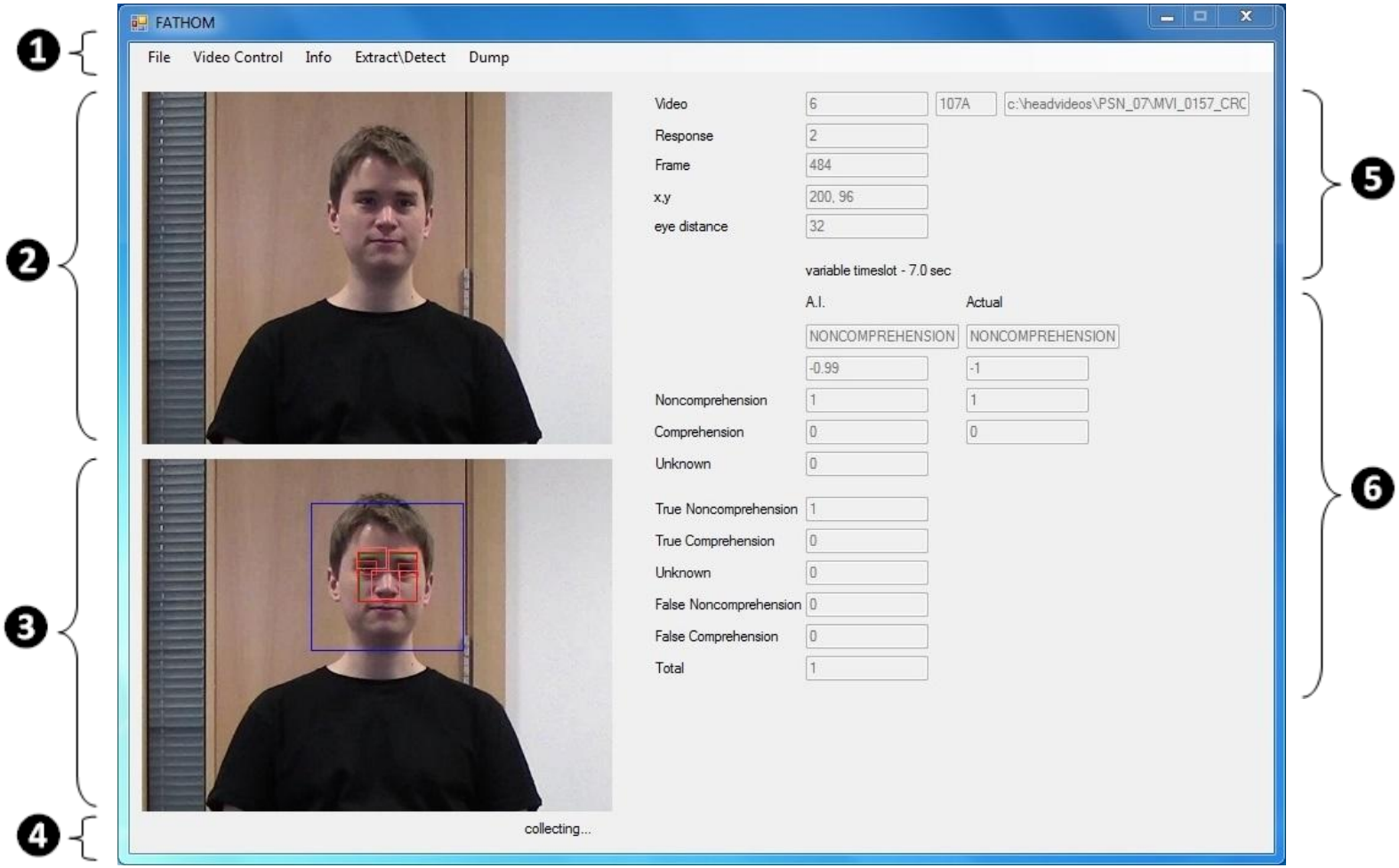


Figure 5.1 FATHOM

following key elements in the main graphical user interface (GUI): ❶ Menu; ❷ Video Window; ❸ Processing Window; ❹ Status Label; ❺ Video Information and ❻ Assessment Information. In Figure 5.1, FATHOM is automatically locating and processing the state of a person's facial nonverbal behavioural channels contained within the currently displayed video frame in the Processing Window. As the digital recording plays back in the Video Window, FATHOM analyses the person's facial nonverbal behaviours on a frame-by-frame basis for a predefined period of time. The predefined period of time is referred to as a timeslot, which is measured in seconds and can be fixed e.g. every *n* second(s) or variable in length. During the processing of each video frame, FATHOM's backpropagation neural networks search for the face and facial features within the search box, which is the blue box in the Processing Window. The red boxes in the Processing Window identify where FATHOM's backpropagation neural networks have located the face, eyes, eyebrows and the nose in the video frame. The Status Label displays the message "collecting..." only when FATHOM is able to successfully detect a sufficient number of facial features within a video frame during the timeslot. The Video Information section provides information about the video such as the frame number currently displayed in the Video Window, the response number under analysis, the face position and the horizontal distance between the eyes. At the end of each timeslot, FATHOM produces a comprehension classification within a scaled bipolar range, where +1 represents the detection of comprehension and -1 represents the detection of non-comprehension. The comprehension classification result is displayed in the Artificial Intelligence (AI) section of the Assessment Information along with the timeslot classification counters, which are incremented accordingly. In the Assessment Information, the Actual section shows the desired comprehension classification, which enables a comparison between FATHOM's neural network comprehension classification (AI) and the desired comprehension classification (Actual) over all encountered timeslots.

FATHOM has been implemented as a Windows Form Application using the C# object-oriented programming language, .NET Framework (Microsoft, 2016a) and the Microsoft Media Foundation (MF) Application Programming Interface (API) (Microsoft, 2016b) on the Windows platform. The Microsoft DirectShow API (Microsoft, 2016c) was not used for video handling because it has been superseded by the Microsoft MF API. The purpose of the FATHOM software application is two-fold: (1) to test for

human comprehension and (2) to extract datasets for training the backpropagation neural networks contained within FATHOM's architecture, which is outlined in Section 5.3.

5.3 Architecture

FATHOM's system architecture is displayed in Figure 5.2. The main functional elements of FATHOM's system architecture are the GUI, the Image Processing Module and the Neural Network Architecture. The end user interacts with FATHOM via the GUI (Figure 5.1) using a mouse and keyboard, which was outlined in Section 5.2. The GUI is driven by a menu, which allows the end user to easily configure the software application for video analysis. During video analysis the GUI provides continuous visual feedback to the end user. Video input is supplied from a video file or a live digital camcorder feed. Throughout this Thesis, the video input was streamed directly from video files. The image processing module is responsible for video handling, executing search algorithms that track facial objects, updating the GUI and it can accumulate channel datasets as described in Section 5.3.2. The neural network architecture (Figure 5.3) in Section 5.3.1 is responsible for image compression, recognising facial features and classifying human comprehension. Communication occurs between the image processing module and the neural network architecture. FATHOM is able to output video analysis results and channel datasets to flat file.

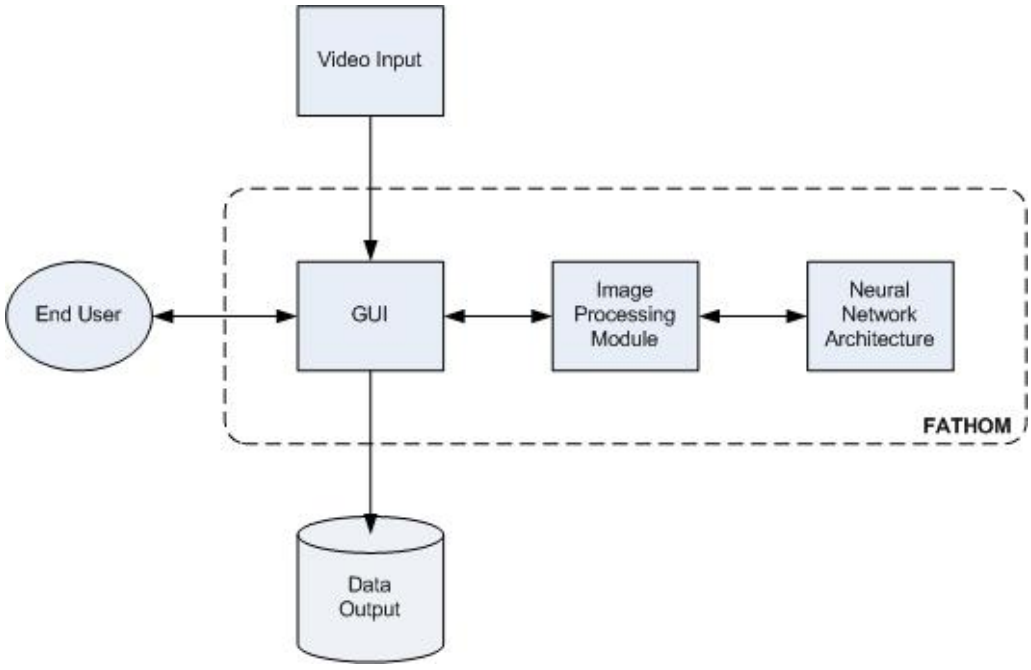


Figure 5.2 FATHOM System Architecture

5.3.1 Neural Networks

FATHOM's neural network architecture is shown in Figure 5.3. There are three distinct types of backpropagation neural networks contained within FATHOM: (1) Object Locators (2) Pattern Detectors and a (3) Comprehension Classifier. Each of the three distinct backpropagation neural network types has their own purpose and are utilised at predefined stages in FATHOM's human comprehension detection process. Firstly, the Object Locator ANNs are used to identify the location of visual nonverbal facial objects i.e. features such as the face, eyes, eyebrows and the nose. The topologies of the four Object Locator ANNs used in FATHOM are displayed in Figure 5.4. Each Object Locator ANN has been derived through the combining of an independently trained Global Compression ANN and a Feature Classifier ANN. The primary function of the Global Compression ANN is to compress the original image presented as a vector at the input layer in the single hidden layer of neurons. The compressed image contained in the single hidden layer can then be decompressed at the output layer, where the number of output neurons is always equal to the number of inputs. The benefit of using neural networks for image compression is that they provide optimised approximations (Cramer, 1998), which reduce the dimensionality of the problem thus lowering complexity and computational cost. Compression ANNs have been successfully applied in studies where human faces were extracted from digital images e.g. SEXNET (Golomb et al., 1990) and EMPATH (Cottrell and Metcalfe, 1990). This Thesis utilises the same basic technique to image compression ANN topology construction as found in EMPATH (Cottrell and Metcalfe, 1990).

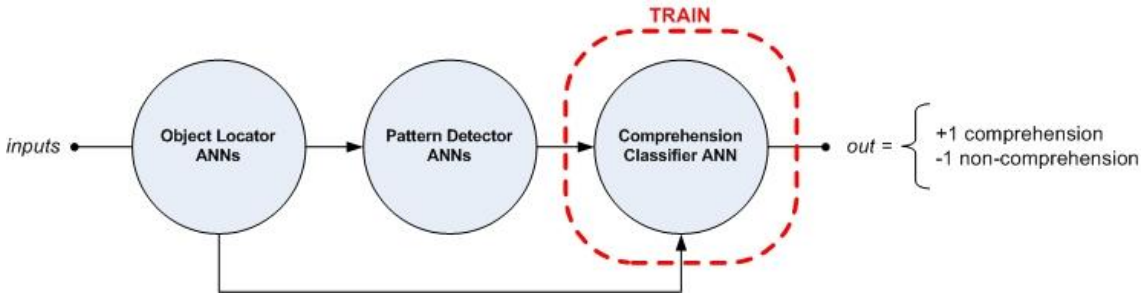


Figure 5.3 FATHOM's Neural Network Architecture

The Feature Classifier ANNs hidden layer is connected directly to the Global Compression ANNs hidden layer, thus forming a single two hidden layer neural network known as the Object Locator ANN. The joining of the Feature Classifier ANN to

the Global Compression ANN is only possible because the number of inputs to the Feature Classifier ANN is equal the number of hidden neurons in the Global Compression ANN. The primary function of the Feature Classifier ANN is to recognise whether the object is present in the compressed image presented at its input layer e.g. the Face Feature Classifier ANN outputs +1 if the compressed image was recognised as being face otherwise outputs -1 to indicate not a face. The Face Object Locator ANN in Figure 5.4 has topology of 120:14:6:1, which was formed by discarding the Global Compression ANNs (topology 120:14:120) output layer (120 outputs) and connecting the Global Compression ANNs hidden layer directly to the Feature Classifier ANN (topology 14:6:1) input layer.

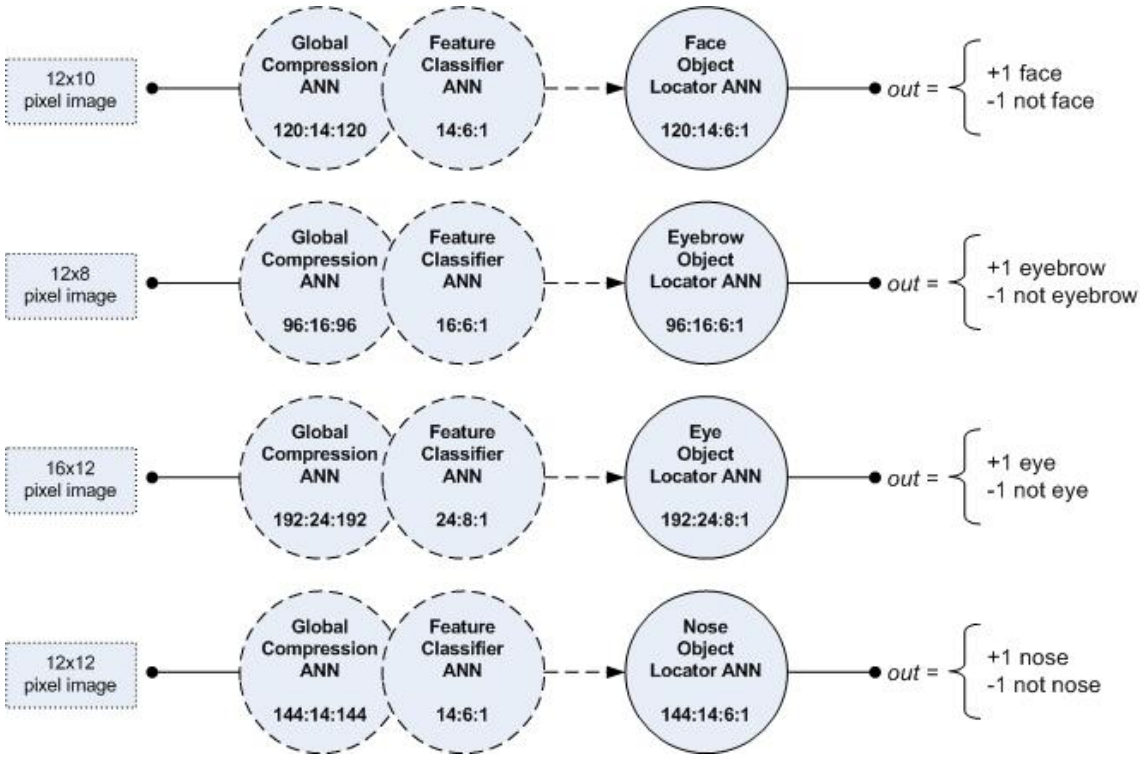


Figure 5.4 Object Locators

In Figure 5.1, the red boxes in the Processing Window (❸) illustrate where each of the four Object Locator ANNs have successfully identified all of the facial objects in a video frame: face, left eye, right eye, left eyebrow, right eyebrow and the nose. To ensure that the Object Locator ANNs reliably find objects, tolerance values have been assigned (Table 5.1). Higher tolerance values determined by Rothwell (2002) are used on the Face and Eye Object Locator ANNs due to the importance of the objects being found.

Table 5.1 ANN Tolerance Values

ANN Type	Tolerance	
	High	Low
Face Object Locator	0.9997	0.99
Eye Object Locator	0.9997	0.99
Eyebrow Object Locator	0.9997	0.8
Nose Object Locator	0.9997	0.8
Comprehension Classifier	+/-0.95	N/A

The Pattern Detectors ANNs are used to identify the state of the object found by its parent Object Locator ANN e.g. determining whether the right eye is closed. Only eye Pattern Detector ANNs have been used in this research. The Pattern Detector ANNs in FATHOM are: eye fully closed (eyefclosed), eye fully left (eyefleft), eye fully right (eyefright), eye half closed (eyehfclosed), eye half left (eyehfleft) and eye half right (eyehfright). The topologies of the six Pattern Detector ANNs are displayed in Figure 5.5. Each Pattern Detector ANN has been derived through the combining of an independently trained Global Compression ANN and a Pattern Classifier ANN. The Pattern Classifier ANNs hidden layer is connected directly to the Global Compression ANNs hidden layer, thus forming a single two hidden layer neural network known as the Pattern Detector ANN. The primary function of the Pattern Classifier ANN is to identify the state of the object in the compressed image presented at its input layer e.g. the eyefclosed Pattern Classifier ANN outputs +1 if the compressed eye image was recognised as being an eye fully closed otherwise outputs -1 to indicate not an eye fully closed. All of the Pattern Detector ANNs have the same neural network topology size of 192:24:10:1. The advantage of using Pattern Detector ANNs is that once they are trained they are more consistent at classifying the state of multiple objects than human coders who are susceptible to fatigue and have limited processing capabilities in comparison to machines. A structural similarity between the Object Locator ANNs and the Pattern Detector ANNs is that they all possess a Global Compression ANN within their topology. However, because not all Object Locator ANNs have an associated Pattern Detector ANNs, in Figure 5.3 there is an arrow that directly connects to the Comprehension Classifier ANN.

The final ANN in FATHOM's neural network architecture is the Comprehension Classifier. The Comprehension Classifier ANN classifies the input vector for each timeslot within a bipolar range, which indicates whether the person was

comprehending (+1) or not comprehending (-1) during the period of time under assessment. The input vector to the Comprehension Classifier is derived from the collation of the monitored channels from the Object Locator ANNs outputs, the Pattern Detector ANNs outputs, geometrical calculations, logical expressions and constant values from flat file. The channels that FATHOM monitors and how they are calculated are discussed in depth in Section 5.4. To ensure that the Comprehension Classifier ANN reliably classifies comprehension and non-comprehension, only a high tolerance value has been assigned (Table 5.1).

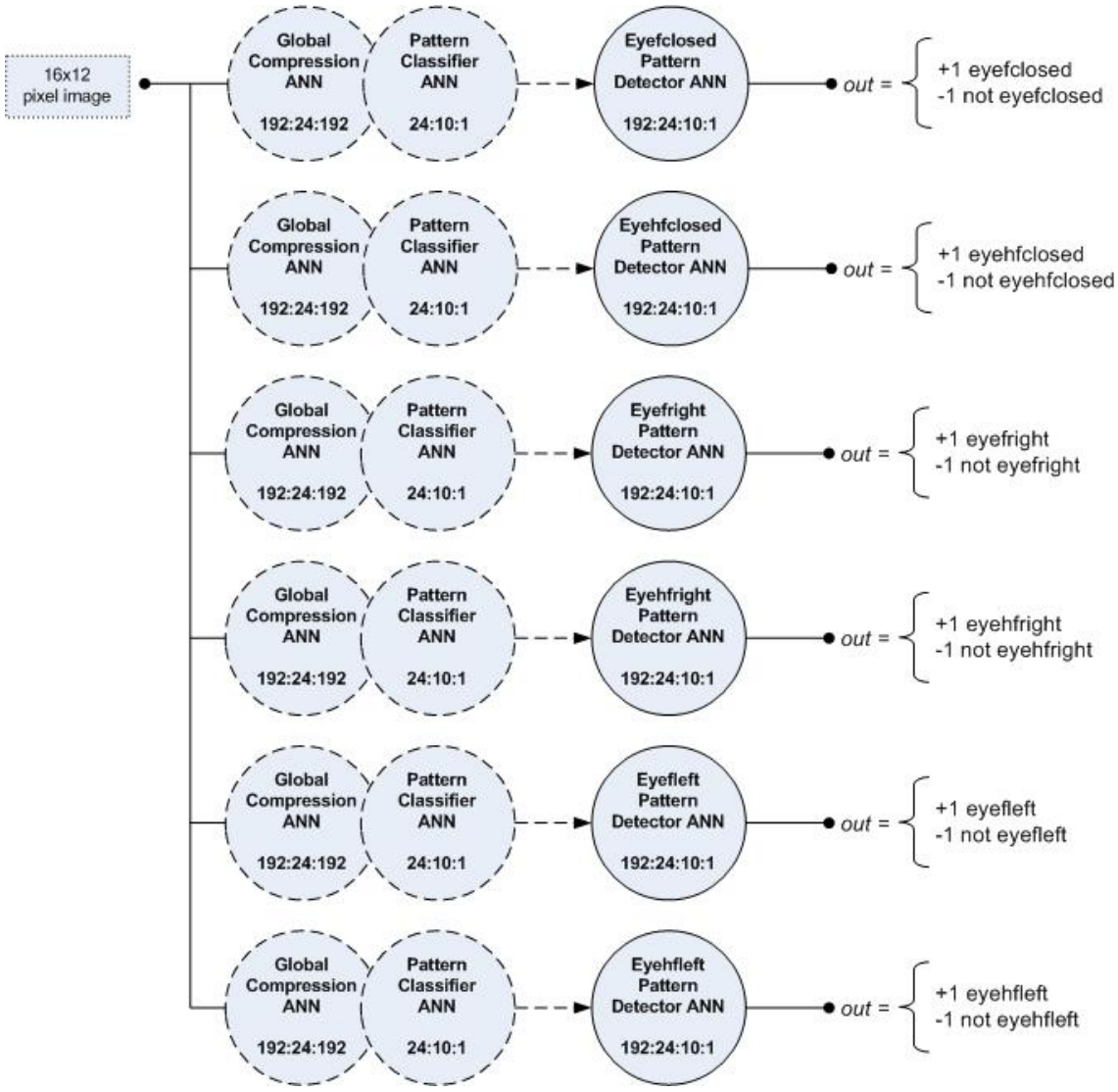


Figure 5.5 Pattern Detectors

The primary focus of this research is on the development of FATHOM's Comprehension Classifier ANN (Figure 5.3) topology through training and validation with datasets contained in Chapters 6 and 7. The optimised backpropagation Object Locator ANNs and the Pattern Detector ANNs have all been absorbed in to FATHOM

from Rothwell's (2002) earlier doctoral research. This has been taken advantage of because the optimised ANNs should be able to continue to robustly recognise all of the generic facial objects (i.e. face, eye, eyebrow and nose) due to still using human subjects in the new experimental datasets and from the characteristic of neural networks having the ability to generalise from previous experience.

A software-based application was built to provide an environment for constructing FATHOM's Comprehension Classifier ANN (Figure 5.3) through training and validation with the datasets contained in Chapters 6 and 7. Figure 5.6 shows the FATHOM neural network training application, which has been used to optimise FATHOM's Comprehension Classifier ANN using the backpropagation learning algorithm and cross-validation. The implementation of FATHOM's neural network training application was built in the object oriented C# programming language as a Console Application on the Microsoft Windows platform.

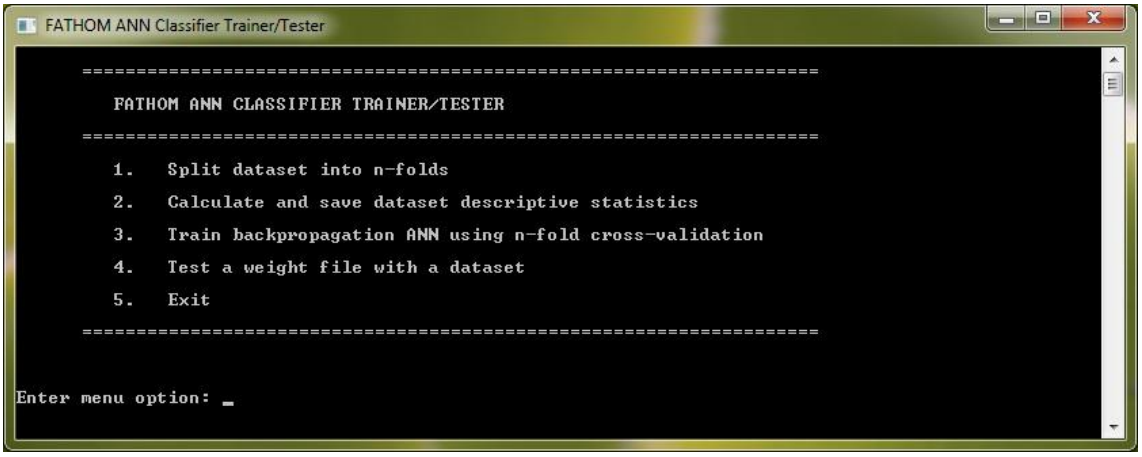


Figure 5.6 FATHOM Neural Network Trainer

5.3.2 Detection Process

The Data Flow Diagram (DFD) in Figure 5.7 provides an overview on how FATHOM (Figure 5.1) automatically processes a digital video containing a single human subject in order to detect human comprehension from the tracked channels using the neural networks in Section 5.3.1. FATHOM processes digital videos from a direct live digital camcorder feed or from video file with a fixed frame dimension of 384 x 288 pixels. The following subsections describe specific elements and individual processes, which FATHOM uses in order to: (1) locate, examine and track the facial objects; (2) detect and classify human comprehension.

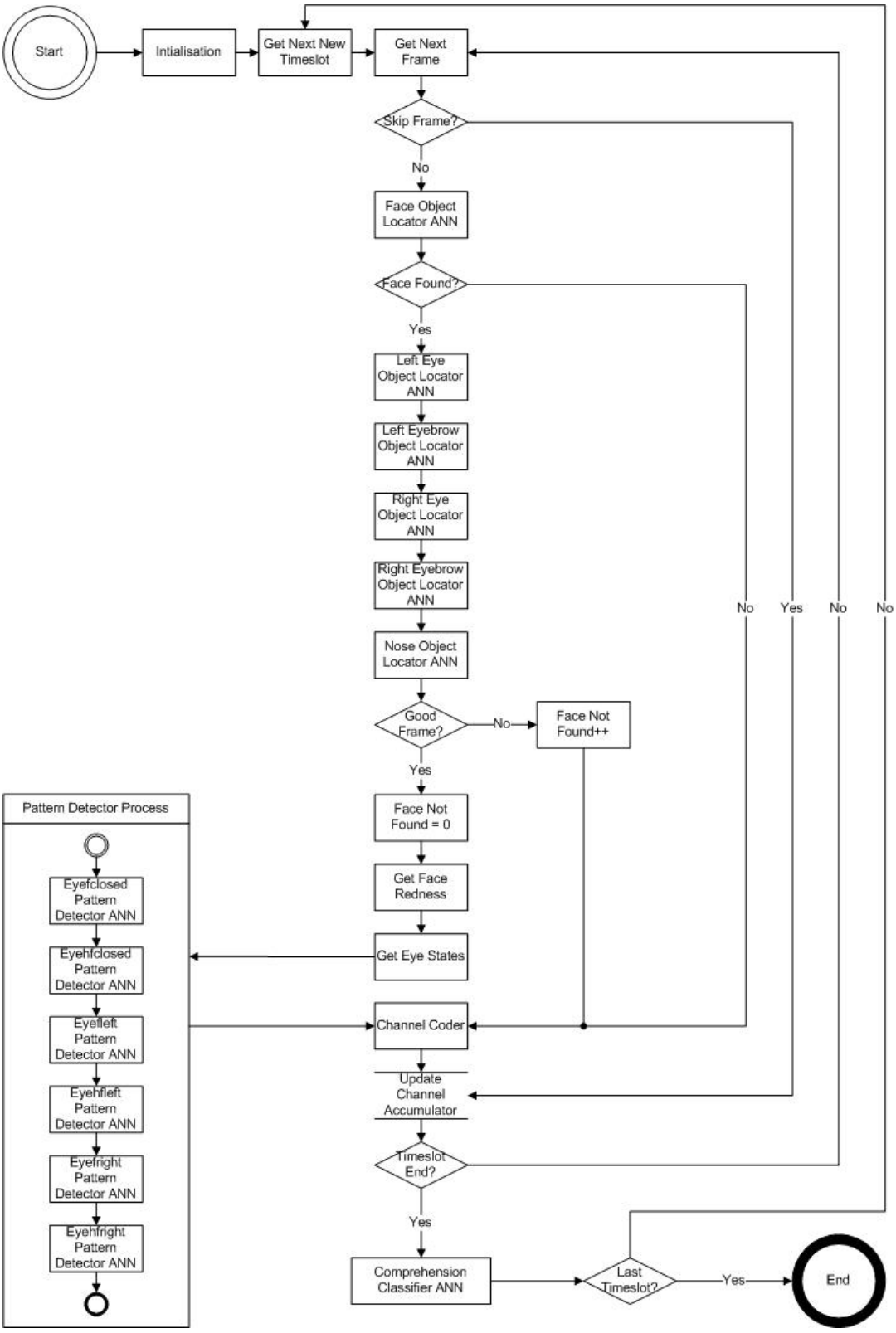


Figure 5.7 FATHOM Architecture Data Flow Diagram

Search Area

In Figure 5.1, the Search Area Window is the outer blue box painted on the video frame shown in the Processing Window (❸). It defines the search area boundaries within which the Object Locator ANNs are permitted to operate their search for each of the facial objects. The dimension of the Search Area Window is determined in three different ways: (1) using predefined default dimensions when no face location knowledge is present i.e. on initialisation; (2) using the scale and location of the face found with border parameters when the face was found in the previous frame; or by (3) increasing the current dimensions when the face had not been found in the previous frame, but had been found recently. Figure 5.8 shows how the Search Area Window dimensions are formed from the border parameters (border width and height), which are predefined in Rothwell (2002) and relatively extend from the facial objects extremity. Therefore, the larger the facial object, the larger the border. Using a Search Area Window with dimensions equivalent to the size of the whole video frame (384 x 288 pixels) would be disadvantageous, as it would make the Object Locator ANNs search for the facial objects extremely time-consuming. Therefore, due to participants being seated at a fixed distance from the digital camcorder with their head in the centre of the top third of the video frame, only a small Search Area Window initialised within this zone was required.

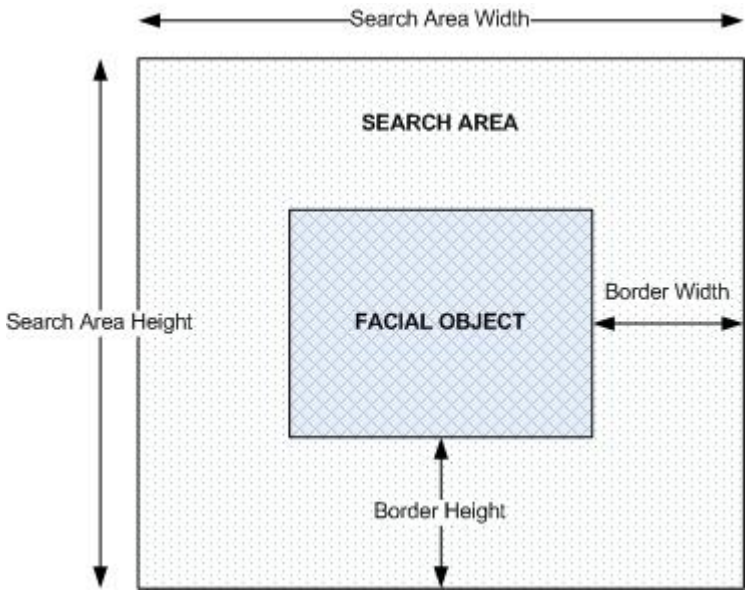


Figure 5.8 Object Search Area Window

Face Search

The face is searched for inside the grey scaled (range 0-255) Search Area Window by systematically presenting small sections of the Search Area Window to the Face Object Locator ANN for assessment. A clockwise spiral search, moving outwards from the centre of the Search Area Window to its perimeter was used to systematically obtain different sections of the Search Area Window for the Face Object Locator ANN to then identify whether and where the face was located. The Spiral Search is one of the basic crime scene search patterns used by crime scene investigators (Baxter, 2015). The step size of the spiral search (*n* pixels) was determined by the face scale found and the radius of the Search Area Window e.g. a large step size is used when the scale and the radius are large. Because the face could be at any scale within the Search Area Window, the small sections of the Search Area Window were tested at different scales at each position in the spiral search e.g. a scale value of 100% relates to a face image size of 12 x 10 pixels and a scale value of 400% relates to a face image size of 48 x 40 pixels. The face search terminates when the spiral search reaches the edge of the Search Area Window or when an acceptable face has been found using the Face Object Locator ANN high and low tolerance values in Table 5.1. When the Face Object Locator ANN's network output was above the high tolerance value (0.9997) then the spiral search terminated immediately to save time. Otherwise, the best face location was determined from the Face Object Locator ANN's best network output found at the end of the spiral search, which was over the low tolerance value (0.99).

All of the small sections (snippets) of the Search Area Window are pre-processed prior to being presented to the Face Object Locator ANN, using some of the input data preparation techniques discussed in Section 4.5.1. Firstly, block averaging is applied to the grey scaled snippet to reduce the size of the image to the fixed size of 12 x 10 pixels. Using block averaging to scale down the snippets has the benefit of reducing the dimensionality of the problem. In earlier postgraduate research, Adolph (1998) demonstrated that 12 x 10 pixel images were large enough for backpropagation ANNs to robustly distinguish face images from non-face images. Furthermore, by using 12 x 10 pixel images, the small image dimensions have the advantages of reducing data pre-processing time and ANN processing duration. Secondly, histogram equalisation is applied to the grey scaled snippet to improve image contrast by using a wider, more uniform distribution of intensity values i.e. it ensures that the intensity values are

spread over the entire grey scale range (0-255). Lastly, the snippets pixels are all normalised in to the bipolar range [-1,1] and transformed into a 120-element vector, ready for immediate presentation to the Face Object Locator ANN's input layer (Figure 5.4).

Once the face has been found, it allows approximations on the locations of other facial objects (i.e. eyes, eyebrows and nose) being found in the Feature Searches, which is outlined in the following subsection. Furthermore, when the next frame is loaded, the Face Search implements a Search Area Window at the location where the face was found in the previous frame with dimensions based upon the face scale found in the previous frame. Otherwise, if the face is not found for a predefined number of consecutive frames then a fixed number of consecutive frames are skipped and the Search Area is reset to the default parameters.

Feature Searches

After the Face Search has located the face object in the frame, the Feature Searches can commence. The purpose of the Feature Searches is to locate the position of the remaining facial objects i.e. the left eye, right eye, left eyebrow, right eyebrow and the nose. The search for the facial objects was conducted in a similar manner to the Face Search, using the same image pre-processing techniques and the outward clockwise spiral search. Due to the face having been located, it helped to determine the start positions for initiating the searches at estimated locations for each of the facial objects. Each facial object search has its own Search Area Window where the dimensions are determined in the same manner as the Search Area Window for the face i.e. using the facial object's dimensions and border parameters as shown in Figure 5.8. The first facial object to be searched for is the left eye. If the left eye is found, then the left eyebrow is immediately searched for. The next facial object to be searched for is the right eye. If the right eye is found, then the right eyebrow is immediately searched for. The grey scaled left/right eye images were scaled down to 16 x 12 pixels and transformed into a 192-element vector for presentation to the Eye Object Locator ANN's input layer (Figure 5.4). The grey scaled left/right eyebrow images were scaled down to 12 x 8 pixels and transformed into a 96-element vector for presentation to the Eyebrow Object Locator ANN's input layer (Figure 5.4). Because the Eye Object Locator ANN was originally validated on left eye images, the pixels in the right eye image are mirrored prior to presentation to the Eye Object Locator ANN (Figure 5.4). The latter

approach is also utilised for the detection of the right eyebrow with the Eyebrow Object Locator ANN (Figure 5.4). The last facial object to be searched for is the nose. The grey scaled nose images were scaled down to 12 x 12 pixels and transformed into a 144-element vector for presentation to the Nose Object Locator ANN's input layer (Figure 5.4). The facial objects high and low tolerance values in Table 5.1 were used to determine whether acceptable facial objects had been found by the Object Locators ANNs. Simple, logical positional rules were adopted in order to aid the search for the location of the facial objects e.g. the left eyebrow is above the left eye and both eyes are above the nose.

Once the Feature Searches have terminated, an assessment is made to determine whether a sufficient number of facial objects were found in the video frame. The assessment helps to ensure that the face and facial features have been positively identified. If all facial objects have been found except for the nose or one eyebrow then the frame is regarded as 'good', the Face Not Found counter is zeroed (Figure 5.7) and the Status Label displays "collecting..." in Figure 5.1. Otherwise, the frame is regarded as 'bad' and the Face Not Found counter is incremented if the face was not located. Once a frame is categorised as 'good', the scaled left/right eye image are transformed into a 192-element vector for presentation to each of the six Pattern Detector ANNs (Figure 5.5). The scaled eye images have the largest dimensions (16 x 12 pixels) in comparison to the other scaled facial object images because it helped to provide a more accurate Eye Object Locator ANN and ensures that there is sufficient detail present for the Pattern Detector ANNs (Figure 5.5) to accurately determine the state of each eye e.g. both the left eye and right eye are closed. At the end of the Feature Searches, the knowledge about the presence, location and states of each facial object is transferred to the Channel Coder for further inspection, which is outlined in the next subsection.

Channel Coder

The Channel Coder extracts channel data from the face and facial objects found by the Face Search and the Feature searches for the current video frame under analysis. The extraction of the channel data varies in approach and complexity. All of the channels extracted by FATHOM are described in detail in Section 5.4. The approaches to obtaining channel data were achieved by using one or a combination of the following techniques: relative object positions, object position changes, Pattern

Detector network output(s), object image pixel functions, simple logical rules and the relationship with channel data from the previous frame(s). For example, to obtain the face blushing/blanching channels, an object image pixel function is applied on the red component of the face object's image pixels to quantify face redness, which is then compared to the previous frames face redness value to determine whether blushing or blanching occurred. Once all of the channels have been extracted and normalised, they are passed on to the Channel Accumulator for further processing.

Channel Accumulator

The Channel Accumulator collates the channel data from all frames contained within the duration of the current timeslot, which may be fixed or variable in length. In this Thesis, only a fixed one second timeslot at 15 frames per second (fps) was used. Once all channel data has been collated from the given timeslot, the Channel Accumulator produces a single, normalised vector known as grouped channel data, which represents the channel data statistics for all extracted channels within the given timeslot. The grouped channel data vector is formatted as

$[\ participant\ number, frame\ number, channel_1, \dots, channel_n, desired\ response\ ]$.

The approach to consolidating the channel data statistics collected for each channel in to a single, normalised value varied slightly, depending upon the complexity of the channel i.e. the averaging (mean) and scaling (multiply by two then minus one) formula has a time related minimum (min) and maximum (max) value, which are both predetermined by the channel type. All channels have a min value of 0 per second. The max value was 2 per second for eye blinks, 5 per second for eye shifts and fps for all other channels. Thus, the maximum and minimum number of times an eye blink could occur in 1 second was two times and never, respectively. The grouped channel data vector was only considered as being 'valid' when channel data was extracted from ≥ 90% of the frames within the given timeslot. Only 'valid' grouped channel data vectors are presented to the Comprehension Classifier ANN, to ensure reliability.

In Chapters 6 and 7, the Channel Accumulator plays a critical role in development of FATHOM's Comprehension Classifier ANN. The grouped channel data vectors that the Channel Accumulator produces during the frame-by-frame processing of digital recordings are collated in to Comma-separated values (CSV) files. The CSV files later formed the training and validation datasets for optimising FATHOM's Comprehension Classifier ANN using FATHOM's neural network training application in Figure 5.6.

Comprehension Detection

The Comprehension Classifier ANN receives the grouped channel data vector at its input layer for processing. The Comprehension Classifier ANN has a single output neuron, which outputs a response between 1 and -1. A positive response indicates that human comprehension has been detected and a negative response indicates that human non-comprehension was detected. To ensure that the Comprehension Classifier ANN reliably detects human comprehension, a high tolerance value of +/- 0.95 has been implemented (Table 5.1). The response from the Comprehension Classifier ANN is displayed in the AI section of FATHOM's Assessment Information (❻) in Figure 5.1. Because the desired response from the grouped channel data vector is shown in the Actual section of ❻ in Figure 5.1, it allows immediate deductions to be made about the performance of the Comprehension Classifier ANN through comparisons with the AI results, on a timeslot by timeslot basis. Furthermore, counters in the lower section of ❻ in Figure 5.1 keep track of vectors classified as true positive (true comprehension), true negative (true non-comprehension), false positive (false comprehension) and false negative (false non-comprehension). Basically, true positives/negatives are correct results, false positives/negatives and unknown are incorrect results.

5.4 Channels

FATHOM monitors forty channels, which are categorised into: face channels (Section 5.4.1), eye channels (Section 5.4.2) and known channels (Section 5.4.3). The review in Section 3.5 repeatedly revealed that facial nonverbal channels play a critical role in human comprehension and non-comprehension detection. However, the review also identified that previous studies on human comprehension detection have focused on few fine-grained nonverbal channels using an automated multichannel approach. Therefore, the channels used in Silent Talker (Rothwell, 2002; Rothwell et al., 2006; Rothwell et al., 2007) have all been integrated into FATHOM. The following subsections introduce the channels and describe how they are derived from the video frame(s). The channels vary in complexity. All of the channels are coded in to the bipolar measurement range [-1, 1] by the Channel Coder/Accumulator (Figure 5.7) and ultimately feature in the $channel_1 ... channel_n$ section of the grouped channel vector. In

Chapter 7, four new known channels have been developed and integrated into FATHOM.

5.4.1 Face Channels

Within FATHOM there are twenty face channels, which cover three main areas: face movement, face angle and face redness. The face movement channels track face movement along the X-axis and Y-axis using the coordinates and dimensions of the face found by the Face Object Locator ANN. The face movement channels are: face vertical movement (fvm), face horizontal movement (fhm), face upward movement (fum), face downward movement (fdm), face left movement (flm), face right movement (frm), face scale (fs), face forward movement (ffm), face backward movement (fbm), face vertical shift (fvs), face horizontal shift (fhs), face vertical shift with noise (fvsn) and face horizontal shift noise (fhsn). The pseudocode in Figure 5.9 shows how the Y-axis coordinates for the current and previous face positions are used to determine whether fvm, fum and fdm occurred in the current frame using a basic formula and a set of logic rules. Horizontal movement of the face is determined in the same manner as vertical movement (Figure 5.9) but using the X-axis coordinates for the current and previous face positions to determine whether fhm, flm and frm occurred in the current frame.

```
initialise verticalMovement = currentFacePosition.Y - previousFacePosition.Y

if verticalMovement > 0 then
       fvm = true
       fum = true
       fdm = false
else if verticalMovement == 0 then
       fvm = false
       fum = false
       fdm = false
else
       fvm = true
       fum = false
       fdm = true
end if
```

Figure 5.9 Vertical Movement Pseudocode

The face scale (fs), face forward movement (ffm) and face backward movement (fbm) channels all use the horizontal distance between the left and right eye X-axis coordinates to determine the scale change in face size due to movement across frames. The scale change is obtained using the current and previous eye distances in

Equation 5.1. Scale values greater than one are clamped at the maximum output range of one.

$$scaleChange = \frac{(currentEyeDistance - previousEyeDistance) \times 10}{previousEyeDistance} \quad (5.1)$$

The scale change value, is then incorporated and normalised in to the range [-1, 1] in Equation 5.2 to produce the fs value. A 10% change in fs corresponds to the maximum value of one and no fs change corresponds to zero. Furthermore, a positive fs value indicates ffm and a negative fs value indicates fbm.

$$fs = (scaleChange \times 2) - 1 \quad (5.2)$$

The face vertical shift (fvs) and the face vertical shift with noise (fvsn) channels both determine movement of the face transitioning upward and then downward or downward then upward. The face horizontal shift (fhs) and the face horizontal shift with noise (fhsn) channels both determine movement of the face from the left and then to the right or from the right and then to the left. The vertical shift channels (fvs and fvsn) compute the differences of the face found Y-axis coordinate between the current and previous frame(s) to identify whether a vertical face shift has occurred. The horizontal shift channels (fhs and fhsn) compute the differences of the face found X-axis coordinate between the current and previous frame(s) to identify whether a horizontal face shift has occurred. The fvs and fhs channels both check the face found axis coordinates for the previous three consecutive frames whereas the fvsn and fhsn only check the previous frame. Thus, enabling the capture of slow and fast head shifts. All of the face shift channels produce a 1 when a face shift occurs and a -1 when a face shift has not occurred.

The face angle channels are: face movement clockwise (fmc), face movement anti-clockwise (fmac), face movement angle-change (fma), face movement up-on-right (fmuor) and face movement up-on-left (fmuol). The fmuor and fmuol determine the current slope direction of the face i.e. the gradient of the head sloping in the uphill (positive value) or downhill (negative value) direction. Therefore, to obtain the gradient direction of the face, the vertical and horizontal difference between the left and right eye coordinates is divided. The fma, fmc and fmac channels compute changes in face rotation (face angle) using the gradient value. Fma determines the change in face angle by obtaining the difference between the current and previous face gradients, multiplying by five then normalising in to the bipolar measurement

range. Hence, a change in eye vertical distance equating to 20% of the eye horizontal distance results in the fma channel maximum value. Clockwise and anti-clockwise face rotation is determined from the change in face angle by obtaining the difference between the current and previous face gradients, multiplying by ten then normalising in to the bipolar measurement range. A positive value indicates fmc and a negative value indicates fmac. A 10% change in face rotation results in the fmc or fmac channel maximum value.

The face redness channels are face blush (fblu) and face blanch (fbla). The fblu channel tracks increases in face redness and fbla tracks decreases in face redness. A baseline of face redness is retrieved from the face image found in the first video fame. Thereafter, only changes in face redness are computed and recorded in the fblu and fbla channels. The change in face redness is determined by multiplying the factional face red change by five then normalising in to the bipolar measurement range. Hence, a 20% change in face redness results in the fblu or fbla channel maximum value.

5.4.2 Eye Channels

There are sixteen eye channels within FATHOM, which are broken down in to eight channels per eye. The left eye channels are: left eye closed (lclosed), left eye blink (lblink), left eye looking left (lleft), left eye looking right (lright), left eye shift (lshift), left eye half looking left (lhleft), left eye half looking right (lhright) and left eye half closed (lhclosed). The right eye channels are: right eye closed (rclosed), right eye blink (rblink), right eye looking left (rleft), right eye looking right (rright), right eye shift (rshift), right eye half looking left (rhleft), right eye half looking right (rhright) and right eye half closed (rhclosed). The state of each eye channel is determined from a Pattern Detector ANN (Figure 5.5) observing the left/right eye image and/or from the application of logical decision(s). The approach to detecting the right eye channels is the same as the left eye channels but the right eye image is reflected (mirrored) prior to analysis. Therefore, only the left eight eye channels are subsequently described.

The first eye channel to be processed is the lclosed channel because other left eye channels rely upon the result in their logical decisions. The left eye image is presented to the eyefclosed Pattern Detector ANN (Figure 5.5), which outputs a value within the bipolar measurement range. A positive value corresponds to eyefclosed (1) and a negative value corresponds to not eyefclosed (-1). The lblink channel is then

determined using a logical decision i.e. if the previous left eye was closed and the current left eye is not closed then a left eye blink occurred (lblink = 1) otherwise a blink did not occur (lblink = -1). In the Channel Accumulator the left eye and right eye blinks are both restricted to a maximum of 2 per second and a minimum of 0 per second. The lleft, lright, lhleft and lhright channels are all determined by presenting the left eye image to their associated Pattern Detector ANN in Figure 5.5, if the left eye is not closed. For example, if the left eye is closed then the lleft channel set to -1 because the left eye is not looking left. On the other hand, if the left eye is not closed then the left eye image is presented to the eyefleft Pattern Detector ANN (Figure 5.5) where a positive output value corresponds to lleft (1) and a negative value corresponds to not lleft (-1). The lshift channel determines movement of the left eye from the left and then to the right or from the right and then to the left. Therefore, the lshift channel cannot be computed until the lleft and lright channels have been processed because it relies upon them in its logical decision. If the current and previous left eye images are both looking in the same direction (left, central or right) then no left eye shift occurred, otherwise a left eye shift occurred. In the Channel Accumulator the left eye and right eye shifts are both restricted to a maximum of 5 per second and a minimum of 0 per second. The last eye channel to be analysed is lhclosed. The lhclosed channel is obtained by presenting the left eye image to the eyehfclosed Pattern Detector ANN (Figure 5.5), where a positive output value indicates eyehfclosed and a negative value indicates not eyehfclosed.

5.4.3 Known Channels

In FATHOM there are four known channels: sex, race, planning and slot. All of the known channels (apart from planning and slot) are text channels, which are obtained from each participant during the data collection phase of each study in Chapters 6 and 7. The known channels were collated for demographic analysis and because Machida (1986) identified differences in patterns of nonverbal behavioural displays of comprehension and non-comprehension based on sex and ethnicity. The sex channel recorded whether the person in the video was male (1) or female (-1). The race channel contained the origin of the person in the video e.g. European (1) or non-European (-1). The planning channel, which is used to distinguish the study condition(s) that the participant participated in. The slot channel represents the current timeslot

(Section 5.2) length in the bipolar measurement range. The sex and slot channels are the only channels, which maintain a constant value after Initialisation (Figure 5.7) throughout all frames for each given video. The reason why the slot channel remains constant is because only a fixed one second timeslot was used in the studies contained within this Thesis. If the timeslot had been set to variable then the slot channel would have not remained constant.

5.5 Conclusion

This chapter has introduced FATHOM a novel human comprehension detection system, which uses a set of neural networks to detect human comprehension from multiple channels of nonverbal behaviour. Within FATHOM's architecture there are three distinct types of neural networks: Object Locator ANNs, Pattern Detector ANNs and the Comprehension Classifier ANN, which each serve their own predefined purpose in the human comprehension detection process outlined in Section 5.3.2 and Figure 5.7. The Comprehension Classifier ANN (Figure 5.3) is a key, modular component of FATHOMs architecture, which requires optimisation and validation on human comprehension detection with datasets from experimental studies. Therefore, the following two chapters describe how FATHOM's Comprehension Classifier ANN has been optimised and validated with datasets extracted from two independent digitally recorded human comprehension studies.

Chapter 6 Human Comprehension Detection During Informed Consent

6.1 Introduction

In Chapter 5, FATHOM a novel human comprehension detection software application was introduced. Evaluating FATHOM with an appropriate dataset will reveal whether neural networks can detect patterns of human comprehension and non-comprehension from multiple channels of nonverbal behaviour. However, FATHOM's Comprehension Classifier ANN (Figure 5.3) topology still needs to be confirmed and evaluated with an appropriate dataset. Therefore, this Chapter shall explain how FATHOM's Comprehension Classifier ANN has been trained and validated in human comprehension detection using a dataset extracted from a real-life study on comprehension of informed consent. Section 3.3 highlighted the critical importance of human understanding during informed consent and the current difficulty of reliably capturing human comprehension using existing assessment tools, which are primarily written- and/or spoken-based approaches. For that reason, FATHOM's Comprehension Classifier ANN has been optimised with a dataset from the informed consent process of a mock Human Immunodeficiency Virus (HIV) prevention trial in an attempt to provide an alternative human comprehension detection tool that overcomes the weaknesses of existing bespoke assessment tools. The primary purpose of this chapter is to establish whether FATHOM can identify nonverbal behavioural patterns of low and high human comprehension.

The study[1] titled 'Enhancing Local Verbal and Non-verbal Communication for Informed Consent Processes in Tanzania Study #10159' (Simpson et al., 2010) was developed by experts at Family Health International 360 (FHI 360) and informed by senior members of the Manchester Metropolitan University (MMU) Intelligent Systems Group (ISG). FHI 360 (2016) is a global non-profit human development organisation. The study was executed in the Mwanza region of Tanzania in collaboration with the National Institute of Medical Research (NIMR). NIMR (2016) is a public health research

[1] The study documentation is owned by FHI 360 and is subject to copyright restrictions so has not been included in this Thesis.

institution, which generates scientific information required in the development of better methods and techniques of enhancing disease management, prevention and control. Ethical approval was obtained from FHI 360's Protection of Human Subjects Committee, NIMR's Medical Research Coordinating Committee in Tanzania and the MMU's Faculty Academic Ethics Committee. The female participants underwent a mock informed consent process for a sexual and reproductive health clinical trial where their comprehension was assessed. The remainder of this chapter shall describe the study design, the FATHOM Comprehension Classifier ANN experiments and include a results discussion. The findings presented in this chapter have also been published in Buckingham et al. (2012) and Crockett et al. (2013).

6.2 Study Design

This study addresses the nonverbal aspect of human comprehension detection in the informed consent process by employing a cross-sectional, quantitative methodology. A high-level overview of the study design is shown in Figure 6.1, which has two distinct, sequential phases: the developmental phase and the exploratory testing phase. The outcomes for each phase are shown in the parallelograms in Figure 6.1. Firstly, in the developmental phase, a video dataset is captured from eighty female participants, which is then utilised to optimise FATHOM's Comprehension Classifier ANN (Figure 5.3) using FATHOM's neural network training application (Figure 5.6). The purpose of the developmental phase is to identify whether FATHOM can identify nonverbal patterns of high and low human comprehension.

Secondly, in the exploratory testing phase, a video dataset is captured from another eighty female participants during a mock informed consent process where their comprehension will be assessed by FATHOM containing the optimised Comprehension Classifier ANN from the developmental phase. After the mock informed consent process, the participants are randomly assigned to complete the open or closed comprehension assessment tool and then the self-perception comprehension assessment tool. The purpose of the exploratory testing phase is to evaluate the performance of FATHOM's optimised Comprehension Classifier ANN from the developmental phase in a field test and to perform a cross-comparison analysis of the comprehension measurements. There are four informed consent comprehension measurements, which are the nonverbal measure (IC-NV), the open-ended assessment

measure (IC-O), the close-ended assessment measure (IC-C) and the self-perception measure (IC-SP). Low-level descriptions of the developmental phase (Section 6.2.1) and the exploratory testing phase (Section 6.2.2) are outlined in the subsequent subsections.

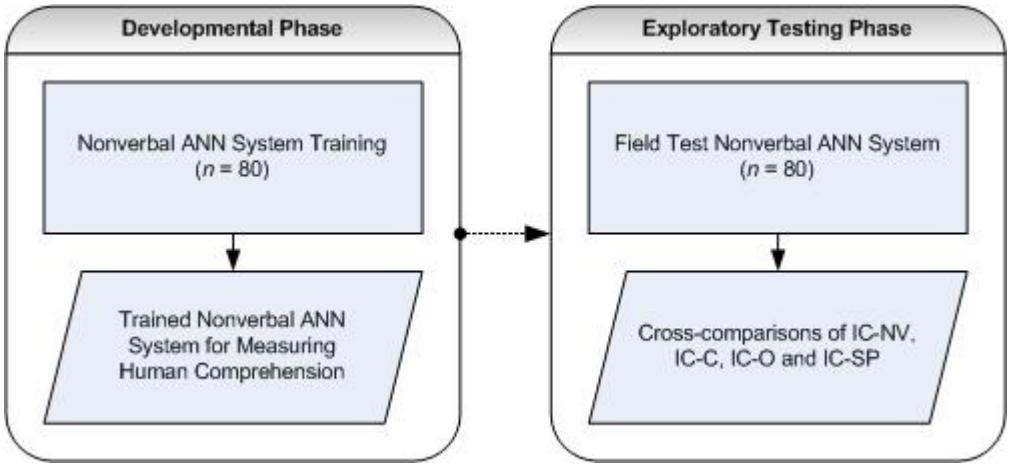


Figure 6.1 Study Design

Source: Adapted from Simpson et al. (2010)

6.2.1 Developmental Phase

A detailed overview of the developmental phase study design is presented in Figure 6.2. The large sample size (n = 80) was chosen to 'account for unexpected heterogeneity in the study population and to maximize accuracy' (Simpson et al., 2010:19). The Sociodemographic Form (Appendix A), is used to collect social and demographic data from the participant. Each participant individually engaged in a learning task that was composed of one high and one low comprehension topic. The high comprehension topic (Task A) was a short talk on condom usage and the low comprehension topic (Task B) was a short talk on HIV mutation. The scripted formative talks for Task A and B are in the Developmental Phase Interviewer Instructions (Appendix B). FHI 360 experts constructed Task A's short talk from Chapter 13 in World Health Organisation (2007) and Task B's short talk from Butler et al. (2007) and Vax Report (2010). Experts in designing informed consent trials were also consulted. Task A was assumed to be a familiar, general HIV prevention topic, therefore considered to be easy to understand so high human comprehension levels were anticipated. On the other hand, Task B was assumed to be an unfamiliar, specialised HIV prevention topic, therefore considered to be hard to understand so low human comprehension levels were anticipated. Immediately, after Task A the participant was asked ten closed and ten open questions about the topic that were intended to be easy. Immediately, after

Task B the participant was asked ten closed and ten open questions about the topic that were intended to be hard. The questions were used to assess how well the participant comprehended each topic i.e. a comprehension measure. All questions for both learning topics are in Appendix B. Experts in designing informed consent trials were consulted over the design of the questions. The presentation order of the learning topics and the two question formats (i.e. open or closed) was randomised, equally. Randomisation was used as an experimental control to account for differences in comprehension relating to question format order (Simpson et al., 2010) and to avoid selection bias. The camcorder symbols in Figure 6.2 indicate when the participant is being filmed.

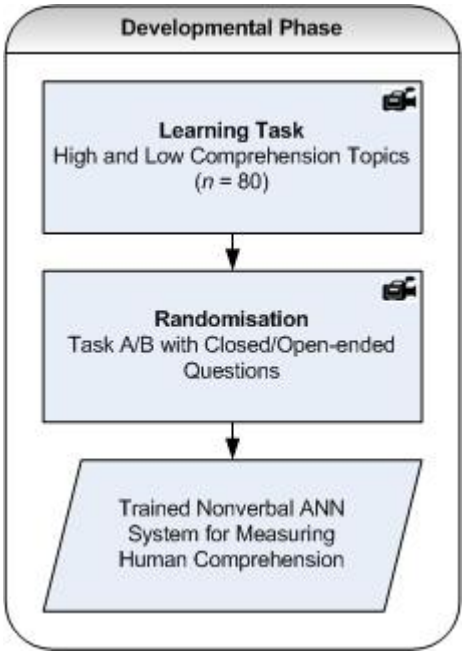


Figure 6.2 Developmental Phase Study Design

Source: Adapted from Simpson et al. (2010)

The study layout in Figure 6.3 was used throughout the developmental phase. Two digital camcorders were used to record head and torso shots (Figure 6.4) of the seated participant during the learning task and the Task A/B closed/open questioning. Study staff were provided with the Developmental Phase Checklist (Appendix C) containing the equipment list and instructions on how to setup the digital camcorder to ensure that consistent high quality digital recordings were obtained. The interviewer followed the instructions and scripts in the Developmental Phase Interviewer Instructions (Appendix B) to make sure that delivery of the learning task and the questions was consistent for each participant. The interviewer also marked the participants' responses to each of the learning task questions in the Developmental Phase

Interviewer Instructions (Appendix B) with a “Y” for a correct answer and an “N” for an incorrect answer. The video technician sets up the digital camcorders before the learning task starts and then leaves the room. During setup, the video technician was responsible for the capturing and playback of an introductory video segment to ensure that the digital recordings had sharp images, adequate lighting and sound quality throughout the session. At the end of the Task A/B questioning, the video technician returned to the room to conclude the digital video recording process. After conducting the study, the interviewer watched the video of the participant and logged the start and end times of the reading of the Task A/B formative script points and the closed/open questions in the Development Phase Summary Sheet (Appendix D) so that it would guide the FATHOM Comprehension Classifier ANN during training with the developmental phase dataset. MMU research staff were not Kiswahili speakers so were solely focused upon the nonverbal behavioural components of the study videos.

Figure 6.3 Study Layout

Figure 6.4 Digital Camcorder Video Shots

Once all participants had completed their participation in the developmental phase, then the videos were transferred to MMU for training with FATHOM's Comprehension Classifier ANN (Figure 5.3) using the backpropagation learning algorithm (Section 4.4). Classification Accuracy (Section 4.5.5) was used to measure the reliability of the performance of FATHOM's Comprehension Classifier ANN at detecting low and high comprehension. The training of the Comprehension Classifier ANN is

discussed in depth in Section 6.7. As soon as, the FATHOM Comprehension Classifier ANN was optimised then it was ready for application and evaluation in the Exploratory Testing Phase.

6.2.2 Exploratory Testing Phase

A detailed overview of the exploratory testing phase study design is presented in Figure 6.5. A large sample size (*n* = 80) was chosen to 'account for unexpected heterogeneity in the study population and to maximize accuracy' (Simpson et al., 2010:19). The Sociodemographic Form (Appendix A), is used to collect social and demographic data from the participant. Each participant individually engaged in a mock informed consent process for a hypothetical clinical trial to evaluate the safety and efficacy of Pre-exposure Prophylaxis (PrEP) to prevent acquisition of HIV. The PrEP trial mock informed consent document (Appendix E) was developed from MacQueen et al. (1999) and Vanichseni (2004) by FHI 360. Immediately after the administration of the mock informed consent document, each participant is randomly assigned to complete either the closed or open-ended informed consent comprehension assessment measure (i.e. IC-C or IC-O). The IC-C is in Appendix F. FHI 360 experts created the IC-C by adapting an existing true or false comprehension assessment tool from a HIV vaccine trial in Bangkok (MacQueen et al., 1999; Vanichseni, 2004). The IC-O is in Appendix G. FHI 360 experts created the IC-O from an existing open-ended tool used in the HIV Prevention Trials Network (HPTN) 035 trial (Karim et al., 2011). Randomisation was used as an experimental control to avoid selection bias. Immediately after completing the IC-C or IC-O, the participant undergoes the self-perception informed consent comprehension assessment measure (IC-SP) in Appendix H. The IC-SP is used to capture the participant's personal perception on how easy or difficult each point of comprehension was in the IC-C/IC-O using a four-point Likert scale i.e. (1) very easy to understand; (2) somewhat easy to understand; (3) somewhat difficult to understand; and (4) very difficult to understand. After the IC-SP, the participant is questioned about their willingness to enrol in the mock clinical trial using the Willingness to Enrol Form (Appendix I). Lastly, a debriefing session is conducted to ensure that the participant fully understood that she did not consent to participate in a real clinical trial. The camcorder symbols in Figure 6.5 indicate when the participant is being filmed. Apart from the informed consent nonverbal measure (IC-NV), all of the

comprehension assessment measures, stemmed from tools utilised in previous HIV trials.

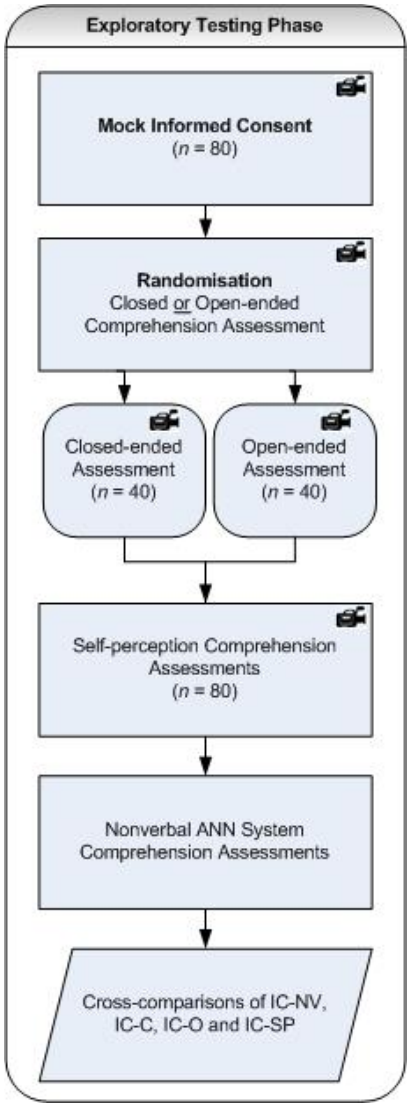


Figure 6.5 Exploratory Testing Phase

Source: Adapted from Simpson et al. (2010)

The study layout in Figure 6.3 was also used throughout the exploratory testing phase. Two digital camcorders were used to record head and torso shots (Figure 6.4) of the seated participant during the mock informed consent process. Experienced counsellors from NIMR, delivered the mock informed consent process. The interviewer followed the instructions contained in Appendix E to Appendix I to ensure that delivery of the mock informed consent process and the comprehension assessment measures was consistent for each participant. The interviewer also marked the participants' responses to each of the questions in the IC-C (Appendix F) and IC-O (Appendix G). The video technician sets up the two digital camcorders before the mock informed consent process starts and then leaves the room. During setup, the video technician was

responsible for the capturing and playback of an introductory video segment to ensure that the digital recordings had sharp images, adequate lighting and sound quality throughout the session. At the end of the willingness to enrol questioning, the video technician returned to the room to conclude the digital video recording process. After conducting the study, the interviewer watched the video of the participant and logged the start and end times of the reading of the informed consent formative script points, the IC-C/IC-O questions and the IC-SP questions in the Exploratory Testing Phase Closed Summary Sheet (Appendix J) or the Exploratory Testing Phase Open Summary Sheet (Appendix K) so that it would guide the evaluation of the optimised FATHOM Comprehension Classifier ANN from Section 6.2.1 with the exploratory testing phase dataset.

Once all participants had completed their participation in the exploratory testing phase, then the videos were transferred to MMU for evaluation with FATHOM's optimised Comprehension Classifier ANN (Figure 5.3) from the developmental phase (Section 6.2.1). Classification Accuracy (Section 4.5.5) will be used to measure the performance of FATHOM's Comprehension Classifier ANN at detecting low and high comprehension. The analysis and evaluation of FATHOM's optimised Comprehension Classifier ANN with the exploratory testing phase videos results in the IC-NV measure. Lastly, cross-comparisons between the four informed consent comprehension assessment measures (IC-C, IC-O, IC-SP and IC-NV) can be performed and the findings discussed.

6.3 Participants

Throughout the developmental and exploratory testing phases of the study, all of the female participants were subject to the same inclusion/exclusion criteria and were recruited from potential future biomedical HIV prevention trial recruitment areas e.g. recreational and entertainment facilities. Eligible female participants had to meet all elements of the following criteria: (1) aged between 18-35 years old; (2) lived in the Magu or Misungwi districts of Mwanza; (3) did not intend to move outside of the study zone in the next month; (4) had at least one vaginal sex act in the past two weeks or had more than one sexual partner in the last thirty days; and (5) had not participated in a clinical research study before. To maintain confidentiality and anonymity, all participants were assigned a unique participant study number.

6.4 Sociodemographics

The sociodemographic profiles for the participants in the developmental and the exploratory testing phases are displayed in Table 6.1. Age is similar across both phases of the study. The participants were predominantly Christian and the majority only had an education up to primary school level. Most participants tended to be married or were cohabiting. Frequently occurring occupations amongst the participants were tailor, bar maid and local food vendor. Their homes were built primarily from burnt/cement brick and thatch/tin roof. The majority of participants knew someone who had died from Acquired Immune Deficiency Syndrome (AIDS), which may be linked with the high prevalence of AIDS/HIV in Africa. The 2010 Joint United Nations Programme on HIV and AIDS (UNAIDS) report on the Global AIDS Epidemic Update reported that approximately 22.5 million people were living with HIV in sub-Saharan Africa in 2009 (UNAIDS, 2010). Most participants possessed a functional literacy level i.e. was able to read the entire sentence "farming is hard work" aloud in Kiswahili.

6.5 Developmental Phase Readability Analysis

The readability of the Task A and Task B formative scripts in the Developmental Phase Interviewer Instructions (Appendix B) have been evaluated. Two readability metrics have been adopted: the Flesch Reading Ease (Flesch, 1948; Flesch, 1949) and the Flesch-Kincaid Grade Level (Kincaid et al., 1975). The Flesch Reading Ease test scores the passage on a 0 to 100-point scale, the higher the Flesch Reading Ease Score (FRES), the easier it is to read the passage. The Flesch Reading Ease formula is:

$$206.835 - (1.015 * ASL) - (84.6 * ASW) \quad (6.1)$$

where ASL is the average sentence length (the number of words divided by the number of sentences) and ASW is the average number of syllables per word (the number of syllables divided by the number of words). The Flesch-Kincaid Grade Level (FKGL) grades the passage using the United States school grading system e.g. a score of 8.3 means that an 8th grader should be able to read and comprehend the passage. The FKGL formula is:

$$(0.39 * ASL) + (11.8 * ASW) - 15.59 \quad (6.2)$$

where ASL is the average sentence length and ASW is the average number of syllables per word. Presently, Swahili readability metrics do not exist. Applying the FRES and FKGL readability metrics to the English text contained in the Task A and B formative

scripts (Appendix B) will provide an estimated measure of readability and enable direct comparisons. Task A's formative script has a 69.5 FRES and a 6.6 FKGL, which means that the average 6th grader (age 11-12 years old) should be able to easily read and understand the script. Task B's formative script has a 53.7 FRES and a 9.6 FKGL, which means that the average 9th grader (age 14-15 years old) should be able to easily read and understand the script. Task A's FRES is higher than Task B and Task A's FKGL is lower than Task B. Therefore, Task B's formative script is more difficult to read and comprehend than Task A, from the perspective of the latter readability metrics.

Table 6.1 Sociodemographics

Sociodemographic Properties		Developmental Phase (*n* = 80)	Exploratory Testing Phase (*n* = 80)
Age	Average	26.8	26.71
	Standard Deviation	4.7	4.721
Education Level	Never attended	6.3%	2.5%
	Pre-primary	1.3%	0%
	Primary	58.8%	63.8%
	O-Level Secondary	32.5%	31.3%
	A-Level Secondary	0%	1.3%
	College	1.3%	1.3%
Religion	Muslim	10%	18.8%
	Catholic	53.8%	42.5%
	Protestant	33.8%	38.8%
	Atheist	2.25%	0%
Marital Status	Never married	11.3%	35%
	Married	28.8%	32.5%
	Cohabiting	43.8%	20%
	Divorced	8.8%	2.5%
	Separated	5%	8.8%
	Widowed	2.5%	1.3%
Occupation	Bar Maid	23.8%	21.3%
	Grocery Maid	1.3%	1.3%
	Hairdresser	7.5%	2.5%
	Local Food Vendor	18.8%	18.8%
	Tailor	30%	37.5%
	Hotel Maid	5%	11.3%
	Petty Trader	7.5%	7.5%
	Shopkeeper	5%	0%
	Other	1.3%	0%
House built with ...	Mud and bricks	0%	6.3%
	Mud and tin roof	17.5%	33.8%
	Burnt/cement brick and thatch or tin roof	82.5%	60%
Residency	Lived continuously for a year or more outside of Tanzania	2.5%	3.8%
Context Information	Has participated in a research study before	1.3%	6.3%
	Personally knows someone who has participated in medical research	18.8%	21.3%
	Personally knows someone who has or has died from AIDS	83.8%	91.3%
Functional Literacy Level	Cannot read at all	13.8%	2.5%
	Able to read only parts of sentence	1.3%	1.3%
	Able to read whole sentence	85%	96.3%

6.6 Developmental Phase Task Analysis

Analysis of the participants marked answers from the closed and open-ended questions asked in Task A and Task B were performed to obtain comprehension measures for comparative purposes and to later guide the optimisation of FATHOM's Comprehension Classifier ANN in Section 6.7. In total, there were 20 questions per Task (10 closed and 10 open questions), which each of the 80 participants verbally answered in Swahili resulting in 3200 answers for marking and analysis. Each question has an exclusive identifier, so that it can easily be uniquely identified. The identifier is structured by amalgamating the abbreviated Task Name, Question Type and Question Number e.g. TAC01 = Task A Closed Question 1 and TBO10 = Task B Open Question 10. Quantitative analyses of the marked answers for Task A and Task B are presented in Sections 6.6.1 and 6.6.2, respectively. A discussion of the findings from the Task A and B analyses is in Section 6.6.3.

6.6.1 Task A Results

Within Task A, there were 1600 answers, which breakdown into 800 closed answers and 800 open answers for analysis. The percentage of participants (n = 80) that correctly answered each closed question are displayed in Table 6.2. In total, 80.625% (645) of the 800 answers for the ten closed questions were marked as being verbally answered correctly. Most participants incorrectly answered TAC03, thus indicating non-comprehension and possible misconceptions on the preventative purposes of male condoms. In TAC10, 55% of participants failed to understand that lubrication can reduce the likelihood of condom breakage. However, despite difficulties with TAC03 and TAC10, the majority of participants knew the answers to the other closed questions, thus signifying comprehension. The percentage of participants (n = 80) that correctly answered each open question are shown in Table 6.3. Unfortunately, one participants marked answer for TAO08 was missing from the dataset resulting in 799 marked open answers for analysis. In total, 80.976% (647) of the 799 answers for the ten open questions were marked as being verbally answered correctly. Interestingly, all participants correctly answered TAO03 correctly, even though the majority incorrectly answered a similar question (TAC03) in the closed format. A comparison between the total number of correct answers from every participant for Task A's closed and open questions are shown in the bar chart in Figure

6.6(a). The bars in the chart are densely concentrated between five and eight correct answers, which means that the participant's comprehension of Task A was consistent.

Table 6.2 Task A Closed Questions

Identifier	Closed Question	Correct (*n* = 80)
TAC01	A male condom is a sheathe or covering to fit over a man's penis.	86.3%
TAC02	Male condoms are mostly made of latex rubber.	98.8%
TAC03	Male condoms can only prevent Sexually Transmitted Infections.	7.5%
TAC04	It is preferable to use a condom when its package is damaged or expired.	100%
TAC05	A condom should be placed on a man's penis when it is soft.	97.5%
TAC06	The condom should be rolled all the way to the base of the penis.	96.3%
TAC07	The condom should be removed immediately after ejaculation.	80%
TAC08	A condom should be used for multiple sex acts.	96.3%
TAC09	A condom should be wrapped in its package and placed in a rubbish bin or latrine.	98.9%
TAC10	Lubrication cannot reduce the likelihood of condom breakage.	45%

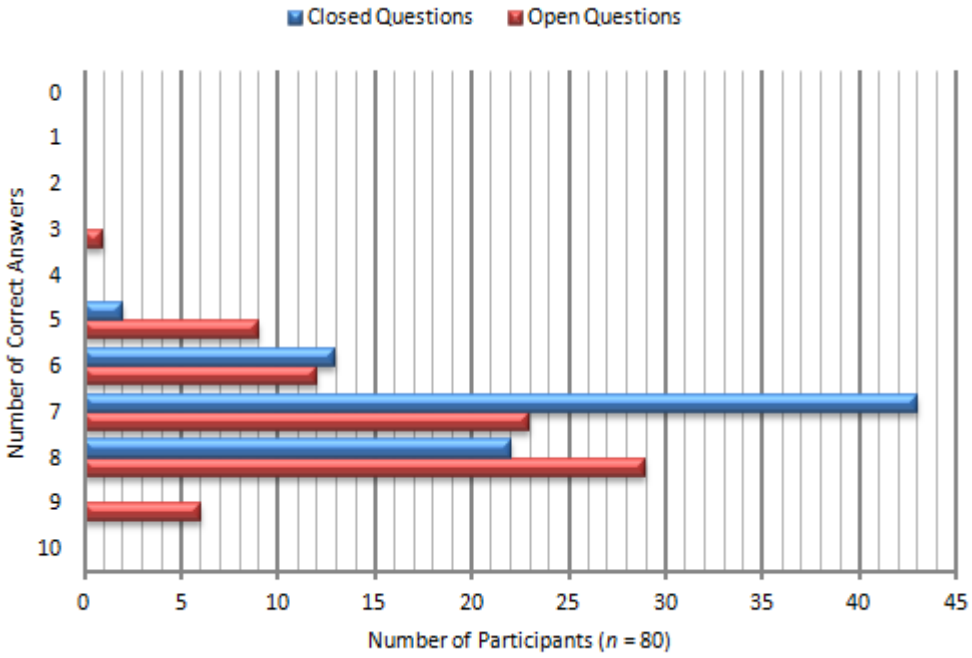
Table 6.3 Task A Open Questions

Identifier	Open Question	Correct (*n* = 80)
TAO01	What is a male condom?	28.8%
TAO02	What are most male condoms made of?	60%
TAO03	What can male condoms prevent?	100%
TAO04	What do you need to check before using a condom?	97.5%
TAO05	When should a condom be placed over a man's penis?	93.8%
TAO06	What do you do after you have placed the condom on the tip of an erect penis?	92.5%
TAO07	When do you take off the condom?	87.5%
TAO08	How often should condoms be used for the greatest effectiveness in HIV prevention?	96.3%*
TAO09	How should condoms be disposed?	95%
TAO10	What can be used to avoid condom leakage?	57.5%

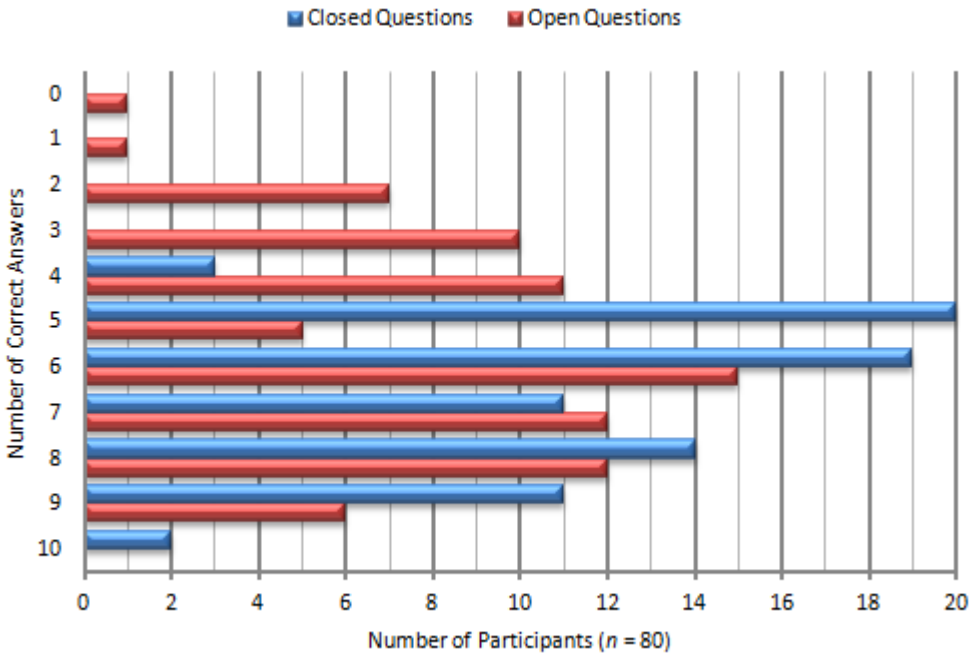
*Missing one answer so (*n* = 79)

6.6.2 Task B Results

Within Task B, there were 1600 answers, which breakdown into 800 closed answers and 800 open answers for analysis. The percentage of participants (*n* = 80) that correctly answered each closed question are displayed in Table 6.4. In total, 66.75% (534) of the 800 answers for the ten closed questions were marked as being verbally answered correctly. Most participants had little difficulty in correctly answering TBC01 and TBC08. However, most participants failed to correctly answer TBC03 and TBC04, which indicates possible comprehension difficulties surrounding the meaning of the complex terms featuring in the questions: mutation, recombination and genome. The percentage of participants (*n* = 80) that correctly answered each



(a) Task A



(b) Task B

Figure 6.6 Bar Charts

open question are shown in Table 6.5. In total, 54.75% (438) of the 800 answers for the ten open questions were marked as being verbally answered correctly. 95% of participants failed to understand TBO05. Overall, the open questions revealed lower levels of understanding than the closed questions. A comparison between the total number of correct answers from every participant for Task B's closed and open questions are shown in the bar chart in Figure 6.6(b). The bars in the chart are dispersed widely between two and nine correct answers, which means that the participant's comprehension of Task B was varied and much lower with when assessed with the open questions.

Table 6.4 Task B Closed Questions

Identifier	Closed Question	Correct (*n* = 80)
TBC01	Viral diversity is when there are different versions of a virus - they are like members of a large family, different from yet related to each other.	96.3%
TBC02	Mutation and recombination are the two major processes leading to the genetic diversity of HIV.	86.3%
TBC03	Mutation is a process by which genetic material is broken and joined to other genetic material.	35%
TBC04	Recombination is when changes to the HIV genome are caused by copying errors in the genetic materials.	35%
TBC05	Mutation and/or recombination cannot change the strain of HIV that a person is originally infected with.	41.3%
TBC06	HIV-1 and HIV-4 are the names of the two major HIV viral types.	62.5%
TBC07	The main groups under HIV type-1 are M, N, and O.	85%
TBC08	The main groups under HIV type-2 are A through H.	95%
TBC09	Viral diversity is not an obstacle to the development of an effective AIDS vaccine.	57.5%
TBC10	Viral diversity means that HIV medications work better for some than others because the medications may better match to fight some strains of virus than others.	73.8%

Table 6.5 Task B Open Questions

Identifier	Open Question	Correct (*n* = 80)
TBO01	What is viral diversity?	53.8%
TBO02	Please name one of two major processes leading to the genetic diversity of HIV.	43.8%
TBO03	What is a mutation?	35%
TBO04	What is recombination?	63.8%
TBO05	How can the strain of HIV virus in an infected person change over time?	5%
TBO06	Please name one of two major HIV viral types.	70%
TBO07	Please name at least one main group under HIV type-1.	65%
TBO08	Please name at least one main group under HIV type-2.	75%
TBO09	What is a major obstacle to the development of an effective AIDS vaccine?	81.3%
TBO10	Why do HIV meds work better for some people than for others?	55%

6.6.3 Discussion

In Task A, the analysis of the marked answers confirmed that the majority of participants easily understood Task A's formative script with the closed and open questions yielding similar levels of comprehension for the total percentage of correctly answered closed and open questions (80.625% vs. 80.976%). The analysis of Task B's marked answers revealed lower levels of comprehension when the participants were assessed using the open questions in comparison to the closed questions (54.75% vs. 66.75%). The closed questions are subject to guessing with a 50% likelihood of getting the right answer, so they may not always reflect the participant's true level of understanding. By comparing the bar charts in Figure 6.6, it can clearly be seen that Task B (Figure 6.6(b)) was more difficult for the participants to comprehend than Task A (Figure 6.6(b)) because the bars are far more distributed for the closed and open assessment tools. The latter finding is also supported by readability metrics performed on the formative scripts for Task A and Task B in Section 6.5.

6.7 Developmental Phase Comprehension Classifier ANN Experiments

The objective of the developmental phase learning tasks was to capture a video dataset for optimising FATHOM's Comprehension Classifier ANN (Figure 5.3) using FATHOM's neural network training application (Figure 5.6). Therefore, this Section will introduce two experiments (Experiment 1 (Section 6.7.1) and Experiment 2 (Section 6.7.2)), where the developmental phase video dataset has been used to train and optimise FATHOM's Comprehension Classifier ANN. Within each experiment, the aim, methodology and results are reported. Detailed descriptions of where the channel datasets were extracted from the videos and the configuration parameters used to train and validate FATHOM's Comprehension Classifier ANN are also included. The approach to the assignment of the desired responses (Section 4.4) for each dataset is what makes these Comprehension Classifier ANN experiments distinctly different and interesting. Most importantly, this Section will identify whether human comprehension and non-comprehension can be detected from the participants multiple channels of nonverbal behaviour when engaging in the learning tasks. Both of the developmental phase experiments have been published in Buckingham et al. (2012) and Crockett et al. (2013).

Throughout the experiments only the head shot videos (see example in Figure 6.4(a)) were used because they encapsulated all of the facial nonverbal behaviours, which needed to be monitored by FATHOM. Before the channel datasets could be collated for each experiment, some preparation was required on the video dataset. The videos were all in Moving Picture Experts Group (MPEG) format and there were between one and three MPEG files per participant. Therefore, in order to work with FATHOM they needed to be merged together and converted to Audio-Video Interleaved (AVI) format so that there was only one AVI file per participant. MPEG Streamclip (Squared5, 2016), a powerful, free video converter was used to merge and convert the videos. The video preparation was a time-consuming process but it only had to be performed once on the video dataset.

During the video preparation, it was noted that the recording quality and setup of the digital camcorder in some of the MPEG videos was not always consistent. As a result, out of the eighty head shot videos, fifty videos were discarded, leaving thirty usable videos for the two experiments. Videos were discarded from the dataset for one or more of the following reasons: (1) poor lighting resulting in undistinguishable facial features due to shadows; (2) corrupt or missing MPEG files; (3) incomplete Developmental Phase Summary Sheet (Appendix D); (4) continuous movement of the digital camcorder angle during the recording; and/or (5) incorrect setup of the head shot recording by the video technician. Feedback regarding the quality of the videos was provided to FHI 360 to help enhance the quality of the video recordings in the Exploratory Testing Phase and for the remainder of the Developmental Phase. Suggestions in the feedback to record the participants in front of a plain, dark background with artificial lighting to help improve the quality of the recordings were later adopted by the study staff in Africa. With hindsight, a mini pilot study on a small number of video recordings should have been executed as a preventative measure.

Firstly, the channel data had to be collated from each of the videos (n = 30). This was achieved by playing each video in FATHOM (Figure 5.1) with the “Extract Channel Data” mode switched on, so that the Channel Accumulator (Section 5.3.2) would automatically collate the channel statistics into grouped channel vectors, which were then output to CSV file. A flat file containing a list of start and end frame numbers, accompanies each video and informs FATHOM of the time period(s) where it is required to track and extract the state of the participants nonverbal channels from the

video frames. The flat file is referred to as a Frame List. Every pair of start and end frame numbers in the Frame Lists, represented the start and end time for a single Task A/B Formative Script Point in the Developmental Phase Summary Sheet (Appendix D) i.e. the interviewer reading a Task A or Task B Formative Script Point of Comprehension to the participant in the video. Within the channel statistics, each grouped channel vector has to have a desired response appended which, can either be the comprehension (+1) or non-comprehension (-1) class. How the desired responses were determined shall be described in the methodology for each experiment because this feature uniquely defined each experiment. When combined together the CSV files from all 30 videos formed the entire dataset and were ready for optimising the FATHOM Comprehension Classifier ANN.

6.7.1 Experiment 1

Aim

The aim of this experiment was to see if FATHOM's Comprehension Classifier ANN could reliably distinguish between the easy and hard learning tasks after training and validation.

Methodology

Task A was considered to be easy to understand so high human comprehension levels were predicted. On the other hand, Task B was considered to be hard to understand so low human comprehension levels were predicted. Analyses of the participants marked answers for each task in Section 6.6 revealed that Task B was more difficult than Task A, thus supporting the latter predictions. Therefore, the dataset had the desired responses assigned as follows, the extracted grouped channel vectors spanning Task A's Formative Script Points of Comprehension were all labelled with a comprehension (+1) desired response and the extracted grouped channel vectors spanning Task B's Formative Script Points of Comprehension were all labelled with a non-comprehension (-1) desired response. The entire dataset had a total of 241,945 vectors, which broke down into 109,160 (45%) comprehension and 132,785 (55%) non-comprehension. The proportions of vectors in each class is nearly balanced so should not inflict a severe bias during training with the FATHOM Comprehension Classifier ANN.

Table 6.6 shows the configuration and parameters used to train and validate FATHOM's Comprehension Classifier ANN using the error-backpropagation learning algorithm (Section 4.4). There were 40 inputs to the ANN because each input vector in the dataset had 40 normalised input signal values. Cybenko (1989) demonstrated that a continuous feedforward single hidden layer ANN can approximate any arbitrary decision region well so a single hidden layer ANN was used. As highlighted in Section 4.5.2, there is no hard-and-fast rule for deciding the starting number of neurons within the hidden layer(s). Heaton (2008) suggests setting the size of the hidden layer between the size of the input layer and the output layer. Therefore, the single hidden layer had 30 neurons. Only one output neuron was necessary for outputting a value between +1 (comprehension) and -1 (non-comprehension). Network weights are typically initialised to small random values in the range of 0 ± 1 or 0 ± 0.5 (Fausett, 1994). Therefore, the random 0 ± 1 weight range was selected as the weight initialisation method. A three-way data split (Section 4.5.5) was used to partition the dataset with equal proportions of each class within each set. The checking epoch value was used to halt training of the Comprehension Classifier ANN every n^{th} epoch so that its performance with the validation set could be evaluated to enable early stopping. The maximum epoch's value was used to ensure that the error-backpropagation learning algorithm terminated in case early stopping was not ascertained. The learning rate (η) is typically a small positive number between zero and one (Section 4.5.4) so the η was set at 0.5. The 10-fold cross-validation experiment was repeated six times. Each time, the neural networks starting weights were randomly initialised and the dataset was divided randomly. The presentation of the vectors during training and validation was also randomised to prevent the neural network from learning patterns on a presentation order basis.

Table 6.6 ANN Training Configuration

Parameter	Value
ANN topology	40:30:1
n-fold cross-validation	10-fold cross-validation
Training Set	60%
Validation Set	30%
Test Set	10%
Stopping criteria	Validation CA
Checking epochs	1000
Maximum epochs	20,000
Learning Rate (η)	0.5
Weight Initialisation	0 ± 1
Lambda (λ)	1

Results and Discussion

Table 6.7 shows the ANNs that yielded the highest total classification accuracy (CA) from the test set in each of the six cross-validation experiments. A breakdown of the CAs for the true positives (TP) and true negatives (TN) is also shown in Table 6.7. A TP is a vector that has been correctly classified as being in the comprehension class (+1) and a TN is a vector that has been correctly classified as being in the non-comprehension class (-1). The CA's were consistent across all the six cross-validation experiments with all CAs above 85%, which strongly suggests that the ANNs were able to distinguish patterns between the easy and hard learning tasks from multiple channels of human nonverbal behaviour. Overall, experiment 5 (red column in Table 6.7), produced the ANN with the largest testing total CA at 89.29%.

Table 6.7 Experiment 1 Cross-validation Results

Set	CA	10-fold Cross-validation Experiments					
		1	2	3	4	5	6
Training	TP	89.89%	89.73%	89.87%	89.88%	88.85%	90.72%
	TN	86.80%	88.78%	89.06%	89.59%	91.26%	90.27%
	Total	88.35%	89.25%	89.46%	89.74%	90.05%	90.50%
Validation	TP	88.55%	87.93%	88.44%	88.63%	87.41%	88.98%
	TN	85.22%	87.45%	87.67%	88.38%	90.48%	89.13%
	Total	86.88%	87.69%	88.06%	88.50%	88.95%	89.06%
Testing	TP	88.81%	87.80%	88.68%	88.69%	87.91%	89.69%
	TN	85.49%	87.84%	88.39%	88.32%	90.67%	88.49%
	Total	87.15%	87.82%	88.53%	88.50%	89.29%	89.09%

Although the analysis of the participants marked answers for each task in Section 6.6 revealed that Task B was more difficult than Task A, it can be seen that Task A did not result in only correct answers (Table 6.2 and Table 6.3) and Task B did not result in only incorrect answers (Table 6.4 and Table 6.5) i.e. not purely comprehension in Task A and not purely non-comprehension in Task B. Thus, the approach to labelling the desired response with the dataset in this experiment has unintentionally introduced some noise within the comprehension and non-comprehension classes. Therefore, the next experiment in Section 6.7.2 introduces an alternative approach to determining the desired response in an attempt to reduce the noise within the datasets classes and to increase reliability.

6.7.2 Experiment 2

Aim

The aim of this experiment was to see if FATHOM's Comprehension Classifier ANN could reliably distinguish between human comprehension and non-comprehension within the easy and hard learning tasks after training and validation.

Methodology

In this experiment, a different approach was used to determine the datasets desired responses in order to overcome the weakness of Experiment 1's (Section 6.7.1) approach. This time, the extracted grouped channel vectors for each Task A or Task B Formative Script Points of Comprehension had their desired response labelled according to whether the participant answered the Task A/B question correctly or incorrectly. Therefore, if the participant got the answer to a Task A/B question correct then the grouped channel vectors spanning the corresponding Task A/B Formative Script Point of Comprehension were all labelled with a comprehension (+1) desired response. Otherwise, if the participant got the answer to a Task A/B question wrong then the grouped channel vectors spanning the corresponding Task A/B Formative Script Point of Comprehension were all labelled with a non-comprehension (-1) desired response. The Task A and Task B closed marked answers were not used to determine the desired response because they are susceptible to guessing. Only the Task A and Task B open marked answers were used as they are a more reliable method for capturing true comprehension and true non-comprehension.

Table 6.8 shows how Task A's open questions were linked to each of Task A's Formative Script Points of Comprehension (Appendix D) to form one-to-one relationships. The one-to-one relationships were possible because the open questions directly related to individual points of comprehension. For example, TAO02 "What are most male condoms made of?" directly links to Task A Formative Script Comprehension Point 2 "Male condoms are made of thin latex rubber". The same approach was used to link Task B's open questions to each of Task B's Formative Script Points of Comprehension (Appendix D) as shown in Table 6.9. One-to-one relationships could not be formed for TBO07 and TBO08 because they both linked to the same comprehension point (red row in Table 6.9) so they were not included. The percentage of participants (n = 30) that correctly answered each open question are also displayed in Table 6.8 and Table 6.9. In total, 82.67% (248) of the 300 answers for the ten Task A open questions were marked as being verbally answered correctly. All participants were able to correctly answer TAO03, TAO08 and TAO09 and none were able to answer TBO05. In total, 57.33% (172) of the 300 answers for the ten Task B open questions were marked as being verbally answered correctly. A comparison between the total number of correct answers from every participant for Task A and Task B's

Table 6.8 Linking Task A's Open Questions to the Points of Comprehension

Point	Points of Comprehension	Identifier	Correct (*n* = 30)
1	Male condoms are sheaths, or coverings, that fit over a man's erect penis.	TAO01	33.3%
2	Most condoms are made of thin latex rubber.	TAO02	63.3%
3	Male condoms can prevent both pregnancy and sexually transmitted infections, including HIV.	TAO03	100%
4	Check the condom package. Do not use if it is torn or damaged. Look at the expiration date on the condom package. Do not use a condom that has expired unless you do not have a newer condom to use.	TAO04	96.7%
5	Before any physical contact, place the condom on the tip of the erect penis with the rolled side out.	TAO05	93.3%
6	Unroll the condom all the way to the base of the erect penis.	TAO06	90%
7	Immediately after ejaculation, hold the rim of the condom in place and withdraw the penis while it is still erect.	TAO07	90%
8	Condoms provide the best protection against HIV when they are used with every act of sex.	TAO08	100%
	If you're having sex again or switching from one sex act to another, use a new condom.		
9	Dispose of the used condom safely. Wrap the condom in its package and put it in the rubbish or latrine. Do not put the condom in a flush toilet because it can cause problems with plumbing.	TAO09	100%
10	Lubrication may be used with condoms. Some kinds of lubrication are less likely to cause condom breakage. Natural vaginal secretions or nonoil based lubricants may be used to reduce likelihood of condom breakage.	TAO10	60%

Table 6.9 Linking Task B's Open Questions to the Points of Comprehension

Point	Points of Comprehension	Identifier	Correct (*n* = 30)
1	Globally, more than 40 million people are infected with HIV. While the vast majority of these people experience similar symptoms, this does not mean that everyone is infected with an identical version of HIV. In fact, there are many, many different versions of HIV. These can be thought of as members of a large family: they are different from, but related to, each other. The broad term for this phenomenon is viral diversity.	TBO01	60%
2	The genetic diversity of HIV stems from two major processes.	TBO02	43.3%
3	One process is called mutation. HIV reproduces (or replicates) in an infected person by making more copies of its genome. When HIV copies itself it frequently makes errors, called mutations.	TBO03	40%
4	The other process, known as recombination, can happen if a person is infected with two different versions of HIV.	TBO04	76.7%
5	It is possible for people who are repeatedly exposed to HIV to become infected with more than one viral strain – including viruses from different groups. Then these viruses can sometimes exchange portions of their genomes to form a new virus that has parts of genes from each parent virus – this is called a recombinant virus. Recombinant virus strains can also be passed from one person to another.	TBO05	0%
6	These include two major viral types (HIV-1 and HIV-2).	TBO06	63.3%
7	Numerous groups (M, N, and O for HIV-1, and A through H for HIV-2) and numerous sub-subtypes and CRFs.	TBO07	73.3%
8		TBO08	80%
9	The diversity of the virus is also one of the obstacles to the development of an effective AIDS vaccine. It is difficult to create a vaccine that will protect people from all the types of HIV.	TBO09	83.3%
10	The viral diversity of HIV is one of the major obstacles to diagnosing, monitoring, and treating HIV. Medications may help some strains of the virus more than other strains.	TAO10	53.3%

open questions are shown in the bar chart in Figure 6.7. The bars in the chart are dispersed more widely for the Task B than Task A. Overall, the participants still had a better understanding of Task A than Task B, which is in line with the findings in Section 6.6.

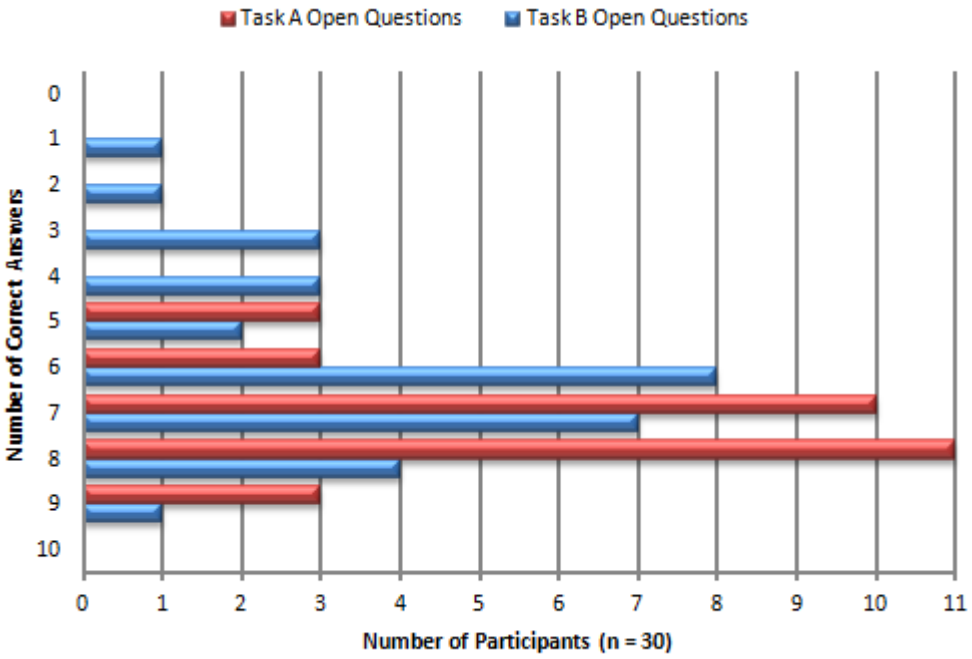


Figure 6.7 Bar Chart

The entire dataset had a total of 71,787 vectors, which broke down into 45,589 (63.5%) comprehension and 26,198 (36.5%) non-comprehension. The proportions of vectors in each class was not as balanced as Experiment 1 (Section 6.7.1) but were purer. The same configuration and parameters from Experiment 1 (see Table 6.6) were used to train and validate FATHOM's Comprehension Classifier ANN using the error-backpropagation learning algorithm (Section 4.4). The 10-fold cross-validation experiment was also repeated six times.

Results and Discussion

Table 6.10 shows the ANNs that yielded the highest total CA from the test set in each of the six cross-validation experiments. A breakdown of the TP CAs and TN CAs is also included. The CA's were consistent across all the six cross-validation experiments with all CAs above 81%, which strongly suggests that the ANNs were able to distinguish human comprehension and non-comprehension patterns in the easy and hard learning tasks from multiple channels of human nonverbal behaviour. Overall, experiment 3

(red column in Table 6.10), produced the ANN with the largest testing total CA at 87.05%.

Table 6.10 Experiment 2 Cross-validation Results

Set	CA	10-fold Cross-validation Experiments					
		1	2	3	4	5	6
Training	TP	89.52%	89.92%	90.54%	89.20%	88.28%	89.53%
	TN	87.67%	89.57%	90.40%	89.85%	90.57%	87.13%
	Total	88.59%	89.74%	90.47%	89.52%	89.43%	88.33%
Validation	TP	85.52%	87.12%	86.97%	86.05%	84.50%	87.00%
	TN	82.97%	83.73%	84.27%	84.73%	85.16%	81.34%
	Total	84.25%	85.42%	85.62%	85.39%	84.83%	84.17%
Testing	TP	86.34%	87.09%	88.27%	86.40%	84.52%	87.31%
	TN	82.75%	83.04%	85.84%	85.34%	86.55%	82.72%
	Total	84.55%	85.07%	87.05%	85.87%	85.54%	85.02%

6.7.3 Conclusions

Across both developmental phase experiments, all twelve 10-fold cross-validation experiments consistently yielded neural networks with CAs above 80%. Thus, repeatedly demonstrating that FATHOM's trained Comprehension Classifier ANN was able to detect patterns of human comprehension and non-comprehension from the dataset, which contained multiple channels of facial nonverbal behaviour. Experiment 1 yielded a Comprehension Classifier ANN with a higher testing total CA than Experiment 2 (89.29% vs. 87.05%). However, the approach to labelling the desired response in Experiment 2 should have produced purer classes of comprehension and non-comprehension than Experiment 1. The classes within Experiment 2's dataset were not as well balanced as Experiment 1, which may explain why testing CAs were lower, particularly for the testing TN CAs. The dataset in Experiment 2 could have been manually balanced but this option was discarded because it would introduce an artificial bias by disturbing the natural occurrence and frequency of the nonverbal behaviours within the dataset.

6.8 Exploratory Testing Phase Comprehension Classifier ANN Experiments

The objective of the exploratory testing phase was to capture a video dataset for evaluating FATHOM's trained Comprehension Classifier ANN (Figure 5.3) from the developmental phase and to compare its performance against the other three informed consent comprehension measures (IC-C, IC-O and IC-SP). However, due to the poor quality of the video recordings in both phases of the study, FATHOM has not

been applied to the videos in the exploratory testing phase. FHI 360 and NIMR have performed cross-comparisons of the IC-O, the IC-C and the IC-SP comprehension measures from the exploratory testing phase and published their findings in MacQueen et al. (2014).

6.9 Summary

This chapter began by introducing the design of two-phase study methodology on a mock informed consent process for a sexual and reproductive health clinical trial, which was executed in Tanzania by FHI 360 and NIMR. Each phase of the study was designed to capture 80 participants in videos that could be used at a later date to train and validate FATHOM's Comprehension Classifier ANN (Figure 5.3) on human comprehension detection during informed consent. Descriptions on how FATHOM's Comprehension Classifier ANN was trained with channel data extracted from the developmental phase videos were provided. The cross-validation results from developmental phase experiments consistently attained testing total CAs above 84%. Thus, demonstrating that FATHOM's Comprehension Classifier ANN was able to detect patterns of human comprehension and non-comprehension from the nonverbal multichannels. These findings are limited to African women and so further studies containing participants of different ethnicities, ages and gender are necessary to enhance the reliability of the results and to confirm whether general patterns exist in the entire population. Unfortunately, the Exploratory Testing Phase analysis with FATHOM was not completed because of the poor video recordings. The next chapter expands research on human comprehension detection using FATHOM by optimising and validating FATHOM's Comprehension Classifier ANN with a dataset from a learning environment study.

Chapter 7 Human Comprehension Detection in a Learning Environment

7.1 Introduction

In the previous chapter, FATHOM's Comprehension Classifier ANN (Figure 5.3) was trained and evaluated with a dataset from an informed consent field study. Although FATHOM's Comprehension Classifier ANN was able to detect human comprehension and non-comprehension from multiple channels of nonverbal behaviour, the findings are limited to 18-35 year old African females. Therefore, this chapter expands research on human comprehension detection using FATHOM by training and validating FATHOM's Comprehension Classifier ANN with a video dataset extracted from a learning environment study conducted at the Manchester Metropolitan University (MMU) in the United Kingdom (UK). Ethical approval was obtained from the MMU Faculty Academic Ethics Committee. In the learning environment study, each of the participants (20 males and 20 females) watched a factual video and then had their comprehension of the factual video assessed in a video recorded question and answer session. The aim of the study was to identify whether FATHOM could distinguish general patterns of human comprehension and non-comprehension from the male and female participant's nonverbal behavioural channels. To avoid the issue of poor video recording quality encountered in the previous study (Chapter 6), a mini pilot study was executed. The remainder of this chapter describes the study design, the FATHOM Comprehension Classifier ANN optimisation experiments and a results discussion. Some of the findings presented in this chapter have been published in Buckingham et al. (2014).

7.2 Study Design

This study addresses the nonverbal aspect of human comprehension detection in a learning environment by adopting a quantitative methodology. An overview of the study design is displayed in Figure 7.1. The outcome for this study is shown in the parallelogram (Figure 7.1). The purpose of the study is to identify whether FATHOM can identify nonverbal patterns of human comprehension and non-comprehension from the male and female participants. To do this, a video dataset is captured from

forty participants (20 males and 20 females), which is then utilised to optimise FATHOM's Comprehension Classifier ANN (Figure 5.3) using FATHOM's neural network training application (Figure 5.6). The sample size (n = 40) was determined as being adequate from a previous nonverbal research study using ANNs (Rothwell, 2002) and because a pilot study was executed to ensure that video recordings were of a consistent high quality to minimise video discarding.

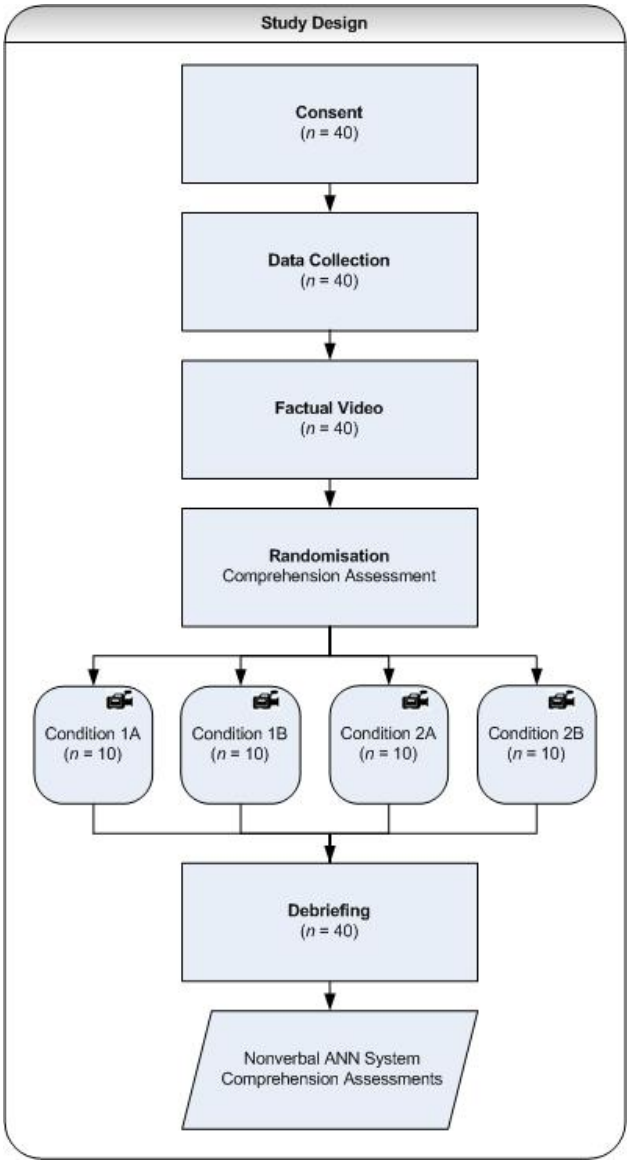


Figure 7.1 Study Design

Each participant independently engaged in the study. On arrival, the participant was given the Information Sheet (Appendix L) and a Consent Form (Appendix M). After consenting to participate, the Data Collection Form (Appendix N) was given to the participant to complete for post study demographical analysis. The participant was

then seated in front of a television screen upon, which a factual video on termites was played back on. Throughout the duration of the factual video (8 min 40 s) the participant's upper body was filmed using one digital camcorder. The factual video was targeted at the general public with no age restriction and covered: functional architectural aspects of the termite mounds, roles within the social structure of a termite colony and locations where termite colonies thrive. The termite video was chosen because it was a subject area, which most participants would have not encountered in their current roles. Furthermore, the video provided comprehensive, educational facts on termites that were not native to the UK. A factual video was used because previous research on automatic human comprehension detection in Section 3.5.2 has predominantly used written passages to deliver subject matter to participants prior to the application of a comprehension assessment tool.

Immediately after watching the factual video, the interviewer sat in front of the participant and asked the participant a set of twenty questions relating to the termite video so that participant comprehension could be measured. Throughout the question and answer (Q&A) session, the participant's upper body was filmed using one digital camcorder. Within the set of twenty questions there were: five easy closed questions, five hard closed questions, five easy open questions and five hard open questions (Table 7.3 - Table 7.6). The questions and answers on the termite video were devised by two experts (Academic Professors) on the subject area. The experts were supplied with the Expert Information Sheet (Appendix O) to focus and guide question generation on the video content. The correct answer(s) were later used as a mark scheme. The order in which the questions were asked by the interviewer was determined by the condition that the participant had been randomised in to. There were four conditions: 1A, 1B, 2A and 2B shown in Table 7.1 to counteract question order effect. For example, if the participant was randomised in to condition 2A then he/she would receive the easy open questions first, followed by the hard open questions, then the easy closed questions followed by the hard open closed questions. Each participant was randomised in to one of the conditions so that each condition had five males and five females (10 participants per condition). The purpose of equally randomising the participants across the conditions was to reduce the chance of producing an imbalanced dataset related to sex.

The interviewer followed the Interviewer Instructions (Appendix P) to ensure consistency in study delivery e.g. if the participant did not respond to a question then the interviewer was instructed to repeat the question. Lastly, a debriefing session was conducted to close the session and to thank the participant for their participation. The camcorder symbols in Figure 7.2 indicate when the participant was being filmed. To ensure consistency in the quality of the digital video recordings the study was executed in the same room with the same equipment and layout (Figure 7.2).

Table 7.1 Question Order

Condition	Closed Questions		Open Questions	
	Easy	Hard	Easy	Hard
1A	1st	2nd	3rd	4th
1B	2nd	1st	4th	3rd
2A	3rd	4th	1st	2nd
2B	4th	3rd	2nd	1st

Once all participants had completed their participation in the study then the videos were prepared and the answers to the questions were marked in preparation for optimising FATHOM's Comprehension Classifier ANN (Figure 5.3) using the error-backpropagation learning algorithm (Section 4.4). Classification Accuracy (Section 4.5.5) was used to measure the reliability of the performance of FATHOM's Comprehension Classifier ANN at detecting low and high comprehension. The training and validation of FATHOM's Comprehension Classifier ANN is discussed in depth in Section 7.7.

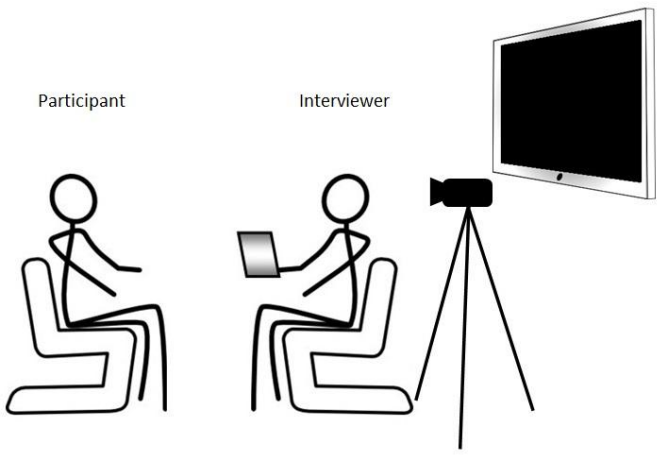


Figure 7.2 Study Layout

7.3 Participants

In this study, the male and female participants were all subject to the same inclusion/exclusion criteria and were recruited from the MMU campus. Eligible participants had to meet all elements of the following criteria: (a) aged ≥ 18 years old; (b) English speaking; and (c) was either a student or member of staff at MMU. To maintain confidentiality and anonymity, all participants were assigned a unique participant study number.

7.4 Demographics

The demographic profile of the forty participants (20 males and 20 females) in this study is shown in Table 7.2. The males had a mean age of 41 years old (Standard Deviation = 14 years) and the females had a mean age of 39 years old (Standard Deviation = 14 years). Most participants had been educated at university level and were predominantly single. Almost half of the participants were teaching staff. None of the participants regarded themselves as having high prior knowledge on the subject of termites.

Table 7.2 Demographics

Demographic Properties		% (*n* = 40)
Age	18-29	27.5%
	30-39	17.5%
	40-39	22.5%
	50-59	22.5%
	60-69	10%
Education Level	A-levels/Foundation Degree	20%
	Bachelors Degree	22.5%
	Masters Degree	22.5%
	PhD	27.5%
	Other	7.5%
Occupation	Student	32.5%
	Teaching Staff	45%
	Non-teaching Staff	22.5%
Prior Knowledge	None	47.5%
	Low	47.5%
	Medium	5%
	High	0%
Religion	Christian	40%
	No religion	55%
	Other	2.5%
	Undisclosed	2.5%
Marital Status	Single	65%
	Married	27.5%
	Divorced	2.5%
	Separated	2.5%
	Undisclosed	2.5%

7.5 Readability Analysis

The narrative from the termites video was transcribed and evaluated using the FRES and FKGL readability metrics, which were outlined in Section 6.5. The transcript had a 63.3 FRES and a 9.8 FKGL, which means that the average 9th grader (age 14-15 years old) should be able to easily read and understand the transcript. The readability of the transcript in this study was more difficult than the both of the Developmental Phase learning task scripts in the informed consent study (Section 6.5).

7.6 Q&A Analysis

Analyses of the participants marked answers from the closed and open questions were performed to obtain comprehension measures for comparative purposes and to later guide the optimisation of FATHOM's Comprehension Classifier ANN in Section 7.7. In total, there were 20 questions (10 closed and 10 open questions), which each of the 40 participants verbally answered resulting in 800 answers for marking and analysis. Each question has an exclusive identifier, so that it can easily be uniquely identified. The identifier is structured by amalgamating the abbreviated Question Type, Difficulty Type and Question Number e.g. CE1 = Closed Easy Question 1 and OH5 = Open Hard Question 5. Quantitative analyses of the marked answers for closed and open questions are presented in Sections 7.6.1 and 7.6.2, respectively. A discussion of the findings from the closed and open results is presented in Section 7.6.3.

7.6.1 Closed Question Results

Within the ten closed questions, five were categorised as easy and five were categorised as hard by the experts. There were 400 answers, which broke down into 200 easy closed answers and 200 hard closed answers for analysis. The percentage of participants (n = 40) that correctly answered the easy closed questions are displayed in Table 7.3. In total, 83% (166) of the 200 answers for the five easy closed questions were marked as being verbally answered correctly. The majority of participants were able to easily answer all easy closed questions correctly apart from CE4. The percentage of participants (n = 40) that correctly answered the hard closed questions are displayed in Table 7.4. In total, 68.5% (137) of the 200 answers for the five hard closed questions were marked as being verbally answered correctly. Most participants failed to answer CH4 and CH3 correctly but had little difficulty correctly answering the other hard closed questions. Overall, the marked closed questions revealed that the

participants had more difficulty answering the hard questions than the easy. However, three of the closed questions that experts deemed hard (CH1, CH2 and CH5) were actually easy for the participants to answer.

Table 7.3 Easy Closed Questions

Identifier	Closed Question	Correct (*n* = 40)
CE1	Termites like cold conditions.	90%
CE2	Magnetic termites build their colonies around geographical magnetic areas.	83%
CE3	A termite queen lays over one thousand eggs per day.	90%
CE4	The shafts in a termite colony penetrate twenty metres below the colony to reach the water table.	53%
CE5	There are two queens in each termite colony.	100%

Table 7.4 Hard Closed Questions

Identifier	Closed Question	Correct (*n* = 40)
CH1	The air within a termite colony is un-breathable.	100%
CH2	The temperature in a termite colony is above 35°C.	93%
CH3	The workers are all male termites.	50%
CH4	There is no king within a termite colony.	2.5%
CH5	The cellar is the coolest part of the termite colony.	98%

7.6.2 Open Question Results

Within the ten open questions, five were categorised as easy and five were categorised as hard by the experts. There were 400 answers, which broke down into 200 easy open answers and 200 hard open answers for analysis. All of the open questions were marked out of *n* marks, where *n* is the maximum number of marks. The maximum mark for the open questions was either 3 or 4 marks, with the same number of total marks (24 marks) in the hard and easy questions. The percentage of participants (*n* = 40) that correctly answered marks to the easy open questions are displayed in Table 7.5. In total, 41.7% (300) of the 720 marks for the five easy open questions were marked as being verbally answered correctly. The participants had the least difficulty answering OE5. The percentage of participants (*n* = 40) that correctly answered marks to the hard open questions are displayed in Table 7.6. In total, 24.7% (178) of the 720 marks for the five hard closed questions were marked as being verbally answered correctly. The participants had the most difficulty answering OH2. Overall, the marked open questions revealed that the participants had more difficulty answering the hard questions than the easy. However, the open questions that experts deemed easy were not that easy for the participants to answer.

7.6.3 Discussion

The analysis of the marked answers confirmed that participant comprehension levels were higher for the easy questions than the closed questions. However, the hard closed questions were still easy for the participants to answer with 68.5% correct. The closed questions are subject to guessing with a 50% likelihood of getting the right answer, so they may not always reflect the participant's true level of understanding. In contrast, the easy open questions were hard for the participants to answer correctly because only 41.7% of the total marks were correct. These findings highlight the problem of experts accurately judging question difficulty, which was also encountered in the previous study contained in Chapter 6.

Table 7.5 Easy Open Questions

Identifier	Open Question	Mark	Correct (n = 40)
OE1	Describe the main qualities you would want from a home?	4	51.9%
OE2	What is the family structure within a termite colony? Describe their basic functions.	3	53.3%
OE3	Describe why magnetic termites from Australia build their colony north/south.	4	36.9%
OE4	Describe the basic structure of a termite colony and explain how the architecture is essential to maintain the colony.	4	10.6%
OE5	What countries or regions can termites be found?	3	64.2%

Table 7.6 Hard Open Questions

Identifier	Open Question	Mark	Correct (n = 40)
OH1	Describe a symbiotic relationship and discuss why this applies to termites.	3	16.7%
OH2	Describe why it is essential for the termites to construct their colony in a tower structure.	3	15%
OH3	Discuss the role of the worker in the termite social structure.	4	26.3%
OH4	There is normal air flow within the termite colony. Describe how this is achieved using the complex architecture.	4	17.5%
OH5	A termite's food supply is dead wood. How do they digest it?	4	43.8%

7.7 Comprehension Classifier ANN Optimisation Methodology

The objective of the study was to capture a video dataset for optimising FATHOM's Comprehension Classifier ANN (Figure 5.3) using FATHOM's neural network training application (Figure 5.6). Therefore, this Section will introduce the methodology used to train and validate FATHOM's Comprehension Classifier ANN with the video dataset in a set of optimisation experiments. Within each experiment, the aim, methodology and results are reported. Detailed descriptions of where the channel dataset was extracted from the videos and the configuration parameters used to train and validate FATHOM's Comprehension Classifier ANN are also included. The approach to the

assignment of the desired responses (Section 4.4) to this dataset is different from the previous Comprehension Classifier ANN experiments in Section 6.7. Most importantly, the subsequent sections will identify whether human comprehension and non-comprehension can be detected from the participants multiple channels of nonverbal behaviour when engaging in the Q&A on the termites video. Findings from an optimisation experiment have been published in Buckingham et al. (2014).

Before the channel dataset could be collated for the optimisation experiments, some preparation was required on the video dataset. The videos were all in MP4 format. Therefore, in order to work with FATHOM they needed to be converted to AVI. MPEG Streamclip (Squared5, 2016) was used to convert the videos. The channel dataset for this set of experiments was only collated from the video when the interviewer was asking each of the twenty comprehension assessment questions (Tables 7.3-7.6) to the participant. Furthermore, in an attempt to ensure that purely high and low comprehension patterns of channel data was extracted from the participant videos for the open questions, a 75% threshold was applied e.g. if the participant got ≥ 75% of the answer right then the question was marked as correct otherwise the question was marked as incorrect. The purpose was to identify whether the participants displayed low and high nonverbal human comprehension patterns when the interviewer asked each question. Being able to predict human comprehension when questions are being asked would be more advantageous than having to wait for a verbal response and would be useful in learning environments with larger audiences where verbal responses are unfeasible. Applying the threshold to the marked open questions resulted in 17% of the easy open questions and 6% hard open questions being marked as correct.

The channel data was collated from all of the videos (n = 40) using FATHOM. None of the videos were discarded because a mini pilot study had been executed to ensure that the digital camcorder recordings were consistent and of high quality. The dataset had the desired responses assigned as follows, if the participant got the answer to the closed/open question right then the grouped channel vectors spanning the interviewer asking that question were all labelled with a comprehension (+1) desired response. Otherwise, if the participant got the answer to the closed/open question wrong then the grouped channel vectors spanning the interviewer asking that question were all labelled with a non-comprehension (-1) desired response. The entire dataset had a

total of 40,808 vectors, which broke down into 16,951 (41.54%) comprehension and 23,857 (58.46%) non-comprehension. This dataset was used to train and validate FATHOM's Comprehension Classifier ANN on human comprehension detection using the error-backpropagation learning algorithm (Section 4.4). To discover the optimal Comprehension Classifier ANN with this dataset, FATHOM's Comprehension Classifier ANN was optimised in a series of experiments, which were executed in the following order:

(1) Maximum Epochs (Section 7.8)
(2) Checking Epochs (Section 7.9)
(3) Weight Initialisation (Section 7.10)
(4) Learning Rate (Section 7.11)
(5) Topology (Section 7.12)
(6) Inputs (Section 7.13)

7.8 Maximum Epochs

Aim

The aim of the first experiment was to determine the maximum epochs value.

Methodology

The properties used in this ANN training experiment are listed in Table 7.7. There were 40 inputs to the neural network because each input vector in the dataset contained 40 normalised input signal values. A single hidden layer as was used because Cybenko (1989) has shown that a continuous feedforward single hidden layer ANN can approximate any arbitrary decision region well. As highlighted in Section 4.5.2, there is no hard-and-fast rule for deciding the starting number of neurons within the hidden layer(s). One rule of thumb is to set the size of the hidden layer between the size of the input layer and the output layer (Heaton, 2008). Therefore, the single hidden layer had 20 neurons, which is half the size of the input layer. Only one output neuron was required for outputting a value between +1 (comprehension) and -1 (non-comprehension). The topology will be investigated further in Section 7.12. A two-way data split (Section 4.5.5) was used to partition the dataset into two equally sized sets for training and testing the Comprehension Classifier ANN. Training with the error-backpropagation learning algorithm was forced to continue to 10,000 epochs by setting the maximum epoch and the checking epoch parameters to 10,000. The

learning rate (η) and the network weights were both initialised using formulae incorporating neuronal fan-in (f). The weight initialisation method used was proposed by Wessels and Barnard (1992) and has been found to perform well against other methods in Thimm and Fiesler (1995). The RMS Error and Classification Accuracy (CA) for the training and test sets were set to output at the end of every epoch so that they could be inspected in the post training analysis.

Table 7.7 ANN Training Configuration: Maximum Epochs Verification

Parameter	Value
ANN Topology	40:20:1
Training Set	50%
Testing Set	50%
Checking Epochs	10,000
Maximum Epochs	10,000
η	$1/f$
Weight Initialisation	$0 \pm 1/\sqrt{f}$

Results and Discussion

Plots of the RMS Error and the Total CA throughout the 10,000 epochs with the training and test sets are shown in Figure 7.3 and Figure 7.4, respectively. During training, the RMS Error falls then plateaus. On the other hand, the testing RMS Error falls but then gradually starts to rise because the Comprehension Classifier ANN has been trained for too long with the training set (i.e. over-fitting) causing poor generalisation performance on the test set. The Total CA plots also reflect the over-fitting on training set. Therefore, the next experiment needs to ensure that the error-backpropagation learning algorithm stops earlier during training in order to prevent over-fitting. These findings also indicate that the maximum epochs value is large enough for ensuring that an ample number of epochs are available for neural network training.

7.9 Checking Epochs

Aim

The aim of this experiment was to find the optimal checking epoch parameter value.

Methodology

The properties used in this ANN training experiment are listed in Table 7.8. The checking epoch value is used to halt training of the Comprehension Classifier ANN every n^{th} epoch so that its performance with the training set can be evaluated to enable early stopping. Different checking epoch values were used, ranging from small

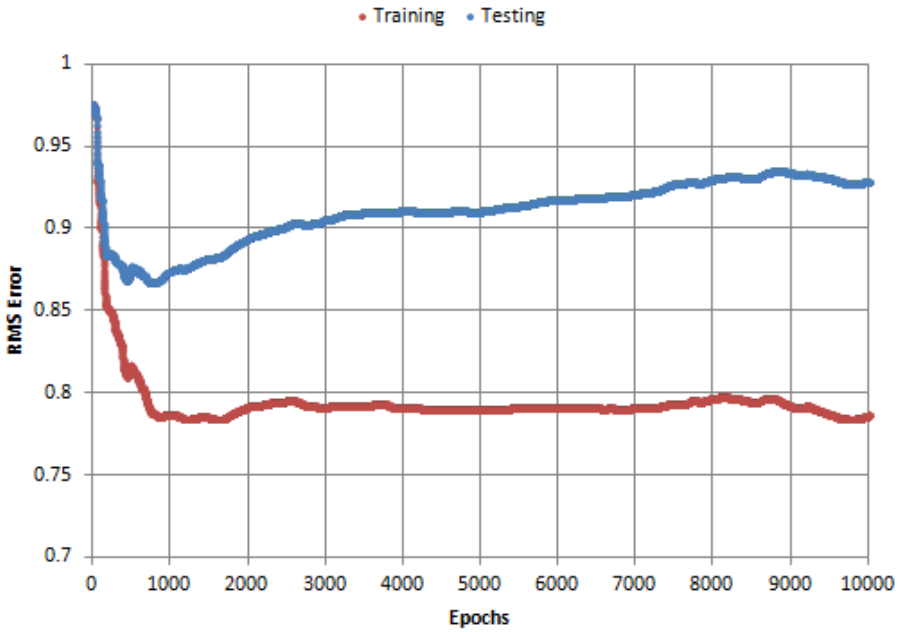


Figure 7.3 Root Mean Square Error Plot

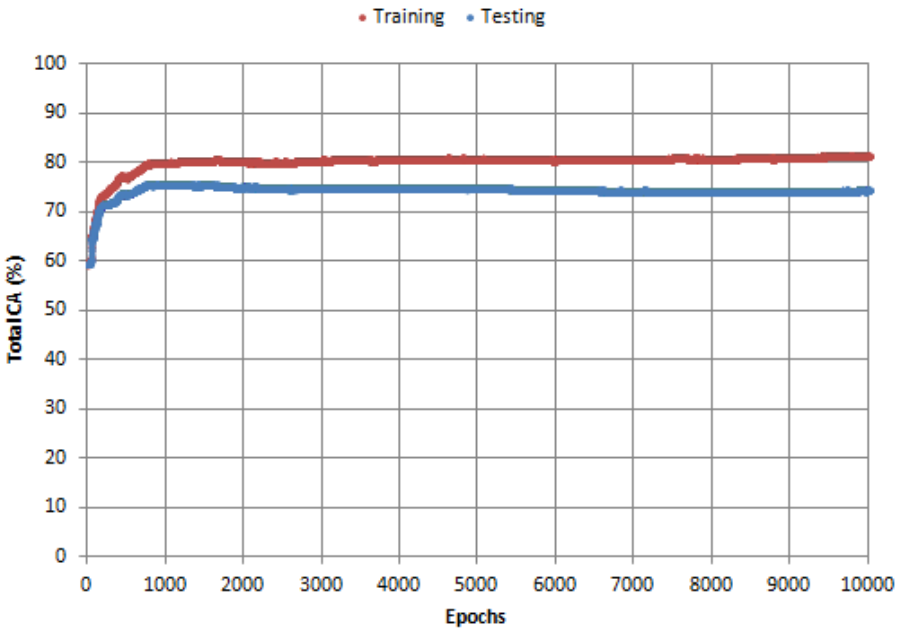


Figure 7.4 Total Classification Accuracy Plot

Table 7.8 ANN Training Configuration: Checking Epochs Verification

Parameter	Value
ANN Topology	40:20:1
k-fold cross-validation	10-fold cross-validation
Training Set	90%
Test Set	10%
Stopping Criteria	Training CA
Checking Epochs	5, 10, 25, 50, 100, 250, 500, 1000, 2000, 2500, 5000
Maximum Epochs	10,000
η	$1/f$
Weight Initialisation	$0 \pm 1/\sqrt{f}$

to large integers that were all less than the maximum number of epochs. In total, there were eleven checking epoch values to be evaluated. If the checking epoch value is too small then under-fitting may occur. In contrast, if the checking epoch value is too large then over-fitting may occur and computational time is increased. The experiment also implemented 10-fold cross-validation with the two-way data split to help counterattack over-fitting. All other parameters remained the same as Section 7.8. Eleven cross-validation experiments were executed, each using a different checking epoch number. The RMS Error and CAs for the test sets were set output so that the neural network performances could be inspected in the post training analysis.

Results and Discussion

Table 7.9 shows the average number of epochs elapsed during training, the test sets average RMS Error and the test sets average CAs for each of the eleven cross-validation experiments, which all used a different checking epoch value. The Normalised Total CA is derived from (TP CA + TN CA) / 2. As expected, the smaller checking epoch values caused training to stop prematurely, resulting in lower average CAs and higher average RMS Errors e.g. every 5^{th}, 10^{th}, 25^{th} and 50^{th} epoch. The larger checking epoch values were inflicted by over-fitting with a plateau in average CAs and increased computational costs e.g. every 500^{th}, 1000^{th}, 2000^{th}, 2500^{th} and 5000^{th} epoch. The optimum checking epoch value was every 250^{th} epoch because it yielded the highest average Total CA using the fewest average number of epochs (green row in Table 7.9). Therefore, the following experiments use the optimal checking epoch value.

Table 7.9 Cross-validation Averages: Checking Epochs Verification

Every n^{th} Epoch	Epochs	RMS Error	Total CA	Normalised Total CA	TP CA	TN CA
5^{th}	101.5	0.9016	69.309	66.713	51.379	82.051
10^{th}	173	0.873	72.353	70.705	60.958	80.45
25^{th}	305	0.861	73.838	72.424	64.067	80.779
50^{th}	480	0.8366	76.031	74.763	67.259	82.263
100^{th}	710	0.8403	76.145	75.409	71.059	79.758
250^{th}	1300	0.8421	76.394	75.631	71.122	80.138
500^{th}	1850	0.841	76.607	75.589	69.578	81.602
1000^{th}	4000	0.8472	76.47	75.258	68.086	82.428
2000^{th}	3000	0.845	76.532	75.541	69.685	81.398
2500^{th}	7000	0.8489	76.458	75.371	68.946	81.795
5000^{th}	8000	0.8441	76.617	75.502	68.917	82.089

7.10 Weight Initialisation

Aim

The aim of this experiment was to verify the weight initialisation method.

Methodology

The properties used in this ANN training experiment are listed in Table 7.10. Four weight initialisation methods have been investigated. All other parameters remained the same as previous experiment in Section 7.9. Network weights are typically initialised to small random values in the range of 0 ± 1 or 0 ± 0.5 (Fausett, 1994). Therefore, the random 0 ± 1 weight range and the user defined weight range (centre ± span) are both tested. The user defined weight initialisation method had the centre set at 0 and span set at 0.5, which produces random values in the range of 0 ± 0.5. Other centre and span values could have been used but the latter weight initialisation ranges are preferred. The weight initialisation method proposed by Wessels and Barnard (1992), 0 ± 3 $/\sqrt{f}$ where f is fan-in of the neuron was also tested. Because the latter weight initialisation is based on the order of 0 ± 1 $/\sqrt{f}$, the weight range 0 ± 1 $/\sqrt{f}$ was also tested. The RMS Error and CAs for the test sets were set output so that the neural network performances could be inspected in the post training analysis.

Table 7.10 ANN Training Configuration: Weight Initialisation Verification

Parameter	Value
ANN Topology	40:20:1
k-fold cross-validation	10-fold cross-validation
Training Set	90%
Test Set	10%
Stopping Criteria	Training CA
Checking Epochs	250
Maximum Epochs	10,000
η	$1/f$
Weight Initialisation	0 ± 1 0 ± 1 $/\sqrt{f}$ 0 ± 3 $/\sqrt{f}$ centre ± span

Results and Discussion

Table 7.11 shows the average number of epochs elapsed during training, the test sets average RMS Error and the test sets average CAs from each of the four cross-validation experiments that adopted a different weight initialisation method. The 0 ± 1 and the centre ± span weight ranges both yielded the lowest number of average training epochs and had an average RMS Error, which was lower than the 0 ± 3 $/\sqrt{f}$

weight initialisation method. However, the $0 \pm 3/\sqrt{f}$ weight initialisation method had the highest average number of training epochs, the highest average RMS error and the lowest average Total CA making it the poorest performing cross-validation experiment. On the other hand, the best performance came from the cross-validation experiment that used the $0 \pm 1/\sqrt{f}$ weight initialisation method because it had the highest average Total CA and the lowest average RMS error. Thus, making the $0 \pm 1/\sqrt{f}$ weight initialisation method the optimum (green row in Table 7.11) so shall be used in the remaining experiments.

Table 7.11 Cross-validation Averages: Weight Initialisation Verification

Weight Initialisation	Epochs	RMS Error	Total CA	Normalised Total CA	TP CA	TN CA
0 ± 1	750	0.839	76.573	75.581	69.72	81.441
$0 \pm 1/\sqrt{f}$	1025	0.8314	76.83	75.642	68.621	82.662
$0 \pm 3/\sqrt{f}$	1225	0.8442	76.209	75.204	69.278	81.133
centre ± span	1000	0.835	76.694	75.558	68.851	82.265

7.11 Learning Rate

Aim

The aim of this experiment was to locate an optimal learning rate (η) value.

Methodology

Table 7.12 lists the properties used in this training experiment. Smaller η result in a slower but smoother gradient descent and larger η result in a faster gradient descent but can lead to an unstable neural network as highlighted in Section 4.5.4. Therefore, the η should be a large as possible without resulting in oscillations (Figure 4.5) (Battiti, 1989). Furthermore, Zurada (1992) found that η values ranging from 0.005 to 10 have been used successfully with the error-backpropagation learning algorithm in experiments reviewed in previous works. Therefore, the majority of the η values to be tested are within the 0.005 to 10 range. $1/f$ where f is fan-in of the neuron was also tested and was the only cross-validation experiment where the η was different across the neural networks layers. All other properties remained the same as previous experiment in Section 7.10. In total, there were 26 different learning rates, ranging from small to large values to be tested. The RMS Error and CAs for the test sets were set output so that the neural network performances could be inspected in the post training analysis.

Table 7.12 ANN Training Configuration: Learning Rate Verification

Parameter	Value
ANN Topology	40:20:1
k-fold cross-validation	10-fold cross-validation
Training Set	90%
Test Set	10%
Stopping Criteria	Training CA
Checking Epochs	250
Maximum Epochs	10,000
η	$1/f$, 1, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0.09, 0.07, 0.05, 0.03, 0.01, 0.009, 0.007, 0.005, 0.003, 0.001, 0.0009, 0.0007, 0.0005, 0.0003 and 0.0001
Weight Initialisation	$0 \pm 1/\sqrt{f}$

Results and Discussion

Table 7.13 shows the average number of epochs elapsed during training, the test sets average RMS Error and the test sets average CAs from each of the learning rate cross-validation experiments. As the η got smaller the average number of training epochs increased. The larger learning rates such as 0.3 to 1 were resulting in unstable neural networks that had not converged with a high average RMS error and lower average CAs. In contrast, the smaller learning rates such as 0.0003 to 0.0009 took longer to converge with little improvement in RMS Error and CAs. The 0.0001 learning rate experiment failed to converge within the maximum number of epochs in three out of ten folds of the 10-fold cross-validation. Although the $1/f$ learning rate had a low average number of training epochs, it had a high average RMS Error. Overall, the 0.005 learning rate was the optimal η value because it yielded the lowest average RMS Error with the highest average Total CA in the fewest epochs (green row in Table 7.13). Therefore, in the subsequent experiments the η was fixed at 0.005 on all the neural network layers.

7.12 Topology

Aim

The aim of this experiment was to determine the optimum number of hidden neurons and hidden layers for the neural networks topology.

Methodology

Table 7.14 lists the properties used in this training experiment. As highlighted in Section 4.5.2, smaller neural network topologies that generalise well are preferred over large neural networks. Therefore, in this set of experiments, a single hidden layer ANN was optimised first and then a two hidden layer ANN. The number of hidden

Table 7.13 Cross-validation Averages: Learning Rate Verification

η	Epochs	RMS Error	Total CA	Normalised Total CA	TP CA	TN CA
$1/f$	1150	0.8383	76.603	75.522	69.119	81.925
0.0001	9225	0.7744	79.379	78.43	72.815	84.041
0.0003	5975	0.7631	80.416	79.489	74.007	84.968
0.0005	5525	0.7648	80.29	79.182	72.652	85.714
0.0007	3550	0.7682	80.162	79.271	73.993	84.548
0.0009	3450	0.7719	79.867	78.674	71.63	85.717
0.001	3400	0.7654	80.313	79.369	73.789	84.946
0.003	2175	0.7812	79.271	78.075	71	85.149
0.005	2275	0.7796	79.391	78.433	72.77	84.097
0.007	1875	0.7828	79.4	78.256	71.488	85.023
0.009	2200	0.7906	78.818	78.016	73.271	82.764
0.01	1250	0.7932	78.566	77.079	68.296	85.861
0.03	1175	0.8286	76.578	75.682	70.391	80.972
0.05	950	0.8603	75.284	74.612	70.647	78.582
0.07	625	0.8836	74.319	73.016	65.313	80.717
0.09	625	0.907	73.806	72.868	67.338	78.4
0.1	1025	0.916	73.79	72.747	66.575	78.915
0.2	675	0.9947	72.8	71.387	63.036	79.736
0.3	1000	1.06	69.615	67.16	52.647	81.675
0.4	925	1.1347	66.511	61.719	33.404	90.036
0.5	800	1.2582	59.005	56.382	40.872	71.89
0.6	300	1.3039	57.324	50.784	12.127	89.443
0.7	300	1.3831	51.779	50.473	42.729	58.218
0.8	275	1.3605	53.386	50.008	30.017	70
0.9	250	1.3366	55.07	50	20	80
1	250	1.3847	51.686	50	40	60

neurons in the optimised single hidden layer ANN was used to form the first hidden layer of the two hidden layer ANN. Because Cybenko (1988) demonstrated that only two hidden layer ANNs with continuous transfer functions are required to approximate any arbitrary nonlinear continuous function, no more than two hidden layers were used. For the single hidden layer ANN experiments, 32 different sized hidden layers were tested, ranging from 2 hidden neurons up to 105 hidden neurons. For the two hidden layer ANN experiments, 15 different sized hidden layers were tested, ranging from 2 hidden neurons up to 35 hidden neurons within the second hidden layer. All other properties remained the same as previous experiment in Section 7.11. The RMS Error and CAs for the test sets were set output so that the neural network performances could be inspected in the post training analysis.

Table 7.14 ANN Training Configuration: Topology Verification

Parameter	Value
ANN Topology	Single Hidden Layer ANN 40:n:1 where n = 2, 4, 6, 8, 10, 12, 15, 17, 20, 22, 25, 27, 30, 32, 35, 37, 39, 40, 42, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 or 105 neurons Two Hidden Layer ANN 40:n:m:1 where m = 2, 4, 6, 8, 10, 12, 15, 17, 20, 22, 25, 27, 30, 32 or 35 neurons
k-fold cross-validation	10-fold cross-validation
Training Set	90%
Test Set	10%
Stopping Criteria	Training CA
Checking Epochs	250
Maximum Epochs	10,000
η	0.005
Weight Initialisation	$0 \pm 1/\sqrt{f}$

Results and Discussion

The average number of epochs elapsed during training, the test sets average RMS Error and the test sets average CAs for the single hidden layer ANN cross-validation experiments are in Table 7.15. The single hidden layer ANNs with a smaller number of hidden neurons tended to had a higher average RMS error and lower CAs because the neural networks were unable to converge to suitable solution due to having too few weights e.g. 40:2:1, 40:4:1, 40:6:1 etc. Naturally, some of the larger neural network took longer to train than the smaller neural networks due to having more neurons and weights e.g. 40:85:1, 40:90:1, 40:95:1 etc. The 40:60:1 topology was selected as the optimal size for a single hidden layer ANN because it yielded a higher average Normalised Total CA than the 40:65:1 topology and the smaller ANNs (green row in

Table 7.15). Furthermore, the performance of 40:65:1 topology did not greatly improve in comparison to the 40:60:1 topology.

Table 7.15 Single Hidden Layer ANN Cross-validation Averages

Topology	Epochs	RMS Error	Total CA	Normalised Total CA	TP CA	TN CA
40:2:1	2725	0.9557	62.202	58.805	38.743	78.871
40:4:1	1400	0.9266	66.244	63.232	45.433	81.033
40:6:1	1225	0.9055	68.677	66.46	53.356	79.564
40:8:1	1275	0.883	70.77	68.927	58.032	79.82
40:10:1	1600	0.862	72.489	71.017	62.334	79.703
40:12:1	1850	0.842	74.504	72.795	62.681	82.906
40:15:1	1675	0.8233	76.096	74.453	64.746	84.161
40:17:1	2250	0.799	77.997	76.922	70.58	83.268
40:20:1	2275	0.7796	79.391	78.433	72.77	84.097
40:22:1	2300	0.7733	79.906	78.86	72.676	85.044
40:25:1	1800	0.756	81.088	80.107	74.314	85.899
40:27:1	2150	0.7412	82.167	81.308	76.22	86.393
40:30:1	2025	0.7283	82.942	82.032	76.647	87.417
40:32:1	2000	0.7271	83.356	82.508	77.495	87.52
40:35:1	1925	0.7172	84.157	83.516	79.718	87.312
40:37:1	1625	0.7145	84.399	83.695	79.547	87.844
40:39:1	1825	0.708	84.953	84.336	80.692	87.978
40:40:1	1800	0.6989	85.277	84.532	80.12	88.941
40:42:1	1500	0.6924	85.492	84.918	81.532	88.304
40:45:1	1775	0.686	86.207	85.549	81.659	89.44
40:50:1	1500	0.6819	86.038	85.605	83.034	88.172
40:55:1	1950	0.6657	87.13	86.61	83.535	89.684
40:60:1	2150	0.6612	87.45	86.977	84.167	89.784
40:65:1	2000	0.6532	87.47	86.871	83.324	90.416
40:70:1	2325	0.6382	88.329	87.829	84.869	90.79
40:75:1	2425	0.6396	88.294	87.682	84.073	91.292
40:80:1	2000	0.6434	87.844	87.232	83.612	90.849
40:85:1	3000	0.6266	88.733	88.112	84.444	91.778
40:90:1	2550	0.6269	88.792	88.205	84.732	91.68
40:95:1	2575	0.6243	88.812	88.34	85.548	91.128
40:100:1	2300	0.6387	88.221	87.594	83.891	91.297
40:105:1	2300	0.6359	88.374	87.881	84.965	90.799

The average number of epochs elapsed during training, the test sets average RMS Error and the test sets average CAs for the two hidden layer ANN cross-validation experiments are presented in Table 7.16. The computational time required for the two hidden layer ANNs tended to be lower than single hidden layer ANNs. This may be because the additional network weights enabled the neural networks to find an optimal solution more easily or were over-fitting the training dataset. Topology

40:60:2:1 had a lower average Total CA and a lower average TP CA than the optimum 40:60:1 topology. However, the smaller single hidden layer topology (40:60:1) was preferred as the optimal neural network topology (green row in Table 7.15) and is used in the subsequent experiments.

Table 7.16 Two Hidden Layer ANN Cross-validation Averages

Topology	Epochs	RMS Error	Total CA	Normalised Total CA	TP CA	TN CA
40:60:2:1	1850	0.6526	87.272	86.709	83.387	90.032
40:60:4:1	1425	0.6416	87.693	87.218	84.415	90.023
40:60:6:1	1050	0.6167	88.656	88.122	84.958	91.287
40:60:8:1	900	0.6157	88.751	88.327	85.805	90.844
40:60:10:1	1300	0.593	89.534	89.242	87.523	90.962
40:60:12:1	1025	0.5828	89.921	89.618	87.836	91.402
40:60:15:1	925	0.5631	90.659	90.163	87.236	93.091
40:60:17:1	900	0.5576	90.801	90.343	87.635	93.051
40:60:20:1	1000	0.5285	91.84	91.502	89.505	93.498
40:60:22:1	1025	0.5321	91.694	91.391	89.606	93.175
40:60:25:1	1075	0.5141	92.323	91.98	89.936	94.022
40:60:27:1	1000	0.5001	92.732	92.389	90.36	94.416
40:60:30:1	1325	0.493	92.926	92.567	90.456	94.681
40:60:32:1	950	0.4855	93.021	92.725	90.987	94.464
40:60:35:1	1550	0.464	93.794	93.501	91.765	95.238

7.13 Inputs

The primary aim of this set of experiments was to determine the optimum number inputs for the FATHOM Comprehension Classifier ANN. To discover the optimal number of inputs, FATHOM's Comprehension Classifier ANN was trained and validated using four different experimental approaches, which were executed in the following order:

(1) All Inputs (Section 7.13.1)

(2) Rotated Pruning of Individual Inputs (Section 7.13.2)

(3) Input Information Gain (Section 7.13.3)

(4) Grouping Inputs by Theme (Section 7.13.4)

The properties used in all of the ANN training experiments in Section 7.13.1 to Section 7.13.4 are listed in Table 7.17. Only the numbers of inputs (n) to the neural network changes across the experiments, all other properties remained the same from the previous optimisation experiments in Sections 7.8 to 7.12. Four new known

channels (Section 5.4.3) have been added to the dataset: age, education level, occupation and prior knowledge. The age channel was selected because the previous works in Section 3.5 on human comprehension detection from nonverbal behaviour has focused upon participants of different ages i.e. children, students and adults. Therefore, age may be a key factor in human comprehension detection. Education level has been included because the highest qualification attained by an individual provides an indicator of previous academic performance and expertise in a subject area(s), which may aid detection of comprehension patterns as it shows the individuals level of understanding evaluated from standardised assessment tools. Occupation was incorporated to see if comprehension patterns were related to whether the participants were a student or a member of teaching/non-teaching staff. The prior knowledge input channel provides the participant's self-perception measure of their prior knowledge on the topic of termites. Including the prior knowledge channel will identify whether participants that regard themselves as having more prior knowledge on the termites are at an advantage or not.

The new channels were extracted from the Data Collection Form (Appendix N) used in this study. Before the new known channels could be appended to the dataset they had to be normalised into the bipolar range (Section 4.5.1). The categories within each known channel and the percentage of participants that selected the each category can be found in the demographic results (Table 7.2). For the prior knowledge channel, the high category was not included in the normalised range because none of the participants selected the category, as shown in Table 7.2. All of the new known channels have been used in the subsequent input optimisation experiments. Lastly, Section 7.13.5 provides a summary of the findings from experiments in Sections 7.13.1 to 7.13.4 and determines the optimal number inputs for the FATHOM Comprehension Classifier ANN.

Table 7.17 ANN Training Configuration: Inputs Verification

Parameter	Value
ANN Topology	n:60:1
k-fold cross-validation	10-fold cross-validation
Training Set	90%
Test Set	10%
Stopping Criteria	Training CA
Checking Epochs	250
Maximum Epochs	10,000
η	0.005
Weight Initialisation	$0 \pm 1/\sqrt{f}$

7.13.1 All Inputs

In this single cross-validation experiment, all of the four new known channels (age, education level, occupation and prior knowledge) were added as inputs along with the existing 40 input channels. However, the 40 existing input channels contained three redundant inputs (planning, race and slot) so they were excluded from the dataset as inputs to the Comprehension Classifier ANN in this experiment. By excluding the redundant input channels this resulted in a total of 41 inputs to the Comprehension Classifier ANN, which had a 41:60:1 topology. Table 7.17 lists the properties used in this training experiment. The RMS Error and CAs for the test sets were set output so that the neural network performances could be inspected in the post training analysis.

The test sets average RMS Error and average CAs for this cross-validation experiment is shown in Table 7.18. By comparing the average Normalised Total CA (91.349%) in Table 7.18 against the optimum topology (41:60:1, green row in Table 7.15) from the previous experiment it shows that the Normalised Total CA has increased by 4.372% with the new four input channels. Furthermore, the average RMS Error decreased and the average TP and TN CAs increased. The results from this experiment are used as the optimal baseline for comparative purposes in the subsequent input optimisation experiments. The following experiment is going to investigate the significance of each input via pruning.

Table 7.18 Cross-validation Averages: All Inputs Verification

Topology	RMS Error	Total CA	Normalised Total CA	TP CA	TN CA
41:60:1	0.5293	91.835	91.349	88.493	94.21

7.13.2 Rotated Pruning of Individual Inputs

In this experiment, the 41 inputs were rotated through so that each input was removed from the dataset once, in order to discover the impact of the pruning of the individual input. Therefore, there were 41 cross-validation experiments all having the same sized neural network topology: 40:60:1. The properties used in this training experiment are listed in Table 7.17. The RMS Error and CAs for the test sets were set output so that the neural network performances could be inspected in the post training analysis.

The test sets average RMS Error and average CAs for the 41 cross-validation experiments are shown in Table 7.19. The red cells in Table 7.19 highlight the neural

Table 7.19 Cross-validation Averages: Rotated Input Pruning Verification

Input Pruned	RMS Error	Total CA	Normalised Total CA	TP CA	TN CA
sex	0.5524	91.257	90.887	88.704	93.07
fvm	0.5175	92.4	92.102	90.332	93.871
fhm	0.513	92.513	92.221	90.497	93.946
fs	0.5262	92.013	91.559	88.881	94.239
fblu	0.5802	90.32	89.814	86.822	92.807
fbla	0.5725	90.657	90.348	88.521	92.174
fum	0.5162	92.381	92.06	90.173	93.948
fdm	0.5223	92.261	91.87	89.558	94.181
flm	0.5246	92.127	91.887	90.473	93.299
frm	0.5325	91.815	91.315	88.349	94.278
ffm	0.5128	92.499	92.136	90.001	94.274
fbm	0.5107	92.635	92.265	90.091	94.441
fvs	0.5252	92.198	91.756	89.147	94.366
fhs	0.5375	91.816	91.41	89.01	93.808
fvsn	0.5324	91.791	91.418	89.227	93.611
fhsn	0.5355	91.814	91.452	89.329	93.577
lblink	0.5121	92.512	92.152	90.019	94.286
lleft	0.5465	91.465	91.114	89.049	93.184
lright	0.5532	91.171	90.784	88.498	93.071
lshift	0.5353	91.697	91.319	89.094	93.544
lclosed	0.5332	91.805	91.494	89.649	93.339
lhleft	0.5563	91.07	90.722	88.661	92.78
lhright	0.5305	91.926	91.591	89.617	93.565
lhclosed	0.5397	91.634	91.379	89.871	92.886
rblink	0.5133	92.547	92.256	90.544	93.969
rleft	0.5476	91.369	90.916	88.243	93.59
rright	0.5305	91.95	91.637	89.778	93.499
rshift	0.5368	91.712	91.259	88.586	93.934
rclosed	0.5397	91.719	91.297	88.786	93.803
rhleft	0.5462	91.294	90.927	88.764	93.091
rhright	0.5305	91.95	91.637	89.778	93.499
rhclosed	0.5558	91.126	90.589	87.417	93.762
fmc	0.5289	91.92	91.529	89.221	93.837
fmac	0.5216	92.242	91.78	89.057	94.505
fma	0.5201	92.222	91.885	89.895	93.876
fmuor	0.5246	92.193	91.797	89.461	94.136
fmuol	0.5169	92.462	92.118	90.078	94.156
age	0.5464	91.46	91.062	88.704	93.418
education level	0.559	90.962	90.501	87.771	93.228
occupation	0.5582	91.028	90.651	88.408	92.89
prior knowledge	0.5612	90.882	90.486	88.155	92.819

networks, which yielded a lower average Normalised Total CA than the 91.349% baseline (in Table 7.18) when the specified input was pruned from the ANN. There are 17 red cells, which suggest that each of those independently pruned inputs play an important role in the detection of human comprehension detection. Therefore, paired sample t-tests were performed for the each of the 41 cross-validation experiments with the baseline ANN (Table 7.18) to determine whether there was a statistical significant difference between the testing Normalised Total CA from the sample with all inputs and the sample with the input pruned. The statistical analyses were performed using IBM SPSS Statistics software. The paired sample t-tests revealed that none of the pairs were significant at $p < 0.05$. With the Bonferroni Correction none of the paired sample t-tests were significant at $p < 0.0012195$. To perform the Bonferroni Correction, the p value is divided by the number of comparisons being made e.g. 0.05/41. Therefore, the latter statistical findings support the hypothesis that the knowledge resides across all of the neural network inputs rather than residing within individual inputs i.e. a holistic approach.

7.13.3 Input Information Gain

Information Gain (InfoGain) 'evaluates attributes by measuring their information gain with respect to the class' (Witten and Frank, 2005:422). InfoGain takes a Minimum Description Length (Rissanen, 1983) approach to make the continuous numeric attributes discrete (Witten and Frank, 2005). The InfoGain formula is calculated as

$$InfoGain(Class, Attribute) = H(Class) - H(Class \mid Attribute) \quad (7.1)$$

where H is entropy. Shannon (1948) developed the Entropy formula. The entropy $H(X)$ of variable X is defined as

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i) \quad (7.2)$$

measured in bits, where $p(x_i)$ is the probability that X is in the state x_i.

Using InfoGain will determine which input(s) in this dataset are the most important and useful for discriminating between the comprehension and non-comprehension classes. The Weka (Hall et al., 2009) (version 3.6.10) implementation of InfoGain was used to evaluate the dataset from this study with all 41 inputs. The ranked results from the InfoGain evaluation with the dataset from this study are displayed in Table 7.20.

Attributes with a larger InfoGain value are ranked higher than attributes with a lower InfoGain e.g. fbm is ranked higher than fhs. The fmour attribute has the highest InfoGain value. Both the fs and fmac attributes have no InfoGain. Interestingly, out of the four new known channels the occupation and age attributes are the most important.

Table 7.20 Information Gain Rank

Rank	Attribute	Information Gain
1	fmuor	0.013253
2	fmuol	0.013253
3	lclosed	0.009682
4	fblu	0.007818
5	fbla	0.005442
6	fbm	0.004675
7	fhm	0.004524
8	occupation	0.004333
9	age	0.003582
10	fhsn	0.003544
11	lright	0.002819
12	lleft	0.002482
13	lhleft	0.002366
14	lshift	0.001901
15	fmc	0.001879
16	frm	0.001839
17	lhright	0.001791
18	rblink	0.001733
19	sex	0.001675
20	flm	0.001662
21	fum	0.001658
22	lhclosed	0.00162
23	rleft	0.001582
24	education level	0.001542
25	lblink	0.001531
26	fma	0.001418
27	rhright	0.001292
28	rclosed	0.001283
29	rhclosed	0.001278
30	fhs	0.001112
31	prior knowledge	0.001111
32	rshift	0.001012
33	fdm	0.001005
34	rhleft	0.00093
35	fvm	0.000702
36	fvsn	0.000644
37	ffm	0.000452
38	rright	0.000315
39	fvs	0.000262
40	fs	0
41	fmac	0

In this experiment, the inputs were divided into two groups using the InfoGain results. The first group named “top” contained all of the attributes that were ranked from 1 to 21 in Table 7.20, which resulted in a total of 21 inputs. The second group named “bottom” contained all of the attributes that were ranked from 22 to 41 in Table 7.20, which resulted in a total of 20 inputs. The top group contained the top half

of the attributes with the highest levels of InfoGain in comparison to the bottom group. Therefore, it was anticipated that when each group was used to train and validate a neural network that the top group's ANN would perform better than the bottom group's ANN. Two cross-validation experiments were executed to compare the performance of the top and bottom groups and to identify whether top group performs better than the baseline ANN with all 41 inputs. The ANN topologies for each group were: 21:60:1 (top) and 20:60:1 (bottom). Table 7.17 lists the properties used in this training experiment. The RMS Error and CAs for the test sets were set output so that the neural network performances could be inspected in the post training analysis.

Table 7.21 shows the test sets average RMS Error and the average CAs results from the two cross-validation experiments. As predicted, the top group performed better than the bottom group, yielding higher CAs and a lower RMS Error. However, the top groups CAs and RMS Error are lower than the baseline ANN in Table 7.18. The experiment in Section 7.13.4 is going to investigate the neural network performances with some other input groups.

Table 7.21 Cross-validation Averages: Information Gain Input Groups Verification

Group	RMS Error	Total CA	Normalised Total CA	TP CA	TN CA
Top	0.6637	87.035	86.538	83.606	89.469
Bottom	0.7821	80.445	79.266	72.308	86.224

7.13.4 Group Inputs by Theme

For this set of cross-validation experiments, the inputs were grouped based on seven themes: demographics, eyes, left eye, right eye, face, face angle and face movement. The input channels within each theme are listed in Table 7.22. The purpose of this experiment was to identify whether any of the themes would produce a similar performance to the baseline ANN and to compare the themes. There were seven cross-validation experiments, one per theme. The only difference between the set of experiments was the number of inputs to the neural networks. The properties used in this training experiment are listed in Table 7.17. The RMS Error and CAs for the test sets were set output so that the neural network performances could be inspected in the post training analysis.

The test sets average RMS Error and the average CAs results from the cross-validation experiments are shown in Table 7.23. None of the cross-validation

experiments performed as well or better than the baseline ANN (Table 7.18). Interestingly, Group 2, which contained 16 input channels solely focused upon the left and right eyes, was the only themed experiment that produced ANNs that could reliably distinguish between the comprehension and non-comprehension classes at above chance levels. Thus, this result further supports the general finding from previous eye tracking studies (Section 3.5) on human comprehension detection, that eye behaviour can provide a window on human cognition. The findings from this set of experiments also support the hypothesis that the human comprehension detection knowledge within the dataset resides across all input channels rather than individual input channels i.e. a holistic approach.

Table 7.22 Inputs Grouped by Theme

Group	Theme	Inputs	Total
1	Demographics	sex, age, occupation, education level, prior knowledge	5
2	Eyes	lblink, lleft, lright, lshift, lclosed, lhleft, lhright, lhclosed, rblink, rleft, rright, rshift, rclosed, rhleft, rhrigh, rhclosed	16
3	Left Eye	lblink, lleft, lright, lshift, lclosed, lhleft, lhright, lhclosed	8
4	Right Eye	rblink, rleft, rright, rshift, rclosed, rhleft, rhright, rhclosed	8
5	Face	fvm, fhm, fs, fblu, fbla, fum, fdm, flm, frm, ffm, fbm, fvs, fhs, fvsn, fhsn, fmc, fmac, fma, fmuor, fmuol	20
6	Face Angle	fmc, fmac, fma, fmuor, fmuol	5
7	Face Movement	fvm, fhm, fs, fum, fdm, flm, frm, ffm, fbm, fvs, fhs, fvsn, fhsn, fmc, fmac, fma, fmuor, fmuol	18

Table 7.23 Cross-validation Averages: Themed Input Groups Verification

Group	RMS Error	Total CA	Normalised Total CA	TP CA	TN CA
1	0.9796	58.348	51.518	11.151	91.885
2	0.8076	78.374	77.583	72.907	82.261
3	0.9623	61.744	56.769	27.368	86.169
4	0.9512	62.569	60.031	45.036	75.025
5	0.9352	65.007	61.747	42.487	81.007
6	0.9829	58.447	50.199	1.462	98.933
7	0.9727	59.92	54.055	19.403	88.707

7.13.5 Summary

Four different approaches have been taken to try and determine the optimum number inputs for the FATHOM Comprehension Classifier ANN. In the first experiment (Section 7.13.1), the introduction of the four new known channels with the existing input channels greatly improved the Normalised Total CA of the Comprehension Classifier ANN by 4.372%, resulting in a 91.349% Normalised Total CA. Thus, forming a baseline for the remaining experiments. In the second experiment (Section 7.13.2), the

pruning of individual inputs resulted in 17 (41%) cross-validation experiments performing below the baseline and 24 (59%) cross-validation experiments with performances above the baseline. Furthermore, the cross-validation experiments that produced ANNs with performances above the baseline, the improved Normalised Total CAs was no greater than 1%. Statistical analyses also found that there was no significant difference between the rotated pruning of individual inputs and the baseline (all inputs) cross-validation experiment. In the third experiment (Section 7.13.3), Information Gain revealed which inputs in the dataset were the most useful for discriminating between the two classes. Although the top group did not perform as well as the baseline, it still yielded a higher Total CA (with only 21 inputs) and lower RMS error than any of the themed input groups. In the last experiment (Section 7.13.4) the themed input group cross-validation performance results were compared. Group 2 (eye) was the only the themed group, capable of distinguishing between the two classes. Throughout the input experiments the findings suggest that human comprehension detection knowledge within the dataset resides across multiple input channels rather than individual input channels. To conclude, the optimal number of inputs was 41 from the all inputs experiment in Section 7.13.1.

7.14 Conclusion

This chapter described a new study methodology using a learning environment to test whether general human comprehension patterns could be detected from the male and female participant's nonverbal channels. The participants were all English speakers from the UK. The study was designed to capture 40 participants in videos that could be used at a later date to train and validate FATHOM's Comprehension Classifier ANN on human comprehension detection during the participants questioning on the factual video. Descriptions on how FATHOM's Comprehension Classifier ANN was trained with channel data extracted from the videos were provided. Four new known channels were added to the channel dataset. Throughout the optimisation experiments (Sections 7.8 – 7.13), FATHOM's Comprehension Classifier ANN was able to detect non-comprehension more easily than comprehension, which may be due to the non-comprehension class being 16.92% larger than the comprehension class. Optimisation of the FATHOM Comprehension Classifier ANN's inputs revealed that the human comprehension detection knowledge resides across multiple inputs, not individual

inputs. The optimisation experiments yielded an optimal FATHOM Comprehension Classifier ANN (topology 41:60:1) capable of reliably detecting human comprehension in a learning environment from male and female nonverbal channels with a 91.835% average Total CA. Applying FATHOM to a new dataset from a different ethnic origin, which also included male participants, has broadened the findings on human comprehension detection using multichannels of nonverbal behaviour. Furthermore, this study has also demonstrated that human comprehension can be detected during question asking whereas the previous study in Chapter 6 focused upon comprehension during the reading of the learning task script. Chapter 9 attempts to replace FATHOM's Comprehension Classifier ANN (Figure 5.3) with a decision tree. The extracted channel dataset from this study is used to optimise the decision tree to identify whether it is a more suitable classifier for human comprehension detection.

Chapter 8 Detecting Human Comprehension Using Decision Trees

8.1 Introduction

In the previous two chapters, FATHOM's Comprehension Classifier (Figure 5.3) is an ANN, which has been successfully trained and validated to detect human comprehension detection from multiple channels of nonverbal behaviour. Although sufficiently sized ANNs are capable of approximating any arbitrary nonlinear continuous function (Cybenko, 1988) they use a "black box" approach so the structure of the ANN does not provide any insight on how the knowledge acquired between the inputs and the classification at the output neuron is used. To circumvent the neural network "black box" problem, a classifier that uses a "white box" approach is required so that the knowledge acquired between the inputs and the classifications outputted are visible and can be analysed. Decision trees are another type of classifier, which use a "white box" approach by pictorially representing knowledge as a set of decision rules in a top-down graph. Given the fact that FATHOM's generic architecture is modular, FATHOM's Comprehension Classifier could easily be replaced with a decision tree. Therefore, the purpose of this Chapter is to identify whether a decision tree is capable of being a Comprehension Classifier and whether the knowledge acquired in the model is comprehensible. This chapter starts by describing the decision tree architecture and then introduces C4.5 (Quinlan, 1993), a well known decision tree induction algorithm. The methodology for optimising a decision tree as the FATHOM Comprehension Classifier is outlined and the results from each optimisation experiment is discussed.

8.2 Decision Trees

Decision Trees are classifier models, which illustrate knowledge learnt from a dataset, as decision rules in a top-down structure. An example of a decision tree is shown in Figure 8.1, which was constructed from Quinlan's (1986) well-known weather-based dataset to decide whether to play or not play outdoors. There are three main components within a decision tree: attribute nodes, leaf nodes and branches. Each attribute node has an associated attribute-based splitting test, which has a branch for each outcome. A branch only leads to one outcome, either an

attribute node or a leaf node. The leaf node represents the classification as the class name. In a decision tree, the topmost attribute node, which has no incoming branches, is referred to as the root node. When classifying an instance from the dataset, the root node is always forms the starting point and one continues traversing a path down the decision tree until reaching a leaf node. For example, to classify the instance x_1 = (**Outlook**=*rain*, **Windy**=*false*) in Figure 8.1, take the following route: (1) start at the **Outlook** root node; (2) take the rightmost *rain* branch; (3) at the **Windy** attribute node take the rightmost *false* branch; and (4) terminate at the Play leaf node, which classifies the instance as a suitable day for playing outdoors. Each path taken from the root node to a leaf node can be represented as a distinct classification rule. Classification rules are also known as decision rules or production rules. For example, the classification rule for x_1 would be

$$IF\ outlook = rain\ AND\ windy = false \\ THEN\ class = play. \tag{8.1}$$

Decision trees come in different shapes and sizes e.g. shallow/deep, bushy/skinny and uniform/skewed (Kotsiantis, 2013). Not all decision trees are of equal worth (Quinlan, 1990) so performance metrics (Section 8.3.2) have to be used in order to distinguish their importance. Tree size is used to report how large a decision tree is, from the total number of nodes in the tree. For example, in Figure 8.1 the tree size is 8. The following Section describes how decision trees are constructed from a given dataset.

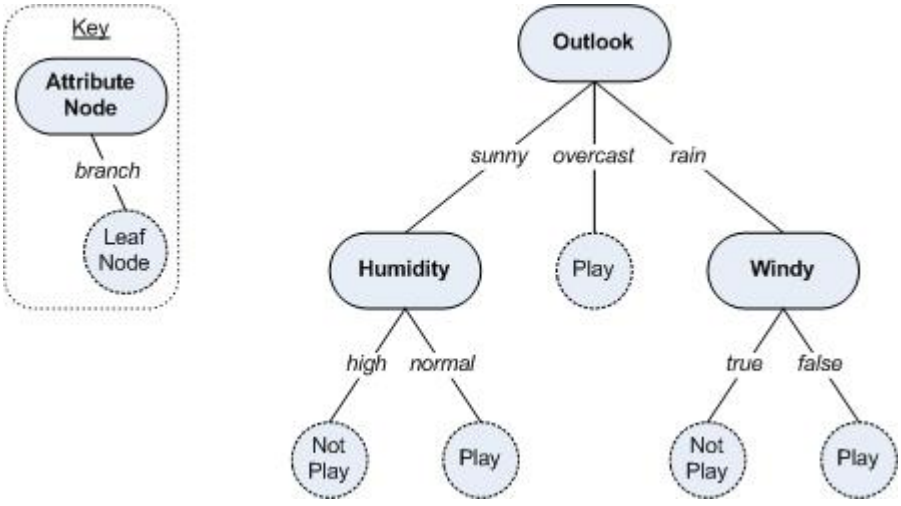


Figure 8.1 Decision Tree

8.3 Decision Tree Induction

The decision tree induction process is composed of two main phases: growing (Section 8.3.1) and pruning (Section 8.3.2). In the growth phase, the decision tree is constructed top-down using a divide and conquer algorithm with recursive partitioning. In the pruning phase, the decision tree is trimmed to reduce complexity and to improve predictive accuracy. The pruning phase can occur during or after the growing phase, depending upon the pruning technique used. Kotsiantis (2013) and Murthy (1998) both provide detailed reviews on decision tree induction.

8.3.1 Growing

The J48 open source Java implementation of the C4.5 algorithm in the Weka (Hall et al., 2009) data mining software tool is used for constructing the decision trees contained within this Chapter. The C4.5 (Quinlan, 1993) algorithm is a descendant of the Iterative Dichotomiser 3 (ID3) algorithm (Quinlan, 1986) and was inspired by the Concept Learning System (Hunt, 1962; Hunt et al., 1966). In 2006, the C4.5 algorithm was rated #1 in the top 10 data mining algorithms (Wu et al., 2008). C4.5 Release 8 was the last non-commercial release of the C4.5 algorithm. C5.0 (See5) supersedes C4.5 and was released as a commercial software application by RuleQuest Research (2010). A comparison between C5.0 and C4.5 can be found at RuleQuest Research (2012).

The training set contains a set of patterns (instances) known as the attribute-class pairs ($\mathbf{p}$ =[($\mathbf{a}^1$,$\mathbf{c}^1$),($\mathbf{a}^2$,$\mathbf{c}^2$) ... ($\mathbf{a}^p$,$\mathbf{c}^p$)]), where $\mathbf{a}^p = [a_1, a_2 \ldots a_n]^T$ is the attribute vector and $\mathbf{c}^p = [c_1, c_2 \ldots c_n]^T$ is the class vector. The attributes can be discrete or continuous values and there must be two or more discrete classes. To construct a decision tree from $\mathbf{p}$, using Hunt et al.'s (1966) divide and conquer algorithm in C4.5 (Quinlan, 1993), the following three conditional steps are repeated until all branches of the decision tree terminate with a leaf node:

1. If $\mathbf{p}$ contains instances that all belong to the same class c_n (i.e. pure node) then create a leaf node and label it with the class name c_n.
2. If $\mathbf{p}$ contains no instances, then create a leaf node and label it with the class name c_n by determining c_n from information not contained in $\mathbf{p}$.
3. If $\mathbf{p}$ contains instances that belong to more than one class $c_1 \ldots c_n$ (i.e. impure node) then:

a. Apply an appropriate splitting test. The splitting test finds the best possible split on a single attribute, which leads to one or more mutually exclusive outcomes that partitions all instances contained in **p** into subsets.
    i. If the attribute has discrete values then each nominal value forms a mutually exclusive outcome.
    ii. If the attribute *A* has continuous numeric values then select a threshold point *T* within the continuous range to split at so that the binary tests form each of the mutually exclusive outcomes e.g. $A \leq T$ and $A > T$.
b. Add an attribute node and label it with the attribute name.
    i. Add a branch for each of the mutually exclusive outcomes.
c. Repeat steps 1-3 for each attribute node.

The splitting test plays a critical role in the construction of a decision tree using the latter non-backtracking greedy search algorithm. The algorithm is greedy because the splitting test partitions the training dataset into the largest possible mutually exclusive subsets. In C4.5, Gain Ratio is used as the splitting test criterion in decision tree learning. First, the expected amount of information from the training set *S* is calculated as

$$info(S) = -\sum_{j=1}^{k} \frac{freq(C_j, S)}{|S|} \times \log_2 \left(\frac{freq(C_j, S)}{|S|} \right) bits \quad (8.2)$$

where the instances in *S* are known to belong to a class, C_j. In information theory, Equation 8.2 is referred to as the Entropy formula (Shannon, 1948). The expected amount of information from the mutually exclusive subsets that have been partitioned by test *X* can then be derived from

$$info_x(T) = -\sum_{i=1}^{n} \frac{|T_i|}{|T|} \times info(T_i). \quad (8.3)$$

The total measure of information gain from splitting on test *X* is obtained from

$$gain(X) = info(T) - info_X(T). \quad (8.4)$$

Quinlan (1993) identified that the Gain criterion in Equation 8.4 had a bias towards tests with many outcomes so Quinlan resolved the problem by adjusting the outcomes

via normalisation to create the Gain Ratio criterion. This was achieved by modifying the *info*(*S*) from Equation 8.3 to form

$$split\ info(X) = -\sum_{i=1}^{n} \frac{|T_i|}{|T|} \times \log_2 \frac{|T_i|}{|T|} \tag{8.5}$$

thus producing splitting information, which is then used in the Gain Ratio formula

$$gain\ ratio(X) = \frac{gain(X)}{split\ info(X)}. \tag{8.6}$$

The split with the largest Gain Ratio is chosen as the best split because it possesses the most useful information for assisting classification. However, for occasions when the splitting test produces subsets that have equal class distributions with zero gain, the C4.5 algorithm uses this condition as stopping criteria (Quinlan, 1996). Within Weka, J48 can be configured to perform the node splitting using a binary split or a multi-way split. The binary split is restricted to producing only two branches at each attribute node e.g. a binary split is used at the humidity and windy attribute nodes in Figure 8.1. The multi-way split can use as many branches as distinct values at each attribute node e.g. the root node in Figure 8.1 uses a multi-way split.

A weakness of the divide and conquer algorithm in C4.5, is that it tends to produce decision trees that overfit the training dataset, which causes poor generalisation on the unseen (test) dataset. The over-fitting, also results in larger decision trees, which decreases the comprehensibility of the decision tree classifier. Therefore, pruning of the decision tree is required to overcome over-fitting. Decision tree pruning is described in the subsequent section.

8.3.2 Pruning

There are two main approaches to pruning decision trees, which are post-pruning and pre-pruning. Post-pruning (bottom-up/backward pruning) removes subtrees from the tree after the decision tree has been fully constructed in the growth phase. Each subtree removed is replaced with a leaf node. On the other hand, pre-pruning (top-down/forward pruning) uses an early stopping rule, which prevents subtree growth at nodes that do not meet a predefined condition during the construction of the decision tree in the growth phase. The purpose of pruning is to reduce the complexity of the decision tree and to improve generalisation by reducing over-fitting. A comparison of pruning techniques can be found in (Quinlan, 1999).

C4.5 adopts a post-pruning approach by using pessimistic pruning (Quinlan, 1993), which is a heuristic technique derived from statistical reasoning (Witten et al., 2011). Once the decision tree has been fully grown with the training set, a pessimistic estimate of error rate is applied using the same training set to make the decision on whether to prune a node or not. Firstly, the resubstitution error rate, *f* is calculated as

$$f = E/N \tag{8.7}$$

where *N* is the number of training instances that reached the node and *E* is the number of training instances that were incorrectly classified at the node. Next, the confidence limits *z* are determined from the predefined desired confidence *c* as

$$\Pr\left[\frac{f-q}{\sqrt{q(1-q)/N}} > z\right] = c \tag{8.8}$$

where *N* is the number of training instances, *f* is the resubstitution error rate (Equation 8.7) and *q* is the true error rate. The pessimistic error rate estimate *e* at the node is calculated as

$$e = \frac{f + \frac{z^2}{2N} + z\sqrt{\frac{f}{N} - \frac{f^2}{N} + \frac{z^2}{4N^2}}}{1 + \frac{z^2}{N}} \tag{8.9}$$

where *N* is the number of training instances, *f* is the resubstitution error rate (Equation 8.7) and *z* is the number of standard deviations related to the confidence *c*. C4.5 defaults to *c* = 25% so *z* = 0.69. Furthermore, Quinlan (1993:85) argues that 'the default value seems to work reasonably well for many tasks' but without any legitimate justification. Within Weka's J48 algorithm, *c* is referred to as the Confidence Factor (CF). Smaller *c* values inflict more severe pruning of the decision tree than larger values.

In C4.5, the Minimum Number of Cases (MNC) has a default value of 2 and acts as an additional pruning parameter by controlling the MNC at each node within the decision tree so that near-trivial splits are avoided (Quinlan, 1993). Hence, splitting does not occur at decision tree nodes with an unreasonable number of cases i.e. a node whose number of cases equal the MNC value. Increasing the MNC value tends to produce decision trees with fewer nodes whereas decreasing the MNC value tends to produce decision trees with more nodes. Within Weka's J48 algorithm the MNC pruning parameter is known as the Minimum Number of Objects (MNO). Pruning applies Occam's Razor principle, which stipulates that 'entities should not be

multiplied beyond necessity' thus giving precedence to simplicity. Smaller decision trees are desired because they are more easily interpreted and produce shorter decision rules that enable decisions to be made more rapidly.

Once the decision tree has been pruned, the decision tree can be evaluated by computing the percentage of correctly and incorrectly classified instances by the tree. The estimate of accuracy, known as classification accuracy (CA) is calculated as the percentage of correct classifications (Kohavi, 1995). To evaluate a decision tree, a test set containing unseen data is used to measure the generalisation performance of the trained decision tree. *K*-fold cross-validation (Stone, 1974) (Section 4.5.5) with a two-way data split is commonly used to train and test decision trees. The CA can be averaged from the *k* folds. C4.5 supports *k*-fold cross-validation and is included as a configuration option in Weka's J48 algorithm.

8.4 Decision Tree Optimisation Methodology

So far, the experiments on human comprehension detection in this Thesis have all optimised a neural network as the Comprehension Classifier (Figure 5.3) in FATHOM. Although the optimised ANNs were able to successfully detect human comprehension from datasets extracted from two different studies, the ANNs work as "black boxes" so the knowledge on the relationships captured between the inputs and outputs are hidden. On the other hand, decision trees are classifiers, which work using a "white box" approach so the knowledge acquired in the relationships between the inputs and outputs are visible and comprehensible. In a decision tree, the knowledge is structured in a treelike form and can easily be expressed as decision rules. Replacing FATHOM's Comprehension Classifier ANN with a decision tree could explain how the channels are used to classify human comprehension and non-comprehension. Therefore, this Section is going to describe a methodology for training and validating a decision tree on human comprehension detection to become FATHOM's Comprehension Classifier.

FATHOM's Comprehension Classifier Decision Tree is trained and validated in a series of optimisation experiments with the channel dataset that was extracted and labelled in the previous study (Section 7.13). The decision trees were all constructed using the J48 open source Java implementation of the C4.5 algorithm in the Weka (Hall et al., 2009) data mining software tool. Within each experiment, the aim, methodology and results are reported. Detailed descriptions of the configuration parameters used to

train and validate FATHOM's Comprehension Classifier Decision Tree are also included. There are three purposes for this set of experiments: (1) to identify whether a decision tree could detect human comprehension and non-comprehension from the participants multiple channels of nonverbal behaviour when engaging in the Q&A on the termites video; (2) to identify whether a decision tree outperforms an ANN as the FATHOM Comprehension Classifier; and (3) to see if a decision tree can produce a simple, comprehensible classifier model that provides knowledge on how the channels are used to deduce comprehension and non-comprehension classifications. To discover the optimal Comprehension Classifier Decision Tree, FATHOM's Comprehension Classifier Decision Tree was optimised in a series of experiments, which were executed in the following order:

(1) Node Splitting (Section 8.5)
(2) Pruning (Section 8.6)
(3) Attributes (Section 8.7)

8.5 Node Splitting

Aim

The aim of this experiment was to determine whether a binary split or a multi-way split should be used as the node splitting method in the decision tree.

Methodology

The properties used in this decision tree training experiment are listed in Table 8.1. There were 41 attributes within the dataset because each input vector contained 40 normalised values and there were only two classes: +1 (comprehension) and -1 (non-comprehension). A two-way data split (Section 4.5.5) was used to partition the dataset for training and testing the Comprehension Classifier Decision Tree. To improve upon the two-way data split, 10-fold cross-validation was also implemented. Both the binary split and the multi-way split were tested as the node splitting method (Section 8.3.1). The default MNO value was used. A range of CF values were used for comparing the binary and multi-way split evaluations. To avoid over-training the CF value was not increased to values above 0.5. In total, there were 18 cross-validation experiments, 9 per node splitting method. The average tree size, the average number of leaves and the average testing CAs for each cross-validation experiment were output so that the decision tree performances could be evaluated in the post induction analysis.

Table 8.1 Decision Tree Training Configuration: Node Splitting Verification

Parameter	Value
k-fold Cross-validation	10-fold cross-validation
Training Set	90%
Testing Set	10%
Split Type	Binary or Multi-way
CF	0.005, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4 and 0.5
MNO	2

Results and Discussion

Table 8.2 shows the average tree size, the average number of leaves and the average testing CAs for each cross-validation experiment. When comparing the binary split results against the multi-way split cross-validation results there is no change in the size of the decision trees and no difference between any of the averaged CAs. Therefore, there is no benefit in using the binary splitting method so the subsequent experiments shall use the multi-way splitting method.

Table 8.2 Binary Splits versus Multi-way Splits

Split Type	CF	Leaves	Tree Size	Total CA	Normalised Total CA	TP CA	TN CA
Binary	0.005	1297	2593	92.950	92.671	91.021	94.320
	0.05	1488	2975	93.869	93.645	92.319	94.970
	0.1	1524	3047	94.089	93.874	92.602	95.146
	0.15	1561	3121	94.107	93.896	92.655	95.138
	0.2	1585	3169	94.241	94.041	92.856	95.226
	0.25	1659	3317	94.381	94.191	93.068	95.314
	0.3	1693	3385	94.450	94.262	93.157	95.368
	0.4	1709	3417	94.459	94.273	93.174	95.372
	0.5	1710	3419	94.469	94.284	93.192	95.377
Multi-way	0.005	1297	2593	92.950	92.671	91.021	94.320
	0.05	1488	2975	93.869	93.645	92.319	94.970
	0.1	1524	3047	94.089	93.874	92.602	95.146
	0.15	1561	3121	94.107	93.896	92.655	95.138
	0.2	1585	3169	94.241	94.041	92.856	95.226
	0.25	1659	3317	94.381	94.191	93.068	95.314
	0.3	1693	3385	94.450	94.262	93.157	95.368
	0.4	1709	3417	94.459	94.273	93.174	95.372
	0.5	1710	3419	94.469	94.284	93.192	95.377

8.6 Pruning

Aim

The aim of this experiment was to find the optimal CF and MNO parameter values.

Methodology

The properties used in this decision tree training experiment are listed in Table 8.3. The CF and the MNO were optimised together because they both affect the pruning of a decision tree. A range of CF values were used but to avoid over-training the CF value was not increased above 0.5. The default CF value (0.25) was also used because Quinlan (1993) found the value to work well for many problems. A set of small and

larger MNO values were used including the default value (2). There were 81 cross-validation experiments in total. The average tree size, the average number of leaves and the average testing CAs for each cross-validation experiment were output so that the decision tree performances could be evaluated in the post induction analysis.

Table 8.3 Decision Tree Training Configuration: Pruning Verification

Parameter	Value
k-fold Cross-validation	10-fold cross-validation
Training Set	90%
Testing Set	10%
Split Type	Multi-way
CF	0.005, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4 and 0.5
MNO	2, 3, 4, 5, 10, 15, 20, 25 and 30

Results and Discussion

The results from the 81 cross-validation experiments are in Appendix Q. The decision trees with the highest average Total CAs highlighted (green rows) in Appendix Q for each CF value tested has been collated together in to Table 8.4. All of the decision trees achieved the highest CAs when the MNO was set at 2. The average Total CA begins to plateau at the 0.3 CF. The decision tree with the optimum pruning values is highlighted in Table 8.4 Pruning Results (green row) with a 0.25 CF and the MNO set at 2. Although the optimum decision tree was able to distinguish between the two classes well, the tree size is large with 3317 nodes and 1659 leaves, making it less intelligible. Furthermore, when compared to the optimum neural network with a 41:60:1 topology, 2520 weights and an averaged 91.349% Normalised Total CA for the same experiment in Table 7.18, the decision tree is considerably larger and more complex. The optimal decision tree result from this experiment is used as the optimal baseline for comparative purposes in the subsequent attribute optimisation experiments.

Table 8.4 Pruning Results

MNO	CF	Leaves	Tree Size	Total CA	Normalised Total CA	TP CA	TN CA
2	0.005	1297	2593	92.950	92.671	91.021	94.320
	0.05	1488	2975	93.869	93.645	92.319	94.970
	0.1	1524	3047	94.089	93.874	92.602	95.146
	0.15	1561	3121	94.107	93.896	92.655	95.138
	0.2	1585	3169	94.241	94.041	92.856	95.226
	0.25	1659	3317	94.381	94.191	93.068	95.314
	0.3	1693	3385	94.450	94.262	93.157	95.368
	0.4	1709	3417	94.459	94.273	93.174	95.372
	0.5	1710	3419	94.469	94.284	93.192	95.377

8.7 Attributes

The primary aim of this set of experiments was to determine the optimum number attributes for the FATHOM Comprehension Classifier Decision Tree. To discover the optimal number of attributes, FATHOM's Comprehension Classifier Decision Tree was trained and validated using two different experimental approaches, which were executed in the following order:

(1) Attribute Information Gain (Section 8.7.1)

(2) Grouping Attributes by Theme (Section 8.7.2)

The properties used in all of the decision tree training experiments in Section 8.7.1 and Section 8.7.2 are listed in Table 8.5. Only the number of attributes used by the decision trees fluctuated across the experiments, all other properties remained the same from the previous optimisation experiments. Section 8.7.3 provides a summary of the findings from the experiments in Section 8.7.1 and 8.7.2.

Table 8.5 Decision Tree Training Configuration: Attributes Verification

Parameter	Value
k-fold Cross-validation	10-fold cross-validation
Training Set	90%
Testing Set	10%
Split Type	Multi-way
CF	0.25
MNO	2

8.7.1 Attribute Information Gain

This experiment takes the same approach as the information gain experiment in Section 7.13.3 but substitutes the neural network with a decision tree and refers to the inputs as attributes. The attributes were divided into two groups using the information gain results in Table 7.20. The first group named "top" contained all of the attributes that were ranked from 1 to 21 in Table 7.20, which resulted in a total of 21 inputs. The second group named "bottom" contained all of the attributes that were ranked from 22 to 41 in Table 7.20, which resulted in a total of 20 inputs. The top group contained the top half of the attributes with the highest levels of information gain in comparison to the bottom group. Therefore, it was anticipated that when each group was used to train and validate decision trees that the top group would perform better than the bottom group. In total, there were two cross-validation experiments. Table 8.5 lists the properties used in both experiments. The average tree size, the average number of

leaves and the average testing CAs for each cross-validation experiment were output so that the decision tree performances could be evaluated in the post induction analysis. Table 8.6 shows the results from the two cross-validation experiments. As predicted, the top group performed better than the bottom group, yielding higher CAs and a smaller decision tree. However, the top groups CAs are lower and the decision tree size is larger than the baseline decision trees in Table 8.4. The bottom group's decision tree is larger and more complex because it has to compensate for the deficit in information gain. The next experiment is going to investigate the decision tree performances with some other attribute groups.

Table 8.6 Information Gain Results

Group	Leaves	Tree Size	Total CA	Normalised Total CA	TP CA	TN CA
Top	1731	3461	93.861	93.661	92.478	94.844
Bottom	2756	5511	88.468	88.126	86.107	90.145

8.7.2 Grouping Attributes by Theme

This experiment takes the same approach as the grouping of inputs into themes experiment in Section 7.13.4 but substitutes the neural networks with a decision trees and refers to the inputs as attributes. The attributes were categorised into 7 themed groups (demographics, eyes, left eye, right eye, face, face angle and face movement) as outlined in Table 7.22. Table 8.5 lists the properties used in this set of experiments. The average tree size, the average number of leaves and the average testing CAs for each cross-validation experiment were output so that the decision tree performances could be evaluated in the post induction analysis. The results from this set of experiments are shown in Table 8.7. None of the cross-validation experiments performed as well or better than the baseline decision tree (Table 8.4). Group 2, which focused upon the left and right eye attributes, yielded the highest CAs. This finding is in line with the results from the Group 2 neural network cross-validation experiment in Section 7.13.4. Therefore, the latter result supports the general finding from previous eye tracking studies (Section 3.5) on human comprehension detection, that eye behaviour can provide a window on human comprehension. Although Group 2 achieved the highest CAs, the decision tree had the largest number of leaves and nodes in comparison to the other groups in Table 8.7. Group 1 and Group 6 both yielded Normalised Total CAs, which were close to chance levels due to having difficulty distinguishing TP's. The findings from this set of experiments support the

hypothesis that knowledge on human comprehension detection within the dataset resides across all of the attributes rather than individual attributes i.e. a holistic approach.

Table 8.7 Theme Results

Group	Theme	Leaves	Tree Size	Total CA	Normalised Total CA	TP CA	TN CA
1	Demographics	8	15	58.498	50.726	4.802	96.651
2	Eyes	3057	6113	86.601	86.148	83.476	88.821
3	Left Eye	2251	4501	73.155	71.424	61.194	81.653
4	Right Eye	2322	4643	72.368	70.597	60.132	81.062
5	Face	2954	5907	84.905	84.459	81.824	87.094
6	Face Angle	389	777	61.718	56.349	24.624	88.075
7	Face Movement	3571	7141	79.249	78.456	73.771	83.141

8.7.3 Summary

Two different approaches have been taken to try and determine the optimum number of attributes for the FATHOM Comprehension Classifier Decision Tree. In the first experiment (Section 8.7.1), the top group did not perform as well as the baseline decision tree (Table 8.4) but it still yielded higher CAs than any of the themed attribute groups experiments (Section 8.7.2). In the second experiment (Section 8.7.2), the decision trees that focused on both eyes (Group 2) yielded the highest CAs out of the all the themes assessed. Throughout the attribute experiments the findings suggest that knowledge on human comprehension detection within the dataset resides across multiple attributes rather than individual attributes. The tree sizes in this set of experiments were predominantly large thus making them less comprehensible. To conclude, the optimal number of attributes was 41 from the baseline decision tree in the pruning experiment (Section 8.6).

8.8 Summary

This chapter began by introducing decision trees, their main components and how decision rules are easily traced from the root node to a leaf node. The C4.5 algorithm, a popular supervised data mining algorithm for growing and pruning decision trees was outlined. Key parameters and their roles within the C4.5 algorithm were also discussed to ensure that appropriate training configurations and parameter values were selected to optimise the decision trees in future experiments. Unlike ANNs, decision trees take advantage of the “white box” approach by illustrating knowledge discovered from a dataset as a set of decision rules in a top-down treelike structure, which makes them

easy to comprehend and allows relationships between the inputs and outputs to be visualised for analysis. Therefore, this Chapter attempted to replace FATHOM Comprehension Classifier ANN with a decision tree. A methodology for optimising FATHOM's Comprehension Classifier Decision Tree using the C4.5 algorithm with the channel dataset extracted from the learning environment study in Section 7.13 was presented. Throughout the optimisation experiments (Section 8.5 - 8.7), FATHOM's Comprehension Classifier Decision Tree was able to detect non-comprehension more easily than comprehension, which was also reported when optimising ANNs with the same dataset (Section 7.14) and may be due to the non-comprehension class being larger than the comprehension class. Optimisation of the FATHOM Comprehension Classifier Decision Tree's attributes revealed that the human comprehension detection knowledge resided across multiple attributes and not individual attributes, which was a finding that was reported when optimising the ANN's inputs with the same dataset (Section 7.13). Far fewer parameters were required for optimising a decision tree than a neural network. The optimisation experiments yielded an optimal FATHOM Comprehension Classifier Decision Tree with 3317 nodes, 1659 leaves capable of reliably detecting human comprehension from 41 attributes with a 94.381% average Total CA. Although the optimised decision tree outperformed the optimised ANN (topology 41:60:1 and a 91.835% average Total CA), the decision tree has a much larger topology so the ANN is preferred as FATHOM's Comprehension Classifier. However, the optimisation experiments in this Chapter have proven that decision trees can be used to detect patterns of human comprehension and non-comprehension from multiple channels of nonverbal behaviour.

Chapter 9 Conclusion and Future Directions

9.1 Introduction

This Thesis has investigated whether an adaptable system can detect human comprehension and non-comprehension from monitoring multiple channels of nonverbal behaviour using artificial neural networks and decision trees. In this Chapter, summaries of the main findings and the contributions of this research are provided in relation to the Thesis aim and objectives listed in Section 1.3. Suggestions on the potential directions of future research are also presented. Lastly, the Chapter concludes with some final remarks.

9.2 Thesis Summary

In Chapter 1, the aim of this Thesis was to develop an adaptable system that can detect human comprehension and non-comprehension from monitoring multiple nonverbal behaviours using artificial neural networks. This section shall summarise how each of the preceding Chapters have contributed towards fulfilling the aim and objectives stated in Section 1.3.

Chapter 2 introduced the concept of nonverbal behaviour and reviewed how discrete channels of nonverbal behaviour can be measured using different extraction techniques. The distinct properties of nonverbal behaviour were identified from a comparison with verbal behaviour. The review highlighted the abundance and continuous availability of nonverbal behaviour for analysis. In Chapter 3, human comprehension was defined and the main tools and techniques used to detect human comprehension in informed consent and learning environments were discussed. Existing approaches on human comprehension detection using nonverbal behaviour were also reviewed in Chapter 3, which identified specific nonverbal behaviours known to be associated with human comprehension or non-comprehension. Chapter 4 began by introducing the architecture of artificial neural networks and how they are inspired by biological artificial neural networks. The limitations of an artificial neuron and the capability of multilayer ANN at solving nonlinear problems were discussed. The error-backpropagation learning algorithm, a well known learning algorithm for training artificial neural networks was described. Common variations in the application of the error-backpropagation learning algorithm were reviewed. Examination of previous

applications using backpropagation neural networks revealed the uncharted opportunity of adapting Silent Talker's patented methodology for use within the human comprehension detection domain. Collectively, Chapters 2 to 4 formed the literature review. The literature review provided motivation for this research because it exposed a niche for a novel, automated adaptable system for human comprehension detection using artificial neural networks and multichannels of nonverbal behaviour.

Chapter 5 presented FATHOM, a novel automated adaptable system that was built to detect human comprehension from multiple channels of nonverbal behaviour using a bank of artificial neural networks. The Chapter included descriptions of how FATHOM was constructed and how it works. The forty nonverbal channels monitored by FATHOM were defined and categorised. The neural network architecture within FATHOM is composed of three types of ANNs: Object Locators, Pattern Detectors and a Comprehension Classifier. The Comprehension Classifier ANN is a key, modular component of FATHOMs architecture, which was later optimised and validated on human comprehension detection with datasets extracted from experimental studies contained in Chapters 6 and 7.

Chapter 6 began by introducing the design of a two-phase field study methodology on a mock informed consent process for a sexual and reproductive health clinical trial, which was executed in Mwanza by FHI 360 and NIMR. Both phases of the field study were designed to capture 80 female participants in videos for optimising and evaluating FATHOM's Comprehension Classifier ANN. Unfortunately, due to the poor quality of the video recordings in both phases of the study, FATHOM was not evaluated with the videos from the exploratory testing phase and 50 videos were discarded from the developmental phase. Descriptions of how FATHOM's Comprehension Classifier ANN was trained and validated with the channel datasets extracted from the developmental phase videos were provided in each experiment. Throughout the developmental phase experiments, the error-backpropagation learning algorithm and cross-validation with a three-way data split was used to optimise the Comprehension Classifier ANN. The purpose of the developmental phase experiments was to determine whether FATHOM's Comprehension Classifier ANN could detect patterns of human comprehension and non-comprehension from the multiple channels of nonverbal behaviour. Results from the developmental phase cross-validation experiments repeatedly revealed testing Total CAs consistently above

84%, which demonstrates that FATHOM's Comprehension Classifier with a 40:30:1 topology was able to find distinct patterns of human comprehension and non-comprehension from the multiple channels of nonverbal behaviour. The findings presented in Chapter 6 are limited to human comprehension detection with African women aged 18-35 years old.

Chapter 7 started by introducing the design of a comprehension study that was executed at MMU in a learning environment. During the study, each participant watched a factual video on termites and then had their comprehension of the termites video assessed in a video recorded Q&A session. The videos from the Q&A session were later used to train and validate FATHOM's Comprehension Classifier ANN on human comprehension detection. None of the videos were discarded. The purpose of the study was to identify whether FATHOM's Comprehension Classifier ANN could detect patterns of human comprehension and non-comprehension from the male and female participants multiple channels of nonverbal behaviour. A methodology was presented for optimising FATHOM's Comprehension Classifier ANN in a series of optimisation experiments using the error-backpropagation learning algorithm and cross-validation with a two-way data split. For each optimisation experiment, the aim, methodology results and discussion were reported. A description of how the channel dataset was extracted from the videos was included. Four new channels were added to the channel dataset. Throughout the optimisation experiments, FATHOM's Comprehension Classifier ANN was able to detect non-comprehension more easily than comprehension. Furthermore, the results also revealed that the human comprehension detection knowledge resided across multiple channels and not individual channels. At the end of the optimisation process, the FATHOM Comprehension Classifier ANN had a 41:60:1 topology and was capable of reliably detecting human comprehension from the male and female nonverbal channels with an average testing Total CA of 91.835%.

Lastly, Chapter 8 introduced decision trees and described C4.5, a popular decision tree induction algorithm. Decision trees illustrate knowledge as a set of decision rules in a top-down graph, which makes them easy to understand. In contrast to artificial neural networks, decision trees use a "white box" approach, so the structure of the tree provides insight on how the knowledge was acquired between the root node and the classification at the leaf node. Therefore, the purpose in Chapter 8 was to identify

whether a decision tree was capable of detecting human comprehension to become an alternative Comprehension Classifier for FATHOM and to determine whether the knowledge acquired in the model was comprehensible. A methodology was presented for optimising FATHOM's Comprehension Classifier Decision Tree in a series of optimisation experiments using the J48 implementation of the C4.5 algorithm and cross-validation with a two-way data split. The channel dataset from Chapter 7 was used train and validate FATHOM's Comprehension Classifier decision tree in human comprehension detection. In each optimisation experiment, the aim, methodology results and discussion were reported. At the end of the optimisation process, the FATHOM Comprehension Classifier Decision Tree had 3317 nodes, 1659 leaves and was capable of reliably detecting human comprehension from multiple channels of nonverbal behaviour with an average testing Total CA of 94.381%. Although the optimised decision tree outperformed the optimised ANN (topology 41:60:1 and a 91.835% average Total CA) in Chapter 7 with the same dataset, the decision tree had a much larger topology, which was less comprehensible so the ANN is preferred as FATHOM's Comprehension Classifier.

9.3 Contributions Summary

The main contributions of the research presented in this Thesis are:

- The creation of a novel adaptable system for detecting human comprehension and non-comprehension through the monitoring of multiple nonverbal behaviours using artificial neural networks.
- The generic, modular architecture of the adaptable system means that other researchers can easily adopt and/or expand the architecture and follow the methodology in future research on human comprehension detection in different environments.
- Experiments conducted on the dataset extracted from the informed consent field study resulted in the following findings:
    - Evidence that human comprehension and non-comprehension patterns reside within multichannels of nonverbal behaviour.
    - Proved that an artificial neural network can be used to classify human comprehension and non-comprehension from multichannels of nonverbal behaviour.

- Experiments conducted on the dataset extracted from the learning environment study resulted in the following findings:
    - Evidence that human comprehension and non-comprehension patterns reside within male and female multichannels of nonverbal behaviour.
    - Evidence that human comprehension and non-comprehension patterns exist in another culture.
    - Proved that human comprehension and non-comprehension can be detected from nonverbal behaviours in a different domain.
- Further experiments conducted on the dataset extracted from the learning environment study resulted in the following finding:
    - Proved that a decision tree can be used to classify human comprehension and non-comprehension from multichannels of nonverbal behaviour.
- General findings from experiments with the informed consent and learning environment datasets were:
    - Provides evidence supporting previous research that facial nonverbal behaviours do emit human comprehension and non-comprehension patterns.
    - Human comprehension and non-comprehension can be detected much sooner from nonverbal behaviours than traditional comprehension assessment measures.

9.4 Future Directions

The research in this Thesis has shown that it is possible to detect human comprehension from nonverbal behaviours using FATHOM and has overcome the limitations of existing approaches to human comprehension detection. However, there are areas within this research that could be enhanced or expanded upon. The time for further experimentation was limited and so application of other prospective approaches still remains unexplored. Therefore, this section shall briefly provide some suggestions that could be investigated in further research.

9.4.1 Channels

At present, FATHOM's channels (Section 5.4) predominantly focus upon facial nonverbal behaviours. New nonverbal channels could easily be added to FATHOM so

that further nonverbal channel analyses could be performed. To add a new nonverbal channel, an Object Locator ANN (Figure 5.4) would need to be created using FATHOM's Neural Network Trainer (Figure 5.6). If the new nonverbal channel had multiple states then a Pattern Detector ANN (Figure 5.5) would also need to be created for each possible state using FATHOM's Neural Network Trainer. Image-based datasets would need to be collated to train and validate the latter ANNs.

In Section 3.5.1, hand gestures have been found to provide knowledge of a child's understanding during their verbal explanations of a concept. Furthermore, Goldin-Meadow and Alibali (2013:257) argue that 'gesture reflects speakers' thoughts, often their unspoken thoughts, and thus can serve as a window onto cognition'. Therefore, incorporation of channels focusing upon nonverbal hand behaviour in FATHOM could potentially enhance human comprehension detection. Other lower body nonverbal behaviours could also be added as new channels in FATHOM. The BAP coding system (Dael et al., 2012) covers the manual coding of body movement spanning from articulation of the head, neck, torso, arms and the lower limbs. Therefore, BAP is an ideal starting point for coding lower body cues as channels in FATHOM.

9.4.2 Neural Network Learning Algorithm Convergence Speed

Although the error-backpropagation algorithm is capable of performing complex pattern detection, it is notoriously slow to converge. This is particularly more so, with large multilayer neural networks. There a multiple reasons as to why the error-backpropagation algorithm may be slow. Firstly, the error-backpropagation algorithm suffers from the limitations and weaknesses of the gradient descent method i.e. not guaranteed to converge, can get trapped in a local minima or oscillate (see Figure 4.5). Secondly, there are a lot of parameters within the neural network that have to be tuned repeatedly by the error-backpropagation algorithm. Thirdly, there are number of parameters within the error-backpropagation algorithm that require tuning. Lastly, larger datasets often require more processing time than smaller datasets. As a result, a large amount of time and computational resources are absorbed when running the error-backpropagation algorithm on datasets.

Researchers have suggested tips and tricks to help improve the convergence speed of the error-backpropagation algorithm e.g. LeCun et al. (1998). However, there are other neural network learning algorithms that could be adopted instead such as the

Extreme Learning Machine (ELM) (Huang et al., 2004) algorithm or the No-Propagation (No-Prop) (Widrow et al., 2013) algorithm. ELM (Huang et al., 2004) is a learning algorithm designed for single hidden layer feedforward neural networks, which randomly initialises the hidden layer and only adjusts the weights between the hidden layer and output layer neurons. When compared against classic feedforward learning algorithms with artificial and real world datasets, ELM has been found to learn extremely fast and reach small training errors (Huang et al., 2004; Huang et al., 2006). Huang et al. (2015) provide a recent review of trends in ELM. No-Prop (Widrow et al., 2013) is a learning algorithm designed for multilayer feedforward neural networks where the hidden layer is randomly initialised and only the weights between the hidden layer and output layer neurons are adjusted using steepest descent. Future research with ANNs in FATHOM could try benefiting from using the ELM or No-Prop algorithm to speed up neural network convergence, save time and free up computational resources sooner.

To avoid large datasets in order to reduce convergence speed, a Thin Slice (Ambady and Rosenthal, 1992) analysis approach could be used, where only short observations of human behaviour are judged. In a meta-analysis, Ambady and Rosenthal (1992) revealed that judgements under 30 seconds of observation were as accurate as those made from 5 minute observations and that the accuracy did not fluctuate significantly between the 30 second, 1, 2, 3, 4 and 5 minute long observations. Therefore, it would be interesting to investigate whether human comprehension patterns reside in different sized thin slices of observed nonverbal behaviours with FATHOM by experimenting with the timeslot (Section 5.2) type and length.

9.4.3 Generalisation

This research presented in this Thesis has shown that it is possible to detect human comprehension from multichannels of nonverbal behaviour using ANNs with datasets from two different ethnicities. There was a predominant focus upon the same multichannels of nonverbal behaviour within both datasets. Therefore, this raises the following research question, which could be pursued in further research: is there a general set of nonverbal behaviours that FATHOM can use to detect human comprehension and non-comprehension across all cultures, sex and ages?

9.4.4 Track Multiple People

At present, FATHOM is only able to track and detect human comprehension from one individual at any one point in time. There may be environments where tracking the human comprehension of multiple people in near real-time may be useful e.g. the classroom. Therefore, an interesting next step would be to scale up FATHOM so that it can detect comprehension from simultaneously monitoring the nonverbal behaviours of a group of people. If FATHOM was able to detect human comprehension from groups of people then it would provide educators with an alternative, automated comprehension assessment tool. However, the implementation of this potential future investigation would present the problem of how to maintain near real-time processing time.

9.5 Conclusion

This Thesis has presented FATHOM a novel adaptable system that is capable of detecting human comprehension and non-comprehension by monitoring multiple nonverbal behaviours using ANNs. Although this research was primarily focused upon human comprehension classification using ANNs, it has also identified that decisions trees inducted using the J48 algorithm can be used to classify human comprehension. This research has taken a novel approach to human comprehension detection by using: (1) a nonverbal multichannel approach; (2) automatic extraction of nonverbal behaviour; and (3) using an ANN or decision tree as the comprehension classifier. FATHOM is only a prototype and the empirical findings are limited to human comprehension detection in informed consent and learning environments. However, there are plenty of directions for future research stemming from this work as highlighted in Section 9.4. Therefore, it is hoped that this research shall encourage other researchers to investigate automated human comprehension detection in other domains using nonverbal behaviour and machine learning algorithms such as ANNs and decision trees.

References

Adolph, J. (1998) *Neural network approaches to face analysis*. M.Sc. Manchester Metropolitan University.

Afolabi, M. O., Okebe, J. U., McGrath, N., Larson, H. J., Bojang, K. and Chandramohan, D. (2014) 'Informed consent comprehension in African research settings.' *Tropical Medicine and International Health*, 19(6), pp. 625-642.

Ahmad, S. and Tesauro, G. (1989) 'Scaling and generalization in neural networks: a case study.' *In* Touretzky, D. S. (ed.) *Advances in Neural Information Processing Systems 1*, San Francisco: Morgan Kaufmann Publishers Inc., pp. 160-168.

Allen, V. L. and Atkinson, M. L. (1978) 'Encoding of nonverbal behavior by high-achieving and low-achieving children.' *Journal of Educational Psychology*, 70(3), pp. 298-305.

Ambady, N. and Rosenthal, R. (1992) 'Thin slices of expressive behavior as predictors of interpersonal consequences: a meta-analysis.' *Psychological Bulletin*, 111(2), pp. 256-274.

App, B., McIntosh, D. N., Reed, C. L., and Hertenstein, M. J. (2011) 'Nonverbal channel use in communication of emotion: how may depend on why.' *Emotion*, 11(3), pp. 603-617.

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

Babad, E. (2009) 'Teaching and nonverbal behavior in the classroom.' *In* Saha, L. J. and Dworkin, A. G. (eds.), *International Handbook of Research on Teachers and Teaching*, Springer US, pp. 817-827.

Bandar, Z., McLean, D. A., O'Shea, J. D., and Rothwell, J. A. (2002) *Analysis of the Behaviour of a Subject*. WO02087443.

Bartlett, M. S., Hager, J. C., Ekman, P. and Sejnowski, T. J. (1999) 'Measuring facial expressions by computer image analysis.' *Psychophysiology*, 36(2), pp. 253-263.

Battiti, R. (1989) 'Accelerated backpropagation learning: two optimization methods.' *Complex systems*, 3(4), pp. 331-342.

Baxter, E. (2015) *Complete Crime Scene Investigation Handbook*. New York: CRC Press.

Beauchamp, T. L. and Childress, J. F. (2001) *Principles of Biomedical Ethics*. 5th ed., New York: Oxford University Press.

Bebis, G. and Georgiopoulos, M. (1994) 'Feed-forward neural networks: why network size is so important.' *IEEE Potentials*, pp. 27-31.

Berger, O., Grønberg, B. H., Sand, K., Kaasa, S. and Loge, J. H. (2009) 'The length of consent documents in oncological trials is doubled in twenty years.' *Annals of Oncology*, 20(2), pp. 379-385.

Birdwhistell, R. L. (1970) *Kinesics and Context: Essays on Body Motion Communication*. Philadelphia: University of Pennsylvania Press.

Buckingham, F. J., Crockett, K. A., Bandar, Z. A., O'Shea, J. D., MacQueen, K. M. and Chen, M. (2012) 'Measuring human comprehension from nonverbal behaviour using artificial neural networks.' *In 2012 International Joint Conference on Neural Networks (IJCNN)*. Brisbane, Queensland, 10th-15th June 2012, IEEE, pp. 1-8.

Buckingham, F. J., Crockett, K. A., Bandar, Z. A. and O'Shea, J. D. (2014) 'FATHOM: A neural network-based non-verbal human comprehension detection system for learning environments.' *In 2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*. Orlando, Florida, 9th-12th December 2014, IEEE, pp. 403-409.

Butler, I. F., Pandrea, I., Marx, P. A. and Apetrei, C. (2007) 'HIV genetic diversity: biological and public health consequences.' *Current HIV Research*, 5(1), pp. 23-45.

Castellano, G., Fanelli, A. M. and Pelillo, M. (1997) 'An iterative pruning algorithm for feedforward neural networks.' *IEEE Transactions on Neural Networks*, 8(3), pp. 519-531.

Church, R. B. and Goldin-Meadow, S. (1986) 'The mismatch between gesture and speech as an index of transitional knowledge.' *Cognition*, 23(1), pp. 43-71.

Cohn, J. F. and Ekman, P. (2008) 'Measuring facial action.' *In* Harrigan, J. A., Rosenthal, R. and Scherer, K. R. (eds.) *New Handbook of Methods in Nonverbal Behavior Research*, Oxford: Oxford University Press, pp. 9-64.

Cootes, T. F., Edwards, G. J. and Taylor, C. J. (2001) 'Active appearance models.', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6), pp. 681-685.

Cootes, T. F., Edwards, G. J. and Taylor, C. J. (1998) 'Active appearance models.' *Lecture Notes in Computer Science*, 1407, Berlin: Springer-Verlag, pp. 484-498.

Cottrell, G. W. and Metcalfe, J. (1990) 'EMPATH: face, emotion, and gender recognition using holons.' *In* Lippmann, R. P., Moody, J. E. and Touretzky, D. S. (eds.) *Advances in Neural Information Processing Systems 3*, San Francisco: Morgan Kaufmann Publishers Inc., pp. 564-571.

Cramer, C. (1998) 'Neural networks for image and video compression: a review.' *European Journal of Operational Research*, 108 (2), pp. 266-282.

Crockett, K., O'Shea, J., Buckingham, F., Bandar, Z., MacQueen, K., Chen, M. and Simpson, K. (2013) 'FATHOMing out interdisciplinary research transfer.' *In 2013 IEEE Symposium on Computational Intelligence for Engineering Solutions (CIES)*. Singapore, 16th-19th April 2013, IEEE, pp. 1-8.

Cybenko, G. (1988) *Continuous valued neural networks with two hidden layers are sufficient.* Massachusetts: Tufts University.

Cybenko, G. (1989) 'Approximation by superpositions of a sigmoidal function.' *Mathematics of Control, Signals and Systems*, 2(4), pp. 303-314.

Darwin, C. (1872) *The Expression of the Emotions in Man and Animals*, London: John Murray.

Dael, D., Mortillaro, M. and Scherer, K. R. (2012) 'The body action and posture coding system (BAP): development and reliability.' *Journal of Nonverbal Behavior*, 36(2), pp. 97-121.

De Villiers, J. and Barnard, E. (1993) 'Backpropagation neural nets with one and two hidden layers.' *IEEE Transactions on Neural Networks*, 4(1), pp. 136-141.

Duch, W. and Jankowski, N. (1999) 'Survey of neural transfer functions.' *Neural Computing Surveys*, 2(1), pp. 163-212.

Dunlosky, J. and Lipko, A. R. (2007) 'Metacomprehension: a brief history and how to improve its accuracy.' *Current Directions in Psychological Science*, 16(4), pp. 228-232.

Ekman, P. (2006) *Darwin and Facial Expression: A Century of Research in Review*. California: Malor Books.

Ekman, P. (2004) 'Emotional and conversational nonverbal signals.' *In* Larrazabal, J. M. and Miranda, L. A. (eds.) *Language, Knowledge, and Representation*, Springer Netherlands, pp. 39-50.

Ekman, P. and Friesen, V. W. (1978) *The Facial Action Coding System (FACS).* California: Consulting Psychologists Press.

Ekman, P. and Friesen, W. V. (1971) 'Constants across cultures in the face and emotion.' *Journal of Personality and Social Psychology*, 17(2), pp. 124-129.

Ekman, P. and Friesen, W. V. (1969a) 'The repertoire of nonverbal behavior: categories, origins, usage, and coding.' *Semiotica*, 1(1), pp. 49-98.

Ekman, P. and Friesen, W. V. (1969b) 'Nonverbal leakage and clues to deception.' *Psychiatry*, 32(1), pp. 88-106.

Faden, R. R., Beauchamp, T. L. and King, N. M. (1986) *A History and Theory of Informed Consent*. New York: Oxford University Press.

Fahlman, S. E. and Lebiere, C. (1990) 'The cascade-correlation learning architecture.' *In* Touretzky, D. S. (ed.) *Advances in Neural Information Processing Systems 2*, San Francisco: Morgan Kaufmann Publishers Inc., pp. 524-532.

Family Health International 360 (2016) *Home.* [Online] [Accessed on 26th April 2016] http://www.fhi360.org/

Fausett, L. (1994) *Fundamentals of Neural Networks*. New Jersey: Prentice Hall.

Flesch, R. (1948) 'A new readability yardstick.' *Journal of Applied Psychology*, 32(3), pp. 221-233.

Flesch, R. (1949) *The Art of Readable Writing*. New York: Harper.

Flory, J. and Emanuel, E. (2004) 'Interventions to improve research participants' understanding in informed consent for research: a systematic review.' *Journal of the American Medical Association*, 292(13), pp. 1593-1601.

Frauendorfer, D., Mast, M. S., Nguyen, L. and 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), pp. 231-245.

Gall, S. N. (1985) 'Help-seeking behavior in learning.' *Review of Research in Education*, 12, pp. 55-90.

Gall, S. N. and Glor-Scheib, S. (1985) 'Help seeking in elementary classrooms: an observational study.' *Contemporary Educational Psychology*, 10(1), pp. 58-71.

Goldin-Meadow, S. (2004) 'Gesture's role in the learning process.' *Theory into Practice*, 43(4), pp. 314-321.

Goldin-Meadow, S. and Alibali, M. W. (2013) 'Gesture's role in speaking, learning, and creating language.' *Annual Review of Psychology*, 64, pp. 257-283.

Golomb, B. A., Lawrence, D. T. and Sejnowski, T. J. (1990) 'SexNet: a neural network identifies sex from human faces.' *In* Lippmann, R. P., Moody, J. E. and Touretzky, D. S. (eds.) *Advances in Neural Information Processing Systems 3*, San Francisco: Morgan Kaufmann Publishers Inc., pp. 572-577.

Graesser, A. C., Lu, S., Olde, B. A., Cooper-Pye, E. and Whitten, S. (2005) 'Question asking and eye tracking during cognitive disequilibrium: comprehending illustrated texts on devices when the devices break down.' *Memory and Cognition*, 33(7), pp. 1235-1247.

Guerrero, L. and Farinelli, L. (2009) 'The interplay of verbal and nonverbal codes.' *In* Eadie, W. F. (ed.) *21st Century Communication: A Reference Handbook*. California: Sage Publications Inc., pp. 239-249.

Gulabovska, M. and Leeson, P. (2014) 'Why are women better decoders of nonverbal language?' *Gender Issues*, 31(3), pp. 202-218.

Gurney, K. (1997) *An Introduction to Neural Networks*. New York: CRC Press.

Haggard, E. A. and Isaacs, K. S. (1966) 'Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy,' *In* Gottschalk, L. A. and Auerbach, A. H. (eds.) *Methods of Research in Psychotherapy*, New York: Appleton-Century-Crofts, pp. 154-165.

Hall, J. A. (2007) 'Nonverbal cues and communication.' *In* Baumeister, R. F. and Vohs, K. D. (eds.) *Encyclopedia of Social Psychology*, California: SAGE Publications Inc., pp. 626-627.

Hall, J. A. (1978) 'Gender effects in decoding nonverbal cues.' *Psychological Bulletin*, 85(4), pp. 845-857.

Hall, J. A., Bernieri, F. J. and Carney, D. R. (2008) 'Nonverbal behavior and interpersonal sensitivity.' *In* Harrigan, J. A., Rosenthal, R. and Scherer, K. R. (eds.) *New Handbook of Methods in Nonverbal Behavior Research*, Oxford: Oxford University Press, pp. 237-281.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I. H. (2009) 'The WEKA data mining software: an update.' *SIGKDD Explorations*, 11(1), pp. 10-18.

Hammerstrom, D. (1993) 'Neural networks at work.' *IEEE Spectrum*, 30(6), pp. 26-32.

Hassoun, M. H. (1995) *Fundamentals of Artificial Neural Networks*. Massachusetts: MIT Press.

Haykin, S. (1999) *Neural Networks: A Comprehensive Foundation*. 2nd ed., New Jersey: Prentice Hall.

Heaton, J. (2008) *Introduction to Neural Networks for C#*. 2nd ed., Missouri: Heaton Research Inc.

Hereu, P., Pérez, E., Fuentes, I., Vidal, X., Suñé, P. and Arnau, J. M. (2010) 'Consent in clinical trials: what do patients know?' *Contemporary Clinical Trials*, 31(5), pp. 443-446.

Hindmarsh, J., Reynolds, P. and Dunne, S. (2011) 'Exhibiting understanding: the body in apprenticeship.' *Journal of Pragmatics*, 43(2), pp. 489-503.

Hornik, K., Stinchcombe, M. and White, H. (1989) 'Multilayer feedforward networks are universal approximators.' *Neural Networks*, 2(5), pp. 359-366.

Hrubes, D. and Feldman, R. S. (2001) 'Nonverbal displays as indicants of task difficulty.' *Contemporary Educational Psychology*, 26(2), pp. 267-276.

Huang, G., Zhu, Q. and Siew, C. (2004) 'Extreme learning machine: a new learning scheme of feedforward neural networks.' *In IJCNN 2004 Proceedings IEEE International Joint Conference on Neural Networks*, vol. 2, Hotel Inter-Continental Budapest, Hungary, 25th-29th July 2004. IEEE, pp. 985-990.

Huang, G., Zhu, Q. and Siew, C. (2006) 'Extreme learning machine: theory and applications.' *Neurocomputing*, 70(1), pp. 489-501.

Huang, G., Huang, G., Song, S. and You, K. (2015) 'Trends in extreme learning machines: a review.' *Neural Networks*, 61, pp. 32-48.

Hunt, E. B. (1962) *Concept Learning: An Information Processing Problem*. New York: Wiley.

Hunt, E. B., Marin, J. and Stone, P. J. (1966) *Experiments in Induction*. New York: Academic Press.

Hush, D. R. and Horne, B. G. (1993) 'Progress in supervised neural networks.' *IEEE Signal Processing Magazine*, 10(1), pp. 8-39.

Jacobs, R. A. (1988) 'Increased rates of convergence through learning rate adaptation.' *Neural Networks*, 1(4), pp. 295-307.

Jecker, J., MacCoby, N., Breitrose, H. S. and Rose, E. D. (1964) 'Teacher accuracy in assessing cognitive visual feedback from students.' *Journal of Applied Psychology*, 48(6), pp. 393-397.

Jefford, M. and Moore, R. (2008) 'Improvement of informed consent and the quality of consent documents.' *Lancet Oncology*, 9(5), pp. 485-493.

Joffe, S., Cook, E. F., Cleary, P. D., Clark, J. W. and Weeks, J. C. (2001) 'Quality of informed consent: a new measure of understanding among research subjects.' *Journal of the National Cancer Institute*, 93(2), pp. 139-147.

Kaastra, I. and Boyd, M. (1996) 'Designing a neural network for forecasting financial and economic time series.' *Neurocomputing*, 10(3), pp. 215-236.

Karim, S. S. A., Richardson, B. A., Ramjee, G., Hoffman, I. F., Chirenje, Z. M., Taha, T., Kapina, M., Maslankowski, L., Coletti, A., Profy, A., Moench, T. R., Piwowar-Manning, E., Mâsse, B., Hillier, S. L., Soto-Torres, L. and HIV Prevention Trials Network (HPTN) 035 Study Team (2011) 'Safety and effectiveness of BufferGel and 0.5% PRO2000 gel for the prevention of HIV infection in women.' *AIDS (London, England)*, 25(7), pp. 957-966.

Kincaid, J. P., Fishburne, R. P, Rogers, R. L. and Chissom, B. S. (1975) *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel*. Virginia: National Technical Information Service. (RBR 8-75)

Kipp, M. (2015) *ANVIL: The Video Annotation Research Tool.* [Online] [Accessed on 8th November 2015] http://www.anvil-software.org

Kipp, M. (2001) 'ANVIL - a generic annotation tool for multimodal dialogue.' *In Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech 2001)*. Aalborg Congress and Culture Centre, Aalborg, Denmark. 3rd-7th September 2001, pp. 1367-1370.

Kohavi, R. (1995) 'A study of cross-validation and bootstrap for accuracy estimation and model selection.' *In IJCAI'95 Proceedings of the 14th International Joint Conference on Artificial Intelligence.* vol. 2. Palais de Congres Montreal, Quebec, 20th-25th August 1995. California: Morgan Kaufmann Publishers Inc., pp. 1137-1143.

Kolen, J. F. and Pollack, J. B. (1990) 'Back propagation is sensitive to initial conditions.' *In* Lippmann, R. P., Moody, J. E. and Touretzky, D. S. (eds.) *Advances in Neural Information Processing Systems 3*, San Francisco: Morgan Kaufman Publishers Inc., pp. 860-867.

Kotsiantis, S. B. (2013) 'Decision trees: a recent overview.' *Artificial Intelligence Review*, 39(4), pp. 261-283.

Knapp, M. L. and Hall, J. A. (1992) *Nonverbal Communication in Human Interaction*. 3rd ed., Fort Worth: Holt Rinehart and Winston.

Kramer, A. H. and Sangiovanni-Vincentelli, A. (1989) 'Efficient parallel learning algorithms for neural networks.' *In* Touretzky, D. S. (ed.) *Advances in Neural Information Processing Systems 1*, San Francisco: Morgan Kaufmann Publishers Inc., pp. 40-48.

Krauss, R. M., Chen, Y. and Chawla, P. (1996) 'Nonverbal behavior and nonverbal communication: what do conversational hand gestures tell us?' *In* Zanna, M. P. (ed.) *Advances in Experimental Social Psychology*, San Diego: Academic Press, pp. 389-450.

Lawrence, J. (1991) 'Data preparation for a neural network.' *AI Expert*, 6(11), pp. 34-41.

LeCun, Y., Bottou, L., Orr, G. B. and Müller, K. (1998) 'Efficient BackProp.' *In* Orr, G. B. and Müller, K. (eds.) *Neural Networks: Tricks of the Trade.* Springer Berlin Heidelberg, pp. 9-50.

LeCun, Y., Denker, J. S. and Solla, S. A. (1990a) 'Optimal brain damage.' *In* Touretzky, D. S. (ed.) *Advances in Neural Information Processing Systems 2*, San Francisco: Morgan Kaufmann Publishers Inc., pp. 598-605.

LeCun, Y., Matan, O., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., Jackel, L. D. and Baird, H. S. (1990b) 'Handwritten zip code recognition with multilayer networks.' *In Proceedings of the 10th International Conference on Pattern Recognition*. Vol. 2. Atlantic City, New Jersey, 16th-21st June 1990, IEEE, pp. 35-40.

Lema, V. M., Mbondo, M. and Kamau, E. M. (2009) 'Informed consent for clinical trials: a review.' *East African Medical Journal*, 86(3), pp. 133-142.

Lindegger, G., Milford, C., Slack, C., Quayle, M., Xaba, X., and Vardas, E. (2006) 'Beyond the checklist: assessing for understanding for HIV vaccine trial participation in South Africa.' *Acquired Immune Deficiency Syndrome*, 43(5), pp. 560-566.

Machida, S. (1986) 'Teacher accuracy in decoding nonverbal indicants of comprehension and noncomprehension in Anglo- and Mexican-American children.' *Journal of Educational Psychology*, 78(6), pp. 454-464.

MacQueen, K. M., Chen, M., Ramirez, C., Nnko, S. E. A. and Earp, K. M. (2014) 'Comparison of closed-ended, open-ended, and perceived informed consent comprehension measures for a mock HIV prevention trial among women in Tanzania' *PLoS One*, 9(8), pp. 1-8.

MacQueen, K. M., Vanichseni, S., Kitayaporn, D., Lin, L. S., Buavirat, A., Naiwatanakul, T., Raktham, S., Mock, P., Heyward, W. L., Des Jarlais, D. C., Choopanya, K. and Mastro, T. D. (1999) 'Willingness of injection drug users to participate in an HIV vaccine efficacy trial in Bangkok, Thailand.' *Journal of Acquired Immune Deficiency Syndromes*, 21(3), pp. 243-251.

Mandava, A., Pace, C., Campbell, B., Emanuel, E. and Grady, C. (2012) 'The quality of informed consent: mapping the landscape. A review of empirical data from developing and developed countries.' *Journal of Medical Ethics*, 38(6), pp. 356-365.

Marton, F. and Säljö, R. (2005) 'Approaches to learning.' *In* Marton, F., Hounsell, D. and Entwistle, N. (eds.) *The Experience of Learning: Implications for teaching and studying in higher education*. 3rd ed., Edinburgh: University of Edinburgh, Centre for Teaching, Learning and Assessment, pp. 39-58.

McLean, D., Bandar, Z. and O'Shea, J. D. (1998) 'The evolution of a feedforward neural network trained under backpropagation.' *In ICANNGA 97 Proceedings of International Conference on Artificial Neural Networks and Genetic Algorithms.* University of East Anglia, Norwich, 2nd- 4th April, Vienna: Springer, pp. 518-522.

McCulloch, W. and Pitts, W. (1943) 'A logical calculus of the ideas immanent in nervous activity.' *Bulletin of Mathematical Biophysics*, 5(4), pp. 115-133.

McQuaid, S. M., Woodworth, M., Hutton, E. L., Porter, S. and ten Brinke, L. (2015) 'Automated insights: verbal cues to deception in real-life high-stakes lies.' *Psychology, Crime & Law*, 21(7), pp. 617-631.

Mehrabian, A. (1968) 'Communication without words.' *Psychology Today*, 2(4), pp. 53-56.

Mehrabian, A. and Wiener, M. (1967) 'Decoding of inconsistent communications.' *Journal of Personality and Social Psychology*, 6(1), pp. 109-114.

Microsoft (2016a) *.NET Framework* [Online] [Accessed on 1st June 2015] https://www.microsoft.com/net

Microsoft (2016b) *Microsoft Media Foundation* [Online] [Accessed on 1st June 2015] https://msdn.microsoft.com/en-us/library/windows/desktop/ms694197(v=vs.85).aspx

Microsoft (2016c) *DirectShow* [Online] [Accessed on 1st June 2015] https://msdn.microsoft.com/en-us/library/windows/desktop/dd375454(v=vs.85).aspx

Miller, G. A. (1956) 'The magical number seven, plus or minus two: some limits on our capacity for processing information.' *Psychological Review*, 63(2), pp. 81-97.

Miller, C. K., O'Donnell, D. C., Searight, H. R. and Barbarash, R. A. (1996) 'The deaconess informed consent comprehension test: an assessment tool for clinical research subjects.' *Pharmacotherapy*, 16(5), pp. 872-878.

Minai, A. and Williams, R. (1993) 'On the derivatives of the sigmoid.' *Neural Networks*, 6(6), pp. 845-853.

Minsky, M. (1961) 'Steps toward artificial intelligence.' *Proceedings of the Institute of Radio Engineers*, 49(1), pp. 8-30.

Minsky, M. and Papert, S. (1988) *Perceptrons: An Introduction to Computational Geometry*. Massachusetts: MIT Press.

Miyake, N. (1986) 'Constructive interaction and the iterative process of understanding.' *Cognitive Science*, 10(2), pp. 151-177.

Moodley, K., Pather, M. and Myer, L. (2005) 'Informed consent and participant perceptions of influenza vaccine trials in South Africa.' *Journal of Medical Ethics*, 31(12), pp. 727-732.

Mukhopadhyay, S. C. (2015) 'Wearable sensors for human activity monitoring: a review.' *IEEE Sensors Journal*, 15(3), pp. 1321-1330.

Murthy, S. K. (1998) 'Automatic construction of decision trees from data: a multi-disciplinary survey.' *Data Mining and Knowledge Discovery*, 2(4), pp. 345-389.

National Institute of Medical Research Tanzania (2016) *Home*. [Online] [Accessed on 24th April 2016] http://www.nimr.or.tz/

Neill, S. (1991) *Classroom Nonverbal Communication*. New York: Routledge.

Newton, D. P. (2000) *Teaching for understanding: what it is and how to do it*. London: RoutledgeFalmer.

Nickerson, R. S. (1985) 'Understanding understanding.' *American Journal of Education*, 93(2), pp.201-239.

Nuremberg Code (1949) 'Trials of war criminals before the Nuremberg military. Tribunals under control counsel law.' 11(10), Washington: U.S. Government Printing Office, pp. 181-182.

Oxford English Dictionary (2016) *Comprehension, n*. [Online] [Accessed on 5th January 2016] http://www.oed.com/

Paris, A., Brandt, C., Cornu, C., Maison, P., Thalamas, C. and Cracowski, J. L. (2010) 'Informed consent document improvement does not increase patients' comprehension in biomedical research.' *British Journal of Clinical Pharmacology*, 69(3), pp. 231-237.

Patterson, M. L. (2014) 'Reflections on historical trends and prospects in contemporary nonverbal research.' *Journal of Nonverbal Behavior*, 38(2), pp. 171-180.

Peterson, C. and Rögnvaldsson, T. (1991) *An Introduction to Artificial Neural Networks*. CERN Summer School of Computing, (CERN Yellow Report 92-02), pp. 113-170.

Plaut D. C. and Hinton, G. E. (1987) 'Learning sets of filters using back propagation.' *Computer Speech and Language*, pp. 35-61.

Plaut, D. C., Nowlan, S. and Hinton, G. E. (1986) *Experiments on learning by back-propagation*. Pennsylvania: Carnegie-Mellon University (CMU-CS-86-126).

Pomerleau, D. A. (1989) 'ALVINN: an autonomous land vehicle in a neural network.' *In* Touretzky, D. S. (ed.) *Advances in Neural Information Processing Systems 1*, San Francisco: Morgan Kaufmann Publishers Inc., pp. 305-313.

Prechelt, L. (2012) 'Early stopping - but when?' *In* Orr, G. B. and Müller, K. (eds.) *Neural Networks: Tricks of the Trade.* Springer Berlin Heidelberg, pp. 53-67.

Priddy, K. L. and Keller, P. E. (2005) *Artificial Neural Networks: An Introduction*. Washington: SPIE Press.

Quinlan, J. R. (1986) 'Induction of decision trees.' *Machine Learning*, 1(1), pp. 81-106.

Quinlan, J. R. (1990) 'Decision trees and decisionmaking.' *IEEE Transactions on Systems, Man and Cybernetics*, 20(2), pp. 339-346.

Quinlan, J. R. (1993) *C4.5: programs for machine learning*. California: Morgan Kaufmann.

Quinlan, J. R. (1996) 'Improved use of continuous attributes in C4.5.' *Journal of Artificial Intelligence Research*, 4(1), pp. 77-90.

Quinlan, J. R. (1999) 'Simplifying decision trees.' *International Journal of Human-Computer Studies*, 51(2), pp. 497-510.

Rayner, K., Chace, K. H., Slattery, T. J. and Ashby, J. (2006) 'Eye movements as reflections of comprehension processes in reading.' *Scientific Studies of Reading*, 10(3), pp. 241-255.

Reed, R. (1993) 'Pruning algorithms - a survey.' *IEEE Transactions on Neural Networks*, 4(5), pp. 740-747.

Refaeilzadeh, P., Tang, L. and Liu, H. (2009) 'Cross-validation.' *In* Liu, L. and Özsu, T. (eds.) *Encyclopaedia of Database Systems*, Springer US, pp. 532-538.

Rissanen, J. (1983) 'A universal prior for integers and estimation by minimum description length.' *The Annals of Statistics,* 11(2), pp. 416-431.

Rothwell, J. (2002) *Artificial neural networks for psychological profiling using multichannels of nonverbal behaviour*. Ph.D. Manchester Metropolitan University.

Rothwell, J., Bandar, Z., O'Shea, J. and McLean, D. (2006) 'Silent talker: a new computer-based system for the analysis of facial cues to deception.' *Applied Cognitive Psychology*, 20(6), pp. 757-777.

Rothwell, J., Bandar, Z., O'Shea, J. and McLean, D. (2007) 'Charting the behavioural state of a person using a backpropagation neural network.' *Neural Computing & Applications*, 16(4), pp. 327-339.

RuleQuest Research (2010) *RuleQuest Research Data Mining Tools*. [Online] [Accessed on 28th March 2016] http://www.rulequest.com/index.html

RuleQuest Research (2012) *Is C5.0 Better Than C4.5?* [Online] [Accessed on 27th March 2016] http://rulequest.com/see5-comparison.html

Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986a) 'Learning representations by back-propagating errors.' *Nature*, 323, pp. 533-536.

Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986b) 'Learning internal representations by error propagation.' *In* Rumelhart, D. E. and McClelland, J. L. (eds.) *Parallel Distributed Processing*, Massachusetts: MIT Press, pp. 318-362.

Sand, K., Kaasa, S. and Loge, J. H. (2010) 'The understanding of informed consent information - definitions and measurements in empirical studies.' *AJOB Primary Research*, 1(2), pp. 4-24.

Sathik, M. and Jonathan, S. G. (2013) 'Effect of facial expressions on student's comprehension recognition in virtual educational environments.' *SpringerPlus*, 2, pp. 455-464.

Scherer, K. R. and Wallbott H. G. (1985) 'Analysis of nonverbal behavior.' *In* van Dijk, T. A. (ed.), *Handbook of Discourse Analysis*, London: Academic Press, pp. 199-230.

Schumm, J. S., Vaughn, S. and Sobol, M. C. (1997) 'Are they getting it? How to monitor student understanding in inclusive classrooms.' *Intervention in School and Clinic*, 32(3), pp. 168-171.

Sejnowski, T. J. and Rosenberg, C. R. (1987) 'Parallel networks that learn to pronounce English text.' *Complex Systems 1*, pp. 145-168.

Sejnowski, T. J. and Rosenberg, C. R. (1988) 'NETtalk: a parallel network that learns to read aloud.' *In* Anderson, J. A. and Rosenfeld, E. (eds.) *Neurocomputing: Foundations of Research*. Massachusetts: MIT Press, pp. 661-672.

Shannon, C. E. (1948) 'A mathematical theory of communication.' *The Bell System Technical Journal*, 27(4), pp. 623-656.

Sietsma, J. and Dow, R. J. F. (1988) 'Neural net pruning-why and how.' *In IEEE International Conference on Neural Networks*. San Diego, California, 24th - 27th July 1998. IEEE, pp. 325-333.

Simpson, K., MacQueen, K. M., Mack, N., Friedland, B. and Nnko, S. (2010) *Enhancing local verbal and non-verbal communication for informed consent processes in Tanzania, study #10159*. North Carolina: Family Health International 360.

Snyder, M. (1974) 'Self-monitoring of expressive behavior.' *Journal of Personality and Social Psychology*, 30(4), pp. 526-537.

Squared5 (2016) *MPEG Streamclip*. [Online] [Accessed on 9th March 2016] http://www.squared5.com/

Stone, M. (1974) 'Cross-validatory choice and assessment of statistical predictions.' *Journal of the Royal Statistical Society*, 36(2), pp. 111-147.

Tam, N. T., Huy, N. T., Thoa, L. T. B., Long, N. P., Trang, N. T. H., Hirayama, K. and Karbwang, J. (2015) 'Participants' understanding of informed consent in clinical trials over three decades: systematic review and meta-analysis.' *Bulletin of the World Health Organization*, 93(3), pp. 186-198.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M. and Sedivy, J. C. (1995) 'Integration of visual and linguistic information in spoken language comprehension.' *Science*, 268(5217), pp. 1632-1634.

The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (1979) 'The Belmont Report: Ethical principles and

guidelines for the protection of human subjects of research.' Washington: U.S. Government Printing Office.

The Language Archive (2015) *ELAN*. [Online] [Accessed on 14th November 2015] http://tla.mpi.nl/tools/tla-tools/elan/

Thimm G. and Fiesler, E. (1995) 'Neural network initialization.' *In* Mira, J. and Sandoval, F. (eds.) *From Natural to Artificial Neural Computation.* Berlin: Springer, pp. 535-542.

UNAIDS (2010) *Global report: UNAIDS report on the global AIDS epidemic 2010*. UN Joint Programme on HIV/AIDS (UNAIDS) [Online] [Accessed on 26th April 2016] http://www.unaids.org/globalreport/documents/20101123_GlobalReport_full_en.pdf

Vanichseni, S., Tappero, J. W., Pitisuttithum, P., Kitayaporn, D., Mastro, T. D., Vimutisunthorn, E., Van Griensvan, F., Heyward, W. L., Francis, D. P., Choopanya, K. and Bangkok Vaccine Evaluation Group (2004) 'Recruitment, screening and characteristics of injection drug users participating in the AIDSVAX B/E HIV vaccine trial, Bangkok, Thailand.' *AIDS (London, England)*, 18(2), pp. 311-316.

Van Amelsvoort, M. and Krahmer, E. (2009) 'Appraisal of children's facial expressions while performing mathematics problems.' *In CogSci2009 Proceedings of the 31st Annual Meeting of the Cognitive Science Society*, Virje University Amsterdam, Amsterdam, 29th July - 1st August. pp. 1698-1703.

Van Amelsvoort, M., Joosten, B., Krahmer, E. and Postma, E. (2013) 'Using non-verbal cues to (automatically) assess children's performance difficulties with arithmetic problems.' *Computers in Human Behavior*, 29(3), pp. 654-664.

Vax Report (2010) *Understanding Viral Diversity VAX Report*. [Online] [Accessed on 1st March 2016] http://www.vaxreport.org/Back-Issues/Pages/UnderstandingViralDiversity.aspx

Velloso, E., Bulling, A. and Gellersen, H. (2013) 'AutoBAP: automatic coding of body action and posture units from wearable sensors.' *In ACII 2013 Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction.* Geneva, Switzerland, 2nd-5th September 2013. Washington: IEEE Computer Society Washington, pp. 135-140.

Vygotsky, L. S. (1980) *Mind in society: The development of higher psychological processes*. Massachusetts: Harvard University Press.

Waring, H. Z. (2002) 'Expressing noncomprehension in a US graduate seminar.' *Journal of Pragmatics*, 34(12), pp. 1711-1731.

Webb, J. M., Diana, E. M., Luft, P., Brooks, E. W. and Breenan, E. L. (1997) 'Influence of pedagogical expertise and feedback on assessing student comprehension from nonverbal behavior.' *The Journal of Educational Research*, 91(2), pp. 89-97.

Werbos, P. (1974) *Beyond regression: new tools for prediction and analysis in the behavioural sciences*. Ph.D. Harvard University.

West, H. F. and Baile, W. F. (2010) '"Tell me what you understand": the importance of checking for patient understanding.' *Journal of Supportive Oncology*, 8(5), pp. 216-218.

Wessels, L. F. A. and Barnard, E. (1992) 'Avoiding false local minima by proper initialization of connections.' *IEEE Transactions on Neural Networks*, 3(6), pp. 899-905.

World Health Organization (2007) *Family Planning: A Global Handbook for Providers*. Maryland: Johns Hopkins.

Widrow, B., Greenblatt, A., Kim, Y. and Park, D. (2013) 'The No-Prop algorithm: a new learning algorithm for multilayer neural networks.' *Neural Networks*, 37, pp. 182-188.

Widrow, B. and Hoff, M. E. (1960) 'Adaptive switching circuits.' *IRE WESCON Convention Record*, 4, pp. 96-104.

Widrow, B. and Lehr, M. (1998) 'Perceptrons, adalines, and backpropagation.' *In* Arbib, M. A. (ed.) *The Handbook of Brain Theory and Neural Networks*, Massachusetts: MIT Press, pp. 719-724.

Widrow, B., Rumelhart, D. E. and Lehr, M. A. (1994) 'Neural networks: applications in industry, business and science.' *Communications of the ACM*, 37(3), pp. 93-105.

Witten, I. H. and Frank, E. (2005) *Data Mining: Practical Machine Learning Tools and Techniques*. 2nd ed., Massachusetts: Morgan Kaufmann Publishers Inc.

Witten, I. H., Frank, E. and Hall, M. A. (2011) *Data Mining: Practical Machine Learning Tools and Techniques.* 3rd ed., Massachusetts: Morgan Kaufmann Publishers Inc.

Won, A. S., Bailenson, J. N., Stathatos, S. C. and Dai, W. (2014a) 'Automatically detected nonverbal behavior predicts creativity in collaborating dyads.' *Journal of Nonverbal Behavior*, 38(3), pp. 389-408.

Won, A. S., Bailenson, J. N. and Janssen, J. H. (2014b) 'Automatic detection of nonverbal behavior predicts learning in dyadic interactions.' *IEEE Transactions on Affective Computing*, 5(2), pp. 112-125.

World Medical Association (2013) 'World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects.' *Journal of the American Medical Association*, 310(20), pp. 2191-2194.

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., Zhou, Z., Steinbach, M., Hand, D. J. and Steinberg, D. (2008) 'Top 10 algorithms in data mining.' *Knowledge and Information Systems*, 14(1), pp. 1-37.

Zhang, G. P. (2000) 'Neural networks for classification: a survey.' *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 30(4), pp. 451-462.

Zuckerman, M., Amidon, M. D., Bishop, S. E., and Pomerantz, S. D. (1982) 'Face and tone of voice in the communication of deception.' *Journal of Personality and Social Psychology*, 43(2), pp. 347-357.

Zurada, J. (1992) *Introduction to Artificial Neural Systems*. Minnesota: West Publishing Co.

Appendix A. Sociodemographic Form

Please note: This document is only available upon request from Family Health International 360 due to copyright restrictions.

Appendix B. Developmental Phase Interviewer Instructions

Please note: This document is only available upon request from Family Health International 360 due to copyright restrictions.

Appendix C. Developmental Phase Checklist

Please note: This document is only available upon request from Family Health International 360 due to copyright restrictions.

Appendix D. Developmental Phase Summary Sheet

Please note: This document is only available upon request from Family Health International 360 due to copyright restrictions.

Appendix E. PrEP Trial Mock Informed Consent Document

Please note: This document is only available upon request from Family Health International 360 due to copyright restrictions.

Appendix F. Informed Consent Closed Comprehension Assessment

Please note: This document is only available upon request from Family Health International 360 due to copyright restrictions.

Appendix G. Informed Consent Open Comprehension Assessment

Please note: This document is only available upon request from Family Health International 360 due to copyright restrictions.

Appendix H. Informed Consent Self-Perceived Comprehension Assessment

Please note: This document is only available upon request from Family Health International 360 due to copyright restrictions.

Appendix I. Willingness to Join Mock Clinical Trial Form

Please note: This document is only available upon request from Family Health International 360 due to copyright restrictions.

Appendix J. Exploratory Testing Phase Closed Summary Sheet

Please note: This document is only available upon request from Family Health International 360 due to copyright restrictions.

Appendix K. Exploratory Testing Phase Open Summary Sheet

Please note: This document is only available upon request from Family Health International 360 due to copyright restrictions.

Appendix L. Participant Information Sheet

You are invited to take part in a research study on the measurement of human comprehension. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read and consider the following information carefully. If you have any queries, please do not hesitate to ask for clarification or more information before consenting to participate.

What will I have to do if I take part?

You will perform a short learning task followed by a small set of questions. You will then perform another short learning task followed by another small set of questions. A camera operator will be present in the room recording the participant during both learning tasks. After there will be a small debriefing session where you have the opportunity to ask any questions and will be asked a few questions about yourself and the study you have participated in.

Do I have to participate?

Participation is voluntary. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason.

How will the recorded media be used?

The audio video recordings of your activities made during this research study will be used only for analysis. No other use will be made of them without your written permission, and no one outside the project will be allowed access to the original recordings.

Will my taking part in this study be kept confidential?

All information that you provide will be strictly confidential and you will not be identifiable in any reports or publications. No information collected will be shown to anyone apart from authorised Manchester Metropolitan University research staff. Transcripts and video recordings will be stored in a locked room on a password protected computer, which only authorised Manchester Metropolitan University research staff have access to. Transcripts will be anonymised and parts in which participants might be identified will be avoided in publications. The anonymous video recordings will not be published or distributed to third parties.

How long will the study data be held for?

The data from the study will be retained till completion of the research project when it is anticipated that the thesis is submitted and destroyed as confidential waste thereafter.

Who has reviewed the study?

Before any research goes ahead it has to be checked by a Research Ethics Committee. They make sure that the research is fair. This study has been approved by the Manchester Metropolitan University Research Faculty Academic Ethics Committee.

Contact details for further information

Fiona Buckingham (Researcher)
Email: fiona.j.buckingham@stu.mmu.ac.uk

Dr. Keeley Crockett (Supervisor)
Email: K.Crockett@mmu.ac.uk

Address: Manchester Metropolitan University, Chester Street, Manchester, M1 5GD.

Thank you for reading this.

Appendix M. Participant Consent Form

Please tick box if you agree with the statement.

Taking Part

1. I confirm that I have read and understood the participant information sheet dated DD/MM/2012. ☐
2. I have been given the opportunity to ask questions about the study and when asked have had them all answered satisfactorily. ☐
3. I agree to participate in the study, I consent to being audio, and video recorded as part of the study. ☐
4. I understand that my participation is voluntary and that I can withdraw from the study at any time without giving any reason. ☐

Use of the information I provide for this project only

5. I understand that any information recorded in the investigation will remain confidential and information that identifies me will not be published. ☐
6. I understand that my words may be quoted in publications, reports and other research outputs. ☐
7. I agree for the video to be used in this study. ☐

Use of the information I provide beyond this project

8. I agree for the video to be used in future studies. ☐

_____	_____	_____
Participant Name [printed]	Signature	Date
_____	_____	_____
Researcher Name [printed]	Signature	Date

Appendix N. Data Collection Form

Participant Information

Date of Birth (*Please state your date of birth*)

_ _ / _ _ / _ _ _ _

Gender (*Please tick the box that best describes your gender*)

☐ Male ☐ Female

Marital Status (*Please tick the box that best describes your marital status*)

☐ Single ☐ Married ☐ Divorced ☐ Separated ☐ Widowed ☐ Not disclosed

Religion (*Please tick the box that best describes your religion*)

☐ Buddhist ☐ Christian ☐ Hindu ☐ Jewish ☐ Muslim ☐ Sikh

☐ Other ☐ No religion or belief ☐ Not Disclosed

Ethnicity (*Please tick the box that best describes your ethnicity*)

Asian or Asian British	*Black or Black British*	*White*	*Mixed*	*Other*
☐ Bangladeshi	☐ African	☐ British	☐ African	☐ Other ethnicity
☐ Chinese	☐ Caribbean	☐ Irish	☐ Caribbean	☐ Not Disclosed
☐ Indian	☐ Other black	☐ Other white	☐ Asian	
☐ Pakistani			☐ Other mixed	
☐ Other Asian				

Education (*Please tick the box that best reflects your highest qualification obtained*)

☐ GCSE ☐ A-Levels ☐ O-Levels ☐ Foundation Degree ☐ Bachelors Degree

☐ Masters Degree ☐ PhD ☐ Other

Occupation (*Please state your current occupation*)

Job title: _____

Prior Knowledge (*Please tick the box that best reflects your level of knowledge on Termites prior to this study*)

☐ None ☐ Low (*novice*) ☐ Medium ☐ High (*expert*)

Thank you for participating in the research study.

Interview Information

PSN: _ _ _

Date: _ _ / _ _ / _ _ _ _

Condition: ☐ 1A

☐ 1B

☐ 2A

☐ 2B

Appendix O. Expert Information Sheet

Background

My research is focused upon measuring human comprehension. In my new study, I am going to be presenting a short video to the participants followed by a small set of questions. I would like to request your expertise on the subject area in order to develop difficulty and easy questions on the video content with expert agreement. The following sections provide further information and instructions on what I am looking for.

Study: Short Video

The short video is on Termites. The video has a duration of 8 minutes and 40 seconds. Please watch the video and whilst doing so think of associated difficult and easy questions that you could ask a potential viewer of the video to help assess whether they comprehended the content of the video.

Study: Question Instructions

Using the video on the subject Termites and your expertise, I would like you to develop a list of twenty questions using the following instructions:

1. The expert independently develops ten open-ended questions.

    - Within the set of ten open-ended questions, five of the open-ended questions need to designed and labelled as difficult and five of the open-ended questions need to designed and labelled as easy.

2. The expert independently develops ten closed questions i.e. True/False.

    - Within the set of ten closed questions, five of the closed questions need to designed and labelled as difficult and five of the closed questions need to designed and labelled as easy.

3. The group of experts shall come together each with their own set of difficult and easy questions relating to the video and present their questions to the group. The group of experts shall collaboratively decide and agree upon the final set of twenty questions with five easy and difficult questions within the ten open-ended and closed question sets.

    - With each question draw up a list of points, which the answer should contain in order to guide marking of the participant responses.
    - Highlight justification as to why you regard a question to be easy or difficult.

When designing the questions it would be useful to have inference questions that encourage deeper assessment of comprehension.

Contact Details for Further Information

If you have any queries, please do not hesitate to ask for clarification or more information.

Fiona Buckingham (Researcher)
Email: fiona.j.buckingham@stu.mmu.ac.uk

Dr. Keeley Crockett (Supervisor)
Email: K.Crockett@mmu.ac.uk

Thank you for reading this.

Appendix P. Interviewer Instructions

Introduction

Morning/Afternoon
Firstly, you will watch a short video on Termites.
Afterwards, I will ask you a set of questions about the Termites video.

Learning Task

Are you ready to watch the video?
Play video on Termites (8 minutes 40 seconds)

Learning Task Questions

I am now going to ask you a set of questions about the Termites video.
Please answer each question as best as you can.

Randomise question presentation on closed/open and easy/hard.

Read the question.
Wait for a participant response.
If the participant does not respond **then** repeat the question.
If the participant still does not respond **then** ask a neutral probe question.
Neutral Probes: What is your best guess? **or** Anything else?

Easy Closed Questions

1. Termites like cold conditions.
2. Magnetic termites build their colonies around geographical magnetic areas.
3. A termite queen lays over one thousand eggs per day.
4. The shafts in a termite colony penetrate twenty metres below the colony to reach the water table.
5. There are two queens in each termite colony.

Hard Closed Questions

1. The air within a termite colony is un-breathable.
2. The temperature in a termite colony is above 35°C.
3. The workers are all male termites.
4. There is no king within a termite colony.
5. The cellar is the coolest part of the termite colony.

Easy Open Questions

1. Describe the main qualities you would want from a home?
2. What is the family structure within a termite colony? Describe their basic functions.
3. Describe why magnetic termites from Australia build their colony north/south.
4. Describe the basic structure of a termite colony and explain how the architecture is essential to maintain the colony.
5. What countries or regions can termites be found?

Hard Open Questions

1. Describe a symbiotic relationship and discuss why this applies to termites**.**
2. Describe why it is essential for the termites to construct their colony in a tower structure.
3. Discuss the role of the worker in the termite social structure**.**
4. There is normal air flow within the termite colony. Describe how this is achieved using the complex architecture**.**
5. A termite's food supply is dead wood. How do they digest it?

Self-Assessment Questions

Ask the following two questions after each closed and open-ended question.

1. How difficult was the last question to answer?
    a. Very easy
    b. Easy
    c. Moderate
    d. Difficult
    e. Very difficult
2. Do you think that you answered the last question correctly or incorrectly?

High Comprehension Benchmark Questions

1. What day is it today?
2. What is the capital of England?
3. Name all of the months beginning with the letter "J".
4. How many years are there in a decade?
5. What is the fifth letter of the alphabet?

Conclusion

Thank you for participating in the study.

Appendix Q. Pruning Experiment Results

CF	MNO	Leaves	Tree Size	Total CA	Normalised Total CA	TP CA	TN CA
0.005	2	1297	2593	92.950	92.671	91.021	94.320
	3	1234	2467	92.536	92.257	90.608	93.905
	4	1161	2321	92.070	91.783	90.083	93.482
	5	1091	2181	91.575	91.253	89.346	93.159
	10	818	1635	89.419	89.017	86.644	91.390
	15	648	1295	87.488	87.073	84.620	89.525
	20	564	1127	85.812	85.217	81.706	88.729
	25	507	1013	84.682	84.048	80.302	87.794
	30	428	855	83.640	82.938	78.786	87.090
0.05	2	1488	2975	93.869	93.645	92.319	94.970
	3	1424	2847	93.371	93.137	91.753	94.522
	4	1307	2613	92.867	92.608	91.080	94.136
	5	1231	2461	92.308	92.008	90.237	93.780
	10	900	1799	89.924	89.581	87.558	91.604
	15	743	1485	87.836	87.445	85.134	89.756
	20	623	1245	86.292	85.771	82.691	88.850
	25	557	1113	85.015	84.485	81.352	87.618
	30	477	953	84.050	83.493	80.202	86.784
0.1	2	1524	3047	94.089	93.874	92.602	95.146
	3	1462	2923	93.599	93.373	92.036	94.710
	4	1335	2669	92.996	92.752	91.310	94.195
	5	1259	2517	92.440	92.155	90.473	93.838
	10	915	1829	89.997	89.659	87.664	91.654
	15	770	1539	87.980	87.589	85.275	89.902
	20	639	1277	86.380	85.860	82.786	88.934
	25	566	1131	85.140	84.611	81.482	87.739
	30	489	977	84.094	83.554	80.361	86.746
0.15	2	1561	3121	94.107	93.896	92.655	95.138
	3	1495	2989	93.648	93.427	92.118	94.735
	4	1343	2685	93.041	92.804	91.405	94.203
	5	1273	2545	92.558	92.288	90.691	93.884
	10	929	1857	90.014	89.671	87.641	91.701
	15	781	1561	88.022	87.636	85.358	89.915
	20	645	1289	86.412	85.895	82.839	88.951
	25	574	1147	85.118	84.578	81.388	87.769
	30	498	995	84.109	83.564	80.349	86.780
0.2	2	1585	3169	94.241	94.041	92.856	95.226
	3	1513	3025	93.719	93.504	92.231	94.777
	4	1353	2705	93.082	92.849	91.470	94.228
	5	1281	2561	92.599	92.333	90.756	93.910
	10	937	1873	90.053	89.707	87.659	91.755
	15	783	1565	88.012	87.632	85.387	89.877
	20	647	1293	86.422	85.908	82.874	88.942
	25	581	1161	85.201	84.678	81.588	87.769
	30	500	999	84.165	83.629	80.461	86.796
0.25	2	1659	3317	94.381	94.191	93.068	95.314
	3	1525	3049	93.734	93.526	92.295	94.756
	4	1373	2745	93.151	92.925	91.593	94.257
	5	1303	2605	92.604	92.337	90.756	93.918
	10	948	1895	90.066	89.727	87.723	91.730
	15	790	1579	88.051	87.666	85.387	89.944
	20	650	1299	86.444	85.944	82.992	88.896
	25	586	1171	85.204	84.687	81.629	87.744
	30	508	1015	84.140	83.609	80.467	86.750

CF	MNO	Leaves	Tree Size	Total CA	Normalised Total CA	TP CA	TN CA
	2	1693	3385	94.450	94.262	93.157	95.368
	3	1544	3087	93.805	93.604	92.413	94.794
	4	1402	2803	93.197	92.978	91.682	94.274
	5	1315	2629	92.653	92.394	90.862	93.926
0.3	10	960	1919	90.090	89.762	87.824	91.701
	15	812	1623	88.061	87.662	85.305	90.020
	20	651	1301	86.424	85.938	83.063	88.813
	25	588	1175	85.192	84.680	81.659	87.702
	30	512	1023	84.143	83.618	80.514	86.721
	2	1709	3417	94.459	94.273	93.174	95.372
	3	1551	3101	93.810	93.606	92.402	94.811
	4	1408	2815	93.207	92.992	91.723	94.262
	5	1321	2641	92.656	92.400	90.885	93.914
0.4	10	972	1943	90.144	89.828	87.959	91.696
	15	817	1633	88.149	87.770	85.529	90.011
	20	664	1327	86.459	85.973	83.104	88.842
	25	589	1177	85.248	84.720	81.600	87.840
	30	518	1035	84.167	83.633	80.479	86.788
	2	1710	3419	94.469	94.284	93.192	95.377
	3	1563	3125	93.820	93.617	92.419	94.815
	4	1418	2835	93.227	93.014	91.759	94.270
	5	1334	2667	92.656	92.402	90.903	93.901
0.5	10	979	1957	90.161	89.850	88.013	91.688
	15	818	1635	88.157	87.774	85.511	90.036
	20	664	1327	86.471	85.988	83.134	88.842
	25	591	1181	85.246	84.718	81.600	87.836
	30	520	1039	84.140	83.611	80.485	86.738

Appendix R. Publications

Buckingham, F. J., Crockett, K. A., Bandar, Z. A. and O'Shea, J. D. (2014) 'FATHOM: A neural network-based non-verbal human comprehension detection system for learning environments.' *In 2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*. Orlando, Florida, 9th-12th December 2014, IEEE, pp. 403-409.

Crockett, K., O'Shea, J., Buckingham, F., Bandar, Z., MacQueen, K., Chen, M. and Simpson, K. (2013) 'FATHOMing out interdisciplinary research transfer.' *In 2013 IEEE Symposium on Computational Intelligence for Engineering Solutions (CIES)*. Singapore, 16th-19th April 2013, IEEE, pp. 1-8.

Buckingham, F. J., Crockett, K. A., Bandar, Z. A., O'Shea, J. D., MacQueen, K. M. and Chen, M. (2012) 'Measuring human comprehension from nonverbal behaviour using artificial neural networks.' *In 2012 International Joint Conference on Neural Networks (IJCNN)*. Brisbane, Queensland, 10th-15th June 2012, IEEE, pp. 1-8.

FATHOM: A Neural Network-based Non-verbal Human Comprehension Detection System for Learning Environments

Fiona. J. Buckingham, Keeley A. Crockett, Zuhair A. Bandar, James D. O'Shea
School of Computing, Mathematics and Digital Technology
Manchester Metropolitan University
Chester Street, Manchester, M1 5GD, UK
fiona.j.buckingham@stu.mmu.ac.uk

Abstract— **This paper presents the application of FATHOM, a computerised non-verbal comprehension detection system, to distinguish participant comprehension levels in an interactive tutorial. FATHOM detects high and low levels of human comprehension by concurrently tracking multiple non-verbal behaviours using artificial neural networks. Presently, human comprehension is predominantly monitored from written and spoken language. Therefore, a large niche exists for exploring human comprehension detection from a non-verbal behavioral perspective using artificially intelligent computational models such as neural networks. In this paper, FATHOM was applied to a video-recorded exploratory study containing a learning task designed to elicit high and low comprehension states from the learner. The learning task comprised of watching a video on termites, suitable for the general public and an interview led question and answer session. This paper describes how FATHOM's comprehension classifier artificial neural network was trained and validated in comprehension detection using the standard backpropagation algorithm. The results show that high and low comprehension states can be detected from learner's non-verbal behavioural cues with testing classification accuracies above 76%.**

Keywords—artificial neural networks; backpropagation; comprehension; FATHOM; non-verbal behaviour

I. INTRODUCTION

Non-verbal behaviour is a form of non-linguistic communication that automatically accompanies verbal conversation. Gestures, facial expressions, and body movement are all examples of non-verbal behaviour [1]. Little work has been done on automatic comprehension detection, yet humans exhibit non-verbal cues consistently while undertaking day-to-day tasks. Thus, the research presented in this paper seeks to examine whether patterns of comprehension and non-comprehension exist within non-verbal behavioural cues.

Previous classroom studies [2-7] have identified non-verbal behavioural indicators of non-comprehension, including facial behaviour, hand and body movements. However, this work has largely relied on subjective human coding [8] with associated inconsistency and upon verbal techniques. Thus there is a role for a non-verbal multichannel, comprehension detection system capable of reliably classifying human comprehension through facial non-verbal behaviour.

Comprehension is often associated with written language [9] and is often defined as "the process of simultaneously extracting and constructing meaning through interaction and involvement with written language" [9]. In this research, we define comprehension as the learner demonstrating through interaction with a tutorial, (via verbal communication and/or non-verbal behaviour), that they understand or grasp the meaning of the tutorial material presented to them at a given point in time. The tutorial in this paper (described in Section V) comprised of each participant watching a factual video and participating in a question and answer (Q&A) session immediately after.

FATHOM [10], is an artificial neural networks (ANN) based system developed specifically to detect levels of comprehension. FATHOM was developed based around an existing physiological profiling system known as Silent Talker [11] and was first trialled during an informed consent assessment process carried out in North-western Tanzania, Africa using a setting similar to that used for a Human Immunodeficiency Virus (HIV)/Acquired Immunodeficiency Syndrome (AIDS) prevention randomized study [10]. The work produced strong evidence [10] that detectable patterns of comprehension and miscomprehension exist within the monitored facial non-verbal multichannels, for the sample of African women with a limited set of non-verbal behavioural features. Initial observations provide grounds to suspect that there will be more, less obvious, micro gestures available for classification.

The aim of the research presented in this paper is to apply FATHOM as a comprehension detection system to a learning task designed to distinguish high and low comprehension states from the learner based on facial non-verbal cues. In order to assess FATHOM's ability, a new exploratory study was designed to capture comprehension levels of adults over the age of 18. The motivation of this work is to ultimately link FATHOM to pedagogical intervention in learner-adaptive online teaching and learning tutorials that could be delivered in 24/7 scenarios to improve the overall learning experience.

978-1-4799-4518-4/14/$31.00 ©2014 IEEE

This paper continues as follows: Sections II and III review non-verbal behaviour, comprehension and learning. Section IV describes FATHOM - a comprehension detection system using ANNs. Section V presents the experimental study methodology and results. The conclusion and further work can be found in Sections VI and VII.

II. Non-verbal Behaviour

Non-verbal behaviour consists of a variety of signals or cues including visual, audio, tactile and chemical which are exhibited by human beings in order to express themselves [12]. The non-verbal cues are revealed often before a verbal response [13] and potentially can provide early signals to the listener about the sender's state whilst they are formulating the actual verbal response. Typically, non-verbal behaviours are being generated before, during and after the sender articulates a verbal response. A large number of non-verbal behavioural channels are available [14] and research has been undertaken to collect data on individuals to try and identify patterns associated with an individual's state. Knapp and Hall [15] stated that as humans often communicate face-to-face – the face was a source of rich information and should be given higher precedence. Mehrabian [16] found that around 55% of non-verbal messages communicated by an individual came from facial behaviour expressions and debated that even when cues from facial behaviour were not consistent with the verbal response that the listener was most impacted by the facial emotion expressed [17].

Little work has been on the automatic detection and classification of non-verbal behaviour. The traditional way still is to use human judges to code the non-verbal behaviour channels [18,19]. Nonetheless, the judges have to be trained, their assessment is subjective and they can concentrate upon only a limited number of channels at one time [11]. The whole process is time consuming with Johnson [20] reporting it can take one hour to code one channel from one minute of film.

Attempts have been made to automatically detect non-verbal cues using more recent technology using the Microsoft Kinect computer vision algorithm [21], however these attempts have been looking at full body gestures as opposed to fine grained channels which are used by FATHOM. Digital technologies such as camcorders and the Microsoft Kinect have the strength of being able to capture large volumes of continuous non-verbal channels, which can be stored as a multimedia file and used for single or repeated post analysis. On the other hand, digital recordings are limited to environments that facilitate the setup of the technology [22]. Therefore, a niche exists for a computer-based system that is able to automatically monitor multiple channels of non-verbal behaviour from digital video recordings such as FATHOM described in Section IV.

III. Comprehension and Learning

Trying to establish whether or not a learner is comprehending a tutorial as a whole or as independent elements is a non-trivial task. Early work by Dollaghan et al. [23] found that in a classroom environment, whilst a child may not comprehend everything they hear, they should be able to recognise that they do not understand and know way to ask for help. From a teacher perspective, it is important that every child knows how to listen, and asks when he or she does not understand. However, in large classes the ability of an individual to monitor and detect comprehension levels of every child is a very great challenge.

Previous work on detecting comprehension has primarily examined language comprehension [24] from reading, writing and listening aspects. However in [5] children participating in a lesson on electricity containing both easy and hard material were videotaped and individual observers were able to distinguish children comprehending and not comprehending from the non-verbal behaviours alone that they exhibited [5]. Amelsvoort et al. [8] examined if non-verbal cues give out information on how a child perceives the difficulty of an arithmetic problem. Again through audio-visual recordings, children were analysed (after the event) by manually identifying head movements using the Active Appearance Model, which were then used to train a classifier. Empirical experiments undertaken in the study found that it was possible to estimate the difficulty level at above chance levels from non-verbal cues. Other research in classroom environments has used human decoders to analyse recordings to identify non-verbal behavioural patterns associated with non-comprehension [3,25]. Research to date involving classroom based studies shows that that comprehension and non-comprehension patterns do reside within non-verbal behaviour alone but the results are often based on only a few non-verbal channels. This is because human observers cannot process more than 7 ± 2 pieces of information [26] at the same time in immediate memory. In addition, the detection of non-comprehension at a certain point in time occurs after the tutorial has taken place so no remedial action can be put in place by the teacher. The solution proposed in this paper is a comprehension detection system, known as FATHOM which can monitor multiple non-verbal behaviours for comprehension detection and make an automatic decision at a predefined point in time on a person's level of comprehension. This removes the subjectivity of human experts and allows more channels to contribute towards the classification of comprehension. An outline of FATHOM is given in the following section.

IV. FATHOM

FATHOM is computer-based comprehension detection system that uses a collection of neural networks to concurrently monitor multiple channels of human non-verbal behaviour [10]. Fig. 1 shows a screenshot of the FATHOM software automatically processing a learner's facial non-verbal behavioural channels contained within the currently displayed video frame. As the video plays, FATHOM analyses the learner's non-verbal behaviours in every video frame within a predefined period of time measured in seconds. The predefined period of time is known as a timeslot, which can be fixed e.g. every *n* second(s) or variable. In this paper, FATHOM was configured to use a fixed one second timeslot. At the end of the timeslot, the overall comprehension/non-comprehension classification level is outputted to FATHOM's user interface.

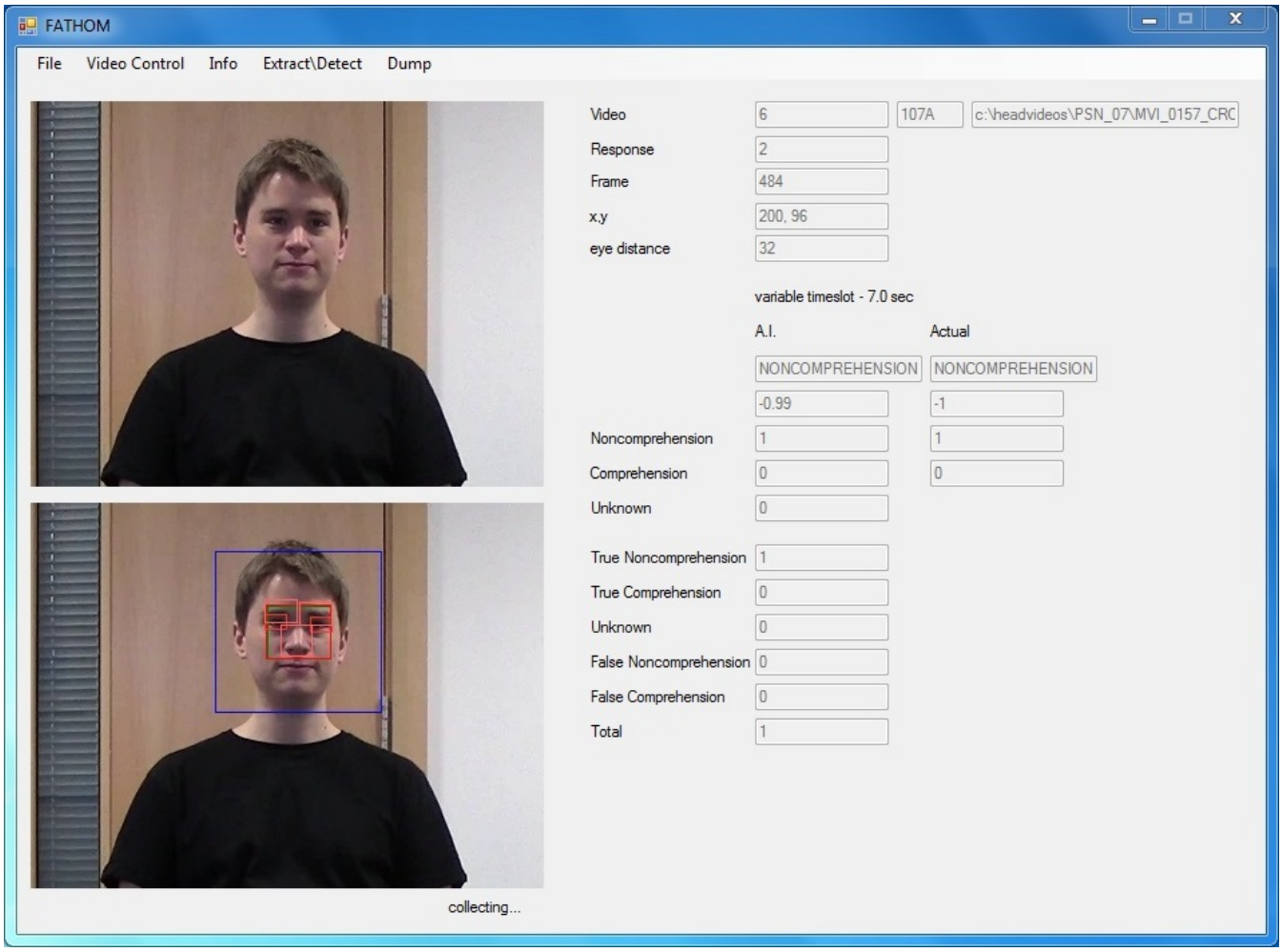


Fig. 1. FATHOM.

FATHOM has forty non-verbal behavioural channels [12], which are categorised as:

- Face (20 channels) e.g. face blushing/blanching.
- Eyes (16 channels) e.g. left/right eye gaze.
- Other (4 channels) e.g. gender.

The neural networks contained within FATHOM are: the object locator ANNs, the pattern detector ANNs and a single comprehension classifier ANN. Firstly, the object locators ANNs identify the location of non-verbal features such as the eyes, eyebrows and the nose. After the object locator ANNs have located their non-verbal feature, the pattern detectors ANNs identify the state of the object such as the right eye gazing to the right. The states of each non-verbal channel is collated for the timeslot and then passed to the final ANN, the comprehension classifier ANN, which outputs a value between +1 and -1, which indicates whether the person was comprehending (+1) or not comprehending (-1) during that period of time.

This research paper is focused upon the training and validation of FATHOM's comprehension classifier ANN with a dataset from an exploratory study (Section V) containing a learning task with associated comprehension assessment questions. FATHOM's comprehension classifier ANN was trained with the standard backpropagation algorithm [27,28] using incremental weight updating and the delta (δ) rule [29]. *n*-fold cross-validation [30] was used to randomly partition the dataset in to *n* equally sized subsets. A single subset is retained as the test set and the remaining subsets are used as the training set to form a single fold. The cross-validation process repeats until all *n* subsets have been used as the test set once (*n*-folds). The training set is used to train the FATHOM comprehension classifier ANN and the test set was used to determine the error rate of the trained FATHOM comprehension classifier ANN. The advantage of *n*-fold cross-validation is that all samples within the dataset are used and the results from the *n* folds can be averaged. Section V describes the experimental study on detecting learner comprehension.

V. Experimental Study: Detecting Learner Comprehension

The primary aim of the experimental study was to identify whether high and low human comprehension associated

multichannels of non-verbal behaviour reside within a video-recorded British (UK-based/English speaking) sample of participants. The participants were filmed whilst watching the factual video and during the Q&A session but only the videoed non-verbal cues exhibited during the reading of the questions in the Q&A session were used for post analysis by FATHOM. The exploratory study builds upon lessons learnt in a previous a research study [10] where an African female video-based non-verbal dataset was used to train and validate a backpropagation neural network in the detection of human comprehension. This section will outline the methodology of the experimental study and present the results.

A. Participants

Forty participants were selected to participate in the study from academic and technical staff at the Manchester Metropolitan University (MMU) in the UK. The sample was composed of 20 males and 20 females. The males had a mean age of 41 years old (SD = 14 years) and the females had a mean age of 39 years old (SD = 14 years). Each participant was invited to individually engage in a short learning task followed by a small set of associated assessment questions whilst being video recorded. All participants completed an informed consent form on their participation and the usage of video recorded material for research purposes. Ethical approval was obtained from the MMU Faculty Academic Ethics Committee.

B. Study Procedure

Prior to the study a short learning topic was selected, which was a factual digital video on Termites with a total duration of 8 minutes 40 seconds. The Termite video was targeted at the general public with no age restriction and covered: functional architectural aspects of the termite mounds, roles within the social structure of a termite colony and locations where termite colonies thrive. Two experts (Academic Professors) on the subject area were recruited to develop ten difficult (hard) questions and ten easy questions related to the video content with expert agreement for the participants to answer. The experts were required to devise five open questions and closed questions within each set of hard and easy questions. At the same time, the experts noted down the correct answer(s) for each question, which was later used as the mark scheme.

The study was conducted at MMU in the same room with the same equipment and layout shown in Fig. 2 to ensure consistency in the quality of the digital video recordings. Each participant watched the termite video and then the interviewer followed a script, asking each participant all of the hard and easy questions in a randomised order. If the participant did not respond to a question then the interviewer was instructed to repeat the question. If there was still no response after repeating the question then the interviewer was instructed to ask a neutral probing questions such as "What is your best guess?". The participant was video-recorded whilst watching the termites video and when answering the questions associated with the video content. The digital camcorder was setup to capture the participant's upper body non-verbal cues in to an MP4 multimedia file for post study analysis and extraction by FATHOM.

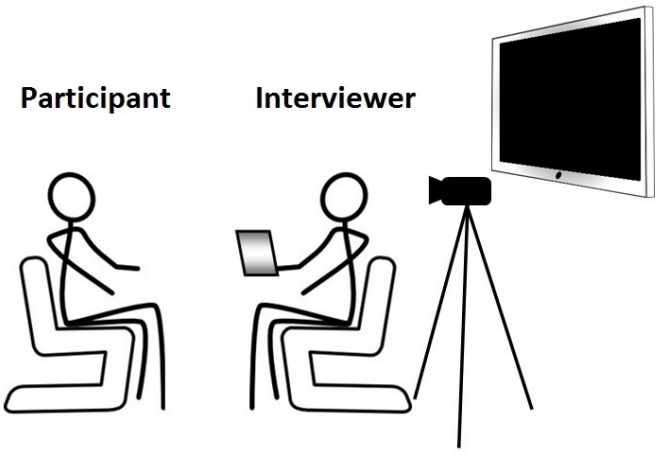


Fig. 2. Study Layout.

To counteract question order effect, the presentation of each subset of open and closed questions were randomised, resulting in four conditions: 1A, 1B, 2A and 2B as shown in Table I. For example, if the participant was randomised in to condition 2A then he/she would receive the easy open questions first, followed by the hard open questions, then the easy closed questions followed by the hard open closed questions. Each participant was randomised in to one of the conditions so that each condition had five males and five females (10 participants per condition). Equally randomising the participants across the conditions in this manner reduces the chance of producing an imbalanced dataset related to gender.

TABLE I. QUESTION ORDER

Condition	Closed Questions		Open Questions	
	Easy	***Hard***	***Easy***	***Hard***
1A	1st	2nd	3rd	4th
1B	2nd	1st	4th	3rd
2A	3rd	4th	1st	2nd
2B	4th	3rd	2nd	1st

C. FATHOM's Comprehension Classifier ANN Training Procedure

After the study, the participant's answers to each of the questions were transcribed from the video recordings and marked using the mark scheme. All forty participant MP4 video files were processed in FATHOM's extraction mode to collate the non-verbal vector-based dataset for cross-validation training of the FATHOM comprehension classifier backpropagation ANN.

Each vector within the dataset represented a 1 second time period and contained the state of all forty non-verbal channels at that point in time, in a normalised format i.e. scaled from +1 to -1. Appended to the end of each vector was a normalised supervisory value, which was used during cross-validation training and testing to determine whether the FATHOM

comprehension classifier backpropagation neural network had correctly classified the input vector. If the supervisory value was +1 then it represented comprehension and if the supervisory value was -1 then it represented non-comprehension.

The extracted dataset was generated from the multichannels of non-verbal behaviours emitted by the participant only when the interviewer was asking the participant each one of the twenty assessment questions. Therefore, if it took 10 seconds for the interviewer to ask the participant an assessment question then successful extraction of all forty non-verbal behaviours would result in ten normalised vectors. Each set of vectors for each question would then be appended with the supervisory value based upon whether the participants answer to that question was marked as correct or incorrect. The purpose was to see if the participants displayed strong non-verbal indicators of comprehension and non-comprehension when the interviewer asked each question.

To ensure that only strongly associated high and low non-verbal indicators of comprehension were extracted from the participant videos for the open questions, the participants marked open answers were all thresholded at 75% e.g. if the participant got ≥ 75% of the answer then the question was marked as correct otherwise the question was marked as incorrect. This threshold was determined by the experts in the field.

D. Results

From the forty participant videos, FATHOM extracted 16,951 comprehension vectors and 23,857 non-comprehension vectors. Therefore, the entire dataset was composed of 40,808 vectors with 41.5% in the comprehension class. Table II provides a breakdown of the percentage of correctly answered questions by all participants.

TABLE II. MARKED QUESTIONS

	Closed Questions		**Open Questions**	
	Easy	***Hard***	***Easy***	***Hard***
Correct (%)	83	68.5	17	6

In this experiment, FATHOM's comprehension classifier ANN was trained and validated in the detection of human comprehension with the extracted non-verbal dataset using the backpropagation algorithm and 10-fold cross-validation. Each of the cross-validation folds was partitioned as follows: 90% training and 10% testing. The topology of the FATHOM comprehension classifier ANN was: forty inputs, a single hidden layer with twenty neurons and an output layer with one neuron (40:20:1). The maximum number of epochs was 10,000 and the learning rate (□) was set at 0.005. The checking epochs parameter was 250, which meant that at every 250th epoch the total classification accuracy would be checked and if it had not improved then the backpropagation training would terminate. On commencement of training, the training set was randomised once and the neural network had the weights initialised in the range of 0±1/*fan-in*, where *fan-in* represents the number of inputs entering the neuron. The latter neural network training parameters were determined from previous exploratory cross-validation sessions. Parameters were not optimised for this exploratory experiment.

The 10-fold cross-validation training phase results are in Table III and the testing phase results are in Table IV. The Root Mean Squared Error (RMSE) is a performance metric used to determine the degree of error by squaring the aggregated difference between the neural network output and the supervisory value for the dataset. The classification accuracies (CA) were calculated as follows:

- **Comprehension CA** is the percentage of comprehension vectors classified correctly.
- **Non-comprehension CA** is the percentage of non-comprehension vectors classified correctly.
- **Total CA** is the percentage of comprehension and non-comprehension vectors classified correctly i.e. Total CA = Comprehension CA + Non-comprehension CA.
- **Total Normalised CA** is calculated as the Total CA / 2.

The best performing FATHOM comprehension classifier ANN was obtained in fold 4 with a testing phase total normalised CA of 79.58% in Table IV. In Table III and Table IV all of the neural networks were able to classify non-comprehension vectors more easily than comprehension vectors, which may be due to the entire dataset being composed of 58.5% non-comprehension class. All of the ANNs during 10-fold cross-validation consistently achieved CAs above 67.5%, thus strongly indicating that comprehension and non-comprehension patterns exist within the dataset.

VI. CONCLUSION

Overall, the cross-validation results for the FATHOM comprehension classifier ANN consistently attained total normalised CAs above 76% in the testing phase, which strongly indicates that comprehension and non-comprehension was detectable from the multichannels of non-verbal behaviour emitted by the male and female participants during the questioning phase of the study on the content of the termite's video. The FATHOM comprehension classifier ANN performed better at classifying non-comprehension than comprehension, which may have been caused by there being more non-comprehension vectors than comprehension. This is also likely to have been down to the design of the task in that the two experts will have their own opinions on what constitutes high and low comprehension questions in the field of termites. Consistency in the quality of the video recordings resulted in a large dataset with no discarding of the participant MP4 multimedia files. The application of FATHOM as a comprehension detection system to the learning task has identified that high and low comprehension states can be detected from learner's facial non-verbal cues, thus satisfying the aim outlined in Section I.

TABLE III. CROSS-VALIDATION TRAINING PHASE RESULTS

Fold	Epoch	RMSE	Total CA	Total Normalised CA	Comprehension CA	Non-comprehension CA
1	750	0.738	81.83	80.85	75.07	86.62
2	1750	0.741	81.72	80.17	71	89.33
3	500	0.75	80.96	79.68	72.12	87.23
4	4750	0.726	82.72	82.38	80.37	84.38
5	3500	0.718	83.09	82.16	76.63	87.69
6	3250	0.721	83.27	82.45	77.62	87.28
7	1750	0.737	82.2	81.77	79.2	84.34
8	2000	0.725	82.55	82.02	78.9	85.14
9	1750	0.73	82.55	81.25	73.53	88.97
10	2750	0.735	82.29	82.04	80.53	83.54
Mean	2275	0.732	82.31	81.47	76.49	86.45

TABLE IV. CROSS-VALIDATION TESTING PHASE RESULTS

Fold	RMSE	Total CA	Total Normalised CA	Comprehension CA	Non-comprehension CA
1	0.772	79.68	78.68	72.8	84.57
2	0.787	78.75	77.13	67.55	86.7
3	0.787	78.08	76.71	68.61	84.82
4	0.771	80.02	79.58	76.99	82.18
5	0.769	80.02	79.15	73.98	84.32
6	0.768	80.34	79.37	73.62	85.12
7	0.788	79.36	78.63	74.27	82.98
8	0.781	79.12	78.51	74.92	82.1
9	0.771	79.93	78.33	68.9	87.76
10	0.802	78.61	78.24	76.06	80.42
Mean	0.779	79.39	78.43	72.77	84.09

VII. FUTURE WORK

Naturally, further application of FATHOM to previously unencountered, larger representative cultural datasets should enhance its ability at comprehension detection and progressively advance towards answering the following hypothesis:

> *Is there a general set of non-verbal behaviours that a backpropagation neural network can use to detect high and low patterns of human comprehension across all cultures, genders and age groups?*

Further work includes pruning the number of inputs to the FATHOM comprehension classifier ANN and performing cross-validation experiments to empirically determine the best neural network topology i.e. optimising the neural network topology with the minimum number of hidden layers and neurons whilst retaining acceptable classification accuracies. Another significant future direction is to perform a non-verbal channel analysis using alternative, artificially intelligent computational models such as Decision Trees [31,32]. Investigating the performance of a range of machine learning models in non-verbal comprehension detection would enable comparisons and identification of suitability.

The real-world applications of the FATHOM comprehension detection system are numerous. For example, use of FATHOM in the academic world would provide educators with a computerised proxy tool for individually assessing a learner's comprehension level from a non-verbal perspective alongside traditional comprehension assessments methods to facilitate more accurate identification of learner comprehension state in near real-time.

ACKNOWLEDGMENT

The authors would like to thank the participants who participated in the study.

REFERENCES

[1] E. Babad, "Teaching and nonverbal behavior in the classroom". In L. J. Saha and A. G. Dworkin, (Eds.) International Handbook of Research on Teachers and Teaching. Boston, Massachusetts: Springer US, pp. 817-827, 2009.

[2] S. Machida, "Teacher accuracy in decoding nonverbal indicants of comprehension and noncomprehension in anglo and mexican-american children". Journal of Educational Psychology, vol. 78(6), pp. 454-464, 1986.

[3] J. M. Webb, E. M. Diana, P. Luft, E. W. Brooks and E. L. Breenan, "Influence of pedagogical expertise and feedback on assessing student comprehension from nonverbal behavior". The Journal of Educational Research, vol. 91(2), pp. 89-97, 1997.

[4] C. J. Patterson, M. J. Cosgrove and R. G. O'Brien, "Nonverbal indicants of comprehension and non-comprehension in children". Developmental Psychology, vol. 16(1), pp. 38-48, 1980.

[5] V. L. Allen and M. L. Atkinson, "Encoding of nonverbal behavior by high-achieving and low-achieving children". Journal of Educational Psychology, vol. 70(3), pp. 298-305, 1978.

[6] T. P. Mottet and V. P. Richmond, "Student nonverbal communication and its influence on teachers and teaching". In J. L. Chesebro and J. C. McCroskey (Eds.), Communication for Teachers. Needham Heights, Massachusetts: Allyn and Bacon, pp. 47-61, 2002.

[7] J. D. Jecker, N. MacCoby and H. S. Breitrose, "Improving accuracy in interpreting non-verbal cues of comprehension", Psychology in the Schools, vol. 2(3), pp. 239-244, 1965.

[8] M. Amelsvoort, B. Joosten, E. Krahmer and E. Postma, "Using non-verbal cues to (automatically) assess children's performance difficulties with arithmetic problems". Computers in Human Behavior, vol. 29(3), pp. 654-66, 2013.

[9] L. S. Pardo, "What every teacher needs to know about comprehension". The Reading Teacher, vol. 53(3), pp. 272-280, 2004.

[10] F. Buckingham, K. Crockett, Z. Bandar, J. O'Shea, K. MacQueen and M. Chen, "Measuring human comprehension from nonverbal behaviour using artificial neural networks". Proceedings of the IEEE World Congress on Computational Intelligence (IEEE WCCI), Australia, pp. 368-375, 2012.

[11] J. Rothwell, Z. Bandar, J. O'Shea and D. McLean, "Silent Talker: a new computer-based system for the analysis of facial cues to deception". Applied Cognitive Psychology, vol. 20, pp. 757-777, 2006.

[12] J. Rothwell, Z. Bandar, J. O'Shea and D. McLean, "Charting the behavioural state of a person using a backpropagation neural network". Neural Computing & Applications, vol. 16, pp. 327-339, 2007.

[13] V. Manusov and A. R. Trees, ""Are you kidding me?": the role of nonverbal cues in the verbal accounting process". Journal of Communication, vol. 52(3), pp. 640-656, 2002.

[14] P. Ekman and W. V. Friesen, "The repertoire of nonverbal behavior categories, origins, usage, and coding". Semiotica, vol. 1, pp. 49-98, 1969.

[15] M. L. Knapp and J. A. Hall, Nonverbal Communication in Human Interaction, 3rd Ed., Fort Worth, Texas: Harcourt Brace, 1992.

[16] A. Mehrabian, "Communication without words". Psychology Today, vol. 2(4), pp. 53-56, 1968.

[17] A. Mehrabian, Silent Messages, 5th Ed., Belmont, California: Wadsworth Publishing Company, 1971.

[18] P. Ekman and V. W. Freisen, The Facial Action Coding System (FACS), Consulting Psychologists Press, Palo Alto, California, US, 1978.

[19] N. Dael, M. Mortillaro and K. R. Scherer, "The body action and posture coding system (BAP): development and reliability". Journal of Nonverbal Behavior, vol. 36(2), pp. 97-121, 2012.

[20] R. C. Johnson, (1999) "Computer program recognizes facial expressions", Electronic Engineering Times, [Online] [Accessed on 17 June 2014] http://www.eetimes.com/at/news/OEG19990405S0017.

[21] A. S. Won, J. N. Bailenson, S. C. Stathatos and W. Dai, "Automatically detected nonverbal behavior predicts creativity in collaborating dyads". Journal of Nonverbal Behavior, pp. 1-20, 2014.

[22] J. M. Montepare, "Nonverbal behavior in the digital age: meanings, models, and methods". Journal of Nonverbal Behavior, pp. 1-3, 2014.

[23] C. Dollaghan and N. Kaston, "A comprehension monitoring program for language impaired children". Journal of Speech and Hearing Disorders, vol. 51, pp. 264-271, 1986.

[24] K. Pezdek, "Comprehension: it's even more complex than we thought". Advances in Psychology, vol. 39, pp. 215-236, 1986.

[25] E. Skarakis-Doyle, N. MacLellan and K. Mullin, "Nonverbal indicants of comprehension monitoring in language-disordered children". Journal of Speech and Hearing Disorders, vol. 55(3), pp. 461-467, 1990.

[26] G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information". Psychological Review, vol. 63, pp. 81-97, 1956.

[27] D. E. Rumelhart, G. E., Hinton and R. J. Williams, "Learning internal representations by error propagation". In D. E. Rumelhart and J. L. McClelland, Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations. Cambridge, Massachusetts: MIT Press, 1986.

[28] D. E. Rumelhart, G. E., Hinton and R. J. Williams, "Learning representations by back-propagating errors". Nature, vol. 323, pp. 533-536, 1986.

[29] M. H. Hassoun, Fundamentals of neural networks. Cambridge, Massachusetts: MIT Press, 1995.

[30] M. Stone, "Cross-validatory choice and assessment of statistical predictions". Journal of the Royal Statistical Society, vol. 36(2), pp. 111-147, 1974.

[31] J. R. Quinlan, "Induction of decision trees". Machine Learning, vol. 1, pp. 81-106, 1986.

[32] J. R. Quinlan, C4.5: programs for machine learning. San Mateo, California: Morgan Kaufmann Publishers Inc, 1993.

FATHOMing Out Interdisciplinary Research Transfer

Keeley A. Crockett, James D. O'Shea, Fiona J. Buckingham Zuhair A. Bandar
The Intelligence Systems Group, School of Computing, Mathematics and Digital Technology
Manchester Metropolitan University
Chester Street, Manchester, M1 5GD, UK
K.Crockett@mmu.ac.uk

Kathleen. M. MacQueen, Mario Chen, Kelly Simpson
FHI 360
Durham, NC 27713, USA
KMacQueen@fhi360.org

Abstract— **This paper presents a case study on the development and deployment of a computerised, non-invasive psychological profiling system which detects human comprehension through the monitoring of multiple channels of facial nonverbal behaviour using Artificial Neural Networks (ANN). Prior work on an earlier system known as Silent Talker, led to collaborations and a funded project with Family Health International 360 (FHI 360) in collaboration with the National Institute of Medical Research (NIMR), to produce the FATHOM system for measuring comprehension of the informed consent process amongst women in Tanzania. This paper describes the process of taking a working research prototype and deploying the system in a real working environment. The paper discusses the process of contract negotiation, global ethics and intellectual property rights. The FATHOM system and initial results are described.**

Keywords- artificial neural networks, FATHOM, human comprehension, nonverbal behaviour, microgesture, microexpression, Silent Talker

I. INTRODUCTION

HIV prevention clinical trials are often executed in developing countries e.g. in Africa where the prevalence of HIV/AIDS is high. Within trials the informed consent process is a legal and ethical requirement, each participant must voluntarily make a truly informed decision on whether to participate in the study. Informed consent requires voluntary consent from comprehension of adequately delivered information about the purpose of the study, any procedures involved and the effects of participation [1] before commencement. The participant's comprehension of the actual informed consent process is a precarious area of concern. Guidelines such as the Nuremburg Code [2] and the Declaration of Helsinki [3] exist to provide guidelines that the research conducted is ethical. However previous studies have shown that participants have problems in understanding informed consent documentation [4-6].

Silent Talker is a patented modular, real time Artificial Neural Network (ANN) based system [7-9] that automatically collects and analyses nonverbal cues to classify an overall behavioral or psychological state. Non-verbal behavior can be defined as the signs and signals (visual, audio, tactile and chemical) used by humans to express themselves. Silent Talker is a lie detector which captures and examines relationships between 24 channels of non-verbal facial behaviors [9]. Banks of ANN's learn the relationships between the channels that indicate deception and truth. For deception, the relationships may indicate some 'discrepancy' or 'incongruence' between the channels. Due to the modularity of the system it can readily be adapted to handle various cues, detect a variety of different states and be tuned to particular situations, environments and applications. Other channels can be added when necessary depending on the problem domain.

FHI 360 is a non-profit organization focused on the implementation of research and programs worldwide. HIV and prevention is one of their research interests. Collaboration between members of the Intelligent Systems Group at Manchester Metropolitan University (MMU) with FHI 360 led to a funded project which aimed to generate critical data for enhancing comprehension by looking at both verbal and non-verbal communication during informed consent for clinical research in sexual and reproductive health, with an emphasis on HIV [10]. The verbal communication element was assessed through the use of qualitative methods in order to develop an elicitation tool that can be used to create culturally and linguistically valid verbal lexicons of key research-related terms and concepts. This element was carried out by FHI 360. The non-verbal element was to be assessed through use of quantitative and qualitative methods to develop culturally valid non-verbal communication patterns of high and low comprehension. This element was to be undertaken by the Intelligent Systems Group at Manchester Metropolitan University (MMU) using Silent Talker as a starting point.

The outcome of the project was the development of FATHOM, a system developed specifically to measure levels of comprehension during an informed consent assessment process carried out in North-western Tanzania, Africa using a setting similar to what would be used for an HIV/AIDS prevention randomized study. Initial results [11] have shown that Artificial Neural Networks (ANN) are able to detect nonverbal comprehension patterns from non-verbal behaviour.

The research is funded through the USAID Cooperative Agreement with FHI 360 for PTA, No. GHO-A-00-09-00016-00. The findings presented do not necessarily reflect FHI 360 or USAID policies.

978-1-4673-5851-4/13/$31.00 ©2013 IEEE

The aim of this paper is to describe the process of taking a working research prototype (Silent Talker) and adapting the system to meet the needs of another domain (FATHOM) identified by industry. In deploying the system in a real working environment there are several important concerns. First, the study itself must be subject to the appropriate ethical procedures. One example to highlight the importance of the interaction between ethical procedures and comprehension is the potential for participants to mistakenly believe they are receiving a treatment for or prevention of HIV. Ethical procedures in trials such as this are complicated by the need to satisfy the national requirements of the countries involved and the participating bodies. The second major concern is the generation of IPR. In this case there was patented upstream IPR from the development of Silent Talker system. However, new IPR could be generated for the new application. Both MMU and FHI 360 would contribute to the newly generated IP and have a financial stake in its exploitation. Therefore this required contract negotiation for the project to proceed.

This paper is organized as follows: Section II provides an overview of the research background. Section III discusses the collaboration with FHI 360 to formulate and adopt a research prototype for use within the field and considers international ethics and IPR. Section IV describes the field study methodology and section V provides an overview of the FATHOM system, methodology and experimental results. Sections VI and VII presents observations from the perspective of the academic team involved with the project and draws conclusions for future projects.

II. Background

A. Non Verbal Behaviour in Detecting Comprehension

Measurement of human comprehension has principally been focused on verbal and written responses [12]. Non-verbal behavioral (NVB) data represents an important untapped source of information about the strengths and weaknesses of different assessment approaches for measuring informed consent comprehension. For example rich information about a person's physiological state can be captured from facial expressions. A person's NVB operates under less conscious control than verbal communication [13], yet forms approximately 93% of messages exhibited from a human [14]. Therefore it could provide a more accurate, reliable measure of human state than verbal responses. Indeed behavior-coding, has long been used in survey research as a way to systematically analyze interactions between interviewers and respondents and identify problematic questions [11]. Such coding may provide insights into the informed consent process by identifying whether patterns of comprehension and non-comprehension exist within NVB alone.

One of the challenges associated with NVB is the encoding and decoding of NVB data channels. This is a well-established idea dating back to the work of David Efron [15] in which human experts viewed film recordings one frame at a time to log a particular channel of non-verbal behavior, for example a specific hand movement. This incurred problems of labour-intensive analysis and subjective interpretation by human judges. According to one estimate it takes over 100 hours to train a human to make reliable judgments using the frame-by-frame-system and 2 hours to analyse 1 minute of video [15].

The work was continued by Paul Ekman in the Facial Action Coding System [15] which currently focuses using micro-expressions (short-lived, unexpected facial expressions such as disgust) to detect lying. This method supports automatic detections of the micro expressions, speeding up the process. The FACS set, derived from 6 gross basic expressions recognized by Darwin: Anger, Disgust, Fear, Happiness, Sadness, and Surprise, was challenged by contemporaries [16], who failed to find evidence of discrete facial expressions associated with the specific emotions. Ekman's supporting work on "Wizards" (human super lie detectors) was also challenged on grounds of its methodology [17]. A recent example uses less specific modeling with a larger range of features, but also takes a more "agnostic" view of classification passing it to higher order classifiers [18]. This work effectively follows the Silent Talker philosophy, which operates at a very fine-grained level known as micro-gestures (for example left eye changes from fully-open to half-open). Silent Talker pioneered the use of supervised learning, i.e. the system is presented with examples of micro-gesture data over the time period in which a lie (or truth telling) takes place, with the correct classification, and learns to generalize from these examples to classify previously unseen data. So Silent Talker makes no *a priori* assumptions about which features or combinations thereof will be required for effective classifiers.

An additional problem is the extensibility of micro-expressions to other domains. Lie detection has been the main thrust of Ekman's work, with some extension to the concealment of emotional responses. There is no evidence that the gross facial expressions provide features to detect comprehension/non-comprehension. We postulate that with its ability to discover combinations of non-verbal micro-gesture features, a system derived from Silent Talker is more likely to succeed in this domain and to be robust to confounding features that could disrupt other approaches to detecting comprehension in clinical trials.

B. Comprehension in the Informed Consent Process

Non-comprehension is defined as as "a state of knowledge that ranges from uncertainty to complete lack of understanding of the materials under discussion" [19]. Informed consent requires experimental participants to have a clear comprehension of the information delivered in the informed consent process so that they can make a valid decision whether or not to take part. Consequently, they must receive sufficient information on the procedures they will follow, the purpose of the research and the use its findings will be put to, any potential benefits of participating and the potential risks they will run [20]. Should participants fail to understand fully prior to consent, investigators may be liable for any consequences. The natural reaction may be for investigators to produce lengthy, detailed and legalistic documentation with the implication that participants have read and understood it prior to giving consent.

In fact a comparison of comprehension assessment methods: self-reports, checklists, vignettes and narratives have revealed the danger of information overloading [21]. Variations in the measurement of understanding [22] were found even after improving an informed consent document through using a working group to enhance the documentation.

Strategies included enhancing the lexico syntactic readability through sentence length reduction and shorter synonyms but neither technique significantly enhanced participant's objective comprehension score. This emphasizes the need for an improvement in comprehension assessment techniques. If it were possible to detect low comprehension of the information about the trial at an early stage it would be possible to adapt the briefing stage to correct this and comply effectively with legal and ethical implications. Prior work on the detection of comprehension has concentrated on verbal and written responses [23]. This approach can be confounded by human factors such as the potential participant's desire to please, fear of looking stupid or perception that there is something to be gained by participating. Using NVB offers the possibility to detect non-comprehension immediately, as the material is delivered and the potential to deliver a more accurate measure of human comprehension in comparison to verbal communication (as it is impossible to consciously control NVB over such a large number of channels). Previous informed consent studies have been limited to predicting comprehension from simple attributes such as age [24], educational level [22], race and ethnicity [25]. The FATHOM project addresses the research question of whether the complex interactions between multiple NVB channels over time that form the basis of the Silent Talker lie detector are transferrable to provide a better predictor of comprehension vs. non-comprehension. Therefore, there is the unexplored need to identify whether more complex, general trends are present within larger experimental groupings of multiple NVB channels including independent predictors. Development of a such a computer-based comprehension measurement system in near real-time would overcome human weaknesses and help reduce the complexities of measuring understanding.

III. The FHI 360 Partnership

Silent Talker technology was brought to the attention of FHI 360 through international publications and through an Associate Member of the Intelligent Systems Group (ISG), Dr Keith Ashcroft who is a leading UK chartered psychologist. Through early conversations it was evident that the technology behind ST could be applied to look for other patterns of non-verbal behavior associated with different domains, such as comprehension. During early discussions between FHI 360 and ISG, Dr Ashcroft played a significant role in bridging the knowledge gap between computational intelligence techniques and a health and human development organization. This was a good and effective learning experience in presenting research in general layman's terms to a majority non-computer science audience.

After initial contact, members of FHI 360 visited the ISG from the US for a demonstration of the existing ST software and to discuss the approach to handling data. Historically, ANN researchers tend to be interested in the Classification Accuracy (CA) interpreted in the light of a baseline measure. The baseline can be set as the CA achieved by classifying all cases as belonging to the majority class or by finding the CA of a OneR classifier (a well-known form of single-node decision tree). There is also an interest in the confidence one can have for individual classifications. This is usually assessed by the distance between the output from the ANN output layer and the decision boundary (i.e. in a unipolar network with 0 representing comprehension and 1 representing non-comprehension, how far is a particular rating from 0.5). That is not to say that ANN researchers are not interested in the statistical significance of results, but when classifiers are producing high levels of CA, it may require an infeasibly large sample to prove that a small improvement is statistically significant.

FHI 360 requirements were much more firmly embedded in the tradition of clinical trials with their focus on participant associated outcomes. Therefore it was decided that all of the development and testing of the FATHOM system would be conducted by the ISG, after which a statistical evaluation of the results would be carried out by FHI 360 statisticians. Four tools for assessing comprehension would be used: non-verbal markers of informed consent comprehension (IC-NV), close-ended (IC-C), open-ended (IC-O) and self-perception (IC-SP) assessments of comprehension [10]. The first stage of analysis would be a background analysis of the characteristics of the population using elementary data analysis measures of frequencies and percentages for discrete variables, and means, standard deviations, minima, maxima, and median for continuous variables. Additionally, the relationship between demographic variables, including gender, and comprehension (high/low) for the IC-NV, IC-C, IC-O, IC-SP, and willingness to enroll in a hypothetical microbicide trial would be assessed by chi-square analyses for each point of, component of, and global comprehension. Comparisons between the different tools for assessing comprehension would be performed using the kappa statistic with 95% confidence intervals and standard two-sided independent samples t-test with a significance level of 5% [10]. Therefore the outcome of the discussions on analysis was that each group (ISG and FHI 360) would use its own choice techniques appropriate to the stage of the work and that there would be independence between development and assessment.

A succession of meetings followed in order to define the project and develop a proposal bid for FHI 360 to secure funding. Members of the ISG worked in partnership on the non-verbal component of this bid in two key areas. The first area was in describing how ST technology could be adapted to measure comprehension within the informed consent process. Secondly, in the costings, in terms of hiring a person with the required skill set to be employed within the University to complete the project. Funding was successfully secured which enabled a detailed protocol for the study to be developed and a contract to be drawn up. FHI 360 produced a contract of work which was scrutinized by the University's legal team, Research and Knowledge Exchange Office and finally signed off by the

University's Finance Director. This process took internally eight weeks to complete.

In any external commercial project involving collaborations between academics, Universities, as stakeholders will also want to be involved with establishing Intellectual Property Rights (IPR). IPR needs to be established not only on the agreed product development, but on who owns the rights to future research development ideas carried out on the product and/or its supporting research by the academic. Within this project the IPR for ST was owned and retained by MMU, but it was agreed that any new IPR would be shared with FHI 360 and a letter of intent was exchanged.

Specific details of the project are described in the protocol in the following section.

A. Establishing the Protocol

The overall aim of the project was to "develop tools and processes for creating culturally and linguistically valid verbal and non-verbal communication to enhance comprehension during informed consent for sexual and reproductive health clinical trials"[10]. The primary objectives were [10]:

1. To develop an elicitation tool for developing culturally and linguistically valid verbal lexicons of key research-related terms and concepts and to assess the tool's usability among Kiswahili-speaking researchers.
2. Train an Artificial Neural Network system to identify high and low comprehension at a high level of reliability, where reliability is defined as classification accuracy.
3. Complete a pilot evaluation of the relationship between non-verbal markers of informed consent comprehension and closed-ended versus open-ended assessments of comprehension.

The role of the ISG was to lead objective two to develop a new ANN system for classifying non-verbal behavior as indicative of high and low comprehension. Objective 3 would require collaborative work in order to develop an evaluation strategy. The system was to be developed from videos of developmental study participants responding to 10 ordered questions for two task scenarios. Each participant would be asked to respond to 20 questions, resulting in a video data base of up to 80 people x 20 questions = 1600 individual responses for analysis. Two types of ANNs would be developed for the new system. Object locator ANNs were developed from face, torso, and non-face/non-torso images cropped from the developmental study video data. Examples of potential objects to be located include left eye, right eye brow, nose, face, shoulder, and forehead. The cropped images were then standardized using well-established techniques to create a vector that forms the input data for the ANN [9]. Pattern detector ANNs were developed to detect changes in the state of objects, e.g. whether an eye is open or fully closed [9].

B. Ethics

In order to undertake this project, ethical approval was required from the Medical Research Coordinating Committee, National Institute for Medical Research (Tanzania), Protection of Human Subjects Committee, FHI 360, Durham NC, USA, and from Manchester Metropolitan University, UK. FHI 360 was responsible for obtaining all ethics approvals and provided a PHSC Expedited Review Approval Notice and Clearance Certificate for Conducting Medical Research in Tanzania to MMU. Following international ethics approval, ethical approval still had to be obtained by the University Faculty Academic Ethics Committee which took approximately 2 months. This process was expedited by the prior work and ongoing support of FHI 360. The field study was undertaken over a five month period. Each participant was compensated for a 4-5 hour time commitment of which approximately one hour was taken up by the actual task which was filmed.

IV. Field Study Methodology

The field study was executed in the Mwanza region of Tanzania in Africa by FHI 360 [26] in collaboration with the National Institute for Medical Research (NIMR) [27]. NIMR recruited sexually active women aged 18-35, inclusive, who are native Kiswahili-speakers and potentially interested in participating in future HIV prevention trials. In total, 292 participants took part in the field study. This section will briefly describe the non-verbal component of the study design.

A. Study Design

of In order for the ANN's to detect the visible non-verbal cues for high and low comprehension a series of brief baseline interviews were conducted to generate a set of training data. For this interview, each participant took part in two brief learning tasks. Task A consisted of a short presentation on a general and assumed to be familiar HIV prevention topic that includes one piece of novel information generally not known by the target population, (e.g.: the use of male condoms to prevent HIV transmission [presumed to be known] and the possibility of a vaginal ring impregnated with antiretroviral to protect women [presumed to be novel]) [10]. After the presentation, the interviewer asked the participant 7 questions, intended to be easy, about the topic presented and 3 technical questions about HIV that were not discussed in the presentation intended to be difficult. Task B will center on a short presentation on a highly specialized and assumed to be unfamiliar HIV prevention topic. After the presentation, the interviewer will ask the participant 3 questions about HIV generally, intended to be easy, and 7 technical questions about this topic in particular, intended to be difficult. Half of the study participants will take part in Task A followed by Task B, and half will take part in Task B followed by Task A. This strategy allowed analysis of whether the knowledge of the nature and order of the questions affected the non-verbal behavior pattern and whether low-comprehension behaviors associated with Task B may "leak" into Task A, e.g., if the psychological effects of confusion persist [10]. Such a strategy is necessary in the ANN's training and testing processes to improve the accuracy of the system for classifying behavior as indicative of high or low comprehension.

V. FATHOM

The FATHOM system builds on the strategies used within Silent Talker to measure human comprehension from NVB in an informed consent field study. FATHOM has been

specifically designed and engineered as a detector of human comprehension and non comprehension through analysing multichannels of NVB using Artificial Neural Networks (ANN). Fig. 1 shows an overview of the nonverbal channels, which FATHOM monitors in order to gauge whether a person comprehends. Each non-verbal behavior channel is operationalized here as numeric vector data that represents the location and pattern of an object. Channels are classified into themes; themes contain one or more channel values. For example, the channel theme eye closure may include specific values for (1) left eye closed, (2) left eye blink, (3) right eye closed and (4) right eye blink.

Structurally, the fundamental components of FATHOM's architecture are the: object locator ANNs, pattern detector ANNs and a classifier ANN. The purpose of the object locator neural networks is to determine the presence of nonverbal features such as the eye so that the pattern detector neural networks can then identify the state of the object found e.g. the right eye is gazing to the left. Once all of the monitored NVB channels have been coded and collated together as an entire response then it is delivered as input to the classifier ANN, which assesses the degree of comprehension or non comprehension found within the response.

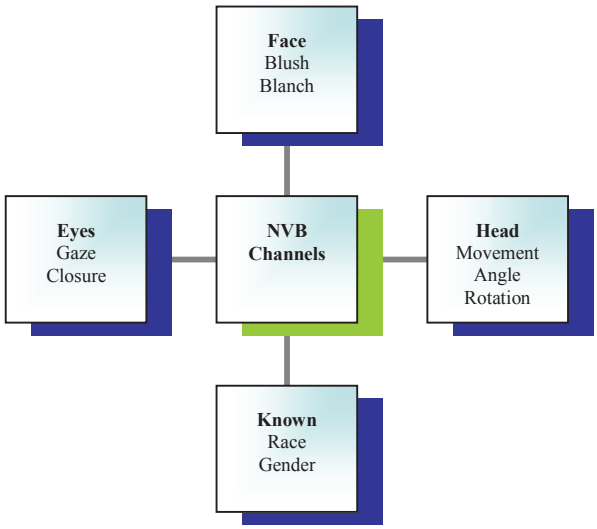


Fig. 1 Nonverbal channels monitored by FATHOM.

The subsequent subsection describes how FATHOM was applied to an informed consent field study and how the BP ANN comprehension classifier was trained to distinguish whether or not a human comprehends.

A. Experimental Methodology

Eighty Tanzanian females were recruited to participate in the interview-based informed consent field study. As described in section IV, the interview was composed of a low comprehension task (Task A) and a high comprehension task (Task B), which were randomised to prevent order effects. A digital camcorder was used throughout the entire interview to capture the participant's NVBs exhibited by their upper body. The digital recordings of the participant interviews were then used to generate data sets for training and validating FATHOM's feed forward Back propagation (BP) ANN classifier at comprehension and non comprehension detection in a set of experiments.

The data set used for training and testing the BP ANN consisted of a set of input vector patterns from sections of the digital participant recordings under experimental analysis. Every input value within each vector was normalised i.e. between +1.0 and -1.0 and represented the state of a NVB channel e.g. right eye closed coded as +1.0 and right eye open coded as -1.0. Associated with each vector was a supervisory value, which was used to assess whether the BP ANN correctly classified the presented input vector pattern during training. The normalised supervisory value was coded as +1.0 for comprehension and -1.0 for non comprehension.

The normalised vectors were divided randomly in to training, validation and testing data sets of a predetermined size. To ensure that the BP ANN did not learn patterns on an order basis the presentation of the vectors in the training, validation and testing data sets was randomised. The training data set was used to update the BP ANN weights incrementally based upon the calculated error using the Delta (δ) rule [28..30]. The validation data set determined when to terminate training with the trained BP ANN weights frozen. The testing data set assessed the BP ANN's ability at classifying previously unencountered vectors.

n-fold cross-validation [28] was also utilized in the training process. The data set is randomly divided into *n* equally sized segments. A single segment is put aside for testing purposes. The remaining segments are used for training and validation. Only when all of the *n* segments have completed the role as the test set once then the cross-validation training terminates.

To evaluate the performance of the BP ANN at classifying vectors, the CA was calculated; a percentage representing the average quantity of vectors correctly classified out of the entire data set presented. Comprehension CA and non comprehension CA were calculated individually and then combined together to identify the overall CA.

B. Experimental Results

Out of the eighty digital recordings containing each participant's performance at learning Tasks A and B, only thirty videos were usable. The unusable videos were discarded from the data set for two reasons. Firstly, fieldworkers failed on some occasions to exactly follow the protocol which was established on the filming of participants. Secondly, some participant's exhibited extreme head tilting which the ANNs had not been trained on. However, if in the future the ANNs were trained on this feature more information will be available for the classifier ANN in order to make a decision on the level of comprehension. Overall, the analysis data set contained 600 participant answers to the questions from the learning tasks for experimental analysis.

In the following experiments, FATHOM's BP ANN was trained and validated in the detection of human comprehension through monitoring the nonverbal channels outlined in Fig. 1. The topology of the BP ANN throughout each experiment was 40:30:1, which was determined as the best performing BP

ANN from previous experimental training sessions. 10-fold cross-validation was adopted with the data set partitioned as: 60% training, 30% validation and 10% testing.

C. Experiment 1

The first experiment was focused upon extracting and analysing the NVB channels displayed by all of the participants during the script reading of the short learning topics for both Task A and B from the digital recordings. The extracted data set containing 241,945 vectors was used to train and test the BP ANN. Task A's script contained 45% comprehension vectors and Task B's script contained 55% non comprehension vectors, thus resulting in a nearly balanced data set. Table I summarises the breakdown of CA results across each phase of the 10-fold cross-validation, which was repeated six times. Overall, the results strongly indicate that comprehension and non comprehension patterns reside within the data set because across the training, validation and testing phases, CA's >87% were consistently attained by the BP ANN.

TABLE I. EXPERIMENT 1: CROSS-VALIDATION CA'S

	Training	Validation	Testing
Comprehension	89.82%	88.32%	88.60%
Non comprehension	89.29%	88.06%	88.20%
Overall	89.56%	88.19%	88.40%

Later examination of the participants marked answers to the open-ended questions for both tasks revealed that the data set was noisy: Task A was not pointing towards being entirely comprehension responses and Task B was not pointing towards being entirely non comprehension responses. Therefore, the following experiment took a new approach in order to overcome the latter disadvantage.

D. Experiment 2

The second experiment still focused upon extracting and analysing the NVB channels displayed by all of the participants during the script reading of the short learning topics for both Task A and B from the digital recordings. But each sentence within each learning task script was mapped with a supervisory value determined by the participant's marked response to the associated open-ended question. For example, an incorrectly answered open-ended question resulted in the vectors encompassing the sentence from the script being assigned a supervisory value of -1.0, denoting non comprehension. On the other hand, a correctly answered open-ended question resulted in the vectors encompassing the sentence from the script being assigned a supervisory value of +1.0, denoting comprehension. The closed questions were susceptible to guessing; therefore they were not used in the script sentence mappings. The extracted data set containing 71,787 vectors (63.5% comprehension and 36.5% non comprehension) was used to train and test the BP ANN. Table II summarises the breakdown of CA results across each phase of the 10-fold cross-validation, which was repeated six times. Again, the results strongly indicate that comprehension and non comprehension patterns exist within the data set because the BP ANN consistently obtained CA's >84%.

TABLE II. EXPERIMENT 2: CROSS-VALIDATION CA'S

	Training	Validation	Testing
Comprehension	89.50%	86.19%	86.66%
Non comprehension	89.20%	83.70%	84.37%
Overall	89.35%	84.95%	85.52%

Discussion

Overall, the results from both experiments show that the BP ANN was repeatedly able to detect patterns of comprehension and non comprehension within the data set's multichannels of NVB by consistently achieving CA's above 84% throughout multiple iterations of 10-fold cross-validation with randomised network weights and vector presentation.

Although FATHOM adapted well to classifying the previously unencountered cultural data set, retraining FATHOM's object locator and pattern detector neural networks for each nonverbally monitored facial channel in Fig. 1 should help to improve the CA. Ensuring consistent digital recordings throughout the field study would have lead to a lower discard rate, which could have enhanced the BP ANNs generalisation and increased reliability.

Completion of the BP ANN training experiments has led to the identification of components of the field study which could be enhanced for application in future work to improve the quality of the data set and the reliability of the results in experimentation. To ensure that a more balanced and clearly defined data set is obtained, the associated complexity of the learning task scripts and questions should be made more difficult/easy to ensure that the distinct extremities of comprehension and non comprehension are collated, thus avoiding a noisy data set. Techniques such as implementing a delay after listening to learning script topic and receiving the associated questions may help to reduce false comprehension e.g. participant responding with memorised keywords from the script and improve the quality of the data set. Adopting techniques that have been found to hinder comprehension such as: rushing points, using jargon and overloading with complex information [29] would help to increase the size of the non-comprehension data set.

VI. OBSERVATIONS

Once FHI 360 had made the decision to invest in the development of FATHOM, one of the key problems was the time it took to agree contracts, obtain ethical approvals, advertise for a person to undertake the programming work, and employ them etc. The amount of development work was underestimated and the project overran to the extent that second phase sample videos, following modification of the informed consent process have to date not yet been run through the system. A key lesson learnt was in the use of the statistical measures of significance that were applied. It was observed

during this project that measures that were acceptable in the ANN research community were not always suitable for clinical trials. The result was that there would be independence between development carried out by the ISG and assessment of results undertaken by FHI 360.

VII. Conclusion

This paper describes how a University research group and an organization (FHI 360) can collaborate on the development and deployment of a computerised, non-invasive psychological profiling system which detects human comprehension through nonverbal behavior. This was achieved by transforming and adapting a working prototype Lie Detection system to meet the requirements of the client, FHI 360. The contribution of the project is twofold. First, the product, FATHOM, has achieved average classification accuracies greater than 83% using facial NVB micro gestures.

More importantly, from the current perspective, it has identified the challenges of collaboration between two different kinds of organization operating in different disciplines who wish to apply academic research in the real world. Consequently it provides a stepping stone to further collaborations and projects to complete commercialization of this system to provide more understanding of participation in clinical trials, particularly for HIV prevention and reduction of its possible transmission.

Acknowledgment

We wish to thank the women in Tanzania who participated in this research and especially for their willingness to be video-recorded. We also wish to thank our collaborators at NIMR in Tanzania, including Soori Nnko, Bahati Andrew, Tusa Erio, Catherine Bunga, and Gerald Lumanyika.

References

[1] UNAIDS, 2010, Global report: UNAIDS report on the global AIDS epidemic 2010 [Online] Available: http://www.unaids.org/globalreport/documents/20101123_GlobalReport_full_en.pdf

[2] Nuremberg Code. In: Trials of war criminals before the Nuremberg military. Tribunals under control counsel law, vol. 11. Washington, DC: U.S. Government Printing Office, No. 10, 1949. p. 181–2.

[3] World Medical Association (WMA). Declaration of Helsinki: ethical principles for medical research involving human subjects, as amended by the 52nd WMA General Assembly, Edinburgh, Scotland; 2000.

[4] National Commission for the Protection of Human Subjects of Biomedical, Behavioral Research. The Belmont report: ethical principles and guidelines for the protection of human subjects of research. Washington, DC: U.S. Government Printing Office; 1979.

[5] D. W. Fitzgerald, C. Marotte, R. I. Verdier, W. D. Jr. Johnson and J. W. Paper, "Comprehension During Informed Consent in a Less-developed Country", Lancet, vol. 360, pp. 1301-1302, 2002.

[6] J. Flory and E. Emanuel, "Interventions to Improve Research Participants' Understanding in Informed Consent for Research: A Systematic Review", JAMA, vol. 292(13), pp. 1593-1601, 2004.

[7] Z. Bandar, D. A. McLean, J. D. O'Shea and J. A. Rothwell, International Patent Number WO02087443. Geneva, Switzerland: World Intellectual Property Organization, 2002.

[8] J. Rothwell, Z. Bandar, J. O'Shea and D. McLean, "Charting the Behavioural State of a Person Using a Backpropagation Neural Network", Neural Computing & Applications, vol. 16, pp. 327-339, 2007.

[9] J. Rothwell, Z. Bandar, J. O'Shea and D. McLean, "Silent Talker: A New Computer-based System for the Analysis of Facial Cues to Deception", Applied Cognitive Psychology, vol. 20, pp. 757-777, 2006.

[10] Simpson, K. McQueen, K. Mack, N. Friedland, B. Nnko, S. Family Health International Protocol. Enhancing Local Verbal and Non-verbal Communication for Informed Consent Processes in Tanzania, Study # 10159.

[11] F. Buckingham, K. Crockett, Z. Bandar, J. O'Shea, K. MacQueen and M. Chen, Measuring Human Comprehension from Nonverbal Behaviour using Artificial Neural Networks, Proceedings, WCCI 2012 IEEE World Congress on Computational Intelligence Australia, pp. 368-375, 2012.

[12] G. Lindegger, C. Milford, C. Slack, M. Quayle, X. Xaba and E. Vardas, "Beyond the Checklist: Assessing for Understanding for HIV Vaccine Trial Participation in South Africa", Acquired Immune Deficiency Syndrome, vol. 43(5), pp. 560-566, 2006.

[13] E. Babad, "Teaching and Nonverbal Behavior in the Classroom" In L. J. Saha, and A. G. Dworkin (Eds.) International Handbook of Research on Teachers and Teaching, Boston, Massachusetts: Springer US, pp. 817-827, 2009.

[14] T. P. Mottet and V. P. Richmond, "Student Nonverbal Communication and its Influence on Teachers and Teaching" In J. L. Chesebro & J. C. McCroskey (Eds.), Communication for Teachers, Needham Heights, Massachusetts: Allyn and Bacon, pp. 47-61, 2002.

[15] P. Ekman and V. W. Freisen, The Facial Action Coding System (FACS), Consulting Psychologists Press, Palo Alto, CA, US, 1978.

[16] James A. Russell, Jo-Anne Bachorowski, Jose-Miguel Fernandez-Dols, Facial and Vocal Expressions of Emotion, Annu. Rev. Psychol. 2003. 54:329–49, doi: 10.1146/annurev.psych.54.101601.145102

[17] Charles F. Bond Jr., Ahmet Uysal, On Lie Detection "Wizards", Law and Human Behavior Volume 31, Issue 1, pp 109-115 , February 2007.

[18] M. Pantic, I. Patras, M. F. Valstar, Learning spatiotemporal models of facial expressions, Proceedings Int'l Conf. Measuring Behaviour (MB'05). Wageningen, The Netherlands, pp. 7 - 10, 2005.

[19] Z. Zeng, M. Pantic, G. I. Roisman and T. H. Huang, "A survey of affect recognition methods: audio, visual and spontaneous expressions" IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31(1), pp. 39–58, 2009.

[20] H. Z. Waring, "Expressing noncomprehension in a US graduate seminar", Journal of Pragmatics, vol. 34(12), pp. 1711-1731, 2002.

[21] V. Manusov and A. R. Trees, "Are You Kidding Me?": The Role of Nonverbal Cues in the Verbal Accounting Process", Journal of Communication, vol. 52(3), pp. 640-656, 2002.

[22] M. H. Hassoun, Fundamentals of Artificial Neural Networks, London: MIT Press, 1995.

[23] B. Widrow, D. E. Rumelhart and M. A. Lehr, "Neural networks: applications in industry", Business and Science. Commun. ACM, vol. 37(3), pp. 93–105, 1994.

[24] M. Hochhauser, 2004, Informed Consent: Reading and Understanding Are Not the Same. Applied Clinical Trials [Online] Available: http://appliedclinicaltrialsonline.findpharma.com/appliedclinicaltrials/article/articleDetail. jsp?id=90594.

[25] E. Skarakis-Doyle, N. MacLellan and K. Mullin, "Nonverbal Indicants of Comprehension Monitoring in Language-disordered Children", Journal of Speech and Hearing Disorders, vol. 55(3), pp. 461-467, 1990.

[26] Family Health International 360, 2011, Who We Are. [Online] Available: http://www.fhi.org/en/AboutFHI/index.htm.

[27] National Institute for Medical Research Tanzania, 2012, About Us [Online] Available: http://www.nimr.or.tz/.

[28] M. Stone, "Cross-validatory Choice and Assessment of Statistical Predictions", Journal of the Royal Statistical Society, vol. 36(2), pp. 111-147, 1974.

[29] Mclean, D. Bandar, Z. O'Shea, J. Crockett, K.A. , "Commercialisation of an artificially intelligent deception detection system in the current security climate," 2010 IEEE WCCI, Barcelona, pp.1-6, 2010.

WCCI 2012 IEEE World Congress on Computational Intelligence
June, 10-15, 2012 - Brisbane, Australia IJCNN

Measuring Human Comprehension from Nonverbal Behaviour using Artificial Neural Networks

Fiona J. Buckingham, Keeley A. Crockett, Zuhair A. Bandar, James D. O'Shea
The Intelligence Systems Group, School of Computing, Mathematics and Digital Technology
Manchester Metropolitan University
Chester Street, Manchester, M1 5GD, UK
fiona.j.buckingham@stu.mmu.ac.uk

Kathleen. M. MacQueen, Mario Chen
Family Health International 360
Durham, NC 27713, USA
KMacQueen@fhi360.org

Abstract— **This paper presents the adaptation and application of Silent Talker, a psychological profiling system in the measurement of human comprehension through the monitoring of multiple channels of facial nonverbal behaviour using Artificial Neural Networks (ANN). Everyday human interactions are abundant with almost unconscious nonverbal behaviours, providing a potentially rich source of information once decoded. Existing comprehension assessments techniques are inhibited by inconsistencies, limited to the verbal communication dimension and are often time-consuming with feedback delay. Major weaknesses hinder humans as accurate decoders of nonverbal behaviour with being error prone, inconsistent and poor at simultaneously focusing on multiple channels. Furthermore, human decoders are susceptible to fatigue and require training resulting in a costly, time-consuming process. ANNs are powerful, adaptable, scalable computational models that are able to overcome human decoder and pattern classification weaknesses. Therefore, the neural networks computer-based Silent Talker system has been trained and validated in the measurement of human comprehension using videotaped participant nonverbal behaviour from a Human Immunodeficiency Virus informed consent field study in Tanzania. A series of experiments on training backpropagation ANNs with different topologies were conducted. The results show that comprehension and noncomprehension patterns exist within the monitored multichannels of facial NVB with both experiments consistently yielding classification accuracies above 80%.**

Keywords- artificial neural networks, backpropagation, human comprehension, nonverbal behaviour, silent talker.

I. INTRODUCTION

The Human Immunodeficiency Virus (HIV) is a virus that attacks the human immune system, which gradually leads to the final stage of the infection known as the Acquired Immune Deficiency Syndrome (AIDS) where the damaged human immune system fails and is unable to fight other infections [1]. HIV can be prevented but the transmission of HIV still occurs and so the fight against reducing the global epidemic of AIDS/HIV remains an ongoing challenge. Alone, sub-Saharan Africa contains an inordinate majority of HIV infections with approximately 22.5 million people living with HIV in 2009 as reported by the 2010 UNAIDS report on the Global AIDS Epidemic Update [2]. Moreover, in sub-Saharan Africa more females are infected with HIV than males [2]. Therefore, sub-Saharan African females are especially vulnerable to HIV and as a result comprehension of the disease needs to be improved to help reduce further outbreaks.

HIV prevention clinical trials are often executed in developing countries where the prevalence of HIV/AIDS is high. Within trials the informed consent process is a legal and ethical requirement, requiring each participant to voluntarily make a truly informed decision on whether to participate in the study. Informed consent requires voluntary informed consent from comprehension of adequately delivered information about the purpose of the study, any procedures involved and the effects of participation [3] before commencement of participation. Although the Nuremburg Code [4], the Declaration of Helsinki [5] and the Belmont Report [6] exist to provide guidelines to ensure that research is ethical, the quality of the assessment of participant's comprehension during the informed consent process still remains a critical area of concern. Previous studies have identified that participants have difficulties in understanding informed consent documentation [7], [8], [9], which can impair decision making. Significant differences have also been found to exist between common comprehension assessment methods [10]. Therefore, the research presented in this paper has focused upon the development of an improved automated method in the measurement of human understanding during the informed consent process in a HIV/AIDS prevention field study with female, African participants.

Measurement of human comprehension has primarily been focused on verbal and written responses [11] with little attention being paid to the nonverbal dimension. Throughout everyday interaction humans are frequently exposed to nonverbal behaviours (NVB) that contain a rich source of information once deciphered e.g. facial expressions. Therefore, there is a niche for the exploration to identify whether patterns of comprehension and noncomprehension exist within NVB alone.

Artificial Neural Networks (ANN) are non-linear parallel computational models comprised of interconnected processing

The research is funded through the USAID Cooperative Agreement with FHI 360 for PTA, No. GHO-A-00-09-00016-00. The findings presented do not necessarily reflect FHI 360 or USAID policies.

U.S. Government work not protected by U.S. copyright

nodes [12]. Experiential knowledge is contained within the neural networks interconnections, which is obtained through the presentation of patterns using training and validation algorithms. ANNs have been adapted to a range of problem domains such as pattern classification [12], [13], image and speech recognition [14], financial analysis [13] and control systems [13]. Application of ANNs in the measurement of human understanding during informed consent could provide a new, more reliable technique of measuring comprehension in comparison to existing comprehension assessment methods.

Silent Talker [15] is an ANN-based psychological profiling system that monitors multiple channels of facial NVB to detect truthful and deceptive human behavioural patterns has the potential ability to be adapted to other human states in different environments. This paper discusses the adaptation and application of *Silent Talker* in the detection of human comprehension and noncomprehension from multichannels of NVB using ANN. This research is centered upon addressing the measurement of comprehension during the informed consent assessment process in a HIV/AIDS prevention field study carried out in North-western Tanzania, Africa.

Overall, the aim of this research is to identify whether human comprehension can be measured from NVB in near real-time using an adapted version of the *Silent Talker* ANN-based system. If the ANNs are able to detect nonverbal comprehension patterns then it could improve and positively impact on the informed consent process in many ways. Moderators of informed consent process would be able to use the tool to identify whether participants comprehended specific features of the informed consent process such as questions in the questionnaire and be able to dynamically adapt their approach until the participant adequately understands the question. As a result, existing comprehension assessment techniques and informed consent documents should dramatically improve as problematic comprehension areas would be identified almost instantly to the moderator allowing immediate correction application. The participants would benefit from enhanced, tailored education from the moderators during the informed consent process, which would facilitate higher comprehension levels and enable execution of more informed, voluntary consent decisions. Due to better education the participants would have an improved understanding of HIV and the prevention methods available, which could potentially lead to a reduction in HIV transmission and the size of the epidemic. Furthermore, other clinical trials could also reap the benefits of comprehension assessment tool through implementation during the informed consent process.

This paper continues as follows: Sections II, III and IV review NVB, comprehension and adapting *Silent Talker*. Section V presents the methodology with the experimental results and discussion contained in Section VI. The conclusion and future work is in Sections VII and VIII.

II. NONVERBAL BEHAVIOUR

NVB, is a form of non-linguistic communication that can accompany verbal responses or standalone. Examples of NVB are facial expressions, gestures, body movement and vocal cues lacking verbal content [16]. With NVB under less conscious control [17] and forming a large proportion of communication it could provide a more accurate, reliable measure of human state than verbal responses.

The nonverbal response is available for assessment by the listener prior to the verbal response [19]. Thus, potentially leaking early signals to the listener about the sender's state whilst a verbal response is still being formulated. Even during and after the verbal response, NVB's would be available for analysis providing a large data set to use in the identification of patterns of human state. The analysis of NVB should not only be limited to the presence of NVB cues but include their absence [20] too, as it could have a significant effect upon overall interpretation of the message.

When the verbal response is available alongside the nonverbal response, it would provide additional vocal cues that could be utilised in the decoding of the expressed emotion. Sauter *et al*. [21] demonstrated that acoustic cues from nonverbal vocalisations such as amplitude and pitch could be used to predict participants rating of emotion from the vocal expression alone. Vocalisations are not considered in this research as the prime focus is upon NVB only.

With there being a large repertoire of NVB channels available [22], researchers have the opportunity to collate data on individual or collections of NVB channels to identify patterns, associated with human state. Focusing upon the most appropriate channel(s) for detecting human state e.g. NVB's associated with deceptive behaviour requires knowledge and experimentation. During face-to-face communication humans tend to primarily focus upon the face as the area contains a rich source of information. Allocating higher precedence to the face than any of the other available channels is regarded as "facial primacy" [20]. Approximately, 55% of the communicated message by a person can be obtained from facial behavioural expressions [18] thus making the face a prime area of interest in the monitoring of human states. Mehrabian [23] argued that even when facial behaviour is not consistent with speech that the listener will be most impacted by the facial emotion expressed. Therefore, there is a great weighting toward NVB being a more reliable source than verbal responses.

To classify NVB, human judges have often been employed to code the NVB channels [24] that have been selected for monitoring. Humans are error prone, susceptible to fatigue and poor at multitasking thus inhibiting them as accurate judges in the classification of nonverbal channels. Moreover, human decoders are not always consistent, requiring time and capital to train.

The Facial Action Coding System (FACS) [25] is a common standard that has been adopted by multiple human decoders in order to aid and ensure reliability in the manual process of categorising facial expressions to measure emotions from NVB. Encapsulated within the FACS there are Action Units (AU) that have been coded to represent a contraction of individual or groups of muscles, which can then be utilised in scoring the facial expression as an emotion. Although the FACS standard exists to ensure consistency, the manual process of encoding NVB still remains a very time-consuming task when using human decoders. More time is required as the number of monitored NVB channels increases and the quantity

of video data raises, especially when taking a frame-by-frame analysis approach, resulting in great expense and delayed findings.

Nowadays, researchers have overcome the weaknesses of manually encoding and decoding NVB through the implementation of automated computer-based solutions to enable consistent real-time processing and classification of multiple NVB channels [26]. Multiple techniques have been used to capture and monitor NVB channels, ranging from invasive to non-invasive methods such as photographs, video recordings and sensors within natural and artificial settings. Therefore, the environment in which the assessment of NVB takes place can severely affect the natural quality of NVB produced, potentially introducing bias and affecting the reliability of the data. Photographs and videos are often used to capture NVB for analysis [27]. Photographs are weak as they only represent the static emotion expressed at a fixed point in time, are often unnatural due to being posed expressions, can miss micro-expressions, pre- and post-expressions of human state. Videos on the other hand are able to overcome the weaknesses of photographs, capturing the natural occurrences of spontaneous behaviour over time.

Overall, nonverbal communication encapsulates a large source of frequently available NVB's in a wide range of formats for analysis and overcomes the weaknesses of verbal behaviour. Interpretation of NVB's has been greatly improved through the use of computer-based technologies, overcoming human decoder weaknesses. Therefore, great potential exists in the development of computer-based programs for use in the detection of human states such as comprehension using NVB.

III. Comprehension

Noncomprehension is regarded as "a state of knowledge that ranges from uncertainty to complete lack of understanding of the materials under discussion" [28]; the opposite of understanding. The research in this paper is focused upon measuring human comprehension during informed consent. The informed consent process ensures that participants make a truly informed, educated decision through clear comprehension of the information delivered in the informed consent process otherwise misconceptions are liable to occur and decision related implications may arise. To ensure valid, ethical informed consent participants should receive adequate information on the nature of the research, the purpose of the study, any potential risks, participation benefits, alternatives and have a full understanding [29] of the concepts prior to consenting to participate.

Informed consent documents are often long, covering a lot of points in detail, which can inhibit comprehension through infliction of information overloading [30]. Comparison of comprehension assessment methods: self-reports, checklists, vignettes and narratives has identified significant differences on the measurement of understanding [13]. Paris *et al*. [31] found that even after improving an informed consent document through using a working group and by enhancing the lexicosyntactic readability through sentence length reduction and shorter synonyms that neither technique significantly enhanced participant's objective comprehension score thus emphasizing the need for an improvement in comprehension assessment techniques. Through early identification of low comprehension during the informed consent process it would provide time to adapt content delivery to facilitate understanding, thus avoiding legal and ethical implications.

Previous research on the detection of comprehension has had a dominant focus upon verbal and written responses [14]. This emphasizes the need to analyse from the nonverbal angle. With NVB forming such a large proportion of communication and under less conscious control it has the potential to deliver a more accurate measure of human comprehension in comparison to verbal communication. Although NVB is susceptible to self-monitoring [32] where human behaviour is controlled in a deceptive manner to enable the person to fit in the social environment, it has been found that nonverbal difficulty displays of behaviour exhibited during problem-solving tasks were more easily distinguishable from low self-monitoring students than high self-monitors [33]. Furthermore, through observation of nonverbal videotaped behaviour of children participating in a lesson on electricity containing difficult and easy content, individual observers were able to distinguish children comprehending and not comprehending from the deliberate and spontaneous NVB behaviours alone [34]. Moreover, it was also found that the observers were able to identify high achieving children as having a greater understanding than the low achieving children in both the easy and hard environments [34]. Therefore, it can be acknowledged that comprehension and noncomprehension patterns do reside within NVB alone but warrants further research as details of the NVB are often omitted and few channels are analysed. Therefore, potential distinct patterns of NVB relating to the level of understanding could be discovered.

Multiple classroom studies have identified NVB patterns associated with understanding and noncomprehension [35], [36], [37], [38] with facial behaviour, general hand and body movements predominantly being monitored using human decoders. However, informed consent studies have been limited to primarily independent predictors of comprehension from known data such as age [31], educational level [29], race and ethnicity [39]. Therefore, there is the unexplored need to identify whether more complex, general trends are present within larger experimental groupings of multiple NVB channels including independent predictors.

Development of a near real-time computer-based application to monitor multiple NVBs for comprehension measurement would overcome human weaknesses and help reduce the complexities of measuring understanding.

IV. Adapting Silent Talker

Silent Talker [15] is an ANN-based software system that successfully monitors multichannels of facial NVB for psychological profiling. Through a simulated theft scenario, *Silent Talker* was able to detect truthful and deceptive NVB with 80% accuracy [15].

When using the patented [40] near real-time *Silent Talker* system, the video camera is focused so that the interviewees NVB facial cues can be captured for analysis by *Silent Talker*, operated by the interviewer. During the interview the

monitored multichannels of NVB in the video frames are analysed by ANNs and at the end of the assessment of the video segment an overall deceptive/truthful classification is outputted to the interviewers display.

Within *Silent Talker* the object locators ANNs identify the location of nonverbal features such as the eye. Once located, the pattern detectors ANNs identify the state of the object such as the right eye gazing to the left. Each of the monitored NVB channels is coded, collated together for the entire answer and provided as input for the classifier ANN to determine the deceptive/truthful assessment of the response.

Currently, within *Silent Talker* the standard set up is 40 channels of NVB available covering the: left and right eye gaze and closure, blanching, blushing, slot (collection time), head movement, angle and rotation and known data such as the participant's gender [41].

Although, *Silent Talker* was originally, specifically designed as a "lie detector" it still has the unexplored, capability of being adapted to other unmonitored areas of human state where NVB patterns are of great interest. Adaptation to a new classification of behaviour using the existing NVB channels would only require retraining of the ANN classifiers. Furthermore, the introduction of new NVB channels and the pruning of existing NVB channels could also be implemented to extend and enhance the adaption of the ANN-based system to a new behavioural state.

Therefore, the research presented in this paper has adapted and applied the classifier ANN within *Silent Talker* through training and validation techniques to measure human comprehension from NVB in an informed consent field study. This paper shall identify whether such a complex ANN-based system can be adapted to measure human comprehension and noncomprehension from NVB's.

V. Methodology

A. Field Study Background

The field study was executed in the Mwanza region of Tanzania in Africa by Family Health International 360 (FHI 360) [42] in collaboration with the National Institute for Medical Research (NIMR) [43]. Eighty female participants between the ages of eighteen and thirty five who had not participated in a clinical trial before were recruited to participate in a video recorded interview on learning task topics of high and low comprehension. A digital camera was used to record the head shot of the participant during the entire interview, generating the Audio-Video Interleaved (AVI) files for analysis. Ethical approval was obtained from NIMR's Medical Research Coordinating Committee and from the Protection of Human Subjects Committee at FHI 360 and from Manchester Metropolitan University's (MMU) Faculty Academic Ethics Committee.

B. Interview Design

Task A was the high comprehension topic, designed to be easy to understand covering condom usage. Task B was the low comprehension topic, intended to be hard to understand on HIV mutation and treatment effectiveness. The purpose of difficult topic was to elicit high levels of distinct noncomprehension NVB's and the easy topic was designed to induce high frequencies of pure comprehension NVB's. The two tasks developed by FHI 360 were necessary to enable the distinct capturing of each type of NVB to reduce noise levels in the training data set. After listening to the short learning task script the participant received the associated ten closed and open-ended questions with randomisation applied to determine which set of questions were to be encountered first. Task order was also randomised so that half of the participants completed task A followed by task B and vice versa. Implementation of randomisation was intended to enable the analysis of any potential relationships between the NVB patterns, task order and question set order to be highlighted.

The resulting participant video files from the learning interview were used to obtain the ANN training data set for use in the experiments with different training calibrations.

C. ANN Classifier Training Procedure

Normalised input vector patterns scaled from +1.0 to -1.0 for each participant's learning task interview were obtained from each video frame in segments from the AVI file under experimental analysis. Each normalised value in a vector represented the state of a monitored NVB as an input for the ANN during training. For example, the left eye was coded as +1.0 when it was closed and -1.0 when open. A supervisory value was appended to the vectors, coded as +1.0 for comprehension and -1.0 for noncomprehension to independently assess whether the ANN correctly classified the presented input vector pattern. The categories of NVB captured were: known data, head movement, blushing, blanching, eye gaze and eye closure. Therefore, each analysed video frame had a corresponding normalised vector containing the state of all of the monitored nonverbal channels with a supervisory value. The normalised vectors were automatically generated by *Silent Talker*. The normalised vectors were randomly split in to training (Tr), validation (Va) and testing (Te) data sets.

Silent Talker's feedforward backpropagation classifier ANN was trained with randomised starting weights using incremental weight updating with the delta rule [44], [45], [46]. Presentation of the Tr, Va and Te sets during training were randomised to ensure that the neural network did not learn patterns on an order basis and to aid convergence speed.

The Tr set was used to update the weights based upon the calculated error. Va determined when to stop training with the trained weights frozen and Te determined the neural networks ability at generalising on an unseen data set. Classification accuracy (CA) was used to evaluate the performance of an ANN as a percentage of the total number of patterns that were correctly classified out of the presented desired data set.

n-fold cross-validation [47] was implemented where the data set is randomly split into *n* segments with one segment retained for testing purposes and the remaining segments used in training and validation. Training ends when all of the *n* segments have completed the role as the test set once.

VI. Results and Discussion

The study yielded thirty usable videos resulting in 600 answers along with the two video segments capturing NVB during the reading of each learning task script. Fifty videos were discarded from the data set due to poor quality of the recordings.

In each of the ANN training experiments the ANNs had the same topology: forty inputs, a single hidden layer of thirty neurons and a single output neuron. The topology was chosen because it was the best performing neural network from previous experimental training configuration sessions. The inputs to the ANNs were the same as the nonverbal input channels used in the original *Silent Talker* experiments. Prior to the 10-fold cross-validation commencement the data set partitioned as Tr (60%), Va (30%) and Te (10%). The individual experiments and their results shall be discussed.

A. Experiment 1: Script Readings

The training data set for Experiment 1 came from the video segments containing the entire reading of the task A and task B scripts. Experiment 1 yielded 241,945 vectors with a near balanced data set: script A = 45% labelled comprehension and script B = 55% labelled noncomprehension. During the reading of script A participants understanding was predicted to be high with high frequencies of NVB comprehension patterns. Understanding was predicted to be low with high levels of associated NVB noncomprehension patterns during the reading of script B.

Table I shows the CA's for each phase of the 10-fold cross-validation iterations. An iteration is the completion of a single run of 10-fold cross-validation. Iteration 5 of the 10-fold cross-validation generated the best performing ANN with an overall testing CA of 89.29%. Table II and III display a breakdown of the comprehension and noncomprehension CA's for each training phase. All of the CA's within the Tr, Va and Te phases of training have consistently achieved CAs greater than 85%, strongly indicating that comprehension and noncomprehension patterns exist within the data set.

Through examination of the participants marked responses to the open-ended questions for script A and B were found to not be purely comprehension responses for task A and not entirely noncomprehension answers for task B. Thus, the two data sets were too noisy. The next experiment implemented a technique to try and overcome the latter weakness.

TABLE I. Experiment 1: Cross-Validation CA's

Iteration	Training CA	Validation CA	Testing CA
1	88.35%	86.88%	87.15%
2	89.25%	87.69%	87.82%
3	89.46%	88.06%	88.53%
4	89.74%	88.50%	88.50%
5	90.05%	88.95%	89.29%
6	90.50%	89.06%	89.09%

TABLE II. Experiment 1: Cross-Validation Comprehension CA's Breakdown

Iteration	Training CA	Validation CA	Testing CA
1	89.89%	88.55%	88.81%
2	89.73%	87.93%	87.80%
3	89.87%	88.44%	88.68%
4	89.88%	88.63%	88.69%
5	88.85%	87.41%	87.91%
6	90.72%	88.98%	89.69%

TABLE III. Experiment 1: Cross-Validation Noncomprehension CA's Breakdown

Iteration	Training CA	Validation CA	Testing CA
1	86.80%	85.22%	85.49%
2	88.78%	87.45%	87.84%
3	89.06%	87.67%	88.39%
4	89.59%	88.38%	88.32%
5	91.26%	90.48%	90.67%
6	90.27%	89.13%	88.49%

B. Experiment 2: Script Reading Points Mapped to the Open-ended Questions

Experiment 2's data set consisted of the reading of the individual sentences within each of the scripts being labelled with a supervisory value through mapping the participants associated open-ended question response to the individual script point. Therefore, if the participant answered the question correctly the corresponding script point vectors would all have a +1.0 supervisory value indicating comprehension and -1.0 when noncomprehension. The close-ended questions were not selected to be mapped to the script points due to being strongly inhibited by the susceptibility to guessing reducing results to chance levels. Experiment 2 yielded a data set containing 71,787 vectors: 63.5% comprehension and 36.5% noncomprehension.

The CA's for each phase of the 10-fold cross-validation iterations is displayed in Table IV. An iteration is the completion of a single run of 10-fold cross-validation. Iteration 3 of the 10-fold cross-validation generated the best performing ANN with an overall testing CA of 87.05%. Table V and VI contain the comprehension and noncomprehension CA breakdown for each training phase. The ANNs consistently achieved CAs above 81%, indicating that comprehension and noncomprehension patterns exist within the data set.

Later inspection of the participants marked open-ended questions found that 70% of responses were correct resulting in high levels of comprehension.

TABLE IV. EXPERIMENT 2: CROSS-VALIDATION CA'S

Iteration	Training CA	Validation CA	Testing CA
1	88.59%	84.25%	84.55%
2	89.74%	85.42%	85.07%
3	90.47%	85.62%	87.05%
4	89.52%	85.39%	85.87%
5	89.43%	84.83%	85.54%
6	88.33%	84.17%	85.02%

TABLE V. EXPERIMENT 2: CROSS-VALIDATION COMPREHENSION CA'S BREAKDOWN

Iteration	Training CA	Validation CA	Testing CA
1	89.52%	85.52%	86.34%
2	89.92%	87.12%	87.09%
3	90.54%	86.97%	88.27%
4	89.20%	86.05%	86.40%
5	88.28%	84.50%	84.52%
6	89.53%	87.00%	87.31%

TABLE VI. EXPERIMENT 2: CROSS-VALIDATION NONCOMPREHENSION CA'S BREAKDOWN

Iteration	Training CA	Validation CA	Testing CA
1	87.67%	82.97%	82.75%
2	89.57%	83.73%	83.04%
3	90.40%	84.27%	85.84%
4	89.85%	84.73%	85.34%
5	90.57%	85.16%	86.55%
6	87.13%	81.34%	82.72%

C. *Summary of Experiments*

Both experiments consistently attained CA's grouped above 80% in the Tr, Va and Te phases of the cross-validation from randomised starting weights and vector presentation. Therefore, the latter results strongly indicate that multichannel NVB patterns of comprehension and noncomprehension do exist within the data set.

Silent Talker performed well at classifying the previously unencountered ethnic data set of Tanzanian women. However, retraining of the pattern detector and object locator ANN's with the unseen ethnic data set would enhance *Silent Talker*'s ability at locating the individual facial features and their state, which would increase the CAs. Furthermore, improvements in the quality of recordings would lead to a larger training data set with lower levels of discarded videos thus potentially increasing the applicability of the results.

From executing the ANN training experiments, it has identified key elements of the field study that can be improved and applied to future work in order to enhance the quality of the data set. Firstly, the style of the open-ended questions and the learning task scripts should utilise techniques to increase the difficulty of task content to raise participant's noncomprehension levels to ensure that a more balanced data set is obtained. For instance, question six in task B was "Please name at least one of the two major HIV viral types." and the corresponding script point was "These include two major viral types (HIV-1 and HIV-2)". Therefore, increasing the complexity of the tasks would make recall and guessing more difficult for participants, ensuring clearer comprehension and noncomprehension data is collated. Introduction of a time delay between script reading and questioning could help to minimise the lack of deeper understanding through the production of recalled script keywords thus enhancing the quality of the responses. Rushing points, overloading the receiver with complex information and using jargon have been found as causes of patient misunderstanding during clinician and patient communication [48] thus providing additional techniques that could be used to impair comprehension. Application of the improved techniques should provide a more balanced data set with low noise levels resulting in clearer comprehension and noncomprehension data sets for experimentation.

VII. CONCLUSION

Application of *Silent Talker* to a new environment in the role as a "comprehension detector" during the informed consent process has proven that it is possible to adapt such a complex ANN-based system to a different human state in a new environment. The experimental results strongly indicated that detectable patterns of comprehension and miscomprehension exist within the monitored facial NVB multichannels of the data set as the cross-validation consistently attained CA's grouped above 80% from randomised weights and vector presentation. The results are limited to African women and so further experiments with participants of different ethnicities, ages and genders are necessary to enhance the reliability of the results and to confirm whether general patterns of comprehension exist outside of the demographics contained in the field study population.

Through adapting *Silent Talker* in the measurement of human understanding during informed consent it has also identified features of the field study and components of *Silent Talker* that can be enhanced in future work to improve the accuracy of comprehension measurement and the quality of the training data set.

VIII. FURTHER WORK

This paper presents a potential comprehension detector prototype that could be used as a proxy tool alongside existing comprehension assessment techniques to aid understanding during the informed consent process in near real-time. Future work includes the introduction of new NVB channels and the pruning of existing NVB channels to extend and enhance the adaption of the ANN-based system to the new behavioural state. Also, comparisons between the nonverbal results from the ANN-based system and existing verbal measures will be analysed.

Further applications of a comprehension monitoring system are numerous. In education, it would provide teachers with a quick and easy method of gauging the student's level of understanding, allowing them to change the lesson content to suit the needs of the individuals without great delay thus saving time for further progression, which would have not been possible before. Within the medical environment it could be used to ensure that misconceptions during diagnosis are rectified and to make sure that patients fully understand treatments and their prescriptions.

ACKNOWLEDGMENT

We wish to thank the women in Tanzania who participated in this research and especially for their willingness to be video-recorded. We also wish to thank our collaborators at NIMR in Tanzania, including Bahati Andrew, Tusa Erio, Catherine Bunga, and Gerald Lumanyika.

REFERENCES

[1] WHO, 2011, HIV/AIDS Fact Sheet [Online] Available: http://www.who.int/mediacentre/factsheets/fs360/en/index.html

[2] UNAIDS, 2010, Global report: UNAIDS report on the global AIDS epidemic 2010 [Online] Available: http://www.unaids.org/globalreport/documents/20101123_GlobalReport_full_en.pdf

[3] G. Lindegger , "Informed consent in HIV vaccine trials" In P. Kahn, AIDS Vaccine Handbook: Global Perspectives, 2nd Ed., New York: AIDS Vaccine Advocacy Coalition, pp.109-116, 2005.

[4] Nuremberg Code. In: Trials of war criminals before the Nuremberg military. Tribunals under control counsel law, vol. 11. Washington, DC: U.S. Government Printing Office, No. 10, 1949. p. 181–2.

[5] World Medical Association (WMA). Declaration of Helsinki: ethical principles for medical research involving human subjects, as amended by the 52nd WMA General Assembly, Edinburgh, Scotland; 2000.

[6] National Commission for the Protection of Human Subjects of Biomedical, Behavioral Research. The Belmont report: ethical principles and guidelines for the protection of human subjects of research. Washington, DC: U.S. Government Printing Office; 1979.

[7] D. W. Fitzgerald, C. Marotte, R. I. Verdier, W. D. Jr. Johnson and J. W. Paper, "Comprehension During Informed Consent in a Less-developed Country", Lancet, vol. 360, pp. 1301-1302, 2002.

[8] J. Flory and E. Emanuel, "Interventions to Improve Research Participants' Understanding in Informed Consent for Research: A Systematic Review", JAMA, vol. 292(13), pp. 1593-1601, 2004.

[9] A. R. Oduro, R. A. Aborigo, D. Amugsi, F. Anto, T. Anyorigiya, F. Atuguba, A. Hodgson and K. A. Koram, "Understanding and Retention of the Informed Consent Process Among Parents in Rural Northern Ghana", BMC Medical Ethics, vol. 9(12), 2008.

[10] G. Lindegger, C. Milford, C. Slack, M. Quayle, X. Xaba and E. Vardas, "Beyond the Checklist: Assessing for Understanding for HIV Vaccine Trial Participation in South Africa", Acquired Immune Deficiency Syndrome, vol. 43(5), pp. 560-566, 2006.

[11] M.E. Falagas, I.P. Korbila, K.P. Giannopoulou, B.K. Kondilis, and G. Peppas, "Informed consent: how much and what do patients understand?," American Journal of Surgery, vol. 198(3), pp. 420-435, 2009.

[12] M. H. Hassoun, Fundamentals of Artificial Neural Networks, London: MIT Press, 1995.

[13] B. Widrow, D. E. Rumelhart and M. A. Lehr, "Neural networks: applications in industry", Business and Science. Commun. ACM, vol. 37(3), pp. 93–105, 1994.

[14] K. Gurney, An Introduction to Neural Networks, London: UCL Press, 1997.

[15] J. Rothwell, Z. Bandar, J. O'Shea and D. McLean, "Silent Talker: A New Computer-based System for the Analysis of Facial Cues to Deception", Applied Cognitive Psychology, vol. 20, pp. 757-777, 2006.

[16] E. Babad, "Teaching and Nonverbal Behavior in the Classroom" In L. J. Saha, and A. G. Dworkin (Eds.) International Handbook of Research on Teachers and Teaching, Boston, Massachusetts: Springer US, pp. 817-827, 2009.

[17] T. P. Mottet and V. P. Richmond, "Student Nonverbal Communication and its Influence on Teachers and Teaching" In J. L. Chesebro & J. C. McCroskey (Eds.), Communication for Teachers, Needham Heights, Massachusetts: Allyn and Bacon, pp. 47-61, 2002.

[18] A. Mehrabian, "Communication Without Words", Psychology Today, vol. 2(4), pp. 53–56, 1968.

[19] V. Manusov and A. R. Trees, ""Are You Kidding Me?": The Role of Nonverbal Cues in the Verbal Accounting Process", Journal of Communication, vol. 52(3), pp. 640-656, 2002.

[20] M. L. Knapp and J. A. Hall, Nonverbal Communication in Human Interaction, 3rd Ed, Fort Worth: Harcourt Brace, 1992.

[21] D. A. Sauter, F. Eisner, A. J. Calder and S. K. Scott, "Perceptual cues in nonverbal vocal expressions of emotion", The Quarterly Journal of Experimental Psychology, vol. 63(11), pp. 2251-2272, 2010.

[22] P. Ekman and W. V. Friesen, "The Repertoire of Nonverbal Behavior Categories, Origins, Usage, and Coding", Semiotica, vol. 1, pp. 49-98, 1969.

[23] A. Mehrabian, Silent Messages, 5th Ed., California: Wadsworth Publishing Company, 1971.

[24] B. McDaniel, S. D'Mello, B. King, P. Chipman, K. Tapp and A. Graesser, Facial Features for Affective State Detection in Learning Environments, Proc. 29th Ann. Meeting of the Cognitive Science Soc, 2007.

[25] P. Ekman and V. W. Freisen, The Facial Action Coding System (FACS), Consulting Psychologists Press, Palo Alto, CA, US, 1978.

[26] A. Sarrafzadeh, S. Alexander, F. Dadgostar, C. Fan and A. Bigdeli "How do you know that I don't understand?" A look at the future of intelligent tutoring systems. Computers in Human Behavior, 24(4), pp. 1342-1363, 2008.

[27] Z. Zeng, M. Pantic, G. I. Roisman and T. H. Huang, "A survey of affect recognition methods: audio, visual and spontaneous expressions" IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31(1), pp. 39–58, 2009.

[28] H. Z. Waring, "Expressing noncomprehension in a US graduate seminar", Journal of Pragmatics, vol. 34(12), pp. 1711-1731, 2002.

[29] J. T. Krankl, S. Shaykevich, S. Lipsitz and L. S. Lehmann, "Patient Predictors of Colposcopy Comprehension of Consent Among English- and Spanish-speaking Women", Women's Health Issues, vol. 21(1), pp.80-85, 2011.

[30] M. Hochhauser, 2004, Informed Consent: Reading and Understanding Are Not the Same. Applied Clinical Trials [Online] Available: http://appliedclinicaltrialsonline.findpharma.com/appliedclinicaltrials/article/articleDetail. jsp?id=90594.

[31] A. Paris, C. Brandt, C. Cornu, P. Maison, C. Thalamas and J. Cracowski, "Informed consent document improvement does not increase patients' comprehension in biomedical research", British Journal of Clinical Pharmacology, vol. 69(3), pp. 231-237, 2010.

[32] M. Snyder, "Self-monitoring of Expressive Behavior", Journal of Personality and Social Psychology, vol. 30, pp. 526–537, 1974.

[33] D. Hrubes and R. S. Feldman, "Nonverbal Displays as Indicants of Task Difficulty", Contemporary Educational Psychology, vol. 26, pp. 267-276, 2001.

[34] V. L. Allen and M. L. Atkinson, "Encoding of nonverbal behavior by high-achieving and low-achieving children", Journal of Educational Psychology, vol. 70(3), pp. 298-305, 1978.

[35] J. D. Jecker, N. MacCoby, and H. S. Breitrose, "Improving Accuracy in Interpreting Non-verbal Cues of Comprehension", Psychology in the Schools, vol. 2(3), pp. 239-244, 1965.

[36] S. Machida, "Teacher Accuracy in Decoding Nonverbal Indicants of Comprehension and Noncomprehension in Anglo- and Mexican-

American Children", Journal of Educational Psychology, vol. 78(6), pp. 454-464, 1986.

[37] C. J. Patterson, M. J. Cosgrove and R. G. O'Brien, "Nonverbal Indicants of Comprehension and Noncomprehension in Children", Developmental Psychology, vol. 16(1), pp. 38-48, 1980.

[38] E. Skarakis-Doyle, N. MacLellan and K. Mullin, "Nonverbal Indicants of Comprehension Monitoring in Language-disordered Children", Journal of Speech and Hearing Disorders, vol. 55(3), pp. 461-467, 1990.

[39] A. S. Fink, A. V. Prochazka, W. G. Henderson, D. Bartenfeld, C. Nyirenda, A. Webb, D. H. Berger, K. Itani, T. Whitehill, J. Edwards, M. Wilson, C. Karson, and P. Parmelee, "Predictors of Comprehension during Surgical Informed Consent" Journal of the American College of Surgeons, vol. 210(6), pp. 919-926, 2010.

[40] Z. Bandar, D. A. McLean, J. D. O'Shea and J. A. Rothwell, International Patent Number WO02087443. Geneva, Switzerland: World Intellectual Property Organization, 2002.

[41] J. Rothwell, Z. Bandar, J. O'Shea and D. McLean, "Charting the Behavioural State of a Person Using a Backpropagation Neural Network", Neural Computing & Applications, vol. 16, pp. 327-339, 2007.

[42] Family Health International 360, 2011, Who We Are. [Online] Available: http://www.fhi.org/en/AboutFHI/index.htm.

[43] National Institute for Medical Research Tanzania, 2011, About Us [Online] Available: http://www.nimr.or.tz/.

[44] P. Werbos, The Roots of Backpropagation: From Ordered Derivatives to Neural Networks and Political Forecasting, Wiley-Interscience New York, NY, USA, 1994.

[45] P. Werbos, "Backpropagation through time: what it does and how to do it", Proceedings of the IEEE, vol. 78(10), pp. 1550-1560, 1990.

[46] P. Werbos, "Beyond Regression: New tools for prediction and analysis in the behavioural sciences", Doctoral Dissertation, Appl. Math., Harvard University, Cambridge, MA, USA, 1974.

[47] M. Stone, "Cross-validatory Choice and Assessment of Statistical Predictions", Journal of the Royal Statistical Society, vol. 36(2), pp. 111-147, 1974.

[48] H. F. West and W. F. Baile, ""Tell me what you understand": the importance of checking for patient understanding", The Journal of Supportive Oncology, vol. 8(5), pp. 216-218, 2010.