Extraction of Information From Eye Gaze Data

Leana Copeland

A thesis submitted in partial fulfilment of the degree of
Bachelor of Software Engineering at
Research School of Computer Science
Australian National University

October 2011

Except where otherwise indicated, this thesis is my own original work.

Leana Copeland

27 October 2011

Acknowledgements

Thank you to Prof. Tom Gedeon who was my supervisor and mentor throughout this project. Thank you for always being supportive and providing kind words and help whenever needed.

I would also like to thank Nandita Sharma, Dingyun Zhu and Fateme Rajabi. Thank you for your help, support and kindness throughout the year, the weekly meetings will be missed.

A further thanks to Nandita Sharma for providing me with data for this study as well as your help to get started using the data.

Thank you to Jaimi McAlister and Katie Johnstone for being the most amazing friends who are always there for me, always cheering me on, and for proofreading for me.

Last but not least, a big thank you to my wonderful Mum, Dad and brother Michael, for supporting me throughout my whole undergraduate degree and always believing in me. Mum and Dad, thank you not just for being the best parents but for also taking the time to proofread for me, I really appreciate it.

Abstract

This study uses data from subjects where eye gaze was recorded as they read text on a screen. The text shown to participants was grouped into different types of content; paragraphs versus questions, relevant versus less-relevant and hard versus easy content. Different analysis techniques are used to identify differences in eye gaze patterns recorded between reading the different types of content. The analysis techniques focus on the words that a participant fixated upon as well as the linear movements of the eye during reading. Hidden Markov models (HMM) are used in this study as a classification technique. HMMs are used to classify sequences based on linear eye movements with differing levels of granularity in the movements. We found that the finer the granularity of modelling of the eye movement the better the classification. In particular, a classification level of 91% between hard and easy paragraphs was achieved.

The results from the analysis techniques show that eye gaze patterns recorded from reading different content types can be differentiated based on both linear measurements of eye movement as well as based on the words that the participant fixated on.

In one of the data sets, the comprehension of the content read was recorded. These results were analysed in this study and show that there are moderate correlations between both the words fixated upon as well as the proportions of backward movements observed in the paragraphs and the reading comprehension scores.

Contents

Chapter 1

Introduction

Reading is a complex process that has only recently become part of the human experience in comparison to spoken language. Reading text, as well as comprehending that text, requires disregarding events around you and focusing your attention on the task almost completely. Reading is not something humans are biologically programmed to do, yet it is one of the most important and ubiquitous skills that a human can acquire in modern society. A good deal of research has been completed on the cognitive processes behind reading. If reading could be better understood it might help our understanding of skill acquisition in general, and in turn improve teaching methods for such skills. Moreover, if eye movement patterns could be better understood, better designed eye-based interactions with computers would be possible, for example better eye movement typing systems could be developed (Salvucci, 1999).

Eye movements provide clues to underlying cognitive processes that occur when performing visual tasks, such as reading, and they help to determine how and when people obtain information as well as what information they use or ignore.

It has been shown in several studies that eye gaze patterns can be used to detect what kind of task the participant is performing (Iqbal & Bailey, 2004, Salojarvi et al., 2005), or when a person is viewing particular expressions on an individual (Kozek, 1997), or when a person is reading or not reading (Gustavsson, 2010). Previous studies have shown that even within the activity of reading, eye gaze patterns can be used to differentiate when individuals are reading different types of content (Vo et al., 2010).

Cognitive models examine the relationships between cognitive and oculomotor processes and eye movements. The intent of the research is to find relationships between text presented to individuals and the eye movements they generate as they read that text. Rather than considering the cognitive psychology point of view, we will be investigating the differences in eye movements as a consequence of the text presented to the individuals.

This study is an analysis of eye gaze data recorded from individuals who read a series of paragraphs. This is accomplished by using various analysis techniques on two data sets. The first data set was collected in an experiment run in 2009 (Fahey, 2009) that investigated the relationships between reading comprehension and eye gaze data from participants who read a series of 10 paragraphs and 5 questions. The second data set comes from a 2011 experiment (Sharma, 2011) related to modelling stress, where again the participants read a series of paragraphs. Only a subset of the Sharma experimental data is used in this study as the majority of it is beyond this study's scope.

The paragraphs are categorised into different groups based on content; for the Fahey experiment the categories are relevant and less-relevant. In the Fahey experiment, the paragraph content was about eye gaze tracking and some of its uses, so the difference between the paragraphs' content was that the relevant paragraphs were taken from discussion sections of a research paper on eye tracking whereas the less-relevant paragraphs were taken from introductory sections of the paper and from student reports on the paper. The paragraphs in the second experiment were categorised based on the difficulty of the content and are labelled either hard or easy. There was no general topic for these paragraphs.

This study hopes to find if there is a difference between the gaze patterns recorded from reading different types of text. The types of text used aim for comparison between relevant and less-relevant paragraphs as well as between hard and easy paragraphs. For the Fahey data set, the eye gaze patterns generated from reading paragraphs will be compared to those generated from reading questions. The assumption is that different eye gaze patterns are generated whilst reading different forms of text.

Various forms of analysis will be used on eye gaze patterns to detect differences in patterns gathered from the different paragraphs and data sets, which includes assessing the data statistically and using hidden Markov models as classifiers.

Finally, since reading comprehension scores were recorded for the Fahey data set, the eye gaze patterns are compared to the reading comprehension scores to determine if there is any correlation between the two, that is, if there is any correlation between reading behaviour (eye gaze patterns) and reading comprehension.

1.1 Hypotheses

The purpose of this study is to detect differences between the gaze patterns observed from reading different types of text. There are several hypotheses that are tested on the data sets used in the study. these hypotheses are:

1. Words affect where a fixation will occur; In particular, longer words will attract the eye and longer words will require multiple fixations to read. Furthermore, fixations will occur on infrequent words more than frequent words.

2. Different levels of forward and backward tracking will be observed between the categories of paragraphs. In particular, more backtracking will be observed in the relevant and harder paragraphs compared to less-relevant and easier paragraphs, and more forward tracking will be observed in less-relevant and easier paragraphs when compared to relevant and harder paragraphs, respectively.

3. There will be a difference in eye gaze patterns observed whilst reading paragraphs as compared to reading the questions.

4. The magnitudes of the forward and backward movements are affected by the different categories of paragraphs.

5. Finally, there will be a correlation between the reading comprehension scores recorded from the Fahey experiment and the above analysis of eye gaze patterns.

These hypotheses are used to answer the overarching hypothesis that there is a difference in eye gaze patterns generated from reading different types of content.

1.2 Project Plan

The following table outlines the planned completion dates for the major milestones of the project:

Activity	Completion Date	Notes
Literature Review/ Introduction Seminar	28th March 2011	*Course set deadline*
Preparation activities such as text analysis and calibration of data	End of Semester 1	
Scoring analysis	Start of Semester 2	
Backward and Forward tracking analysis	Start of Semester 2	
Midterm Seminar	29th August 2001	*Course set deadline*
Reading comprehension analysis for Fahey data set	End of term 3	
hidden Markov model analysis	End of Term 3	
Thesis	27th October 2011	*Course set deadline*
Final Seminar	7-8th November 2011	*Course set deadline*

1.3 Document Outline

This document reports the analysis and results from existing eye gaze tracker data. The chapters of the document are as follows:

Chapter 1 : Introduction

Introduction to the thesis, including project place, document outline and glossary.

Chapter 2 : Literature Survey

Literature survey to provide background information and introduce the direction of the analysis.

Chapter 3 : Preliminary Work

Details of all work needing to be done before analysis could be performed such as data cleaning and calibration.

Chapter 4 : Analysis

Description of the analysis techniques used in the study.

Chapter 5 : Results

Outline of the results achieved from the analysis of the eye gaze data.

Chapter 6 : Discussion

Discussion of the results achieved from the analysis of the eye gaze data.

Chapter 7 : Conclusion and Further work

Summary of the results and discussion of the analysis performed in the study as well as discussion of future work that could be performed on the data.

1.4 Glossary

Some terms that will be frequently used throughout this document include:

Content Type

Text content shown to participants in both experiments are categorised into several types. For the Fahey data set: relevant, less-relevant, hard, easy and questions. For the Sharma experiment, these types are hard and easy.

Easy Paragraph

Content of paragraph shown to participants is conceptually easy to understand and readability tests show that content is easy to read.

Eye Gaze Pattern

The combination of all the eye gaze points recorded for a participant for each screen showing a text paragraph.

Eye Gaze Point

The eye gaze trackers take measurements of where the participants eye is looking on the screen at regular intervals (in this case 60Hz). Gaze points are used to determine fixations and saccades.

Fixation

When the eye finishes a saccade and stays relatively still to take in visual information for processing.

Hard Paragraph

Content of paragraph shown to participants is conceptually difficult to understand and readability tests show that content is quite difficult to read.

Less-Relevant Paragraph

Content of paragraph shown to participants is from an introductory section of a particular research paper or student reports on that research paper. Content is believed to be easier to understand than the content in the relevant paragraphs. *(Only applies to Fahey data set.)*

Oculomotor

Relating to or causing movement of the eyeballs.

Questions

Content shown to participants is in the form of a multiple choice question, where a questions is asked and four answers are provided below the question. *(Only applies to Fahey data set.)*

Relevant Paragraph

Content of paragraph shown to participants is from the discussion section of a particular research paper. Content is believed to be more difficult to understand than the content in the less-relevant paragraphs. *(Only applies to Fahey data set.)*

Saccade

A small rapid movement of the eye as it jumps from one fixation to another; little to no visual information is taken in during a saccade.

Two-tailed, paired t-test

The t-test assesses whether the means of two groups are statistically different from each other. T-tests are used throughout the study to compare the participants results recorded from one category of content to another:

- *Two-tailed test:* Used where there is no basis to assume that there may be a significant difference between the groups.

- *Paired t-test:* Used when each data point in one group corresponds to a matching data point in the other group.

Chapter 2

Literature Survey

This survey is a compilation of reading and investigation conducted before the analysis of the data, setting the scene for the research and providing background information for the reader. There will be a brief introduction to the human eye, human language as well as current reading models. Finally there will be a discussion of hidden Markov models (HMMs) and their uses to analyse eye gaze data.

2.1 Previous Work

This study is an extension of the work by Fahey in 2009 which investigated the connection between the way that people read and the way that they understand information. Fahey performed an experiment where participants read a series of paragraphs, seven of which were from "Keyboard before Head Tracking Depresses User Success in Remote Camera Control" (Zhu, et al., 2009) and the remaining three of which were written by students who were required to write about the paper for course work. Five paragraphs from the paper's discussion section were selected because they were considered to be "relevant" in terms of detail, and the remaining paragraphs were considered "less-relevant" due to their generality and lack of detail. Three of the less-relevant paragraphs were selected from the introduction section of Zhu's paper, the remaining two were written by students required to write a report on Zhu's paper.

Participants read these paragraphs followed by a series of questions based on these paragraphs while an eye tracking system recorded eye gaze information. The participants then answered the questions and wrote a sentence about the information they read. The participants were scored on the sentence out of 3 based on a criterion set out to assess whether they grasped the big picture of the

10 paragraphs they had read. The participants were also scored on whether they answered the questions right or wrong, scoring one point for a correct answer and zero points for a wrong answer. Finally, the participants were asked to rank the paragraphs as relevant or less-relevant, and they were given a point for correctly identifying the paragraph as being relevant (or less-relevant) and zero points for incorrect identification. The highest possible score therefore was 18. Fahey found that the highest score achieved was 16, the lowest was 4 and the mean and standard deviation were 9.6 and 3.3 respectively.

Statistical analysis was performed on the gaze patterns and the following statistical information measured and analysed:

- Average time taken to read a paragraph.
- Average horizontal distance between fixations.
- Average vertical distance between fixations.
- Average number of gaze points per paragraph.
- Average number of fixations per paragraph.
- Average length of fixations.

These measurements were plotted against the scores of the participants, which showed that there were no statistically significant correlations (Fahey, 2009). However, Fahey does note that the plots show that participants who had lower average distances between fixations when reading the paragraphs generally received better scores.

Finally, Fahey showed that a neural network could be trained to recognise the difference between relevant and less-relevant paragraphs using the eye gaze data.

Sharma (2011) performed an experiment to model stress. In this experiment, participants were asked to read a series of paragraphs whilst aspects of their physiological response were monitored. These physiological responses include eye gaze, galvanic skin response, blood pressure, etc. The paragraphs of text were categorised as follows; hard, easy, stressful, calm and neutral. Participants were shown each paragraph for 60 seconds. After they had read all the paragraphs, participants were asked to fill out a survey, which consisted of answering questions about their reaction to the paragraphs that they remembered.

Only a small subset of Sharma's dataset was available to be analysed in this study. This dataset that will be analysed is the eye gaze data for both the hard and easy paragraphs.

2.2 Description Data already recorded

The eye gaze data for both Fahey and Sharma experiments were recorded from two Sony VFCB-EX480B infrared Face-Lab cameras, and the tracking software used was Seeing Machines Face-Lab 4.5. The system was manually calibrated for each participant (Fahey, 2009, Sharma 2011).

For Fahey's study the raw data that was collected from the gaze tracker is in a Microsoft Access database called "db1.mdb". The gaze point information has been extracted from the database and each participant's data is kept in a file *participant_num.txt* where *participant_num* is their assigned participant number. There is also a file called *Participants.txt* that includes information about the participants such as their gender, age range and whether they are a Computer Science student or not.

The gaze points were recorded by the gaze tracker at 60Hz. The gaze points were then translated into fixation points by calculating the distance between gaze points. If two gaze points were within a threshold of 15 pixels they were considered to be part of a fixation. All other gaze points were considered to be part of a saccade and were therefore discarded.

GNU DevIL was used to draw the fixations and saccades onto the paragraphs and questions for each participant. There are separate JPEG images stored for each of the participants' gaze points on each paragraph and question (Fahey, 2009). Visual inspection of these images of the fixations show that there is much variation between participants. It is also apparent that some calibration of the data may be required.

2.3 Physiology of the Eye

Eyes are the small but complex organs which enables vision in humans. In most cases, reading is possible because our eyes give us the ability to see; with the exception of Braille. In a simplified explanation the eye performs two basic processes, the first of which is to focus light onto the retina via the cornea, pupil and lens. Next, the retina converts this visual image into neural impulses that are relayed to and interpreted by the brain. So, when humans read it is most likely that the eyes are the starting point of the process; again with the exception of Braille for blind people. Since the eye is an important part of reading we will discuss further how letters on a piece of paper or on a computer screen make their way into the human brain for interpretation.

First light enters the eye through the cornea, which is a transparent covering over the iris and pupil at the front of the eye. Light then travels through the opening in the centre of the coloured part of the eye, which are the pupil and the iris respectively. The iris dilates and constricts the pupil to regulate the amount of light that enters the eye (Burton et al., 2005).

The lens then focuses the light that enters through the pupil onto the retina which is a light sensitive layer of tissue at the back of the eye (Burton et al., 2005). The retina is a very important part of the eye because it transforms the light that enters the eye into neural impulses, which is information that the brain can interpret.

The retina is a complex multilayer structure; the layer at the back of the retina contains two types of light receptors which are rods and cones. These are specialised neurons capable of phototransduction. Rods are responsible for vision in low level light and are used in peripheral vision. Cones, on the other hand, are responsible for vision in higher levels of lights and for our ability to see colour.

When a rod or cone absorbs light energy, it generates an electrical signal stimulating the next layer of the retina which is made up of bipolar cells. These cells combine information from the rods and cones to produce graded potentials on the next layer of the retina, the ganglion cells. The ganglion cells integrate information from the bipolar cells. The long axons of the ganglion cells are bundled together to form the optic nerve, which carries visual information to the brain. Information from the optic nerve first passes through the optic chiasma where information from the left half of each visual field goes to the right hemisphere and similarly for the right. The information then passes to the visual cortex in the occipital lobes for processing (Burton et al., 2005).

Peripheral vision uses the whole retina, but detailed vision is handled by the fovea. The fovea is the small central region of the retina that is sensitive to fine detail. The fovea only sees the central 2° of the visual field (Rayner & Bertera, 1979). The fovea does not have rods, only cone light receptors. The parafovea extends approximately 1mm from the central fovea (Rayner & Bertera, 1979). The fovea is necessary in humans for reading, watching television, driving and other activities that require detailed vision. The fovea takes up less than 1% of the retina yet processing of this information accounts for over 50% of the activity of the visual cortex in the brain (Mason & Kandel, 1991).

Peripheral vision is sensitive to sudden movement, and mostly used to gather information about the present surroundings. The human brain prioritises the information to give attention only to what its somehow deems important. Itti and Bladi (2009) found that humans orient their attention and gaze toward surprising events or items, in the context of watching television. Humans orient themselves toward a stimulus in order to use foveal vision on the most important stimuli.

2.4 Human Language

Language is a complex system used by humans to communicate. The system includes symbols, sounds, meanings and rules. Thought shapes language and language can help to shape thought. The *Whorfian Hypothesis of Linguistic Relativity* states that people with a language that provides numerous terms for distinguishing subtypes within a category actually perceive the world differently than those with a more limited collection (Burton et al., 2009). Language is ever evolving to express new ideas and concepts. Language consists of basic elements that form a hierarchical structure, the lowest level being phonemes. Phonemes are the minimum elements of sound that form coherent speech such as in English, vowels, consonants and how they are pronounced. Phonemes make up morphemes, which are the smallest units of meanings, like words. Morphemes make up phrases that in turn combine with more words to make up sentences. The rules that govern this combination are syntax. Syntax is a part of grammar, which is the system of generating acceptable language expressions. Semantics is used alongside syntax to understand what people are saying (Burton et al., 2009). Semantics are the rules behind the meanings of the morphemes, words, phrases and sentences. Language is generative and diverse, allowing humans to express themselves in an infinite number of ways. From a finite set of elements that make up language – phonemes - an infinite number of words, phrases and sentences can be generated. Human languages are forever growing, changing and evolving to each society's needs.

Various parts of the brain are specialised for language, for example, Wernicke's Area is a region of the temporal lobe that is associated with speech comprehension and Broca's Area is a region of the temporal lobe associated with speech production (Burton et al., 2009).

Humans acquire language in early childhood, after which it is believed that language cannot be acquired anymore (see the case of Genie – Transcript, 1997). Due to the complexity of human language, Noam Chomsky concluded that children learn language so quickly and efficiently because language is innate, that the human brain is biologically programmed for language. For the same reason, Steven Pinker argued that learning language is instinctive.

2.4.1 Written Language

Human language developed at least 45,000 years ago whereas written language only occurred about 3500 BC (Barton, 2007, Burton et al., 2009). Children instinctively learn to speak a language but must be taught how to read and write. Writing systems were built on to spoken language and it is argued that reading demonstrates "cultural engineering" (Burton, et al., 2009) as opposed to biological evolution.

It is believed that writing developed due to economic and social needs. Writing has been seen to develop as the formation of towns and growing complexity of societies also developed. Writing was invented independently in three different areas; the Fertile Crescent of Mesopotamia and Egypt, China and pre-Columbian America (Barton, 2007). The earliest known writing system from Mesopotamia being cuneiform, the first of which can be dated back to about 3500 BC. The initial uses of cuneiform were for commerce and trade (Barton, 2007). The Chinese invented pictograph writing, dating back to around 1200BC, was used initially for administrative and historical purposes. Finally, the Mayans invented their own system of writing, Maya script that dates back to around 300 BC.

Writing can be seen to have evolved out of cultures and societies needing more reliable methods of sending information, maintaining financial accounts, other administration purposes and keeping historical records, which would have become too complex for the human memory. Written language allows humans to communicate and share information across time and space, and to do so more accurately than what can be handled by the human capability of memory and story re-telling.

2.5 Reading

Reading is the combination of language and the use of our eyes to form a complex process. When a human reads, the eyes quickly and almost unconsciously move to acquire the text on display so that the brain can piece them all together and make logical sense out of it. Reading requires several cognitive processes to work together including visual information processing, word recognition, attention, and oculomotor control (Richter et al., 2005). The eyes move in such a way that very little is actually seen accurately. As the fovea is responsible for the small detail vision it is thus the fovea that is responsible for accurate vision of text (Rayner & Bertera, 1979).

Although Chomsky and Pinker convey the widely believed idea that language is innate in humans, we see that up to 30% of Australian children have difficulty learning to read even with normal schooling (Burton et al., 2009). The process of learning to read is less natural compared to learning to speak, as written language was invented as an addition to spoken language. Nevertheless, this complex skill has become a central, everyday task in modern society.

Reading written language requires complex interpretation of symbols in order to derive meaning from them. This is termed reading comprehension. Capable readers quickly and unconsciously recognise words; if there are letters written on a page in front of that reader, then they cannot help but read them as words (Reicher, 1969). This is described through the "Word Superiority Effect", the phenomena where humans more accurately recognise a letter in the context of a word than they do when presented as a singular letter. Humans recognise words because processing at the word level is happening even before they finish processing at the letter level (Reicher, 1969). If a word is not familiar it requires more cognitive processing in order to discern the meaning or the nature of the word. Reading, therefore, requires continuous education to ensure this processing time is minimised.

The SWIFT (Richter et al., 2005) and E-Z reader (1-5) models (Fisher et al., 1998) are two examples of cognitive models of reading. These studies attempt to describe the relationships between cognition and oculomotor functions in reading and how these processes effect, and essentially control, eye movements.

The two models are different in the way that they explain how these processes determine saccadic and fixation behaviour during reading. The E-Z Reader model states that attention is allocated serially to one word at a time, whereas the SWIFT model says that several words are processed at the same time so that attention is spatially distributed.

2.5.1 Reading Comprehension

Reading comprehension is the capacity to make sense out from written language. This requires assimilating symbols to make them into words, and then sentences and deducing meaning from the bigger picture they are conveying.

Carver (1992) presents the idea that the reading rate has a direct correlation to reading comprehension. He established 5 different "gears" which a human uses whilst reading which are based solely on reading rate (words per minute) but are representative of different cognitive processes. These "gears" are scanning, skimming, 'rauding', learning and memorizing (Carver, 1992). He proposes that people are constantly shifting between these gears, but that goals and cognitive processes drive this gearshift. Note that the word rates are based on typical US college students. Starting from the quickest reading method, there is scanning which is typically 600 wpm. Carver describes that the goal of scanning is to find target words. Next there is skimming which is used to get an overview of the text without reading the entire text which averages 450 wpm.

Carver also presents his idea of 'Rauding', which is a word made from reading and auding. Auding is the process of hearing, recognising, and interpreting spoken language (Carver, 1992). He states that reading and auding is essentially the same thing because they are comprehension of complete thought in sentences. Rauding is performed at 300 wpm and corresponds to normal reading. Learning is performed at 200 wpm and is used for remembering ideas and information. Memorizing is the slowest "gear" at 138 wpm; this is the process whereby recall is used, for example when you write something down to try to remember it later.

Although these ideas stray away from the fixation-saccade model of reading being examined in this study, they give insight into how reading comprehension is performed and provides background for establishing a model for determining reading comprehension.

2.5.2 Eye Movement During Reading

The eyes move in such a way as to assimilate the text on display in order to comprehend what is written. The eyes do not move in a smooth pattern taking in constant information, i.e. the eyes do not move from left to right, line by line. Instead eye movement is somewhat sporadic and complex, with the eyes moving at high velocity before stopping for a period to take in information before moving on again. This process was first described by Louis Émile Javal, in 1879.

These movements are now known as saccades and fixations respectively. A saccade is high velocity movement of the eye that transports the eye to a fixation. A fixation is where the eye stops and takes in information. These movements are not reserved simply for reading, they are observed in normal vision that requires any amount of detailed vision. It has been found that generally fixation duration during reading is 200-250 milliseconds, with a range of 100-500 milliseconds and saccadic movement is generally between 1 and 20 characters with an average of 7-9 characters (Rayner, 1983).

Most reading models, whether they be psychologically based or mathematically based, tend to focus on fixations; in particular the analysis of fixation duration and location. Some research has been done on saccade analysis in terms of the saccades being used for lexical processing time (Yatambe, et al., 2009). This is due in part to the belief that saccades are the means of bringing the eye to new fixations points. Little to no information is seen during a saccade so fixations are when the eye takes in information. However, saccades cannot be discounted, as lexical processing occurs during saccades (Yatambe et al., 2009). Yatambe et al. found that during long saccades, readers perform more lexical processing than during short saccades. They concluded that saccade duration should be included when calculating reading time.

In terms of what causes saccadic movement Engbert & Kliegl (2001) found that initiations of saccades are not completely driven by lexical processing, that in fact saccades can be autonomous.

Due to the complexity of eye movements and the variance that is observed not just between people but in the same person reading the same text (Rayner & McConkie, 1976) it is unclear whether eye movements are directly guided by high level lexical processing (Richter, Engbert, Nuthmann, & Kliegl, 2005) or by low level oculomotor processing.

These two aspects of processing are often assessed through research on factors that affect eye movement. The E-Z Reader and SWIFT models take into account lexical difficulty. Two properties of lexical difficulty commonly assessed are word frequency and word predictability. Furthermore, longer eye fixations have been observed for misspelled words (Rayner, 1998), which in essence is a product of word frequency and word predictability. Word predictability is the probability of guessing a word from the sequence of previous words of the sentence (Fisher, Reichle, Pollatsek, & Rayner, 1998). Word frequency can be computed easily as it is independent of context, but word predictability incorporates many aspects of a reader's knowledge of language and is strongly dependent on context. Word predictability must be estimated from experiments, obtained from incremental reading tasks (Fisher et al., 1998). These factors stipulate that fixations often occur on low frequency words as well as on words that are harder to predict in context. A further example of these factors can be seen through observation that in "normal" text about 80% of the nouns, verbs and adjectives in the text are fixated on and about 20% of the articles, conjunctions, prepositions and pronouns in the text are fixated on (Fisher et al., 1998). It is assumed that most functional words are easier to predict and are shorter words compared to content words such as nouns. This leads to the eye-mind hypothesis put forward by Just and Carpenter (1978) where they propose that a person will think about a word for exactly as long as they fixated on that word.

The factors that affect oculomotor movement are low-level information such as word length (Reilly & O'Regan, 1998). The longer a word, the more likely it is that a fixation will occur on it.

Within the SWIFT model for reading, the researchers set out certain quantitative measures for the factors that influence eye movements. These include; fixation duration, fixation probabilities, effects of word length versus word frequency, within word fixation positions, lag and successor effects and fixation time before word skipping (Fisher et al., 1998). These are interesting phenomena as we see that eye movement patterns are very complex in that words may be fixated on several times, saccadic movement happens in any direction, fixation duration is variable and words are quite often skipped altogether.

McConkie et al. (1988) showed that there is a general tendency for the eye to land around the middle of the word. They found that these landing positions of fixations were not affected by word length when they investigate the data in terms of prior fixations. The further material is presented from the fovea the less accurately it can be seen and correctly interpreted (Rayner & McConkie, 1976). Intuitively fixating in the middle of a word would provide the best accuracy for viewing the whole word. In 1981, O'Regan proposed the "convenient viewing position" hypothesis, which is that, for the reason just stated, people learn to fixate at the centres of words because this is optimal for viewing words. Furthermore, the likelihood of re-fixation increases as the fixation points become further from the "convenient viewing position" (O'Regan, 1984, McConkie et al., 1989).

2.5.3 Reading and the Fovea

Interestingly, there have been no apparent special adaptions of the human body for reading; the fovea developed a long time before written language was developed. In fact, on an evolutionary scale, the fovea first appeared in the temporal retina of fish, which humans eventually inherited (Azuma, 2000). So it must be the human brain's cognitive systems that are under evolutionary pressure to optimise the process of reading (Fisher et al., 1998).

The further a word is presented from the fovea the greater the decrease in ability to identify that word (Rayner & McConkie, 1976). This is why eye movement is characterised by saccades and fixations. The eye must move so as to orient the fovea for high definition vision in different places to assimilate information.

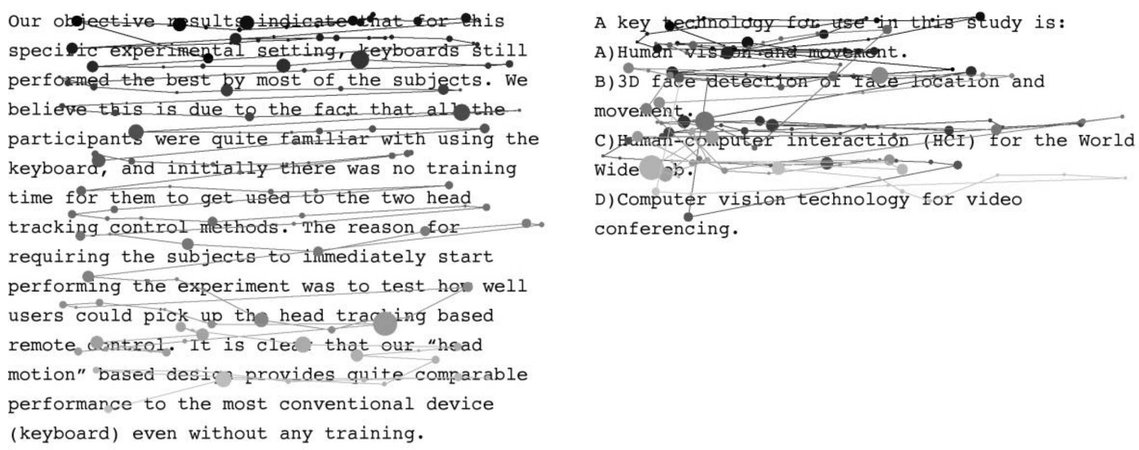
To differentiate the parts of the retina involved in vision there is first the foveal region where 2° of visual acuity extends across the fixation point. The parafoveal region is just outside the foveal region and it comprises 10° around the fixation point. The peripheral region is the rest of the visual field. Raynor and McConkie (1976) showed that it is the fovea and paraforveal regions that are critical for reading. In this experiment masks were placed over fixations points exhibited by participants to take out their foveal and then their parafoveal view. The results showed that although they could see words they would get the words wrong in the sentence. Longer fixation times were recorded when vision was masked and reading time increased markedly. The larger the mask, the greater the percentage of words incorrectly identified. They found that even though the participants were aware that words were in the parafovea and peripheral view they could not report what the words were. The results showed that masking of the fovea resulted in more severe reading difficulties compared to masking of only the parafovea. They concluded that information necessary for meaningful identification of a word is obtained from the fovea and near parafovea. Additionally, information such as that used to guide further eye movements to the next location to read is collected by the parafovea.

2.6 Analysing Eye Movement

Modelling eye movement patterns is challenging, as the trajectories are very complex even for simple sentences. Eye movement patterns can be quite different depending on what task is being performed. The task a person is performing can be predicted based on their eye movement (Simola et al., 2008). We see from

Fahey's data (2009) that even in reading tasks there are differences in eye movement patterns. There appears to be a difference in eye movement patterns when you visually compare easy gaze paths for reading the paragraphs to reading the questions, in most cases. This was not noted by Fahey but will be examined later in this thesis.

Figure 1 shows there is a difference between eye movements of participants recorded reading paragraphs compared to reading questions. This difference is expected, as when individuals answer questions they may study the questions and the answers more closely than they study the paragraph material the questions are based on.



```
Our objective results indicate that for this
specific experimental setting, keyboards still
performed the best by most of the subjects. We
believe this is due to the fact that all the
participants were quite familiar with using the
keyboard, and initially there was no training
time for them to get used to the two head
tracking control methods. The reason for
requiring the subjects to immediately start
performing the experiment was to test how well
users could pick up the head tracking based
remote control. It is clear that our "head
motion" based design provides quite comparable
performance to the most conventional device
(keyboard) even without any training.
```

```
A key technology for use in this study is:
A)Human vision and movement.
B)3D face detection of face location and
movement.
C)Human-computer interaction (HCI) for the World
Wide Web.
D)Computer vision technology for video
conferencing.
```

Figure 1: Eye movement trajectories of one participant; to the left is the eye movement whilst reading a paragraph and the right is the eye movement pattern whilst reading a question.

Salvucci and Anderson (2001) put forward the idea of eye movement as protocols, which they describe as "tracing", that is plotting eye movements to predictions of a cognitive model. The three tracing methods are target, fixation and point tracing. These three methods can be used differently in applications such as equation solving, reading and eye typing. They detailed that for reading, only fixation and point tracing are relevant. These three scenarios are an example of the different patterns that can be generated from eye movement. All three are essentially reading; when solving an equation you must first read the equation and when typing you must read the letters before you type. They generate quite different patterns where fixation trends tend to be focussed on the elements of the equation or the keyboard. There are many more applications of analysis of eye movement in tasks such as viewing faces, driving, watching television where complex patterns can be seen in eye movement. Salvucci and Anderson (1998) showed that "tracing" eye movement data is effective at interpreting the intent of eye movements using hidden Markov models. Tracing maps user actions and has been shown to generate accurate interpretations of these actions in areas such as eye typing and has been proposed to improve flexibility and design of eye based user interfaces (Salvucci, 1999).

2.6.1 Fixation Identification

Fixation identification plays an important part in the analysis of eye movement data and can have a large impact on results. The raw eye gaze data from the experiment consists of recordings at 60Hz, which has to be converted into fixation points.

During fixations, the eye does not stay completely still. The eye can make very small rapid movements, or occasional drifts and sometimes micro saccades to bring the eye back to the original position (Salvucci & Goldberg, 2000). These mean very little to high level analysis. Nevertheless they can make it harder to establish when fixations begin and end. Poor fixation identification may result in too few or too many fixations, which could in turn have dramatic effects on observations and further analysis.

Salvucci and Goldberg (2000) did a comparison study on fixation identification algorithms. They divided them into two characteristic groups: spatial and temporal, and then categorised the criteria based on these characteristics; Spatial: Velocity based; Dispersion based and area based; Temporal: duration sensitive and locally adaptive.

Analysis of fixation locations is often complex for several reasons: equipment noise, user variability and the size of the data set. One has to therefore infer from the data as best as possible. There are generally two major difficulties faced when interpreting eye movements: incidental fixations and off centre fixations (Salvucci, 1999). Although this has less impact on analysis of reading eye gaze patterns, it is important to keep in mind whilst modelling the data.

Incidental fixations are fixations that are accidental; these types of fixations are not of much interest when looking at reading eye gaze patterns. The process of reading is assumed to be driven cognitively so all fixation points are of some relevance. However, they are important to note for consideration in terms of random eye movements that are observed where the participant may have become distracted and lost track of where he/she was and then have to scan ahead to see how much is left to read.

Gaze points recorded by eye trackers can be off centre over visual targets. This creates off centre fixations. Also humans can fixate within 1° visual angle of the target and still encode information in the fovea. To add to this, eye trackers have a typical accuracy of approx 1°. Which further adds to the problem of mapping user actions to user intentions based on eye movement (Salvucci, 1999). This is a problem in terms of analysis of reading eye gaze patterns because essentially calibration needs to be done to bring the fixation points in line with what the participant is actually fixating on. If the points are not bought in line with actual fixation points there could be misinterpretation of the gaze patterns.

2.7 Markov Models

A Markov model is a stochastic model that assumes that the probability of moving to future states in the system depends only upon the present state of the system. In essence, the past states are irrelevant when considering the future states in the system.

A Markov model can be described as a finite set of all possible states $S_1..S_N$. The simplest Markov model is a Markov chain where the chain occupies one of these states at each of the time points, t = 1,2,3, ... At time t the process is in state S_i then at time t+1 the process moves to any possible state S_j with a certain probability, denoted p_{ij}. The probabilities p_{ij} $\forall$ i, j = 1, . . . N, are called the transition probabilities of the Markov chain and are arranged in what is called a matrix of transition probabilities (Isaev, 2006):

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1} & p_{12} & \cdots & p_{NN} \end{pmatrix}$$

The rows of the matrix can be read as the states from which a transition can be made. The columns can be read as states which can be transitioned to, meaning that the probabilities in any particular row in the of matrix sum up to 1.

A Markov chain is the process of generating all possible sequences of a finite length. Such sequences are denoted $x = x_1 \ldots x_L$. We denote P(x) as the probability of the sequence x.

This process can be called "observed" since the output of the process is a set of states at each instant in time, t, where each state corresponds to an observable event. This can be too restrictive to model certain problems, which is where HMM's come in.

2.7.1 Hidden Markov Models

Hidden Markov models (HMMs) are Markov models with unobservable or "hidden" states. In the Markov model described above, each state is directly visible to the observer and therefore the state transition probabilities are the only parameters. In an HMM the states are not directly observable but the output is visible. The essential difference between a Markov chain and an HMM is that every state in an HMM emits symbols, and hence there is no longer a one-to-one correspondence between the states and the symbols (Rabiner & Juang, 1986).

The basic overview of an HMM is that there are a set of states and associated with these states are transition probabilities, which are the probabilities of moving from one state to another (or staying in the same state). At each state there are a number of possible observations. Associated with these observations are output probabilities which are affected by the current state of that the system.

The five basic components that make up an HMM, λ, are (Rabiner, 1989, Rabiner & Juang, 1986):

1. N is the number of states ($S_1..S_N$) that compose the model
2. M is the number of distinct symbols from alphabet,∑, emitted from a state
3. π is the initial distribution of states π(i), $1\leq i \leq N$
4. A is the state-transition probability matrix, which is NxN; $a_{ij} = P(q_t = S_i \mid q_{t-1} = S_j)$.
5. B is the observation probability matrix which is NxM; i.e. $B_i(k)$ is the probability that state *i* emits symbol *k*, where $1\leq i \leq M$.

The HMM, λ, can therefore be written as: λ = (N,M,A,B, π) with a simplified notation of λ = (A,B, π) as N and M can be obtained from the matrices A and B.

A good example of a HMM is the occasionally dishonest casino, where sometimes a fair die is replaced by a loaded one. The casino does not tell you if the die is loaded or fair. In this example, if you see a sequence of rolls, you do not know which rolls used a loaded die and which used a fair die, because this is kept secret by the casino (Isaev, 2006). In other words the state sequence is hidden.

The three principal problems for HMMs (Rabiner, 1989):

1. Given an observation sequence and a model how do you efficiently compute the probability of the observation sequence from that model?

2. Given an observation sequence and a model, how do you choose the most probable state sequence that produced it?

3. Given a set observation sequence and a model, how do we find the most likely state transitions and symbol output probabilities?

The algorithms that address the first problem are known as the Forward Algorithm and Backward Algorithm. In this study, it is the Forward Algorithm that will be used to calculate the probability that a given sequence comes from one of a set of models, hence classifying the sequence. Problem 3 is the issue of training of the models using real data. The solution used in this study is the Baum Welch Algorithm which is an expectation maximization algorithm that employs the Forward-Backward Algorithm. The algorithms are discussed in more detail below. The solution to problem 2 is known as the Viterbi Algorithm. The Viterbi Algorithm is used to find the most likely sequence of states that generated an observation sequence. This is an important algorithm used when looking at HMM's and although not used in this study, future work may benefit from the use of the Viterbi Algorithm.

2.7.1.1 Forward Algorithm

The solution to this problem 1 is called the Forward Algorithm. This solution is particularly useful because if given two models λ_1 and λ_2, the Forward Algorithm can be used to find the most probable model that a particular sequence came from, which is the process of classification.

The Forward algorithm is used to calculate the probability of an HMM (Rabiner, 1989). For a long observation sequence it is not realistic to simply enumerate all possible paths because of the sheer number of them. This is where the Forward Algorithm comes in. The algorithm computes a set of forward probabilities which provide the probability of ending up in any particular state given the first k observations in the sequence.

We can define the forward variable $\alpha_t(i) = P(O_1, O_2, \ldots, O_3, S_t, = i \mid \lambda)$. Where $\alpha_t(i)$ is the probability of observing the sequence $O_1, O_2, \ldots, O_T$ such that the state S_t is i. The algorithm for calculating the Forward probability (Rabiner, 1989):

1. Initialization: $\alpha_1(i) = \pi_i b_i(O_1)$

2. Induction: $\alpha_{t+1}(j) = \left(\sum_{i=1}^{N} \alpha_i(i)\alpha_{ij}\right) b_k(O_{t+1})$

3. Termination: $P(0|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$

The Forward Algorithm is also used in the process of solving problem 3 where it is used in the method of training HMMs.

2.7.1.2 Parameter Estimation

One of the most difficult problems faced when using HMMs is that of specifying the model. That is, the definition of the states and their connectivity as well as the assignment of parameter values and the transition and emission probabilities.

Training sequences are used to estimate the parameters of a given HMM, but parameter estimation is dependent on whether the paths are known or unknown for the sequences.

If the paths are known then it is simple to estimate the probability parameters, because they are directly known. We can calculate the state transition probabilities via $a_{ij} = \frac{a_{ij}}{\sum_l a_{il}}$ and the emission probabilities are calculated via: $B_j(k) = \frac{B_j(k)}{\sum_b B_j(b)}$ (Isaev, 2006).

This is very straight forward, but we do not always have the state paths for training sequences. So how do we calculate the transition and emission probabilities without this knowledge? Clearly the above formulas cannot be directly used.

Baum Welch Algorithm

Training of the HMM is the important last problem however, unlike above, we rarely know the states from which observations came. This is where the Baum-Welch algorithm comes in.

Given a sequence of observed symbols $x = x_1, x_2, x_3, . . ., x_L$ we would like to choose parameter values that maximize the likelihood $P(x_1) \times ... \times P(x_L)$. The method commonly used to find $P(x_1) \times ... \times P(x_L)$ is the Baum-Welch training algorithm. This algorithm makes use of the Forward-Backward algorithm. The problem with the Baum-Welch algorithm is that it may converge to a point close to a point of local maximum and not a global maximum point.

As described by Rabiner (1989), the Baum-Welch algorithm works by first defining two values $\xi(i,j)$ and $\gamma_t(i)$. $\xi(i,j)$ is the probability of state i at time t and moving to state j at time t+1 and is defined as:

$$\xi(i,j) = \frac{\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}$$

$\gamma_t(i)$ is the probability that in state i at time t, given the observation sequence and is defined as:

$$\gamma_t(i) = \sum_{j=1}^{N}\xi(i,j)$$

The Baum-Welch algorithm uses these two values to recalculate the parameters of the model:

1. The initial state distribution is the expected frequency in state i at time t=1: $\pi_i = \gamma_1(i)$
2. the state transition matrix entry a_{ij} is the expected number of transitions from state i to state j divided by the expected number of transitions from state i: $a_{ij} = \frac{\xi(i,j)}{\sum\gamma_t(i)}$
3. The observation emission of symbol k at state j, $b_j(k)$, is the expected number of times in state j of observing symbol k divided by the expected number of times in state j: $b_j(k) = \frac{\sum_{t,O_{t=k}}\gamma_t(i)}{\sum\gamma_t(i)}$

2.7.2 Classification Using HMMs

Classification is common in machine learning and pattern recognition; it refers to the procedure whereby given a set of classes, an input is assigned to one of the classes based on the classifier algorithm. Since classification requires that inputs be classified based on behaviour learnt from a training set of properly labelled sequences, it is termed a supervised procedure and is a form of pattern recognition.

In a broader sense, classification is an important task in machine learning in general. Hidden Markov Models are a type of statistical machine learning used for pattern recognition. It is the solutions to first and third principal problems of HMMs (see first part of the HMM explanation) that can be used to classify sequences data. Firstly, a model λ = (N,M,A,B,π) is designed with some knowledge of the area to which it is being applied. The choice of structure is not straightforward and may involve some trial and error. Once the model has been chosen it must be trained using training data. The Baum-Welch algorithm is used for this purpose. Then the forward algorithm can be used to calculate the probability that a given sequence came from a particular model.

Several different models can trained individually from data within the field of the classification dominion to create specialised models that recognise sequences from those particular domains. Then the likelihood that a given sequence came from its corresponding model can be calculated to give an estimate of how well the model classifies the data.

2.7.3 HMMs and Eye Movement

HMMs are quite relevant for modelling real life events, it is common to obtain observations from the world but not be able to observe the underlying structure that explains those observations.

Many different HMM can be constructed from the same observations but the challenge is finding one that fits well. The problem is how to construct an HMM to explain the observed sequence of events when the underlying structure in not known, i.e. like in the occasionally dishonest casino example where it is unknown whether the die is fair or loaded.

HMM's are commonly used in speech recognition (Rabiner, 1989), hand-writing recognition, part-of-speech tagging, and bioinformatics (Isaev, 2006). Rabiner (1986, 1989) provides thorough tutorials on HMMs and their application in speech recognition.

Salvucci likens eye movements conceptually to these speech and handwriting recognition schemes in the way that recognition systems translate a user's speech input to the most likely interpretation of the persons intentions. Salvucci has made many models on eye movement data using HMMs, including identification of fixations and saccades (Salvucci & Goldberg, 2000), inferring intent of fixation points (Salvucci & Anderson, 1998; Salvucci, 1999) and application of tracing protocols to equation solving, reading and eye typing (Salvucci & Anderson, 2001). In terms of application to reading, the HMM's were based on the E-Z Reader models 3 and 5, see Fisher et al., (1998) for more information.

Simola et al. (2008) found that they could make predications on what tasks were being performed by a participant using a discriminant HMM (dHMM). The dHMM consisted of three states corresponding to three types of eye behaviour. The first behaviour is scanning which was observed at the beginning of the tasks. Then there is reading eye movements followed by decision.

HMM's have been used in real time classification of gaze data to recognise if an individual is reading or not reading (Gustavsson, 2010). In the experiment it was shown that HMM's not only perform well at classifying whether an individual is reading or not but perform better than neural network classifier designed to perform the same task. The HMM classifier produced 95% classification accuracy.

Kozek (Kozek, 1997) investigated the use of HMMs in diagnosing schizophrenia, where the HMMs were used to classify whether the participant was schizophrenic or not. Kozek's experiment was composed of participants who did not have schizophrenia and medicated schizophrenics. The participants viewed a series of different images, which consisted on an abstract art, a neutral face, a happy face and a sad face. The HMMs were used to classify the eye gaze patterns based on the image that the participant viewed. The classification of the eye gaze patterns based on viewing the abstract art versus the neutral face gave 80-90% accuracy of classification. The classification levels were decreased to a maximum of 68% classification when the gaze patterns generated from viewing the neutral face versus the happy face, as there is less difference between the two images. The least difference is between the neutral face and sad face, where the classification accuracy became around 60%. These divisions of separation of the types of material shown to participants are similar to what will be used in this experiment where the content of the paragraphs will be used to define the divisions of separation of material.

Looking at vision from another perspective, Rimey and Brown (1990 , 1991) have had success with the use of HMMs and augmented HMMs in the field of computer vision, in particular object vision.

Chapter 3

Preliminary Work

The initial phase of the project was concerned with analysing aspects of the data sets and structuring the eye gaze data for subsequent analysis. This phase included three steps; calibration of fixations, finding the words that a participant had fixated upon in each paragraph and a textual analysis of the paragraphs' content. These three steps help in the further analysis, which is detailed in subsequent sections.

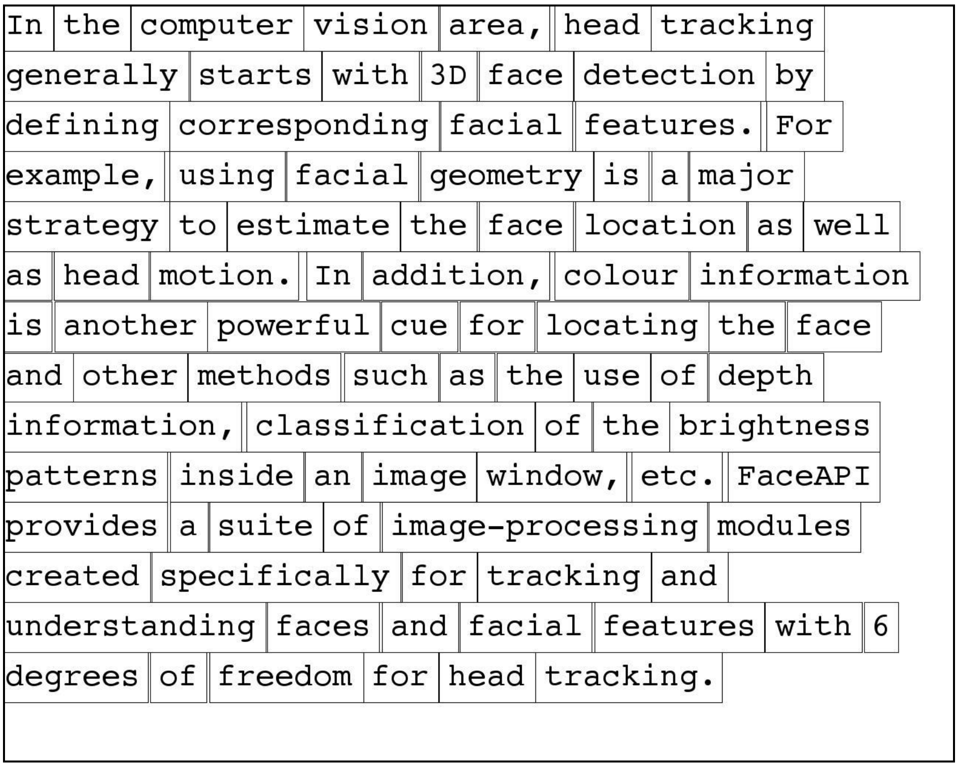
3.1 Find Fixated Words

Retro-fitting the eye gaze data recorded from each participant to corresponding screenshots was done to find the fixated words. This involved differentiating the parts of the screen into word boxes and the white space around the word. The fixation circles were generated from the eye gaze data and each fixation circle was checked to see if it landed in a word box. The result is a list of word boxes that a fixation had occurred in, which gives the sequence of words that a participant fixated on in a given paragraph.

3.1.1 Extraction of Words from Image

The first part of the study was investigating data that had been collected from previous work (Fahey, 2009). In this experiment the eye gaze of participants was monitored as they read text from a monitor. The data from the experiment consists of eye gaze points that were collected at intervals of 1/60th of a second. Each time point has x, y coordinates corresponding to where the participant was looking at that time as recorded by the eye gaze tracker. For the Fahey study, these x, y coordinates correspond to pixel values in the range of 1280x1040, as this was the size of the screen the participants viewed. Displayed to the participants was a series of paragraphs and questions, which are presented all in the same font as 1280x1040 black and white images. For the Sharma study (2011), the pixel range is 1680x1050; however, the paragraph of text is shown in a much smaller box on the screen and uses a smaller font compared to the Fahey data set. It is only this part of the screen that is of relevance in this study so points that lie out of the text's range are disregarded. This is to keep the two data sets consistent as with the Fahey data set only text is shown to the participants, and the text takes up the entire screen. Again all paragraphs in the Sharma set are displayed in the same font and same size.

An important difference between this study and Fahey's study is that the words that were "fixated" on are being assessed as opposed to the fixations alone. The previous work looked solely at statistical measures such as the average number of fixations, the average horizontal and vertical movements, etc., (Fahey, 2009) to see if these provide insight to participant comprehension. Also, work has been done where the screenshots that were presented to participants were partitioned into boxes and the eye gaze is characterised by the movement between these boxes (Vo et al., 2010). These movements were classified as forward and backtracking. In that experiment a 4x5 grid was used and produced about 80% classification levels using artificial neural networks as the classifier.

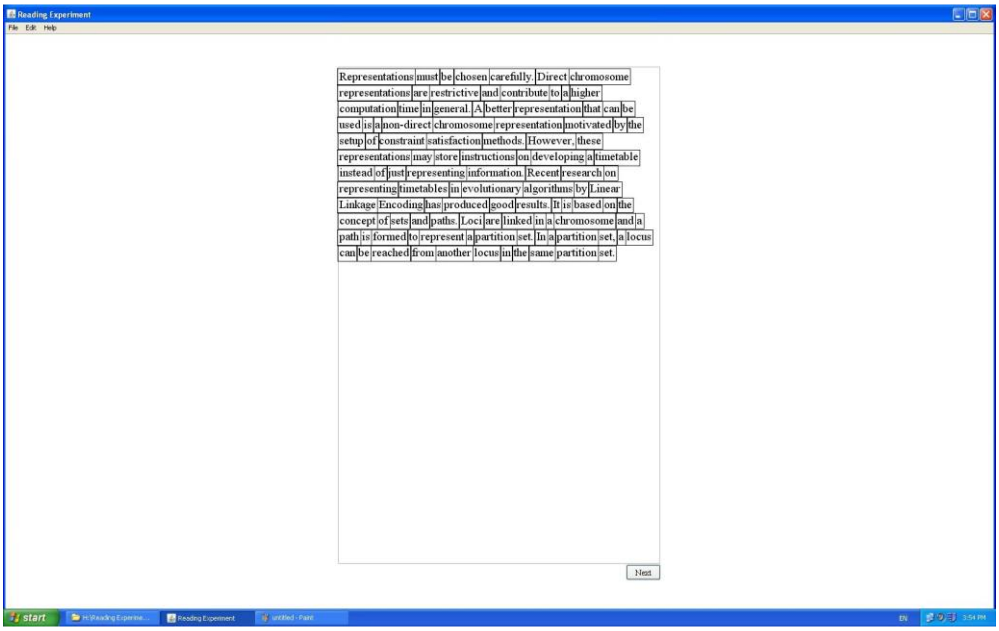


In the computer vision area, head tracking
generally starts with 3D face detection by
defining corresponding facial features. For
example, using facial geometry is a major
strategy to estimate the face location as well
as head motion. In addition, colour information
is another powerful cue for locating the face
and other methods such as the use of depth
information, classification of the brightness
patterns inside an image window, etc. FaceAPI
provides a suite of image-processing modules
created specifically for tracking and
understanding faces and facial features with 6
degrees of freedom for head tracking.

Figure 2 : Graphical representation of the boxes that are drawn around each word in a paragraph from the Fahey data set.

In order to discover the words that a participant fixated on, retro-fitting of the data had to be completed. To infer which word is fixated upon the coordinates of each of the words in the paragraph on the screenshot had to be mapped out. The mapping was done by "drawing" boxes around each presumed word on screen. These boxes' coordinate values for where they lie on the screen are stored. The boxes are kept in a list and given an index, for instance the first word of the paragraph has index = 0 and the subsequent word has index = 1, and so on until the last word which will have the index = total number of words in paragraph - 1.

The contents for the paragraph are also in plain text files which can be read in as a string. The index obtained by "boxing" the words on the screenshots corresponds to the actual word that the box contains, and hence we now have a way to find the words that an individual fixated upon when we map fixations to the boxes.



Representations must be chosen carefully. Direct chromosome
representations are restrictive and contribute to a higher
computation time in general. A better representation that can be
used is a non-direct chromosome representation motivated by the
setup of constraint satisfaction methods. However, these
representations may store instructions on developing a timetable
instead of just representing information. Recent research on
representing timetables in evolutionary algorithms by Linear
Linkage Encoding has produced good results. It is based on the
concept of sets and paths. Loci are linked in a chromosome and a
path is formed to represent a partition set. In a partition set, a locus
can be reached from another locus in the same partition set.

Figure 3: Graphical representation of the boxes that are drawn around each word in a paragraph from the Sharma data set.

3.1.2 Fixation Identification

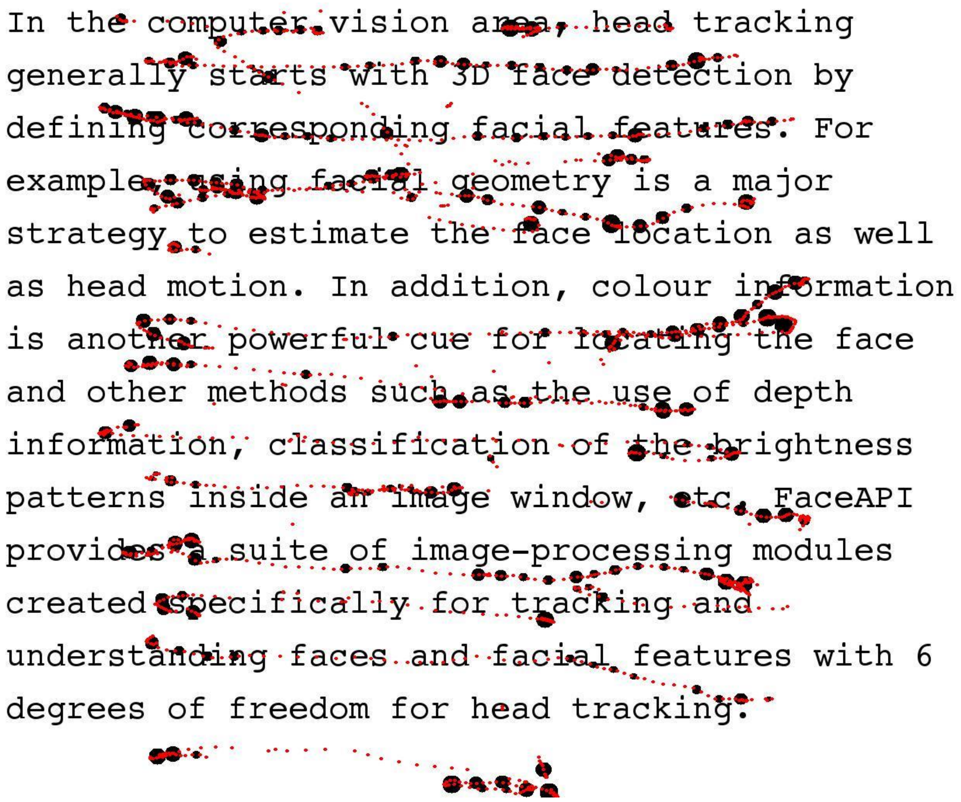
The raw data from the experiment consists of x, y coordinates for where the eye was looking at that point in time. In both experiments, the eye tracker measured eye movement at 60Hz so the eye position is captured around every 17 milliseconds. The eye could have been doing one of two things; bringing in information (in a fixation), or jumping to the next fixation point (in a saccade). There is an average fixation duration of about 200–250 ms and a range from 100 ms to over 500 ms, so on average about 12 to 15 gaze points make up a fixation with a range of about 5 to 30 fixations in each trial. Furthermore, the eye never stays still, even in a fixation. During a fixation the eye quivers and there are microsaccades, but these are still considered to be part of the fixation because the eye is taking in information. During saccades information is not taken in, this is why we do not see the world swivel by.

The only information of interest is the fixation data, as we want to obtain which words the participant fixated on. So the main problem is deciding which gaze points make up a fixation and which ones are part of a saccade. There are many methods available to identify fixations as described by Salvucci (2000).

In order to differentiate between the two types of movement, each x-y point is compared to the existing fixation circle. The distance between the point and the fixation's centre is calculated and if the distance is less than a certain pixel tolerance then the point is classified to be part of the same fixation. The fixation circle is then recalculated to include the new point, this involves recalculating the minimum enclosing circle of the fixation. If the distance between the two points is greater than the tolerance then the new point is classified to be part of a saccade, and the current fixation is "closed off". To find the next fixation, a point is compared to its predecessor and if they are within the given threshold then a fixation circle is started. This form of fixation identification is called a dispersion threshold technique and has been shown to be one of the most accurate and robust techniques available (Salvucci, 2000).

Figure 4 shows raw gaze points recorded from one participant, these are the red dots, and the black circles indicate the fixations which were calculated using a 13 pixel threshold. There are quite visible hot spots that are undoubtedly fixations and there are trails of small (red) dots which are quite likely to be saccades. It is when those trails include very close gaze points, but no clear groupings of gaze points are seen, that it becomes unclear whether the gaze points form a fixation or not.

It is important to note that not all fixations may lie on words; this could be due to the fact that the data was not calibrated well or the algorithm for determining a fixation is too lenient and is recording a fixation point rather than as a saccade or it could just be that an individual actually looked at the white space between two lines or between two words. In any case, these fixation points are discounted, in some areas of the analysis presented in this study.



In the computer vision area, head tracking
generally starts with 3D face detection by
defining corresponding facial features. For
example, using facial geometry is a major
strategy to estimate the face location as well
as head motion. In addition, colour information
is another powerful cue for locating the face
and other methods such as the use of depth
information, classification of the brightness
patterns inside an image window, etc. FaceAPI
provides a suite of image-processing modules
created specifically for tracking and
understanding faces and facial features with 6
degrees of freedom for head tracking.

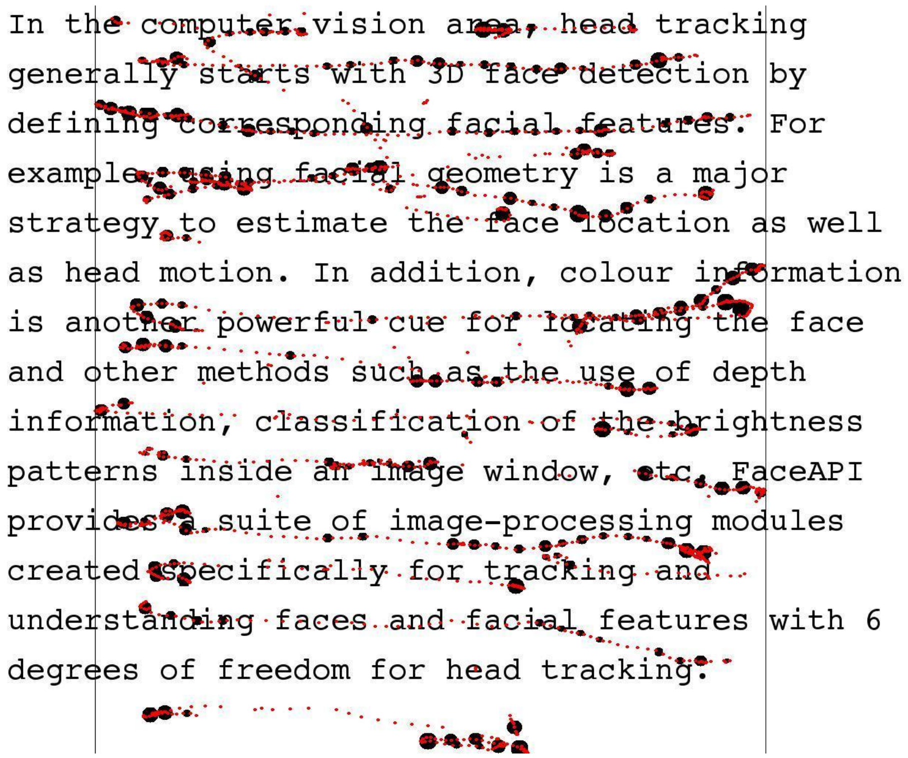
Figure 4: The raw gaze data and the calculated fixations circles plotted for one participant on paragraph 1.

3.1.3 Word Retrieval

After finishing the work described above we essentially have a screen that is partitioned into the segments that contain text and other segments that are white space. From the eye gaze data, we have obtained the fixations for each paragraph. The word retrieval is about putting the two types of information together. To do this, each fixation circle is tested to see if it lies in one of the word boxes. If it does, then that box index is added to a list of fixated boxes. The resulting list of box indices is then cross referenced back to the text words, as mentioned above, and a list of words is generated which is the words that the participant fixated on.

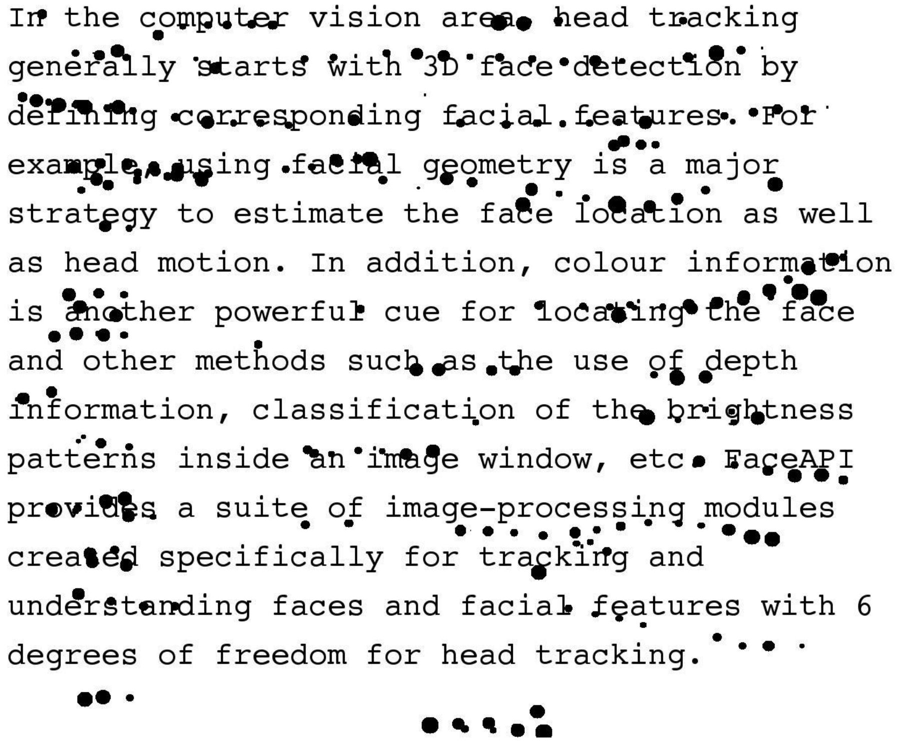
3.2 Calibration

Visual analysis of the participants' eye gaze mapped onto the screenshots shows that it looks as though the points have been squashed together onto a smaller screen size. This is consistent among participants as well as all paragraphs. Automated calibration on the points was designed and performed which in essence spreads the points further in width. This calibration involved finding the horizontal eye movement limits (as shown in Figure 5) and then performing a linear transformation of the fixations based on these limits.



In the computer vision area, head tracking generally starts with 3D face detection by defining corresponding facial features. For example, using facial geometry is a major strategy to estimate the face location as well as head motion. In addition, colour information is another powerful cue for locating the face and other methods such as the use of depth information, classification of the brightness patterns inside an image window, etc. FaceAPI provides a suite of image-processing modules created specifically for tracking and understanding faces and facial features with 6 degrees of freedom for head tracking.

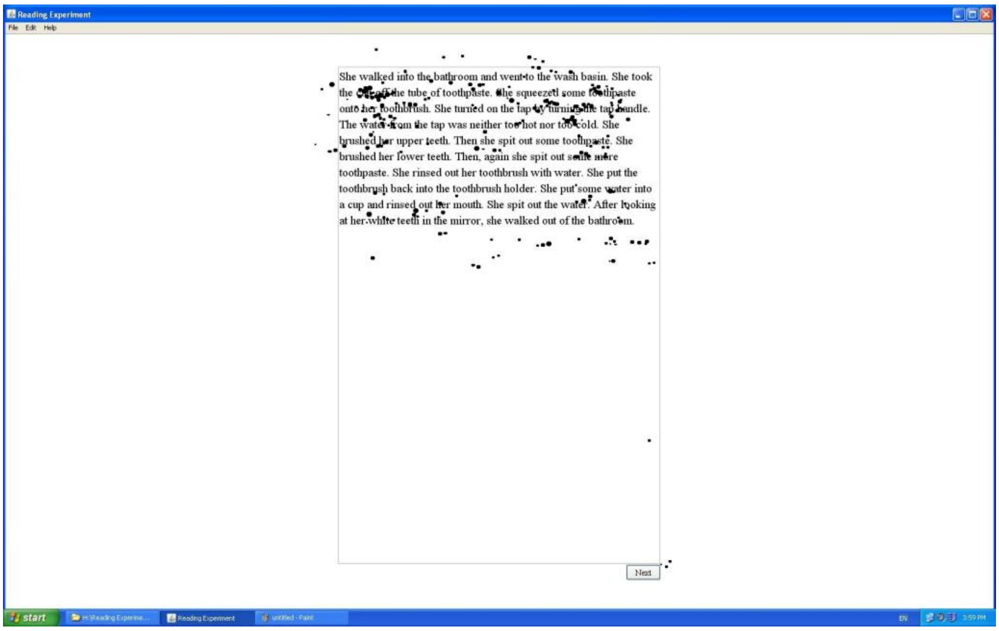
Figure 5*:* The vertical lines show the limits of the horizontal eye movement.



In the computer vision area, head tracking generally starts with 3D face detection by defining corresponding facial features. For example, using facial geometry is a major strategy to estimate the face location as well as head motion. In addition, colour information is another powerful cue for locating the face and other methods such as the use of depth information, classification of the brightness patterns inside an image window, etc. FaceAPI provides a suite of image-processing modules created specifically for tracking and understanding faces and facial features with 6 degrees of freedom for head tracking.

Figure 6: The results of the calibration.

The calibration of the fixations gives more realistic data to work with. It must be noted however, that the calibration was only performed on the Fahey data set and the calibration was only in the horizontal axis. The points at the bottom of the screenshot in Figure 4 are believed to be an artefact.



She walked into the bathroom and went to the wash basin. She took the cap off the tube of toothpaste. She squeezed some toothpaste onto her toothbrush. She turned on the tap by turning the tap handle. The water from the tap was neither too hot nor too cold. She brushed her upper teeth. Then she spit out some toothpaste. She brushed her lower teeth. Then, again she spit out some more toothpaste. She rinsed out her toothbrush with water. She put the toothbrush back into the toothbrush holder. She put some water into a cup and rinsed out her mouth. She spit out the water. After looking at her white teeth in the mirror, she walked out of the bathroom.

Figure 7: Example of the fixations plotted on the screenshot for the Sharma data set.

Further, as time did not permit, the Sharma data set was not calibrated. Some participants were excluded from the results as the data recorded from them was considerably offset. Although the data from these participants would likely prove to be useful, the data would give inaccurate results in its current state and so the participants were excluded on this basis. It is important to note that the results generated from the Sharma dataset are not calibrated at all and so the results of the analysis performed on them is less certain than that of the Fahey data set since the fixations have been calibrated for that dataset.

3.3 Analysis of the Text

A key difference between this study and previous work is that analysis of the words that were fixated upon. As part of the preliminary analysis the content displayed to the participants is assessed in several ways. As part of the previous experiment (Fahey, 2009) the paragraphs chosen were classified into one of two categories; relevant and less-relevant. The relevant paragraphs are believed to provide more information than the less-relevant paragraphs based on their source and therefore are harder to comprehend.

To analyse the content of the paragraphs, a program was written to perform statistical analysis on the contents. It is important to point out that the words were all converted to lower case and the punctuation has been stripped out. This is a design aspect that was chosen to keep the analysis simple, though it is foreseen that in subsequent iterations, both capitalisation and punctuation should be analysed.

Further, the physical aspects of the words are analysed, that being the word frequency and word length. It has been shown previously that these characteristics can affect eye movement during reading. Both the E-Z Reader (Reichle et al. 1998) and SWIFT (Engbert et al. 2005) models take into account lexical difficulty. Two properties of lexical difficulty commonly assessed are word frequency and word predictability. Word frequency can be computed easily as it is independent of context, but word predictability incorporates many aspects of a reader's knowledge of language and is strongly dependent on context. These factors stipulate that fixations often occur on low frequency words as well as on words that are harder to predict in context. With these points in mind the word frequencies of the content displayed to participants are calculated. A scoring method is developed in order to test the hypothesis that word frequency will be a factor that affects eye movement during reading of text. As discussed, word predictability is a much more difficult factor to assess. The research itself is centred on differentiating between types of text based on the fact that some types of text are easier to read than others. Predicting a word in a piece of text that is easy to read would be more accurate and faster than that in a piece of text that is difficult to read. If a word is predictable in context then it is likely to be skipped with a saccade and a fixation will not occur on it, with the opposite being true for a word that is not predictable in context. Further, if a word was wrongly predicted, the eye will be bought back to that word in order to take it in. We would expect that different eye gaze patterns would be produced from different forms of text based on these predictability factors. The paragraphs were subjectively chosen as being relevant or less-relevant, or hard or easy in the case of the Sharma experiment. We ran readability measures over each paragraph and question in order to confirm this and separate the paragraphs based on this factor.

Lastly, it has been found that one of the factors that affects oculomotor behaviour is word length (Reilly & O'Regan, 1998). Here the longer a word is the more likely it is that a fixation will occur on it. In order to account for this in the study the length of every word in the content displayed to participants was calculated and used to form another scoring method in order to test this.

3.3.1 Paragraph Length

An important characteristic of the paragraphs is their length. The tables below summarise the respective lengths and averages for both experiments.

3.3.1.1 Fahey Experiment

The following table details the lengths of all 10 paragraphs and 5 questions from the Fahey study.

Table 1*:* Paragraph and question length for the Fahey dataset

Relevant Paragraph	Length	Less-Relevant Paragraph	Length	Question	Length
1	94	6	62	1	61
2	101	7	64	2	70
3	92	8	47	3	86
4	108	9	82	4	55
5	100	10	78	5	39
Average	**99**	**Average**	**67**	**Average**	**62**

As seen above, the paragraphs are not of the same length. There is a range from 47 words to 108 words in each paragraph. Notably, less-relevant paragraphs have about 33% less words on average than relevant paragraphs, with ranges that do not overlap.

Table 2: Comparison of paragraph lengths of Fahey dataset.

	Relevant	Less-Relevant	Questions
Length (in words)	99	67	62
Range (in words)	92 to 108	47 to 82	39 to 86

The paragraph length is an important aspect that has implications on the analysis of the eye gaze data. Since the paragraphs are not all the same length, all results must be normalised for the first data set.

3.3.1.2 Sharma Experiment

The following table details the length of all 6 paragraphs from the Sharma Study.

Table 3 : Hard Paragraph Lengths vs Easy Paragraph Lengths

Hard Paragraph	Length	Easy Paragraph	Length
1	126	1	121
2	120	2	112
3	113	3	123
Average	**120**	**Average**	**119**

The paragraph lengths in the Sharma experiment are relatively equal, with virtually the same number of words on average between the two categories, 120 to 119 words, for hard and easy paragraphs respectively.

3.3.2 Word Frequency

3.3.2.1 Fahey Experiment

For the Fahey data set, two increments were used in the analysis of the text. The first looked principally at the paragraphs alone and the second looked at the combination of the paragraphs and the questions.

Paragraphs

The frequency of each word was tallied and the results are shown below in Table 4 and graphically in Figure 8. There is a range of 1 to 57 for different frequencies with 18 frequency bins in total. We can observe that the majority of words occur at low frequencies (<5).

Table 4 : Summary of frequency of word in the paragraphs

Frequency	Number Words
1	241
2	63
3	24
4	9
5	9
6	4
7	3
8	3
10	1
12	2
15	1
16	1
17	1
21	1
22	1
25	1
32	1
57	1

There are 367 different words that make up the total of 828 words for the paragraphs. We can observe that about 66% of those words occur only once.

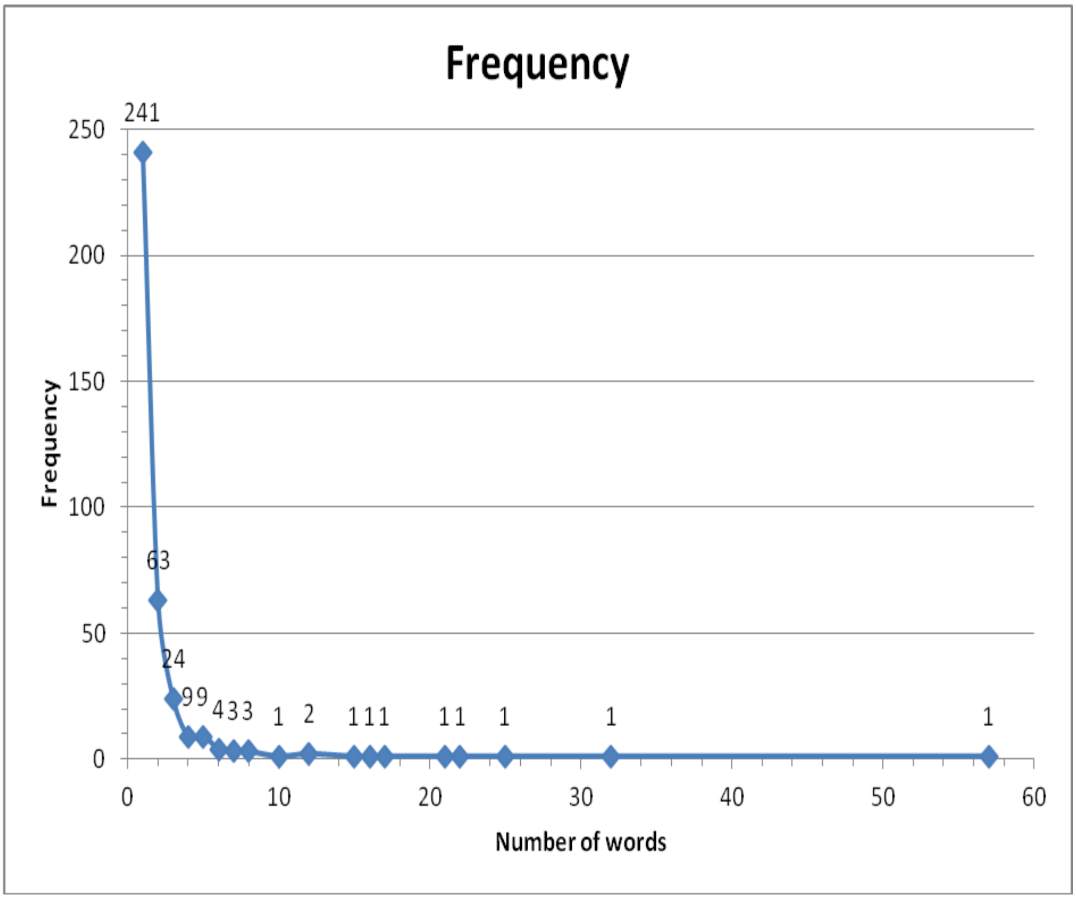


Figure 8: Graphical representation of frequency of words

Table 5: Summary of highest frequency words

Word	Frequency
camera	10
tracking	12
in	12
of	15
for	16
is	17
a	21
and	22
head	25
to	32
the	57

Table 5 summarises some of the higher frequency words; here we can observe that "the" is the most frequency word by far, with a frequency of 57. This means that the word "the" is approximately 16% of the total different words (367) and about 7% of the total number of words (828). The majority of the most frequent words are as expected, being common joining words such as "and", "a" and "in". We can also observe that the words "head" and "tracking" are amongst the most frequent words. Since the content of the paragraphs is about head tracking and eye gaze these words are expected to be of high frequency.

The analysis shows that there is a great differentiation between the frequencies of words, but that majority of words occur quite infrequently (<6 times).

Paragraphs and Questions

In order to compare the questions to the paragraphs, a text analysis is performed on the content of both the Paragraphs and the Questions combined. The results are quite similar to the analysis on the paragraphs alone, as expected. The frequency of each word was tallied and the results are shown below in Table 6 and graphically in Figure 9. The range of frequencies has increased to be between 1 and 79 for different frequencies with 22 frequency bins in total. We still observe that the majority of words occur at low frequencies (<5).

Table 6*:* Summary of frequency of word in the paragraphs

Frequency	Number of Words
1	253
2	76
3	26
4	14
5	17
6	8
7	3
8	5
9	2
10	2
11	1
12	1
13	1
17	1
18	2
20	1
21	1
25	1
31	1
32	2
39	1
79	1

Now we see that there are 420 different words that make up the total of 1139 words for the paragraphs and questions combined.

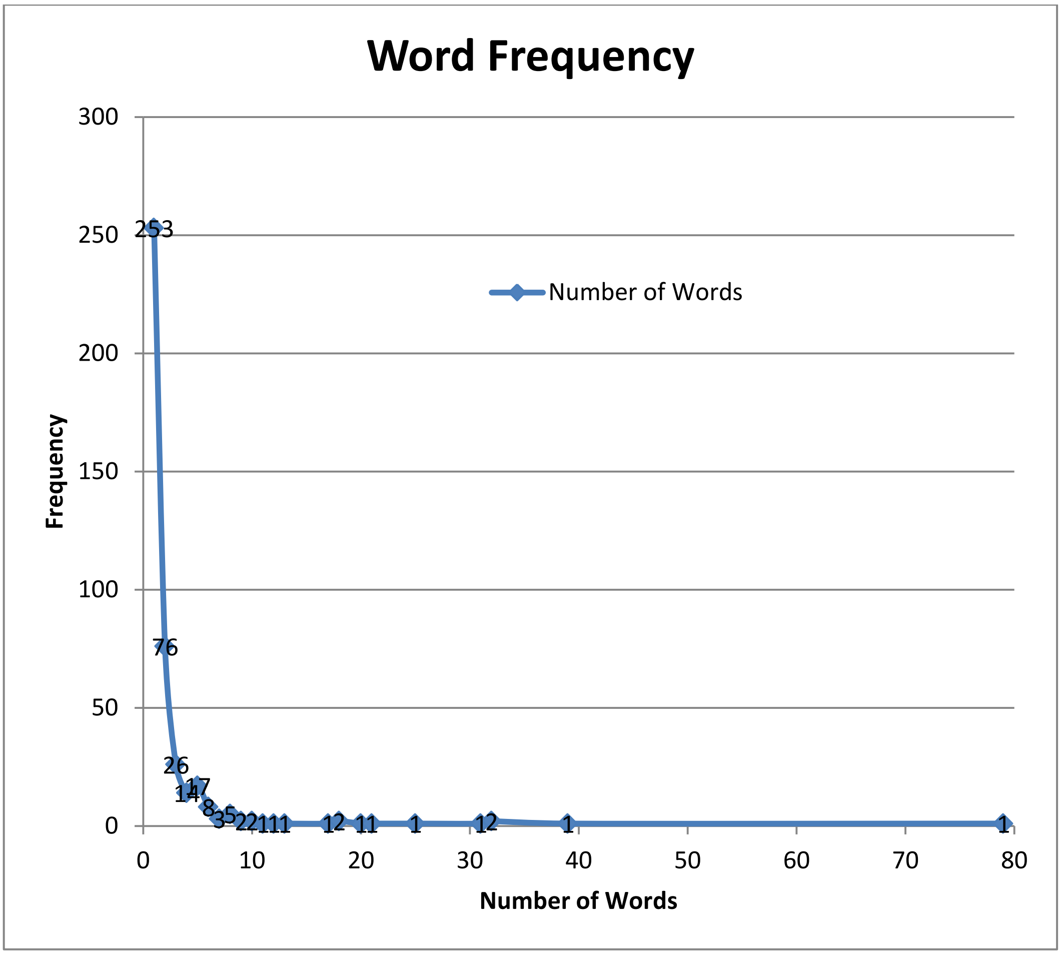


Figure 9*:* Graphical representation of frequency of words

The table below summarises some of the higher frequency words; here we can observe that "the" is the most frequent word by far, with a frequency of 79. The majority of the most frequent words are still as expected, being common joining words such as "and", "a" and "in". We can also observe that again the words "head" and "tracking" are amongst the most frequent words. Further many of the same words compose the most frequent list as did in the paragraphs only analysis. This is expected as the questions are based on the paragraphs and we would expect that the same focus words would be used.

Table 7*:* Summary of highest frequency words

Word	Frequency
user	10
flicking	10
was	11
motion	12
camera	13
tracking	17
in	18
that	18
is	20
for	21
of	25
a	31
and	32
head	32
to	39
the	79

3.3.2.2 Sharma Experiment

The frequency of each word was tallied and the results are shown below in Table 8 and graphically in Figure 10. There is a range of 1 to 52 for different frequencies with 17 frequency bins in total, which is almost identical to the paragraph only analysis for the Fahey data set. We can observe that the majority of words occur at low frequencies (<5).

Table 8: Summary of frequency of word in the paragraphs

Frequency	Number of Words
1	267
2	52
3	18
4	9
5	4
6	1
7	4
8	1
9	1
10	1
11	1
13	1
16	1
19	1
20	1
21	2
52	1

There are 366 different words that make up the total of 715 words for the paragraphs. We can observe that about 73% of those words occur only once which is an increase of about 10% compared to the Fahey data set which contains more words. Analysing this further we look at the words that occur most frequently.

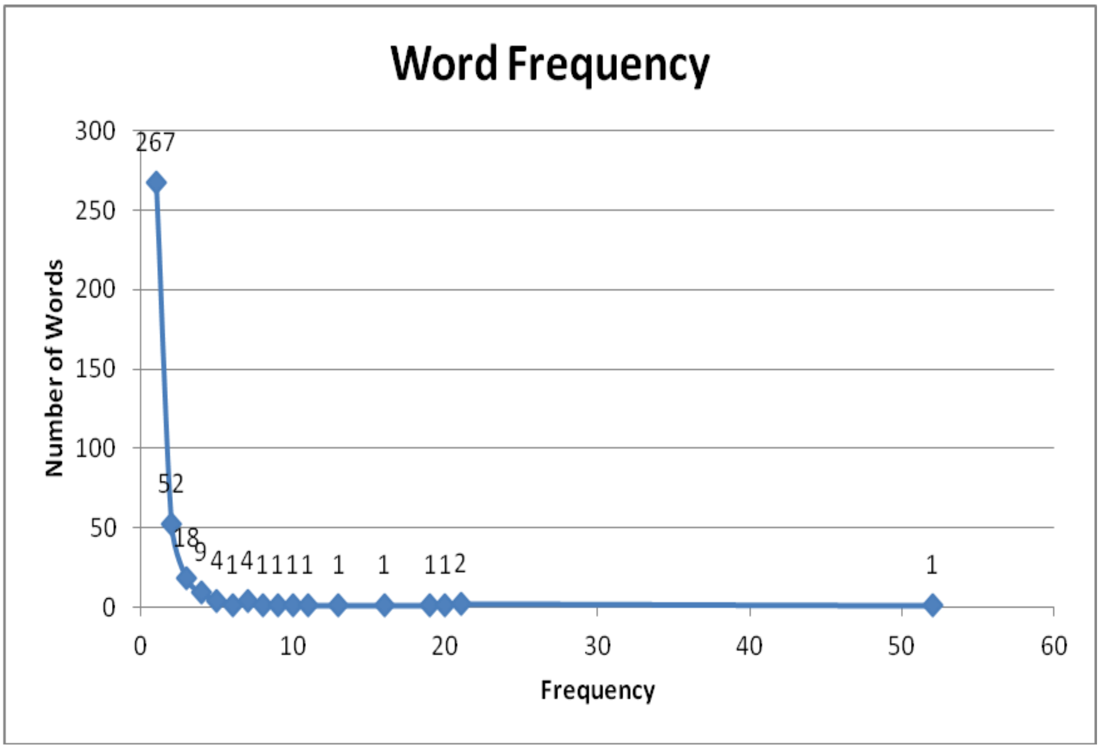


Figure 10: Graphical representation of frequency of words

As with the Fahey data we see that "the" is the most frequent word by far, with a frequency of 52. However, unlike in the Fahey data set we see that all of the most frequent words are common joining words such as "and", "of" and "to".

Table 9: Summary of highest frequency words

Word	Frequency
from	8
are	9
is	10
by	11
she	13
in	16
a	19
to	20
and	21
of	21
the	52

In this experiment, the paragraphs presented to the participants were all about different subject matter. As a result, we see more words occurring either once or at low frequency even though there are less words in the total paragraphs' content. This means that some words that may actually be quite normal, or commonly used words, are occurring at low frequency simply because the different subject matters speak about completely different topics. As a result these commonly used words are behaving similarly to uncommon words which are also occurring at the same frequency. This is the downside of the analysis for this data set and should be kept in consideration when assessing the results obtained from this analysis.

The pattern is, however, similar to that of the Fahey data set whereby there are a high number of words that occur at lower frequencies compared to those that appear at high frequencies. This just happens to be accentuated due to the differing subject matter. Again, "the" is the most commonly used word.

3.3.3 Word Length

As with the frequency analysis, the Fahey data set is analysed in two stages, first only analysing the paragraphs and secondly looking at the combination of the paragraphs and questions.

3.3.3.1 Fahey Experiment

Paragraphs

The length of each word was calculated and the frequencies were tallied for each length bin. There is a range from 1 to 16 letters per word, giving a total of 16 length bins. Firstly, the number of words with a given length was first tallied, shown in the second column of the Table 10 below. Secondly, the number of different words with that frequency were tallied, shown in the third column of the table below. The results are summarised in Table 10 and Figure 11.

Table 10 : Summary of lengths of words in the paragraphs

Length	Number of Words	Number of Different Words
1	22	2
2	122	19
3	133	28
4	135	51
5	84	47
6	72	44
7	70	51
8	86	49
9	35	27
10	23	15
11	15	12
12	17	10
13	9	7
14	2	2
15	1	1
16	2	2

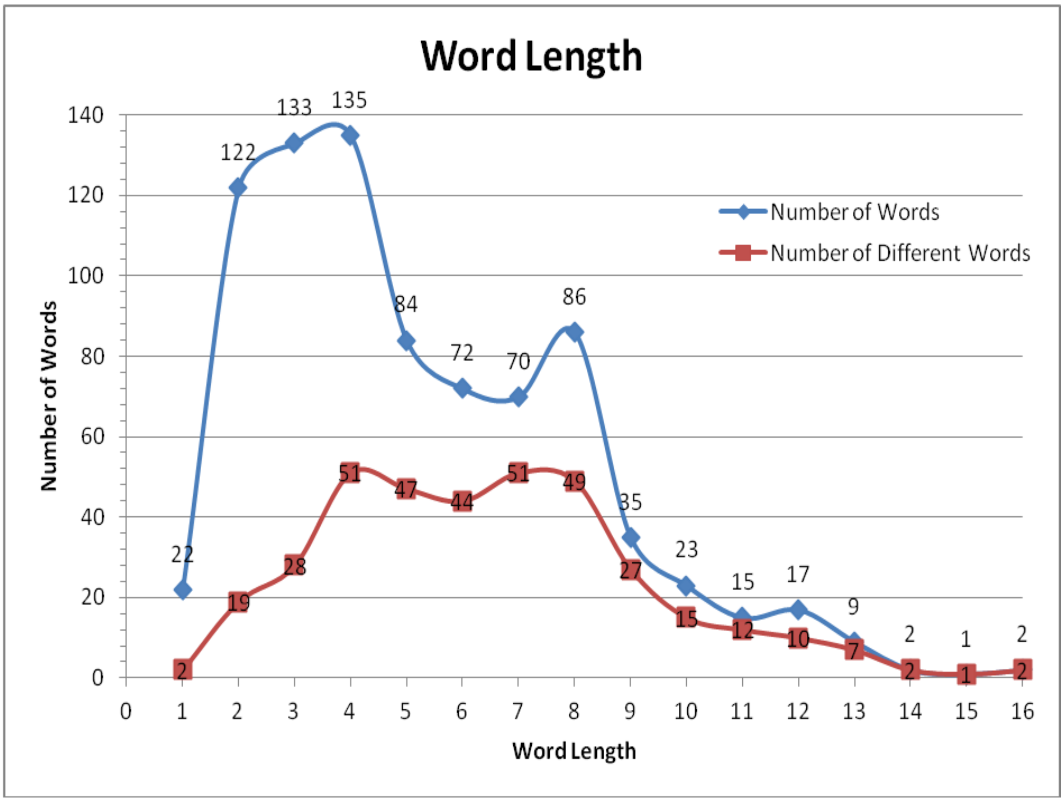


Figure 11 : Graphical representation of Word Length

We can see that there is a correlation between the total number of words with a given length and the number of different words with a given length. The difference between the two values generally decreases with increasing word length.

The shorter length words tend to occur at higher frequencies compared to the longer words. An example of this is the word "the" which is a three letter word that occurs 57 times. This helps to explain why there is such a great difference between the total number of words and the absolute number of words with length 3. The same rationale can be used to explain the patterns at lengths one, two and four.

Below we see that the shortest words are the number 6 and "a"; where "a" occurs quite frequently at 21 times.

Table 11*:* Summary of shortest words in the paragraphs

Word	Length	Frequency
6	1	1
a	1	21

The longest words are summarised in the Table 12 below, where it can be observed that 3 out of the 5 longest words are concatenations of two words. This is an important fact to note; concatenation were not disturbed when other punctuation was taken out.

Further, we observe that the word "correspondingly" is the second longest word. This word comes from the word correspond which occurs in the text. This is another important consideration to take into account; no word stemming was applied to the text. This will be left for future analysis or if there is time in the study, stemming will be accounted for. For the current perspective this is not of concern as the main purpose is looking at the physical nature of the raw words to see if this could affect eye movement.

Table 12: Summary of longest words in the paragraphs

Word	Length
classification	14
human-computer	14
correspondingly	15
image-processing	16
labour-intensive	16

In summary, we observe that the shorter words have greater frequency than longer words, in general. There is also a difference between the total number of words with a given length compared to the absolute number of words with that length. There is a pattern that shows that the difference between the two values decreases with increasing word length.

Paragraphs and Questions

In order to compare the questions to the paragraphs, a text analysis is performed on the content of both the Paragraphs and the Questions combined. The results are quite similar to the analysis on the paragraphs alone, as expected.

The length of each word was calculated and the frequencies were tallied for each length bin. There is still a range from 1 to 16 letters per the word, giving a total of 16 length bins. Firstly, the number of words with a given length was tallied, shown in the second column of the table below. Secondly, the number of different words with that frequency were tallied, shown in the third column of the table below. The results are summarised in Table 13 and Figure 12.

Table 13: Summary of lengths of words in the paragraphs

Length	Number of Words	Number of Different Words
1	32	2
2	179	23
3	193	30
4	183	54
5	112	57
6	105	51
7	87	63
8	111	52
9	44	31
10	33	20
11	19	14
12	25	11
13	10	7
14	3	2
15	1	1
16	2	2

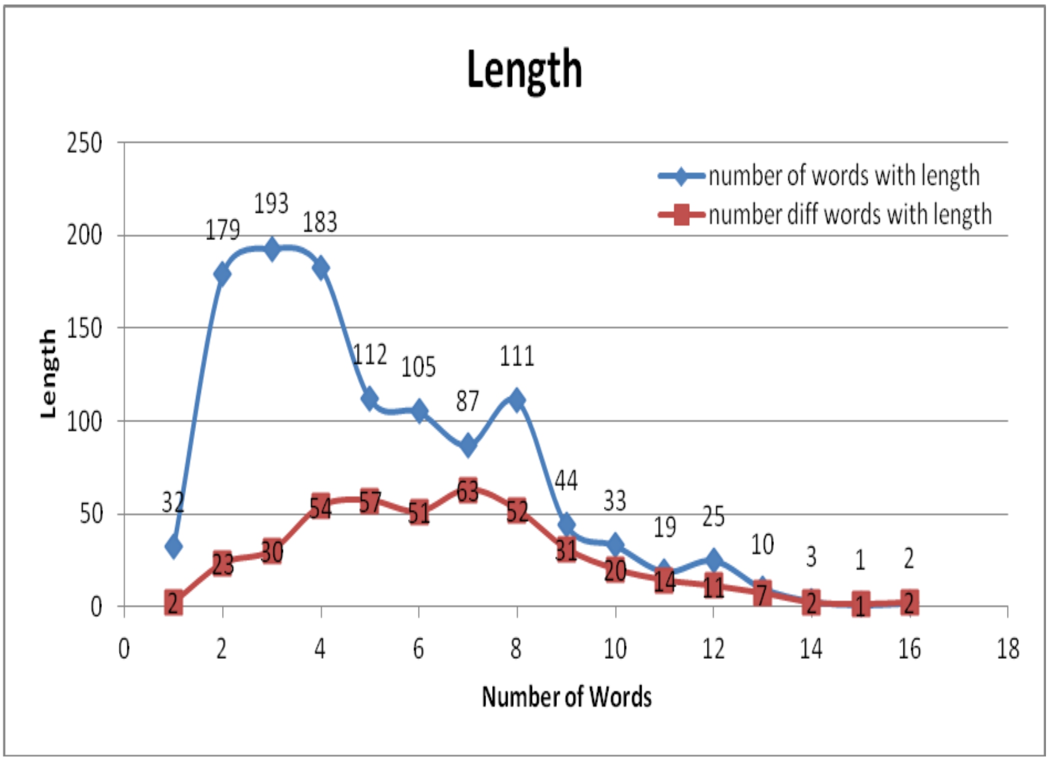


Figure 12: Graphical representation of Word Length

We can see that there is a graphical correlation between the total number of words with a given length and the number of different words with a given length. The difference between the two values generally decreases with increasing word length.

As with the paragraph only data set, the lower length words still occur at higher frequencies compared to the longer words. Interestingly, the inclusion of the questions to the paragraphs has caused little change to the words that occur at the ends of the distribution. Again, we see that the shortest words are the number 6 and "a"; however, "a" now occurs quite frequently at 31 times in total. The longest words are also the same as in table.

3.3.3.2 Sharma Experiment

Similar analysis was performed on the Sharma data set, which showed similar results to the Fahey data set. The length of each word was calculated and the frequencies were tallied for each length bin. There is a range from 1 to 18 letters per word, giving a total of 16 length bins. Again, the number of words with a given length was first tallied, shown in the second column of the table below. Then, the number of different words with that frequency were tallied, shown in the third column of the table below. The results are summarised in Table 14 and Figure 13.

Table 14: Summary of lengths of words in the paragraphs

Length	Number of Words	Number of different words
1	20	2
2	108	17
3	162	39
4	106	64
5	63	50
6	60	46
7	48	41
8	46	31
9	36	29
10	31	21
11	13	11
12	9	7
13	5	3
14	3	2
15	4	2
18	1	1

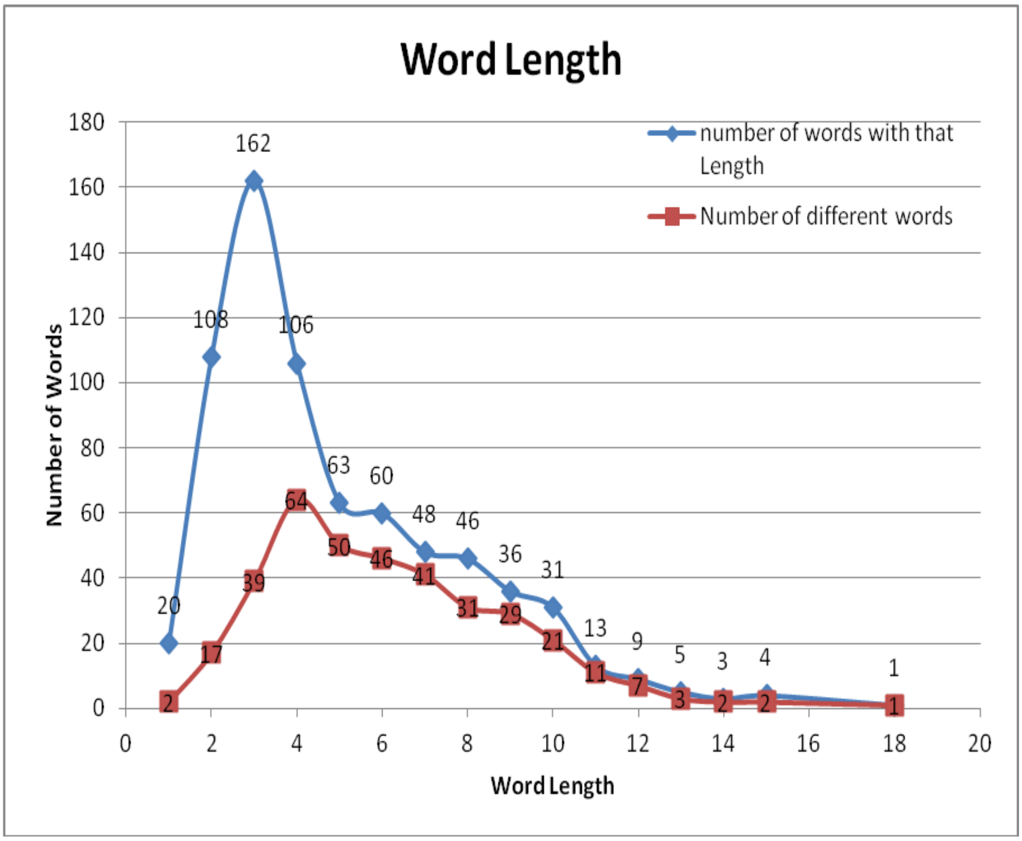


Figure 13: Graphical representation of Word Length

Once again, we can see that there is a correlation between the total number of words with a given length and the number of different words with a given length. The difference between the two values generally decreases with increasing word length.

The lower length words tend to occur at higher frequencies compared to the longer words. An example of this is the word "the" which is a three letter word that occurs 57 times. This helps to explain why there is such a great difference between the total number of words and the absolute number of words with length 3. The same rationale can be used to explain the patterns at lengths one, two and four.

Below we see that the shortest word is "a" which occurs quite frequently at 21 times, coincidently the same frequency at which it occurs in the Fahey data set.

Table 15: Summary of shortest words in the paragraphs

Word	Length	Frequency
a	1	21

The longest words are summarised in the table below.

Table 16 : Summary of longest words in the paragraphs

Word	Length
characteristic	14
representation	14
transformations	15
representations	15
nearest-neighbours	18

In summary, we observe that the shorter words have greater frequency then longer words, in general. There is also a difference between the total number of words with a given length compared to the absolute number of words with that length. There is a pattern that shows that the difference between the two values decreases with increasing word length.

3.3.4 Text Readability

The text readability of the paragraphs used in each experiment was evaluated using quantitative evaluation of the text, namely, readability measures. Readability is defined as the ease of which one can understand and read text. This is used as a measure of how easy or difficult word predictability will be in the paragraphs. This can only be used as a rough measure as word predictability is extremely difficult to measure as it is based on context and on the participants' prior knowledge. Unlike the above analyses, the readability of the text will be used to help characterise and define categories of content.

These readability measures include: Flesch-Kincaid Reading Ease and Grade Level, Gunning-Fog Score, Coleman-Liau Index, SMOG Index and Automated Readability Index. These measures are used to indicate the level of US education appropriate for reading certain pieces of text.

3.3.4.1 Fahey Experiment

Table 17: Readability analysis for the Fahey paragraphs

	Flesch-Kincaid Grade Level	Gunning-Fog Score	Coleman-Liau Index	SMOG Index	Automated Readability Index	Average Grade Level	Flesch-Kincaid Reading Ease
Paragraphs							
1	14.8	17.9	15.8	12.9	15.7	15.45	31.3
2	13.2	16.4	12.4	11.6	13.7	13.46	45.5

3	12.6	16.6	13.8	11.9	13.7	13.72	45.6
4	13.7	16	14.2	10.8	16.1	14.16	44.7
5	14	16	15.5	11.6	13.8	14.18	31
Average	**13.66**	**16.58**	**14.34**	**11.76**	**14.6**	**14.19**	**39.62**
6	19.5	25.9	15.9	18.6	19.7	19.92	11
7	16.5	20.4	16	14.1	20.3	17.46	34.6
8	11.6	14.9	17.1	10.5	10.8	12.98	32.9
9	13.3	15.5	10.9	10.6	13.9	12.84	49.5
10	16.6	20.2	17.1	14.5	18.2	17.32	23.3
Average	**15.5**	**19.38**	**15.4**	**13.66**	**16.58**	**16.10**	**30.26**
Questions							
1	9.1	11.3	11.9	8.8	8.4	9.9	56.8
2	10.5	11.6	14.3	8.3	11.4	11.22	51.3
3	9.3	9.5	11.2	4.4	11	9.08	67
4	8.2	12	11	8.8	6.9	9.38	60.6
5	9	13	11.7	9.2	5.6	9.7	48.60
Average	**9.2**	**11.5**	**12.0**	**7.9**	**8.7**	**9.9**	**58.9**

As the text analysis show that the relevant paragraphs do not correspond to hard content and the less-relevant paragraphs do not correspond to easy content, the paragraphs are re-categorised for subsequent analysis on the first data set:

Table 18 : Paragraphs from the Fahey study categorised into hard and easy based on the readability tests performed

Category	**Paragraph**	**Flesch-Kincaid Grade Level**	**Gunning-Fog Score**	**Coleman-Liau Index**	**SMOG Index**	**Automated Readability Index**	**Average Grade Level**	**Flesch-Kincaid Reading Ease**
Hard	6	19.5	25.9	15.9	18.6	19.7	19.92	11
Hard	7	16.5	20.4	16	14.1	20.3	17.46	34.6
Hard	10	16.6	20.2	17.1	14.5	18.2	17.32	23.3
	Average	**17.5**	**22.2**	**16.3**	**15.7**	**19.4**	**18.2**	**23.0**
Easy	2	13.2	16.4	12.4	11.6	13.7	13.46	45.5
Easy	8	11.6	14.9	17.1	10.5	10.8	12.98	32.9
Easy	9	13.3	15.5	10.9	10.6	13.9	12.84	49.5
	Average	**12.7**	**15.6**	**13.5**	**10.9**	**12.8**	**13.1**	**42.6**

Note: The rows highlighted grey will be considered as hard paragraphs and the non-highlighted rows are the easy paragraphs in subsequent analysis of the Fahey data set.

3.3.4.2 Sharma Experiment

Table 19 : Readability analysis for the Sharma paragraphs

Paragraph	**Flesch-Kincaid Grade Level**	**Gunning-Fog Score**	**Coleman-Liau Index**	**SMOG Index**	**Automated Readability Index**	**Average Grade Level**	**Flesch-Kincaid Reading Ease**
E1	2.9	4.8	9.4	3.6	3.1	4.76	89.7
E2	14.6	16.8	17.2	13.9	16.1	15.72	30
E3	5.8	8.3	7.6	4.7	6	6.48	84.4

Average	**7.8**	**10.0**	**11.4**	**7.4**	**8.4**	**9.0**	**68.0**
H1	11.8	15.4	14.9	11.3	13.7	13.42	48.1
H2	18	23.4	13.5	16.8	18	17.94	23
H3	12.3	15.9	16.7	11.4	11.7	13.6	32.1
Average	**14.0**	**18.2**	**15.0**	**13.2**	**14.5**	**15.0**	**34.4**

The results of the readability tests on the paragraphs in the Fahey experiment are contrary to what was believed about those paragraphs. It can be seen that on average, the less-relevant paragraphs are defined as harder to read compared to the relevant paragraphs, with Flesch-Kincaid Reading Ease of 30 versus 40, for less relevant paragraphs versus relevant paragraphs. Note that the Flesch-Kincaid Reading Ease is inverse to the other scores where the scoring system can be summarised:

Table 20 : Flesch-Kincaid Reading Ease scoring level definitions

Score	Level of Difficulty
90-100	Very Easy (easily understandable by an average 11-year-old student)
80-89	Easy
70-79	Fairly Easy
60-69	Standard (easily understandable by 13- to 15-year-old students)
50-59	Fairly Difficult
30-49	Difficult (best understood by university graduates)
0-29	Very Confusing

These results are the opposite of the selection intention of the paragraphs, since the relevant paragraphs are considered to have more difficult content than the less-relevant paragraphs. The authors of the paper that was the source of most of the paragraphs in Fahey's study were consulted and a possible explanation is now available. The authors commented that reviewers often seek improvement and clarification of the substantive paragraphs showing results ('relevant' paragraphs in Fahey's study). Meanwhile space restriction means that introductory paragraphs end up compressed and hence harder to read. Furthermore, the readability tests are quantitative measures and with the analysis of the complexity of text one must consider qualitative measures as well.

The questions from the Fahey data set have a more defined division from the paragraphs using the readability measures as compared to the division between the categories of paragraphs. The questions are all easier to read compared to the paragraphs based on the readability measures.

There is an anomaly in the second dataset, which is that paragraph E2 has a Flesch-Kincaid Reading Ease of 30 which would indicate that this is a very hard piece of text to read and that the intended audience is that of university graduates. The paragraph is in fact an advertisement for an annual flower shower in Canberra, Floriade. The advertisement was taken straight from the show's website and it is believed that it is aimed at a wide target audience and thus should be easy to read. Subjectively, it is fair to say that it is an easy piece of text to read especially in comparison with the three hard paragraphs for this section.

For these reasons, the results of the readability tests should be taken with some caution in regard to how reliable the categorisation of paragraphs as hard or easy in respect to content. The results should be interpreted more as a measure of word predictability. This is why for the study we break the Fahey paragraphs into two lots, one with the initial categorisation of relevant and less-relevant and the other with the categorisation of hard and easy.

Chapter 4

Analysis of Eye Gaze Data

The second phase of the project was taking the information that was generated in the first phase and analysing it using several different methods. The different analyses were designed to test the different hypotheses set out at the beginning of the experiment.

4.1 Score Analysis

The first hypothesis is that the difference in paragraph content will affect the positioning of fixations on specific types of words, namely based on frequency and length. To test this hypothesis, the words that a participant fixated on had to be determined by matching the eye gaze to pixels on the screen to text content. The method of performing this task is outlined in the *Preparatory Work* section.

Each word was given three score values, these scores were obtained from the textual analysis performed on the text content. The first score was the frequency score which is the number of times the word occurred in all the paragraphs. The second score was the length of the word and the last score combined the first two scores as Score = (1/frequency)*length.

Words that appear at a high frequency will have higher frequency scores than those that appear at lower frequencies. Similarly for long words versus short words under the length based scoring method.

Through the textual analysis it was found that most long words appeared less often than short words, that is, in general the scores are the inverse of each other. So the combined score of (1/frequency)*length is designed to combine the two physical aspects of the word without corrupting either scores' reliability. This scoring system allows for the case where long words are more frequent and short words are infrequent, so the upside of this scoring system is that the effect of highly frequent, long words even out.

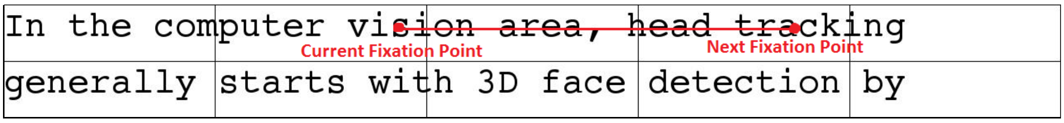
A significant issue that had to be resolved before analysis could be performed was that for the first data set the paragraphs were not the same length. There are 10 paragraphs in total for this data set and on average the less-relevant paragraphs were one third the length of the relevant paragraphs, with non-overlapping ranges between the two (see Table 1 and Table 2). This clearly means that there will be more fixations recorded for reading relevant paragraphs compared to reading non-relevant paragraphs simply because there are more words there to read. All results for the Fahey data set needed to be normalised, which was done by dividing by the number of fixations recorded for the paragraph.

Another way to normalise the data set would have been to divide scores by the number of words per paragraph. The results are quite similar under this normalisation, though there are some differences. Normalising by the number of fixations observed per paragraph focuses on the pattern of fixations rather than the number of words in the paragraph. It was chosen as it also at least partially normalises for re-reading of text, which many subjects indicated they did, as the time given for each paragraph was much longer than required for a single reading.

The measure used to assess if there is a difference between the two paragraphs is direct comparison between each participant's average normalised score for the relevant paragraphs against their average normalised score for the non-relevant paragraphs. A paired two-tail t-test was used to test if the values from the two groups are statistically significantly different. The same method was applied to compare the hard to easy paragraphs and the paragraphs to questions.

4.2 Backward and Forward Tracking Analysis

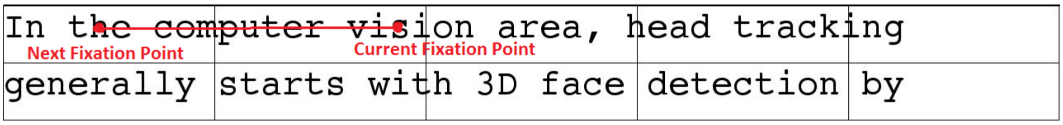
The second hypothesis states that more backtracking of eye movement will be observed in harder paragraphs compared to easier paragraphs, similarly for relevant versus less-relevant paragraphs respectively. To test this hypothesis the position of the fixations on the screen were assessed. Here, the fixation movement is measured as being one of three possibilities; forward movement, backward movement, or staying on the same word. Forward movement is where the eye moves forward in the text, backward movement is where the eye moves backward over text that has already been passed over and no movement implies that there was another fixation on the same word.



In the computer vision area, head tracking

generally starts with 3D face detection by

Figure 14: Example of the forward eye movement.



In the computer vision area, head tracking

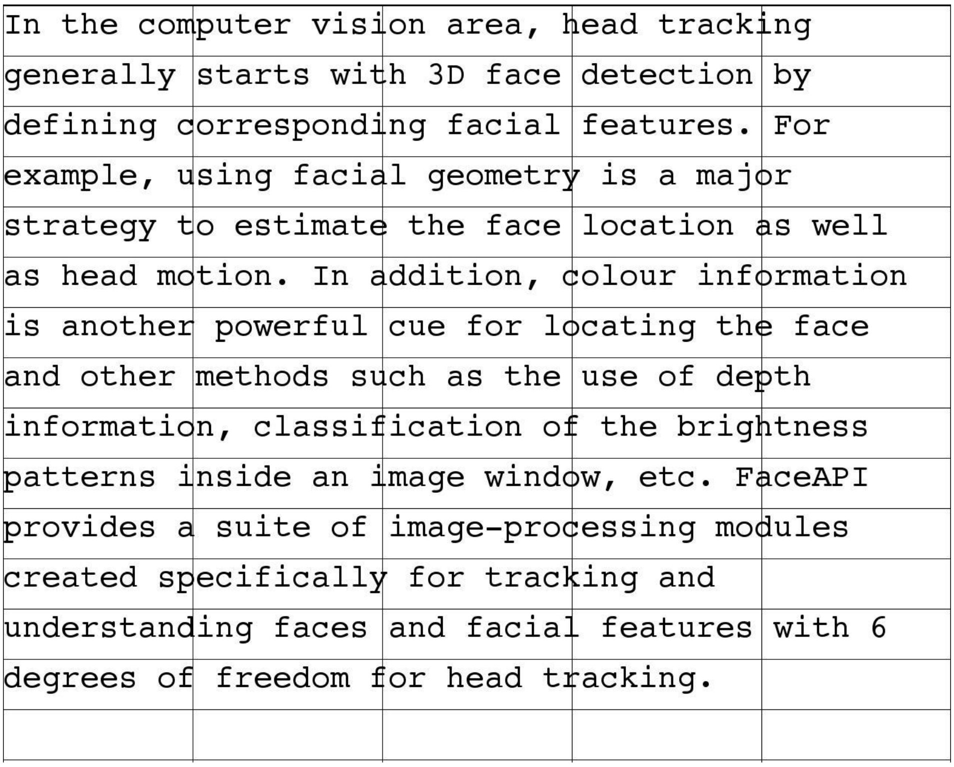
generally starts with 3D face detection by

Figure 15: Example of a backward movement.

The forward and backward tracking was assessed using two systems; the first of which is a grid system as used by Vo (2010) and the second was the word box system used in the scoring analysis. With the grid system, the screenshots are divided into equal sized boxes that form a grid. This system does not allow for scoring but does allow for measurement of back tracking and forward tracking.

Each fixation was checked to see which box it fell in, generating a list of numbers. To measure the forward and backward movements, each number in the list is subtracted from its predecessor. This gives a positive number (forward movement), a negative number (backward movement), or a zero (no movement from the box).

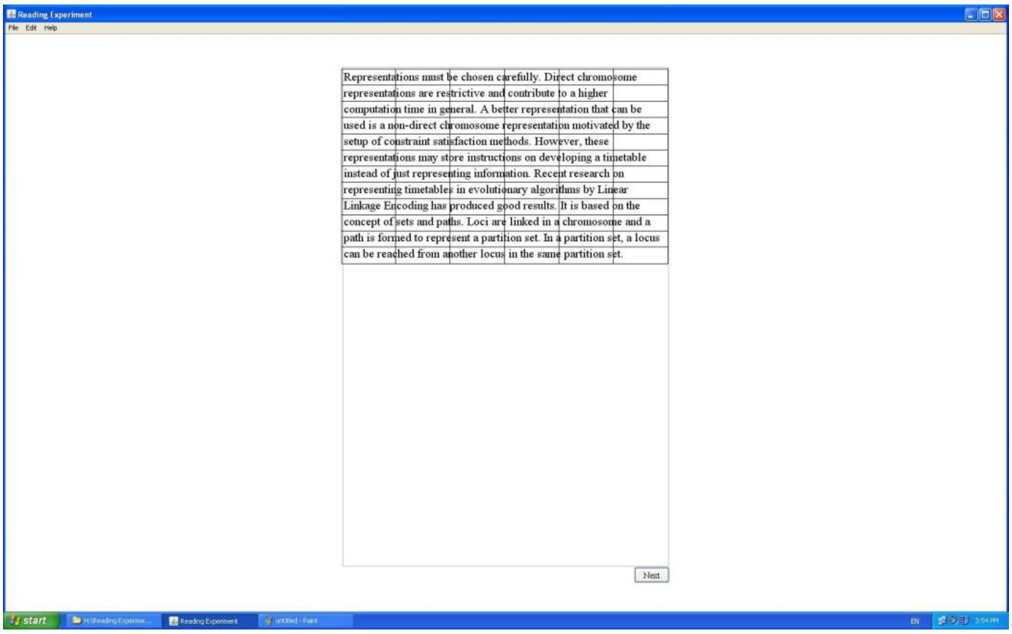
It is believed that a reader normally follows a reading pattern of scanning left to right and moving from one line to the next line, so the grid is designed so that each box only encompasses one line. The horizontal range is kept the same, as this is approximately 1 word per box. The lines that are placed on the screen are equally spaced. For Fahey's data set the grid used is 15x5, giving 75 boxes in total. This can be seen graphically in the diagram below.



In the computer vision area, head tracking
generally starts with 3D face detection by
defining corresponding facial features. For
example, using facial geometry is a major
strategy to estimate the face location as well
as head motion. In addition, colour information
is another powerful cue for locating the face
and other methods such as the use of depth
information, classification of the brightness
patterns inside an image window, etc. FaceAPI
provides a suite of image-processing modules
created specifically for tracking and
understanding faces and facial features with 6
degrees of freedom for head tracking.

Figure 16: 15x5 grid example for the Fahey data set

The grid used for Sharma's study is similar, however, since the paragraph only makes up a small area of the screen the grid is only applied to the area of the screen that includes the paragraph. The grid used is basically the same as it is 15x6. The reason for the finer granularity is because the paragraph is displayed in a smaller box, the text is smaller so the smaller boxes are designed to try to encapsulate only one to two words as in the grid used for the Fahey data set. This is a more natural approach to displaying the data as the display size of the text is more natural. This can be seen graphically in the figure below:



Representations must be chosen carefully. Direct chromosome
representations are restrictive and contribute to a higher
computation time in general. A better representation that can be
used is a non-direct chromosome representation motivated by the
setup of constraint satisfaction methods. However, these
representations may store instructions on developing a timetable
instead of just representing information. Recent research on
representing timetables in evolutionary algorithms by Linear
Linkage Encoding has produced good results. It is based on the
concept of sets and paths. Loci are linked in a chromosome and a
path is formed to represent a partition set. In a partition set, a locus
can be reached from another locus in the same partition set.

Figure 17: 15x6 grid example for the Sharma data set

These grids still cut some words in half and also incorporate some white space, as seen in Figure 15 at the very bottom of the screenshot and to the very right of the screenshot. To avoid inclusion of fixations that landed on white space or fragmentation of words, the word boxes technique was used as a secondary sequence generation technique for the analysis. The word boxes technique is the same system of dividing the screen up is used as in the scoring analysis. This is where the words have boxes drawn around them (see Figure 1 and Figure 2) as described in the Preliminary Work section.

4.3 Reading Comprehension Scores

In the Fahey experiment (2009), participants were tested on their understanding of the content for which they had just had their eye gaze monitored whilst reading. In essence, a score on their reading comprehension was given to each participant based on three criteria. These scores are summarised below:

Table 21 : Summary of the participant scores from the previous study

Participants	Paragraph score	Question score	Sentence score	Total score
0675819234	10	4	2	16
1234678590	6	3	2	11
1627384950	6	1	0	7
2347581906	8	2	1	11
3451890267	4	0	0	4
4592136780	8	2	0	10
5120364789	8	4	0	12
6718902345	0	4	0	4
6789123045	4	2	1	7
7892036451	6	2	0	8
8906345712	6	2	0	8
Mean	6	2.4	0.5	8.9
SD	2.7	1.3	0.8	3.6
Median	6	2	0	8
Mode	6	2	0	11

The measures outlined in Section 2.1 were used to assess if there was any correspondence between the participants' total score and the participants' eye gaze data. The study found that there was no correlation between any of these measures and the participant's total score (Fahey, 2009).

4.3.1 Pearson Moment Correlation Coefficient

To test for correlation between the scoring method analysis and forward and back tracking analysis with the reading comprehension scores generated from the original experiment, the Pearson correlation coefficient is used. The Pearson Moment correlation coefficient measures the linear dependence between two variables and produces a value between -1 and 1. The formula for the Pearson Moment Correlation Coefficient is:

$$r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}}$$

The correlation values obtained from comparing the scores and the linear eye movements generated from this study to Fahey's reading comprehension scores to determine if there are any linear correlations between the findings from this study to the participants' reading comprehension.

4.4 HMM Analysis

The hidden Markov model (HMM) analysis of the eye gaze patterns is used to classify the sequences of eye movement from each paragraph. Here the forward algorithm is used to calculate the probability that each particular sequence was generated by the model λ_m from a set of models λ_1, λ_2, ... λ_n. The HMMs are trained using the Baum-Welch algorithm from the eye gaze data collected in each experiment. Classification of a data sequences using HMMs is accomplished by calculating the probability of the sequences given each model and then choosing the most likely model, i.e. the forward algorithm.

All HMMs used in the analysis have three hidden states as this has given good results in previous experiments that looked at reading and eye gaze data (Gustavsson, 2010). Different HMMs are designed to model different types of data and similar HMMs are trained with different training sets in order to test different categories of data.

Table 23 describes these alphabets; under each alphabet size are the divisions of movement that each symbol in the HMM models. For example, the first HMM has an alphabet M=3 which represents back movement (<1), stationary (0) and forward movement (>1). This is the same as the grid analysis described above. The remaining alphabets model the eye movement with increasing granularity. The symbols within the alphabets refer to the magnitudes of movement between boxes. The alphabet with 7 symbols were designed to provide more definition of the linear eye movements whilst still keeping the definition relatively loose (i.e. increments of 5 box jumps to ±10). The alphabet with 17 symbols was designed to increase the range of movements described as well as increase the definition of movements. Here the range of movements is defined up to ±20 box jumps with approximately 3 box jumps defined in each symbol. The alphabet of 43 symbols is designed to further increase the definition in the movements but no extension on the range of movements.

Table 22 : Alphabet sizes for HMMs used in the analysis

Alphabet Size				
M=3		M=17	M=43	
<-1	<-10	<-20	<-20	
0	-6 to -10	-19 to -20	-20	1
>1	-1 to -5	-16 to -18	-19	2
	0	-13 to -15	-18	3
	1 to 5	-10 to -12	-17	4
	6 to 10	-7 to -9	-16	5
	>10	-4 to -6	-15	6
		-1 to -3	-14	7
		0	-13	8
		1 to 3	-12	9
		4 to 6	-11	10
		7 to 9	-10	11
		10 to 12	-9	12
		13 to 15	-8	13
		16 to 18	-7	14
		19 to 20	-6	15
		>20	-5	16
			-4	17
			-3	18
			-2	19
			-1	20
			0	>20

Each HMM within a type is designed to model eye gaze patterns collected from certain content types. There are three different classifiers for the Fahey 2009 data set, as described below:

Table 23: The groupings of the HMM classifiers used on the Fahey data set

Classifiers	Classes
Classifier 1: $\lambda = (\lambda_1, \lambda_2)$ **Purpose:** to distinguish between gaze patterns records from Paragraphs and Questions	λ_1= Paragraphs λ_2= Questions
Classifier 2: $\lambda = (\lambda_1, \lambda_2)$ **Purpose:** to distinguish between gaze patterns records from Relevant and Less-Relevant	λ_1= Relevant λ_2= Less-Relevant
Classifier 3: $\lambda = (\lambda_1, \lambda_2)$ **Purpose:** to distinguish between gaze patterns records from Hard and Easy	λ_1= Hard λ_2= Easy

In each case, the training sequences are labelled with the paragraph from which they were recorded. The sequences are evaluated using the forward algorithm to find the most likely class that they belong to within the classifier. Theoretically, if two classes are distinct and there was total distinction between gaze patterns recorded between the different forms of content, we would observe 100% recall of sequences to the class they belong to.

One of the problems with using HMMs is that long sequences can produce very small probabilities, so close to zero that underflow becomes a problem and the computer reports a probability of zero. It can be readily observed that a high number of stationary observations are recorded, causing long sequences

As an additional test case, the stationary points were concatenated together to generate shorter observation sequences. These shorter observation sequences were then passed through the above classifiers.

Finally, as with the backward and forward movement analysis, the two cases of grid system and word boxes are analysed to check for the most efficient way of distinguishing if there are differences in gaze patterns based upon the content of what is read.

Since the Sharma data set only consists of hard and easy content, the number of classifiers is reduced to one where hard paragraphs are compared to easy paragraphs. This is the same as classifier 3 for the Fahey set.

Chapter 5

Results

For the Fahey data set, results from the analyses were all obtained using a 13-pixel threshold and calibrated fixations. For the Sharma data set, a 5-pixel threshold and non-calibrated fixations were used. For both data sets and whenever a t-test was performed, a significance level of p<0.01 has been chosen as the significance level for all T-test probabilities unless otherwise specified. All t-tests performed are paired two tail tests.

Note that for the Fahey data set the hard paragraphs are paragraphs 6, 7, and 10 and the easy paragraphs are paragraphs 2, 8 and 9 as classified by the text complexity analysis performed in the Preliminary Analysis section (see Table 18).

5.1 Scoring Analysis

5.1.1 Fahey Data set

The results of the analysis were obtained by finding the score and the number of fixations observed for each paragraph for each participant. The number of fixations is needed in order to normalise the data for this set as the paragraphs have unequal length.

The scoring system works as follows; the words that a participant fixated upon in each paragraph are found. Associated to each word are three scores, one based on frequency, one based on the length and lastly a combination of the two. The scores for each word in each paragraph are tallied together. The tallied scores from the relevant paragraphs are compared to the tallied scores from the less-relevant paragraphs using a t-test to assess for statistical significance in the differences. This is also performed for the hard paragraphs compared to the easy paragraphs, and for the paragraphs compared to the questions.

Summarised in Table 24 are the t-test probability values for this dataset.

Table 24: T-test probabilities of comparisons content types used the three scoring schemes

	T-Test Probability		
Comparison Classes	**Frequency**	**Length**	**Combined Score**
Relevant vs. Less-Relevant	1.6E-08	7.0E-06	1.8E-09
Hard vs. Easy	9.3E-05	2.3E-06	4.5E-04
All Paragraphs vs. Questions	0.075	0.337	1.26E-08

These results show that the relevant and less-relevant paragraphs can be distinguished using all three scoring methods, as there is a strong significant difference ($p<0.01$) between the scores generated from the relevant paragraphs versus the less-relevant paragraphs. Similarly for hard and easy paragraphs, the results show that the hard paragraphs can be distinguished from the easy paragraphs using all three scoring methods.

Interestingly, the results show that the paragraphs cannot be distinguished from the questions using the frequency or length scoring method but can be distinguished using the combined scoring method. It was believed that the paragraphs versus the questions would display a stronger difference and be easier to detect as compared to difference in content within the category of paragraphs. However, although the frequency and length scoring methods alone cannot distinguish the paragraphs from questions, using the combined scoring method does show that there is a significant difference ($p<0.01$) between the paragraphs and the questions.

Please refer to Appendix E for the raw tallied scores used to produce t-test values.

5.1.2 Sharma Data set

The results from the Sharma data set were obtained similarly to the above for the Fahey data set. The outcomes are summarised in the table below:

Table 25: T-test probabilities of comparisons content types used the three scoring schemes

	T-Test Probability		
Comparison Class	**Frequency**	**Length**	**Score**
Hard vs. Easy	0.033	0.129	0.129

The results show that there is no significant difference detected between the hard versus easy paragraphs for this data set from all of the three scoring systems. This is not what was expected as there was significant difference (p<0.01) between the hard and easy paragraphs in the Fahey data set. This is discussed further in Section 6.1.

Refer to Appendix E for the raw tallied scores used to produce t-test values.

5.2 Backward and Forward Tracking Analysis

In the backward and forward tracking analysis, three measures were taken to assess the eye gaze data. These measures are; forward, backward and stationary movements. Refer to Section 4.2 for more details on the analysis.

For the backward and forward tracking analysis two different systems are used to generate the sequences of eye movement; these being the grid system and the word boxes system. The grid used is dependent on the dataset as the grids are specialised for the datasets; a 15x5 grid is used for the Fahey data set and a 15x6 grid is used for the Sharma data set.

5.2.1 Fahey Data set

A two-tailed, paired t-test was used to identify if there was any significant difference in backward, forward and stationary movement between each category of paragraph comparison groups (i.e. relevant versus less-relevant). The results of the t-tests are summarised in the table below.

Table 26: T-test probabilities of comparison of content types for the three movement types

	T-Test Probability		
15x5 Grid	Forward	Backward	Stationary
Relevant vs. Less-Relevant	0.049	0.724	0.283
Hard vs. Easy	0.871	0.111	0.160
Question vs. Paragraphs	0.008	0.011	0.004
Word Boxes			
Relevant vs. Less-Relevant	0.655	0.079	0.213
Hard vs. Easy	0.904	0.926	0.852
Questions vs. all Paragraphs	0.531	0.910	0.703

All t-test results show that there is no significant difference in any of the movement types between relevant versus less-relevant or hard versus easy paragraphs under both the grid and word box schemes. These results are unexpected based on previous research and are discussed in more depth in Section 6.2.

However, the t-test results show that under the 15x5 grid, there is a significant difference ($p<0.01$) in the forward and stationary movements between questions and paragraphs. Further, if we consider $p<0.05$ as the significance level, under all movement types there is a difference between eye gaze patterns recorded from reading paragraphs opposed to questions.

Please refer to Appendix F for the raw tallied scores used to produce t-test values.

5.2.2 Sharma Data set

A two-tailed, paired t-test was used to identify if there was any significant difference between each category of movement; backward, forward and stationary. This comparison was between each participant's average movement in the hard paragraphs versus easy paragraphs. The results of the t-tests are summarised in the table below.

Table 27: T-test probabilities of comparison of content types for the three movement types

	T-Test Probability		
6 x 15 Grid	Forward	Backward	Stationary
Hard vs. Easy	0.278	0.719	0.834
Word Boxes			
Hard vs. Easy	0.292	0.491	0.548

All t-test results show that there is no significant difference in any of the movement types between the hard and easy paragraphs. This is consistent with the above results that are based on the content of the paragraphs, backward and forward tracking cannot differentiate the two.

Please refer to Appendix F for the raw tallied scores used to produce t-test values.

5.3 Reading Comprehension

5.3.1 Scoring

To investigate if there is a correlation between eye gaze patterns and reading comprehension scores the overall average normalised values for each scoring method is compared against the four comprehension scores that were taken from the initial experiment using the Pearson correlation coefficient. The results are outlined in the table below:

Table 28: Pearson correlation coefficient values for the total scores obtained for all paragraphs versus the total reading comprehension scores.

	Total score
Scoring Method	**Pearson's R**
Frequency	0.03
Length	-0.22
(1/Frequency)*Length	-0.08

The results show that for each scoring method there is little to no correlation between the normalised scores for all the paragraphs and Fahey's reading comprehension scores. This suggests that there is no linear correlation between the average scores obtained across all paragraphs and Fahey's reading comprehension scores.

To analyse this further the paragraph categories are assessed individually. The results of this analysis are shown in the table below:

Table 29: Pearson correlation coefficient values for the total scores obtained for the content groupings of paragraphs versus the total reading comprehension scores.

	Total Score
Scoring type	**Pearson's R**
Frequency	
Relevant paragraphs	-0.36
Less Relevant paragraphs	0.38
Hard paragraphs	0.35
Easy paragraphs	-0.53
Length	
Relevant paragraphs	-0.14
Less Relevant paragraphs	-0.01
Hard paragraphs	-0.49
Easy paragraphs	0.44
Score	
Relevant paragraphs	-0.14
Less Relevant paragraphs	-0.14
Hard paragraphs	-0.29
Easy paragraphs	0.34

These results are discussed in detail in the discussion, as there are complex relationships between the scores obtained for the kinds of paragraphs and Fahey's reading comprehension scores.

5.3.2 Backward and Forward Tracking

The following results show the Pearson correlation coefficient for the total reading comprehension score from the Fahey experiment versus the normalised overall average of forward movement, backward movement and stationary movement displayed by each participant.

Table 30: Pearson correlation coefficient values for the total scores obtained for all paragraphs versus the total reading comprehension scores.

	Total score
Movement type	**Pearson's R**
15x5 grid - all paragraphs	
Forward	-0.03
Backward	-0.50
Stationary	0.37
Word Boxes - all paragraphs	
Forward	0.15
Backward	-0.46
Stationary	0.28

The results show that there is a negative correlation of about -0.5 between the backward movement and the total score and sentence score. These are the strongest correlations measured for the data set. More specifically, no significant correlations are found between any measures of forward or stationary movement versus any of the comprehension scores.

To analyse this further the paragraph categories are assessed individually. The results of this analysis are shown in the table below:

Table 31: Pearson correlation coefficient values for the total scores obtained for all paragraphs versus the total reading comprehension scores.

	Total score
Movement type	**Pearson's R**
15x5 grid - relevant paragraphs	
Forward	0.06
Backward	-0.48
Stationary	0.33
15x5 grid - less-relevant paragraphs	
Forward	-0.23
Backward	-0.49
Stationary	0.44
15x5 grid - hard paragraphs	
Forward	-0.02
Backward	-0.52
Stationary	0.38
15x5 grid - easy paragraphs	
Forward	-0.21
Backward	-0.58
Stationary	0.50
Word boxes - relevant paragraphs	
Forward	0.19
Backward	-0.47
Stationary	0.25
Word boxes - less-relevant paragraphs	
Forward	0.04
Backward	-0.41
Stationary	0.28
Word boxes - hard paragraphs	
Forward	0.48
Backward	-0.49
Stationary	0.23
Word boxes - easy paragraphs	
Forward	0.48
Backward	-0.49
Stationary	0.23

The results are almost identical to those found when considering the paragraphs together. We can see that in all cases there is a negative correlation of about -0.5 between the backward movement and the total score and sentence score. This suggests that any connections found between reading comprehension and the eye gaze patterns are not caused by the individual categories of paragraphs, but by the eye movements found in regards to reading the paragraphs combined.

5.4 Hidden Markov Model Analysis

As introduced in the Analysis section, there are several HMM classifiers used in the analysis. These types differ based on the size of the alphabet of emission symbols. The HMM's used in the analysis have the following alphabet sizes: 3, 7, 17 and 43, see Table 22. All HMMs have 3 hidden states. In all cases, both sequences that have had the stationary points left non-concatenated are used, as well as with stationary points concatenated. For the Fahey data set this was a 15x5 grid is used to generate sequences and for the Sharma data set this was a 15x6 grid is used.

5.4.1 Fahey data set

The following table shows the results from HMM classifiers consisting of two classes; paragraphs and questions. The idea is that there is a difference in eye gaze patterns generated from reading a paragraph versus a question.

Table 32: HMM Classifier with two classes: paragraphs and questions. Grid and word box sequence generation techniques with raw and concatenated sequences assessed.

HMM Classifier Classes: Paragraphs vs. Questions

HMM alphabet size (M)	Stationary points concatenated	Sequences generated from Grid system	Sequences generated from Word Boxes system
		Percentage correctly assigned	Percentage correctly assigned
3	No	**43**	**56**
3	Yes	**56**	**52**
7	No	**61**	**61**
7	Yes	**68**	**74**
17	No	**61**	**64**
17	Yes	**72**	**79**
43	No	**76**	**72**
43	Yes	**73**	**81**

The results from the HMM analysis show that with the 3 symbol alphabet there is only a by-chance classification of sequences into their respective classes; paragraphs or questions. These results are predictable from the simple HMM as the backward and forward analysis showed that there was no statistical difference in the eye movements based on sequences generated from either the grid or the word box techniques.

The classification results are improved with the use of the 7 and 17 symbol alphabets where we obtain better than chance classification of sequences, 61-79% accuracy. We expect that the HMM that models the eye movement with finer granularity would perform better than the models with less granularity. Again, we see this to be true, where the use of the 43 symbol alphabet increases the classification accuracy to 72-81%. In most cases the sequences generated from the word boxes have a high percentage of accurate classification, as well as the sequences with concatenated stationary points.

Since the paragraphs and questions can be differentiated using the HMM analysis, the next step is to examine if the paragraphs can be differentiated based on content. That is, relevant versus less-relevant.

The following table show the results from HMM sets consisting of two classes again; one for relevant paragraphs and the other for less-relevant paragraphs. The idea is to examine if there is a difference in eye gaze patterns generated from reading a relevant paragraph versus a less-relevant paragraph.

Table 33: HMM Classifier with two classes: relevant and less-relevant. Grid and word box sequence generation techniques with raw and concatenated sequences assessed.

HMM Classifier Classes: Relevant vs. Less-Relevant

HMM alphabet size (M)	Stationary points concatenated	Sequences generated from Grid system	Sequences generated from Word Boxes system
		Percentage correctly assigned	Percentage correctly assigned
3	No	**57**	**56**
3	Yes	**55**	**55**
7	No	**56**	**59**
7	Yes	**60**	**57**
17	No	**66**	**67**
17	Yes	**73**	**67**
43	No	**70**	**75**
43	Yes	**78**	**74**

The results from this analysis show that there is a slightly better than chance classification of sequences back to their parent class using both the 3 and 7 symbol models using either concatenated and non-concatenated sequences.

The classification is improved to 66-73% accuracy from the 17 symbol alphabet and further improved to 70-78% accuracy from the 43 symbol alphabet. Like the case of distinguishing between paragraphs and questions, we see that the sequences with concatenated stationary points have higher classification levels.

We expected to obtain lower levels of classification between relevant and less-relevant sequences as compared to the paragraphs versus the questions. This is because the difference between the relevant and less-relevant paragraphs is by far the subtlest of the three groupings. The structure and readability of the questions compared to the paragraphs is quite different.

Next, the question of whether easy paragraphs can be differentiated from hard paragraphs is investigated. The following results show the results from HMM classifiers consisting of a class for hard paragraphs and a class for easy paragraphs. The idea is that there is a difference in eye gaze patterns generated from reading a hard paragraph versus an easy paragraph.

Table 34: HMM Classifier with two classes: hard and easy. Grid and word box sequence generation techniques with raw and concatenated sequences assessed.

HMM Classifier Classes: Hard vs. Easy

HMM alphabet size (M)	Stationary points concatenated	Sequences generated from Grid system	Sequences generated from Word Boxes system
		Percentage correctly assigned	Percentage correctly assigned
3	No	**61**	**41**
3	Yes	**67**	**56**
7	No	**61**	**61**
7	Yes	**64**	**61**
17	No	**64**	**59**
17	Yes	**79**	**73**
43	No	**79**	**77**
43	Yes	**91**	**85**

The results from this analysis show that there is a marginally better classification of sequences into the hard and easy classes compared to classifying the sequences into relevant and less-relevant classes in respect to the alphabet sizes 3, 7 and 17.

The results from the 43 symbol alphabet are the best observed for the data set where we have 77-91% classification accuracy. Again, we see that sequences with concatenated stationary points have the highest classification levels, which is consistent with the results from the previous classifiers. This shows that we can differentiate between eye gaze on the hard and easy content quite well. These classification levels are better than those observed for the paragraphs versus the questions, where it was hypothesised that we would see the highest level of differentiation and hence the highest levels of classification.

5.4.2 Sharma Data set

The following results show the results from HMM classifiers consisting a class for hard paragraphs and a class for easy paragraphs. The idea is again that there is a difference in eye gaze patterns generated from reading a hard paragraph versus an easy paragraph.

Table 35: HMM Classifier with two classes: hard and easy. Grid and word box sequence generation techniques with raw and concatenated sequences assessed.

HMM Classifier Classes: Hard vs. Easy

HMM alphabet size (M)	**Stationary points concatenated**	**Sequences generated from Grid system**	**Sequences generated from Word Boxes system**
		Percentage correctly assigned	**Percentage correctly assigned**
3	No	**45**	**52**
3	Yes	**55**	**62**
7	No	**67**	**77**
7	Yes	**53**	**53**
17	No	**73**	**73**
17	Yes	**53**	**58**
43	No	**48**	**72**
43	Yes	**65**	**58**

The results from this analysis show that the HMM with a 3 symbol alphabet gives by-chance classification of sequences generated from either the grid system or the word box techniques with no concatenation of stationary points. With concatenation of stationary points these results are improved to above chance classification. These results are better than those for the hard versus easy sequences and their related models from the analysis on the Fahey data set.

The use of the HMM sets with 7 symbol alphabet and with no concatenation, gives an overall 67-77% accuracy of classification of sequences to the correct classes. The use of concatenation of sequences reduces this accuracy back down to chance. We see this trend again under the 17 symbol alphabet where without concatenation there is 73% classification but concatenation reduced classification back down to chance. The results from the 43 symbol alphabet are not as expected. There is not only worse classification results compared to the other symbol sets but also the results are inverse for the sequences generated from the grid compared to the sequences generated from the word boxes scheme. This again brings into question the use of the sequences without calibration performed on the raw eye gaze data, and particularly on a data set using small font size letters. Perhaps, the smaller font sizes meant that the 43 symbols set for the HMM was too fine a granularity and using less symbols such as with the 17 symbol alphabet better define the movements on the eye on the screen whilst reading.

Chapter 6

Discussion

The purpose of this study is to differentiate between eye gaze patterns recorded from individuals reading different types of content. The analysis is centred on the hypotheses formulated for the study (Section 1.2). These hypotheses are used to answer the overarching hypothesis that there is a difference in eye gaze patterns generated from reading different types of content.

6.1 Scoring

Based on previous research it is believed that words affect where fixations occur (Fisher et al., 1998, Rayner, 1998, Reilly & O'Regan, 1998, Richter et al. 2005). In particular, longer words will attract the eye and longer words will require multiple fixations to read (Reilly & O'Regan, 1998). Furthermore, fixations will occur on infrequent words more than frequent words (Fisher et al., 1998, Richter et al., 2005). To assess this proposition a scoring system is constructed that is based on the content of the text read.

Another aspect of the text that affects the positioning of fixations is how predictable the words are in context (Fisher et al., 1998). This is not a simple measure that can be calculated as the word frequency and length can be, as predictability is dependent on the individual's prior knowledge and the context of what is being read. This factor is not tested the same way as the frequency and length propositions. Instead it is used to explain why we expect that there are different gaze patterns likely from different types of content.

There are three scoring systems used, one is based on the frequency of which a word occurs in the total content of the data set. The next is based on the length of the words, and the last scoring method is a combination of the two.

The scores are normalised for the Fahey data set as the paragraph and question length are not the same. A t-test is performed on results to check if there is any statistically significant difference between the categories. The t-test probability of $p<0.01$ is the chosen significance level.

The results from the Fahey data set show that there is significant difference ($p<0.01$) between scores tallied for the relevant compared to the less-relevant paragraphs. Similar results are found for the hard and easy paragraphs for all scoring methods. Surprisingly, there was no significant difference between the scores tallied for paragraphs compared to those tallied for questions when using the frequency method and the length method. This is not an expected result since it is believed that the difference between reading a general paragraph and reading a question with multiple-choice answers would be quite different. The combined scoring method does produce results where there is significant difference ($p<0.01$) between the scores tallied for paragraphs compared to those tallied for questions.

Looking at the results in more depth (Appendix E), they show that on average, normalised scores for relevant paragraphs are larger than those for less-relevant paragraphs, indicating that participants are looking at more frequent words in relevant paragraphs as compared to the words they are looking at in less-relevant paragraphs.

In regards to the length-based scores, the average normalised length scores tallied for each participant for the less-relevant paragraphs are higher than for the relevant paragraph. This indicates that participants are looking at longer words in the less-relevant paragraphs compared to these in the relevant paragraphs. It was found that most of the least frequent words are long and most of the more frequent words are short. Since we found that participants are just as likely to look at infrequent words in relevant paragraphs it therefore makes sense that we obtain this result for the length based scoring.

Finally, it was found that on average, normalised scores for less-relevant paragraphs are larger than for relevant paragraphs, indicating that the participants are looking at infrequent long words in the less-relevant paragraphs more than in the relevant paragraphs which is consistent with the above two findings.

For the Sharma data set, there was no statistical difference between eye gaze patterns recorded from hard and easy paragraphs. This is contradictory to the results found for the Fahey data set for the hard and easy paragraphs. Given that the scoring method was initially designed for the Fahey data set this is not completely unexpected. The scoring methods were designed to take into account the fact that the content of the paragraphs are based on the same topic and thus would in general have the same types of words, therefore all paragraphs and even questions could be compared to each other. However, the paragraphs from the Sharma data set are all based on different topics, which causes a skew in word scores. In essence, words that should be low frequency in everyday use may occur at unusually high frequencies just because of the topic of the paragraph. Conversely, words that would be considered more common but happen to occur infrequently due to chance will have the same frequency value as the uncommon words. This could account for the lack of statistically significance difference between categories of paragraphs based on these scoring methods.

Due to the fact that no calibration is performed on the fixations for the Sharma data set this could also be causing results to be skewed. Also, the text shown to participants in Fahey's study took up the entire screen as compared to the text shown to participants in Sharma's study, which used only a small part of the screen. We therefore have great resolution for the Fahey data set, that is, a larger font and word size. Further, normalisation by the number of fixations recorded per paragraph should be performed on the Sharma data as this at least partially normalises for re-reading of text, which many subjects indicated they did, as the time given for each paragraph was longer than required for a single reading.

6.2 Backward and Forward Tracking

The next part of the analysis looked into the different levels of forward and backward tracking of the eye gaze, observed between the categories of text. In essence, this is still looking at the effect words have on fixation location. This exploration particularly looks at the factor of word predictability. In the case where words are difficult to predict we would expect to see more fixations and more backtracking of the eye since predictions may be wrong.

Further, it has been found that regression fixations (backward movement of the eye) are related to comprehension and text difficulty (Just & Carpenter, 1978). This is not always consistent amongst readers; the harder the text is the higher the number of fixations observed from some individuals, others simply exhibit fixations that occur at longer durations, whilst some individuals display an increase in both number of fixation and duration of fixations (Just & Carpenter, 1978).

The proposition is that when reading something difficult more fixations would be recorded because either the words or the concepts are harder for the reader to grasp meaning that they will either display a tendency to fixate on a word longer or exhibit back tracking in order to re-read parts of the text or certain words (Just & Carpenter, 1978, Fisher et al., 1998). Also, part of reading involves your brain anticipating what word follows; your brain does this from context of what you have currently read (Fisher et al., 1998). The second proposition is that in easy paragraphs words are easier to anticipate and concepts are easier to understand so less regressions will be observed. For these reasons we would expect more backtracking - regressions - will be observed in the relevant and harder paragraphs compared to less-relevant and easier paragraphs, respectively. Conversely, we would expect to see more forward tracking in less-relevant and easier paragraphs compared to relevant and harder paragraphs.

Neither the Fahey data set nor the Sharma data set showed that there was a statistically significant difference between the different categories of paragraphs based on any of the movement types. In the case of the Fahey data set, there was statistically significant difference ($p<0.01$) found between the gaze patterns from the paragraphs to those from the questions. As the questions differ greatly to the paragraphs, in readability (Table 17) and structure, this result is expected. Based on previous analysis using a grid system to analyse backtracking the results from comparisons of paragraphs based on content are not what was expected (Vo et al., 2010). Vo et al. (2010) achieved about 80% classification accuracy of eye gaze sequences based on artificial neural networks under a similar grid system and using the backward and forward tracking counts but without normalisation. In Vo et al.'s experiment a 4x5 grid was used as opposed to the smaller grids used in this study. Originally the analysis on the Fahey data set considered a small grid so that it would be more akin to the Vo et al.'s experiment. These results are outlined in the Appendix A, however, it found that there was no statistical difference between eye movements observed between the types of text content.

It is expected that when reading something difficult or complicated you often have to read over it again, or when reading a hard word you have to focus on that word and the words around it to comprehend it. Whereas, when reading something basic you find that you can read it quite fast and that you in fact skip words because your brain can interpret what comes next from context. This was not what was found.

One possible explanation for this is that the participants were shown the paragraphs for the same length of time, which was 90 seconds each in the Fahey experiment and 60 seconds each in the Sharma experiment. It was noted in the Fahey study that this length of time was far longer than the participants needed to read each paragraph. Further the participants were being tested on the content of the paragraphs so they may have tried to "learn" all they could from each paragraph regardless of its difficulty. This could cause the same sorts of behaviours observed whilst reading the easy paragraphs for memorisation, as observed with the hard paragraphs for comprehension.

In this case, the backward and forward tracking analysis would inevitably find no difference between the two types of paragraphs since the behaviour of participants has been altered from the real life situation. However, if the magnitudes of these movements were taken into account then it would seem more likely that a difference would be observed. To examine this analysis further the use of hidden Markov models is used (Section 6.4).

6.3 Reading Comprehension

The correlation between the reading comprehension scores recorded from the Fahey experiment and the statistical analyses of eye gaze patterns in this study are examined. This part of the study could only be performed on the Fahey data set as no reading comprehension tests were performed in the Sharma experiment, as this was not an objective in that experiment. Fahey found no significant correlations between the statistical measures he used and the reading comprehension scores. These measures were for example, distance between fixations and number of fixations per screenshot.

As a preliminary task, the average normalised scores for frequency, length and combined scoring method for each participant for all paragraphs was compared to their total score using the Pearson correlation coefficient. Under no scoring scheme was there any correlation between scores obtained using this study's scoring method and the reading comprehension scores from Fahey's study.

More to the point the study, we want to see if there is a correlation between the specific types of content and the reading comprehension scores obtained. The results from the frequency based scoring system shows that there is a correlation between high frequency scores and high comprehension scores when reading the less-relevant paragraphs but conversely there is a high correlation between lower frequency scores and high comprehension scores when reading the relevant paragraphs.

However, the opposite results are found when the paragraphs are grouped based on text readability measures. Here, high frequency scores correlate to high comprehension scores when reading the easy paragraphs and there is correlation between lower frequency scores and high comprehension scores when reading the hard paragraphs. In hard paragraphs, the low frequency words are likely to confuse the reader and not necessarily give away the deeper meaning of the content. The low frequency words could be complicated and just make comprehending the content more difficult. If the individual understands the content then these more complicated words may be of some benefit but if the individual is unable to understand the content, then complicated words will just make it worse. It is likely that the high and mid frequency words convey the meaning or help string together meaningful sentences, which is why they occur at high frequency. This is not to say that the words "the" and "and" convey much meaning but more that the words such as "tracking" and other high and mid-range frequency words are useful for comprehension in the context of the Fahey data set. We see the opposite trend in the easy paragraphs most likely because as stated before if the individual finds understanding the content achievable then these more complicated words may be of some benefit and since most participants should be able to understand the easy paragraphs, this is why we see a stronger correlation there.

When looking at the length based scoring we see that there is no correlation between the scores from relevant or less-relevant paragraphs and the total comprehension scores. There is however a negative correlation between the length scores from hard paragraphs and the total score, indicating that for hard paragraphs if the participant looked at the longer words then this correlated with a high comprehension score. However, in the easy paragraphs if the participant looked at shorter words than this correlated with high comprehension scores. These are practically the same results as with the frequency based scoring. We know that many of the long words are low frequency and that many of the short words are high frequency. So in hard paragraphs the long words make it harder to comprehend the deeper meaning of the content and in short paragraphs, the long words aid the comprehension of the text.

When looking at the combined scoring method, again we see that there is no correlation between reading comprehension scores and scores obtained from either of the relevant or less-relevant paragraphs. We do see a weak negative correlation between the combined score from the hard paragraphs and the comprehension scores and a weak positive correlation between the combined scores from the easy paragraphs and the comprehension scores. This means that for the hard paragraphs, high scores correlate with high reading comprehension scores and for the easy paragraphs, low scores correlate with high comprehension scores. High combined scores are obtained from looking at long infrequent words and lower combined scores are obtained from looking at shorter more frequent words. So in the hard paragraphs, participants who looked at the longer infrequent words had higher comprehension scores and in the easy paragraphs, participants who looked at shorter more frequent words had higher comprehension scores. This is consistent with the above results for the frequency and length scoring systems alone, so our reasoning for why this is the case still stands.

There is an additional observation to consider here which are that there are significantly different results obtained from the relevant and less-relevant paragraphs versus the hard and easy paragraphs.

Finally, the results from the backtracking analysis were compared to the reading comprehension scores. These results show that when using both the grid and word box techniques to generate movement sequences, there is a medium negative correlation between the backward movements observed over all paragraphs and the total reading comprehension scores. Furthermore, there is a weaker positive correlation between the stationary movements and the total reading comprehension scores.

These results indicate that higher comprehension scores are associated with lower percentages of backtracking movements observed. Further, the results indicate that higher comprehension scores are associated with higher percentages of stationary movements observed. Less backtracking movement could indicate that the individual had less trouble understanding what they are reading and hence less regressions are needed to reread parts of the text just prior to what was read. Instead theses participants with higher comprehension scores had fixations that occurred on the same word. These results are as we would expect based on the literature (Just & Carpenter, 1978). The results are also consistent among the four groupings that further show that these results are what were expected.

Furthermore, when grouping the paragraphs into the subdivisions of relevant and less-relevant as well as hard and easy, the same results are found as when the paragraphs are combined. That is, the same -0.5 correlation coefficient is found for backward tracking when considering only relevant paragraphs and the same for the less-relevant paragraphs. This suggests that there is no connection between the eye movements observed in the types of paragraphs and the reading comprehension scores, when we consider the basic forward, backward and stationary movements of the eye.

In Appendix B is a table summarising the Pearson correlation coefficient for the total comprehension score as well as for the individual comprehension scores; paragraph, question and sentence scores. The correlation coefficients are consistent across scoring methods. It is assumed that the total score gives the highest level of information about the participants' comprehension of the text and therefore is the score of most interest in this study.

6.4 HMM

In the Fahey data set there are three different types of categorisations, first there are the questions that can be compared to the paragraphs. Then there is the initial categorisation of paragraphs into relevant and less-relevant. Finally, there is the categorisation of paragraphs based on the readability analysis into hard and easy paragraphs. There are two ways of generating observation sequences; the first is using the grid system and the second is using the word box system. The sequences are then either left raw or have the stationary points concatenated. Finally, 4 different alphabet sizes for the HMMs were assessed. The first has the basic 3 symbol alphabet which takes into account backward, stationary and forward movement and the remaining alphabets have increasingly more definition about the eye movements.

If we first consider the paragraphs versus the questions we can look at the highest level of differentiation between contents. For this reason it is plausible that there will be the greatest difference between the gaze patterns produced from reading a paragraph compared to reading a question. Using the 3 symbol alphabet we see only by-chance classification of the sequences to original models using either concatenated and non-concatenated sequences. These results confirmed expectations based on the backtracking analysis and that showed no differentiation between the paragraphs and questions based on the simple measures of back, stationary or forward movement.

The results show that using HMM's with an alphabet of 7 symbols using either concatenated and non-concatenated stationary points and either the grid or word boxes technique yields 60-74% accuracy of classification of sequences.

Classification is slightly improved with the use of the 17 symbol alphabet (61-79%) but greatly improved with the 43 symbol alphabet (72-81%). In most cases, the use of concatenation of stationary points improves classification levels. For example, we see that the concatenated sequences generated from the word boxes scheme produce the highest classification of 81% under the 43 symbol alphabet.

The alphabets size are only limited by the number of movements that are capable of being performed. The larger movements (i.e. movement of >20 in the 43 symbol alphabet) were considered the same in the design of this experiment. This was because it is believed that jumps this large would be jumps back to the beginning of a sentence or the paragraph in order to re-read whole sentences and could therefore be considered the same. Future work would benefit from extending the alphabets to include all movement types possible.

Next, we look at classifying the relevant paragraphs and the less-relevant paragraphs. Here the questions are not considered. Under the grid and word boxes generation schemes, using both concatenated and non-concatenated sequences we obtained correct classification of 55-60% for sequences from the HMMs with 3 and 7 symbol alphabets. As with the paragraphs versus the questions, we see that the classification is improved with the inclusion of more information about the movements. We obtain 66-73% classification levels from the 17 symbol models and 70-78% classification levels from the 43 symbol alphabet. Given that this case of differentiation is by far the subtlest these results are very good because it shows that there is high enough differentiation between the gaze patterns generated from reading a relevant paragraph versus those generated from reading less-relevant paragraphs.

When we look at the paragraphs after they have been categorised based on the readability analysis we see quite similar overall results as seen for the relevant versus less-relevant. Using both the grid and the word-boxes, we have 61-67% accuracy at classification under the 3 symbol alphabet models. These are the highest classification levels under the 3 symbol alphabet models we have seen so far and shows that even at the lowest level of movement characterisation there is enough differentiation between sequences generated from viewing hard versus easy paragraphs, that we can get above chance classification of the sequences. These results are unexpected as it was expected that the highest levels of classification from the paragraphs versus the questions case.

Interestingly the classifications were not improved under the use of the 7 symbol alphabet models, with classification only being 61-64%, which is better than the results obtained from the relevant versus the less-relevant case. The results are however improved with the use of the 17 symbol alphabet where we have classification of 59-79%. The best classification results from the study were obtained from the 43 symbol alphabet where we see 77-91% classification. Under both generating schemes and the use of concatenation, we see that there is near perfect classification (85-91%). These results show that there is high differentiation between the gaze patterns produced from reading hard paragraphs versus those generated from reading easy paragraphs.

Analysis of the Sharma data set showed that in most cases the best results are seen when classifying raw (not concatenated) sequences generated from either the grid or the word box schemes and using the models with a 7 symbol alphabet. Under the grid scheme we get an overall accuracy of 67% and for the word boxes we get 77% accuracy. Under the 17 symbol alphabet there are 73% correct classification with no concatenation of stationary points. It must be kept in mind that the analysis of the Sharma dataset is only preliminary and the results may not be as stable as the results with the Fahey data set as no calibration has been performed on the data set. Much further work is required on this data set to bring the analysis in line with the results from Fahey data set.

Interestingly for the Sharma data set, we can see that the sequences with concatenation of stationary points perform worse than those that have no concatenation. This is contrary to what we observe for the Fahey data set where concatenated sequences increase classification accuracy. There are several differences between the Sharma data set and the Fahey data set. The most notable being that the in Fahey's study the text shown to participants took up the entire screen and as a result was quite large. Meanwhile, the text shown to participants in Sharma's study only took up a small part of the screen and as a results is quite a lot smaller than that from the Fahey study. Examples of the screenshots shown to participants in both studies can be seen in Figures 15 and 16. It is reasonable to say that more fixations are required to view the words on the screenshots shown to participants in the Fahey study compared to the Sharma study as the words are so much larger. The whole point of a fixation by the eye is to take in information and the eye can only take in so much information accurately at a fixation point. This means that when a smaller text size is used, the eye can possibly take in more information about a given word as opposed to if it was displayed at a larger size. This raises the point of how models such as the analysis model used here can be quite dependent on the context in which they are used and that models have to be designed for the specific purpose for which they are intended.

Based on the content which the participants read, the sequences of eye movement can be differentiated from each other to some degree, which means that participants produced eye gaze patterns from the various paragraphs, and questions that varied enough that it is detectable. At the lowest level this is the result that was aimed for. The HMM analysis unfortunately provides no insight as to why this may be, unlike the previous statistical analyses performed, but does show that in all three cases sequences can be differentiated well beyond chance classification. We can say that the magnitudes of the forward and backward movements are affected by the different categories of paragraphs to some degree at least. In most cases, we observe that more sequences are correctly classified with the use of more definition in the eye movement (i.e. larger symbol sets). We also see that contrary to the initial belief, the highest level of differentiation can be seen between the hard versus the easy case as opposed to the paragraphs versus the questions. Furthermore, for the Fahey data set, we see that the use of concatenation improves classification results. This is an example of noise reduction and it is believed that further noise reduction of other forms would improve the results.

Kozek (1997) showed the effectiveness of using HMMs to differentiate gaze patterns generated from viewing images of faces. The results of this experiment fit in well with this study. Kozek ran a series of three experiments, the first of which participants were presented two images, one was an abstract piece of art made up of lines and curves loosely approximating those in a human face but not arranged to look like a face, and the other was the image of a face with no expression (denoted neutral). For this experiment there was 81-92% correct classification of sequences. There was quite a large difference between the two images so high differentiation was predicted for the sequences and hence good classification levels. The second experiment consisted of participants viewing a neutral face and happy face (of the same person). There is less difference between the two images so less differentiation in gaze sequences was predicted. As a result the best classification obtained through the HMM analysis was 68%. In the final experiment, participants viewed the neutral face and a sad face (again from the same person). The difference between the sad face and the neutral face was less, so here the greatest classification level was 61%. The results of which are comparable to results from this experiment, where we obtained differing levels of classification based on what was viewed. However, unlike Kozek the greatest classification levels were not found when the biggest differentiation between sequences was expected, i.e. paragraphs versus questions, but instead found when comparing the hard to easy paragraphs. We also obtained higher levels of classification than Kozek (1997) did from the remaining two cases. This may be due to the extra calibration performed in this study, as Kozek had no similar data cleaning or noise reduction stage.

HMM's have been used to differentiate eye gaze patterns generated from either reading or not reading (Gustavsson, 2010). The greatest level of classification obtained from the use of HMM's was 95%. There is a big difference between eye movements when reading and not reading, as the patterns generated whilst reading are very particular in structure. These results are therefore expected. What is unexpected is that we get results almost as good as these based on differentiating eye gaze patterns generating from reading different types of content.

The HMM analysis of the eye gaze patterns has proved to be quite successful at differentiating eye gaze patterns based of movements. The results show that in all three cases, eye movements are different when reading particular types of content. The results show that these differences are most detectable where there is fine granularity taken between movements by the recognition algorithm.

Chapter 7

Conclusion and Further Work

This study investigated differences in eye gaze patterns recorded from individuals who read a series of paragraphs with different types of content. Several methods were used to assess these differences. The analysis of the eye movements based on the length and frequency of the words in the text read proved to be effective at differentiating the patterns generated from reading different types of content for the Fahey data set. We found that for that data set, under all three scoring schemes presented in the study, the categories of relevant versus less-relevant and hard versus easy could be differentiated with high statistical significance ($p<0.01$). The paragraphs and questions could be differentiated under the combined scoring method with statistical significance but not under the scoring schemes separately.

For the Sharma data set, the hard versus easy paragraphs could not be differentiated with statistical significance, but as this is only a preliminary analysis and no calibration has been performed on the data, explanation is provided in the discussion as to why this may be the case.

When considering only the simple forward, backward and stationary eye movements, this did not show any statistically significant difference between the patterns generated from reading different types of content for either data set. However, looking at the eye movements with finer granularity with the use of hidden Markov models, has proved to be very effective at differentiating the patterns generated from reading different types of content from both data sets and under all cases of content division. For the Fahey data set we obtained 91% classification accuracy of eye movement sequences into the hard versus easy categories.

The results from these three analysis techniques show that eye gaze patterns generated from reading different content types can be differentiated based on both the eye movement and on the words which they fixate on.

As the Fahey data set had information recorded for the comprehension of the content read, there was an analysis performed to see if there was correlation between the comprehension scores and the scoring methods used in this study, as well as the backward tracking results. Fahey had not found any obvious correlations between the measures that he used in his study and the reading comprehension scores. There were medium correlations found between each of the analysis techniques used. In particular, we can see that there is a medium negative correlation between the percentage of backtracking observed and the reading comprehension scores. This indicated that lower levels of backtracking are associated with high comprehension scores. Also, the analysis of the scoring methods showed that there are medium correlations between the words viewed in the paragraphs and the comprehension scores achieved. This shows that there are some connections between the eye movements displayed by a person when they read text and their comprehension of that text.

There is much further work that could be performed on the data sets. Due to time limitations, calibrations were not performed on the Sharma data set. This would be the first and foremost piece of work to accomplish in the future. As shown by the Fahey dataset, calibration of the data should provide more accurate results. The Sharma experiment is itself more natural in that the text is displayed in a more realistic manner and the text is shown for a shorter length of time. With this in mind we would expect truer to life results from this data set.

Furthermore on calibration, a more dynamic calibration would improve accuracy of the results. Even in the case of the Fahey data set, only horizontal calibrations were performed, not vertical. Further analysis could be completed on data that has been calibrated in both directions.

Additionally, normalisation of the Sharma data set should be attempted. As the text was shown to participants for equal periods of time in both data sets, it was proposed that normalising by paragraph length may not be a good indicator of time spent reading. Instead normalising by the number of fixation recorded from reading could be of benefit, as good results were achieved on the Fahey data set using this kind of normalisation.

As stated previously, only a very small subset of Sharma's experimental data was used in this study. This is the eye gaze data for only the hard and easy paragraphs. However, there were five categories of paragraphs shown to participants; hard, easy, stressful, calm and neutral. Additional analysis on the remaining 3 categories which were not assessed could provide more insight into the differences in eye gaze patterns generated when reading different types of text.

Additional reduction of noise should be considered as cleaning the data further would inevitably produce strong results. Eye gaze data is inherently noisy and this can be a big problem whilst using the data. This study did very little to clean the data, but as seen from the results from the Fahey data set using HMMs, in most cases, the concatenation of stationary points improved classification levels.

As shown from the results from the HMM analysis, greater definition of the magnitudes of directional movements proves to be more effective at differentiating sequences based on the content class from which they are drawn. It would be of benefit to analyse the data with HMMs with even larger alphabet sizes, with more symbols accounting for forward and backward movement.

As the text is displayed to individuals in a much smaller box than compared to the Fahey data set, the individuals may not have been looking at the text for the whole time. For the Sharma data set the HMM alphabet could be modified to include symbols for when the participants are looking at other areas of the screen, including the "Next" button, or the blank space around the text. It is expected that reading different types of content will cause different eye movements outside of the text area. For example, since reading an easy paragraph is assumed to be much quicker than reading a hard paragraph, we would expect that participants might want to skip to the next paragraph sooner and therefore look at the "Next" button more.

It is also envisioned that these magnitudes could form a pattern that indicate reading type such as learning vs. skimming vs. reading. This is based on the proposal by Carver (1992) who suggested that there are five reading patterns relating to the duration taken to read text (words per minute analysis).

The HMMs used to classify sequences represented the eye movements linearly. This could be extended to look at the eye movements in different dimensions. That is to look at fixation length, saccade length, saccade velocity, and other features of the eye movements themselves.

As alluded to previously, the extended duration of viewing of each paragraph meant that participants could re-read each paragraph many times. Perhaps the most relevant way of looking at the eye gaze data would be to look at each participants' first read through of each paragraph. This however, would be a complex task, as participants may not simply read top to bottom and then top to bottom again. The participants may reread certain sentences or re-read certain parts of the paragraph but not necessarily in order. Furthermore, individuals from different language backgrounds may read differently. So the case of finding the first run through of the text may not be as simple as finding when the participant gets to the bottom of the text and counting that as the first run through.

This study looked solely into aspects of the words that comprised the text shown to participants and the movements of the eye. No account was taken of fixation duration, saccade length or velocity, as well as overall reading duration. These factors have been shown to be important in the analysis of eye movement during reading (Fisher et al., 1998, Richter et al., 2005). Further work could be done in looking at the aspects of the eye movement as well as the aspects of the text.

Although the data provides near endless potential for data analysis, a new experiment dedicated to looking at the difference in eye gaze patterns based on text content could provide more insight into the findings from this study. Potentially, an option would be to repeat Fahey's experiment with a wider array of participants, i.e. not just male computer science students between 18 and 25, and use paragraphs of equal length. Furthermore, the paragraphs could be specially selected to have a much greater differentiation in content, whether that be in structure, complexity or subject matter. Participants could be asked to indicate when they had finished reading and told that their score would be a function of time as well as comprehension.

This study only touched on the connections between eye gaze and reading comprehension. Much more analysis could be performed in this area. The formulation of a model to predict a participants' understanding of what they read would be an overall goal and something of great interest. The model could be used in an experiment to see if a prediction can be made based on their eye gaze patterns generated from reading the paragraphs to how well they will go answering the questions, ranking the paragraphs and composing a big picture sentence about the content. Furthermore, the model could be used in real time to predict the participants' understanding and then compare it to the result. This is far beyond the scope of what was possible in this study but a very worthwhile future goal of such research.

References

Azuma, N., Hirakiyama, A., Inoue, T., Asaka, A., & Yamada, M. (2000). Molecular cell biology on morphogenesis of the fovea and evolution of the central vision. 104 (20), 960-985.

Barton, D. (2007). Literacy: an introduction to the ecology of written language. Wiley-Blackwell.

Burton, L., Westen, D., & Kowalski, R. (2009). Psychology 2nd Edition. Wiley.

Carver, R. P. (1992). Reading rate: Theory, research, and practical implications. Journal of Reading, 36 (2), 84-95.

Engbert, R., & Kliegl, R. (2001). Mathematical models of eye movement in reading: a possible role for autonomous saccades. Biological Cybernetics, 85, 77-87.

Fahey, D. (2009). A Preliminary Investigation into using eye-tracking to analyse a person's reading behaviour. Honours Thesis. School of Computer Science. Australian National University, ACT, Australia.

Fisher, D. L., Reichle, E. D., Pollatsek, A., & Rayner, K. (1998). Toward a model of eye movement control in reading. Psychological Review, 105 (1), 125-157.

Gustavsson, C. J. (2010). Real Time Classification of Reading in Gaze Data. Masters Thesis. School of Computer Science and Engineering. Royal Institute of Technology. Stockholm, Sweden.

Iqbal, S., & Bailey, B. (2004). Using Eye Gaze Patterns to Identify User Tasks. The Grace Hopper Celebration of Women in Computing.

Isaev, A. (2006). Introduction to Mathematical Methods in Bioinformatics. Springer.

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. Vision Research , 49, 1295-1306.

Just, M. A., & Carpenter, P. A. (1978). Inference Processes During Reading: Reflections from Eye Fixations. In J. Senders, D. Fisher, & M. R.A, Eye Movments and the higher psychological functions. Lawrence Eribaum Associates.

Kozek, K. K. (1997). Classification of eye tracking data using hidden markov models. Honours thesis, University of New South Wales, NSW, Australia.

Mason, C. K. (1991). Central Visual Pathways. In E. R. Kandel, & T. M. Jessel, Principles of Neural Science. New York: Elsevier.

McConkie, G., Kerr, P., Reddix, M., & Zola, D. (1988). Eye movement control during reading: I. The location of initial eye fixations on words. Vision Research, 20 (10), 1107-1118.

McConkie, G., Kerr, P., Reddix, M., & Zola, D. (1989). Eye movement control during reading: II. Frequency of refixating a word. Perception & Psychophysics , 46 (3), 245-253.

O'Regan, J. K. (1984). How the eye scans isolated words. In A. G. Gale, & F. Johnson, Theoretical and applied aspects of eye movement research (pp. 159–168).

O'Regan, J. K. (1981). The "convenient viewing position" hypothesis. In R. W. D. F. Fisher, Eye movements: Cognition and visual perception (pp. 289–298). Erlbaum.

Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of the IEEE, 77 (2), 257-286.

Rabiner, L. R., & Juang, B. (1986). An introduction to hidden markov models. IEEE ASSP Magazine.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. Psychological Bulletin, 124 (3), 372-422.

Rayner, K. (1983). Eye movements, perceptual span, and reading disability. Annals of Dyslexia, 33 (1), 163-173.

Rayner, K., & Bertera, J. H. (1979). Reading without a fovea. Science, 206 (4417), 468-469.

Rayner, K., & McConkie, G. W. (1976). What guides a reader's eye movements? Vision Research, 16 (8), 829-837.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. Journal of Experimental Psychology, 81 (2), 275–280.

Reilly, R. G., & O'Regan, J. K. (1998). Eye movement control during reading: A simulation of some word-targeting strategies. Vision Research, 124 (3), 303-317.

Richter, E. M., Engbert, R., Nuthmann, A., & Kliegl, R. (2005). Swift: A dynamical model of saccade generation during reading. Psychological Review, 112 (4), 777-813.

Rimey, R. D., & Brown, C. M. (1991). Controlling eye movement with hidden markov models. International Journal of Computer Vision, 7 (1), 47-65.

Rimey, R. D., & Brown, C. M. (1990). Selective Attention as Sequential Behavior: Modeling Eye Movements With an Augmented Hidden Markov Model. Proceedings: DARPA Image Understanding Workshop, 840-849.

Salojärvi, J., Puolamäki, K., Simola, J., Kovanen, L., Kojo, I., & Kaski, S. (2005). Inferring relevance from eye movements: Feature extraction (Tech Rep No. A82). Helsinki University of Technology, Publications in Coputer and Information Science. (http://www.cis.hut.fi/eyechallenge2005/)

Salvucci, D. (1999). Inferring Intent in Eye-Based Interfaces: Tracing Eye Movments with Process Models. Human Factors in Computing Systems: CHI 99 Conference Proceedings (pp. 254-261). ACM Press.

Salvucci, D., & Anderson, J. (2001). Automated Eye-Movement Protocol analysis. Human-Computer Interaction , 16, 39-86.

Salvucci, D., & Anderson, J. (1998). Tracing Eye Movement Protocols with Cognitive Process Models. Department of Psychology, (p. Paper 48).

Salvucci, D., & Goldberg, J. (2000). Identifying fixation and saccades in eye-tracking protocols. Proceedings of the 2000 symposium on Eye tracking research & applications, (pp. 71-78).

Genie Case - Secret Of The Wild Child –Transcript (March 4, 1997). URL: http://www.pbs.org/wgbh/nova/transcripts/2112gchild.html Accessed: 25 October 2011

Sharma, N. (2011). A Computational Stress Model for Reading. Personal communication.

Simola, J., Salojrvi, J., & Kojo, I. (2008). Using hidden markov model to uncover processing states from eye movements in information search tasks. Cognitive Systems Research, 9 (4), 237-251.

Vo, T., Mendis, B. S., & Gedeon, T. (2010). Gaze Patterns and Reading Comprehension. Neural information processing. Models and applications, 6444, 124-131.

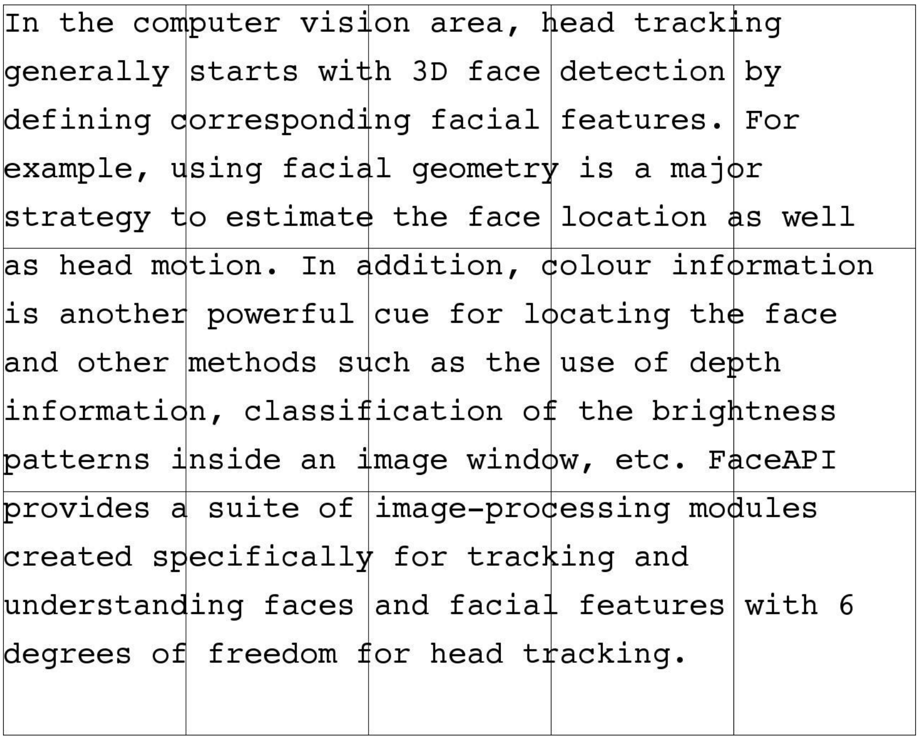
Yatambe, K., Pickering, M. J., & McDonald, S. A. (2009). Lexical processing during saccades in text comprehension. Psychonomic Bulletin & Review, 16 (1), 62-66.

Zhu, D., Gedeon, T., & Taylor, K. (2009). Keyboard before Head Tracking Depresses User Success in Remote Camera Control. In Proceedings of INTERACT (2), 319-331.

Appendix A

Appendix A – 3x5 Grid

The first grid used in the analysis was to get preliminary results on the backtracking and forward tracking analysis. The grid used in the preliminary analysis was a 3x5 grid. The grid system was based on the same techniques used in previous experiments that showed good results (Vo et al., 2010) where a 4x5 grid is used.



In the computer vision area, head tracking
generally starts with 3D face detection by
defining corresponding facial features. For
example, using facial geometry is a major
strategy to estimate the face location as well
as head motion. In addition, colour information
is another powerful cue for locating the face
and other methods such as the use of depth
information, classification of the brightness
patterns inside an image window, etc. FaceAPI
provides a suite of image-processing modules
created specifically for tracking and
understanding faces and facial features with 6
degrees of freedom for head tracking.

Figure 18: Graphical representation of the 3x5 grid imposed on the screenshot

The results from the 5x3 grid show that there is no significant difference between any of the types of content. A significance level of p<0.001 was chosen.

Table 36: T-test probabilities of comparison of content types for the three movement types

	T-Test Probability		
5 x3 grid	Forward	Backward	Stationary
Relevant vs. Less-Relevant	0.29	0.69	0.14
Hard vs. Easy	0.84	0.03	0.03
All Paragraphs vs. Questions	0.85	0.38	0.61

After no significant difference was detected, the grid system was reconsidered as it was predicted that there would be some difference. When we read, although not strictly, we read sentence by sentence. Each grid boxes in the 5x3 grid encompasses 5 lines of text each. So we would expect that with this grid we would obtain a high number of backtracks recorded just simply from the fact that reading sentence by sentence a participant is going to come back to the same grid boxes over and over again. This is why the grid in the analysis, in the main part of the study was designed so that each grid box only incorporates one line of text.

Appendix b

Appendix B – Reading Comprehension Scores

The reading comprehension score recorded for each participant from the Fahey dataset is composed of three parts; the paragraphs score, question score and sentence score (Refer to Fahey 2009 for more details on the scores). All assess different aspects of the comprehension of the text. The Pearson correlation coefficient was also calculated for the different scoring methods and the counterparts of the total comprehension score. The results are shown in the table below.

Table 37 : Pearson correlation coefficient values for the total scores obtained for the content groupings of paragraphs versus the three counterparts of reading comprehension score.

	Paragraph score	**Question score**	**Sentence score**
Scoring Type		**Pearson's R**	
Frequency Based Scoring			
Normalised ave for Relevant paragraphs	-0.08	-0.47	-0.56
Normalised ave for less-Relevant paragraphs	0.04	0.78	0.30
Length Based Scoring			
Normalised ave for Relevant paragraphs	0.00	-0.20	-0.28
Normalised ave for less-Relevant paragraphs	-0.05	-0.23	0.47
Combined Scoring			
Normalised ave for Relevant paragraphs	0.08	-0.41	-0.25
Normalised ave for less-Relevant paragraphs	-0.04	-0.41	0.17

There is no correlation between any of the scoring methods and the paragraph score. There was correlation found between the scoring systems and the question and sentence scores. The correlations are not consistent which could be a result of the fact that the question and the sentence score vary very little amongst participants as the they are out of 5 and 3 respectively.

Table 38: Pearson correlation coefficient values for the total scores obtained for all paragraphs versus the three counterparts of reading comprehension score.

	Paragraph score	**Question score**	**Sentence score**
Movement type		**Pearson's R**	
15x5 grid Results for all paragraphs			
Forward	-0.05	-0.01	0.05
Backward	-0.30	-0.39	-0.58
Stationary	0.25	0.26	0.40
Word Boxes for all paragraphs			
Forward	0.10	0.05	0.25
Backward	-0.28	-0.36	-0.51
Stationary	0.19	0.21	0.25

The Pearson correlation coefficients calculated for the counterparts of the total reading comprehension score are consistent with those found for the total comprehension scores.

As found in the primary analysis of the data, there was little to no difference in correlation coefficients obtained from the paragraphs combined compared to the subtypes of paragraphs. For this reason, the Pearson correlation coefficients were not calculated for any of the divisions of the total reading comprehension score.

Appendix C

Appendix C – 3 Class HMM Analysis

The Fahey data set consists of paragraphs and questions. The questions are divided into two types of content, either relevant and less-relevant or hard and easy. So there is the potential to use three classes in the HMM analysis as opposed to just two. This provides the ability to distinguish not just between two types of content but three, which could eventually be extended to distinguish between many types of content. The following results are a preliminary analysis. Only the HMM's with alphabet size 3 and 7 were used in this analysis.

Table 39 : HMM Classifier with two classes: relevant, less-relevant and questions. Grid and word box sequence generation techniques with raw and concatenated sequences assessed.

HMM Classifier Classes: Relevant vs. Less-Relevant vs. Questions

		Sequences generated from Grid system	Sequences generated from Word Boxes system
HMM alphabet size (M)	Stationary points concatenated	Percentage correctly assigned	Percentage correctly assigned
3	No	35	35
3	Yes	43	43
7	No	42	49
7	Yes	50	55

Table 40 : HMM Classifier with two classes: hard, easy and questions. Grid and word box sequence generation techniques with raw and concatenated sequences assessed.

HMM Classifier Classes: Hard vs. Easy vs. Questions

		Sequences generated from Grid system	**Sequences generated from Word Boxes system**
HMM alphabet size (M)	**Stationary points concatenated**	**Percentage correctly assigned**	**Percentage correctly assigned**
3	No	40	34
3	Yes	46	43
7	No	42	55
7	Yes	51	61

The analysis shows that there is less than chance classification level of sequences using the HMM with alphabet size 3. This result is worse than any results produced by using the 2 class classification scheme. With no concatenation of stationary points, there is still below chance classification of sequences using the HMM with alphabet size 7. These results are improved with the use of concatenation of stationary points and the alphabet size 7, however, the classification is still around chance. These results are not as good as those observed when only using two classes.

Appendix D

Appendix D – Paragraph Content

Fahey Data Set

The following are the paragraphs and questions shown to participants in Fahey’s study. Paragraphs 1 to 5 were labelled relevant and 6 to 10 were labelled less-relevant.

Paragraph 1

In the computer vision area, head tracking generally starts with 3D face detection by defining corresponding facial features. For example, using facial geometry is a major strategy to estimate the face location as well as head motion. In addition, colour information is another powerful cue for locating the face and other methods such as the use of depth information, classification of the brightness patterns inside an image window, etc. FaceAPI provides a suite of image-processing modules created specifically for tracking and understanding faces and facial features with 6 degrees of freedom for head tracking.

Paragraph 2

The head tracking method, called “motion”, operates according to natural human head motion. Assuming initially that the user's head is directly facing the screen, when the user rotates the head to either left or right by a certain angle, the camera will pan in the corresponding direction. It will keep panning the view along that direction until the user moves their head back to the original position. When the user tilts their head up or down by a certain angle, the camera will correspondingly carry out the tilt function and not stop tilting until the head returns to the original position.

Paragraph 3

The head tracking method based on human quick head movements, called “flicking”. Head flicking based interactive control for camera functions is mostly like a switch. When a user quickly rotates his head to either the left or the right direction then moves back to the original position, we consider this to be a head “flicking” along the corresponding orientation, which appropriately turns on the camera to start panning along this direction. When the user flicks to the opposite direction, it will switch the camera movement off and stop at the current position.

Paragraph 4

Our objective results indicate that for this specific experimental setting, keyboards still performed the best by most of the subjects. We believe this is due to the fact that all the participants were quite familiar with using the keyboard, and initially there was no training time for them to get used to the two head tracking control methods. The reason for requiring the subjects to immediately start performing the experiment was to test how well users could pick up the head tracking based remote control. It is clear that our “head motion” based design provides quite comparable performance to the most conventional device (keyboard) even without any training.

Paragraph 5

In this research, we focus on importing computer vision technology to undertake head tracking in interface design for teleoperation activities. The common remote control situation described above is modelled by using a physical game analogue: playing a table soccer game with two handles. This has the advantage of being more compelling for our student experimental subjects than a more abstract task. We use student experimental subjects as we have limited access to the operators. We then propose a novel design applying natural human head gestures for controlling a Pan-Tilt-Zoom camera as an effective approach to solve the camera control problem.

Paragraph 6

Head tracking is a key component in applications such as human computer interaction, person monitoring, driver monitoring, video conferencing, and object-based compression. Recently, one of the most popular ways of applying head tracking is to couple the virtual camera to a user's head position in order to achieve a more realistic and immersive experience of perspective in virtual reality or visual gaming.

Paragraph 7

There have also been attempts to develop head tracking based "hands-free pointing" interface for controlling the mouse cursor, by which a user can point their nose where they wish to place the cursor on a monitor screen. "hMouse" is another head tracking driven camera mouse system, which provides alternative solutions for convenient device control with potential applications for people with disabilities and the elderly.

Paragraph 8

Technology is constantly changing and evolving around us. And, as technology evolves, so does the media that is viewed and accessed through new mediums. New technologies allow for greater flexibility and more accessibility. Such technologies are now rising that pose the threat of making past technologies obsolete.

Paragraph 9

Participants in a research experiment used various sensory methods in order to find pictures of human faces and note the numbers on them with a remote camera view. Participants were then asked to write a research report on their experiences and how or whether they relate to their studies of the World Wide Web. Human-computer interaction (HCI) is always pressing forward into the future, and this experiment is one of many that will help it do so in the years to come.

Paragraph 10

The accuracy and complexity of human vision and movement is a concept which has been studied for years. Extensive study and research has been conducted through out history to create human computer interface designs such has head-tracking technology, joysticks and other control methods of machines to help with (in particular) labour-intensive industries. However technology is constantly striving to imitate natural human movement, which means research and study, is continuously being conducted to help improve and test new designs.

Question 1

The purpose of the user study described was to:

A) Find an optimal camera technique for playing novel games using a remote camera.

B) Determine the usefulness of some new user interfaces that worked by tracking a persons head.

C) To demonstrate that a keyboard is superior to head tracking interfaces.

D) To test some new devices that enable head tracking.

Question 2

An important design choice in the user study described was:

A) That the participants received no training time on either of the new head tracking interfaces.

B) That a cursor could be controlled by the participant pointing their nose.

C) That new technologies were tested that allowed for greater flexibility and more accessibility.

D) That only two new interfaces would be tested because more would cause confusion in the participants.

Question 3

Two types of head tracking were described called "motion" and "flicking", which where:

A) "motion" was used to follow the motion of the table soccer game and "flicking" was used to flick between pictures of peoples faces.

B) "motion" was like the heads natural motion and "flicking" was like flicking a switch on and off.

C) "motion" was like the motion of the table soccer game and "flicking" was like the heads natural flicking action.

D) Neither "motion" nor "flicking" were described in any meaningful way.

Question 4

The device used to track the participants head in the user study was:

A) A device that is known as the "hMouse".

B) A visor with acceleration sensors attached.

C) A system of electrodes that measure the position of the muscles in the participants head and neck.

D) A camera and various computer vision technologies.

Question 5

A key technology for use in this study is:

A) Human vision and movement.

B) 3D face detection of face location and movement.

C) Human-computer interaction (HCI) for the World Wide Web.

D) Computer vision technology for video conferencing.

Sharma Data Set

The following are the paragraphs shown to participants in Sharma's study.

Paragraph Easy 1

She walked into the bathroom and went to the wash basin. She took the cap off the tube of toothpaste. She squeezed some toothpaste onto her toothbrush. She turned on the tap by turning the tap handle. The water from the tap was neither too hot nor too cold. She brushed her upper teeth. Then she spit out some toothpaste. She brushed her lower teeth. Then, again she spit out some more toothpaste. She rinsed out her toothbrush with water. She put the toothbrush back into the toothbrush holder. She put some water into a cup and rinsed out her mouth. She spit out the water. After looking at her white teeth in the mirror, she walked out of the bathroom.

Paragraph Easy 2

Floriade is Australia's celebration of spring. Featuring more than a million blooms as a backdrop, Commonwealth Park in Canberra, explodes into colourful bloom, with fantastic entertainment, displays and a whole program celebrating a set theme each year. Floriade 2011 will run from Saturday 17 September to Sunday 16 October 2011. Be captivated by exhibitions and displays to inspire a taste of fresh and healthy living, spring sensations including fashion displays, floral art, crafts and fine art, horticultural displays and practical tips and advice on home gardening and outdoor living from noted celebrities and experts. See dazzling entertainment including headline music acts, be enthralled by street performers, comedy acts, community performances and dance.

Paragraph Easy 3

Which part of your body lets you read the back of a cereal box, check out a rainbow, and see a cricket ball heading your way? Which part lets you cry when you are sad and makes tears to protect itself? Which part has muscles that adjust to let you focus on things that are close up or far away? The eye. Your eyes are at work from the moment you wake up to the moment you close them to go to sleep. They take in tons of information about the world around you - shapes, colours, movements, and much more. Then they send the information to your brain for processing so the brain knows what is going on outside of your body.

Paragraph Hard 1

The location of the landmark points are extracted by AAM from every gallery image. Landmark points are normalised to remove any minor variation. This is done by applying Euclidean transformations and taking landmark points representing the corner of the eyes and tip of the nose as reference points. These normalised landmark points can be used to generate synthetic images by predicting new landmark locations via the learnt regression model and warping the texture from the frontal gallery image. PAW is used for warping the texture from frontal images to the new landmark locations. All invalid pixel locations inside the convex hull of the canonical shape are filled by their nearest-neighbours and background pixels are filled with the mean value of texture to complete the synthetic image.

Paragraph Hard 2

To date, the contribution of M(y)-cell activity to magnocellular dysfunction in dyslexia is unknown. Given the precedent in glaucoma research, in which there is good evidence for large diameter cell loss in the retina to suggest that a magnocellular deficit in dyslexia may have its origins in a visual response mediated at least partially by the M(y) system. The current study therefore is designed to assess the proficiency of the M(y) system in reading disabled children by exploiting the spatial frequency doubling response characteristic of the M(y)-cells. If the functional integrity of the magno system is compromised in dyslexic children at the level of M(y)-cells, then dyslexic readers should be less sensitive to the frequency doubling illusion than normal readers.

Paragraph Hard 3

Representations must be chosen carefully. Direct chromosome representations are restrictive and contribute to a higher computation time in general. A better representation that can be used is a non-direct chromosome representation motivated by the setup of constraint satisfaction methods. However, these representations may store instructions on developing a timetable instead of just representing information. Recent research on representing timetables in evolutionary algorithms by Linear Linkage Encoding has produced good results. It is based on the concept of sets and paths. Loci are linked in a chromosome and a path is formed to represent a partition set. In a partition set, a locus can be reached from another locus in the same partition set.

Appendix E - Scoring Analysis Results

The following are the raw results from scoring analysis which the paired two tail t-tests were performed on for the two data sets.

Fahey Data Set

Table 41: Average normalised Frequency scores for each paragraph type and each participant.

	Average Frequency scores for:			
Participant	Relevant Paragraphs	Less-Relevant Paragraphs	Hard paragraphs	Easy Paragraphs
0675819234	11.6	9.2	9.7	9.6
1234678590	11.8	8.0	8.7	9.9
1627384950	11.8	7.2	7.2	9.0
2347581906	13.0	8.4	9.0	10.5
3451890267	13.4	7.8	7.8	11.3
4592136780	12.7	7.9	8.3	10.0
5120364789	12.8	8.5	7.4	10.8
6718902345	12.2	9.0	9.0	11.8
6789123045	12.3	8.1	8.7	10.4
7892036451	12.8	8.4	8.9	10.5
8906345712	12.5	8.3	8.2	10.5

Table 42: Average normalised Length scores for each paragraph type and each participant.

	Average Length scores for:			
Participant	Relevant Paragraphs	Less-Relevant Paragraphs	Hard paragraphs	Easy Paragraphs
0675819234	5.0	5.7	5.7	5.4
1234678590	5.2	5.8	6.1	5.2
1627384950	5.4	5.7	6.0	5.4
2347581906	5.3	5.8	5.8	5.4
3451890267	5.1	5.7	5.9	5.2
4592136780	5.3	5.4	5.8	5.0
5120364789	5.2	5.4	5.6	5.1
6718902345	5.2	5.6	6.0	4.9
6789123045	5.1	5.6	5.9	5.0
7892036451	5.0	5.5	5.7	5.1
8906345712	5.2	5.8	6.1	5.3

Table 43: Average normalised Combined scores for each paragraph type and each participant.

	Average Combined scores for:			
Participant	Relevant Paragraphs	Less-Relevant Paragraphs	Hard paragraphs	Easy Paragraphs
0675819234	2.4	3.4	3.3	3.2
1234678590	2.5	3.7	3.8	3.0
1627384950	2.8	3.8	3.7	3.5
2347581906	2.7	3.5	3.4	3.2
3451890267	2.6	3.5	3.6	2.9
4592136780	2.6	3.6	3.5	3.1
5120364789	2.4	3.3	3.5	2.7
6718902345	2.5	3.4	3.6	2.6
6789123045	2.2	3.6	3.6	2.8
7892036451	2.3	3.2	3.0	2.9
8906345712	2.6	3.6	3.8	2.8

Table 44: Average normalised Frequency scores for all paragraphs and questions and for each participant.

	Average Frequency scores for:	
Participant	**All Paragraphs**	**Questions**
0675819234	14.0	14.5
1234678590	13.4	14.1
1627384950	12.8	13.1
2347581906	14.5	14.7
3451890267	14.2	14.9
4592136780	13.9	14.7
5120364789	14.5	13.6
6718902345	14.4	14.8
6789123045	13.8	15.2
7892036451	14.4	13.9
8906345712	14.1	14.7

Table 45: Average normalised Length scores for all paragraphs and questions and for each participant.

	Average Length scores for:	
Participant	**All Paragraphs**	**Questions**
0675819234	5.3	4.9
1234678590	5.5	4.9
1627384950	5.6	5.8
2347581906	5.5	5.4
3451890267	5.4	6.0
4592136780	5.4	5.3
5120364789	5.3	5.5
6718902345	5.4	5.2
6789123045	5.4	4.8
7892036451	5.2	5.2
8906345712	5.5	5.3

Table 46: Average normalised Combined scores for all paragraphs and questions and for each participant.

	Average Combined scores for:	
Participant	**All Paragraphs**	**Questions**
0675819234	2.6	1.7
1234678590	2.8	1.8
1627384950	3.0	2.0
2347581906	2.8	1.8
3451890267	2.8	2.2
4592136780	2.8	1.8
5120364789	2.6	1.7
6718902345	2.6	1.8
6789123045	2.6	1.7
7892036451	2.4	2.0
8906345712	2.8	1.9

Sharma Data set

Table 47: Average Frequency scores for hard and easy paragraphs and for each participant.

	Average Frequency scores for:	
Participants	**Hard paragraphs**	**Easy Paragraphs**
EH4	2429	2230
EH5	1110	1105
EH6	5131	2698
EH8	752	405
EH9	783	1341
HE3	3049	2168
HE4	4524	2533
HE8	2595	1452
HE9	2227	1895
HE10	1476	983

Table 48: Average Length scores for hard and easy paragraphs and for each participant.

	Average Length scores for:	
Participants	**Hard paragraphs**	**Easy Paragraphs**
EH4	969	1388
EH5	529	594
EH6	2823	1710
EH8	392	243
EH9	423	499
HE3	1360	1175
HE4	2138	1539
HE8	1389	640
HE9	1082	1092
HE10	719	550

Table 49: Average Combined scores for hard and easy paragraphs and for each participant.

	Average Combined scores for:	
Participants	**Hard paragraphs**	**Easy Paragraphs**
EH4	569	825
EH5	318	371
EH6	1755	1083
EH8	266	163
EH9	261	290
HE3	810	695
HE4	1319	1005
HE8	828	352
HE9	661	662
HE10	445	344

Appendix F - Backward and Forward tracking Analysis Results

The following are the raw results from backward and forward tracking analysis which the paired two tail t-tests were performed on for the two data sets.

Fahey Data Set

Table 50: The average proportions of forward, backward and stationary movements recorded for each participant for the relevant and less-relevant categories of paragraphs from sequences generated from the 15x5 grid scheme.

	15x5 Grid Results					
	Relevant			**Non-Relevant**		
Participants	**Forward**	**Backward**	**Stationary**	**Forward**	**Backward**	**Stationary**
0675819234	0.24	0.12	0.64	0.23	0.10	0.66
1234678590	0.26	0.26	0.48	0.26	0.25	0.49
1627384950	0.23	0.24	0.52	0.26	0.26	0.48
2347581906	0.25	0.28	0.46	0.27	0.26	0.47
3451890267	0.24	0.27	0.48	0.26	0.25	0.49
4592136780	0.27	0.27	0.45	0.27	0.28	0.45
5120364789	0.29	0.29	0.41	0.3	0.30	0.4
6718902345	0.23	0.24	0.52	0.26	0.24	0.49
6789123045	0.35	0.27	0.37	0.34	0.29	0.35
7892036451	0.25	0.27	0.48	0.27	0.25	0.47
8906345712	0.27	0.27	0.46	0.27	0.28	0.45

Table 51: The average proportions of forward, backward and stationary movements recorded for each participant for the hard and easy categories of paragraphs from sequences generated from the 15x5 grid scheme.

15x5 Grid Results						
	Hard			Easy		
Participants	Forward	Backward	Stationary	Forward	Backward	Stationary
0675819234	0.24	0.09	0.66	0.24	0.11	0.65
1234678590	0.26	0.27	0.47	0.27	0.23	0.49
1627384950	0.24	0.26	0.5	0.26	0.25	0.49
2347581906	0.28	0.27	0.44	0.26	0.26	0.48
3451890267	0.26	0.25	0.48	0.27	0.27	0.45
4592136780	0.27	0.28	0.44	0.28	0.28	0.44
5120364789	0.32	0.29	0.38	0.28	0.27	0.43
6718902345	0.26	0.25	0.48	0.25	0.23	0.51
6789123045	0.33	0.31	0.35	0.35	0.28	0.36
7892036451	0.27	0.27	0.46	0.26	0.25	0.48
8906345712	0.27	0.28	0.45	0.27	0.28	0.45

Table 52: The average proportions of forward, backward and stationary movements recorded for each participant for the relevant and less-relevant categories of paragraphs from sequences generated from the word boxes grid scheme.

Word Boxes Results						
	Relevant			Non-Relevant		
Participants	Forward	Backward	Stationary	Forward	Backward	Stationary
0675819234	0.30	0.14	0.56	0.28	0.12	0.59
1234678590	0.28	0.29	0.42	0.29	0.28	0.42
1627384950	0.25	0.28	0.47	0.27	0.28	0.44
2347581906	0.26	0.34	0.40	0.30	0.30	0.39
3451890267	0.26	0.30	0.43	0.28	0.25	0.47
4592136780	0.30	0.30	0.40	0.28	0.30	0.42
5120364789	0.32	0.31	0.36	0.30	0.31	0.38
6718902345	0.25	0.28	0.47	0.26	0.26	0.47
6789123045	0.41	0.27	0.31	0.37	0.31	0.3
7892036451	0.27	0.32	0.41	0.31	0.27	0.42
8906345712	0.28	0.29	0.42	0.28	0.27	0.44

Table 53: The average proportions of forward, backward and stationary movements recorded for each participant for the hard and easy categories of paragraphs from sequences generated from the 15x5 grid scheme.

Word Boxes Results						
	Hard			**Easy**		
Participants	**Forward**	**Backward**	**Stationary**	**Forward**	**Backward**	**Stationary**
0675819234	0.30	0.11	0.59	0.27	0.14	0.59
1234678590	0.30	0.28	0.41	0.30	0.29	0.40
1627384950	0.28	0.28	0.44	0.25	0.27	0.48
2347581906	0.31	0.32	0.37	0.30	0.30	0.40
3451890267	0.26	0.26	0.48	0.29	0.29	0.41
4592136780	0.28	0.30	0.42	0.29	0.31	0.40
5120364789	0.33	0.30	0.37	0.31	0.31	0.37
6718902345	0.25	0.28	0.46	0.26	0.27	0.46
6789123045	0.35	0.36	0.28	0.4	0.28	0.32
7892036451	0.30	0.28	0.42	0.3	0.29	0.41
8906345712	0.27	0.26	0.46	0.27	0.29	0.44

Table 54: The average proportions of forward, backward and stationary movements recorded for each participant for all paragraphs and for questions from sequences generated from the 15x5 grid scheme.

15x5 Grid Results						
	All paragraphs			**Questions**		
Participants	**Forward**	**Backward**	**Stationary**	**Forward**	**Backward**	**Stationary**
0675819234	0.24	0.11	0.65	0.25	0.18	0.57
1234678590	0.26	0.25	0.49	0.28	0.26	0.46
1627384950	0.24	0.25	0.50	0.25	0.27	0.48
2347581906	0.26	0.27	0.46	0.27	0.27	0.46
3451890267	0.25	0.26	0.49	0.24	0.25	0.51
4592136780	0.27	0.28	0.45	0.27	0.29	0.43
5120364789	0.29	0.29	0.41	0.31	0.31	0.38
6718902345	0.24	0.24	0.51	0.27	0.26	0.47
6789123045	0.35	0.28	0.36	0.35	0.32	0.33
7892036451	0.26	0.26	0.47	0.29	0.27	0.44
8906345712	0.27	0.27	0.46	0.29	0.30	0.41

Table 55: The average proportions of forward, backward and stationary movements recorded for each participant for all paragraphs and for questions from sequences generated from the word boxes scheme.

Word Boxes Results						
	All paragraphs			Questions		
Participants	**Forward**	**Backward**	**Stationary**	**Forward**	**Backward**	**Stationary**
0675819234	0.29	0.13	0.58	0.28	0.19	0.53
1234678590	0.28	0.29	0.42	0.29	0.29	0.42
1627384950	0.26	0.28	0.46	0.28	0.28	0.44
2347581906	0.28	0.32	0.40	0.28	0.28	0.44
3451890267	0.27	0.28	0.45	0.25	0.26	0.49
4592136780	0.29	0.30	0.41	0.29	0.29	0.42
5120364789	0.31	0.31	0.37	0.31	0.30	0.38
6718902345	0.25	0.27	0.47	0.27	0.27	0.45
6789123045	0.39	0.29	0.31	0.36	0.31	0.33
7892036451	0.28	0.3	0.42	0.31	0.29	0.39
8906345712	0.28	0.28	0.43	0.30	0.30	0.39

Sharma Data set

Table 56: The average proportions of forward, backward and stationary movements recorded for each participant for hard paragraphs and easy questions from sequences generated from the 15x5 grid scheme.

15x6 Grid Results						
	Hard Paragraphs			Easy Paragraphs		
Participants	**Forward**	**Backward**	**Stationary**	**Forward**	**Backward**	**Stationary**
EH4	52.7	36.3	106.7	73.0	43.3	214.7
EH5	21.0	11.7	66.0	21.7	19.0	87.7
EH6	108.3	84.0	384.0	98.7	86.7	283.7
EH8	35.3	32.0	22.3	21.3	21.7	17.7
EH9	23.0	15.0	44.3	31.7	20.3	76.0
HE3	93.0	75.7	136.7	92.7	82.7	113.0
HE4	101.0	65.7	244.3	81.0	57.0	248.7
HE8	69.3	51.7	174.3	43.7	35.0	83.0
HE9	56.3	51.3	127.3	59.7	59.7	138.3
HE10	56.7	51.0	54.0	41.3	37.3	56.3

Table 57: The average proportions of forward, backward and stationary movements recorded for each participant for hard paragraphs and easy questions from sequences generated from the word boxes scheme.

Word Boxes Results						
	Hard Paragraphs			Easy Paragraphs		
Participants	**Forward**	**Backward**	**Stationary**	**Forward**	**Backward**	**Stationary**
EH4	60.7	38.0	91.7	87.3	46.0	171.7
EH5	24.0	12.3	59.0	30.0	21.0	75.0
EH6	122.7	84.0	331.3	94.3	67.3	195.0
EH8	29.7	29.7	14.3	20.3	20.3	12.0
EH9	21.3	14.0	38.7	39.3	24.0	58.7
HE3	94.7	69.7	109.3	88.3	76.3	97.3
HE4	114.3	67.7	221.3	82.7	52.7	197.7
HE8	72.7	47.7	147.3	45.3	30.7	66.0
HE9	60.7	47.3	99.7	62.3	57.7	115.7
HE10	56.0	47.7	39.3	35.7	34.0	45.3