Running head: A FRAMEWORK TO DETERMINE SIMULTANEOUS GAZE PATTERNS AND FIXATIONS

A Methodological Framework for Capturing Relative Eye Tracking Coordinate Data to Determine Gaze Patterns and Fixations from Two or More Observers.

Joanna C Scurr, Jenny Page, and Heather Lunt

University of Portsmouth

Corresponding Author: Joanna Scurr

Department of Sport and Exercise Science,

Spinnaker Building,

University of Portsmouth,

Cambridge Road,

Portsmouth,

UK.

Email: Joanna.scurr@port.ac.uk

Abstract

Few studies have published the methodologies used to analyse simultaneous gaze behaviours and recurrent fixations whilst observers are viewing dynamic scenes and moving their heads. This study aimed to develop a methodological framework to assess simultaneous gaze behaviours and recurrent fixations in pre-determined areas of interest, whilst accounting for head movement and non-standardised observer positioning. Gaze coordinates were recorded during six trials where a single participant focused on the centre of a video image and moved his head in 6DOF. Markers were positioned on the image corners. Eye tracking equipment recorded the video image and gaze behaviours (cross hair) which were uploaded to SIMI motion analysis. Corner markers were digitized to determine image position as the head moved and used to calculate new gaze coordinates relative to head movement. To account for non-standard participant positioning, any gaze coordinates outside the image were excluded. Across all trials the error between measured and calculated coordinates was acceptable (<3.5%CV). Frequencies and durations of fixations (≥100 ms within 1° visual angle) within six areas of interest were reported and compared well to manual calculations. This methodology was then assessed using participant dyads (n=6) simultaneously observing the same image. Recurrent fixations were determined using a hierarchical model, and compared well to manual analysis. This study presents a valid and reliable methodological framework for determining fixation frequency, duration and location from multiple observers, accounting for head movement and non-standard positioning. This framework facilitates the analysis of simultaneous oculomotor variables, improving ecological validity and reducing environmental constraints.

Introduction

There are a number of occupations that require two people to work together to observe and make decision about a common display. For example, CCTV operators often work in teams when viewing and reviewing footage. Airplane pilots and coxswains of boats often work in teams when navigating to a destination. Beach lifeguards often works in pairs when making decisions about preventative action to promote safety and potential rescues. In these situations, one observer may not be aware of what the other is attending to at any moment. Despite the numbers of occupations that may be facilitated by coordinated approaches to observation, the research in this area is limited.

The work that has been completed in interpersonal coordination has focused on how individual's respond to others physical movements (Bangerter, 2004; Clark & Krych, 2004; Richardson, Marsh, Isenhower, Goodman, & Schmidt, 2007a; Schmidt & O'Brien, 1997) or verbal communication (Richardson & Dale, 2005; Richardson, Dale & Kirkham, 2007b). However, as with the occupations previously mentioned, it is often visual information (what we see in our environment) rather than physical contact or auditory information that is the primary means by which behavior is coordinated (Marsh, Richardson, Baron, & Schmidt, 2006). The literature that does exist in the area of visual interpersonal coordination has primarily focused on training an individual's gaze behaviour using another's scan path (e.g., Litchfield, Ball, Donovan, Manning, & Crawford, 2008; 2010; van Gog, Jarodzka, Scheiter, Gerjets & Pass, 2009; Velichkovsky, 1995). For example, Velichkovsky (1995) found that time to complete three puzzles was significantly shorter by novices when they were able to speak freely with an expert as well as see the expert's gaze behavior compared to when they could only speak freely with the expert. Furthermore in a second experiment, it was also found that novices were quicker at completing the three puzzles when experts were able to view the novices gaze behavior and speak freely compared to just being able to speak freely. Conversely, van Gog et al. (2009) found that success on a computer-based leapfrog task was not significantly influenced by the addition of eye movement data (in conjunction with the solution to the task) compared to being shown the solution alone. Additionally, when the task was completed for the second time the group who saw the solution to the task, heard the experts underlying thought processes and saw the eye movements performed significantly worse than the group who saw the solution to the task and heard the experts underlying thought processes. However, there has been little research investigating visual interpersonal coordination when two individuals are performing a task simultaneously using a visual ‘common ground’ (Richardson & Dale, 2005; Richardson et al., 2007b). More research in the area of visual interpersonal coordination may allow a greater understanding of how two or more people can work together to optimise their visual performance.

A key issue in such research is the problem of combining two or more people’s eye movements into a uniform scale when the occupations require their employees to be in non-standard positions (not directly aligned with the field of view) and to move their heads to interact with their environments. Given the paucity of research in this area, a systematic approach is required to analyzing combined gaze behaviours when the observers are able to move their heads. Before considering potential solutions, a number of factors inherent in eye-movements studies need to be considered such as the different systems that are available, the oculomotor measures that are important and the issues of working with dynamic scenes.

A variety of eye-tracking systems have been used to assess gaze behaviours when performing a range of individual tasks including reading (e.g., Reichle, Reineberg, & Schooler, 2010; Kanonidou, Proudlock, & Gottlob, 2010; Bucci, Nassibi, Gerard, Bui-Quoc, & Seassau, 2012), everyday tasks (Land, Mennie, & Rusted, 1999; Pelz & Canosa, 2001), lifeguarding (Page, Bates, Long, Dawes, & Tipton, 2011) and medical diagnostics (e.g., Wood, Knapp, Rock, Cousens, Roobottom, & Wilson, 2013; Litchfield et al., 2008; 2010); these include commercially available eye tracking systems such as the Tobii x50 eye tracker (Tobii Technology, Stockholm, Sweden; Litchfield et al., 2008); EyeLink 1000 eye tracker (SR Research, Ottawa, Ontario, Canada, Reichle et al., 2010); EyeLink Eye Tracker (SensoMotoric Instruments GmbH, Berlin, Germany; Kanonidou et al., 2010); and Mobile Eyebrain Tracker (Mobile EBT®, e(ye)BRAIN, www.eye-brain.com; Bucci et al., 2012). Once eye movement data have been collected a variety of software packages are used to analyse the eye movements including MeyeAnalysis (e(ye)BRAIN, www.eye-brain.com, France; Bucci et al., 2012); Pegasus software (Reichle et al., 2010); Spike2 software (Cambridge Electronic Design, Ltd., Cambridge, UK; Kanonidou et al., 2010) and manual coding (Land et al., 1999; Pelz & Canosa, 2001). However, the systems that analyse data collected from non-computer based stimuli (e.g., in situ data collection where head movement is permitted) do not offer the ability to semi automatically analyse data from two participants observing the same field of view.

Existing data analyses software of the various systems often calculate many oculomotor variables including, number of fixations, average fixation duration, dwell time, saccades, number of blinks and pupil diameter. Perhaps the most frequently cited oculomotor variable is the fixation. The use of visual fixations in eye movements research dates back to the 1930's (Vernon, 1931) and is still used as a key variable (Mason, Pluchino, Tornatora, & Ariasi, 2013). The frequency of the use of fixation measurements may be because it is thought to indicate the amount and importance of processed information. Indeed, Carpenter and Just (1976, p. 139) suggested that rapid mental operations of the central processor can be revealed by an analysis of eye fixations during a task involving visual input. Furthermore, Mackworth and Morandi (1967) made comparisons between visual fixations on, and verbal estimates of, the relative importance of regions within photographs. They found that the regions that were rated highly for informativeness produced the highest fixation frequency. Given the importance of visual fixations, any analysis of simultaneous gaze behaviours should report this variable.

Despite the importance of visual fixations, the algorithms used within software programmes to establish visual fixations are often not readily available. Additionally, within the eye-tracking literature there are a variety of definitions for fixations which include: fixating on a target (diameter 0.5^{o}) for >250 ms (Bucci et al., 2012), gaze with a spatial deviation threshold of 19 pixels for 10 samples (Ryan, Duchowski, Vincent & Battisto, 2010) and keeping the gaze within a radius of 50 pixels for at least 100 ms (Litchfield et al., 2008). Interestingly, Reichle et al. (2010) categorised fixations of <80 ms and

>1,000 ms due to the nature of the task (mindless reading) even though fixations of these durations are typically discarded as outliers in eye movement studies (Inhoff & Radach, 1998; Liversedge, Paterson, & Pickering, 1998). Moreover, Kuhn and Findlay (2010) analysed a fixation based on analysis of one frame using a 500 Hz eye-tracker. Given the variation in existing systems, any methodological framework should allow flexibility when determining fixation length and diameter. Furthermore, of importance to this framework, is the ability to determine recurrent fixations. For the purpose of this framework, recurrent fixations are defined when both observers fixate within an intra-individual visual angle with a diameter of 1° for a period of >100 ms, and these fixation occur at the same time within an inter-individual visual angle with a diameter of 1°, these limits are based on recurrence analysis (Richardson & Dale, 2005). It may also important when designing a framework to analyse simultaneous participants data to understand the distance between gaze points when participants are not recurrently fixating, and therefore Euclidian distances between gaze points can also be analysed to understand differences and similarities in smooth pursuit eye tracking strategies.

Once the algorithms for recurrent fixations are established there is often a need to state the location of fixations within the field of view. Some analyses software require the head to be static and in the same position for all participants and therefore are able to produce areas of interest (AOI) templates that can be employed for all participants within the experiment. Other analyses systems that are designed for mobile eye tracking (e.g., Gazetracker) require that AOIs are added post hoc and although the AOIs can be moved in relation to head movement this often requires a frame by frame approach. Although effective, the time taken for manual coding can be prohibitive for research which involves long duration tasks. Other systems (e.g., tobii eye trackers) use automatic marker detection which allows the aggregation of multiple participants' eye movements. However, if a system does not have built in capacity for marker detection and participants are moving their heads, analyses of multiple participant data is problematic. Ryan et al. (2010) suggested that adding trackboxes is a solution to the head movement issue. The trackboxes follow particular features (bright and dark spots) within a display and therefore enable automatic analyses, but this process could be complicated by dynamic, rapidly changing fields of view. However because Ryan et al. (2010) were primarily concentrating on individual eye movement data, they did not attempt to explain how analyses from two (or more) people can be combined. Therefore any methodological framework should permit the addition of AOIs whilst accounting for head movement and enabling non-standard observer positioning.

Although there are studies that have used semi-automatic analyses of eye movement data to measure fixations from individual participants when the head was freely moving (Ryan et al., 2010) and other studies that have analysed simultaneous eye movements, whilst limiting head movement (Brennan, Hanna, Zelinsky & Savietta, 2012; Clark & Gergle, 2012; Richards & Dale, 2005; Richardson et al., 2007), no processes have been published which combine these elements, whilst enabling non-standard observer positioning, automatic detection of areas of interest and offering explicit definitions of individual and recurrent fixations. Therefore, the primary aim of the framework was to address the problem of analysing eye movement whilst viewing dynamic scenes, allowing for head movement and subsequently combining these data to determine when and where recurrent fixations occur. The study aims to develop a methodological framework that assesses simultaneous gaze behaviours and recurrent fixations in pre-determined areas of interest, whilst accounting for head movement and non-standard observer positioning. This aim was investigated using a three stage process;

Part 1: developing a process to determine gaze coordinates independent to the movement of the head.

Part 2: establishing the frequency and duration of fixations within pre-determined areas of the field of view.

Part 3: identifying gaze patterns and recurrent fixations from two observers, simultaneously observing the same display from a non-standard position.

Part 1: Determining Gaze Coordinates Relative to the Movement of the Head

1.1 Method.

For part 1 a single male participant was used, with no sight impairments. A video image (blank blue screen, Figure 1) was positioned directly in line with the participant's eye (horizontally and vertically) and the participant maintained a standardised position and orientation throughout all trials (Figure 1). The participant then undertook seven validation trials (Figure 2), each lasting 5 s. During these trials the participant focused on the marker in the centre of the video image whilst tilting his head to the left (trial 1, Figure 2), right (trial 2), up (trial 3), down (trial 4), turning his head to the left (trial 5), right (trial 6), and finally, focusing on each corner of the image in turn (trial 7).

INSERT FIGURE 1 NEAR HERE

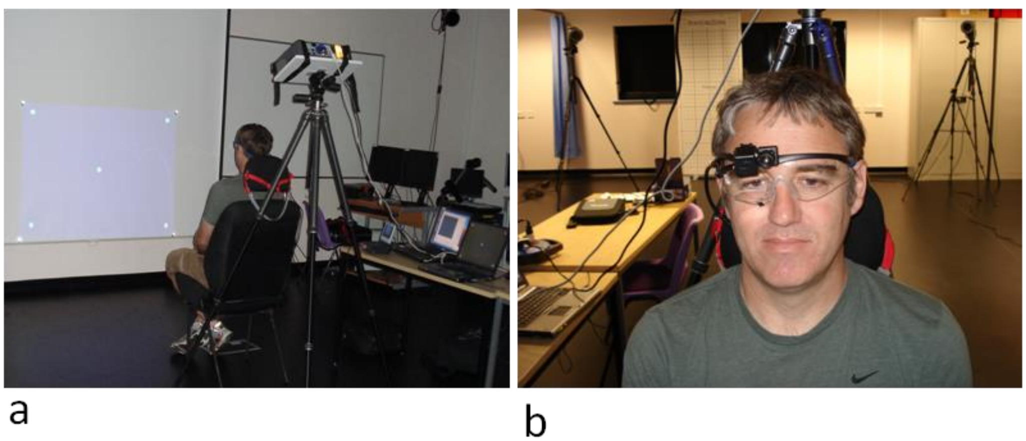


Figure 1. a) Experimental set up and b) central head position.

INSERT FIGURE 2 NEAR HERE

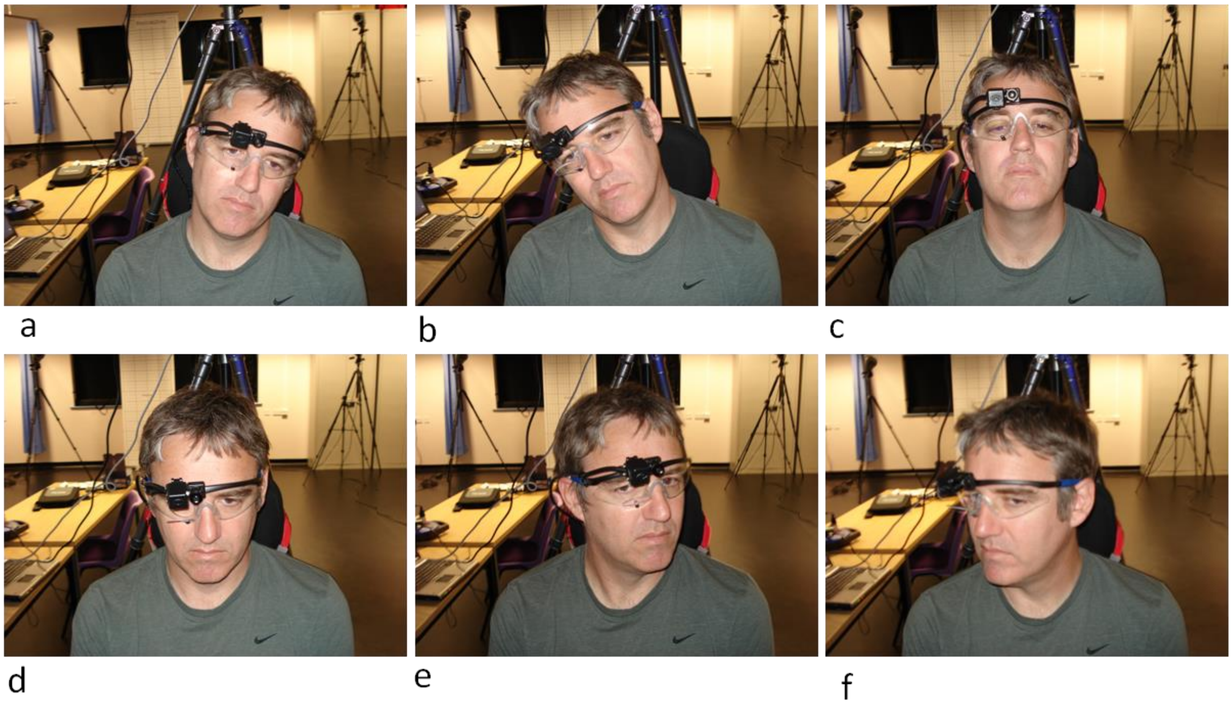


Figure 2. Head movements for validation trials where a) head tilt left, b) head tilt right, c) up, d) down, e) head turn left, and f) head turn right.

The movement of the participant's right eye was tracked using an Applied Science Laboratories MobileEye™ (ASL, Bedford, MA). When recording eye tracking data white LED markers of 3 mm diameter were attached to the projection screen in each corner of the video image. An additional LED marker was placed approximately in the centre of the video image (Figure 1 & 3). These markers were highly contrasting with the image to enable easy identification of the markers during the analysis phase. The markers remained stationary throughout all trials. The distance of the participant's eye to the image centre was noted as 2.5 m, the distance between the top horizontal and left vertical markers was noted as 1.62 m and 1.23 m.

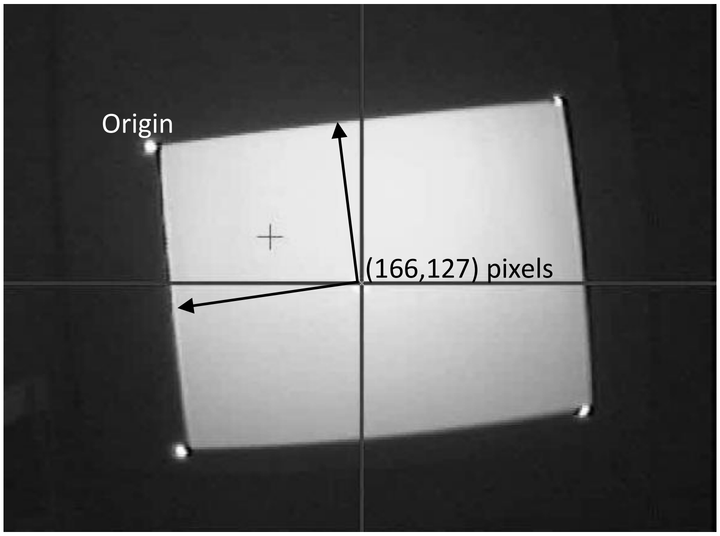


Figure 3. Example image from trial 1 (head tilt left). The grey screen represents the video image. The LED markers on each corner and in the centre of the screen are visible. The large cross represents the position of the participants gaze (on the centre marker).

The eye tracking equipment recorded an AVI file (25 Hz) for each validation trial, showing the video image and the gaze behaviours indicated by a cross hair (Figure 3). Video files were uploaded into the digtising software SIMI Motion Analysis (version 6.5.309, Germany). The AVI file produced by the eye tracking software had an image resolution of 768 x 576 pixels, whilst the SIMI software changed the AVI resolution to 640 x 480 pixels. Within SIMI Motion Analysis, the markers on each corner of the image were semi-automatically digitised throughout each trial. The semi-automatic tracking parameters used were colour recognition, with a 95% threshold for positive marker acceptance and a 75% threshold for negative marker acceptance. The search parameters used a maximum marker size of 10 pixel radius, with an actual marker size of no greater than 10 pixel radius, which included the illumination area around each LED (Figure 3).

Once digitised the raw horizontal and vertical coordinates for each corner marker were filtered using a second order Butterworth filter with a cut-off frequency of 1.05 Hz. This cut-off was determined by studying the frequency components of all markers in all trials using Fast Fourier Transformation technique and assessing the Power Density Spectrum (performed in MATLAB). With a cut off frequency of 1.05 Hz, 99% of the signal power fell below this level.

The filtered horizontal and vertical coordinates for each corner of the image, in each validation trial, were exported from SIMI into Microsoft Excel. The raw eye tracking data derived from EyeVision (the horizontal and vertical scene data) were also exported into Microsoft Excel from the same starting point.

Data analysis

To begin with both sets of data were converted to similar resolution ratios (768 by 576 pixels). To eliminate the influence of head movement, the position of the cross hair was calculated relative to the position of the markers defining the video image. To do this the new relative horizontal coordinates (*x'*) of the cross hair (*cross*) were calculated using equations 1 to 5;

(1) $$tan\theta_1 = \frac{crossy - TLy}{crossx - TLx}$$

(2) $$tan\theta_2 = \frac{TLy - TRy}{TRx - TLx}$$

(3) $$\theta_3 = \theta_1 + \theta_2$$

(4) $$H = \sqrt{(crossx - TLx)^2 + (crossy - TLy)^2}$$

(5) $$Hcos\theta_3 = x'$$

and the new relative vertical coordinates (*y'*) were calculated using equations 6, 7, 3, 4, and 8 (sequentially).

(6) $$tan\theta_1 = \frac{crossx - TLx}{crossy - TLy}$$

(7) $$tan\theta_2 = \frac{TLx - BLx}{BLy - TLy}$$

(8) $$Hcos\theta_3 = y'$$

Where *TL* was the top left marker, *TR* the top right, *BL* the bottom left and *H* was the vector between the cross hair and the origin (*TL*) of the new local coordinate system (defined by the corners of the video image). New relative horizontal and vertical coordinates for the cross hair (with the movement of the head eliminated) were calculated every frame throughout each trial.

To account for any perspective error, if the participant was not positioned directly in front of the image, once the new gaze coordinates had been established relative to the top left of the image; if horizontal or vertical gaze coordinates fell outside of the boundaries of the image they were marked as 'outside image' and excluded from subsequent calculations. To do this, trigonometry was used to calculate the coordinates of the top, bottom, left and right sides of the image using the digitised positions of the corners of the image in each frame. This method does not assume that the sides of the video image are perpendicular.

The total error in these new relative horizontal and vertical gaze data may be influenced by technical error (the accuracy of the automatic tracking of eye position using the eye tracking equipment, and the semi-automatic tracking of the markers defining the corners of the video image), and biological variance (accuracy of the participant to gaze at the appropriate point). The total error was assessed using the validation trials 1 to 6. During these trials the participant was asked to focus on a marker positioned approximately in the centre of the video image whilst moving their head in the various directions. The coefficient of variance of the new relative horizontal and vertical coordinates across all frames in trials 1 to 6 were calculated using equation 9;

(9) %CV = (standard deviation/mean)*100

The mean relative horizontal and vertical coordinate values were also compared to the predicted coordinates of the centre marker measured experimentally during data collection (in pixels).

To assess the contribution of technical error in the semi-automatic tracking of the image corner markers in SIMI Motion Analysis, the coordinates of the corner markers in trials 1 to 6 were used to calculate the length of the top of the video image and the side of the video image in each frame as the participant moved their head. As these absolute lengths should not change, the %CV in the length values for each trial were calculated and compared to the predicted distances between the markers measured experimentally (in pixels). Acceptable levels of %CV were defined as <10% (Atkinson & Nevill, 1998; Stokes, 1985).

1.2 Results.

The predicted horizontal and vertical coordinates of the marker positioned in the centre of the image (measured during data collection), relative to the top left of the image (origin), were 166 by 127 pixels (Figure 3). This compared to mean (standard deviation) recorded relative horizontal and vertical coordinates of 160 (3.7) by 130 (3.8) pixels in trials 1 to 6 (Figure 4). This suggests that the overall difference in the predicted and the recorded coordinates were 3 pixels horizontally and 6 pixels vertically (which equates to 14 mm and 28 mm, respectively for this trial), with a coefficient of variance of 2.3% and 2.9%. The gaze coordinate data for this participant were skewed with their gaze appearing predominantly to the left of the centre marker (closer to the origin) and slightly below the centre marker in all trials (Figure 4).

** INSERT FIGURE 4 NEAR HERE**

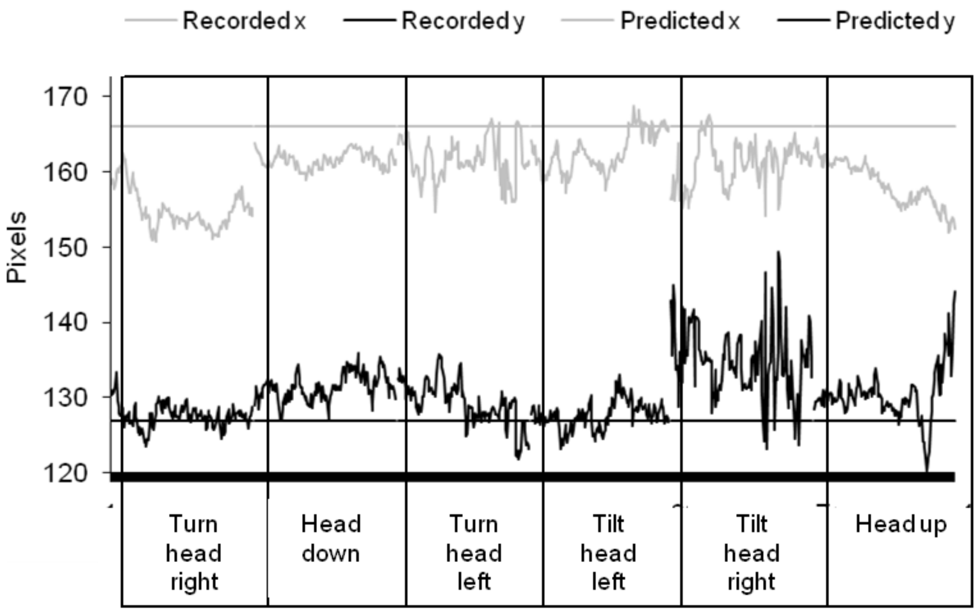


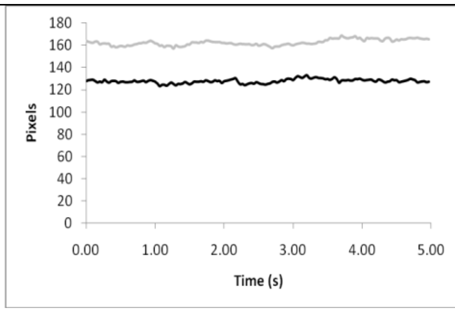
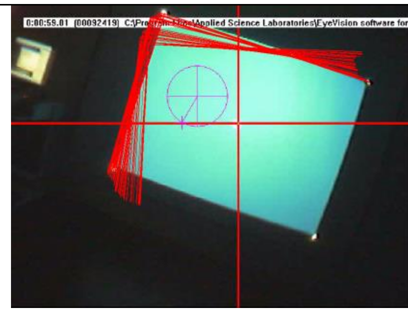
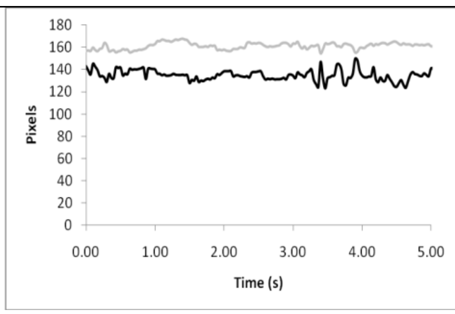
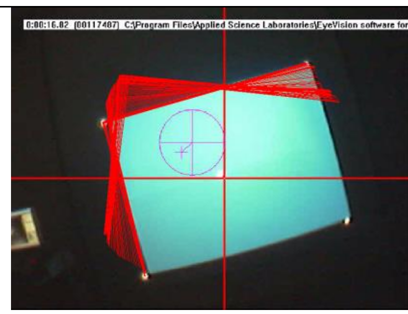
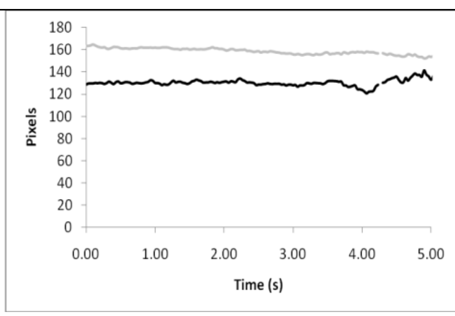
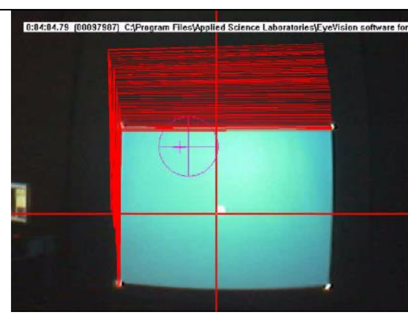
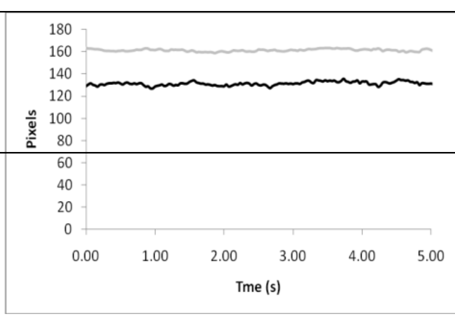
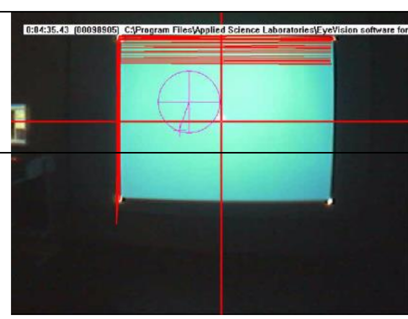
Figure 4. The predicted and the recorded coordinates of the relative horizontal (*x'*) and vertical (*y'*) position of the eye as it focused on the centre marker during six head movement validation trials.

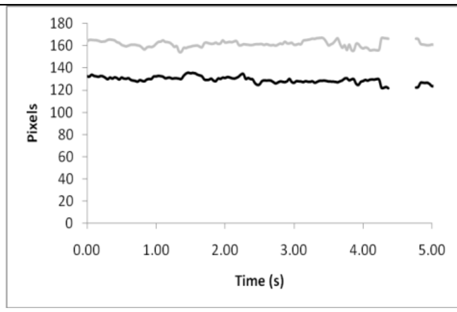
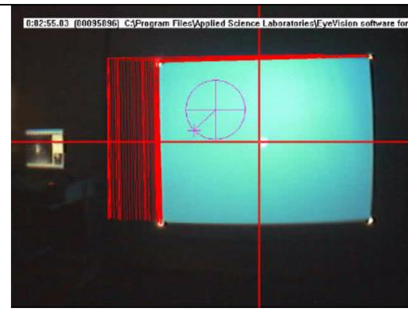
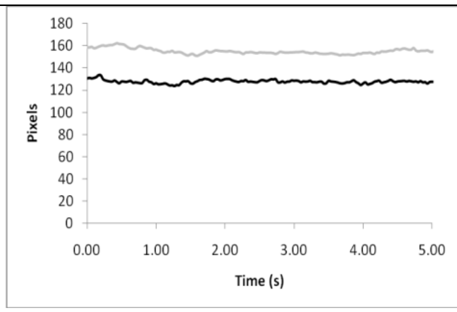
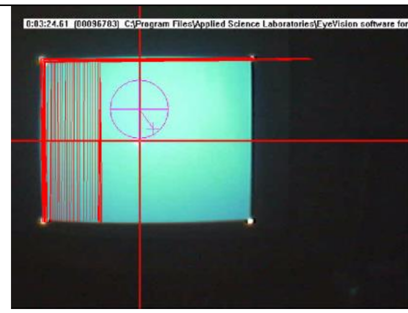
To assess the reliability of these coordinates as the head was moved, the %CV between the predicted and recorded coordinates was assessed in each of the six validation trials and found to be <3.5% across all trials (Table 1).

This variance incorporates technical error and biological variance. To understand the contribution of the semi-automatic tracking in SIMI to the total error reported in table 1, the mean length of the top of the image (top left marker to top right marker) was calculated as 338 (3.8) pixels and the mean length of the side of the image (top left marker to bottom left marker) was 254 (1.7) pixels throughout all trials. The coefficient of variance in these values was 1.1% (top) and 0.7% (side) across all trials (18 mm and 8 mm). Another source of technical error is that produced by the eye tracking equipment, the MobileEye™system has an accuracy of + 1° of visual angle and precision of 0.5° (the diameter of the cursor centre was 2°), however, the reliability of the system is not reported.

** INSERT TABLE 1 NEAR HERE**

Table 1: For validation trials 1 to 6 the new relative horizontal (grey) and vertical (black) gaze coordinates are displayed over the duration of each trial, with the visual output (from SIMI Motion Analysis) and the descriptive statistics for the mean (*x*), standard deviation (SD), and coefficient of variance (%CV).

Trial	Data	Visual output	Descriptive statistics (in pixel units)
Tilt head left			Horizontal *x* = 162 SD = 3 %CV = 1.6 Vertical *x* = 128 SD = 2 %CV = 1.4
Tilt head right			Horizontal *x* = 161 SD = 3 %CV = 1.9 Vertical *x* = 135 SD = 5 %CV = 3.5
Head up			Horizontal *x* =158 SD = 3 %CV = 1.8 Vertical *x* = 130 SD = 3 %CV = 2.6
Head down			Horizontal *x* = 162 SD = 1 %CV = 0.7

			Vertical *x* = 131 SD = 2 %CV = 1.4
Turn head left			Horizontal *x* = 162 SD = 3 %CV = 1.7 Vertical *x* = 129 SD = 3 %CV = 2.2
Turn head right			Horizontal *x* = 155 SD = 3 %CV = 1.7 Vertical *x* = 128 SD = 2 %CV = 1.2

1.3 Discussion.

The first stage of this study successfully reports an algorithm to determine gaze patterns relative to the movement of the head. This was achieved using markers positioned on the corners of the video image that the participant was observing, the position of the markers were digitised to provide known coordinates establishing the horizontal and vertical axis of a new local coordinate system. This local coordinate system represents the position of the head in each sample. Gaze coordinates were then calculated in the local coordinate system to eliminate the influence of head movement. Using this method, the top, bottom, left and right side of the video image were not assumed to be orthogonal, instead these segments were defined individually using the markers position in each frame. This method accounts for perspective error enabling non-standard positioning of participants (not directly in line with the video image), and also accounting for head movement.

To assess the validity of this algorithm, experimental data was collected with the participant fixating on a marker of known coordinates (roughly in the centre of the local coordinate system), whilst tilting and rotating their head. As the marker coordinates did not change whilst the head was moved, the outcome of the algorithm should report similar coordinates to the known marker coordinates. The results showed a 3 pixel and 6 pixel difference between predicted and recorded coordinates, equating to coefficients of variance of 2.3% and 2.9%, horizontally and vertically. These values are within the acceptable criteria set for this study and therefore this algorithm is deemed valid. Also across the six head movement validation trials the reported coefficient of variance was always within the acceptable range (<3.5%) suggesting that the algorithm was also reliable across multiple trials.

Despite the low variance between the predicted and recorded coordinates of the centre marker it should be noted that this variance is a combination of technical error and biological variance. Therefore the output may be very different if assessed using a different participant. It may be that the participant used in part 1 of this study had an exceptionally stable gaze. Future research could utilise this protocol to determine gaze variance across multiple participants.

Across the validation trials it is interesting to note that considerably more variance occurred in the vertical gaze coordinates of the eye as this participant tilted his head to the right. It is speculated that this may be a participant specific anomaly, but further research is warranted using similar head movement trials across multiple participants to determine whether some individuals find it more difficult to fixate with their head in a certain orientation.

To understand the technical error associated with the proposed method, the digitisation of the image corner markers was assessed across all validation trials, reporting a coefficient of variance of 1.1% horizontally and 0.7% vertically. These values are considered acceptable and therefore the semi-automatic digitisation process is deemed valid for this application.

In conclusion, part 1 of this study has established a valid and reliable method to determine gaze coordinates relative to the movement of the head, whilst reducing the need for participants to be positioned in direct alignment with the field of view. Part 1 improves the ecological validity of eye tracking data collection by allowing head movement and non-standard observer placement (as necessary in many occupational environments) The non-standard positioning of individuals is crucial for subsequent eye tracking data collection with multiple observers as they cannot all be simultaneously positioned directly in line with the field of view.

Part 2: Determining the Frequency and Duration of Fixations within Pre-Determined Areas of the Field of View

2.1 Method.

A fixation was defined as a cluster of consecutive gaze points of 100 ms or greater (four frames or more at a sampling frequency of 25 Hz) within a visual angle with a diameter of 1° (Boraston, Corden, Miles, Skuse, & Blakemore, 2008). The visual angle was calculated in the horizontal dimension as the differences in the size and resolution of the image horizontally compared to vertically mean that similar outputs apply for this calculation if undertaken in the horizontal or the vertical dimension. Therefore, to calculate the number of pixels within this visual angle; the gaze point was assumed to be positioned in the centre of the visual angle, the distance from the participant's eye to the image equated to the adjacent side of a right angle triangle, the tangential angle of this triangle was then half of the visual angle (eg. 0.5°). Using these two values, trigonometry then determined the diameter of the fixation window (in pixels). For the validation trials described in part 1 this gave a fixation window of 9 pixels in diameter, which equated to 43 mm for these experimental conditions. Using this fixation window diameter, the new relative coordinates for the gaze point must fall within this window for 100 ms or longer to be classified as a fixation. To assess this, the resultant distance (pixels) that the gaze point moved between frames was calculated, if this displacement value fell within the fixation window diameter for 100 ms or longer a fixation was identified.

To aid interpretation of the fixation data, a grid system consisting of six boxes was then applied to the video image, dividing it into equal thirds horizontally (the top third, the middle third and the bottom third) and into halves vertically (left and right). The number of fixations and the duration of each fixation within each pre-determined area of the field of view were calculated, as well as the total number of fixations and the overall mean duration of a fixation within each trial. Gaze points were discarded if they fell outside of these boxes, if coordinate data were lost due to blinking, or where the fixation was less than 100 ms.

2.2 Results.

For the final validation trial (trial 7), the participant fixated consecutively in three areas of the video image; bottom left, followed by top left and then top right (Figure 5).

INSERT FIGURE 5 NEAR HERE

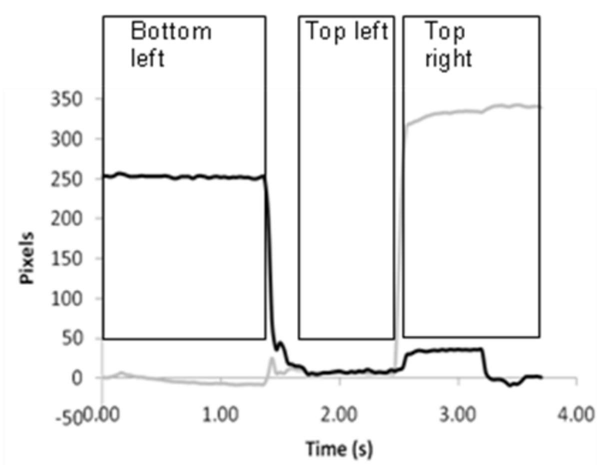


Figure 5. The new relative horizontal (grey) and vertical (black) gaze coordinates as the particpant fixated on three corners of the image.

The output from the algorithm automatically identified four fixations, the first in the bottom left box (0 s to 1.37 s, 1.37 s duration), the second in the top left box (1.57 s to 2.46 s, 0.9 s duration), the third in the top right box (2.56 s to 3.2 s, 0.63 s duration) and the final fixation also in the top right box (3.23 s to 3.7 s, 0.47 s duration). The algorithm output in terms of the frequency, location and duration of fixations was similar to the predicted output (based on the instructions given to the participant), the output of visual observation of the video data and the output from manual calculations.

2.3 Discussion.

Part two of this study, using the relative gaze coordinates established in part one, determined a second algorithm to automatically identify fixations, documenting their frequency, duration and position in pre-determined areas of the field of view. This type of analysis reduces the labour intensive and subjective manual categorisation of gaze patterns that may be undertaken currently. The results of this study reported no difference in the output of the algorithm compared to subjective assessment of fixations from the video image and manual calculation. Part two has developed a valid process for assessing the frequency, duration and positional categorisation of fixations using relative gaze coordinates. This process can then be used to assess these variables for multiple participants simultaneously observing the same field of view.

Part 3: Identifying Simultaneous Gaze Patterns and Recurrent Fixations from Two Observers

3.1 Method.

For part three of this study, test data from 12 participants (in six dyads) was collected. The participants were all male, with no visual impairments. For data collection each dyad sat side by side at a distance of 2.2 m to the video image. Again LED markers were positioned on each corner of the image (dimensions: 1.41 m by 1.1 m). Eye movement and image marker coordinates were recorded using similar methods during one trial of between 4 s and 8 s. During trials lifeguards were instructed to watch a beach scene as if they were at the beach lifeguarding. They were then told that one or more of the swimmers may or may not disappear and if they saw this to verbally state ‘person drowned’ and point a laser pen to the location where the person was..... Manual coding of the data showed that on occasion both participants in the dyad gazed in similar regions of the image, this suggests that some recurrent fixations were present in each data set. For the purpose of this study a recurrent fixation is defined as a period where both participants gaze was within 1^{o} visual angle for ≥100 ms and where these fixations occurred at the same point in time and in the same location (within 1^{o} of visual angle), these limits are based on recurrence analysis (Richardson & Dale, 2005). To quantify the qualitative observations of recurrent fixations, the image corner markers were digitised using the methods described above, filtered using a second order Butterworth filter with a cut-off frequency of 1.05 Hz and exported to Excel. As before, the corresponding raw eye tracking data (scene x and y) were also exported to Excel. All participants’ eye tracking data were then processed to calculate the relative horizontal and vertical coordinates using the procedures described above. Following this, a hierarchical model was implemented (Figure 6).

INSERT FIGURE 6 NEAR HERE

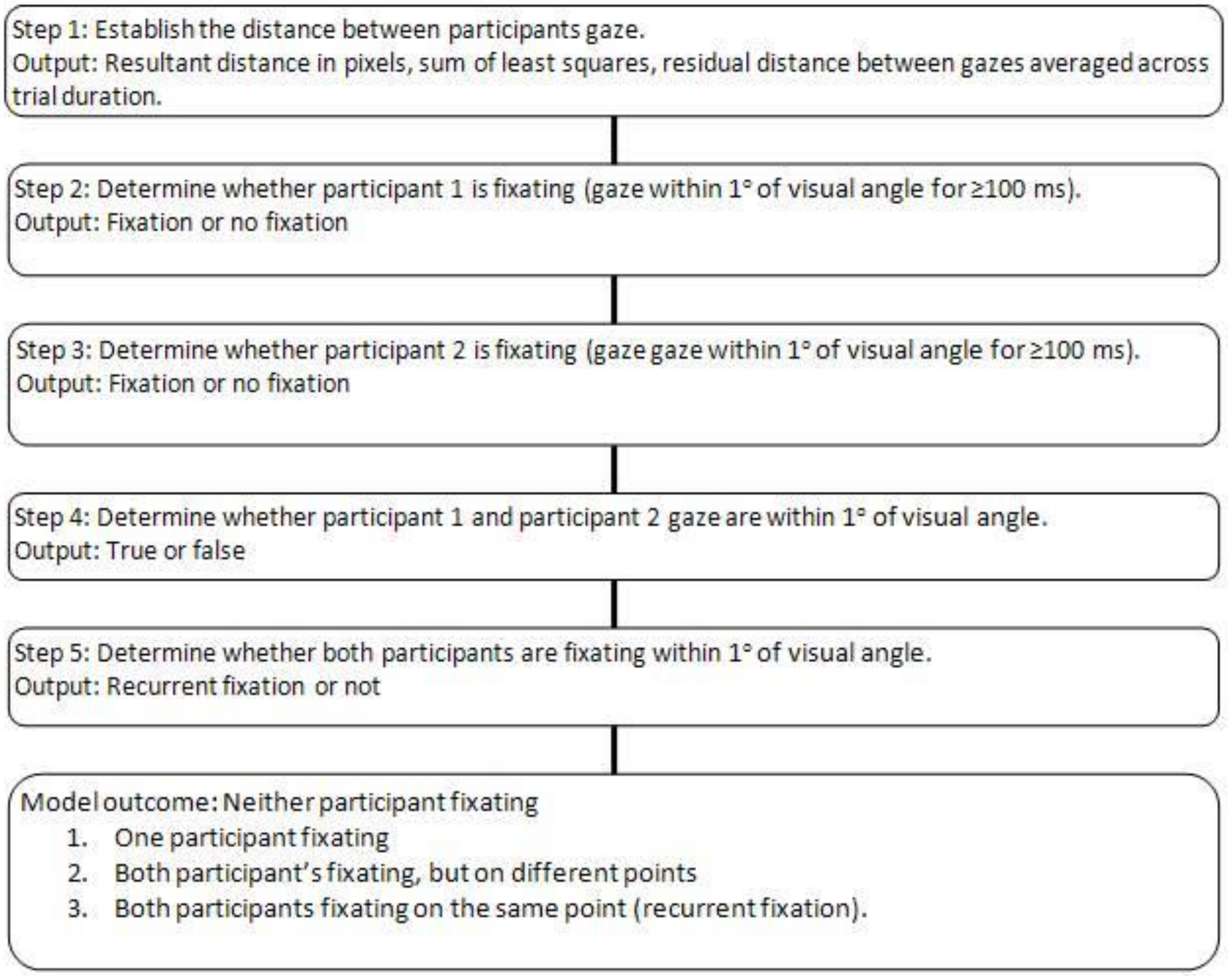


Figure 6. Hierarchical model showing the data analysis process for determining recurrent fixations from two participants observing the same image.

3.2 Results.

Visual inspection of the video clips showing the movement of the cross hair for each dyad identified one possible recurrent fixation within five of the six trials (Table 2). The output of the algorithm and manual calculations identified some additional recurrent fixations not identified through manual coding. The frequency and duration of individual and recurrent fixations calculated manually and using the algorithm were the same. Despite receiving the same instructions, the time spent fixating varied within a dyad by as much as 44% of the trial (participants 5 and 6) and ranged from 19% to 90% across different participants (Figure 7). Although on average, individual participants spent just over half of the trial fixating (51%), only 7% of the trials were spent fixating within the same visual angle. For participants 7 and 8, of the time that each spent fixating, 41% and 30% of this fixation time was spent fixating within the same visual angle. This is supported by the residual value (demonstrating the distance between participant's gazes across the trial), which was lowest in this dyad (32 pixels), showing the closest gaze patterns of the dyads tested in this study (155 mm). Conversely, participants 3 and 4 only spent 2% and 5% of their fixation time fixating within the same visual angle and demonstrated the greatest residual (108 pixels) or distance between gazes (521 mm).

Table 2: Individual and combined gaze parameters from six dyads (12 participants) during simultaneous observation trials of 4 s to 8 s in duration.

Participant	Number of fixations		Total duration of fixations in seconds (% of trial)		Number of recurrent fixations			Total duration of recurrent fixations in seconds (% of trial)		Mean residual between participants gaze in pixels (mm)
	Manual calculation	Algorithm	Manual calculation	Algorithm	Visual inspection of video data	Manual calculation	Algorithm	Manual calculation	Algorithm	
1	3	3	2.52 s (63%)	2.52 s (63%)	1	2	2	0.27 s (7%)	0.27 s (7%)	68 pixels (317 mm)
2	3	3	0.87 s (22%)	0.87 s (22%)						
3	15	15	2.76 s (46%)	2.76 s (46%)	0	1	1	0.08 s (1%)	0.08 s (1%)	108 pixels (521 mm)
4	4	4	1.16 s (19%)	1.16 s (19%)						
5	4	4	2.92 s (74%)	2.92 s (74%)	1	1	1	0.4 s (10%)	0.4 s (10%)	70 pixels (334 mm)
6	5	5	1.16 s (30%)	1.16 s (30%)						
7	2	2	1.88 s (37%)	1.88 s (37%)	1	1	1	0.76 s (15%)	0.76 s (15%)	32 pixels (155 mm)
8	8	8	2.52 s (50%)	2.52 s (50%)						
9	5	5	2.16 s (43%)	2.16 s (43%)	1	2	2	0.28 s (6%)	0.28 s (6%)	79 pixels (383 mm)

10	7	7	3.48 s (69%)	3.48 s (69%)						
11	10	10	5.44 s (90%)	5.44 s (90%)	1	1	1	0.24 s (4%)	0.24 s (4%)	90 pixels (392 mm)
12	10	10	3.96 s (66%)	3.96 s (66%)						
Mean	6.3	6.3	2.57 s (51%)	2.57 s (51%)	0.83	1.3	1.3	0.33 s (7%)	0.33 s (7%)	75 pixels (350 mm)

INSERT FIGURE 7 NEAR HERE

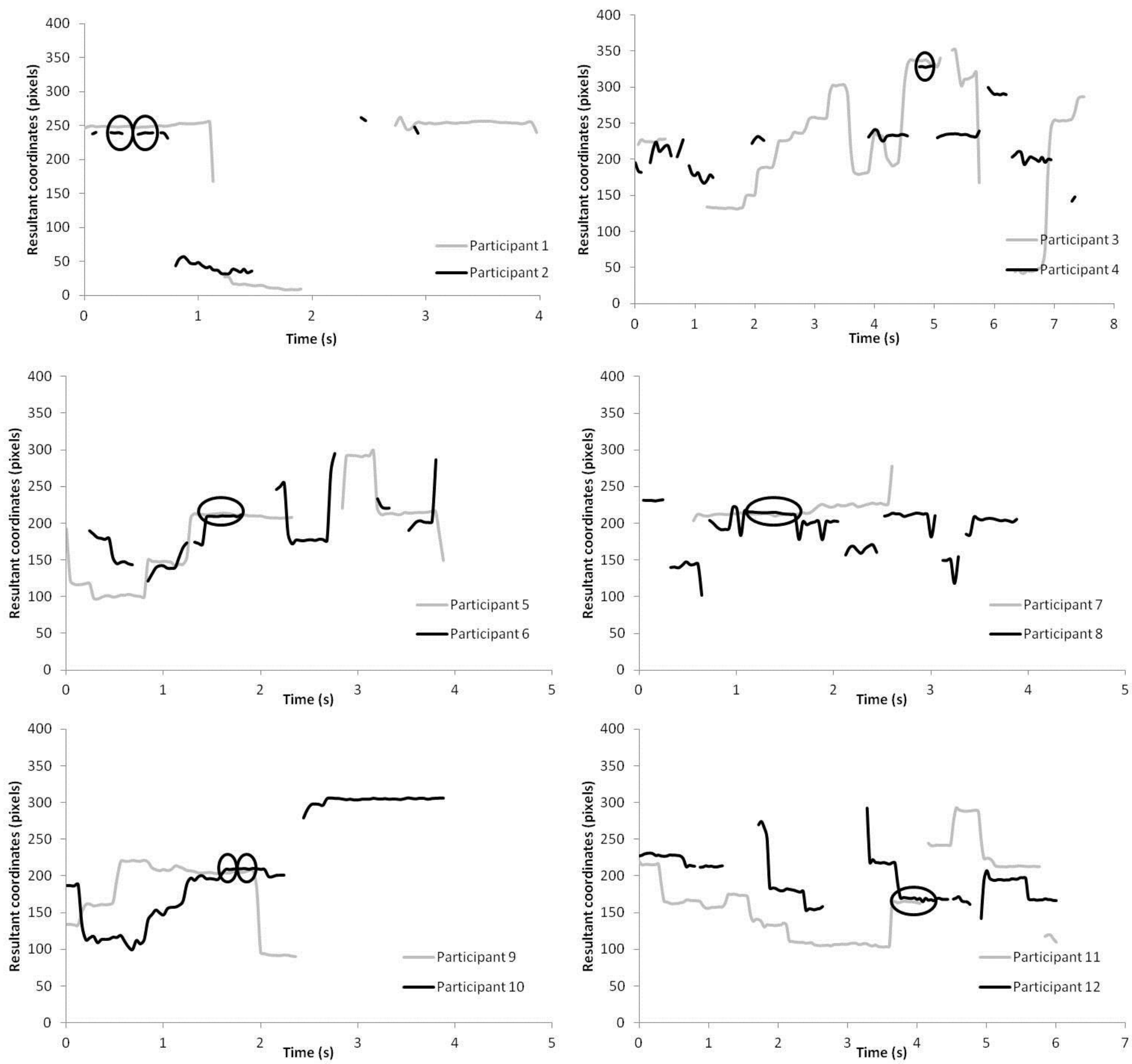


Figure 7. Relative resultant simultaneous gaze patterns and recurrent fixations from six dyads (12 participants) during simultaneous observation trials of 4 s to 8 s in duration (recurrent fixations are circled).

3.3 Discussion.

The establishment of a method to assess fixation frequency, duration and location whilst accounting for head movement and non-standard positioning, enabled the analysis of gaze patterns from multiple participants whilst viewing a dynamic scene. The novel hierarchical model then processes these gaze patterns to determine recurrent fixations from multiple observers. The output of the model identified the same frequency, duration and location of recurrent fixations from multiple dyads compared to manual calculations, whilst offering an improvement on manual coding of the cross hair which was not able to detect all recurrent fixations.

The output of the model was able to discriminate between dyads that followed a more similar gaze pattern during the trial (more recurrent fixations and lower residual between gaze coordinates) compared to those that fixated on different parts of the field of view. This demonstrates an important application of the model, either in training participants to undertake similar or repeatable observation procedures or in assessing differences in gaze patterns within professions where multiple observers are required to survey different parts of the same field of view.

Limited literature has documented exact methods by which simultaneous gaze patterns have been analysed despite the practical importance of understanding interactions between individuals' gaze patterns for occupational purposes. The exception to this is the work of Richardson and Dale (2005) and Richardson et al. (2007b), however, given that participants were watching a small screen during their study, it can be assumed that limited head movement took place. The results from part 3 of this study suggest that this methodological framework has produced a valid and reliable process for the assessment of the frequency, duration and location of recurrent fixations from two participants observing the same field of view simultaneously.

Conclusion

The primary aim of the study was to develop a methodological framework that assessed simultaneous gaze behaviours and recurrent fixations in pre-determined areas of interest, whilst accounting for head movement and non-standard observer positioning. This was achieved using a three stage process which involved ; developing a process to determine gaze coordinates independent to the movement of the head, establishing the frequency and duration of fixations within pre-determined areas of the field of view and identifying simultaneous gaze patterns and recurrent fixations from two observers, simultaneously observing the same display.

There are a number of key strengths of this methodological framework. Firstly, the framework enables the calculation of visual fixations and recurrent fixations and given the importance of visual fixations in understanding attention processes, this will enable conclusions to be drawn in relation to differences and similarities in atttentional processes when working alone and together. Secondly, the framework allows for flexibility in fixation parameters which is important based on the use of different eye tracking systems (e.g., with different capture rates) and different study designs (e.g., long vs. short duration tasks). Thirdly, the framework enables identification of AOIs which adjust relative to head movement, this will reduce the time needed for analysis of AOIs allowing researchers to design longer duration tasks, improving the ecological validity of such assessments. Finally, the framework enables the collection of data from observers that are not positioned in direct alignment with the field of view. This increases the ecological validity of eye tracking assessment and also enables the collection of simultaneous data from two or more observers, which was not possible previously as both observers could not be in direct alignment with the field of view at the same time.

The development of this methodological framework may encourage more research in the area of visual interpersonal coordination, facilitating greater understanding of how multiple people can work together to optimise their visual performance. Future research can now use this framework to analyse data from occupational groups to help understand and train factors that may influence efficient visual attention when individuals are working together. Such research will enable greater understanding of whether recurrence takes place when viewing particular scenes, the impact that recurrence has on task effectiveness and also the types of physical and verbal communication that can influence recurrence. This has strong implications for the optimisation of any tasks that require observation by more than one person.

References

Atkinson, G., & Nevill, A. M. (1998). 'Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine', *Sports Medicine, 26,* 217-238.

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science, 15*, 415– 419.

Boraston, Z. L., Corden, B., Miles, L. K., Skuse, D. H., & Blakemore, S. (2008). Perception of genuine and posed smiles by individuals with autism. *Journal of Autism and Developmental Disorders, 38*, 574-580.

Brennan, S. E., Hanna, J. E., Zelinsky, G. J., & Savietta, K. J. (2012). Eye gaze cues for coordination in collaborative tasks. *DUET 2012 Workshop: Dual eye tracking in CSCE. 2012 ACM Conference on Computer Supported Cooperative Work*, Seattle, WA.

Bucci, M. P., Nassibi, N., Gerard, C-L., Bui-Quoc, E., & Seassau, M. (2012). Immaturity of the oculomotor saccade and vergence interaction in dyslexic children: Evidence from a reading and visual search study. *PLoS ONE*, 7(3), 1-8.

Clark, A. T. & Gergle, D. (2012). Know what I'm talking about? Dual eye-tracking in multimodal reference resolution. *DUET 2012 Workshop: Dual eye tracking in CSCE. 2012 ACM Conference on Computer Supported Cooperative Work*, Seattle, WA.

Clark, H.H., & Krych, M.A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language, 50*, 62–81.

Fuchs, A. F. (1971). The saccadic system. In P. Bach-y-Rita, C. C. Collins, & J. E. Hyde (Eds.), *The control of eye movements* (pp. 343-362). New York: AcademicPress.

Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 29–53). Amsterdam; Elsevier.

Just, M. A., & Carpenter, P. A. (1976). The role of eye-fixation research in cognitive psychology. *Behaviour Research Methods & Instrumentation, 8*, 139-143.

Kanonidou, E., Proudlock, F.A., & Gottlob, I., (2010). Reading strategies in mild to moderate strabismic amblyopia: An eye movement investigation. *Investigative Ophthalmology & Visual Science, 51*, 3502-3508.

Kuhn, G., & Findlay, J. M. (2010). Misdirection, attention and awareness: Inattentional blindness reveals temporal relationship between eye movements and visual awareness. *The Quarterly Journal of Experimental Psychology, 63*, 136–146.

Land, M., Mennie, N., & Rusted, J. (1999). Eye movements and the roles of vision in activities of daily living: making a cup of tea. *Perception, 28*, 1311-1328.

Litchfield, D., Ball, L. J., Donovan, T., Manning, D. J., & Crawford, T. (2008). Learning from others: Effects of viewing another person's eye movements while searching for chest nodules. In B. Sahiner & D. J. Manning (Eds.), *Proceedings of spie medical imaging 2008: Image perception, observer performance, and technology assessment* (Vol. 9). San Diego, CA, USA : SPIE.

Litchfield, D., Ball, L. J., Donovan, T., Manning, D. J., & Crawford, T. (2010). Viewing another person's eye movements improves identification of pulmonary nodules in chest x-ray inspection. *Journal of Experimental Psychology: Applied, 16*, 251–262.

Liversedge, S. P., Paterson, K. B., & Pickering, M. J. (1998). Eye movements and measures of reading time. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 55–75). Amsterdam; Elsevier.

Mackworth, N.H., & Morandi, A.J. (1967). The gaze selects informative details within pictures. *Perception & Psychophysics, 2*, 547–52.

Marsh, K. L., Richardson, M. J., Baron, R., & Schmidt, R.C. (2006). Contrasting approaches to perceiving and acting with others, *Ecological Psychology, 18*(1), 1-38.

Mason, L., Pluchino, P., Tornatora, M.C., & Ariasi, N. (2013). An eye-tracking study of learning from science text with concrete and abstract illustrations. *The Journal of Experimental Education, 81*, 356-384.

Page, J., Bates, V., Long, G., Dawes, P., & Tipton, M. (2011). Beach lifeguards: Visual search patterns, detection rates and the influence of experience. *Ophthalmic and Physiological Optics, 31*, 216 –224.

Pelz, J. B., & Canosa, R. L. (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research*, *41*, 3587-3596.

Reichle, E.D., Reineberg, A.E., & Schooler, J.W. (2010). Eye movements during mindless reading. *Psychological Science*, *21*, 1300-1310.

Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science, 29(6),*1045–1060.

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007b). The art of conversation is coordination: Common ground and the coupling of eye movements during dialogue. *Psychological Science, 18*, 407–413.

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R. L., & Schmidt, R.C. (2007a).

Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science, 26*, 867–891.

Ryan, W. J., Duchowski, A. T., Vincent, E. A. & Battisto, D. (2010). Match-moving for area-based analysis of eye movements in natural tasks. *ETRA 2010 Proceedings of the 2010 Symposium on Eye-Tracking Research and Applications*, Austin, Texas.

Schmidt, R. C., & O'Brien, B. (1997). Evaluating the dynamics of unintended interpersonal coordination. *Ecological Psychology, 9,* 189–206.

Stokes, M. (1985) Reliability and repeatability of methods for measuring muscle in physiotherapy. *Physiotherapy Theory and Practice. 1*, 71-6.

van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Pass, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior, 24,* 785-791.

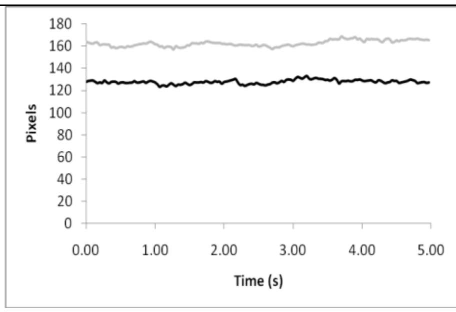
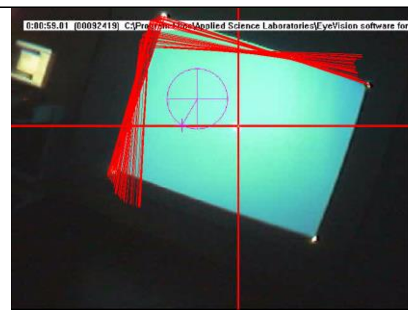
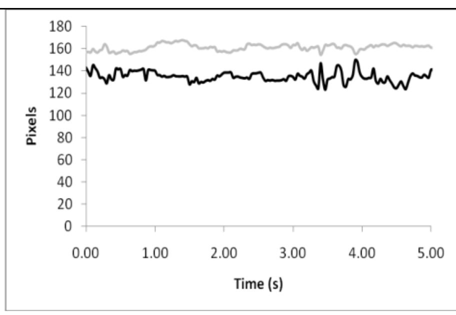
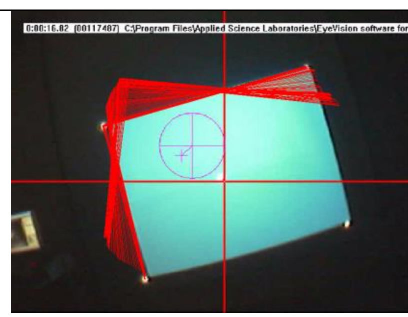
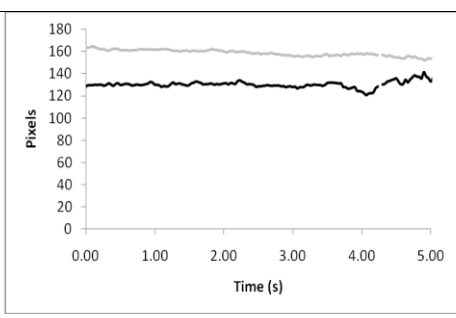
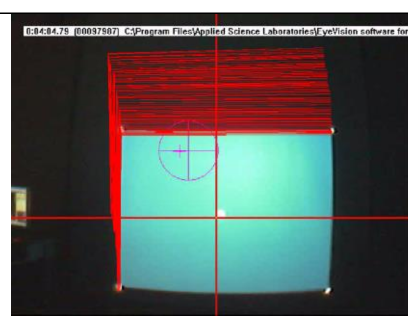
Velichkovsky, B. M. (1995). Communicating attention: Gaze position transfer in cooperative problem solving. *Pragmatics & Cognition, 3,* 199-224.

Vernon, M. D. (1931). *The experimental study of reading*. London; Cambridge University Press.

Wood, G., Knapp, K. M., Rock, B., Cousens, C., Roobottom., C. & Wilson, M. (2013). Visual expertise in detecting and diagnosing skeletal fractures**.** *Skeletal Radiology, 42*, 165-172.

Table

Table 1: For validation trials 1 to 6 the new relative horizontal (grey) and vertical (black) gaze coordinates are displayed over the duration of each trial, with the visual output (from SIMI Motion Analysis) and the descriptive statistics for the mean (*x*), standard deviation (SD), and coefficient of variance (%CV).

Trial	Data	Visual output	Descriptive statistics (in pixel units)
Tilt head left			Horizontal *x* = 162 SD = 3 %CV = 1.6 Vertical *x* = 128 SD = 2 %CV = 1.4
Tilt head right			Horizontal *x* = 161 SD = 3 %CV = 1.9 Vertical *x* = 135 SD = 5 %CV = 3.5
Head up			Horizontal *x* =158 SD = 3 %CV = 1.8 Vertical *x* = 130 SD = 3 %CV = 2.6

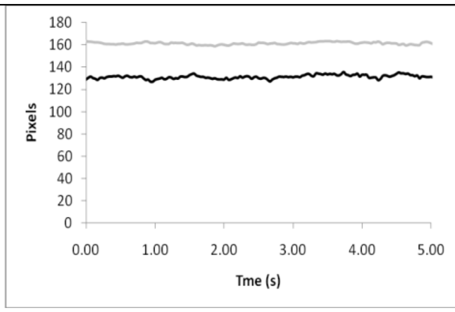
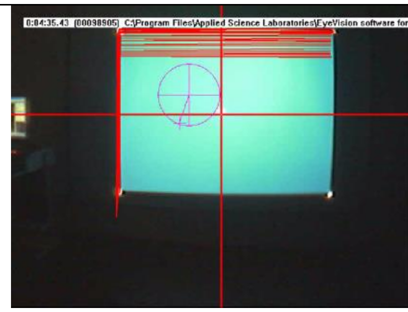
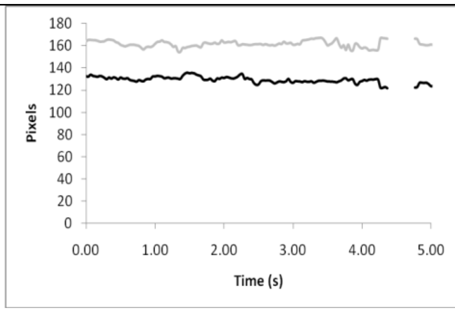
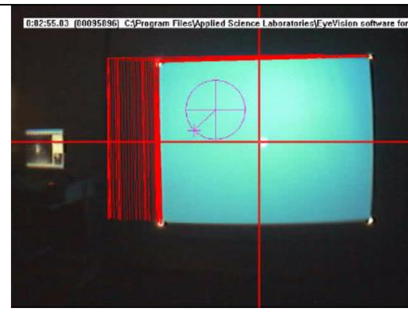
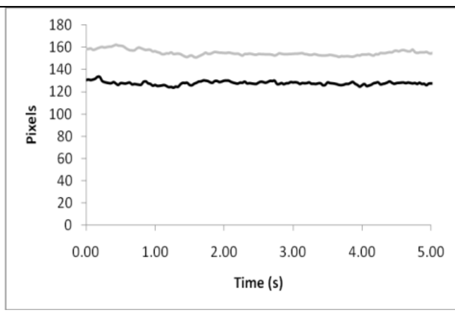
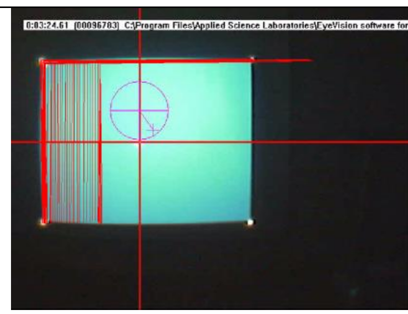
Head down			Horizontal x = 162 SD = 1 %CV = 0.7 Vertical x = 131 SD = 2 %CV = 1.4
Turn head left			Horizontal x = 162 SD = 3 %CV = 1.7 Vertical x = 129 SD = 3 %CV = 2.2
Turn head right			Horizontal x = 155 SD = 3 %CV = 1.7 Vertical x = 128 SD = 2 %CV = 1.2

Table 2: Individual and combined gaze parameters from six dyads (12 participants) during simultaneous observation trials of 4 s to 8 s in duration.

Participant	Number of fixations		Total duration of fixations in seconds (% of trial)		Number of recurrent fixations			Total duration of recurrent fixations in seconds (% of trial)		Mean residual between participants gaze in pixels (mm)
	Manual calculation	Algorithm	Manual calculation	Algorithm	Visual inspection of video data	Manual calculation	Algorithm	Manual calculation	Algorithm	

1	3	3	2.52 s (63%)	2.52 s (63%)	1	2	2	0.27 s (7%)	0.27 s (7%)	68 pixels (317 mm)
2	3	3	0.87 s (22%)	0.87 s (22%)						
3	15	15	2.76 s (46%)	2.76 s (46%)	0	1	1	0.08 s (1%)	0.08 s (1%)	108 pixels (521 mm)
4	4	4	1.16 s (19%)	1.16 s (19%)						
5	4	4	2.92 s (74%)	2.92 s (74%)	1	1	1	0.4 s (10%)	0.4 s (10%)	70 pixels (334 mm)
6	5	5	1.16 s (30%)	1.16 s (30%)						
7	2	2	1.88 s (37%)	1.88 s (37%)	1	1	1	0.76 s (15%)	0.76 s (15%)	32 pixels (155 mm)
8	8	8	2.52 s (50%)	2.52 s (50%)						
9	5	5	2.16 s (43%)	2.16 s (43%)	1	2	2	0.28 s (6%)	0.28 s (6%)	79 pixels (383 mm)
10	7	7	3.48 s (69%)	3.48 s (69%)						
11	10	10	5.44 s (90%)	5.44 s (90%)	1	1	1	0.24 s (4%)	0.24 s (4%)	90 pixels (392 mm)
12	10	10	3.96 s (66%)	3.96 s (66%)						
Mean	6.3	6.3	2.57 s (51%)	2.57 s (51%)	0.83	1.3	1.3	0.33 s (7%)	0.33 s (7%)	75 pixels (350 mm)

Figure Captions

Figure 1. a) Experimental set up and b) central head position.

Figure 2. Head movements for validation trials where a) head tilt left, b) head tilt right, c) up, d) down, e) head turn left and f) head turn right.

Figure 3. Example image from trial 1 (head tilt left). The grey screen represents the video image. The LED markers on each corner and in the centre of the screen are visible. The large cross represents the position of the participants gaze.

Figure 4. The predicted and the recorded coordinates of the relative horizontal (x') and vertical (y') position of the eye as it focused on the centre marker during six head movement validation trials.

Figure 5. The new relative horizontal (grey) and vertical (black) gaze coordinates as the particpant fixated on three corners of the image.

Figure 6. Hierarchical model showing the data analysis process for determining recurrent fixations from two participants observing the same image.

Figure 7. Relative resultant gaze coordinates for participants 1 and 2 during a validation trial (recurrent fixations are circled).