A methodological framework for capturing relative eyetracking coordinate data to determine gaze patterns and fixations from two or more observers

Joanna C. Scurr & Jenny Page & Heather Lunt

Abstract

Few studies have published methodologies that can be used to analyze simultaneous gaze behaviors and recurrent fixations while multiple observers are viewing dynamic scenes and moving their heads. In this study, we aimed to develop a methodological framework to assess simultaneous gaze behaviors and recurrent fixations in predetermined areas of interest, while accounting for head movement and nonstandard observer positioning. Gaze coordinates were recorded during six trials in which a single participant focused on the center of a video image and moved his head in six degrees of freedom. Markers were positioned at the image corners. Eyetracking equipment recorded the video image and gaze behaviors (crosshair), which were then uploaded to SIMI Motion Analysis software. The corner markers were digitized in order to determine image position as the head moved, and were used to calculate new gaze coordinates relative to this head movement. Calculations accounted for the perspective error due to nonstandard participant positioning. Across all trials, the error between the measured and calculated coordinates was acceptable (<3.5 %CV). The frequencies and durations of fixations (≥100 ms within 1 ° of visual angle) within six areas of interest are reported, and they compared well to manual calculations. This methodology was then assessed using participant dyads (N=5), each simultaneously observing the same images. Recurrent fixations were determined using a hierarchical model, and these also compared well to manual analysis. This article presents a valid and reliable methodological framework for determining fixation frequency, duration, and location from multiple observers, while accounting for head movement and nonstandard positioning. This framework facilitates the analysis of simultaneous oculomotor variables, improving ecological validity and reducing environmental constraints.

Introduction

A number of occupations require two people to work together to observe and make decisions 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 individuals respond to others' physical movements (Bangerter, 2004; Clark & Krych, 2004; Richardson, Marsh, Isenhower, Goodman, & Schmidt, 2007a, b; Schmidt & O'Brien,

1997) or verbal communication (Richardson & Dale, 2005; Richardson et al. 2007a, b). However, as with the occupations previously mentioned, often visual information (what we see in our environment) rather than physical contact or auditory information 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 behavior 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 found that the time to complete three puzzles was significantly shorter for novices when they were able to speak freely with an expert as well as see the expert's gaze behavior, as compared to when they could only speak freely with the expert. Furthermore, in a second experiment, novices were also quicker to complete the three puzzles when experts were able to view the novices' gaze behavior and speak freely, as compared to just being able to speak freely. Conversely, van Gog et al. 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), as 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 expert's underlying thought processes, and saw the expert's eye movements performed significantly worse than the group who only saw the solution to the task and heard the expert's underlying thought processes. However, little research has investigated visual interpersonal coordination when two individuals are performing a task simultaneously using a visual "common ground" (Richardson & Dale, 2005; Richardson et al. 2007a, b). More research in the area of visual interpersonal coordination may allow a greater understanding of how two or more people can work together to optimize 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 their occupation requires employees to be in nonstandard 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 would be required for analyzing combined gaze behaviors when the observers are able to move their heads. Before considering potential solutions, a number of factors inherent in eye movement 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 eyetracking systems have been used to assess gaze behaviors when performing a range of individual tasks, including reading (e.g., Bucci, Nassibi, Gerard, Bui-Quoc, & Seassau, 2012; Kanonidou, Proudlock, & Gottlob, 2010; Reichle, Reineberg, & Schooler, 2010), everyday tasks (Land, Mennie, & Rusted, 1999; Pelz & Canosa, 2001), lifeguarding (Page, Bates, Long, Dawes, & Tipton, 2011), and medical diagnostics (e.g., Litchfield et al., 2008, 2010; Wood et al., 2013); these systems have included commercially available eyetracking systems such as the Tobii x50 eyetracker (Tobii Technology, Stockholm, Sweden; Litchfield et al., 2008), the EyeLink 1000 eyetracker (SR Research, Ottawa, Ontario, Canada; Reichle et al., 2010), an EyeLink eyetracker (SensoMotoric Instruments GmbH, Berlin, Germany; Kanonidou et al., 2010), and the 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 analyze the eye movements, including MeyeAnalysis [e(ye)BRAIN, www.eye-brain.com; 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 analyze data collected from non-computerbased stimuli (e.g., in situ data collection where head movement is permitted) do not offer the ability to semiautomatically analyze data from two participants observing the same field of view.

Existing data analysis software often calculates 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 number of fixations. The use of visual fixations in eye movements research dates back to the 1930s (Vernon, 1931) and is still used as a key variable (Mason, Pluchino, Tornatora, & Ariasi, 2013). Fixation measurements may be used so frequently because they are thought to indicate the amount and importance of processed information. Indeed, Just and Carpenter (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 frequencies. Given the importance of visual fixations, any analysis of simultaneous gaze behaviors should report this variable.

Despite the importance of visual fixations, the algorithms used within software programs to establish visual fixations are often not readily available. Additionally, the eyetracking literature has produced a variety of definitions for fixations, which include fixating on a target (diameter 0.5 °) for >250 ms (Bucci et al., 2012), gaze with a spatial deviation threshold of 19 pixels for ten 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) categorized as fixations gazes fixated on a target for <80 ms or >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) analyzed fixations based on analysis of one frame using a 500 Hz eyetracker. Given the variation in existing systems, any methodological framework should allow for 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 intraindividual visual angle with a diameter of 1 ° for a period of >100 ms, and these fixations occur at the same time within an interindividual visual angle with a diameter of 1 °; these limits are based on recurrence analysis (Richardson & Dale, 2005). It may also be important when designing a framework to analyze simultaneous participant data, to understand the distance between gaze points when participants are not recurrently fixating, and therefore Euclidean distances between gaze points can also be analysed to understand differences and similarities in smooth-pursuit eyetracking strategies.

Once the algorithms for recurrent fixations are established, there is often a need to state the locations of fixations within the field of view. Some analysis software requires the head to be static and in the same position for all participants, and therefore is able to produce templates for areas of interest (AOI) that can be employed for all participants within the experiment. Other analysis systems that are designed for mobile eyetracking (e.g., Gazetracker) require that AOIs be added post hoc, and although the AOIs can be moved in relation to head movements, this often requires a frame by frame approach. Although manual coding is effective, the time that it requires can be prohibitive for research that involves long-duration tasks. Other systems (e.g., Tobii eyetrackers) use automatic marker detection, which allows the aggregation of multiple participants' eye movements. However, if a system does not have the built-in capacity for marker detection and participants are moving their heads, analyses of multiple participant data are problematic. Ryan et al. (2010) suggested that adding track boxes is a solution to the head movement issue. The track boxes 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 and does not consider multiple observers. Therefore, any methodological framework should permit the addition of AOIs while accounting for head movements and enabling nonstandard observer positioning.

Although some studies have used semiautomatic analyses of eye movement data to measure fixations from individual participants when the head was freely moving (Ryan et al., 2010), and other studies have analyzed simultaneous eye movements while limiting head movement (Brennan, Hanna, Zelinsky, & Savietta, 2012; Clark & Gergle, 2012; Richardson & Dale, 2005; Richardson et al. 2007a, b), no processes have been published that combine these elements while enabling nonstandard observer positioning, featuring automatic detection of AOIs, and offering explicit definitions of individual and recurrent fixations. Therefore, the primary aim of the framework established within this study was to address the problem of analyzing eye movements while viewing dynamic scenes, allowing for head movement and for subsequently combining these data to determine when and where recurrent fixations occur. Here we aimed to develop a methodological framework with which to assess simultaneous gaze behaviors and recurrent fixations in predetermined AOIs, while accounting for head movement and nonstandard 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 predetermined areas of the field of view;
- Part 3: identifying gaze patterns and recurrent fixations from two observers, simultaneously observing the same display from nonstandard positions.

Part 1: Determining gaze coordinates relative to the movement of the head

Method

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

The movement of the participant's right eye was tracked using an Applied Science Laboratories MobileEye (ASL, Bedford, MA). When recording the eyetracking 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 center of the video image (Figs. 1 and 3). These markers were highly contrasted 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 center was noted as 2.5 m, and the distances between the top horizontal and left vertical markers were noted as 1.62 m and 1.23 m.

The eyetracking equipment recorded an AVI file (25 Hz) for each validation trial, showing the video image and the gaze behaviors indicated by a cross hair (Fig. 3). These video files were uploaded into the digitizing software SIMI Motion Analysis (version 6.5.309; SIMI Reality Motion Systems GmbH, Germany). The AVI file produced by the eyetracking software had an image resolution of 768×576 pixels, and the SIMI software changed the AVI resolution to 640×480 pixels.

Within SIMI Motion Analysis, the markers on each corner of the image were semi automatically digitized throughout each trial. The semiautomatic tracking parameters used were color 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 a ten pixel radius, with an actual marker size of no greater than a ten-pixel radius, which included the illumination area around each LED (Fig. 3).

Once digitized, 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 the 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 eyetracking data derived from EyeVision (the horizontal and vertical scene data) were also exported from SIMI into Microsoft Excel.

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 Eqs. 1–5, and the new relative vertical coordinates (y') were calculated using Eqs. 6, 7, 3, 4, and 8 sequentially), where TL was the top left marker, TR the top right, and BL the bottom left, and H was the vector between the crosshair 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 crosshair (with the movement of the head eliminated) were calculated every frame throughout each trial.

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

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

$$\theta_3 = \theta_1 + \theta_2 \tag{3}$$

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

$$Hcos\theta_3 = x' \tag{5}$$

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

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

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

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 the horizontal or vertical gaze coordinates fell outside 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 digitized 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 eyetracking equipment, and the semiautomatic tracking of the markers defining the corners of the video image), and biological variance (the accuracy of the participant when gazing at the appropriate point). The total error was assessed using Validation Trials 1–6. During these trials, the participant was asked to focus on a marker positioned approximately in the center of the video image while moving the head in the various directions. The coefficient of variance of the new relative horizontal and vertical coordinates across all frames in Trials 1–6 were calculated using Eq. 9.

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

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

To assess the contribution of technical error in the semiautomatic tracking of the image corner markers in SIMI Motion Analysis, the coordinates of the corner markers in Trials 1–6 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 his head. Since these absolute lengths should not change, the %CVs 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).

Results

The predicted horizontal and vertical coordinates of the marker positioned in the center of the image (measured during data collection), relative to the top left of the image (origin), were166 × 127 pixels (Fig. 3). This was compared to the mean (with standard deviation) recorded relative horizontal and vertical coordinates of 160 (3.7)×130 (3.8) pixels in Trials 1–6 (Fig. 4). This suggests that the overall differences in the predicted and recorded coordinates were three pixels horizontally and six pixels vertically (which equated to 14 mm and 28 mm, respectively, for this trial), with coefficients of variance of 2.3 % and 2.9 %. The gaze coordinate data for this participant were skewed, with the gaze appearing predominantly to the left of the center marker (closer to the origin) and slightly below the center marker in all trials (Fig. 4).

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 semiautomatic 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 coefficients of variance in these values were 1.1 % (top) and 0.7 % (side) across all trials (i.e., 18 mm and 8 mm). Another source of technical error was that produced by the eyetracking equipment. The MobileEye system has an accuracy of +1 ° of visual angle and a precision of 0.5 ° (the diameter of the cursor center was 2 °); however, the reliability of the system is not reported.

Discussion

In the first stage of this study, we were successfully able to report 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 positions of the markers were digitized, so as to provide known coordinates establishing the horizontal and vertical axis of a new, local coordinate system. This local coordinate system represented 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 sides of the video image were not assumed to be orthogonal; instead, these segments were defined individually using the marker positions in each frame. This method accounts for perspective error, enabling the nonstandard positioning of participants (i.e., not directly in line with the video image), and also accounting for head movement.

To assess the validity of this algorithm, experimental data were collected with the participant fixating a marker of known coordinates (roughly in the center of the local coordinate system), while tilting and rotating his head. Since the marker coordinates did not change while the head was moved, the outcome of the algorithm should report coordinates similar to the known marker coordinates. The results showed three-pixel and six-pixel differences between the predicted and recorded coordinates, equating to coefficients of variance of 2.3 % and 2.9 %, respectively, for horizontal and vertical. These values are within the acceptable criteria set for this study, and therefore this algorithm was deemed valid. Also across the six head movement validation trials, the reported coefficients of variance were 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 center marker, it should be noted that this variance was a combination of technical error and biological variance. Therefore, the output might be very different if it were assessed using a different participant. It might be that the participant used in Part 1 of this study had an exceptionally stable gaze. Future research could utilize 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. We speculate that this may be a participant-specific anomaly, but further research using similar head movement trials across multiple participants will be warranted, 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 digitization of the image corner markers was assessed across all validation trials, reporting coefficients of variance of 1.1 % horizontally and 0.7 % vertically. These values are considered acceptable, and therefore the semi automatic digitization process was deemed valid for this application.

In conclusion, Part 1 of this study established a valid and reliable method to determine gaze coordinates relative to the movement of the head, while reducing the need for participants to be positioned in direct alignment with the field of view. In Part 1, we improved the ecological validity of eyetracking data collection by allowing head movement and nonstandard observer placement (as is necessary in many occupational environments) The nonstandard positioning of individuals will be crucial for subsequent eyetracking data collection with multiple observers, since they could not all simultaneously be positioned directly in line with the field of view.

Part 2: Determining the frequency and duration of fixations within predetermined areas of the field of view

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, are the same as those vertically, so that this calculation would produce similar outputs 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 center of the visual angle, the distance from the participant's eye to the image was equated to the adjacent side of a right-angle triangle, and the tangential angle of this triangle was then half of the visual angle (e.g., 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 nine 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, and if this value fell within the fixation window diameter for 100 ms or longer, a fixation was identified.

To the 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 vertically (the top third, middle third, and bottom third) and into halves horizontally (left and right). The number of fixations and the duration of each fixation within each predetermined 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.

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 (Fig. 5). The output from the algorithm automatically identified four fixations, the first in the bottom left box (0 to 1.37 s, 1.37 s duration), the second in the top left box (1.57 to 2.46 s, 0.9 s duration), the third in the top right box (2.56 to 3.2 s, 0.63 s duration), and the final fixation also in the top right box (3.23 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 (manual coding), and the output from manual calculations.

Discussion

Part 2 of this study, using the relative gaze coordinates established in Part 1, determined a second algorithm to automatically identify fixations, documenting their frequency, duration, and position in predetermined areas of the field of view. This type of analysis reduces the labor-intensive and subjective manual categorization of the gaze patterns that currently may be undertaken. The results of this study reported no difference in the output of the algorithm relative to subjective assessment of fixations from the video image and manual calculation. In Part 2, we developed a valid process for assessing the frequency, duration, and positional categorization 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

Method

For Part 3 of this study, the test data from ten participants (in five dyads) was collected. The participants were male professional lifeguards with no visual impairments. For the data collection, each dyad of participants sat side by side at a distance of 2.2 m from the video image. Again LED markers were positioned at each corner of the image (dimensions: 1.41×1.1 m). Eye movement and image marker coordinates were recorded using similar methods during one trial of between 4 and 8 s. During the 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 might or might not disappear, and that if they saw this they were to verbally state "person drowned" and point a laser pen to the location where the person had been. Manual coding of the data showed that on occasion, both participants in the dyad gazed at 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 was defined as a period during which both participants' gazes were within 1 ° of visual angle for ≥100 ms, so that these fixations occurred at the same point in time and in the same location (within 1 ° of visual angle; Richardson & Dale, 2005). To quantify the qualitative observations of recurrent fixations, the image corner markers were digitized 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 eyetracking data (scene x and y) were also exported to Excel. All participants' eyetracking data were then processed to calculate the relative horizontal and vertical coordinates using the procedures described above. Following this, a hierarchical model was implemented (Fig. 6).

Results

Visual inspection of the video clips showing the movements of the crosshairs for each dyad identified one possible recurrent fixation within four of the five 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 the individual and recurrent fixations calculated manually and using the algorithm were the same. Despite the participants' receiving the same instructions, the times spent fixating varied within a dyad by as much as 44 % of the trial (for Participants 3 and 4) and ranged from 19 % to 90 % across different participants (Fig. 7). Although, on average, individual participants spent just over half of the trial fixating (52 %), only 7 % of the trials were spent fixating within the same visual angle. For Participants 5 and 6, of the time that each spent fixating, 41 % and 30 % of this fixation time, respectively, was spent fixating within the same visual angle. This is supported by the residual value (demonstrating the distance between the participants' 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 1 and 2 only spent 2 % and 5 % of their fixation time fixating within the same visual angle, and they demonstrated the greatest residual (108 pixels) and distance between gazes (521 mm).

Discussion

The establishment of a method to assess fixation frequency, duration, and location while accounting for head movement and nonstandard positioning enabled the analysis of gaze patterns from multiple participants while they viewed a dynamic scene. The novel hierarchical model then processed 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, as compared to manual calculations, while offering an improvement on manual coding of the crosshair, which was not able to detect all recurrent fixations.

The output of the model was able to discriminate between dyads that followed more similar gaze patterns during the trial (more recurrent fixations and lower residual between gaze coordinates) than did those who fixated on different locations within parts of the image. This demonstrates important applications of the model, either to train participants to undertake similar or repeatable observation procedures or to assess differences in gaze patterns within professions in which 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 can be analyzed, despite the practical importance of understanding interactions between individuals' gaze patterns for occupational purposes. The exception is the work of Richardson and Dale (2005) and Richardson et al. (2007a, b); however, given that participants were watching a small screen during these studies, it can be assumed that limited head movement took place. The results from Part 3 of this study suggest that the present 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 this study was to develop a methodological framework that assessed simultaneous gaze behaviors and recurrent fixations in predetermined areas of interest, while accounting for head movement and nonstandard 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 predetermined areas of the field of view, and identifying simultaneous gaze patterns and recurrent fixations from two observers simultaneously observing the same display.

This methodological framework has a number of key strengths. First, 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 attentional processes when people are working either alone or together. Second, the framework allows for flexibility in fixation parameters, which is important for compatibility with different eyetracking systems and different study designs (e.g., long- vs. short-duration tasks). Third, the framework enables identification of AOIs that adjust relative to head movement, reducing the time needed for analysis of such AOIs, allowing researchers to design longer-duration tasks, and improving the ecological validity of such assessments. Finally, the framework enables the collection of data from observers who are not positioned in direct alignment with the field of view. This increases the ecological validity of eyetracking assessment and also enables the collection of simultaneous data from two or more observers, which may not have been possible previously, since 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 optimize 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 people view particular scenes, of the impact that recurrence has on task effectiveness, and also of the types of physical and verbal communication that can influence recurrence. This will have strong implications for the optimization 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. doi:10.1111/j.0956-7976.2004.00694.

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, February). Eye gaze cues for coordination in collaborative tasks. Paper presented at the 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(e33458), 1–8. doi:10.1371/journal.pone.0033458

Clark, A. T., & Gergle, D. (2012, February). Know what I'm talking about? Dual eye-tracking in multimodal reference resolution. Paper presented at the 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.

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, The Netherlands: Elsevier.

Just, M. A., & Carpenter, P. A. (1976). The role of eye-fixation research in cognitive psychology. Behavior Research Methods & Instrumentation, 8, 139–143. doi:10.3758/BF03201761

Kanonidou, E., Proudlock, F. A., & Gottlob, I. (2010). Reading strategies in mild to moderate strabismic amblyopia: An eye movement investigation. Investigative Ophthalmology and Visual Science, 51, 3502–3508. doi:10.1167/iovs.09-4236

Kuhn, G., & Findlay, J. M. (2010). Misdirection, attention and awareness: Inattentional blindness reveals temporal relationship between eye movements and visual awareness. Quarterly Journal of Experimental Psychology, 63, 136–146. doi:10.1080/17470210902846757

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. doi:10.1068/p2935

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, pp. 691–715). San Diego, CA: SPIE. doi:10.1117/12.768812

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, The Netherlands: Elsevier.

Mackworth, N. H., & Morandi, A. J. (1967). The gaze selects informative details within pictures. Perception & Psychophysics, 2, 547–552. Marsh, K. L., Richardson, M. J., Baron, R., & Schmidt, R. C. (2006). Contrasting approaches to perceiving and acting with others. Ecological Psychology, 18, 1–38.

Mason, L., Pluchino, P., Tornatora, M. C., & Ariasi, N. (2013). An eyetracking study of learning from science text with concrete and abstract illustrations. 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, 1045–1060.

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007a). 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. (2007b). 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. Paper presented at ETRA 2010: 2010 Symposium on EyeTracking 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, 1,71–76.

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 and Cognition, 3, 199–224.

Vernon, M. D. (1931). The experimental study of reading. London, UK: 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.

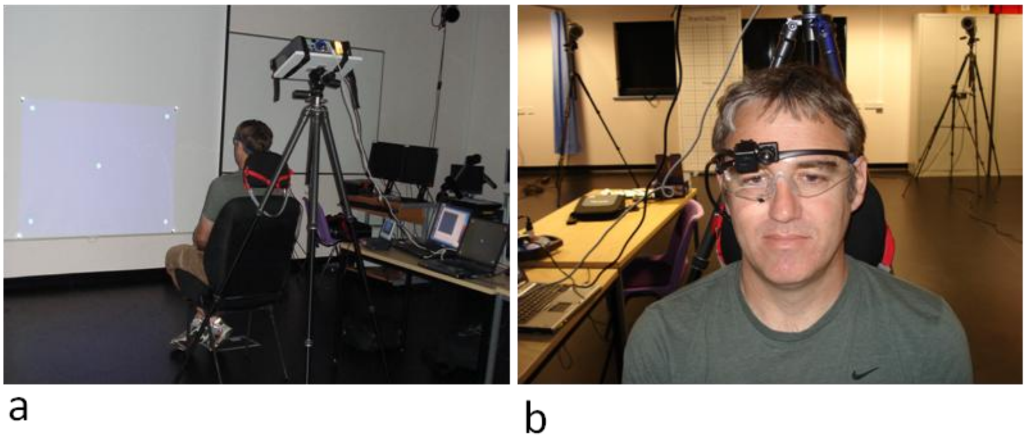


Figure 1. a) Experimental set up and b) central head position.

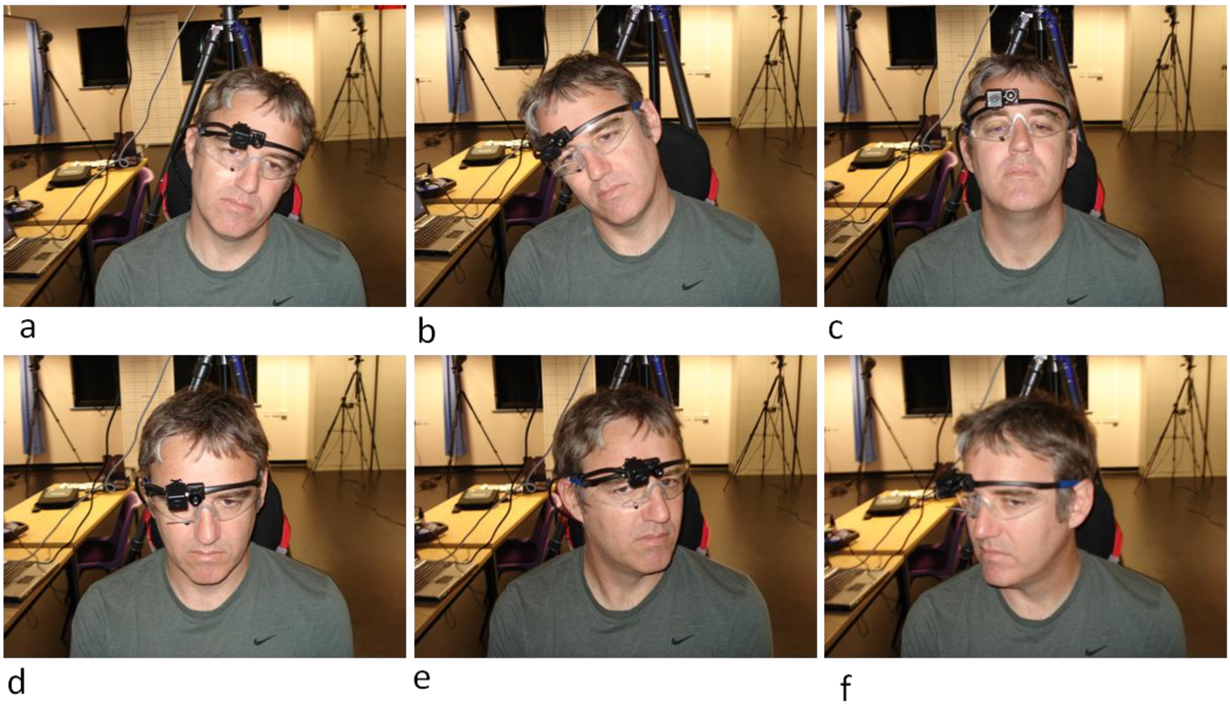


Figure 2. Head movements for validation trials where a) head tilt left, b) head tilt right, c) up, d) down, e) head turn left, and f) head turn right.

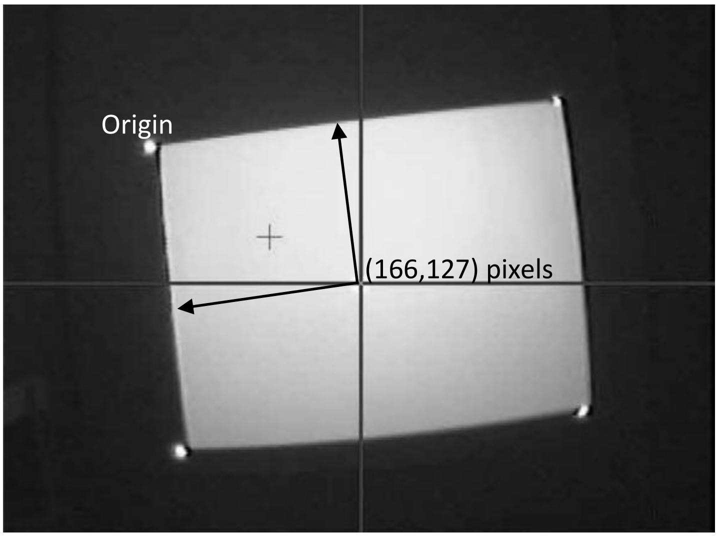


Figure 3. Example image from trial 1 (head tilt left). The grey screen represents the video image. The LED markers on each corner and in the centre of the screen are visible. The large cross represents the position of the participants gaze (on the centre marker).

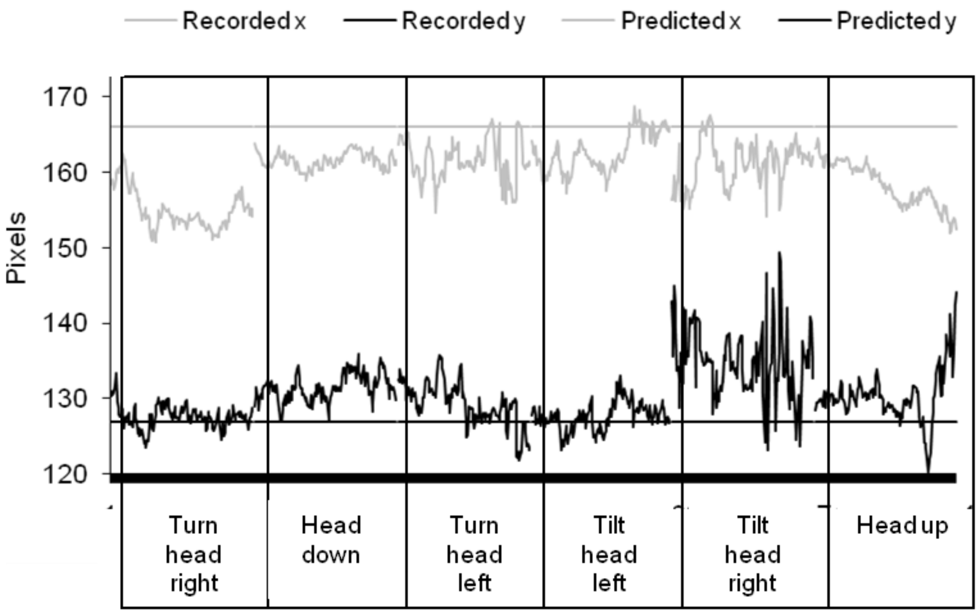


Figure 4. The predicted and the recorded coordinates of the relative horizontal (*x'*) and vertical (*y'*) position of the eye as it focused on the centre marker during six head movement validation trials.

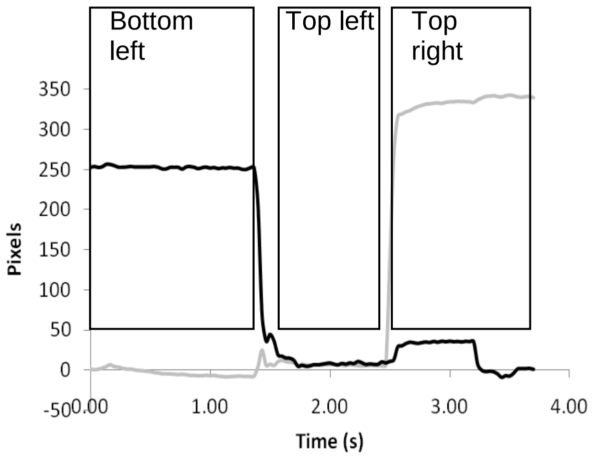


Figure 5. The new relative horizontal (grey) and vertical (black) gaze coordinates as the particpant fixated on three corners of the image.

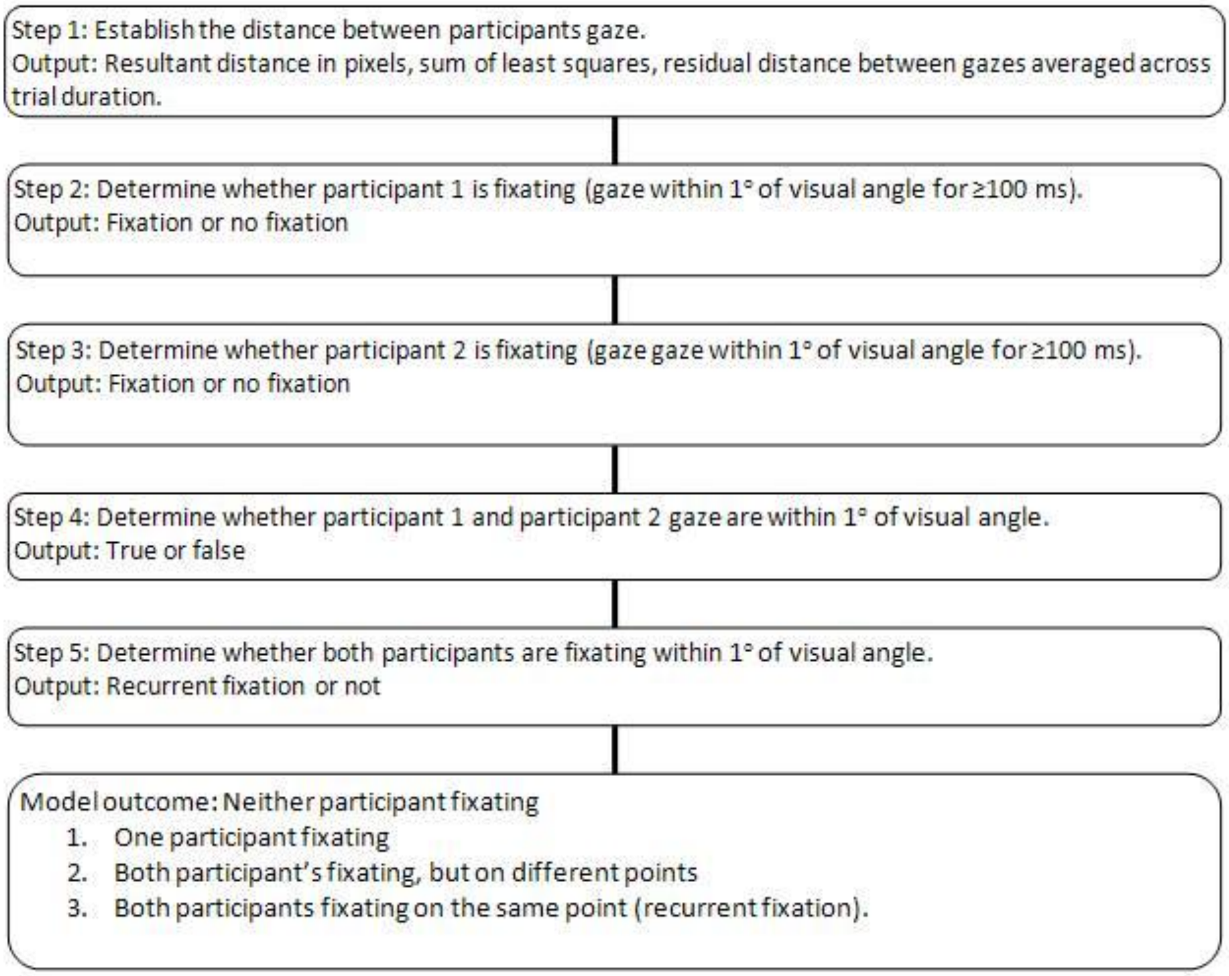


Figure 6. Hierarchical model showing the data analysis process for determining recurrent fixations from two participants observing the same image.

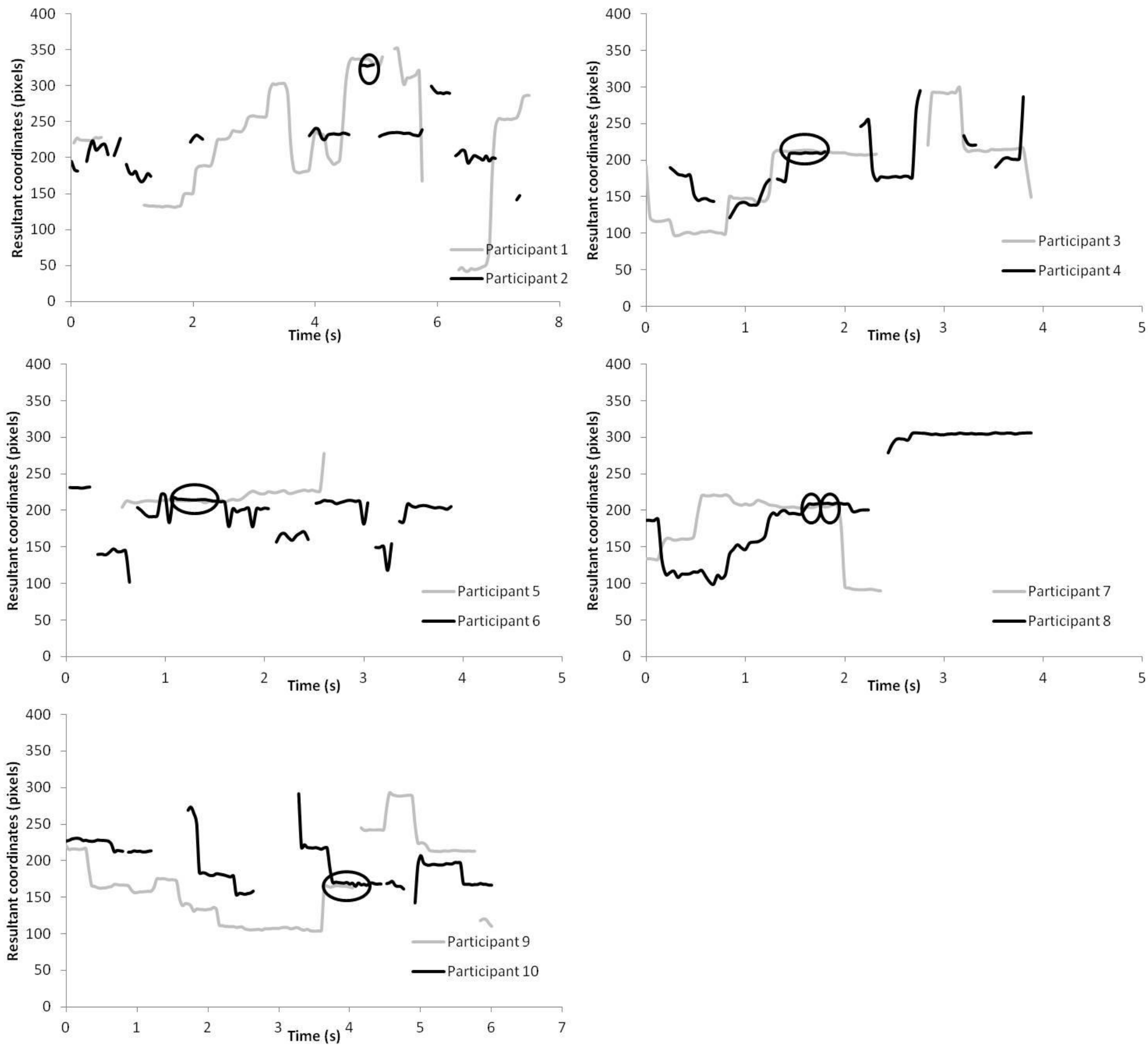
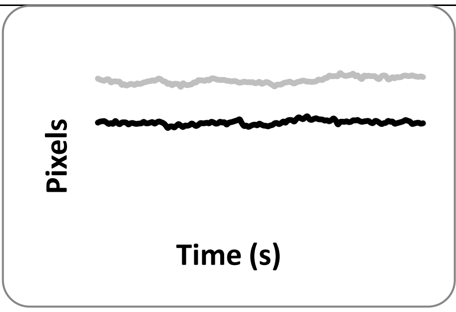
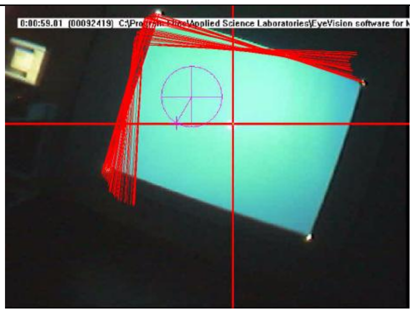
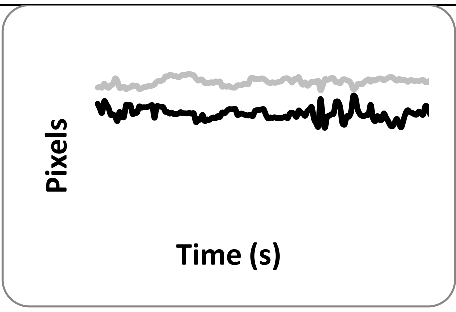
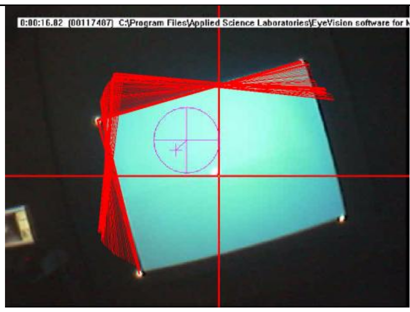
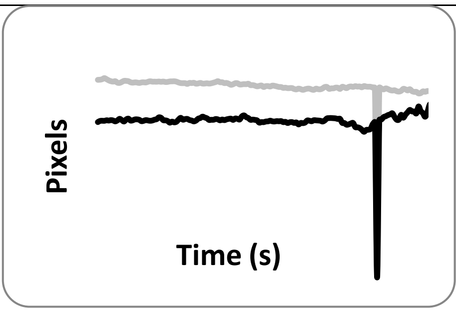
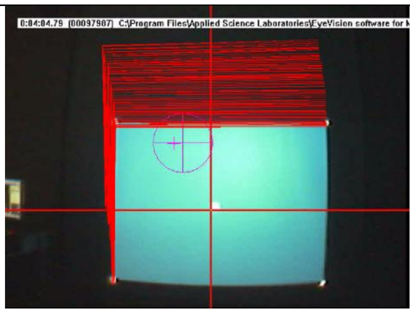
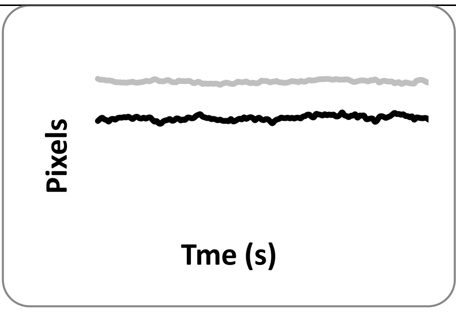
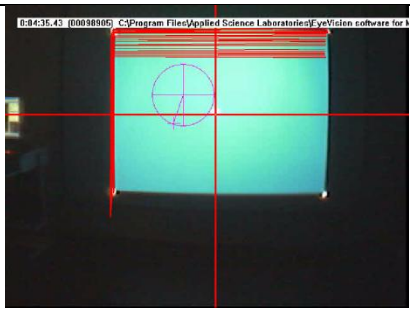
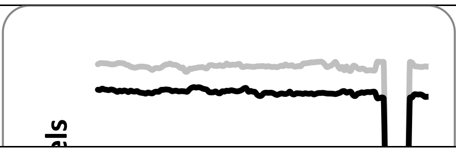
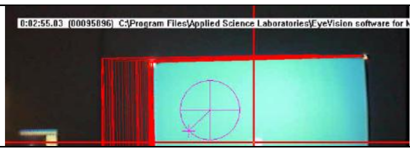


Figure 7. Relative resultant simultaneous gaze patterns and recurrent fixations from five dyads (10 participants) during simultaneous observation trials of 4 s to 8 s in duration (recurrent fixations are circled).

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

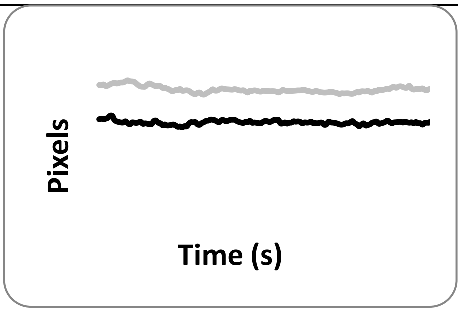
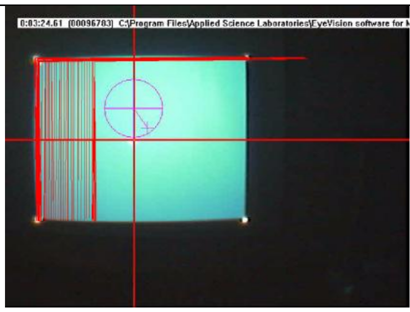
			Vertical *x* = 129 SD = 3 %CV = 2.2
Turn head right			Horizontal *x* = 155 SD = 3 %CV = 1.7 Vertical *x* = 128 SD = 2 %CV = 1.2

Table 2: Individual and combined gaze parameters from five dyads (10 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	Manual coding	Manual calculation	Algorithm	Manual calculation	Algorithm	
1	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)
2	4	4	1.16 s (19%)	1.16 s (19%)						
3	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)
4	5	5	1.16 s (30%)	1.16 s (30%)						
5	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)
6	8	8	2.52 s (50%)	2.52 s (50%)						
7	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)
8	7	7	3.48 s (69%)	3.48 s (69%)						
9	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)
10	10	10	3.96 s (66%)	3.96 s (66%)						
Mean	7	7	2.74 s (52%)	2.74 s (52%)	0.8	1.2	1.2	0.35 s (7%)	0.35 s (7%)	76 pixels (355 mm)