Published in final edited form as:

Psychon Bull Rev. 2010 October ; 17(5): 710–717. doi:10.3758/PBR.17.5.710.

Eye-response lags during a continuous monitoring task

Christina J. Howard*, Tom Troscianko*, and Iain D. Gilchrist*

*Department of Experimental Psychology, University of Bristol, 12a Priory Road, Bristol, BS8 1TU, UK

Abstract

We measured the temporal relationship between eye movements and manual responses while experts and novices watch a video-taped football match. Observers used a joystick to continuously indicate the likelihood of an imminent goal. We measured correlations between manual responses and between-subjects variability in eye position. To identify the lag magnitude, we repeated these correlations over a range of possible delays between these two measures and searched for the most negative correlation coefficient. We find lags in the order of two seconds, and an effect of expertise on lag magnitude, suggesting that expertise has its effect by directing eye movements to task-relevant areas of a scene more quickly, facilitating a longer processing duration before behavioural decisions are made. This is a powerful new method for examining the eye movement behaviour of multiple observers across complex, moving images.

Introduction

It is well known that observers can extract the general meaning or context of a scene e.g. 'busy street' or 'mountains' in a single glance (Biederman, Rabinowitz, Glass & Stacy, 1974). Responding to the 'gist' of real-world scenes occurs within around 450 ms and with presentations as short as 26 ms (Rousselet, Joubert & Fabre-Thorpe, 2005).

Whilst a coarse, 'gist-like' analysis appears to proceed fairly rapidly, semantic analysis of elements within a scene is somewhat slower. Fixations on semantically task-relevant objects in a scene are typically in the order of 300 ms in duration (Henderson, Weeks & Hollingworth, 1999) but occur after several fixations to non-informative objects. Indeed initial fixations in a scene appear to be determined on the basis of non-semantic factors, as fixation locations are similar for unprocessed scenes and for low-pass filtered versions of the same scenes (Mannan, Ruddock & Wooding, 1995). If continuous monitoring of moving scenes shares similarities with evaluation of static scenes, detailed semantic analysis of moving images may also take some time to proceed. As a result, responses might lag behind visual events.

It remains an open question how evaluation of the semantic content of a scene proceeds when the scene is in constant motion - as in viewing real-time video. The experiments presented here address this issue directly, as well as allowing analysis of the temporal relationship between eye movements and manual responses.

We measure eye movements to investigate attentional deployment in the context of monitoring a video of a football¹ match for imminent goals. We chose this sports monitoring

Tel: (+44) 117 9288450 Christina.Howard@bris.ac.uk.

¹We use the term *football* to refer to association football which in the US is referred to as *soccer*. Five-a-side football is a variation of association football in which each team fields five players rather than the usual eleven.

task as a constrained real-world scenario in which well defined events (goals) occur within a stream of complex actions. Our observers were likely to vary in their expertise on the task allowing analysis of the effects of expertise on visuo-motor lags. Manual responses here take the form of pushing a joystick to reflect the current likelihood of an imminent goal.

In their musical sight reading task, Furneaux and Land (1999) found that experts and novices both moved their eyes to the notes they would be playing around one second in the future i.e. the magnitude of the visual-motor buffer was not affected by expertise, though experts could fit more information into the buffer. Others have shown that fixation patterns are dependent on the cognitive complexity and working memory load required to perform a task, such that fixations can be used to regain task-relevant information (Droll & Hayhoe, 2007; Hardiess, Gillner & Mallot, 2008). It seems reasonable that in a more cognitively demanding task, expertise may moderate the cognitive and mnemonic load for observers and so affect the lag. For these reasons we measured the effect of expertise on the relationship between eye movements and responses in this task.

Method

The stimulus was a video of a real five-a-side football match lasting 40 minutes and displayed by a computer. The match was recorded at a 5-a-side football centre in Bristol using a Sony DCR-SR32 video camera. Observers viewed the match in two blocks corresponding to the play before and after half-time. We filmed the match from stationary camera at a high vantage point to capture the whole pitch. No attempt was made to remove visual information from around the pitch in an attempt to keep the stimulus close to the experience of watching real live or televised sports, where the surrounding environment is likely to be visible. It was, therefore, representative of real world inspection tasks in which non task-related salient events frequently occur.

Observers made a constant judgement about the current likelihood of a goal being scored in the next thirty seconds by moving a joystick as described in the Procedure section. Observers' eye positions were recorded throughout the task using the ASL Mobile Eye head-mounted eye-tracker and Eye Vision software.

The video recording measured 22×18 degrees of visual angle from the observers' vantage point and was projected in a dimly lit room against a white background using a Canon SX6 projector onto a screen at a distance of 1.6 m. Black chequerboard-shaped markers subtending 4×4 degrees were placed at each corner of the video display so that a computer algorithm could be used after data collection to stabilise eye position recording for changes in head position.

Observers

Twenty one observers responded to advertisements for the study in return for course credit or a small monetary reward. All were naïve as to the purpose of the experiment and had normal or corrected-to-normal vision. The mean age of observers was 21 years (range 18 - 32 years) of whom nine were female.

Procedure

Observers were given written instructions as follows. They were asked to watch the football match and to monitor the video for potential goals. They were asked to move a joystick to correspond to what they perceived as being the current likelihood of a goal being scored in the next 30 seconds. We chose 30 seconds as a unit of time in which this prediction task should be both reasonably possible and also challenging. They were told that at all times, the joystick should reflect what they perceived as being the current likelihood of an imminent

goal. If they thought that a goal was 100% likely in the next 30 seconds, they were told to move the joystick fully forwards. If they perceived that there was currently no chance of a goal being scored in the next 30 seconds, they were told not to move the joystick at all. They were informed that they could push the joystick to any level in between these two extremes. Observers were asked to keep their hand on the joystick at all times.

At the end of the task, observers completed a 17-item football quiz covering football-related general knowledge, history and teams. They were asked to report the number of hours of football they had watched over the last two weeks, how many matches they had watched over the last year, and rated how much they like watching football between 1 (not at all) to 7 (very much). These four measures were later used to assess the effect of expertise on performance.

Results

Expertise

Four measures of expertise were examined: football general knowledge as a score out of 17 questions on the quiz, hours of football watched over the last two weeks, number of matches watched over the last year, and self-reported enjoyment of watching football.

The mean number of matches watched in the last year was 15.9, ranging from 0 to 100 matches. Observer 18 gave the highest response to this question, more than two standard deviations above the mean and more than two standard deviations away from the second highest value (35). As a result this participant was removed from further analysis, after which the mean number of matches watched was 11.5 and the mean rating giving on the enjoyment scale was 4.1. Observers reported having watched between 0 and 5 hours of football in the last two weeks with a mean of 1.5 hours. Scores on the quiz ranged from 0 to 8, with a mean of 3.0. All four normalised measures were correlated at the 1% significance level with the mean of the other three. Hence, we created a new composite measure of expertise by taking the mean for each observer of their normalised scores on the four measures.

Ratings of perceived suspicious events

Joystick ratings were sampled at 100Hz and recorded from 0 (neutral) to 5 (maximum goal likelihood). One observer gave maximum ratings for the last 6.7 minutes of the first block which was atypical for their own behaviour and compared to other observers who never held the joystick at maximum for more than 200 seconds. After removing these ratings, the total joystick dataset had a mean value of 0.74, standard deviation 1.18 units. We calculated the total overall mean and standard deviation of responses for each observer and used these to normalise each observer's data.

Consistency of eye gaze position

For each frame in each video stimulus, we calculated a measure of spread of eye positions across participants. The spread was calculated as follows: for each frame we first calculated the interquartile ranges (IQR) for horizontal and vertical eye positions. We used IQR as a measure of variability to minimise the influence of position outliers. We then calculated for every frame the mean of these two IQRs. This 'spread value' is a measure of the extent to which all observers were looking at the same part of the screen at the same time. Note that a low degree of eye position spread will result whenever most observers are looking at the same static event or the same moving event, since the difference between observers' eye positions will remain small even if they are all tracking the same moving object. The spread measure had a mean value across the whole dataset of 2.79 degrees, standard deviation 0.97

degrees. Eye position spread was negatively correlated with joystick responses ($r = -0.06$, $p < 0.01$), showing that observers tended to look at the same things as one another when high responses were given.

Total convergence - individual response lags

We first calculated estimates of the temporal lag between eye position spread in the total data set and the responses made by individuals. This relationship is illustrated in Figure 1 and represents how long it took each individual to respond to events that caused a reduction in total eye position spread.

To test for a lag, we measured correlations after artificially shifting the eye spread data forwards and backwards in time. For instance, to test for a 100 ms lag, we shifted the eye spread data 100 ms forwards in time relative to the response data and recalculated the correlation value. At the best estimate of the lag, this correlation should be maximally negative. Note that these correlations and resulting lag calculations represent not a single moment in time, but all moments in the eye spread measures and manual responses over the whole duration of the stimulus.

We performed this series of correlations separately for each observer using their normalised joystick responses against the total eye position spread measure. Missing data created by these artificial time shifts was replaced with the mean spread value. This is equivalent to cross-covariances, or the cross-correlation, which produced similar results. The results of these lag analyses are shown in the upper panel of Figure 2, and the mean of all these curves is shown in the lower panel.

Negative correlations indicate that observers rate the likelihood of a goal as high at the times when observers tend to be looking at the same area of the video. A lagged dip in this correlation curve indicates a delay between this convergence of eye positions and pushing the joystick. Individuals' lag curves show some noise in terms of the negativity and clarity of the dip, though minima are clustered at lags of around 2500 ms. In addition, the negative correlations will be attenuated by spread reductions that are a result of task-irrelevant but salient events.

However, the general trend in the individual curves is for lagged negative correlations i.e. a tendency to push the joystick shortly after all observers' eye movements converge. The lower panel shows the mean of all twenty total convergence - individual response lags where this trend is confirmed by a highly significant maximum negative correlation at 2040 ms ($r = -0.11$, $p < 0.01$).

During the football match, players tend to move towards and away from goals in a roughly periodic manner which may affect perceived goal likelihood. To assess the extent of periodicity in the joystick ratings, we thresholded the median normalised ratings across observers such that ratings above the normalised mean of zero were coded as 1, otherwise 0. The distribution of intervals between the midpoints of these supra-threshold periods displayed a clear peak at 10700 ms. Thus when the ratings data are displaced by half this period (5350 ms) from the most negative correlation (when eye spreads and ratings are maximally in antiphase), we should expect the ratings and eye spreads to be in phase and produce a positive correlation, which we do indeed observe in the lower panel of Figure 2 at around -3310 ms.

The lags ranged from 1120 ms to 4240 ms with a mean of 2280 ms, standard deviation 770 ms. As an alternative method, we also calculated cross-covariances, or the cross-correlation between the two mean-removed eye spread and response sequences. This produced very

similar lags ranging from 1130 ms to 4240 ms with a mean of 2300 ms, standard deviation 770 ms.

When we repeated this lag analysis both for frames when a goal was imminent (i.e. would be scored within the next 30 seconds) and not imminent (i.e. would not be scored within the next 30 seconds), we found no stable effect on the lag (lags of 2300 ms and 2200 ms respectively).

Figure 3 shows the significant negative correlation between the expertise measure detailed in the Expertise section and lag ($r = -0.46$, $p = 0.042$), where relative experts displayed shorter lags than relative non-experts. The same negative relationship was produced from lags calculated using cross-covariance ($r = -0.45$, $p = 0.047$).

Group convergence - group response lags

The total convergence - individual response lags calculated up until this point have represented the time difference between the eye position spread for the total data set, and the joystick responses for each individual, or how long it took each individual to respond to events that caused a reduction in the whole group eye position spread.

The total eye position spread measure does not, however, examine possible differences in when the experts' eye positions converge, and when the non-experts' eye positions converge. To investigate this, we created two groups of observers, high and low expertise, based on a median split of values on the composite expertise measure. We can now start to examine the time delay between the eye position spread of each group, and their manual responses as a group, or group convergence - group response lags. Figure 4 shows these lag curves for both groups. Each curve shows the mean of the correlation values at each lag for all observers in that group. Note the more pronounced dip in the curve for relative experts, suggesting a stronger relationship between group convergence and group responses for experts than non-experts.

The responses of the non-expert group were best correlated with their group eye position spread at a lag of 1360 ms ($r = -0.08$, $p < 0.01$) compared to a lag nearly a second longer of 2260 ms ($r = -0.10$, $p < 0.01$) in the experts. Using cross-covariances to derive lags produced very similar lags of 1340 ms for non-experts and 2470 ms for experts. These group convergence - group response lags thus show the opposite of the expertise dependence seen in the total convergence - individual response lags where expertise was associated with shorter mean lags at 2020 ms for experts and 2550 ms for non-experts. Together these results suggest that relative to the depicted events, experts fixate relevant events earlier and respond sooner, but that the lag between group eye movements and responses is longer for experts than non-experts. Figure 5 illustrates this relationship and shows a cartooned set of frames (not actual data) created to demonstrate lags between eye spread events and responses.

Let us consider a sample frame in which there is low eye position spread in the expert group and we call this as the first landmark event, Time 1. This is followed shortly afterwards by Time 2, a second landmark time at which the total group of twenty observers display low spread as the non-experts also start to fixate relevant events. At Time 3 the non-expert group display low eye position spread, and by this time the expert group are no longer fixating these events. Shortly after this at Time 4, experts initiate joystick pushing to reflect their perception of high goal likelihood. Lastly, Time 5 follows as the non-experts push the joystick. The total convergence - individual response lags reflect the time difference between Time 2 and Time 4 for the experts, and the time difference between Time 2 and Time 5 for the non-experts. The group convergence - group response lags reflect the time

difference between Time 1 and Time 4 for the experts, and Time 3 and Time 5 for the non-experts.

Because these correlations were calculated over the whole stimulus duration, this relationship holds for whichever frame is assigned the label Time 1. For instance, if we took a frame on which experts' eye position spread was high, we would expect to see high levels of eye position spread in the non-experts shortly thereafter, followed by a reduction in joystick pushing in the experts, followed by a reduction in joystick pushing in non-experts. Similarly, mid-levels of eye position spread in the experts will be followed by mid-level eye spread values in the non-experts, and shortly thereafter, midlevels of joystick ratings in the experts, followed by mid-levels of ratings in the non-experts.

Interestingly, relative experts and non-experts did not differ in their absolute performance in this task. For each individual, we calculated the difference between joystick positions whenever a goal was scored within the next 30 seconds and the joystick positions whenever no goal was scored in the next 30 second period. There was no reliable difference between experts (3.2% difference score) and non-experts (2.8% difference score) $t(18) = 0.16$, $p < 1$. The mean difference score represented only 3.0% of the range of the joystick suggesting that the task was difficult for observers. However, even in the absence of a difference in performance levels, experts and non-experts displayed differences in their behaviour during the task. Even though expertise is not affecting *how well* the task is performed, it still had a profound effect on *how* the task was performed.

Discussion

We here report individuals' scene evaluation responses lagging behind total fixation convergence by around one to four seconds. This is much longer than one might expect if observers were simply performing gist perception of the kind previously observed for static scenes (e.g. Biederman, Rabinowitz, Glass & Stacy, 1974) which occurs within half a second (Rousselet, Joubert & Fabre-Thorpe, 2005; Thorpe, Fize & Marlot, 1996).

It seems probable that scene analysis at a more detailed semantic level extends the processing time required after fixation in the order of seconds. This would be consistent with the findings of Henderson, Weeks & Hollingworth (1999) who reported fixations on semantically task-relevant objects after several fixations to non-relevant objects. Another possibility is that the long lag reflects processing of moving video stimuli which of course contain multiple complex motion signals. We know from tasks using much simpler visual stimuli that continuous processing of multiple moving luminance gratings can result in processing lags of up to a quarter of a second even when monitoring as few as four objects (Howard & Holcombe, 2008). When the stimuli to be monitored are dissimilar to each other, potentially visually degraded and containing multiple elements, for instance, the degraded outline of many partially occluded human figures in different clothing, cognitive and perceptual load is likely to be very high, perhaps contributing to the large lags we report here.

The lags measured here between eye movements and manual responses are longer than, and should not be confused with those traditionally observed in the vision for action literature. In a range of naturalistic tasks it has been shown that eye movements lead actions by around half to one second such as when making tea (Land, Mennie & Rusted, 1999), preparing sandwiches (Hayhoe, 2000), steering a car (Land, 1996) and performing musical sight reading (Furneaux & Land, 1999). This visual buffer for action may well have contributed to the lags we report here, though there are several marked differences between the task presented here and the naturalistic tasks used elsewhere. First, eye movements here were not

made towards the subject of actions. Second, the task here is also very different from tasks in which the visual information is determined by manual behaviour, such as in visual tracking of self-moved objects (e.g. Vercher & Gauthier, 1992). And third, the manual task here is a complex evaluative one and so involves an increased level of cognitive processing between fixation and judgement. Importantly, the task here was to continuously and explicitly evaluate the scenes presented rather than to act within them.

Here we show expert observers detecting relevant events in the video earlier in terms of their joystick responses, and also fixating those relevant sections earlier than non-experts. It seems likely that their early fixations are a result of experience, or ‘knowing what to look for’. Earlier fixation of task-relevant events could potentially be the whole cause of earlier joystick responses, allowing experts to take their time to process the stimuli before making a manual response. If, as found by Furneaux and Land (1999), expertise allows more information to be held in a visuo-motor buffer, it may also be the case that experts here were able to hold more information in visual memory as events unfolded. Because of this, they may have had to use fewer fixations to regain lost information in the manner shown to occur under high load by Droll and Hayhoe (2007) and Hardiess, Gillner and Mallot (2008).

We present a novel method of measuring lags between eye movements and responses which enables the use of tasks with moving video stimuli from real-life scenarios, as well as on-line continuous evaluations of these stimuli. This method allows a detailed exploration of the temporal relationship between dynamic visual stimuli, eye movements and manual responses. Here we observe differences in when events are fixated, manual response times, and the lags between these events for relative experts and novices at the task of football monitoring. Our results have implications for the time course of scene perception and the continuous monitoring of dynamic real-life environments.

Acknowledgments

This work is supported by a grant from the UK Engineering and Physical Sciences Research Council and from the Wellcome Trust.

References

- Biederman I, Rabinowitz JC, Glass AL, Stacy EW. On the information extracted from a glance at a scene. *Journal of Experimental Psychology*. 1974; 103(3):597–600. [PubMed: 4448962]
- Droll JA, Hayhoe MM. Trade-offs between gaze and working memory use. *Journal of Experimental Psychology: Human Perception and Performance*. 2007; 33(6):1352–1365. [PubMed: 18085948]
- Furneaux S, Land MF. The effects of skill on the eye–hand span during musical sight-reading. *Proceedings of the Royal Society of London B*. 1999; 266:2435–2440.
- Hardiess G, Gillner S, Mallot HA. Head and eye movements and the role of memory limitations in a visual search paradigm. *Journal of Vision*. 2008; 8(1):7, 1–13. [PubMed: 18318610]
- Hayhoe M. Vision using routines: a functional account of vision. *Visual Cognition*. 2000; 7:43–64.
- Henderson JM, Weeks PA Jr, Hollingworth A. The effects of semantic consistency on eye movements during scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*. 1999; 25(1):210–228.
- Howard CJ, Holcombe AO. Tracking the changing features of multiple objects: Progressively poorer perceptual precision and progressively greater perceptual lag. *Vision Research*. 2008; 48:1164–1180. [PubMed: 18359501]
- Land MF. The time it takes to process visual information when steering a vehicle. *Investigative Ophthalmology & Visual Science*. 1996; 37:S525.
- Land MF, Mennie N, Rusted J. The roles of vision and eye movements in the control of activities of daily living. *Perception*. 1999; 28:1311–1328. [PubMed: 10755142]

- Mannan S, Ruddock KH, Wooding DS. Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-D images. *Spatial Vision*. 1995; 9(3):363–386. [PubMed: 8962841]
- Rousselet GA, Joubert OR, Fabre-Thorpe M. How long to get to the ‘gist’ of real-world natural scenes? *Visual Cognition*. 2005; 12(6):852–877.
- Thorpe S, Fize D, Marlot C. Speed of processing in the human visual system. *Nature*. 1996; 381:520–522. [PubMed: 8632824]
- Vercher J-L, Gauthier GM. Oculo-manual coordination control: Ocular and manual tracking of visual targets with delayed visual feedback of the hand motion. *Experimental Brain Research*. 1992; 90:599–609.

Figure 1. Cartooned illustration of a total convergence - individual response lag. In this example, manual responses for one observer are shown on the left and eye positions for the all participants on the right. This observer's responses are at a maximum at time 5, corresponding to the minimum total eye position spread at time 3.

Figure 2.

upper panel: Total convergence - individual response lags for twenty observers. Crosses show the most negative correlation for each observer.

lower panel: Mean of all twenty total convergence - individual response lags. The maximum negative correlation occurs at 2040 ms. Error bars indicate standard error of the twenty individual lags.

Figure 3. Negative relationship between expertise and total convergence - individual response lags.

Figure 4.

Group convergence - group response lags: time differences between group eye position spread and group manual responses for non-experts (shown in grey) and relative experts (shown in black). Error bars represent standard errors obtained by bootstrapping.

Figure 5.

Cartooned timeline of events represented by the two lag types for non-experts (shown in grey) and relative experts (shown in black). At Time 1, experts' group eye position spread is low, and experts respond with the joystick correspondingly 2260 ms later. At Time 3, the non-experts' group spread is low and non-experts push the joystick 1360 ms later. At Time 2, the spread measure for the total data set is low, and experts push the joystick relatively soon, 2020 ms later, whilst the non-experts, because they take longer to fixate the relative events, push the joystick 2550 ms later.