VISUAL SAMPLING IN A ROAD TRAFFIC MANAGEMENT CONTROL ROOM TASK

Sandra Starke[1], Neil Cooke[1], Andrew Howes[2], Natan Morar[1] & Chris Baber[1]

[1] *School of Electronic, Electrical and Systems Engineering and*
[2] *School of Computer Science, University of Birmingham*

Control room tasks like road traffic management require continuous visual assessment paired with active interventions. This includes monitoring a range of information sources and resolution of live scenarios in real time using Standard Operating Procedures. We investigated how three operators resolved a simulated 'object in the road' task, requiring information gathering and integration from multiple displays. Visual sampling was quantified from eyetracking and related to a Hierarchical Task Analysis. Operators followed heterogeneous strategies to accomplish the same main goal, differing in the attended ROI sequence and task component weighting. Differing preferences for information sampling were accommodated by prescribed data entry fields in the incident log.

Introduction

In road traffic management control rooms, operators perform monitoring, supervisory and executive roles: after receiving an alert from an external source about incidents such as a congestion or accident, their role encompasses monitoring a multitude of information sources, assessing the severity of an incident, interacting with the road network *via* lane closures / speed limits to manage the situation, and recording an incident log. Little is known how different observers navigate multiple information sources to complete these activities.

Control room operators rely on visual processing of information, which is facilitated through 'visual sampling', the sequence of overt visual attention to regions of interest (ROIs) executed by head and eye movement: due to the physiological properties of the eye, humans have to perform continuous eye movements ('saccades') to build up a mental representation of the world around them using information from 'fixations' (gaze directed at stationary object) or 'smooth pursuit' (tracking a moving object). The use of eye tracking devices to understand the underlying cognitive processes has a long tradition especially under the consideration of 'active vision' (Findlay and Gilchrist, 2003), the active direction of gaze towards regions that hold task-relevant information.

Visual search refers to individuals sampling a scene with the goal of finding information. Search patterns can be systematically changed in response to the goal (Yarbus, 1967) or training (Chapman et al., 2002) of the searcher. 'Bottom-up' search is typically feature driven, i.e., based on image salience (prominent visual features) or distractions; 'top-down' search is typically based on the user task (Doshi and Trivedi, 2012). For the task of this study, we assume that visual sampling is not exclusively guided by image saliency but by the goals of the observer, reflecting an 'internal agenda' (Yarbus, 1967, Hayhoe and Ballard, 2005). This assumption has been argued in context of many real-life tasks, where visual search patterns are closely linked to an observer's 'schema', or mental task representation, being controlled top-down (Land, 2009). This raises the challenge of developing a combination of eye-tracking metrics which can be related to the goals that the operator is seeking to achieve. This is the first aim of the study reported in this paper.

The redirection of gaze can be accomplished by a combination of eye-, head- and body movement depending on the desired saccade amplitude. While gaze shifts $> 40°$ commonly require a contribution from head movements *per se*, literature values on head contributions to smaller gaze shifts vary (Fuller, 1992). In general, eye-head coupling is highly subject specific (Thumser and Stahl, 2009) and even small gaze shifts $< 15°$ can have a head contribution (Goossens and Opstal, 1997). While knowledge about the cognitive reasons for eye-head coupling remains sparse, head movements have been linked to processing demands (Stein, 1992) and the temporal lag between eye- and head movement appears linked to top-down/bottom-up attention shifts (Doshi and Trivedi, 2012). Thus, the second aim of this study is to consider ways in which eye and head movements are combined in visual search.

It is proposed that the notion of an 'internal agenda' is beneficial to Ergonomics studies of operator performance because there ought to be correspondence between eye movement, head movement and the goal structure of a task. This pilot study provides a descriptive framework of operator behaviour in a road traffic management control centre performing a standard task. We investigated whether operators follow comparable workflows with respect to visual information sampling when resolving incidents, given the multitude of sources they can use, or whether there are individual differences in strategy between operators. Specifically, we were interested in the relationship between head movements, gaze shifts and their relation to goal structure. This work is part of a larger study into interactive human decision making where decisions are facilitated by continuous ongoing visual attention.

Methodology

Location and scenario
Data were collected from three expert operators (one female, two male) at the road traffic management facility at DIR Centre Est, Grenoble, France. The operators had a minimum of three years post-training experience. The protocol

for data collection had been approved by the University of Birmingham Ethics Committee. Figure 1 shows the control room, which has, on the desk in front of the operator, five computer monitors displaying a user interface for incident logs (LG), a schematic map of the traffic network (MP), a live CCTV feed linked to the camera that the operator is currently controlling (TV) and access to the internet or other information (D1 and D5). In the background is a large display panel (BS) with a 4x4 colour array of CCTV feeds from different traffic cameras and a smaller display panel (SS) presenting colour CCTV feeds which can be interacted with from the desk. Also located on the desk are standard PC peripherals and phones / radios for communication with stakeholders from outside the facility such as traffic patrol staff and emergency services personnel.

The facility is a fully operational traffic control centre, hence forbidding the interruption of incoming data with a recorded scenario. For this reason, operators were asked to engage in a pretend task, initiated by a call from a member of staff which simulated a standard alert of 'object in the road'. The rationale behind this approach was that operators were assumed to navigate this pretend task in a manner representative of their 'average' behaviour, hence reducing the confounding influence of otherwise highly specific live events.
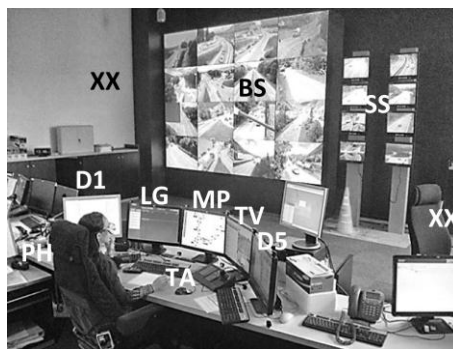

Figure 1: Road traffic control room with participant wearing mobile eyetracker (Tobii glasses v.1).

Collection and analysis of eyetracking data
Eyetracking data were recorded using Tobii glasses with a sampling frequency of 30 Hz. Prior to each recording session, the eyetracker was calibrated to each individual participant. To later relate the gaze data in the local glasses reference frame to the global reference frame of the control room, infrared markers were placed in strategic positions around the display units. All participants received an explanation of the task and had the opportunity not to participate.

Following data collection (task duration: 3.8 to 4.5 min.), gaze data were automatically mapped onto ten regions of interests (ROIs, Figure 1) defined by the IR markers using Python. Head orientation was automatically inferred from

the video data. Subsequent analysis was performed in Matlab (MathWorks). Data were pre-processed using conditional rules, such as the removal of data points in one data stream (e.g., eyes) if the other stream (e.g., head) had a dropout. Times at which observers attended to external stimuli (such as unrelated phone calls) were cropped out. The resulting dataset had a trackability ranging from 81 % to 87 %. For gaze data, cumulative percentage viewing time per ROI, frequency of switches between ROIs and viewing networks were calculated. Agreement between the ROI attended to by eye and head was calculated for each sample (excluding empty samples) and expressed in % total tracked viewing time.

Defining Operator Goals

Hierarchical task analysis (HTA) was performed using observations and interviews with subject matter experts, corroborated with eyetracking recordings across participants performing both real and pretend tasks to define goals, sub-goals and plans. Goals and subgoals of the operators were mapped to ROIs, i.e., the display which operators were most likely to use for a given subgoal.

Relating visual sampling to task goals

To relate the observer's navigation through the task to the HTA, video recordings of each observer where mapped to the identified steps in the task analysis. For the eyetracking data, 100 equally spaced bins were created for the duration of the task. For each bin, the ROI with the highest viewing time within that bin was extracted, arriving at a smoothed signal holding the main stimulus participants attended to, 'filtering out' high frequency switches between ROIs. Results from this mapped gaze data and HTA were then contrasted qualitatively.

Results

Attended ROIs

The cumulative % viewing time (Figure 2) for different displays varied largely: participant 1 spent most time viewing LG / SS, participant 2 favoured LG / MP and participant 3 favoured LG / TV. None of the participants allocated noteworthy time (viewing time for ROI ≤ 1.6 %) to D1, D5, PH, BS and XX.
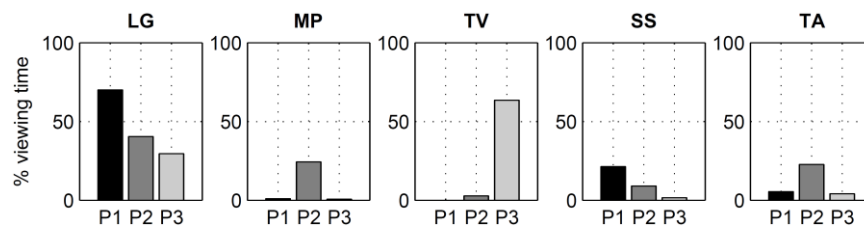
Figure 2: Cumulative percentage viewing time allocated to those five regions of interest (ROIs) within the display configuration of the control room that received noteworthy attention. Black - participant 1, dark grey – participant 2, light grey – participant 3.

Sequential information sampling

Participants followed different search patterns when redirecting gaze between the different display options (Figure 3). The switch frequency between ROIs averaged 1 switch every 5 to 10 seconds (0.10 to 0.27 Hz) for the eyes and 0.11 to 0.27 for the head (Figure 4). However, switches were not evenly distributed, often showing bursts. The match between ROIs attended to by eyes and head ranged from 88 % to 97 % viewing time (Figure 4).
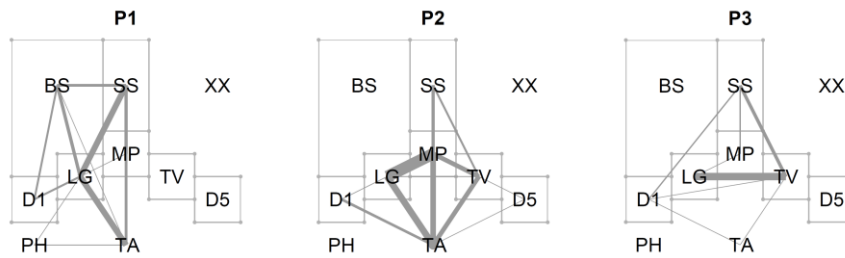

Figure 3: Viewing networks within the schematic control room layout of participant 1 (left), 2 (middle) and 3 (right), illustrating switches between ROIs. Line thickness proportional to switch count.
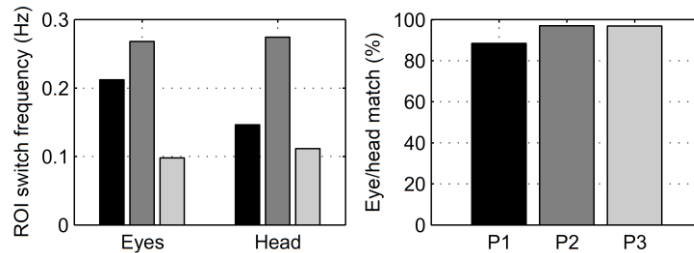

Figure 4: Left – average switch frequency between ROIs. Right – percentage of time for which head and eyes pointed at matching ROIs. For colour coding please refer to Figure 2.

Task navigation

Operators navigated through the task systematically, viewing information sources relevant to the task (Figure 5). Rather than frequently switching information sources during individual subtasks, operators tended to adhere to a single source. For each operator, task navigation was intersected by entries into the digital log (LG). After navigating the first three subgoals of the HTA, namely 1. Receive notification, 2. Determine incident type, 3. Determine incident location, participants differed in their simulation of 4. Determine incident impact, 5. Initiate response and 6. Monitor road user compliance. There were differences at further sub-task levels (Figure 5) in the style of task execution between operators (e.g. participant 2 took written notes on paper while participant 3 monitored the CCTV feed on TV).
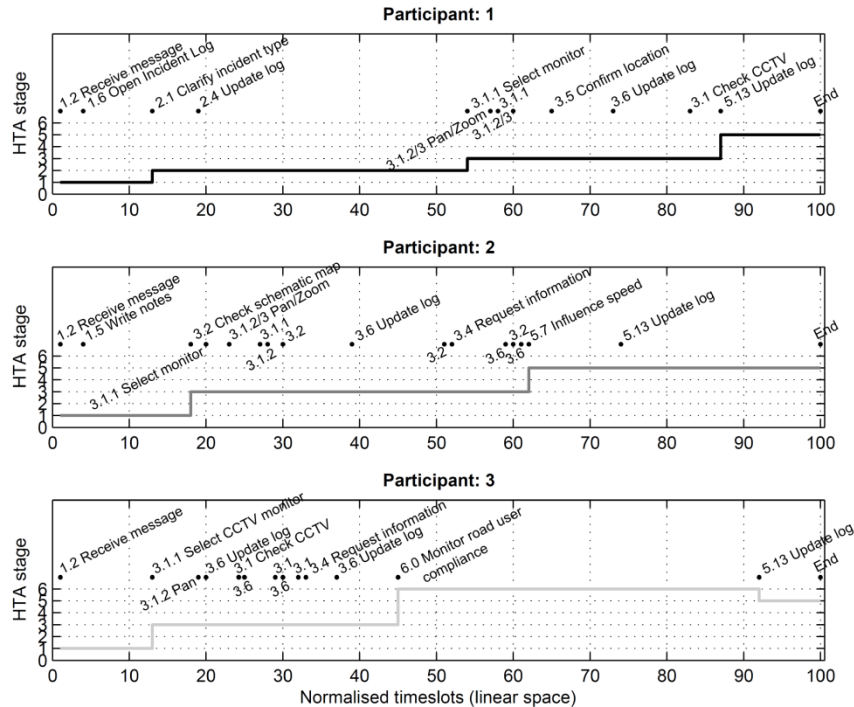
Figure 5: Stages of the hierarchical task analysis (HTA) normalized to 0-100% task time. For colour coding please refer to Figure 2.

Discussion and conclusions

In this study, operators favoured different information sources to complete the same goal. The difference in CCTV monitoring between participant 1 and 3 can be explained, as the selected CCTV feed on SS can also be viewed on TV – the two participants had developed different preferences regarding the viewing modality. However, while participant 2 emphasised information gathering from the schematic road map on MP, the other two participants largely ignored this information source. The lack of attention to the large bank of CCTV feeds which we found matches a recent review in the CCTV crime monitoring domain: those authors highlighted that with increasing number of feeds (16 on BS in our study) both observer accuracy and confidence rapidly diminish (Stainer et al., 2013). Instead, our three observers relied on the interactive bank of smaller CCTV screens (SS), selecting a relevant camera and adjusting the view in order to extract desired information. We observed that operators may attend to one source of information for the majority of time, similar to operators searching a bank of CCTV screens attending primarily to a selected feed on a separate monitor (Stainer et al., 2013) or air traffic controllers attending to a preferred information source (Stein, 1992).

We found that operators aligned head and eye direction with respect to the attended ROIs for the majority of time. On the one hand, this finding may be due to chance (N = 3): the contribution of head movements to gaze shifts is generally highly subject specific; while some subjects move their head a lot, others do not (Boyer, 1995, Goossens and Opstal, 1997). For example, in an airtraffic control task the distribution of operators moving their head when shifting gaze from a monitor to an input device was approximately 50:50 (Boyer, 1995). On the other hand, we propose that our findings result from the top-down gaze control based on the internal task schema of the operator: humans executing domestic tasks typically move their body in the direction of objects of interests prior to moving the eyes (Land, 2009). In our study, operators were familiar with the control room layout and we conclude that head movement indicated their higher level goal structure similar to Land's (2009) task scenarios. Alignment of head and gaze direction can be affected by physiological factors, i.e., increasing gaze eccentricity causes stress on the eye muscles, which compensatory head movement helps relieve (Sanders, 1963). Hence, operators might seek to move their heads to minimise gaze eccentricity. However, we believe that the key factor in aligning eye and head in the control room is the arrangement of ROIs and their relation to operator goals. We noticed that occasionally the eyes briefly flick to a new ROI while the head remained directed at the old target. We hypothesise that this relates to the operator's goal structure, and that analysis of disjunctions between ROIs attended to by eyes and head serve to quantify changes between primary and secondary goals.

In static displays, re-sampling has been linked to refreshing memory content (Peterson and Beck, 2013) and has received attention in context of 'embedded' (Spivey and Dale, 2013) or distributed cognition, where the location of information is stored in memory to access detailed information when needed. This raises an interesting question of what needs to be encoded during visual search: do operators encode the content of the displays, perhaps in some form of mental model, or do they encode the location of pieces of information, in an approach more closely aligned with distributed cognition? In control rooms where information content continuously evolves, re-sampling is essential to maintain awareness of dynamic situations. In our study, operators clearly performed task-directed sequential sampling. The different information sampling strategies we found may indicate that in an over-determined system, operator decision making may lead to differences in task navigation and that there may be several rationally optimal solutions to a task. Since the task was a simulated event, operator behavior may differ to that observed in a real event (although initial comparison with data collected during performance of real tasks suggests that such differences may not be as clear cut as one might imagine). In the present scenario, we believe that the eye- and head-movement data can be used to infer observers' perceived importance of individual regions of interest relative to the goal structure that they were employing. Following the structured incident log ensured that despite heterogeneous sampling approaches, operators extracted information relevant to resolving the incident systematically. Our results suggest that operators, in a control room, have developed idiosyncratic strategies for

information search, that these strategies reflect the emphasis that operators place on the ways in which they navigate the goal structure of the tasks they are performing, and that eye and head movement give insight into these strategies in ways that either recording operator activity through HTA or interviewing operators might miss.

References

Boyer, D. J. 1995, *The Relationship Among Eye Movements, Head Movements, and Manual Responses in a Simulated Air Traffic Control Task*, U.S. Department of Transportation Federal Aviation Administration.

Chapman, P., G. Underwood and K. Roberts. 2002, Visual search patterns in trained and untrained novice drivers, *Transportation Research Part F: Traffic Psychology and Behaviour* 5(2): 157-167.

Doshi, A. and M. M. Trivedi. 2012, Head and eye gaze dynamics during visual attention shifts in complex environments. *Journal of Vision* 12(2).

Findlay, J. M. and I. D. Gilchrist. 2003, *Active vision: The psychology of looking and seeing*, Oxford: Oxford University Press.

Fuller, J. 1992, Head movement propensity. *Experimental Brain Research* 92(1): 152-164.

Goossens, H. H. L. M. and A. J. V. Opstal. 1997, Human eye-head coordination in two dimensions under different sensorimotor conditions. *Experimental Brain Research* 114(3): 542-560.

Hayhoe, M. and D. Ballard. 2005, Eye movements in natural behavior. *Trends in Cognitive Sciences* 9(4): 188-194.

Land, M. F. 2009, Vision, eye movements, and natural behavior. *Visual Neuroscience* 26(01): 51-62.

Peterson, M. S. and M. R. Beck (2013) Eye movements and memory, In S. P. Liversedge and I. D. Gilchrist (eds.) *The Oxford Handbook of Eye Movements*, Oxford: Oxford University Press.

Sanders, A. F. 1963, *The selective process in the functional visual field*, Institute for perception RVO-TNO.

Spivey, M. J. and R. Dale (2013) Eye movements both reveal and influence problem solving, In S. P. Liversedge and I. D. Gilchrist (eds.) *The Oxford Handbook of Eye Movements*, Oxford: Oxford University Press.

Stainer, M. J., K. C. Scott-Brown and B. Tatler. 2013, Looking for trouble: A description of oculomotor search strategies during live CCTV operation. *Frontiers in Human Neuroscience* 7: article 615.

Stein, E. S. 1992, *Air Traffic Controller Visual Scanning*, U.S. Department of Transportation Federal Aviation Administration.

Thumser, Z. and J. Stahl. 2009, Eye–head coupling tendencies in stationary and moving subjects. *Experimental Brain Research* 195(3): 393-401.

Yarbus, A. L. 1967, *Eye movements and vision*, New York: Plenum Press.