Eye Metrics for Task-Dependent Automation

Puck Imants
Faculty of Psychology and Eduction
VU University Amsterdam
Van der Boerchorststraat 1, 1081 BT
Amsterdam, The Netherlands
p.imants@vu.nl

Tjerk de Greef
Department of Intelligent Systems
Delft University of Technology
Mekelweg 4, 2628 CD Delft,
The Netherlands
t.e.degreef@tudelft.nl

ABSTRACT
Future air traffic is expected to grow increasingly, opening up a gap for task dependent automation and adaptive interfaces, helping the Air Traffic Controller to cope with fluctuating workloads. One of the challenging factors in the application of such intelligent systems concerns the question what the operator is doing in order to optimize support and minimize automation surprises. This study questions whether eye metrics are able to determine what task the operator is engages in. We therefore examined A) if the eye-path would differ for three different ATC tasks and B) whether this effect can be quantified with six eye-metrics. In an experiment, the six eye-metrics were calculated and used as a dependent variable. The results show that some tasks can be inferred by eye movement metrics and other metrics infer workload, although none inferred by both task and workload.

Author Keywords
Eye-tracking data; scan-path; fixations; adaptive automation; air traffic control

ACM Classification Keywords
H.4.2 Types of Systems: Decision Support; K.4.3 Organizational Impacts: Automation

INTRODUCTION
Air Traffic Controllers (ATCo) are responsible for safe and efficient air traffic and need to find ways to expedite an orderly flow of aircraft. An important indicator for the workload of the controller is the flow of air traffic, at times leading to cognitive overload or underload. These cognitive conditions increase risks due to intentional tunneling, cognitive lockup, and the out-of-the-loop syndrome [1], at unfortunate moments contributing to fatal accidents [2].

Task dependent automation and adaptive interfaces can help the ATCo to cope with the increased flight movements [3]. One of the challenging factors in the application of such intelligent systems concerns the question what the operator is doing to optimize support maximally. We examined A) if the eye-metrics would differ for the three different ATC tasks and B) whether this effect can be quantified with specific eye-metrics.

ECCE '14, September 01 - 03 2014, Vienna, Austria

http://dx.doi.org/10.1145/2637248.2637274

AIR TRAFFIC CONTROL IN THE WILD
The task of expediting an orderly flow in air traffic control can be divided in three, at the work floor representative tasks, namely monitoring, planning and controlling. Monitoring involves checking if all flights are separated according to separation procedures ensuring safe and orderly air traffic. Therefore the ATCo has to identify airplanes and interpret the overall traffic situation. Where monitoring is about assessing the current situation, planning associates with anticipating future safety issues. When controlling the air traffic, the ATCo issues clearances and instructions to ensure the separation of the air traffic.

EYE METRICS
For this study, inspiration was taken from Yarbus [4] who showed that the eye-path is different for different tasks. Generally, people move their eyes quickly over a scene by a sequence of saccades and fixations (see Figure 1). Saccades are rapid movements that move the eyes around a particular scene. The processing of visual information occurs in the fixation that follows [5] (cf. a longer fixation time denotes more information processing). A fixation is a fairly stable gaze at a particular point of interest for a certain time.

Based upon data from an eye-tracking device, various alternative eye metrics are calculated (also see Figure 1). First, the *Scan-path length* is a metric based upon the repetitive saccade-fixation-saccade sequence [6]. The length is the sum of the distance between two fixations, summarized over all fixations. The scan-path length is considered a productivity measure: lengthy scan-paths (cf. extensive search behavior) indicate a less efficient scanning behavior. Secondly, the *average fixation duration* divides sum of fixation times by the number of fixations. A high average duration of fixations in a scan-path implies that more time is spent interpreting or relating the information to internalized representation [7]. Third, the *convex hull area* is the smallest area covering the scan. A small convex hull area indicates a small area where the search took place whereas a large convex hull area indicates a large search area [7]. Fourth, the *fixation clusters* metric describes a set of fixations in a cluster based on temporal and spatial characteristics. When a cluster consists of a high amount of fixations and/or the total fixation duration of the cluster is longer, the target on which the cluster was fixated could be more complex compared to targets for which the fixation clusters have few fixations and/or a shorter duration [7, 8]. If a fixation point was less then 100 pixels from the average position of the fixations in the cluster, the fixation belonged to the cluster. Fifth, the *spatial density* metric quantifies the spatial distribution of fixation points indicating the widespread of fixations. Evenly spread samples throughout the display indicate extensive search whereas samples in a small area reflect direct search. The interface can be divided into grid areas with the spatial density index equal to the number of cells containing at least one sample divided by the total number of grid cells (for this study the screen was divided in 13 by 11 cells, each cell having a width and a height of 100 pixels). Sixth, *Nearest Neighbor index* (NNI) describes the ratio between the distribution of fixation points on the screen (see Formula 1, d(NN) and the random distribution of points on the screen [9] (d(ran)). If the ratio is equal to 1 the distribution is random, if the ratio is smaller than 1 the distribution suggest grouping and if when the ratio is larger than 1 this suggest regularity.

$$d(\mathrm{NN}) = \sum_{i=1}^{N} \left[min \frac{d_{ij}}{N} \right]$$

$$d(\mathrm{ran}) = 0.5 \sqrt{\frac{\mathrm{A}}{\mathrm{N}}}$$

$where\ N\ is\ the\ total\ number\ of\ fixations$

$d(i,j)$: minimal distance (pixels) between two nearest fixations

A is the area of the region, in pixels

Formula 1: calculus for the Nearest Neighbor Index

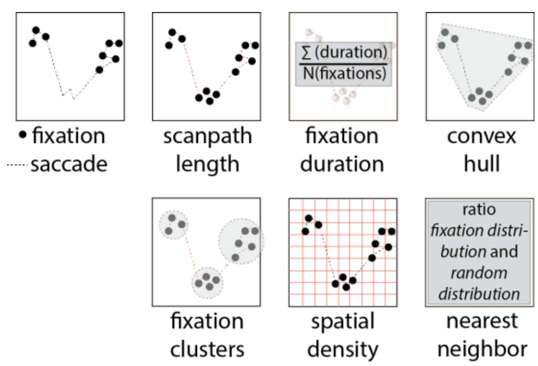

Figure 1. Different eye metrics yield quantified measures of eye movement during tasks executed by the operator.

WORKLOAD

Workload is a variable "…*that modulates or indexes the tuning between the demands of the environment and the capacity of the operator…*" [10]. Workload increases when the capacity of the human decreases or the task demands increase. Several studies indicate that eye movement metrics can be used to infer workload. Examples of metrics studied in relation to workload are blink frequency [8, 11, 12], blink interval [13] and pupil diameter [8, 11]. Research by [9] also indicated that the Nearest Neighbor Index is a sensitive metric for workload.

EXPERIMENT & METHOD

The experiment studies whether the three specified ATC tasks yield different scan-paths and whether this can be quantified using the described metrics. A randomized repeated measures within-subject design was used, where task (monitoring, planning and controlling) and workload (high and low) were the independent variables. The dependent variables were the six eye-metrics and the subjective workload using the Rating Scale Mental effort (RSME) scale [14].

We invited nine naïve participants and taught them a simpler version of the ATC tasks. Subjects participated voluntary but were compensated for their effort with a 15 Euro gift voucher. For one of the participants the experiment was terminated due to tracking complications.

Differentiated in the number of approaching and leaving aircraft, two high and two low workload scenarios were created in ATC-lab. Air traffic control sectors are made up of a series of flight paths, waypoints and airports (see Figure 2). Circles represent the aircraft flying in the air traffic control sectors, with a data block label containing call sign, type, ground speed, current flight level and cleared flight level.

Each run existed of the three ATC-subtasks (monitoring, planning and controlling) for one scenario in the following order; first the participant had to monitor the traffic, for the next sample the participant had to plan future actions and in the last sample the participant gave instructions, i.e. controlled the air traffic. Between the runs, the scenario and workload alternated.

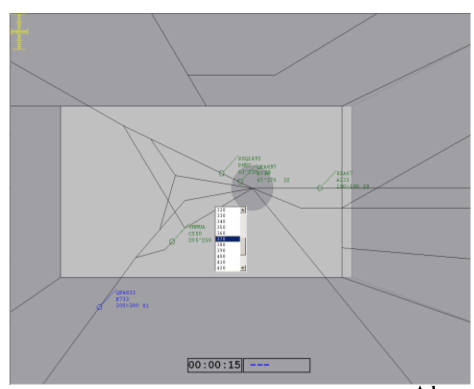

Figure 2. Screenshot of ATC-Lab^Advanced

For the experiment, a set of safety and efficiency rules need to be followed (relevant for all three tasks). A simplified set of rules was devised based upon the Convention on International Civil Aviation. Clearly, these rules are too complicated for an experiment with non-experts and were therefore simplified (Table 1).

Safety Rules
• The vertical separation minimum is 1000 ft • Aircraft flying on different tracks are considered separated • For aircraft on the same track, a vertical separation of 1000ft is required until both aircraft pass each other on the track • The landing interval and minimum separation between arriving aircraft is 2 minutes
Altitude Level Rules
• Aircrafts leaving the sectors East, South, West or North must be assigned the highest flight level possible regarding other (incoming and outgoing) traffic • Aircrafts handed over to approach need to be assigned 7000ft/flight-level (FL) 70

Table 1. Safety rules ensuring safe and efficient flying

Eye data was collected with the EyeLink II eye-tracker, which is head-mounted with 500 Hz binocular eye monitoring and a 0.5º average accuracy. After each run the eye-tracker was corrected for drift and calibrated.

One computer controlled the EyeLink II and a second computer ran ATC-Lab and collected the data from the eye tracker. Both were connected via TCP/IP. The experiment room was dimly lit to facilitate eye tracking.

RESULTS

The Eyelink II software preprocesses the eye-data into fixations, saccades and blinks. Specially prepared Matlab modules calculated the eye metrics before analyzing with the statistical Analysis Toolpack of Microsoft Excel.

Each participant was asked to rate their subjective effort using the RSME-forms in order to crosscheck the high or low workload scenarios. To see if the data from the two low workload scenarios could be analyzed together, a student t-test was used. Next, the student t-test was used to check if the low workload scenarios yielded a significant lower-experienced workload than the high workload scenarios.

The eye movement metrics were individually analyzed with a two-factor (2x3) ANOVA with replication with factors workload (low, high) and task (monitoring, planning, controlling). For the eye movement metrics with a significant effect on task, the Tukey test was performed to differentiate between a q-value. This value was then compared to the critical q-value. A posteriori Tukey test was performed for the eye movement metrics with a significant effect of task.

Verifying Scenarios for workload

There was no significant difference in experienced workload between the two high workload scenarios ($t(24) = 1.64$, $p = 0.11$) or between the two low workload scenarios ($t(24) = 1.55$, $p = 0.13$). Consequently, the data can be taken together for further analyzing. When done, subjects rated workload for the high workload scenarios significantly higher compared to the low workload ones ($t(24)_{\text{scenario 1}} = 6.89$, $p < .001$; $t(24)_{\text{scenario 2}} = 7.72$, $p < .001$).

Eye movement metrics

Data revealed a significant effect of task on the *amount of clusters*, *cluster size*, *cluster duration* and *fixation duration*. Also, the metrics *scan-path length*, *nearest neighbor index*, and *spatial density* showed a significant effect of workload.

Effect of Task

The average duration of fixations (ms) in a scan-path is significantly different between planning and controlling, where the average duration is lower when planning ($M = 3.25 \times 10^2$, $SD = 63.34$) compared to controlling ($M = 3.61 \times 10^2$, $SD = 60.71$).

The average number of clusters when planning and monitoring is significantly different from the average number of clusters when controlling, where the average number of clusters is lower when controlling ($M = 55.22$, $SD = 15.53$) compared to monitoring ($M = 77.47$, $SD = 9.04$) and planning ($M = 77.97$, $SD = 17.01$).

The average duration of fixations (in ms) in the clusters when controlling is significantly different from the average duration for planning and monitoring, where the average duration is higher when controlling ($M = 1881.83$, $SD = 559.86$) compared to monitoring ($M = 1239.93$, $SD = 170.64$) and planning ($M = 1228.24$, $SD = 213.35$).

The average cluster size (in pixels) in the controlling task is significantly different from the average cluster size in the planning and monitoring tasks, where the average size is biggest when controlling ($M = 1652.82$, $SD = 704.87$) compared to monitoring ($M = 941.21$, $SD = 286.68$) and planning ($M = 996.53$, $SD = 404.79$).

	Workload ($F(1,90)$)	Workload (p)	Task ($F(2,90)$)	Task (p)
Amount of clusters (a)	0,24	0,63	25,85*	< 0.01*
Cluster size (b)	3,67	0,06	20,51*	< 0.01*
Cluster duration (c)	0,01	0,94	33,67*	< 0.01*
Fixation duration (d)	1,65	0,20	3,10*	0,05*
Scan-path length (e)	51,81*	< 0.01*	0,54	0,58
Nearest Neighbor (f)	4,07*	0,05*	1,75	0,18
Spatial density (g)	6,41*	0,01*	2,91	0,06

Table 2. Eye-metric statistics. The convex hull area did not show a significant effect as well as there was no significant interaction effect. Both are therefore not included. Values with a * are significant with p < .05

Effect of workload

The scan-path is significantly longer in the high workload scenarios compared to the low workload scenarios, which can be expected being that the high workload condition contained more aircraft. The Nearest Neighbor Index is significantly higher for high workload scenarios compared to the low workload scenarios but both are lower than 1, suggesting grouping. The spatial density is significant higher for high workload scenarios compared to the low workload scenarios.

DISCUSSION & CONCLUSION

Regarding workload, [9] indicated that the Nearest Neighbor Index is a sensitive metric, which is confirmed by the data of this experiment. A higher index corresponds to a higher workload. Furthermore`, data reveals that the scan-path length and spatial density are sensitive for workload, with a higher spatial density index and a longer scan-path indicating high workload. Several studies showed that the visual field, the area visible at a single glance, narrows with increasing workload with longer fixation times and smaller saccades resulting in distinctly different scan-paths and longer search times in visual task performance [15]. A longer scan-path and higher spatial density index are found to indicate less efficient search [8].

Furthermore, the results of this study indicate that various eye movement metrics show a difference between the three ATC subtasks. First, the *duration of fixations* is associated with the ease of processing information, where longer fixation durations indicate difficulty with processing or extracting information [7, 8]. In this study the controlling task resulted in longer average fixation durations compared to the planning task. The subject had to choose a new speed or height for the aircraft that needed to be controlled, choosing the appropriate speed or height from the pull down menu. The subject possibly had to check his or her mental picture and relate information from the screen to internalized representations, before selecting the appropriate speed or height resulting in longer average fixation durations. Secondly, the amount of clusters shows a difference between controlling & planning and controlling & monitoring. Fixations belong to the same cluster when they are close to each other spatially and are in temporal order [7]. A higher amount of clusters indicate more areas needing attention. The results show that both planning and monitoring result in more fixation clusters than controlling. Third, although controlling yields less clusters, the average *fixation duration* in the clusters while controlling is longer compared to planning and monitoring. Fourth, the *cluster size* is bigger for controlling compared to planning and monitoring. The longer duration and larger size is possibly explained by the airplane attribute selection. The pull down menus appears near to the controlled aircraft resulting in a larger area containing information and thus a bigger fixations region. Possibly, the fixations in the cluster have a longer duration for the same reason, relating the internal representation of the traffic sample to the information on the screen.

The results from this study indicated that eye movements could be effective in differentiating subtasks and encourage detailed study of these metrics and the possibilities they create for adaptive or task dependent automation.

ACKNOWLEDGMENTS

We are grateful for the support we received through the Mediated Presence activity that is sponsored by the knowledge and innovation community EIT-ICT Labs with its mission is to drive leadership in innovation in ICT for economic growth and quality of life.

REFERENCES

1. Endsley, M., Kiris, E.: The Out-of-the-Loop Performance Problem and Level of Control in Automation. Hum. Factors. 381–394 (1995).
2. Flugunfalluntersuchung, B. für: Investigation Report AX001-1-2/02. , Braunschweig (2004).
3. Eurocontrol: Long-Term Forecast: IFR Flight Movements 2010-2030. , Brussels (2010).
4. Yarbus, A.L.: Eye Movements During Fixation on Stationary Objects. Eye Movements and Vision. pp. 103–127. Plenum Press (1967).
5. Land, M.F., Tatler, B.W.: Looking and Acting. Oxford University Press, Oxford (2009).
6. Goldberg, J.H., Kotval, X.P.: Eye movement-based evaluation of the computer interface. In: Kumar, S. (ed.) Advances in occupational ergonomics and safety. pp. 529–532. ISO press, Amsterdam, the Netherlands (1998).
7. Goldberg, J.H., Kotval, X.P.: Computer interface evaluation using eye movements: methods and constructs, (1999).
8. Poole, A., Ball, L.J.: Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects. In: Ghaoui, C. (ed.) Encyclopedia of Human-Computer Interaction. pp. 211–219. Idea Group (2005).
9. Di Nocera, F., Terenzi, M., Camilli, M.: Another look at scanpath: Distance to nearest neighbour as a measure of mental workload. Shaker Publishing. (2006).
10. Kantowitz, B.H.: Defining and Measuring pilot Mental Workload. In: Comstock, J. (ed.) Mental-State Estimation. pp. 179–188. National Aeronautics and Apace Administration, Scientific and Technical Information Division, Hampton, VA (1988).
11. Van Orden, K.F., Limbert, W., makeig, S., Jung, T.: Eye activity correlates of workload during a visuospatial memory task. Hum. Factors. 43, 111–121 (2001).
12. Verwey, W.B., Veltman, H.A.: Detecting short periods of elevated workload: A comparison of nine workload assessment techniques., (1996).
13. Ryu, K., Myung, R.: Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic, (2005).
14. Zijlstra, F.R.H.: Efficiency in work behaviour, (1993).
15. Rantanen, E.M., Goldberg, J.H.: The effect of mental workload on the visual field size and shape. Ergonomics. 42, 816–834 (1999).