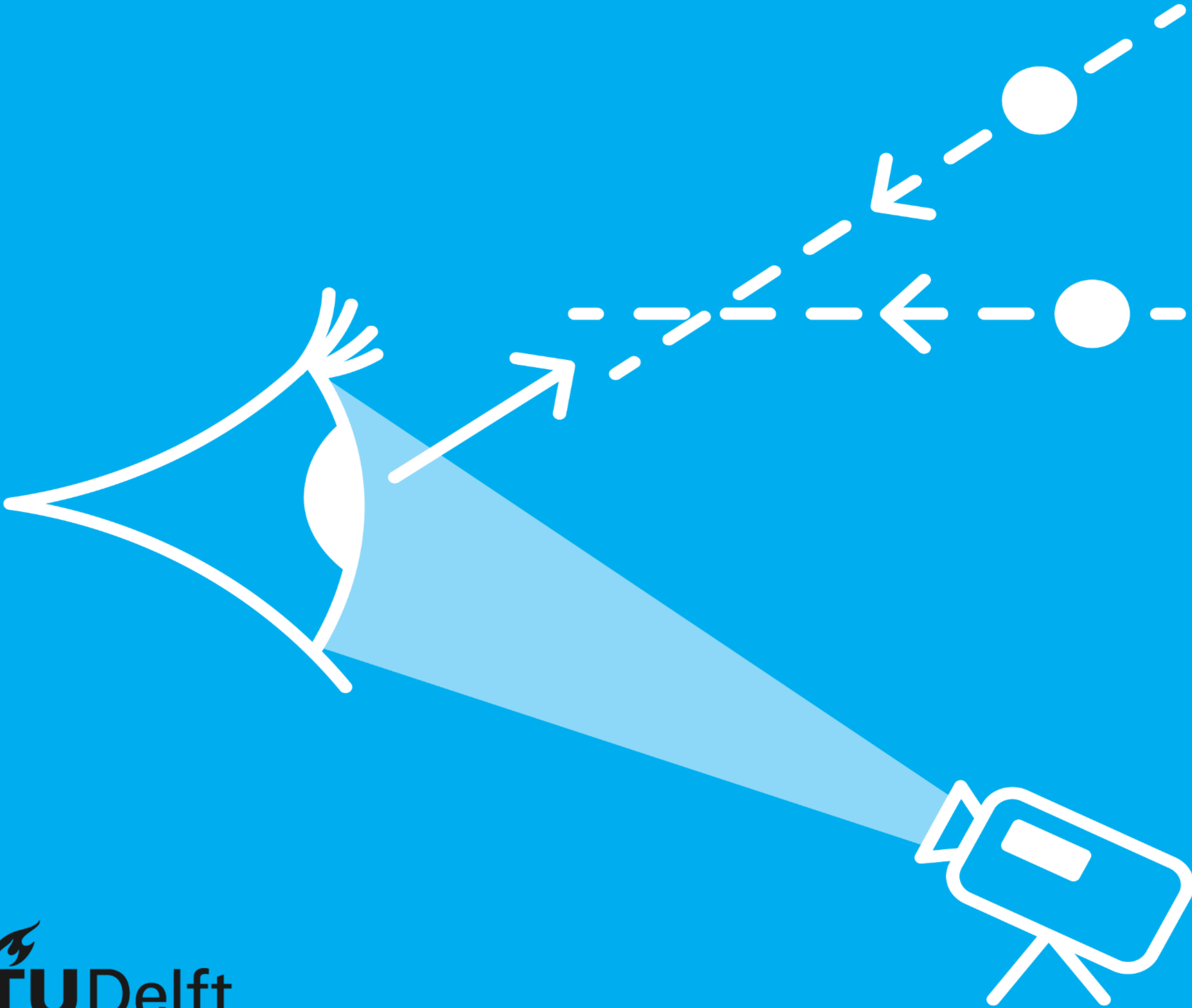
Sampling behavior while detecting conflicts between linear moving stimuli

Anouk Looijestijn

MSc Thesis



TUDelft

Sampling behavior while detecting conflicts between linear moving stimuli

by

Anouk Looijestijn

in partial fulfilment of the requirements for the degree of

Master of Science
in Mechanical Engineering

at the Delft University of Technology
to be defended publicly on Wednesday October 17, 2018 at 13.00 PM

Student ID:	4222938	
Thesis committee:	Dr. ir. J.C.F. de Winter,	TU Delft, faculty of 3mE, supervisor
	Ir. Y.B. Eisma,	TU Delft, faculty of L&R, supervisor
	Dr. D. Dodou,	TU Delft, faculty of 3mE

Contents:

Sampling behavior while detecting conflicts between linear moving stimuli

A.E. Looijestijn, Y.B. Eisma and, J.C.F. de Winter

Abstract

Introduction. Air traffic controllers (ATCo's) are responsible for a safe and efficient air traffic flow, and therefore, they are required to be excellent in conflict detection. Various studies have uncovered relationships between conflict geometry (e.g., conflict angle) and operators' abilities to detect conflicts. However, little is known about the underlying perceptive and cognitive processes during a conflict detection task. Knowledge of these processes could give insight into how ATCo's could be supported. Examples may be the change of the flight radar's design and the adjustment of the education program. In order to discover how people look at typical air traffic control (ATC) situations, a simplified ATC scenario was presented to novice participants in which two dots (representing aircraft) moved towards each other.

Methods. The eye movements of 35 participants were recorded during an experiment in which they had to indicate whether a conflict was present or not by pressing the spacebar. Each participant watched 36 different videos with a duration of 20 seconds and different air traffic geometries. The independent variables were: (1) conflict angle between the two approaching dots (30, 100, 150 deg), (2) configuration (one of the dots moved horizontal, diagonal, or vertical), (3) closest distance of approach (collision, or no collision), and (4) discrete vs continuous moving stimuli (2 Hz or 30 Hz). The effects of these variables on conflict detection performance, self-experienced difficulty, and sampling behavior were investigated.

Results. The results show that continuous moving trials obtained a significantly higher performance score, and more and shorter fixations were found compared to discrete moving trials. No significant difference in self-experienced difficulty was found between continuous and discrete trials. Furthermore, in accordance with Neal and Kwantes (2009) and Remington et al. (2000), participants performed significantly better with a conflict angle of 30 degrees compared to larger angles. Also, a lower difficulty score was reported with a conflict angle of 30 degrees compared to larger angles. However, it was found that the performance score with a conflict angle of 100 degrees was lower and the self-experienced difficulty was higher, compared to 150 degrees. The eye movement variables showed a monotonic relation with the conflict angle: when the number of fixations increased, the fixation duration decreased with the higher conflict angles. Furthermore, participants sampled more often from one dot to the other and exhibited less pursuit movement with increasing conflict angles. Moreover, it was found that conflict detection was significantly easier when one of the dots moved diagonally. Also, significantly more pursuit movements were used when one of the dots moved vertically, compared to one of the dots moving horizontally. Finally, the results show that for trials in which no conflict occurred, participants exhibited more fixations and sampled more from one dot to the other, compared to trials with a conflict. No significant difference in self-experienced difficulty was found between trials with or without a conflict.

Conclusions. We conclude that novice participants are better at detecting conflicts with continuous moving stimuli compared to discretely moving stimuli. Eye movements indicate that participants are able to take up more information about the movements of the dots in continuous situations. If further research shows the same increase in performance with continuous motion in real ATC, flight radars could be adjusted accordingly. Also, it is concluded that the conflict angle has influence on eye movements. Also, indications are found that conflict angles close to 0 and 180 degrees are easier for detecting conflicts. Further research with various conflict angles is recommended. Furthermore, while in general it is concluded that configuration has not much influence on sampling behavior, indications are found that diagonal movements might be easier for conflict detection. Moreover, we conclude that in our experiment pursuit movements are preferred with vertical movements compared to horizontal movement. Finally, we conclude that participants sample from one dot to the other when the dots are further away from each other, but when they come closer to each other, pursuit movement is often used to follow both dots at the same time.

Keywords
Collision detection, Eye movements, Eye tracking, Visual attention

1. Introduction

The use of automation in modern technical systems becomes ever more prevalent (Hancock, 2014). As a consequence, humans often need to perform the task of supervising automated systems (Parasuraman, Sheridan, and Wickens, 2000). One example of such a supervisory control task is Air Traffic Control (ATC) (Sheridan, 1984). An air traffic controller (ATCo) has to monitor the flight radar to “expedite and maintain a safe and orderly flow of traffic” (Oprins, Burggraaff, and Weerdenburg, 2006, p. 297) as well as to maintain separation standards (5 NM lateral, 1000 ft vertical). To accomplish these goals, ATCo’s have to be excellent at detecting conflicts between aircraft.

Influence of air traffic geometry on conflict detection

Various studies have investigated the relationships between conflict detection performance and a variety of archetypal ATC traffic configurations (Boag et al., 2006; Eyferth, Niessen, and Spaeth, 2003; Thomas and Wickens, 2006). Remington et al. (2000) showed in an experiment in which 4 experienced participants were asked to find a conflicting pair from multiple (12 to 20) aircraft, that with small angles (< 90 degrees), the response time (i.e., the time from the beginning of the trial until first response) was faster, in comparison with the large angles (> 90 degrees). They argued that conflicts are easier to detect with small angles because the conflicting aircraft (and their symbols and flight labels) are closer to each other on the display. Mackintosh et al. (1998) confirmed that acute conflict angles were easier to detect than obtuse angles. Moreover, Loft et al. (2009) found that for both lateral conflict and non-conflict situations, participants were more likely to intervene with smaller conflict angles compared to larger angles (45^o, 90^o, and 135^o were used). Loft et al. (2009) argued that with smaller angles, the overlapping area of the trajectory estimation is larger and therefore there is more uncertainty, increasing the likelihood of a conflict. Finally, Neal and Kwantes (2009) demonstrated in an experiment with three different conflict angles (45^o, 90^o, and 135^o), that smaller angles of intersection between two conflicting airplanes increased the probability of correct conflict detection and increased the response time for conflict scenarios compared to larger angles. However, they found the opposite effect for situations without a conflict; there the response time increased with decreasing angles.

Eye tracking to obtain insight in the processes underlying the conflict detection task

Neal and Kwantes (2009) explain their results with the distance to velocity ratio strategy (Xu and Rantanen, 2003), which implies that people use the difference in arrival time of two aircraft at their conflict point to estimate whether they will be in conflict or not. As another possible strategy for detecting lateral conflicts, Xu and Rantanen (2003) mention the cognitive motion extrapolation strategy (DeLucia and Liddell, 1998), in which participants are expected to extrapolate the trajectory of an object. However, still little is known about the exact perceptive processes that underlie conflict detection. Eye tracking could be used to discover more about the relationships between visual stimulus structure and operators’ perceptive and cognitive processes. It could help, for example, to explain why larger conflict angles are experienced to be more difficult. Eye tracking is becoming more and more popular to gain insight into visual sampling strategies, partly because contemporary eye-tracking devices are able to measure eye movements more accurate than ever before (Duerrschmid and Danner, 2018).

Eye tracking in air traffic control studies

Only in a few ATC studies, eye-tracking data were recorded. Manske and Schier (2015) performed a study with six experienced ATCo’s who had to give aircraft arrival and landing clearances. It was found that with higher task demands slightly more visual scans were conducted by an operator than with lower task demands. In the low task demands situations, the clearance decision was independent of other traffic while in the high task demands situations arriving and departing traffic ahead had to be taken into account. Furthermore, with a simplified ATC simulation task, a longer scan path length (summed distance between all fixations) and a higher spatial density (indicating the spreading of fixations) were found when more aircraft were added to the flight scenario (Imants and de Greef, 2014). These results could be expected since more aircraft require more fixations to different locations to get an overview of the situation. Furthermore, Wang et al. (2015) found that novices glanced to a smaller number of areas of interest (AOIs), exhibited a higher fixation duration and higher saccadic velocities, compared to experienced ATCo’s during the same ATC simulation task. Moreover, Hunter, and Parush (2009) recorded eye movements of participants who looked at a static simulation of two aircraft with different velocities on a converging trajectory. They conclude that participants were more likely to scan between the two aircraft than towards the collision site.

Knowledge gap

To the best of our knowledge, no eye-tracking studies have been performed that quantify the relationships between eye movements and dynamic conflict detection. There are practical and theoretical reasons why more insight into those relationships may be beneficial. Firstly, knowledge of sampling behavior during conflict detection could be used in practice to make the ATC task more efficient, for example by optimizing flight displays such that performance increases. Also, the training trajectory of new ATCo's could be adjusted and subtasks could be replaced or altered so that the operator can do his/her job better. Secondly, insight into the relationships between eye movements and dynamic conflict detection improves general understanding of how humans perceive linear moving stimuli during a supervisory control task. Different models are developed that predict where humans will sample during supervisory control (see Moray, 1986, for an overview). Wickens et al. (2003) recently extended the existing theories by developing a descriptive model that gives the probability that a certain AOI will be sampled based on salience, effort, expectancy, and value (SEEV). Increasing the understanding of how humans sample linear moving stimuli, could be a first step to further refine these models.

Aim of the study

For the aforementioned reasons, a dynamic conflict detection study will be performed, in which eye movements will be tracked. The aim of this study is to investigate in an empirical way the effects of different air traffic geometries on (a) conflict detection performance and (b) eye movements during a dynamic conflict detection task, such that the perceptual processes and underlying cognitive processes during conflict detection may be examined.

Hypotheses

We expected that a smaller conflict angle (CA) would be considered easier and would lead to improved conflict detection results compared to larger angles (as in Neal and Kwantes, 2009; Remington et al., 2000). Furthermore, since the effort to sample between two aircraft is smaller when the CA is smaller, we expected that participants would make more comparisons between the aircraft for smaller CAs (SEEV model, Wickens et al., 2003). Moreover, although current ATC systems are discrete, we expected that continuous movements would be experienced to be easier, get a higher performance score, and an earlier response. Braddick (1974) argues that humans use different visual mechanics to analyze continuous (short range) or discrete (long range) motion. Both types of motion analyses require different physiological processes (Hildreth and Ullman, 1982). During discrete movement, the previous location of the object has to be remembered, in order to estimate the trajectory and velocity. Therefore we expected that it is more difficult to estimate the relative position of the dots when they move discretely compared to continuous.

2. Method

2.1 Participants

19 males and 16 females between 18 and 31 years old (M = 22.8, SD = 2.91) participated in this research. They were all students at the TU Delft or recently graduated from the TU Delft. The data of one participant were excluded since he/she did not perform the task as instructed. This research was approved by the Human Research Ethics Committee of the TU Delft on 03-04-2018 with the title: "Sampling behavior when observing moving stimuli with a converging heading". A written informed consent form was signed by all participants before the start of the experiment. The majority of the participants were given a financial compensation of five euros to compensate for their time.

2.2 Experimental task

Participants were asked to watch a total of 36 movies with a duration of 20 seconds each, in which they were presented with two dots (representing aircraft) moving towards each other in a linear way (see Figure 1). Participants were instructed to press the spacebar (and keep it pressed) when they thought the dots would collide in the future. Participants were allowed to change their mind and consequently release the spacebar. In real-life air traffic control, aircraft have to be separated at least five nautical miles (NM) from each other (Rantanen and Nunes, 2005). However, in this experiment, no 5-NM zone existed around the dots. A collision was defined as two dots overlapping. After each trial, the participants were asked to indicate to what extent they agreed with the following statement: *'The task was difficult'* on a scale from 0 to 10. Subsequently, their performance score from the previous trial was displayed. The performance score was computed as the percentage of time that the spacebar was correctly pressed or released (depending on whether the trial contained a conflict or not). Before the start of the experiment, a nine points calibration and validation of the eye tracker were performed and one discrete trial, with a different configuration and conflict angle than the ones in the experiment, was done as training. In this training trial, a collision was presented, in order to check that all participants understood the task. A break of a

Figure 1: Screenshots of one stimulus at four moments in time. The dots move towards the middle of the screen.

few minutes was held halfway through the experiment. In total the experiment took about 30 minutes per participant.

2.3 Eye tracking

During the experiment, the eye movements of the right eye were measured with the SR Research Eyelink 1000 Plus eye tracker with a sampling rate of 2000 Hz. Participants were asked to put their head on a head support, that was accordingly adjusted to the length of each participant so that they could sit comfortably. Figure 2 shows the test setup. The participants were informed to keep their head in the support, except during the break. If needed, an extra break could be added at any time. The 24-inch screen with a resolution of 1920 x 1080 pixels (531 x 298 mm) was positioned 95 cm from the head support, and the eye-tracking camera/IR light source was located 65 cm from the head support, as indicated by the eye-tracker manual (SMI cooperation, 2014). The refresh rate of the screen was set at 60 Hz.

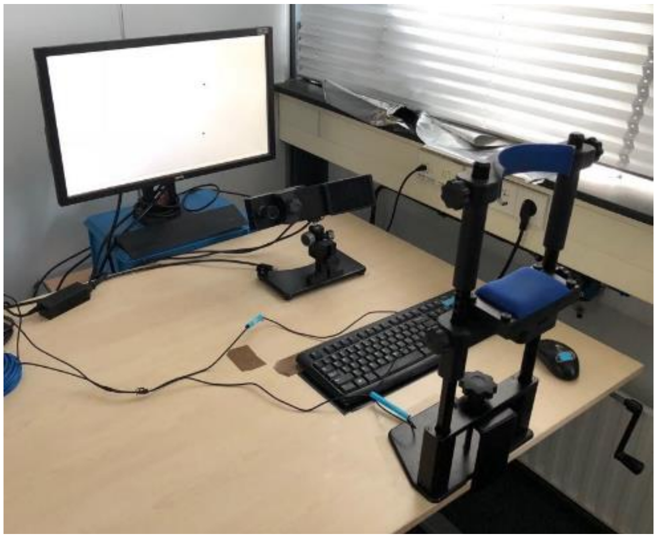


Figure 2: Test setup: on the left the computer screen which shows the stimuli, in the middle the eye tracker, on the right the head support.

2.4 Independent variables

The experiment featured four independent variables.

1. The **distance at the closest point of approach** is the distance between the two dots at the moment that the dots are closest to each other. This distance was varied to differentiate between situations in which a collision occurred or situations in which no collision occurred, in order to make the task unpredictable for the participants. Half of the trials had a distance at the closest point of approach of zero, which resulted in a collision, and in the other half, the distance at the closest point of approach was set to be 115 pixels.
2. The spatial **configuration** of the first dot was set to be horizontal, vertical, or diagonal.
3. The **conflict angle** between the two dots was varied. The configuration determined the starting position of the first dot, and with the conflict angle, the initial position of the second dot was determined. Conflict angles are categorized into three categories: 0^o-60^o (overtake), 60^o-120^o (crossing), and 120^o- 180^o (head-on) (Thomas and Wickens, 2006). For this experiment, one angle from all three categories was used as CA; namely 30^o, 100^o, and 150^o. In this study, an angle of 90^o resulted in similar situations when combined with different configurations, therefore 100^o was used instead.
4. Whether the signal was **discrete or continuous** was also varied. For the continuous option, the frame rate was set to 30 frames per second. The discrete option updated the location of the dots two times per second.

So, a total of 2x3x3x2 variables were varied throughout this experiment, resulting in a total of 36 different trials. Table 1 gives an overview of the variables per trial. The sequence of these trials was randomized for each participant in order to minimize the effect of learning and fatigue.

2.5 Design of the stimuli

The WriteVideo function in MATLAB was used to create all 36 videos. A white screen of 1920x960 was created (RBG: [0.9 0.9 0.9]), on which two round markers with size 14 (i.e. approximately 18 pixels) are projected (RBG: [0.1 0.1 0.1]). Dot 1 always started 960 pixels from the center of the screen at an angle to the x-axis of 0, 45, or 90 degrees, depending on the configuration (α). Dot 1 moved through the middle and ended 96 pixels from the mid-point. The speed of both dots was set at 52.8 pixels/second (1065 pixels in 20 seconds). The speed of both dots remained the same during the entire experiment. The direction of dot 2 was determined by the CA and with a random sign (β) relative to dot 1. β defined whether the angle of dot 2 would become larger or smaller than dot 1. To keep the stimuli unpredictable, the sign was defined randomly, but the same for the continuous and discrete stimuli: β = [-1 -1 -1 1 -1 -1 1 1 1 -1 1 1 -1 1 -1 -1 1]. The velocity vectors of the dots are computed with equations 1 and 2.

$$V1 = speed \bullet \begin{bmatrix} cos(\alpha) \\ sin(\alpha) \end{bmatrix} \tag{1}$$

$$V2 = speed \bullet \begin{bmatrix} cos(\alpha - (\beta * AOC)) \\ sin(\alpha - (\beta * AOC)) \end{bmatrix} \tag{2}$$

The relative velocity and the time to closest point of approach (t_{cpa} = 960 pixels/ speed) were used to compute the relative distance (d_{rel}) (Ellerbroek, 2013, Equation A.2):

$$d_{rel} = t_{cpa} \bullet |V1 - V2| + \sqrt{{R_{pz}}^2 - {d_{CPA}}^2} \qquad (3)$$

R_{pz} represents the radius of the protected zone, which was set as the radius of the dot. The relative distance vector (x_{rel}) of dot 2 with respect to dot 1 was computed with equation 4 (Ellerbroek, 2013, Equation A.3), in which V_{rel} is the relative velocity between the dots.

$$x_{rel} = \begin{bmatrix} d_{rel} & d_{CPA} \\ -d_{CPA} & d_{rel} \end{bmatrix} \bullet \frac{Vrel}{|Vrel|} \qquad (4)$$

The start and end position of the dots was used to update their position at each time step.

2.6 Dependent variables

The dependent variables were defined as follows:

- *Performance score (%)*

The performance score was computed as the percentage of trial duration that the participant had the spacebar correctly pressed or not-pressed. It should be noted that the performance score for conflict and non-conflict cannot be compared since each trial started without the spacebar being pressed. It was therefore easier to obtain a higher score in the non-conflict trials.

- *Self-experienced difficulty score*

A score between 0 and 10 was given by the participants after each trial to obtain the subjective difficulty. The score was averaged per independent variable.

- *Spacebar presses against time (%)*

For all scenarios that contained a conflict, the time it took to press the spacebar (and keep it pressed) was listed for all scenarios and participants per independent variable. If the spacebar was pressed and later released again, this press was not considered. In this way, the period in which the participant was doubting was not taken into account.

- *Number of false positives*

For all scenarios that did not contain a conflict, the number of false positives is counted as the number of trials in which the spacebar was pressed.

- *Total number of fixations per trial*

Fixations give an indication of the information uptake of a participant (Kilingaru et al., 2013). For example, a higher fixation rate could mean that the task is experienced to be more difficult (Sharma et al., 2016). Therefore, the total number of fixations per independent variable was computed. A median filter with a 100 ms interval was used to smooth the raw eye-tracking data. When there was no data available, for example during a blink, a linear interpolation was used. Fixations were defined in the same way as in Eisma, Cabrall, and De Winter (2017). The gaze speed was filtered with a Savitzky-Golay filter with order 2 and a frame length of 41. Fixations were only counted if they lasted longer than 40 ms.

- *Averaged fixation duration (s)*

The fixation time was computed from the end time of a previous saccade and the begin time of the next saccade. The average duration per fixation was computed in seconds per independent variable. A longer fixation duration could imply that more time was needed to process the information (Backs and Walrath, 1992).

- *Percentage time on AOIs (%)*

Percentage time on AOIs was defined as the percentage of time that the gaze was on one of the AOIs. The percentage time on AOIs was not influenced by how a fixation is defined. If the gaze was 30 pixels or less from one or both of the

Table 1: Overview of the 36 different trials.

Trial no	Conflict	CA (degrees)	Configuration	Disc/Cont
1	yes	30	horizontal	continuous
2	yes	30	diagonal	continuous
3	yes	30	vertical	continuous
4	no	30	horizontal	continuous
5	no	30	diagonal	continuous
6	no	30	vertical	continuous
7	yes	100	horizontal	continuous
8	yes	100	diagonal	continuous
9	yes	100	vertical	continuous
10	no	100	horizontal	continuous
11	no	100	diagonal	continuous
12	no	100	vertical	continuous
13	yes	150	horizontal	continuous
14	yes	150	diagonal	continuous
15	yes	150	vertical	continuous
16	no	150	horizontal	continuous
17	no	150	diagonal	continuous
18	no	150	vertical	continuous
19	yes	30	horizontal	discrete
20	yes	30	diagonal	discrete
21	yes	30	vertical	discrete
22	no	30	horizontal	discrete
23	no	30	diagonal	discrete
24	no	30	vertical	discrete
25	yes	100	horizontal	discrete
26	yes	100	diagonal	discrete
27	yes	100	vertical	discrete
28	no	100	horizontal	discrete
29	no	100	diagonal	discrete
30	no	100	vertical	discrete
31	yes	150	horizontal	discrete
32	yes	150	diagonal	discrete
33	yes	150	vertical	discrete
34	no	150	horizontal	discrete
35	no	150	diagonal	discrete
36	no	150	vertical	discrete

trajectory lines, the corresponding time sample was considered to be on an AOI as can been seen in Figure 3.

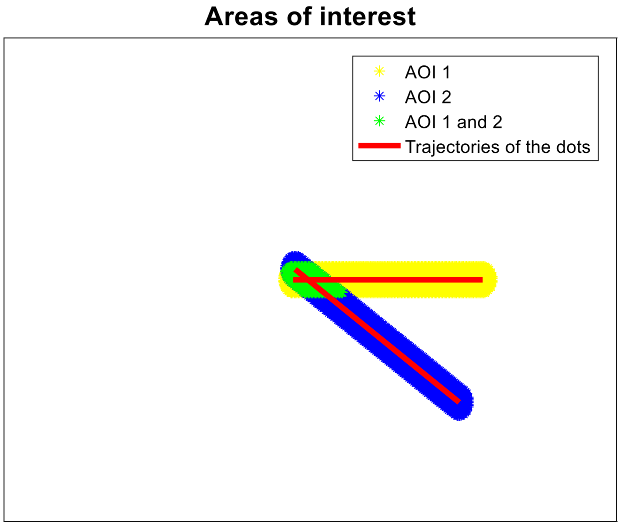


Figure 3: The trajectories of the dots are shown in red with the corresponding AOIs of 30 pixels around the trajectories. The green area shows were both AOIs are overlapping.

- *Number of comparisons between the dots*

The number of comparisons between the dots was defined as how often participants sample from one dot to the other. The previous fixation had to be on a different AOI than the current fixation.

- *Percentage of fixations that contain pursuit movement (%)*

Smooth pursuit is defined by the ISO norm (ISO 15007-1, 2014, p3) as: " smooth, continuous movement of the eyes made to closely follow/pursue a moving object of signal". By plotting all fixations in the x and y plane, it became apparent that some of the intervals that were defined as fixations by the algorithm did also contain pursuit movement. For each fixation, the range (travelled distance in x- and y-direction) was computed. If this range was larger than 30 pixels, the fixation contains pursuit. The threshold of 30 pixels was obtained by visualizing data from 9 trials (3 different participants and 3 different angles). The percentage of all fixations that did contain pursuit per independent variable was listed. Before the percentage was computed, an extra fixation was added after the last saccade, such that the time period after the last saccade was considered as a fixation. The reason for this is that from the pilot study, it became apparent that for some trials (especially with a CA of 30 degrees), the last saccade occurred early in the trial, for example after 12 seconds. The standard algorithm does not take the time after the last saccade into account. However, to get a better insight into the pursuit movements, it was decided to add a fixation in this period. Adding this fixation had an influence on the percentage of fixations that contain pursuit and allowed to draw conclusions about what happened in the period after the last saccade.

- *Duration of fixations that contain pursuit movement (s)*

The average fixation duration of all fixations that contain pursuit was computed per independent variable. A larger fixation duration could indicate more pursuit movement per fixation. Again, the time period after the last saccade was considered as a fixation.

2.7 Data analyses

MATLAB was used to analyse the data and to visualize the results. Afterwards SPSS was used to perform a repeated-measures ANOVA on each dependent variable. Results with p values smaller than 0.05 were considered significant. Post hoc paired comparisons T-tests with a Bonferroni adjustment were done on the results of conflict angle and configuration results when the ANOVA showed a significant result.

3. Results

3.1 Performance and difficulty scores

Performance score

Figure 4 shows the averaged performance and difficulty scores. The colors indicate the different independent variables. Participants perform better on the continuous moving stimuli compared to the discrete moving stimuli ($F(1,34) = 4.592$, $p = 0.039$). Furthermore a significant difference is found for CA ($F(2,68) = 25.869$, $p < 0.001$). Post hoc analyses revealed that the performance on CA30 was significantly higher than both CA100 ($p < 0.001$) and CA150 ($p = 0.005$). Furthermore, participants have a significantly higher performance score on CA150 compared to CA100 ($p = 0.005$). The effects of configuration on performance score (horizontal/vertical or diagonal) is also significant ($F(2,68) = 5.251$, $p = 0.008$). Post hoc tests revealed that the diagonal configuration resulted in a significantly higher performance score than the horizontal configuration ($p = 0.017$).

Self-reported difficulty

Out of the four independent variables, only CA shows a statistically significant effect of self-reported difficulty ($F(2,68) = 30.095$, $p < 0.001$). Post hoc analyses revealed that participants report the difficulty of CA30 significantly lower than CA100 ($p < 0.001$) and CA150 ($p = 0.005$). Furthermore, CA150 has a significantly lower self-reported difficulty compared to CA100 ($p < 0.001$). Figure 5 shows the performance scores per conflict angle for conflict vs no conflict and continuous vs discrete trials.

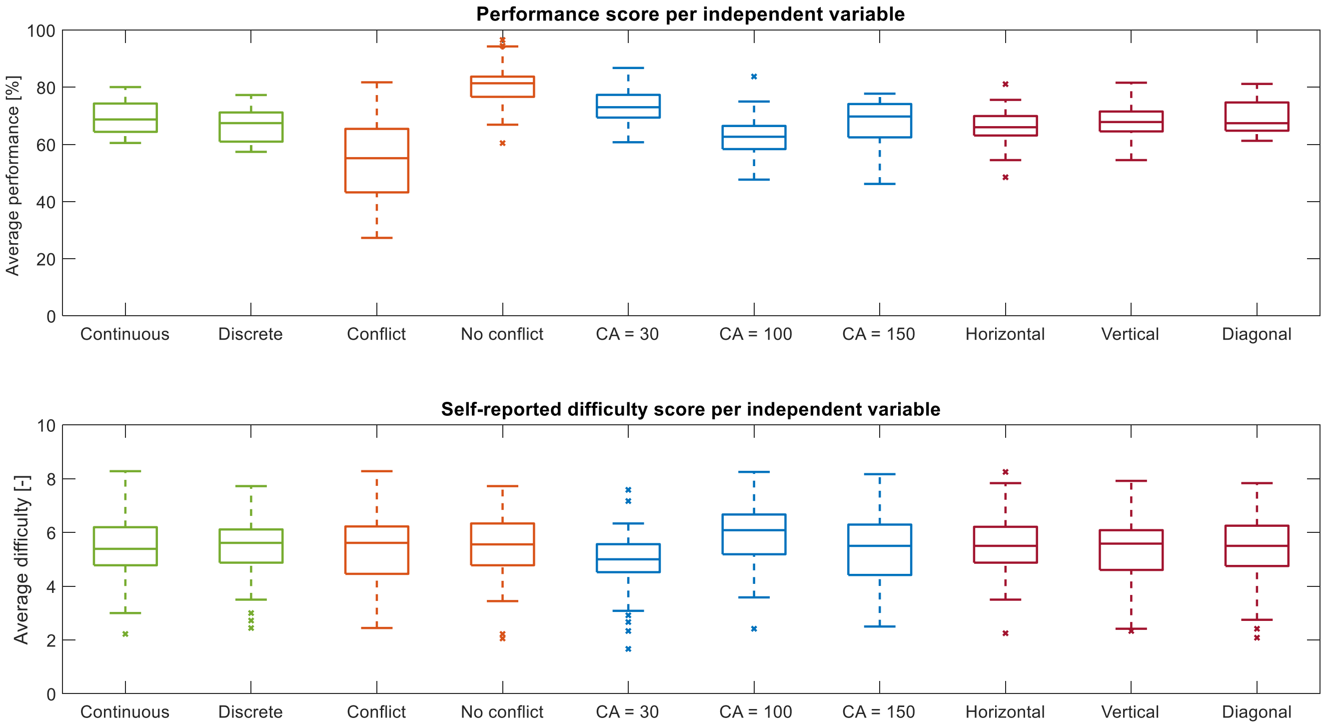


Figure 4: Performance and difficulty scores per independent variable. Each plot is created from 35 data points, which are the average values for each participant per independent variable.

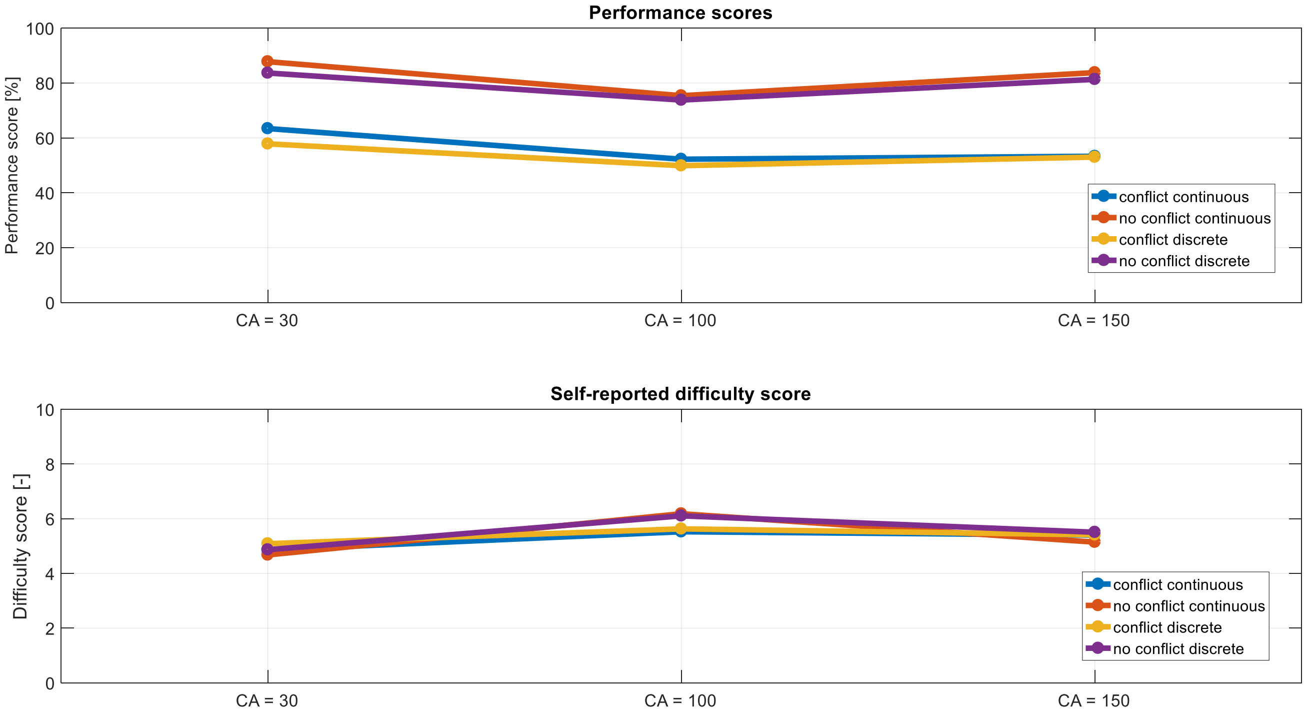


Figure 5: Performance and self-reported difficulty scores for the different conflict angles, separated for conflict/non-conflict and continuous/discrete.

Percentage spacebar presses over time

Figure 6 shows the percentage of detected conflicts over time for the conflict trials. At the end of each trial the conflict/non-conflict was shown, so at the end 100 percent of the participants pressed the spacebar, as can be seen in the figure.

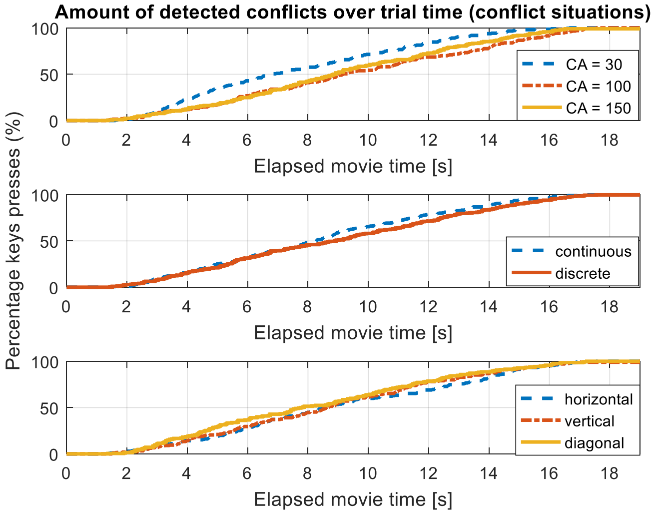


*Figure 6: Percentage of spacebar presses over time per independent variable. For each conflict angle and configuration 35 participants responded in 6 conflict scenarios per angle, therefore 35x6 = 210 data points are plotted. The continuous and discrete graph contains of 35*9 = 315 data points.*

Number of false positives per variable

From the spacebar press and release data, the number of false positives during non-conflict trials was determined and shown in Figure 7. A false positive is counted when a participant pressed the spacebar while no conflict existed in the specific trial. If the spacebar was pressed multiple times in one trial, it is counted as one spacebar press. The total number of false positives is found to be 334. That means that in about half of the trials without a conflict, the spacebar was pressed. The figure shows the division of those false positives over the trials with different conflict angles. These results are in line with the findings from the difficulty and performance scores. Continuous trials got a significantly higher performance score and here we observe a smaller number of false positives (162 vs 172). Furthermore, an angle of 100 degrees was perceived as more difficult and obtained a lower score, which is in accordance with the higher number of false positives found (152 for CA100, 106 for CA150, and 86 for CA30). The horizontal configuration got also more false positives than the other configurations (126 vs 104).

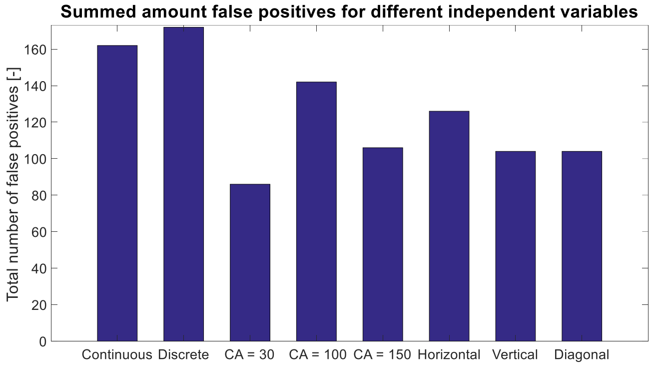


Figure 7: Amount of false positives per independent variable.

Learning effects

The average difficulty and performance scores for all participants per trial are shown in Figure 8. The trials are shown on the x-axis in the sequence as presented to the participants. From these graphs, the development in performance and self-experienced difficulty can be seen over time. Since the participants started all trials without the spacebar pressed, the performance scores for non-conflict trials are higher. Therefore, the performance scores of trials with and without conflict are plotted separately. A repeated measures ANOVA with trial sequence as independent variable has been performed to see whether the learning effects were significant. The performance score for non-conflict trials showed a significant difference over time (F(17,578) = 1.999, p = 0.010). The difficulty score and the

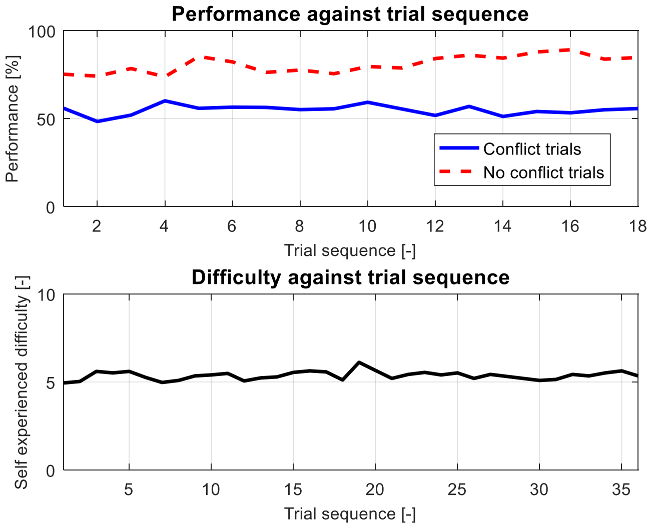


Figure 8: Learning effects are indicated by showing the performance and difficulty scores against trial sequence. For example, the self-reported difficulty score of all first trials are averaged and that value is shown with the black line at the first trial on the x-axis.

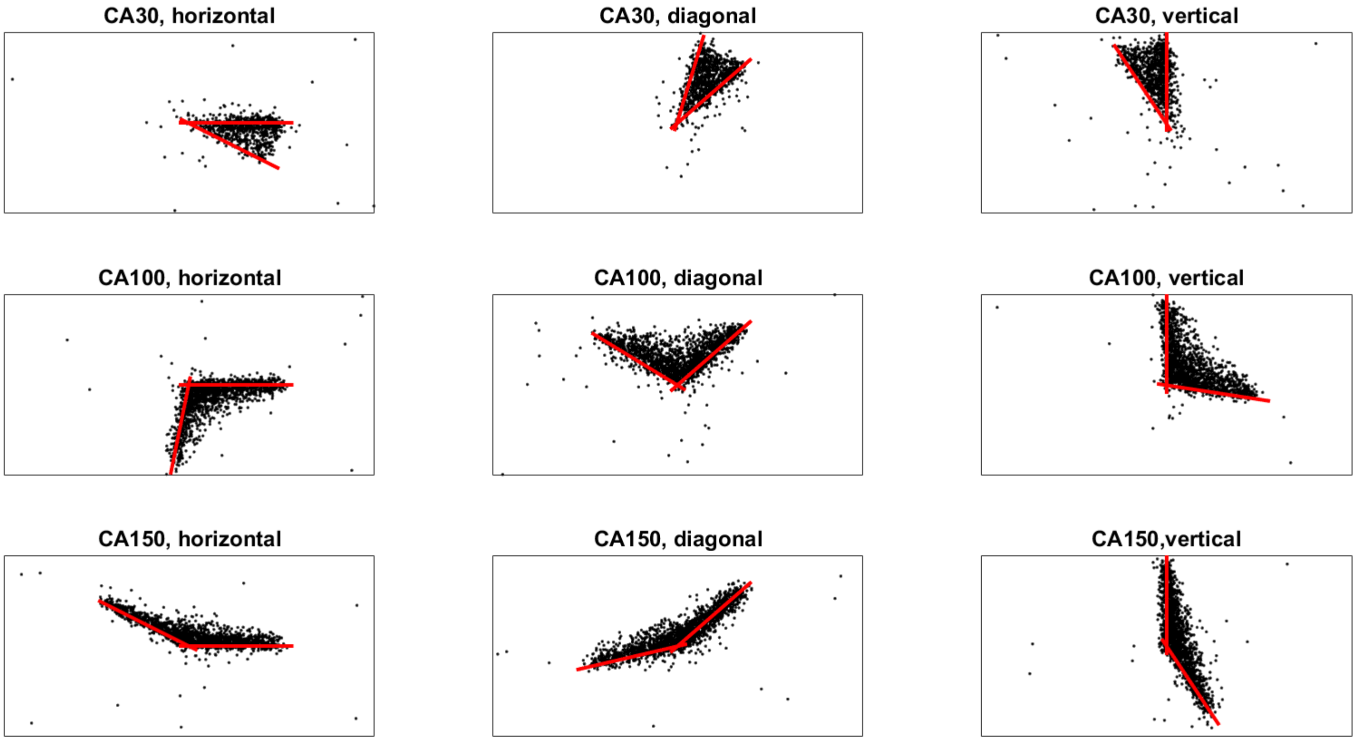


Figure 9: Fixation distribution of all participants per conflict angle and configuration. The fixations of the continuous and discrete trials are displayed on top of each other, so each subfigure shows the fixations of 70 (35 times 2) trials. The movements of the dots are indicated with the red lines.

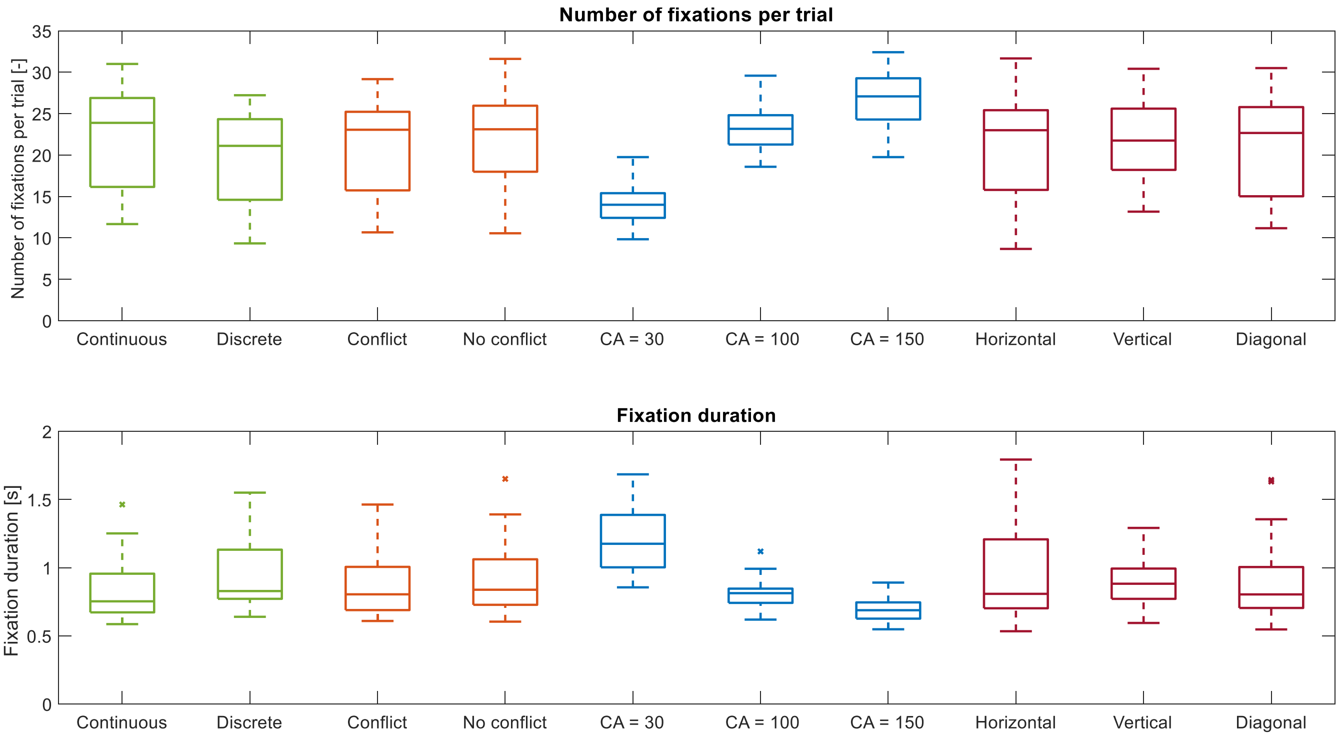


Figure 10: Total number of fixations per trial and the fixation duration, per independent variable. For each independent variable, the average for each participant has been used, so each boxplot is created from 35 data points.

performance score for conflict trials did not show a significant effect (p = 0.969 and p = 0.547, respectively).

3.2 Fixation placement

Figure 9 shows all the fixations of the trials with a conflict for the different conflict angles and configurations. The aggregated results of both continuous and discrete trials with identical conflict angle and configurations are shown. The red lines indicate the trajectories of the dots.

3.3 Fixation amount and duration

Figure 10 shows the average number of fixations during a trial for each independent variable. Also, the average fixation duration is shown.

Number of fixations

A significantly higher number of fixations is found in continuous moving stimuli in comparison to discrete stimuli (F(1,34)= 53.912, p < 0.001). Furthermore, the number of fixations in conflict trials is significantly lower compared to the non-conflict trials (F(1,34) = 44.046 p < 0.001). Moreover, a significant difference is found for CA (F(2,68) = 150.511 , p < 0.001). Post hoc tests show a significant difference between all levels of conflict angle (p < 0.001 for all three pairwise comparisons). The effect of configuration is also significant (F(2,68) = 5.457, p = 0.006). Post hoc analyses revealed that the vertical configuration has more fixations compared to the horizontal configuration (p = 0.022). No significant difference was found for the other combinations.

Fixation duration

The average fixation duration is found to be higher for discrete trials compared to the continuous trials (F(1,34) = 21.185, p < 0.001). Furthermore, the presence of a conflict has no significant effect on fixation duration (F(1,34) = 1.225, p = 0.276). The effect of CA to the fixation duration is significant (F(2,68) =70.014 p < 0.001). Pairwise comparisons revealed a significant decrease in fixation duration if the CA increases (p < 0.001 for all comparisons). Configuration does not have a significant effect on the fixation duration (p = 0.078).

3.4 Percentage time per AOI and number of comparisons between the AOIs

Figure 11 shows the percentage of time that the gaze was on one of the AOIs and the number of comparisons between the AOIs.

Percentage time per AOI

Participants spent significantly more time looking at the AOIs with continuous moving stimuli than for discrete moving stimuli (F(1,34) = 10.285, p = 0.003). Furthermore, non-conflict trials resulted in a higher percentage time on the AOIs compared to conflict trials (F(1,34) = 6.135, p = 0.018). A significant effect of conflict angle on AOI glance time is

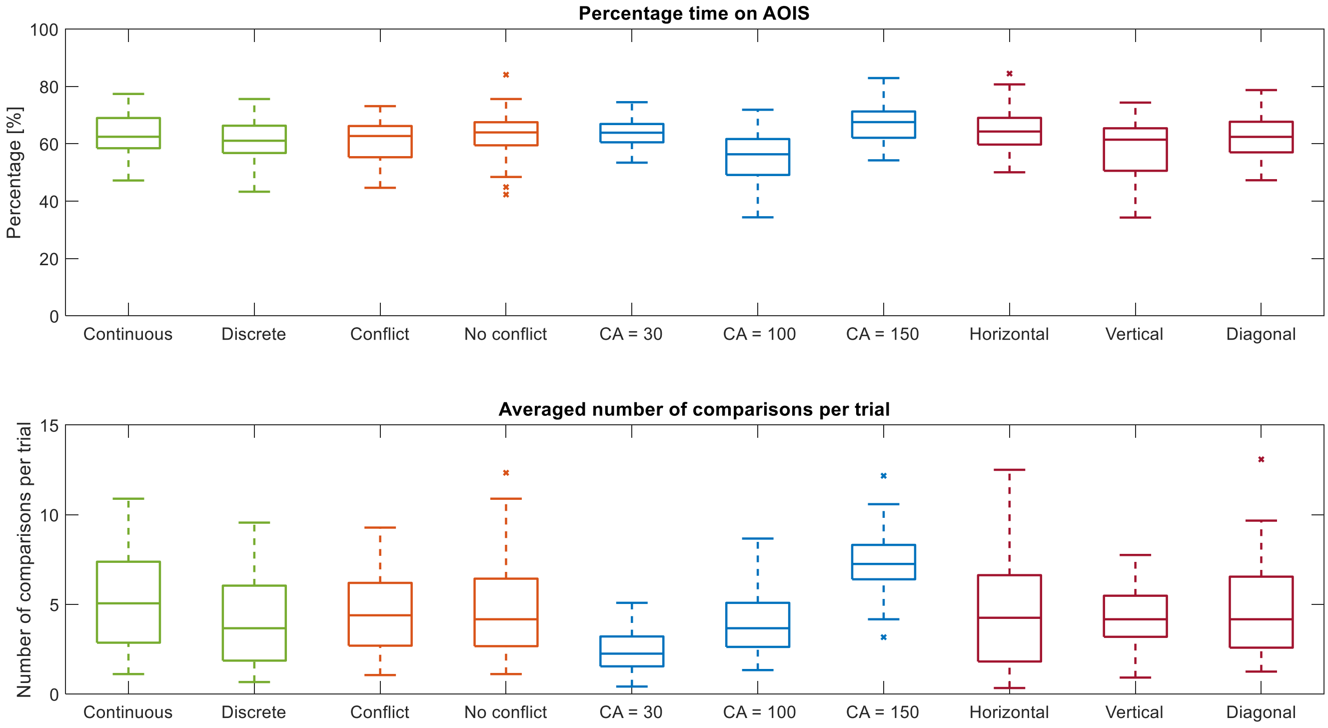


Figure 11: Percentage time that the eyes were on the AOIs and the number of comparisons between the AOIs per independent variable.

found (F(2,68) = 53.088 p < 0.001). Pairwise comparisons show that CA100 got significantly less time on AOIs compared to CA30 and CA150 (p < 0.001 for both). No significant difference between CA30 and CA150 was found. Configuration shows a significant effect (F(2,68) = 11.728, p < 0.001). Post hoc analyses show that the vertical configuration receives significantly less time on AOIs than the horizontal and diagonal configurations (p = 0.001 and p = 0.002 respectively). No significant difference between horizontal and diagonal configuration is found.

Number of comparisons

Participants perform significantly more comparisons between the dots in the continuous moving trials compared to the discrete moving trials (F(1,34) = 34.123, p < 0.001). No significant difference is found between conflict and non-conflict trials. Furthermore, there is a significant effect of CA on the number of comparisons (F(2,68) = 71.130, p < 0.001) and the post hoc analyses show an increasing number of comparisons for increasing conflict angles (p < 0.001 for all three paired comparisons). Configuration does not show a significant effect on the number of comparisons (F(2,68) = 1.944, p = 0.151).

Distribution of gaze samples

Figure 12 shows heat maps off the time all participants together spend per area. The light area in the middle of all plots shows the end of the trial when the conflict area was sampled. Furthermore, the plot shows that most gazes are centered around the trajectories of the dots. With a CA of 150 degrees there are not many gazes towards the area between the dots compared to the other angles.

Number of fixations per AOI

The number of fixations is computed per AOI to see whether there was a difference between the dot that was placed in the horizontal, vertical, or diagonal position and the other dot. A significant difference between the number of fixations per trial on AOI1 and AOI2 is found (F(1,34) = 213.29, p < 0.001). The total number of fixations to AOI1 was 8848 while AOI2 obtained a total of 5060 fixations. Figure 13 shows the number of fixations per AOI for each configuration. From the plot, it becomes clear that for each configuration the first dot (so the one that is actually in the configuration), obtained more fixations on average.

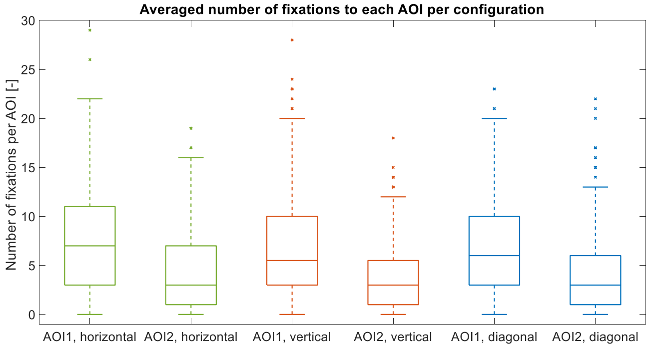


Figure 13: Number of fixations per AOI, per configuration. The different configurations are shown with different colors. The boxplots are based on 420 samples each (35pps x 12 trials).

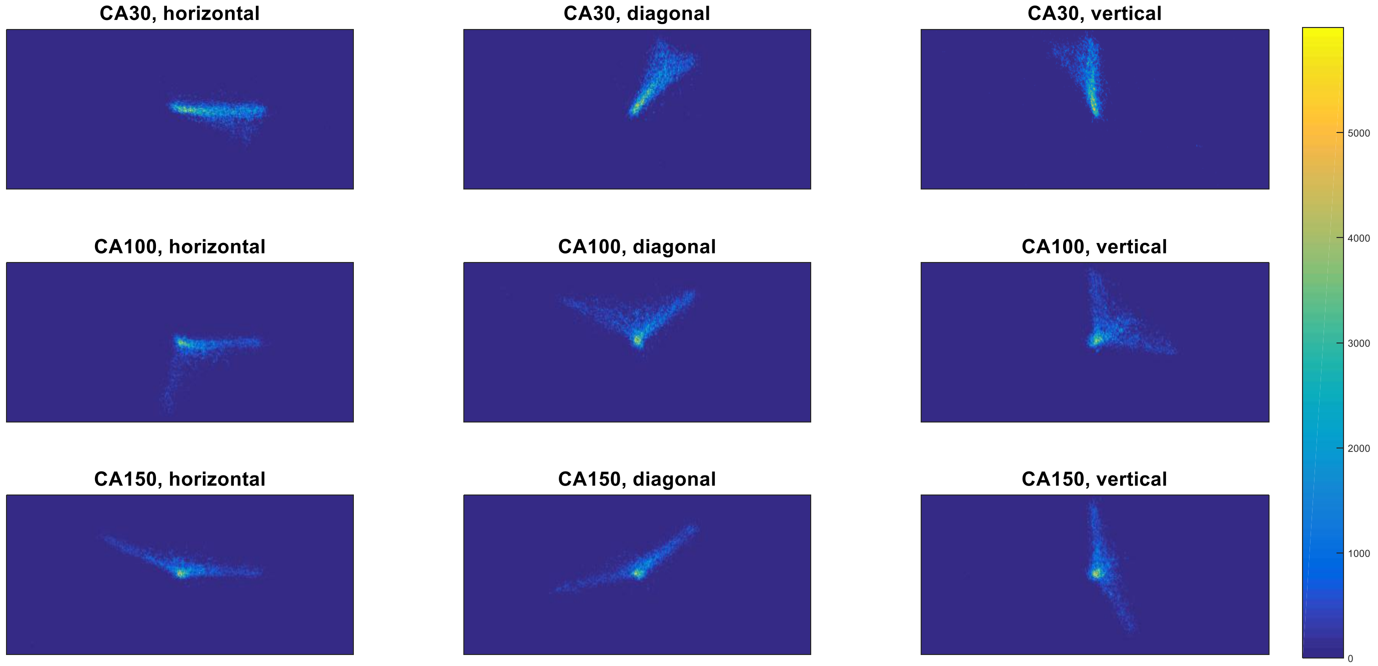


Figure 12: Heat maps of all gaze moments of all participants per conflict angle and configuration. All samples (2000 Hz) are divided in 455x256 bins. The color indicates how much samples each bin contains.

3.5 Pursuit movement

Pursuit fixations are a subset of the described fixations above (i.e., all fixations that are described above could possibly contain a pursuit movement). Pursuit fixations typically have a longer fixation duration and contain more movement (Holmqvist et al., 2011, chapter 5.8). If pursuit occurs, this means that another sampling strategy is used. Figure 14 shows the percentage of all fixations that contain pursuit movement and the average duration of the fixations that contain pursuit for each independent variable.

Percentage of fixations with pursuit

No significant difference in the percentage of fixations that contain pursuit is found between continuous and discrete (p = 0.925) and conflict and non-conflict trials (p = 0.890). For conflict angle, there is a significant effect (F(2,68) = 72.975, p < 0.001). Post hoc analyses show that the percentage of fixations with pursuit movement decreases significantly with CA (p < 0.001 for all paired comparisons). Furthermore, the effect of configuration is significant (F(2,68) = 3.876, p = 0.025). The vertical configuration has significantly more fixations with pursuit movement than the horizontal configuration (p = 0.050). The other paired comparisons do not show significant results for configuration.

Pursuit duration

The average duration of the fixations that contain pursuit is also shown in Figure 14. There were eight trials without pursuit (from five different participants and five different trials), so they are not considered in the analysis. The fixation duration for fixations that contain pursuit is found to be significantly higher for continuous moving stimuli compared to discrete moving stimuli (F(1,29) = 19.985 , p < 0.001). Furthermore, conflict scenarios have a higher pursuit duration than scenarios without a conflict (F(1,29) = 6.541 , p = 0.016). The effect of CA on pursuit duration is found to be significant (F(2,58) = 78.909, p < 0.001). Post hoc analyses show that the durations of fixations that contain pursuit, decrease with CA (p < 0.001 for all paired comparisons). Furthermore, the repeated measures ANOVA showed a significant effect of configuration (F(2,58) = 3.466, p = 0.038). The post hoc analyses show no significant difference between the separate configuration conditions.

4. Discussion

The aim of this study was to investigate in an empirical way the effects of different air traffic geometries on (a) conflict detection performance, and (b) eye movements during a dynamic conflict detection task, such that the perceptual processes and underlying cognitive processes during conflict detection could be examined.

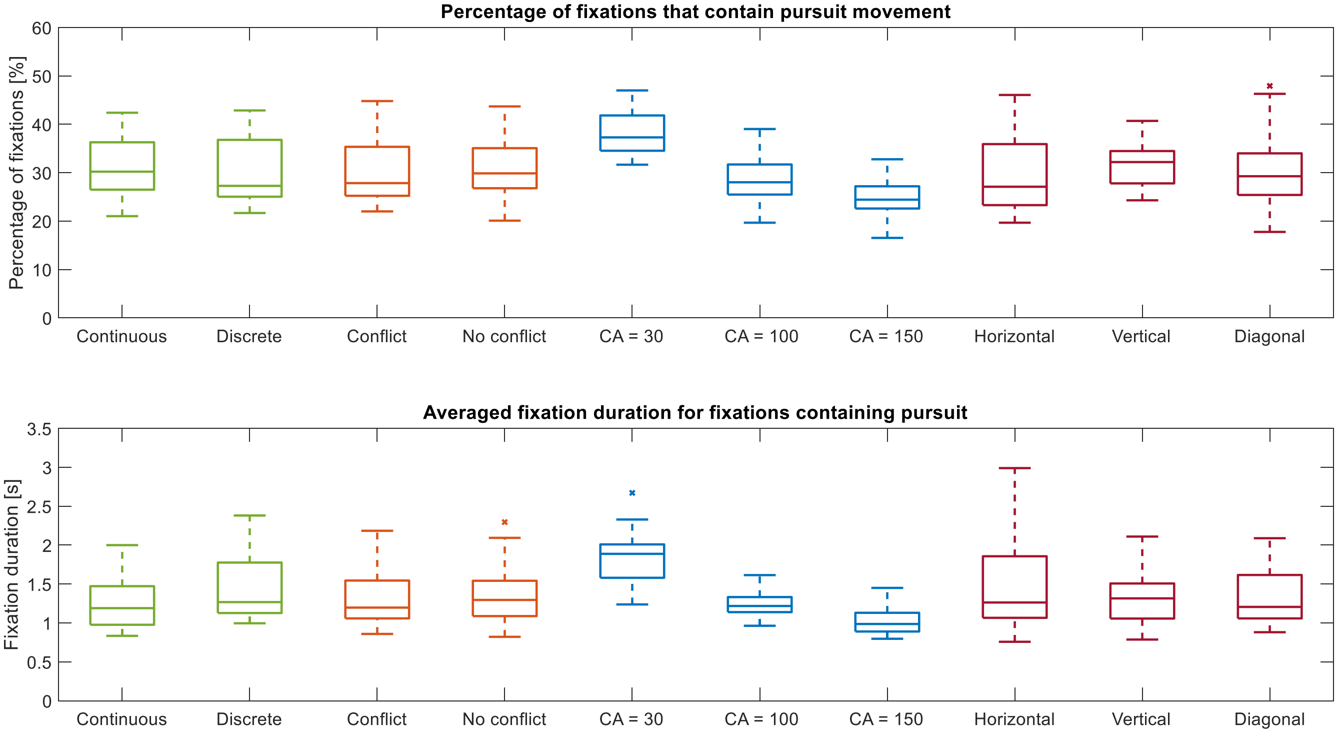


Figure 14: The percentage of fixations that contain pursuit movement and the average duration of those fixations per independent variable. For each participant, the average value is computed for each independent variables and those averages are shown in the box plot. So each box is created from 35 samples.

4.1 Performance and difficulty analyses

Continuous/discrete

In the hypotheses, it was stated that conflict detection would be easier with continuous moving stimuli than with discrete moving stimuli, because the information bandwidth of continuous signals is higher and therefore it would be easier to extract the trajectory and velocity of the dots, compared to discrete signals. Therefore, a higher performance score, a lower difficulty score, and a lower response time were expected for the continuous moving stimuli compared to discrete moving stimuli. In accordance with this hypothesis, continuous moving stimuli yielded a significantly higher performance score than discrete moving stimuli. Moreover, Figure 6 shows that between 8 and 16 seconds of the trials, a higher percentage of the participants had pressed the spacebar in continuous conflict situations compared to discrete conflict situations, indicating that participants were indeed better in detecting conflicts in the continuous situation. The averaged self-experienced difficulty score did not show a significant difference between continuous and discrete. A possible explanation could be that the subjective experience of difficult for continuous or discrete movement depends per person. Remarks from the participants after the experiment (see Appendix J) confirmed this idea.

Conflict angle

In the hypotheses, we stated that a smaller conflict angle would lead to a higher performance score and a lower difficulty score compared to a higher conflict angle (based on Neal and Kwantes, 2009; Remington et al., 2000, and Loft et al., 2009). Participants indeed performed with a significantly higher performance score, and reported a significantly lower difficulty score on the CA30 trials, compared to the higher angles. However, unexpected, participants had a significantly lower difficulty score and a higher performance score on the CA150 trials, compared to the CA100 trials. Also, the total number of false positives was higher for a conflict angle of 100 degrees compared to the other angles. These results indicate that in this experiment, conflict detection was hardest for participants with an angle of 100 degrees. This is confirmed by Figure 6, in which it can be seen that with an angle of 100 degrees the spacebar was pressed later in conflict scenarios than with the other angles.

It is possible that with an angle of 100 degrees, it is particularly difficult to predict if conflicts exist, compared to 30 and 150 degrees. According to Xu and Rantanen (2003), one cognitive strategy for conflict detection is the cognitive motion extrapolation theory (DeLucia and Liddell, 1998), which involves extrapolating the trajectories of the objects. With a conflict angle of 100 degrees, it might be harder to extrapolate the trajectories since the relative position of the dots has to be estimated both in horizontal and vertical direction. When the angle is closer to 0 degrees or 180 degrees, the distance between the dots in the vertical direction is small, therefore it might be easier to see if there is a conflict or not. With exactly 0 degrees, the two dots would either overlap from the beginning (if there is a conflict), or the second dot would start ahead of the first one (if there is no conflict), showing immediately whether there is a conflict or not. If the conflict angle is 180 degrees, the dots will always collide, also when the velocity would not be equal. So, it is possible that conflict angles of 30 and 150 degrees might be easier for conflict detection compared to 100 degrees because these angles are close to conflict angles for which it is easy to detect a conflict. It could be that for acute angles, performance increases with decreasing angle, while for obtuse angles, performance might increase for increasing angle. It would be interesting to investigate this theory further because ATCo's could be helped with information about which conflict angles are easier. For example, the flight scheme could be adjusted so that most common flight routes follow an easy angle.

In previous literature, no indications were found that conflict angles of 100 degrees would be harder than larger conflict angles. Remington et al. (2000) showed a difference in response time between acute and obtuse angles in general, however, they did not investigate the difference between specific angles. Therefore, their results do not specify whether a CA of 100 degrees might be harder for participants than other obtuse angles. Neal and Kwantes (2009) and Loft et al. (2009) did investigate specific angles, however, they examined different conflict angles (i.e., 45°, 90°, and 135°) than in this study. Loft et al. (2009) found that smaller angles yielded a higher probability of intervention for both conflict and non-conflict situations, indicating that smaller angles are regarded as a conflict earlier, compared to the larger angles. However, our results about the number of false positives show that trials with a conflict angle of 30 degrees got the least false positives. So, this indicates that in our experiment, the high performance score with 30 degrees was caused by the fact that participants were able to estimate the situation more accurately and not by a higher probability of intervention in general for small angles. The difference in results could be caused by the different structure of the experiment. Loft et al. (2009) tasked participants to indicate the likelihood that they would intervene for each stimulus instead of letting them indicate whether there would be a conflict or not. In the experiment of Neal and Kwantes (2009), participants were asked to react if they thought there would be a conflict, as in our experiment. However, the performance score was computed in a different way, which could have influenced the response of the participants. In the experiment of Neal and Kwantes (2009), participants were not allowed to change their response. At the moment the participant reacted whether he would intervene or not, the trial was ended. More points could be earned if reacting quickly, however, if the reaction was wrong, the same amount of points was subtracted from their score. So, in the study of Neal and Kwantes (2009), the results are based on the first reaction of participants, while in our study, the reaction during the entire trial is taken into account. It is possible, that with small conflict angles, people tend to think there is a conflict earlier (as found by Neal and Kwantes (2009) and Loft et al. (2009)), however, when considering the entire trial time, people are better in detecting conflicts with angles close to 0 and 180 degrees (as found in our study), with a preference for acute angles.

Moreover, Neal and Kwantes (2009) found that in situations without a conflict, the response time decreased if the angle of conflict increased, indicating that it could be easier to see that there is no conflict when the conflict angle is larger. Figure 5 is used to get more insight into the factors that influence the performance score in our experiment. It becomes apparent that both for conflict and non-conflict trials the performance increases and the difficulty decreases with decreasing conflict angle. The difference with the results of Neal and Kwantes (2009) could be caused by the fact that the stimuli, and thus, the predictability of the conflicts, were different in both studies. In the study of Neal and Kwantes (2009), the stimuli contained more visual aids, for example, the aircraft's airspeed was displayed and the flight routes were indicated with lines, thereby clearly showing the intersection point. However, in our experiment, there was little context that could help to estimate whether there would be a conflict or not. Therefore, participants could not be sure whether there was a conflict or not, until the very end of the trial, which may have caused the lack of effect of conflict/non-conflict on the difficulty score. Furthermore, the study of Neal and Kwantes (2009) was different from our study because they showed two aircraft pairs per trial. It could be that the presence of another aircraft pair influenced the results. Moreover, in the experiment of Neal and Kwantes the stimuli had to remain separated by 5-NM, which was indicated with a circle around the aircraft, while in our experiment the dots collided only if they were overlapping each other. It is possible, that these circles caused the aircraft to appear closer to each other with small conflict angles. So, the circles might cause participants to think there could be a conflict sooner, both for conflict and non-conflict situations. Further research must show what the influence of showing trajectory lines and 5-NM (no-go) circles is on conflict detection performance. Finally, it can be concluded from Figure 5, that discrete or continuous movement does not influence the performance and difficulty score per conflict angle. Both for conflict and non-conflict trials, the conflict angle of 100 degrees got the worst performance score and obtained the highest self-experienced difficulty.

Configuration

Participants obtained a significantly higher performance score when one of the dots was moving diagonally compared to the case in which one of the dots was moving horizontally, indicating that it was easier for humans to estimate the trajectory of a diagonal moving object. However, Eisma et al. (2017) found indications that diagonal sampling was less preferred by participants when they were tasked to sample six dials. It should be noted that the task from Eisma et al. (2017) did not contain linear moving stimuli and by diagonal sampling participants would have to skip dials, making the eye movements less efficient in this case. In our task, there was no choice for what to sample, there was only one pair of moving dots that had to be monitored. More research would be needed to ensure that diagonal sampling, in general, is easier. If this is indeed found to be the case, it could be implemented in the training program of ATCo's.

Conflict/non-conflict

Finally, the difference in performance score between conflict and non-conflict trials is caused by the fact that the ranking was not equal for both conditions and is therefore not relevant. Furthermore, the result that there was no significant difference between the difficulty of conflict and non-conflict trials confirms that it was difficult to predict the outcome of the trials, as was the intention of the experimental setup.

The small learning effect that is found for the performance score on the non-conflict trials indicates that participants became better at detecting non-conflict trials when they had done more trials. However, since the trials were shown in a random sequence to the 35 participants, this learning effect is not considered to have had much influence on the results.

4.2 Fixation distribution

Different kinds of scanning strategies are observed during the experiment. Some participants tended to keep their eyes (close) to one dot, others sampled a lot from one dot to the other, and some participants kept their eyes in the middle of the screen. Which strategy was used was different per individual as argued by Stankovic et al. (2008). It is noteworthy that over all trials, almost all fixations and gazes are directed at or between the two trajectories of the dots, which is in line with the conclusions of Hunter and Parush (2009). They found, with a static stimulus, that people scanned more often between two aircraft than towards the closest point of approach. We now confirmed this finding for the dynamic case as well. It is found that participants sample mostly from one dot to the other, and sample the collision site at the end of the trial, when the dots are close to it. Moreover, the large number of fixations between the dots could indicate that people are able to follow the movement of the dot from peripheral vision without the need to fixate on the dot itself. Preference for fixating in the middle, instead of switching attention from one dot to the other, can be related to the theory of Wickens et al. (2003), who predict that humans will try to minimize the visual effort. When both dots can be monitored by looking in the middle, humans might prefer that because the visual effort is less. The trials with an angle of 150 degrees show, at the beginning of the trials, fewer fixations in the middle and more fixations closer to the dots compared to the other angles. The fewer amount of fixations in the middle could indicate that the dots are too far away from each other to be monitored simultaneously by glancing in the middle. Also, the heat maps show for a conflict angle of 100 degrees more gazes in the middle between the trajectories than for 150 degrees. It is possible that participants try to estimate the relative position and movement of the dots by comparing them with an imaginary line (bisector) in between. Finally, the finding that many of the fixations contain pursuit movement indicates that participants followed each dot for a short time before a saccade was made towards the other dot. So, when the dots were too far away from each other to sample between the dots, participants sampled from one dot to the other, moving along with the dots a little bit during each fixation. Especially with conflict angles of 30 degrees, a long pursuit movement is often observed at the end of the trial indicating a special sort of sampling behavior for small angles.

4.3 Eye-movement analyses

Continuous/Discrete

A possible explanation for the significantly higher number of fixations and a higher number of comparisons during continuous trials compared to discrete trials could be the fact that the frequency bandwidth was higher for continuous trials. The framerate of the discrete moving stimuli was 2 Hz, while the continuous moving stimuli had a framerate of 30 Hz. Senders (1983) found that a human observer would ideally sample a signal that has a frequency bandwidth of W with a sampling frequency greater than $2W$. According to this theory, the continuous signals should indeed receive more fixations and comparisons between the dots than discrete signals. Furthermore, the fact that the duration per fixation was significantly lower for continuous trials compared to discrete trials indicates that more time was needed with discrete stimuli to obtain information about the speed and direction of the dot. Overall, the combination of performance and eye movement results on continuous and discrete movements could indicate that with continuous moving stimuli participants get more information with more glances and are able to judge better whether there is a conflict or not.

Conflict angle

The significantly higher amount of fixations for increasing conflict angles was predicted since trials with larger AOIs were expected to be harder. Therefore, it was expected that more fixations would be needed to perceive the situation correctly. Furthermore, the fixation duration is found to be significantly lower when the CA gets higher, which makes sense in relation to the higher amount of fixations for higher CAs since if more fixations are done at the same time, the fixation time will be shorter. The higher fixation durations for trials with a smaller conflict angle can be explained by the findings of pursuit movement. Larger CAs got a significantly lower percentage of fixations that contain pursuit movement. The length of the pursuits decreased significantly with increasing conflict angle, indicating that for smaller angles, more pursuit movement was used.

Furthermore, the observed relationship between conflict angle and the number of fixations is similar to the amount of comparisons. Participants made significantly more comparisons between the AOIs with increasing conflict angle. However, it was expected that the effort to sample between higher angles would be higher and therefore (according to the SEEV model of Wickens et al., 2003), it was expected that fewer comparisons would be made for higher angles. It should be noted, that the SEEV model assumes that if there is a choice in what to sample, an operator will prefer the low effort case. However, in this experiment there was no choice, urging participants to sample the presented aircraft pair, even though the effort was high. It could be, in a situation with three dots, that the smaller angle gets more glances than the larger angle. With two dots it seems participants wanted to have a correct estimate of the trajectory and velocities of the dots, even if that would cost some extra effort. It could be argued that the expectancy of the location of the dot that is not sampled at a certain moment, would be higher for larger angles since the dots are further away from each other, therefore increasing the uncertainty about the location of the dot and (again following the SEEV model), therefore increasing the sample probability.

Moreover, from the significantly higher number of fixations for increasing conflict angle, the same relation was expected for time on AOI. However, the percentage time on AOIs with a conflict angle of 30 degrees was unexpectedly high, indicating that with an angle of 30 degrees, participants often have their eyes on the AOIs, without exhibiting new fixations. These contradicting results can be explained with the findings of pursuit movements, since participants used significantly more fixations that contained pursuit movement for 30 degrees compared with the others angles, resulting in time on the AOIs without extra fixations.

Finally, it is notable that the unexpected outcome of the conflict angle of 100 degrees in the performance score was not resembled in the eye-movement variables. So, although this study gives reasons to believe that conflict detection with an angle of 100 degrees might be harder for participants, there are no indications that the according eye movements of this angle are different. So, the lower performance on the 100 degree trials cannot be explained by sampling behavior.

Configuration

The significantly lower time on the AOIs for the vertical configuration compared to the horizontal configuration could mean that the exact trajectory of the dots was harder to predict when one of them was moving vertically. Therefore, participants did not look to the right trajectory resulting in a gaze directed towards the AOIs a shorter percentage of time. The result that the vertical configuration got significantly more pursuit fixations compared to the horizontal configuration is unexpected since humans are in general better in horizontal than in vertical smooth pursuit (Holmqvist et al., 2011, p309). It may be argued that this difference was caused by the fact that in this study, horizontal and vertical only referred to one of the two dots. As a recommendation, it could be interesting to further investigate the use of smooth pursuit in conflict detection tasks, since it gives much insight into operators' visual sampling techniques. Furthermore, no significant effect on fixation duration and the number of comparisons was found. Furthermore, there was only a significant difference between the diagonal and horizontal trials in case of the performance score. Thus, the difference in performance cannot be related to the difference found in eye movements. The absence of a significant effect for most dependent variables on all configurations indicates that the configuration does not have a significant influence on visual sampling behavior, as predicted. Finally, the result that significantly more fixations were directed towards AOI 1, indicates that participants had a preference for sampling the dot that followed a horizontal, vertical, or diagonal line. It could be that it was easier for participants to estimate the trajectory of the first dot and that they compared the location of the second dot to the estimated trajectory. It seems participants used the first dot as a visual reference for the movement of the second dot.

Conflict/non-conflict

The number of fixations and the time spent by participants on the AOIs per trial, was significantly higher without a conflict, compared to trials with a conflict. No significant difference between the percentages of fixations that contain pursuit was found. However, the average duration of the fixations with pursuit was higher for conflict scenarios. These results indicate that the lower number of fixations could be caused by longer pursuit movement for conflict scenarios. It could be that in the conflict situation, participants decided at a certain moment that there would be a conflict and stopped sampling actively, instead of using a pursuit movement to follow the dots until they reached the conflict area. At the end of the trials, relatively few fixations and few comparisons were found in conflict trials compared to non-conflict trials (see Appendix B6). Since CA100 also showed a difference in difficulty score between conflict and non-conflict trials (Figure 5), this could indicate that participants thought that non-conflict trials with a conflict angle of 100 degrees were particularly difficult. However, this difference is not found in the performance score (Figure 6).

4.4 Practical applications of the study

This study shows the results of a simulation of a simplified ATC task. It differs from a real ATC task in several ways. For example, the displayed speed of the aircraft on a real flight radar is much lower, the screen update frequency is lower and the number of aircraft that are visible is much higher. Furthermore, it should be noted that our performance score is artificial. In reality, an ATCo would indicate a situation to be a conflict earlier than in our experiment since the cost of allowing two aircraft to collide is much higher in reality compared to an unnecessary conflict resolution maneuver. However, by simplifying the task, the independent variables could be varied in a more systematical manner, making it possible to draw conclusions on human sampling behavior without the influence of confounds like speed labels, flight routes, and other aircraft.

The results of the difference between continuous and discrete moving stimuli show that detecting conflicts might be easier with continuous movements. Furthermore, the eye movement analysis shows that with continuous movements, more fixations and comparisons can be used to optimally process the information. This could be implemented in ATC by increasing the update frequency of the aircraft positions on the screen. The ideal frequency for experienced ATCo's must be found before we know whether continuous signals or discrete signals with a higher update frequency are optimal. This study shows that increasing the information intake of operators could mean an increase in conflict detection performance. However, it should also be considered that in this experiment, all the information about the speed of the dots had to be obtained from the changing location of the dots. In real life, however, this information is provided in the flight label. If continuous movements are indeed found to increase conflict detection performance of ATCo's, real time air traffic information may be used to achieve a high framerate radar update.

Furthermore, the results on configuration suggest that diagonal movement could make it easier to detect a conflict for novices. If further research shows this is also the case for experienced ATCo's, it can be applied in the design of the flight radar. For example, it could be considered to position flight routes that create many conflicts in a diagonal position by rotating the screen. This adaption could make the task of the ATCo easier and therefore increase performance.

Moreover, this study confirms that in general, a smaller conflict angle makes it easier to predict whether there will be a conflict or not. In practice this could be useful information for the training of ATCo's, to be able to increase the difficulty level of practice simulations. Furthermore, when designing flight schemes and routes, it might be possible to adjust the scheme such that common flight routes are positioned towards each other with a smaller angle. Lastly, in a more theoretical context, this study provides new information about the influence of conflict angle on performance when there is no conflict present. Our results do not confirm the results from Neal and Kwantes (2009), therefore, more research needs to be done to figure this out in more detail.

Finally, it can be argued that the results of this study could be applied to other relative judgment tasks. Due to the abstract design of the stimuli, the conclusions that are drawn are applicable in a more general way and not specific for air traffic control. However, it should be noted that, due to the design of our stimuli in 2D, our conclusions are restricted to the detection of horizontal conflicts.

5. Conclusions and recommendations

This study is the first to show in a dynamic simulation the relationships between eye movements and conflict geometries. The various findings reported in this study could be further interpreted in the framework of existing theories on cognitive processes during conflict detection and supervisory control in general, such that insight can be obtained about how humans perform such tasks. Various studies have already investigated the relationships between conflict detection performance and atypical simplified air traffic situations. Furthermore, eye tracking is developed and is becoming a reliable way to investigate perception processes of participants while they perform tasks. However, eye tracking had not been measured yet while participants were performing a dynamical conflict detection task. The relationships between eye movements and different conflict geometries provides insight into how humans perform a conflict detection task. This knowledge could prove to be useful in order to make the training of ATCo's more efficient. Also, the radar display could be adjusted to improve conflict detection performance. Furthermore, this experiment could help refine existing theories about sampling behavior.

The first aim of this study was investigating the effects of different air traffic geometries on conflict detection performance**.** We can conclude that the smallest angle was, as reported by previous literature, perceived as being the easiest and obtained the highest performance score. However, we also conclude that an increase in CA does not always increase performance and consequently reduces the difficulty. An angle of 100 degrees obtained worse performance and a higher difficulty score than an angle of 150 degrees. So, in general, it might be true that with higher CA it is more difficult for participants to detect a conflict, but this study indicates that there are exceptions to this general trend. It could be possible that if the angle becomes closer to 180 degrees, conflict detection becomes easier again. Furthermore, it was concluded that it might be easier for participants to detect conflicts when the stimuli move continuously. This could be useful to implement in real ATC tasks such that performance can improve. Finally, indications are found that sampling for a diagonal case was easier for participants. If this finding is confirmed with experienced participants, it can be used in flight radar display design. Also, this information about which traffic situations are easier for participants, can be used in the education program for ATCo's. For example, the difficulty level of training situations can be adjusted more precisely.

The second aim of this study was to investigate in an empirical way the relationships between eye movements and conflict detection for different ATC traffic geometries, such that the underlying processes during conflict detection could be examined. It can be concluded that participants direct almost all their attention to the area between the two moving objects. Participants direct their fixations ahead and between the moving objects and not directly at them. Furthermore, it is concluded that CA has a strong influence on eye movements. Increasing the CA leads to more fixations with shorter fixation durations. Also, if the CA is increased, participants tend to sample more from one object to the other and exhibit less pursuit movement. Moreover, it can be concluded that there is a difference in eye movements between continuous and discrete moving stimuli. With continuous moving stimuli, participants adjust their sampling behavior with more fixations, lower durations compared to discrete moving stimuli. Finally, it is concluded that for trials in which no conflict occurred, participants exhibit more fixations and sample more from one dot to the other compared to trials with a conflict.

Recommendations

Eisma et al. (2017) investigated eye movements with rotating stimuli. This study investigated eye movements with linear stimuli. The next step could be to investigate eye movement behavior while looking at non-linear moving stimuli. For example, the effect of a change in speed during a trial could be investigated. Ascending and descending aircraft draw much attention (Eyferth et al., 2003), so it could be predicted that more fixations are directed towards the dots when they do not move linearly.

Furthermore, we found a difference in both performance and eye movement behavior between continuous and discrete moving stimuli. Currently, ATC works with discretely updating flight radars. Further research must show whether the preference for discrete or continuous is person-dependent or that indeed, as our results indicate, continuous movement is in general easier for humans to process. Also, our results confirm that conflicts are easier to detect between small angles but that it is not certain whether the difficulty is increasing with larger angles. It is possible that conflict detection becomes easier again when the angle comes closer to 180 degrees. It would be interesting to try different CAs (e.g. 50, 70, 110, 130, 150, 170 degrees) to investigate what could be the cause of the increasing difficulty of CA100.

Moreover, it is recommended to study the effects of the independent variables in a task that is more like real ATC to investigate whether the same results are obtained. More aircraft could be added to the scenarios so that interaction effects can be taken into account. With more aircraft, we predict that less pursuit movement will be exhibited since there are more different areas that require attention. Also, we expect that in a situation with three dots or more, a smaller conflict angle would get more glances than larger angles. More glances are expected for small angles because sampling between higher angles is expected to cost more effort, resulting in fewer glances, according to the SEEV model (Wickens et al., 2003). Furthermore, it is shown (e.g. by Wang et al., 2015), that novices exhibit different sampling behaviors compared to experienced ATCo's. Therefore, we recommend repeating this study with experienced participants instead of students, to see whether the same results are obtained.

Finally, our results show per independent variable which percentage of fixations contain pursuit movements and the duration of those fixations. The finding that smooth pursuit is used for small angles to look between objects to see how they both are moving, gives new insight into sampling behavior. However, it has not been investigated which part of the fixations exactly is pursuit. The longer fixations that were found to contain pursuit, most likely consisted partly of small fixations and partly of pursuit movement. For further research it is recommended to divide each fixation in pursuit and fixation time stamps (for example with the algorithm as described by San Agustin (2010)), to get a better understanding of what is happening at the end of the CA30 trials.

References

Backs, R. W., & Walrath, L. C. (1992). Eye movement and pupillary response indices of mental workload during visual search of symbolic displays. *Applied Ergonomics*. https://doi.org/10.1016/0003-6870(92)90152-L

Bjorklund, C., Alfredson, J., & Dekker, S. (2006). Mode Monitoring and Call-Outs: An Eye-Tracking Study of Two-Crew Automated Flight Deck Operations. *The International Journal of Aviation Psychology*, *16*(3), 263–275. https://doi.org/10.1207/s15327108ijap1603_2

Boag, C., Neal, A., Loft, S., & Halford, G. S. (2006). An analysis of relational complexity in an air traffic control conflict detection task. *Ergonomics*, *49*(14), 1508–1526. https://doi.org/10.1080/00140130600779744

Braddick, O. (1974). A short-range process in apparent motion. *Vision Research*, *14*(7), 519–527. https://doi.org/10.1016/0042-6989(74)90041-8

Carbonell, J. R. (1966). A queueing model of many-instrument visual sampling. *IEEE Transactions on Human Factors in Electronics*, *HFE-7*(4), 157–164. https://doi.org/10.1109/THFE.1966.232984

Das, L., Iqbal, M. U., Bhavsar, P., Srinivasan, B., & Srinivasan, R. (2018). Toward Preventing Accidents in Process Industries by Inferring the Cognitive State of Control Room Operators through Eye Tracking. *ACS Sustainable Chemistry & Engineering*, *6*, 2517–2528. https://doi.org/10.1021/acssuschemeng.7b03971

de Greef, T., Lafeber, H., van Oostendorp, H., & Lindenberg, J. (2009). Eye Movements as Indicators of Mental Workload to Trigger Adaptive Automation. https://doi.org/10.1007/978-3-642-02812-0

DeLucia, P. R., & Liddell, G. W. (1998). Cognitive Motion Extrapolation and Cognitive Clocking in Prediction

Motion Tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 901–914. https://doi.org/10.1037/0096-1523.24.3.901

Duerrschmid, K., & Danner, L. (2018). Eye Tracking in Consumer Research. *Methods in Consumer Research*, *2*, 279–318. https://doi.org/doi.org/10.1016/B978-0-08-101743-2.00012-1

Eisma, Y. B., Cabrall, C. D. D., & De Winter, J. C. F. (2017). Visual Sampling Processes Revisited: Replicating and Extending Senders (1983) Using Modern Eye-Tracking Equipment. *IEEE Transactions on Human-Machine Systems Manuscript*.

Ellerbroek, J. (2013). *Airborne conflict resolution in three dimensions*.

Eyferth, K., Niessen, C., & Spaeth, O. (2003). A model of air traffic controllers' conflict detection and conflict resolution. *Aerospace Science and Technology*, *7*(6), 409–416. https://doi.org/10.1016/S1270-9638(03)00064-6

Fidopiastis, C. M., Drexler, J., Barber, D., Cosenzo, K., Barnes, M., Chen, J. Y. C., & Nicholson, D. (2009). Impact of automation and task load on unmanned system operator's eye movement patterns. *Foundations of Augmented Cognition, Neuroergonomics and Operational Neuroscience*, 229–238. https://doi.org/10.1007/978-3-642-02812-0_27

Flemisch, F. O., & Onken, R. (2000). Detecting Usability Problems with Eye Tracking in Airborne Battle Management Support. *RTO MP*, *57*(23).

Hancock, P. A. (2014). Automation: how much is too much? *Ergonomics*. Taylor & Francis. https://doi.org/10.1080/00140139.2013.816375

Haslbeck, A. (2012). The relationship between pilots ' manual flying skills and their visual behavior : a flight simulator study using eye tracking, (November 2015), 3131–3138.

Hildreth, E. C., & Ullman, S. (1982). Ad-a128 398 the measurement of visual motion(u) massachusetts inst.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

Hunter, A. C., & Parush, A. (2009). Using Eye Movements to Uncover Conflict Detection Strategies. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, *53*, 1729–1733. https://doi.org/10.1518/107118109X12524444081872

Ikuma, L. H., Harvey, C., Taylor, C. F., & Handal, C. (2014). A guide for assessing control room operator performance using speed and accuracy, perceived workload, situation awareness, and eye tracking. *Journal of Loss Prevention in the Process Industries*, *32*, 454–465. https://doi.org/10.1016/j.jlp.2014.11.001

Imants, P., & de Greef, T. (2014). Eye Metrics for Task-Dependent Automation. In *Proceedings of the 2014 European Conference on Cognitive Ergonomics - ECCE '14*. https://doi.org/10.1145/2637248.2637274

ISO 15007-1. (2014). Road vehicles - Measurement of driver visual behavior with respect to transport information and control systems - Part 1: Definitions and parameters. Geneva, Switzerland: International Standards Organization.

Jamson, A. H., Merat, N., Carsten, O. M. J., & Lai, F. C. H. (2013). Behavioral changes in drivers experiencing highly-automated vehicle control in varying traffic conditions. *Transportation Research Part C: Emerging Technologies*, *30*, 116–125. https://doi.org/10.1016/j.trc.2013.02.008

Kilingaru, K., Tweedale, J. W., Thatcher, S., & Jain, L. C. (2013). Monitoring pilot "Situation Awareness." *Journal of Intelligent and Fuzzy Systems*, *24*(3), 457–466. https://doi.org/10.3233/IFS-2012-0566

Kodappully, M., Srinivasan, B., & Srinivasan, R. (2016). Towards predicting human error: Eye gaze analysis for identification of cognitive steps performed by control room operators. *Journal of Loss Prevention in the Process Industries*, *42*, 35–46. https://doi.org/10.1016/j.jlp.2015.07.001

Koffskey, C., Ikuma, L. H., Harvey, C., & Aghazadeh, F. (2014). Using Eye-Tracking to Investigate Strategy and Performance of Expert and Novice Control Room Operators. In *Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting* (pp. 1667–1671). https://doi.org/10.1177/1541931214581348

Lin, Y., Zhang, W. J., & Watson, L. G. (2003). Using eye movement parameters for evaluating human-machine interface frameworks under normal control operation and fault detection situations. *International Journal of Human Computer Studies*, *59*(6), 837–873. https://doi.org/10.1016/S1071-5819(03)00122-8

Loft, S., Bolland, S., Humphreys, M. S., & Neal, A. (2009). A Theory and Model of Conflict Detection in Air Traffic Control: Incorporating Environmental Constraints. *Journal of Experimental Psychology: Applied*, *15*(2), 106–124. https://doi.org/10.1037/a0016118

Mackintosh, M.-A., Dunbar, M., Lozito, S., Cashion, P., Mcgann, A., Dulchinos, V., … Gent, R. Van. (1998). Self-Separation from the Air and Ground Perspective. *USA/Europe Air Traffic Management R&D Seminar*, (December), 1–16.

Manske, P. G., & Schier, S. L. (2015). Visual Scanning in an Air Traffic Control Tower - A Simulation Study. *Procedia Manufacturing*, *3*, 3274–3279.

https://doi.org/10.1016/j.promfg.2015.07.397

Moray, N. (1986). Monitoring Behavior and Supervisory Control. In *Handbook of perception and human performance, Vol. 2* (pp. 40-01, 40–46).

Moray, N., & Rotenberg, I. (1989). Fault management in process control: eye movements and action. *Ergonomics*, *32*(11), 1319–1342. https://doi.org/10.1080/00140138908966910

Neal, A., & Kwantes, P. (2009). An Evidence Accumulation Model for Conflict Detection Performance in a Simulated Air Traffic Control Task. *Human Factors*, *51*(2), 164–180. https://doi.org/10.1177/0018720809335071.

Oprins, E., Burggraaff, E., & Weerdenburg, H. Van. (2006). Design of a Competence-Based Assessment System for Air Traffic Control Training. *International Journal of Aviation*, *8414*(3), 297–320. https://doi.org/10.1207/s15327108ijap1603

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans.*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

Rantanen, E. M., & Nunes, A. (2005). Hierarchical Conflict Detection in Air Traffic Control. *The International Journal of Aviation Psychology*, *15*(4), 339–362. https://doi.org/10.1207/s15327108ijap1504

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Remington, R. W., Johnston, J. C., Ruthruff, E., Gold, M., & Romera, M. (2000). Visual search in complex displays: factors affecting conflict detection by air traffic controllers. *Human Factors*, *42*(3), 349–366. https://doi.org/10.1518/001872000779698105

Robinski, M., & Stein, M. (2013). Tracking Visual Scanning Techniques in Training Simulation for Helicopter Landing. *Journal of Eye Movement Research*, *6*(2:3), 1–17. https://doi.org/10.16910/jemr.6.2.3

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the Symposium on Eye Tracking Research & Applications - ETRA '00*, 71–78. https://doi.org/10.1145/355017.355028

San Agustin, J. (2010). Off-the-Shelf Gaze Interaction. *Communication*, (September), 179.

Sarter, N. B., Mumaw, R. J., & Wickens, C. D. (2007). Pilots' monitoring strategies and performance on automated flight decks: an empirical study combining behavioral and eye-tracking data. *Human Factors*, *49*(3), 347–357. https://doi.org/10.1518/001872007X196685

Senders, J. W. (1964). The Human Operator as a Monitor and Controller of Multidegree of Freedom Systems. *IEEE Transactions of Human Factors in Electronics*, *HFE-5*(1).

Senders, J. W. (1983). *Visual sampling processes*.

Sharma, C., Bhavsar, P., Srinivasan, B., & Srinivasan, R. (2016). Eye gaze movement studies of control room operators: A novel approach to improve process safety. *Computers and Chemical Engineering*, *85*, 43–57. https://doi.org/10.1016/j.compchemeng.2015.09.012

Sheridan, T. (1984). Supervisory contol of remote manipulators, vehicles and dynamic process: Experiments in command and display aiding. In *MASSACHUSETS INST OF TECH CAMBRIDGE MAN/MACHINE SYSTEMS LAB*.

Sheridan, T. (1986). Human supervisory control of robot systems. *Proceedings 1986 IEEE International Conference on Robotics and Automation*, *3*, 808–812. https://doi.org/10.1109/ROBOT.1986.1087506

SMI cooperation. (2014). BeGaze User Manual.

Stankovic, S., Raufaste, E., & Averty, P. (2008). Determinants of conflict detection: a model of risk judgments in air traffic control. *Human Factors*, *50*(1), 121–34. https://doi.org/10.1518/001872008X250584.

Thomas, L. C., & Wickens, C. D. (2004). Eye-tracking and Individual Differences in off-Normal Event Detection when Flying with a Synthetic Vision System Display. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *48*(1), 223–227. https://doi.org/10.1177/154193120404800148

Thomas, L. C., & Wickens, C. D. (2006). Display Dimensionality, Conflict Geometry, and Time Pressure on Conflict Detection and Resolution Performace Using a Cockpit Display of Traffic Information. *The International Journal of Aviation Psychology*, *16*(3), 321–342.

Trouvain, B., & Schlick, C. M. (2007). A Comparative Study of Multimodal Displays for Multirobot Supervisory Control. *Engin. Psychol. and Cog. Ergonomics, HCII*, 184–193.

Valerie, A., Huemer, M. S., Hayashi, M., Renema, F., Elkins, S., McCandless, J. W., & McCann, R. S. (2005). Characterizing Scan Patterns in a Spacecraft Cockpit Simulator: Expert Vs. Novice Performance. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 49, pp. 83–87). https://doi.org/10.1177/154193120504900119

Wang, Y., Cong, W., Dong, B., Wu, F., & Hu, M. (2015). Statistical analysis of air traffic controllers ' eye movements. In *11th USA/Europe Air Traffic Management Research and Development Seminar* (pp. 1–9).

Wickens, C. D., Goh, J., Helleberg, J., Horrey, W. J., &

Talleur, D. A. (2003). Attentional Models of Multitask Pilot Performance Using Advanced Display Technology. *Human Factors*, *45*(3), 360–380. https://doi.org/10.1518/hfes.45.3.360.27250

Wickens, C. D., Helleberg, J., Goh, J., Xu, X., & Horrey, W. J. (2001). *Pilot Task Management: Testing an Attentional Expected Value Model of Visual Scanning*.

Wickens, C. D., McCarley, J. S., Alexander, A. L., Thomas, L. C., Ambinder, M., & Zheng, S. (2008). Attention-Situation Awareness (ASA) Model of Pilot Error. In *Human performance modeling in aviation* (pp. 213–239). https://doi.org/10.1201/9781420062984.ch9

Xing, J. (2006). Color and visual factors in ATC displays, (June), 1–15.

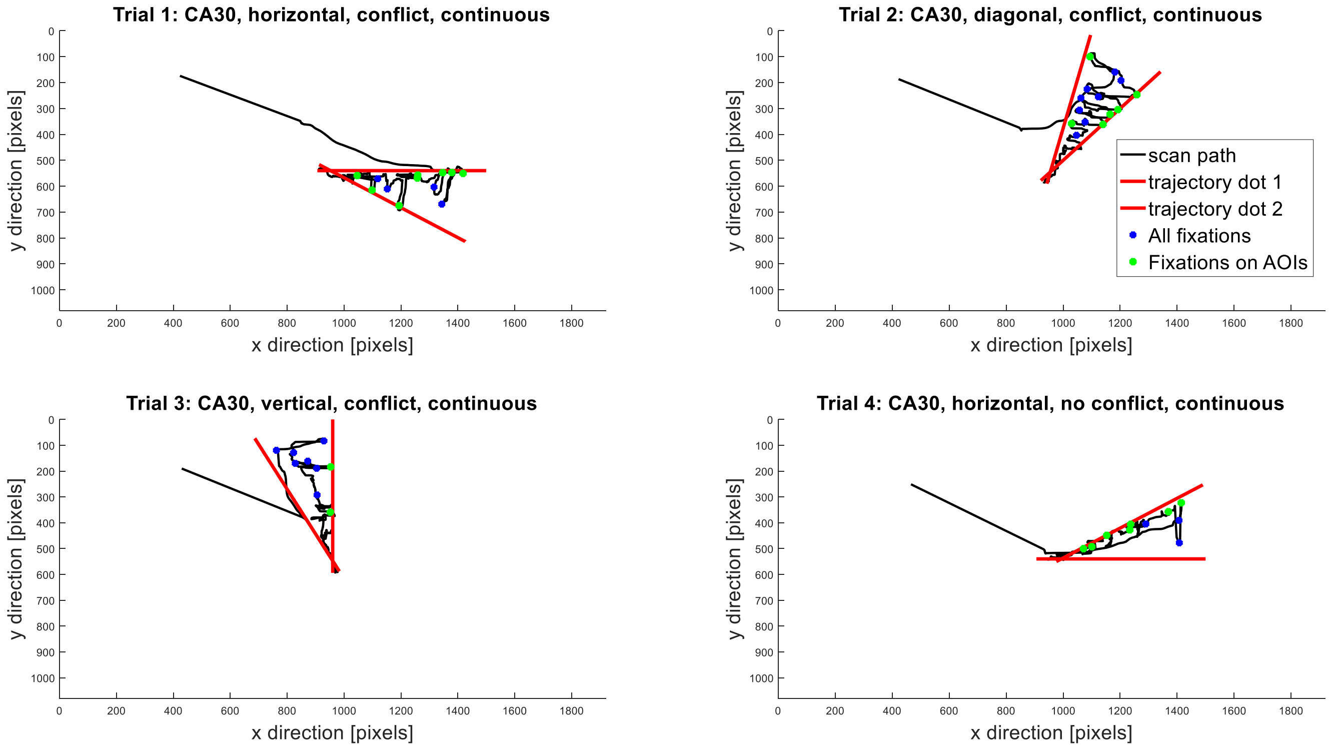
Xu, X., & Rantanen, E. (2003). Conflict detection In air traffic control: A task analysis, a literature review, and a need for further research. *Proceedings of the International Symposium on Aviation Psychology*, (March).

Appendix A: Line AOIs vs circular AOIs

In this appendix, it is explained why the trajectory lines are chosen as AOIs. The area of 30 pixels around the linear trajectories of the dots are used as AOIs for this paper. So, each fixation that lies within 30 pixels of one of the two red lines in Figure A1 is counted as a fixation on an AOI. Figure A1 shows the scan path and fixations of the first 4 trials of the first participant. In green, the fixations are shown which are considered to be on an AOI with the line AOI definition. The extra fixation after the last saccade that was added for the pursuit results is not displayed in these graphs.

It is also considered to use circular AOIs, which would be a circle with a radius of 30 pixels around the center of the moving dot. The advantage of these circle AOIs would be that the number of fixations towards the dots are counted and fixations far from the dots, but on the trajectory line, are not taken into account. So for the circle AOIs, the locations of the AOIs change with time, while for the line AOIs they are constant during a trial. Figure A2 shows the same 4 trials as Figure A1, however, now the fixations are colored green if they lie within 30 pixels radius of the center of the moving dots. It was found that with circle AOIs only few fixations are considered to be on an AOI.

So apparently the fixations were not located directly on top of the dots. Increasing the size of the circle AOI is tried, but still, the line AOI gave a better representation of the number of fixations on the AOIs. Therefore, it is concluded that for this experiment line AOIs would give the best insight in the results.



Trial 1: CA30, horizontal, conflict, continuous

Trial 2: CA30, diagonal, conflict, continuous

Trial 3: CA30, vertical, conflict, continuous

Trial 4: CA30, horizontal, no conflict, continuous

Figure A1: Overview of fixations on AOIs when line AOIs are used. The first four trials of the first participants are displayed.

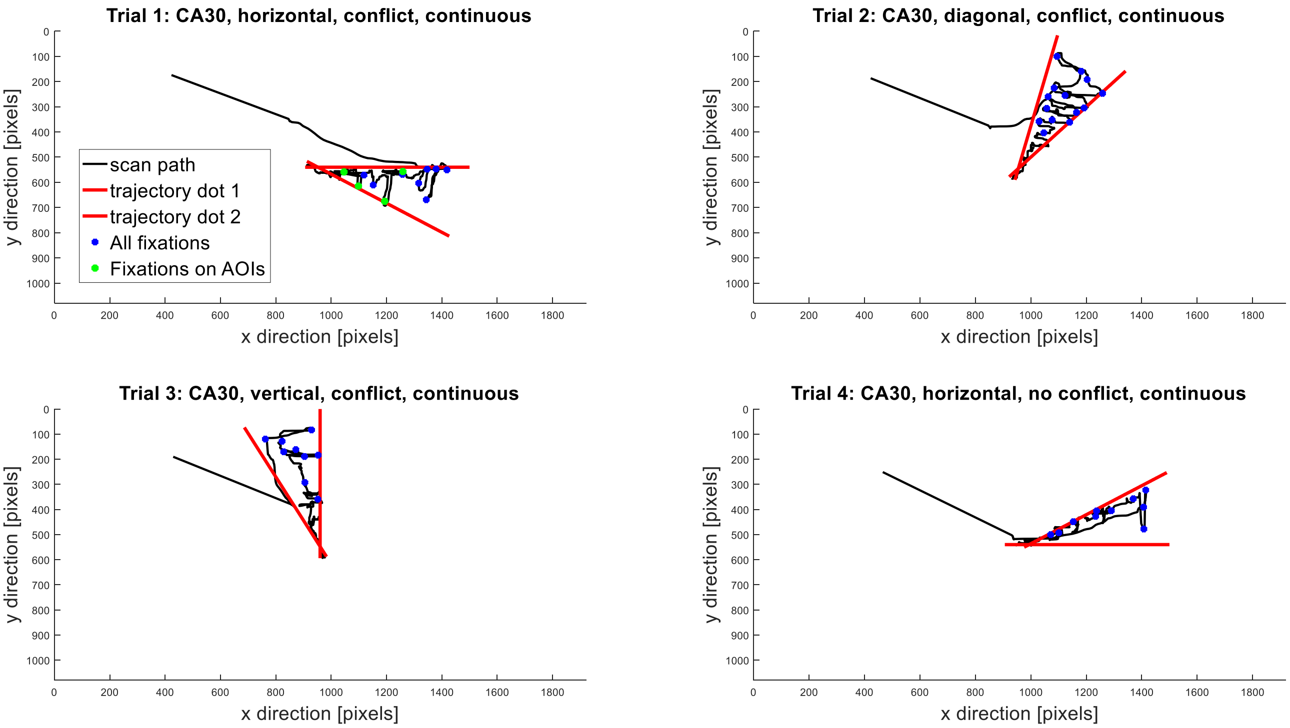


Figure A2: Overview of fixations on AOIs when circular AOIs are used. The first four trials of the first participants are displayed.

Appendix B: Additional results

This appendix shows the results found with some additional dependent variables. These results are not incorporated in the main article because they do no answer the research question or because they do not add extra elements to the conclusion. The additional dependent variables are:

- Number of fixations per AOI for all trials
- Pupil size over time (mm)
- Number of fixations to the AOIs
- Averaged fixation duration of fixations to AOIs (s)
- Saccadic duration (s) per conflict angle
- Percentage of fixations over time (%)
- Percentage of comparisons over time (%)
- Pursuit duration against trial time
- Fixations in the xy plane

B.1 Fixations per AOI

Figure B1 shows the number of fixations to AOI 1 and AOI 2 for all 36 trials. In green the fixations are counted which lay in both AOIs, which could for example happen at the end of the trial when both dots were close to the center of the screen. In this graph the fixations are not divided into the different independent variables, however it gives some insight in how the fixations are divided over the two AOIs.

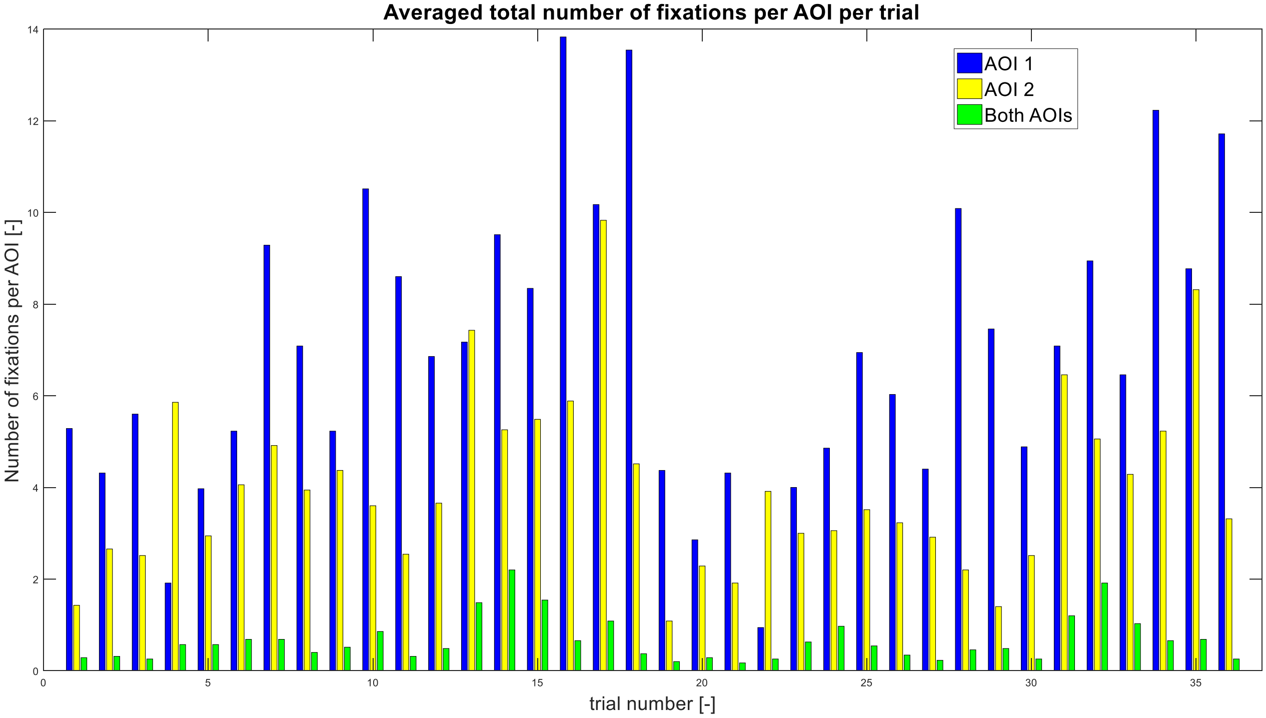


Figure B1: Total number of fixations to each AOI per trial, averaged over all participants.

B.2 Pupil size over time

Since the background of the videos is quite uniform for all 36 trials, the pupil size can be compared. Figure B2 shows the average pupil size against trial time separately for the different independent variables. Interpolation was used to fill the time samples that contained no pupil size, for example during blinks.

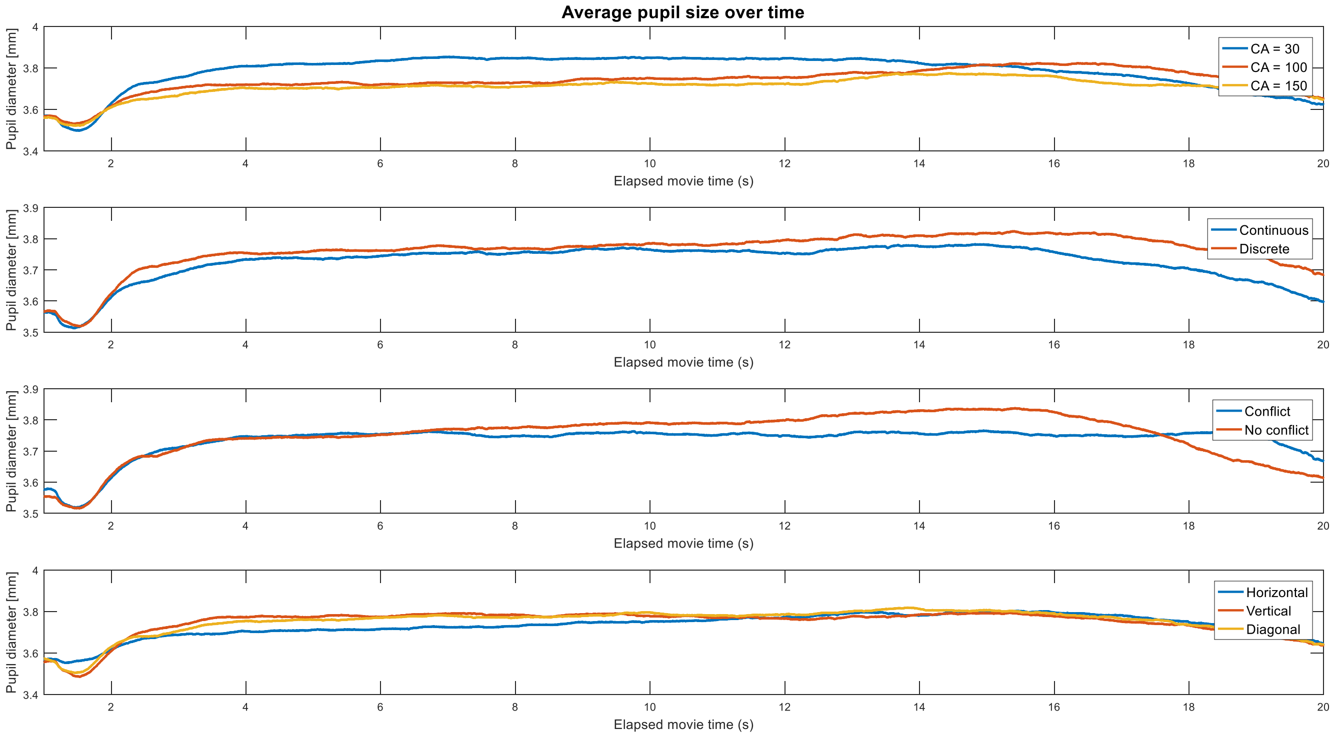


Figure B2: Pupil size over time for the different independent variables.

B.3 Fixations to AOIs

The average number of fixations to the line AOIs and their average duration can be found per independent variable in Figure B3.

Number of fixations on AOIs

With continuous moving stimuli there were significantly more fixations on AOIs than with discrete moving stimuli (F(1,34)= 50.367, p < 0.001). The number of fixations was significantly lower with a conflict than without a conflict (F(1,34) = 23.012, p < 0.001). Furthermore, a significant difference was found for CA (F(2,68) = 267.989, p < 0.001). Post hoc tests revealed a significant increase in number of fixations with increasing CA (p < 0.001 for all comparisons). A significant effect of configuration was found (F(2,68) = 3.303, p = 0.043). However, the post hoc tests did not show a significant difference between the pairs. Concluding, the same significant effects are found compared to the results when all fixation were taken into account. The only difference is that when all fixations were considered, the difference between horizontal and vertical was found to be significant and with only fixations on AOIs it was not, which could be because the number of fixations were higher.

Fixation duration on AOIs

The lower part of Figure B3 shows the duration of fixations directed to an AOI per independent variable. 15 of the 1260 trials did not have any fixation on the AOIs, therefore, these trials were not taken into account. Discrete trials showed a significantly higher fixation duration than continuous trials (F(1,27) = 6.555 , p = 0.016). No significant difference was found between conflict and non-conflict situations (F(1,27) = 1.683 , p = 0.205). The CA had a significant effect on the fixation duration (F(2,54) =47.537, p < 0.001) and post hoc analyses showed that the fixation duration decreased if CA increased (p < 0.001 for all comparisons). The effect of configuration was significant (F(2,54) =4.402, p = 0.017). The pairwise comparisons however did not show a significant effect between configurations. Again the results are very similar to the case when all fixations were considered. The only difference lies in the configuration, were no significant effect was found for all fixations.

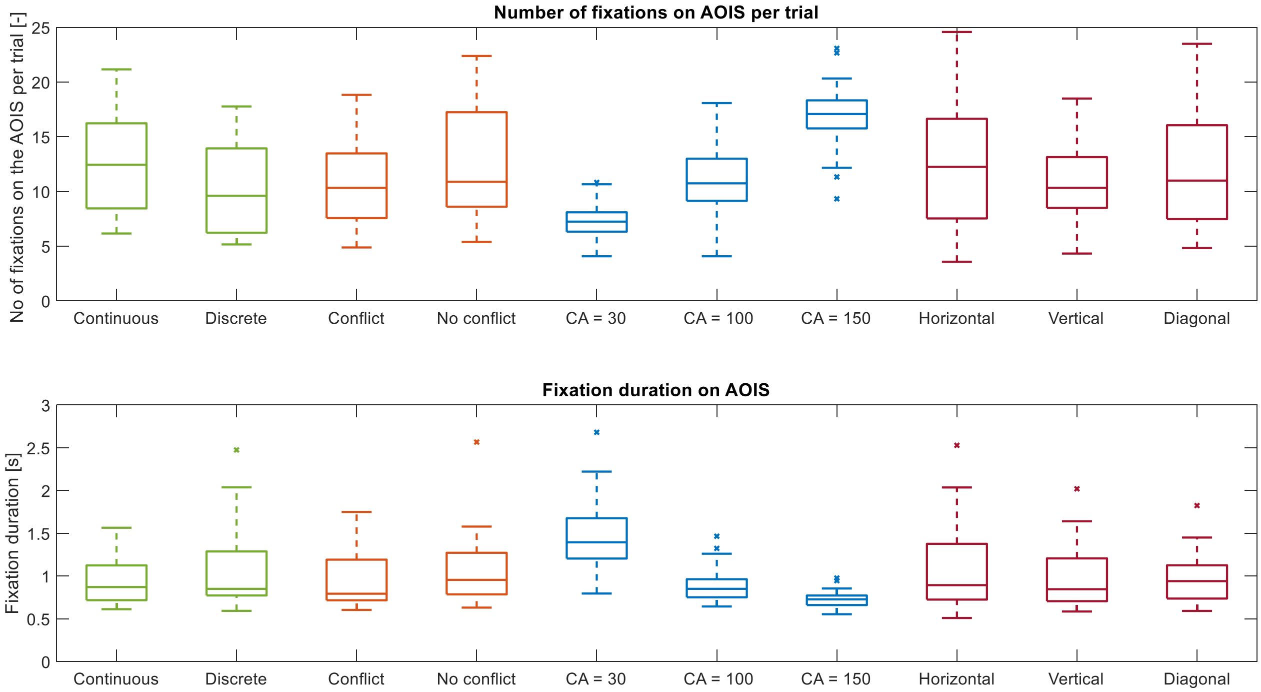


Figure B3: The number of fixations and fixation durations of fixations on AOIs per independent variable. The average per independent variable per participant has been used. So each box plot is created from 35 data points.

B.4 Saccadic duration

Figure B4 shows the saccadic duration for all trials, divided by CA.

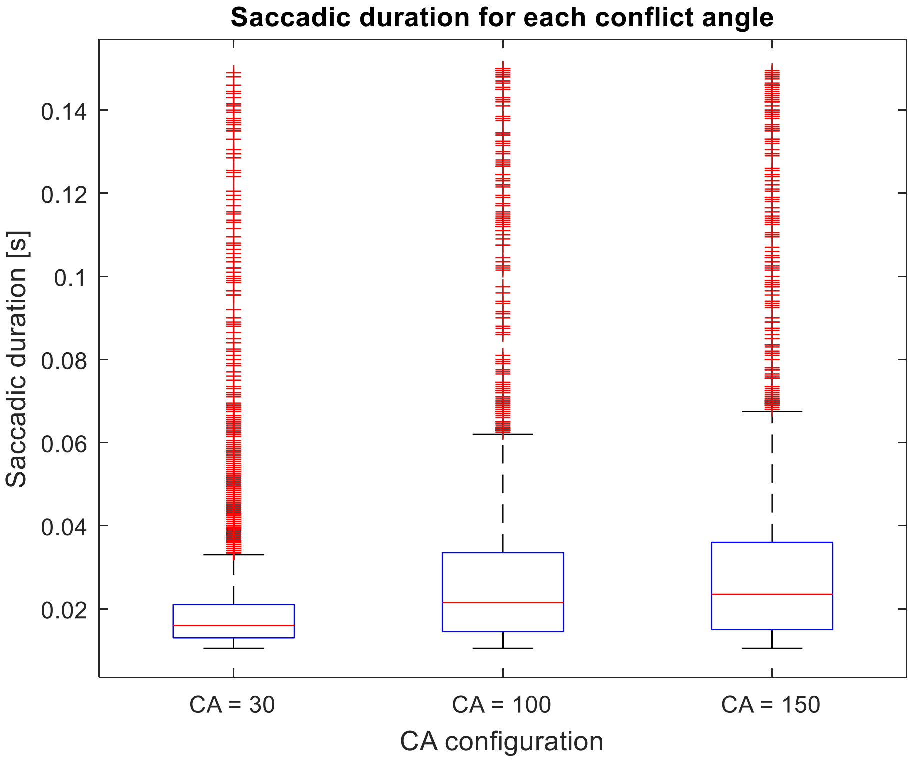


Figure B4: Saccadic duration in seconds per conflict angle.

B.5 Fixations and comparisons over time

Figure B5 shows the percentage of fixations and percentage of comparisons against the trial time. The different conflict angles and conflict/non-conflict scenarios are shown separately. With an angle of 30 degrees, more fixations and comparisons are performed in the first half of the trials compared to the other angles. Furthermore, for the trials with a conflict angle of 30 degrees and a conflict, a higher percentage of fixations and comparisons is reached earlier in the trial compared to non-conflict trials. With a conflict angle of 100 degrees, the percentage of fixations and comparisons is higher for non-conflict scenarios compared to conflict scenarios. The percentage of fixations and comparisons over time shows not much difference between conflict and non-conflict for the trials with a conflict angle of 150 degrees.

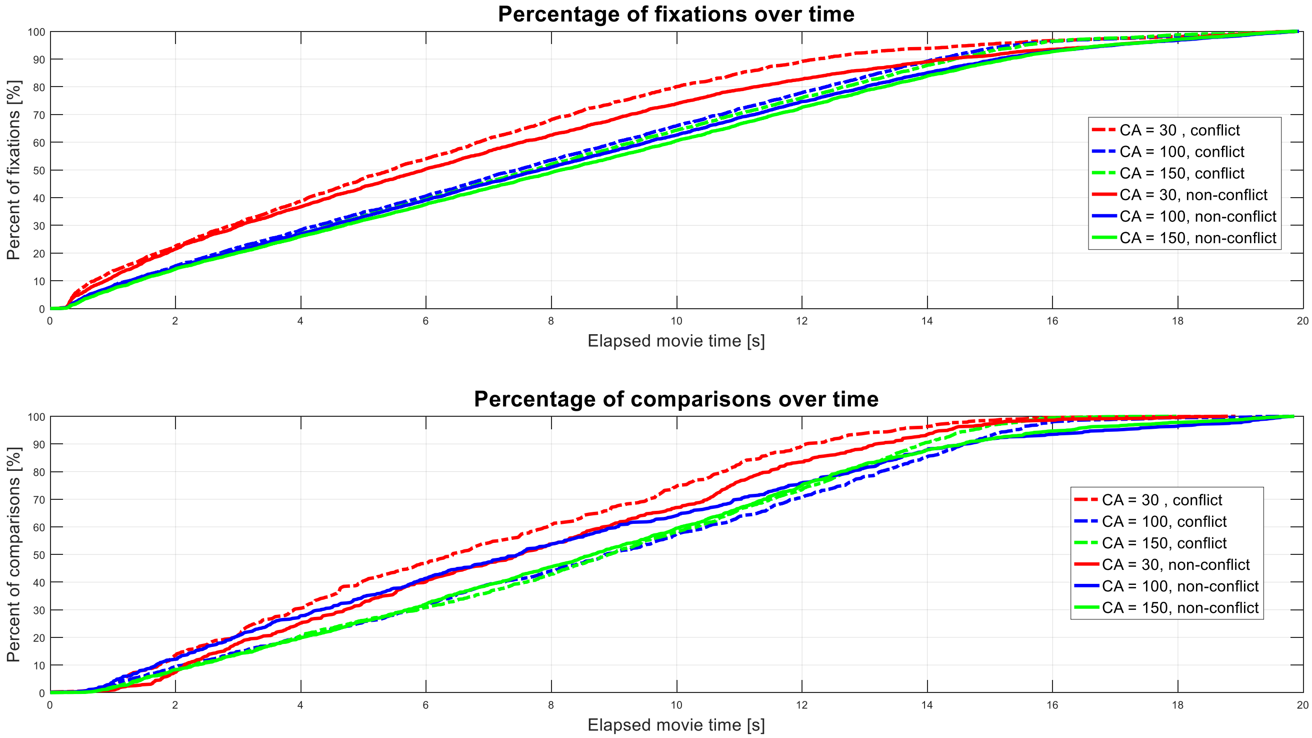


Figure B5: The percentage fixations and the percentage of comparisons are plotted against the trial duration. The different angles of conflict and conflict/non-conflict scenarios are plotted separately. Each graph consists of the data from 35 participants and 6 trials, so 210 data points.

B.6 Pursuit duration over trial time

Figure B6 shows the duration of all pursuit movements against trial time. The fixations of the different conflict angles are indicated with different colors: green shows the conflict angle of 30 degrees, red shows the conflict angle of 100 degrees and blue the conflict angle of 150 degrees. The figure shows that with a conflict angle of 30 degrees, the pursuits are longer and also, that often at the end of the trial a long pursuit movement was performed.

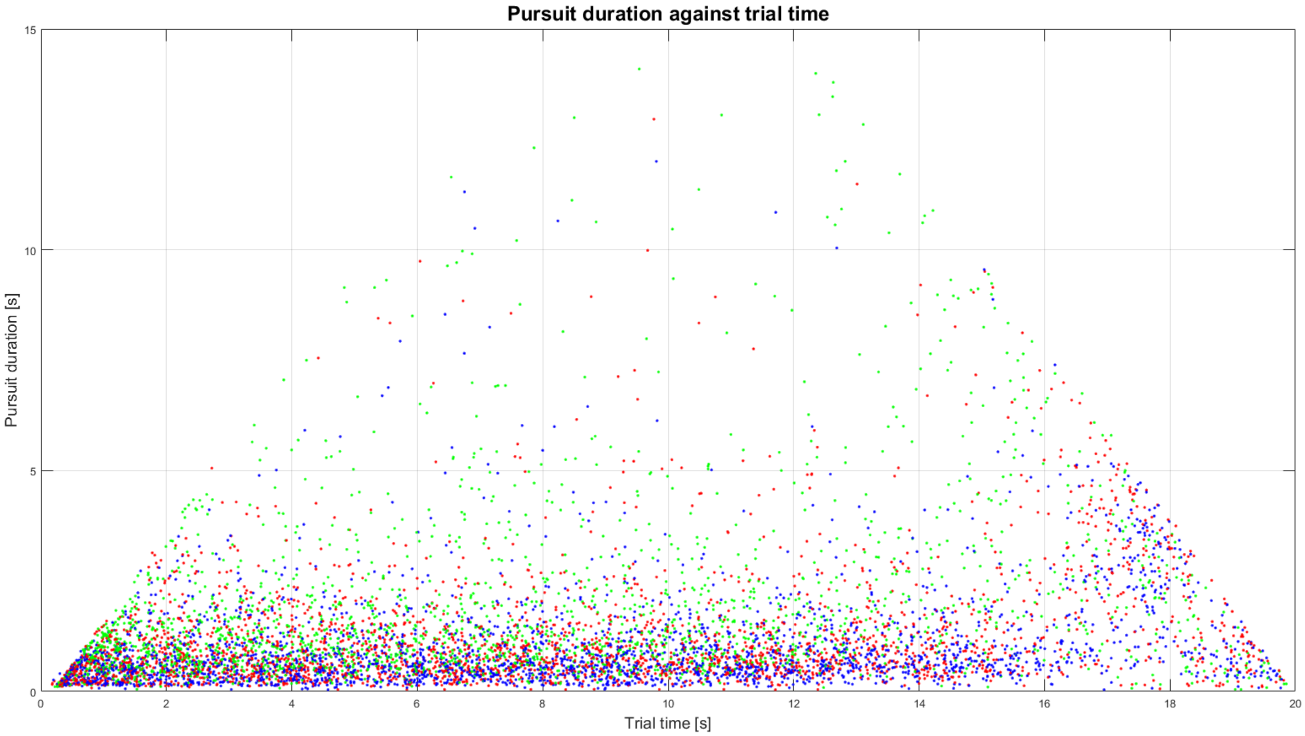


Figure B6: Pursuit duration is showed against the moment in the trial that the pursuit movement occurs. Since the moment in the middle between the previous and next saccade is taken as the moment of fixation, the data points show a triangular shape. The green dots show the fixations with a conflict angle of 30 degrees, the red dots show the fixations with a conflict angle of 100 degrees and the blue dots show the fixations with a conflict angle of 150 degrees.

B.7 Fixations in the xy plane

Figure B7 and B8 show all fixations of one trial in the xy plane to give some insight in how much movement each fixation contained. From these two trials it can be observed that there is some movement in the fixations. Furthermore, the long fixation with pursuit movement at the end of the trial can be seen in these two examples.

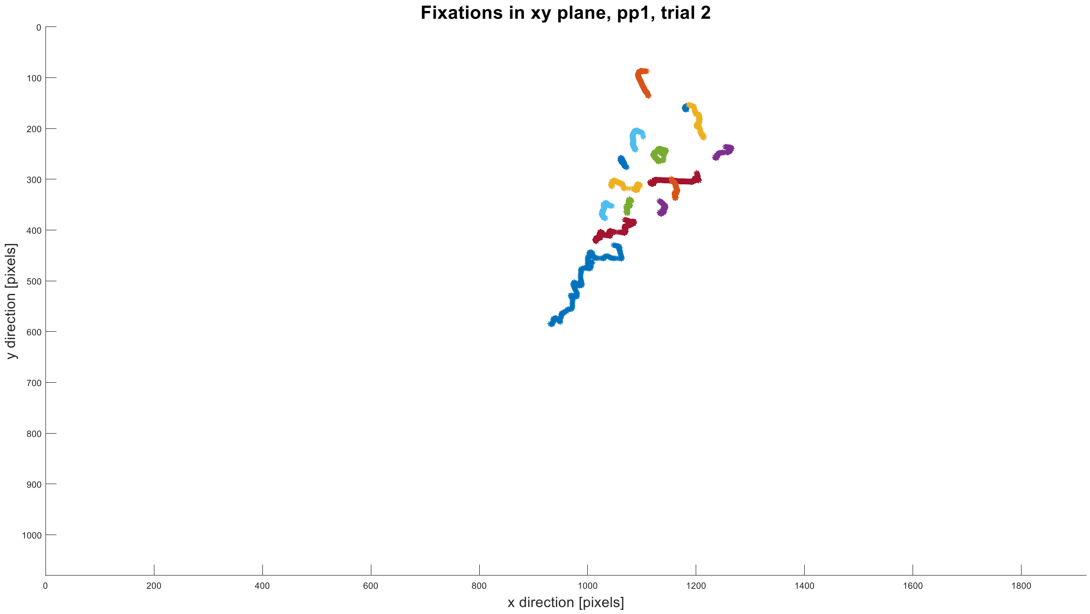


Figure B7: The fixations as found in the second trial of the first participant. Each fixation is indicated with another color.

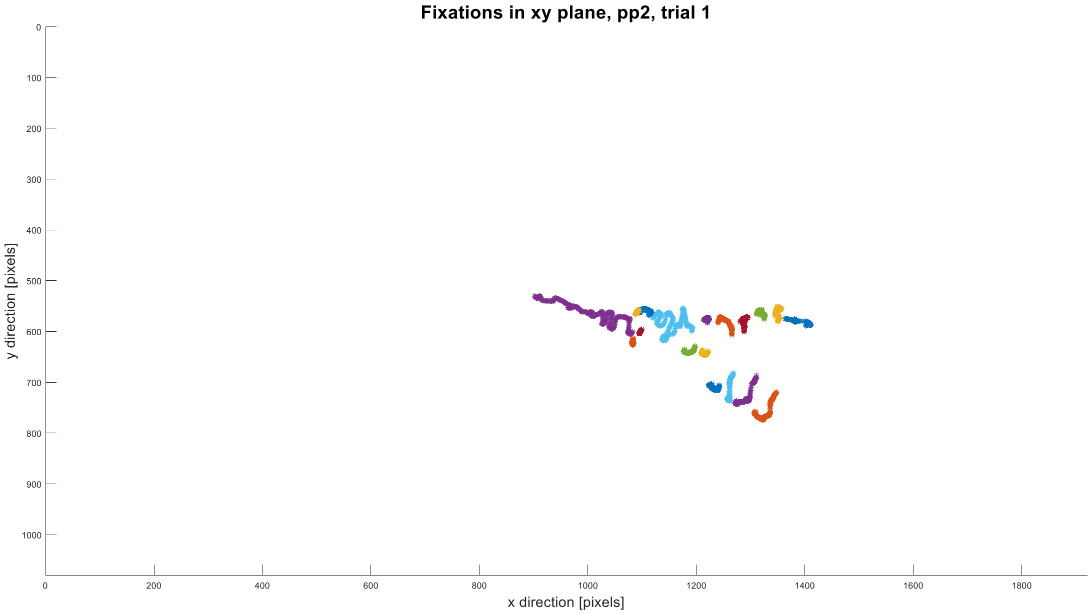


Figure B8: The fixations as found in the first trial of the second participant. Each fixation is indicated with another color.

Appendix C: Written consent form

Researcher: Anouk Looijestijn
Supervisors: Yke Bauke Eisma and Dr.ir. Joost C.F. de Winter
Title: Sampling behavior during a simplified air traffic control task
Date: 04-06-2018

Dear reader,

Thank you for considering to participate in this research. This form contains some relevant information about the study. If you would like to participate, please read this form carefully. There will be opportunity to ask questions before you start the experiment.

Purpose of the research:
This research is designed to measure eye movements during a simplified air traffic control task. Air traffic controllers have to continuously monitor the flight radar to make sure all air traffic remains safely separated from each other. Therefore, they have to be excellent in detecting conflicts between aircraft. To improve the performance of air traffic controllers and obtain general insight into how humans perform collision detection tasks, more knowledge is needed about how humans visually sample while they search for conflicts.

What does participation in this research involve?
Participation in this research involves watching 36 videos which each last 20 seconds. The videos will show two moving dots (representing aircraft), moving towards each other. You will be asked to indicate whether you think the dots will eventually collide or not. During this task, eye movements will be recorded. To measure the eye movements, you will need to place your head on a head support. The experiment starts with a calibration of the eye tracker and one training trial. Halfway the experiment there will be a small break. In total the experiment will take about 30 minutes.

Task description:
During each trial, you have to indicate by pressing the spacebar whether you think the two moving dots would collide if they continue their trajectory, or not. **Keep** the spacebar **pressed** when you think the dots will collide in the current trial. It is allowed to change your mind, so if you have pressed the spacebar you can release it again if you think there will be no collision. It is always about the question: 'is there a collision in this trial'. So if you already saw a collision happening, you should keep the spacebar pressed. After each trial you will be asked how difficult you thought the task was. Next, your score will be displayed. The score is computed as the percentage of time that the spacebar is correctly pressed or released.

Glasses/lenses:
Unfortunately, the eye tracker does not work well with glasses. The data will be less reliable or even useless when wearing them. Therefore, no glasses can be worn during the experiment. If you are not able to see properly at 1 meter distance without your glasses, unfortunately you cannot join this experiment.

Confidentially:
Eye movements will be recorded during this research. Furthermore the conflict detection response and self-experienced difficulty will be recorded. Participants will not be identified by name in any reports. It will be impossible to trace results back to you.

Risk:
The risk related with participating in this study is very small. The head support will be adjusted for each participant so that it fits comfortably. Furthermore, if needed, you can always request for a small pause, apart from the scheduled break.
This study has been reviewed and approved by the Human Research Ethics Committee of the TU Delft on 03-04-2018 with the title: 'Sampling behavior when observing moving stimuli with a converging heading.'

Participation is voluntary
Please note that participation is voluntary. You do not have to take part in this research if you do not wish to. A financial compensation of 5 euros will be given for the time spend. You may stop participating in the research at any time without providing a reason. You will be asked to sign an informed consent form before the start of the study.

Location of the experiment:
The experiment will be conducted in the Eye lab, room F-2-360 at the faculty of 3ME (Building 34, Mekelweg 2) of the TU Delft.

Please feel free to ask questions!

Consent form:

I confirm that I have read and understood the participant information form for the above mentioned study.

I understand that my participation is voluntary. I may withdraw and discontinue participation at any time without giving any reason.

I agree to storage and use of collected data for the purposes of this study. The results of the study will not be made available in a way that could reveal the identity of individuals.

I have had all my questions answered to my satisfaction.

I agree to take part in the above study.

___________________________	___________________________
Signature	Date
___________________________	___________________________
Name	Signature of Investigator

Anouk Looijestijn – MSc student, Mechanical Engineering
Email: a.e.looijestijn@student.tudelft.nl
Tel: 06-16303185

Yke Bauke Eisma – PhD student, Aerospace Engineering
Email: Y.B.Eisma@tudelft.nl

Dr.ir. Joost C.F. de Winter – Supervisor, Mechanical Engineering
Email: j.c.f.dewinter@tudelft.nl

Appendix D: Program design in experiment builder

The experiment builder program from the SR Research Eyelink 1000 Plus Eye tracker was used to create a program that showed the 36 trials in random order while eye movements were measured. The movies were added to the program in xvd format. The split avi function was used to convert them. The program started with a screen that asked the gender and age of each participant. Afterwards the task instructions were shown to the participants. If the participant had no questions, the training trial started. During the training, no eye movements were recorded. Afterward the training, a nine points calibration and validation was performed. If needed, the illumination was adjusted.

If the validation was correct, the program continued with the trial sequence block during which the eye movements were recorded. A drift correction was performed before each new trial. All spacebar presses and releases were recorded during the trials and was immediately used by the program to compute the performance score per trial. The computation of the score was dependent on which movie was shown, since whether the spacebar should be pressed or not, depended on whether there was a conflict or not. Furthermore, the possibility that the spacebar was pressed multiple times was considered when computing the score. After each trial the difficulty question appeared on the screen. Participants could indicate by pressing 0 to 10 on the keyboard, how difficult they thought the trial was. Directly afterwards, the computed performance score was shown.

By counting the number of trials done, it was possible to insert a break halfway the experiment. After the break a new 9 point calibration and validation was incorporated so that the participant could remove his/her head from the support and move around freely. The key q could be used to quit the program at any time. Furthermore, there was always an option to do an extra calibration if the drift correction was really off, or if the participant would need an extra break.

The eye tracking measurements were logged in edf files, from which a sample report was created. The resulting excel files contained the following columns: 1.) session label, 2.) timestamp, 3.) right gaze x, 4.) right gaze y, 5.) right pupil size, 6.) trial start time, 7.) video frame index, 8.) video name and 9.) right in blink. From participant 31 the left eye was measured, so for this participant the left gaze x, left gaze y, left pupil size and left in blink was logged. For each participant a separate excel file with those parameters at each sample is created. Furthermore, for each participant the program gave two .txt files as output. The first one contained data on all spacebar pressed and released and the second file contained the difficulty and performance scores per trial.

Appendix E: Pilot experiments

At the moment that the eye tracking program was working, a pilot study was done to check the experiment. The goal of the pilot was to get some first results and to check whether the experiment was designed correctly. Furthermore, some design variables, like the velocity and size of the dots, had to be tuned.

From the first pilot it was concluded that 36 trials were doable, however a break in the middle was required so that participants could stay focused. Furthermore, it was concluded that 1 practice trial was enough to make sure the task was clear for the participant. Moreover, the total time per participant could be estimated from the pilot. Going through all movies and difficulty questions took about 20 minutes, 10 minutes were taken into account for signing the consent form, asking questions and debriefing.

Some mistakes in both the eye tracking program and the stimuli movies were found. The eye tracking program was optimized such that it was easier to go through all steps and the movies were adjusted so that no information about whether there would be a conflict or not, was given away by the movie structure. Furthermore, the idea raised to add the performance score after each trial, so that participants would stay motivated to perform as best as they could. Some more pilots were done, and it was concluded that the task was difficult enough. Even though the experimental setup was familiar for the pilot participant, it was not possible to predict the outcome of the movies and participants would need to stay focused during each trial to answer correctly. Finally, it was found during the pilots, that starting the trial with pressing the spacebar and then releasing when a conflict was detected was counterintuitive and would lead to unnecessary mistakes. Therefore, this was inverted for the real experiment. The spacebar was not pressed at the beginning of a trial. Participants had to press it when they thought a conflict would occur. This is also more in resemblance with real ATC, where also something has to be done when a conflict occurs.

When everything was eventually working properly, the pilot data was used to obtain some first results. These pilot results did not show anything weird or unexpected and all data that was needed to answer the research question could be obtained. So it was concluded that the experiment was designed correctly and the real measurements could start.

Appendix F: Remarks from participants

After the experiment it was asked to the participants what they thought about the experiment and whether they had any remaining questions. The remarks about the different test conditions are stated below. Note: these remarks were not part of the experiment and were not recorded in a consistent way. The comments of participants were interpreted by the researcher. For example a participant could indicate discrete movement with ‘blinking dots’.

- ‘I thought discrete was easier’ (5x)
- ‘I thought discrete was harder’ (9x)
- ‘I thought configuration with smaller angles were easier’ (5x)
- ‘I thought the intermediate angle was most difficult’
- ‘I thought the largest angle was most difficult’ (2x)
- ‘With smaller angles I could earlier see whether there was a conflict’
- ‘The task was hard but fun to do’
- ‘Sometimes it seemed the velocity was not equal for the dots’ (2x)
- ‘I was not sure whether the dots moved linear’

Appendix G: MATLAB, stimuli design

The following link gives access to the MATLAB files that were used to create the stimuli movies:
https://drive.google.com/drive/folders/1vseIkGs730l1DrzHgS25wOtfNH_0nYFu?usp=sharing

Appendix H: MATLAB, data analysis

A short description of the main files used to perform the data analysis is given in this appendix. The following link gives access to the MATLAB files:
https://drive.google.com/drive/folders/1IMD2_g-1LKJAgWSNghIds1ue0nMwpnJ5?usp=sharing

Import data and convert to .mat files

The MATLAB file ‘Part1_Read_data.m’ loads all the excel files and converts the data to .mat files. The movie title is changed to the movie number (see table 1). Furthermore for each participant the .txt file called ‘RESULTS_FILE_KEYTIMES.txt’ (containing all spacebar presses and releases) is loaded. The data from these files is stored in the .mat file called XSpace.mat. This file contains a matrix with the following 10 columns: 1.) participant number,2.) gender, 3.) age, 4.) trial number, 5.) key pressed, 6.) key released, 7.) key press/release time, 8.) time on display, 9.) EDF time and 10.) video name. The gender input is changed to 1 for female and 2 for male. Furthermore, the fifth column is changed to 1 when the spacebar is pressed and into a 2 when the spacebar is released. Finally, the performance score and self-experienced difficulty is stored into the file called XDiff.mat, which contains a matrix with the following columns: 1.) participant number, 2.) trial number, 3.) self-experienced difficulty score, 4.) performance score and 5.) whether the trial contained a conflict or not (1 indicates a conflict, 2 indicates no conflict). The size of this matrix with 35 participants is 35x36x5 since there is one entry for each trial for each participant.

Eye movement analyses

The MATLAB file ‘Part2_Eye_movement_analysis.m’ first imports the trajectories of the dots for all trials so that the AOI locations can be found. Furthermore, the .mat files as stored by the read data script are imported and the different variables are stored separately as vectors. For each participant it is determined at which frame numbers, each movie was displayed. With those information, a loop is created which loops over all participants and all trials. For each trial, the saccades and fixations are defined as explained in the method section. Furthermore, the fixation location is compared to the AOIs to define which fixations were towards the AOIs. The pursuits and comparisons are also computed in this loop. At the end of the loop, the computed variables are stored in matrices, most of them with size 35x36. Those matrices are directly used for the data analysis in SPSS. After the loop, the dependent variables are grouped per independent variable, to create the graphs as showed in the paper.

Performance and difficulty analyses

Finally, the MATLAB file ‘Part3_Performance_analysis.m’ analyses the performance and difficulty scores. XDiff.mat and XSpace.mat are loaded and the data is sorted per independent variable. This script generates the graph for the difficulty and performance scores, the graph on the learning effect, the spacebar presses over time graph and the graph that shows the number of false positives.

Appendix K: Literature review

A literature review of eye movements in supervisory control

A. E. Looijestijn and J. C. F. de Winter

***Abstract*—As technology keeps developing, humans more often have to supervise automated systems. Information about operators' eye movements may be useful to get insight in their cognitive processes. According to the SEEV model, eye movements in supervisory control are a function of stimulus Salience, required Effort, stimulus Expectancy, and the Value associated with the stimulus. Elsewhere, Wickens et al. (2001) described four key characteristics of eye movements during supervisory control experiments: (1) the process is dynamic, (2) operators look for critical events, (3) the dependent variable is the attention distribution, (4) the operator has to look at the right critical event at the right time. This study aims to examine to what extent eye-movement analyses in 23 empirical supervisory control studies reflect the SEEV model and the four supervisory control characteristics. Results showed that independent variables used in empirical studies are: (1) level of automation, (2) signal bandwidth, (3) task demands, (4) error occurrence, (5) learning or operator experience level, and (6) task context. Time-related variables (e.g. glance duration) and frequency related variables (e.g. glance rate) were the most used dependent variables. The operators' attention distribution was often defined using areas of interest. The dynamic element (i.e. when an operator looks at a particular location) is found in specific situations in which the time element is important, for example when an error occurred at a certain time, the eye movements were compared before and after that time. However, the dynamic element was not found as dependent variable in the empirical studies. It is recommended to investigate whether the SEEV model can be refined with knowledge from empirical data and then use it to judge the dynamic sampling behavior of operators.**

Index Terms—Eye movement, Tracking, Supervisory Control, Visual Sampling, Gaze Pattern

INTRODUCTION

In the last decades, automation has developed considerably (e.g., Hancock, 2014). Consequently, the task of humans has changed from direct control to the control of automation, also known as supervisory control (Sheridan, 1986). Knowledge about how humans process information during supervisory control tasks would be beneficial to prevent human errors and improve human-machine interfaces.

The first models that predict human sampling behavior during a visual sampling task, developed by Senders (1964) and Carbonell (1966), were based on the variables expectancy (i.e., signal bandwidth) and value (i.e., the cost of missing a signal). Wickens et al. (2003) added the variables effort (the amount of eye/head movements required to complete the supervisory task) and salience (distinctness of the stimulus) by developing a descriptive model that outputs the probability that an area of interest (AOI) will be sampled. The salience, effort, expectancy, and value (SEEV) model has been validated in several studies (Horrey, Wickens, and Consalus, 2006; Wickens et al. 2008). Elsewhere, Wickens et al. (2001) described four key features that distinguish supervisory control task models from visual search (i.e., locating a target) models: (1) the process that is supervised is dynamic, not static, (2) operators have to look for critical events instead of targets, (3) the dependent variable is the visual attention distribution, (4) the challenge is not to find a target (where to look), but to know when to look where, so that the dynamic process stays under control.

Models of eye movements in supervisory control, such as the SEEV model, are potentially powerful tools in human-machine systems research. However, it is presently unclear to what extent the empirical literature on eye movements uses knowledge from the SEEV model and whether the results found in empirical literature are in accordance with the SEEV model and the supervisory control characteristics as stated by Wickens et al. (2001). This study aims to compare the variables from current empirical supervisory control studies with the variables as used in the SEEV model, and with the elements of supervisory control as stated by Wickens et al. For this purpose, an overview will be given of the independent and dependent variables used in empirical studies in which eye movements are measured during a supervisory control task.

METHOD

We selected papers describing empirical studies in which eye movements were measured while participants performed a supervisory control task. In order for the study to be included in the review, eye movements had to be measured and the task had to be supervisory control. The participant had to be tasked to check whether automation performs as required, and act if needed. Google Scholar and ResearchGate were used to retrieve scientific material. The following keywords were used: eye tracking, eye movement, visual sampling, and sampling behavior. They were each combined with the following terms: supervisory control, automation, plant operator, human controller, air traffic control. So, a total of 4x5 searches were conducted in both databases. The first 20 articles from each search were studied to see whether they fulfilled the above mentioned criteria. A total of 23 studies that fit the criteria were retrieved.

RESULTS

Independent variables

This section describes all the independent variables that were used in the selected studies. Table 1 provides an overview and shows the number of studies that have used each variable.

3.1. Level of automation/support

Four studies described the effect of different levels of system automation on operator's eye movements. In a study with six participants who had to monitor the tactical flight situation on a display and re-plan if the route was endangered, the level of automation was increased by adding speech input to the information that was given to the operator or by highlighting changes on the display. It was found that with more advanced technical support, the outside window was sampled more often and the navigation task on the secondary display was sampled fewer times (Flemisch and Onken, 2000).

Trouvain and Schlick (2007) studied eye movements of 14 participants during supervision and navigation of two ground based robots. The authors compared a unimodal head-up-display with two multimodal displays. It was found that with the multimodal displays, which used tactile (TAC) or auditory (AUD) information in addition to visual information, the gaze direction moved more towards the peripheral regions, indicating more attention to scanning of the surroundings. In an experiment with nine participants, Fidopiastis et al. (2009) varied the level of automation during the operation of an uncrewed ground vehicle by comparing manual teleoperation navigation with semi-automated waypoint navigation, where the vehicle followed the route by itself. In the manual situation, most fixations were directed to the top down viewing area (navigation overview map), while in the semi-autonomous case (with same task and target recognition aiding system) most glances were to the main viewing area (direct vision from the vehicle). Finally, in a study with 29 participants, Jamson et al. (2013) observed a significant difference in visual attention between manual car driving and supervisory control of an automated driving system. It was found that in the latter case, participants glanced less often (54.0% vs. 74.5% of the time) to the road ahead. The results of these four studies suggest that, when the level of automation is higher, operators have more freedom to look around or look towards a secondary task or display.

3.2. Bandwidth

Two studies that investigated the relationship between signal bandwidth and sampling behavior found that a higher bandwidth signal was sampled more often. Sarter, Mumaw, and Wickens (2007) performed research on the detection of mode transitions on a highly automated commercial flight deck. They found that participants ($n = 20$) looked significantly more often to the raw data (such as airspeed, altitude, and attitude) on higher bandwidth instruments than to low bandwidth flight mode annunciations (FMAs). Dwell duration to the FMAs was found to be shorter than to the other instruments (average of 0.40 s vs. 0.60 s). Eisma et al. (2017) confirmed (replicating work from Senders, 1983) with 86 participants that fast-moving pointers were looked at more often than slowly moving dials. These results are in line with the SEEV model, where expectancy is operationalized as signal bandwidth; according to the SEEV model, a higher expectancy increases the probability that an AOI will be glanced at.

3.3. Task demands

This paragraph presents the effects of task demands on eye movements as described in nine of the 23 studies. To start with, Eisma et al. (2017) investigated physical task demands. Their study indicated that an increase of visual effort causes less ideal sampling: Dials with a higher bandwidth, which are in general sampled more often, are sampled less if they are placed at a position in which the eyes have to move further, i.e., towards the outside of the display. The other eight papers focused on cognitive task demands. Firstly, Manske and Schier (2015) performed an air traffic control simulation study with six experienced air traffic controllers as participants. The task demands were altered between complex and simple, depending on whether the decision could be made independent of other traffic or not. It was found that in the complex case slightly more scans were conducted (9.55 vs. 8.83 AOIs were observed on average). These findings are in line with the conclusions of Imants and de Greef (2014), who used a simplified air traffic control (ATC) simulation experiment with nine participants. These authors found a longer scan path when the task demands were higher. Furthermore, they found that an increase in task demands caused increases of the nearest-neighbour index and spatial density. The nearest neighbour index indicates the ratio between the distribution of fixation points on the screen and the random distribution of points on the screen, whereas the spatial density is an index of the closeness of fixations. Das et al. (2018) found, in a control room experiment, a lower average fixation duration during consistent events (where the system performed in a predictable manner) than during inconsistent events (423 ms vs. 806 ms). The saccade duration increased with task demands for three out of the four participants (41 ms for consistent vs. 44 ms for inconsistent events on average). Ikuma et al. (2014) reported on a control room experiment, with twelve participants, in which safety-critical alarms had to be addressed. During inactive periods (i.e., when no alarm went off), a higher number of simulated events (thus higher task demands) resulted in significantly fewer glances to the main graphical display, more glances to the faceplate (used to adjust variables of the system and appeared only when clicked on) and more glances to the alarm bar. Moreover, Jamson et al. (2013) showed that participants in an automated car allocated more visual attention to the road in the case of a high traffic density as compared to low traffic density, while with manual driving there was no statistically significant difference between high and low traffic density). Furthermore, De Greef et al. (2009) performed a research in which 18 participants had to operate a combat management workstation aboard naval vessels with three different levels of task demands, which were altered by changing the number of airplanes and vessels per scenario and by increasing the number of airplanes and vessels with special or ambiguous behavior. They found that fixation time increased significantly if the task demands increased.

To summarize, one study found less ideal sampling (i.e., under-sampling of high-bandwidth signals) when the physical task demands were higher. Furthermore, an increase in cognitive task demands leads to an increase in fixation duration.

3.4. Error Occurrence

Several studies programmed an error in the automated system to see what the effect of the error would be on eye movements. Moray and Rotenberg (1989) measured eye movements of twelve participants who had to control a simulated thermal-hydraulic system. They found no significant difference in fixation durations with or without error occurrence; however, faulty subsystems were examined more frequently. Furthermore, significantly more glances to the guidance, navigation, and control displays were found by Valerie et al. (2015) in a space shuttle cockpit simulator during nominal runs (with no malfunctions), than during runs with three malfunctions.

Sharma et al. (2016) investigated the effects on eye movements after successes and failure by operators (N = 72). They found that after a control room operator was unable to reject a disturbance, the dwells were more evenly distributed over all AOIs, whereas operators who rejected the disturbance successfully focussed more on areas containing valves that were related to the disturbance. Furthermore, Thomas and Wickens (2004) researched the scanning patterns of eight pilots in a flight simulator with eight flight scenarios, from which one or two had an unexpected event (unannounced blimp or wrongly indicated runway). Participants who did not detect the unexpected event scanned the main information display significantly more often and scanned the outside word less often than participants who did detect the event.

Summarizing, an increase in the number of saccades and more short dwells to various AOIs are found after a mistake was made. With the occurrence of system malfunctions, fewer glances to the control display were found. These effects could indicate a more panicky reaction if an error or incident occurs.

3.5. Learning/expertise

In six studies the relationship between eye movements and the operators' level or experience or learning was examined. Robinski and Stein (2013) measured the visual scanning techniques of 33 participants in a helicopter simulation. Participants were either experienced (flight instructors with an average of 1,800 hours of (simulator-based) flight experience) or inexperienced (student pilots with an average of 150 hours of flight experience). Two conditions were examined: (1) low task demands, landing on a terrain-pinnacle and (2) high task demands, landing on a frigate on the open sea. During take-offs and landing in terrain flight, the experienced group looked more outside than the inexperienced group (outside looks instructors = 68.3%, outside looks students = 48.5%). According to Robinski and Stein, this difference indicated that the flight instructors used their peripheral field of view more effectively. The student pilots performed shorter and more frequent scans to view the environment than the instructors (especially during high workload phases like landing). During the high-workload task of landing at sea the scanning technique was different: In the frigate scenario, instructors dedicated 7.9% of all glances to the outside environment, while for students 18.8% of the glances were to the outside environment. Bjorklund et al. (2006) assessed the difference between captain (commander) and first officer (second in command). The captain visually verified a flight mode transition in 72% of the cases versus 47% for the first officer. Moreover, Eisma et al. (2017) found that when repeating their experiment, attention was less equally distributed over the different dials and more distributed according to the theory of ideal scanning as a function of bandwidth. Valerie et al. (2005) found no significant difference between experience levels (astronaut vs. airline transport pilots) during a spacecraft control task. However, it should be noted that the definition of experience level was questionable (the pilots practised in the task) and the sample size was small (*n* = 11), which could explain why no significant difference was found. An experiment in a control room experiment by Koffskey et al. (2014, N = 9) found that more experienced operators scanned more often to critical areas. Lastly, a study by Wang et al. (2015, N = 25) performed with 25 participants who are categorised into 5 different levels of experience (increasing from only simulator experience till at least twelve years of work experience in air traffic control) an air traffic control task. They found that with more experience a higher number of AOIs is viewed, and the fixation duration is smaller. The five participants with only simulator experience (n = 4) showed higher mean saccadic velocity than the other categories. According to Wang et al. this does not suggest that novices are better in quickly searching critical targets but merely that they select targets randomly and are easily distracted by disturbances.

In summary, it has been found that when participants repeat the same experiment, they become more efficient samplers of visual information. Furthermore, between operators with various experience levels, a difference in scanning technique is found. It seems that with more experience, operators can perform a more efficient scanning strategy, using their knowledge about the system.

3.6. Task Context

The context of the task has an influence on eye movements. For exemple, Valerie et al. (2005) found a significant effect of flight segment on the proportion of fixations on the guidance, navigation, and control displays. The flight segments differed from each other in time duration and task requirements. For example, the first phase consisted mostly of navigation and system parameter checks. Furthermore, Haslbeck (2012) performed a study in which 57 participants had to fly a flight simulator with different flight phases (phase 1: automated descent on runway, phase 2: manual flown approach, 3: go-around manoeuvre, 4: manual landing). The number of glances per area of interest differed significantly for both experienced participants (first officers) and less experienced participants (long haul captains) between flight phases. Similarly, Diez et al. (2001) found a relationship between the phase of the flight and the percentage of looking time to different instruments, such as the mode control panel, the primary display, and the navigational display. Kodappully et al. (2016) investigated eye patterns of 11 participants in a control room. They found a different scanning pattern for each phase of the task. For example, at the start of the experiment, many glances with high durations were found to AOIs with the primary tags (to be used to bring the system back in equilibrium if an error would occur). This scanning behavior could indicate that participants wanted to orientate on the situation. Lastly, Lin, Zhang, and Watson (2003) compared an ecological interface design and a function-behavior-state display and found that the total percentage of fixations to the ecological interface display was significantly higher than to the function-behavior-state display.

Summarizing, the operational context, such as the phase of the task, influences the scanning pattern. How the scanning behavior is influenced is found to be different per application.

Table 1: Overview of the independent variables. Some studies investigated multiple independent variables, that is why the total number of studies is higher than 23.

Independent variable	Used in # studies	References
Level of automation/support	4	Jamson et al. (2013), Flemisch and Onken (2000), Trouvain and Schlick (2007), Fidopiastis et al. (2009)
Bandwidth	2	Sarter, Mumaw and Wickens (2007), Eisma et al. (2017)
Task demands	9	Eisma et al. (2017), Manske and Schier (2015), Imants and de Greef (2014), Das et al. (2018), Ikuma et al. (2014), Jamson et al. (2013), de Greef et al. (2009), Sibley, Doddi and Jasper (2015), Lin, Zhang and Watson (2003)
Error occurrence	4	Sharma et al. (2016), Valerie et al. (2015), Moray and Rotenberg (1989), Thomas and Wickens (2004)
Learning/experience	6	Robinski and Stein (2013), Bjorklund et al. (2006), Eisma et al. (2017), Valerie et al. (2005), Koffskey et al. (2014) , Wang et al. (2015)
Task context	5	Valerie et al. (2005), Haslbeck (2012), Diez et al. (2001), Kodappully et al. (2016), Lin, Zhang and Watson (2003)

Dependent variables

This section investigates which dependent variables are used in emperical research for supervisory control tasks.

Time-related variables

13 of the 23 studies have used a dependent variable that is time related. An example of a time-related variable is the dwell time. The dwell time is defined as: “the sum of consecutive individual fixation and saccade times to an AOI in a single glance” (ISO 15007-1, 2014, p. 5). Kodappully et al. (2016) and Sharma et al. (2016) used dwell time per AOI as the dependent variable in their study. Bjorklund, Alfredson, and Dekker (2006) used the dwell time of glances longer than 150 ms within a visual angle of 1.5° per AOI as a variable. Thomas and Wickens (2004), Diez et al. (2001), Sarter, Mumaw, and Wickens, (2007), Flemisch and Onken (2000) and Eisma et al. (2017) measured percentage dwell time per AOI. Moray and Rotenberg (1989) used both the dwell time and proportion of time allocated to each area. Furthermore, Eisma et al. (2017) and Sarter, Mumaw, and Wickens (2007) used the dwell time per glance (in seconds).

Apart from the dwell time, also the glance duration is used as dependent variable. A glance is defined by (ISO 15007-1, 2014) as a period in which the visual gaze is located within one AOI and can consist of a combination of multiple fixations and saccades. The glance duration is the dwell time plus the transition time (between AOIS or accommodation of the eyes). The summation of all glance durations is called the total glance time (ISO 15007-1, 2014). Haslbeck and Diez et al. (2001) used glance duration per AOI, and Wang et al. (2015) used the mean glance duration in general (so not per AOI) as a dependent variable. Ikuma et al. (2014) used percentage time on the AOIS. Lastly, Das et al. (2018) and de Greef et al. (2009) used fixation durations as a measure. It was defined by the Greef et al. as “the time that fixations lasted within a radius of 40 pixels and a minimum of 100 milliseconds”, (de Greef et al. (2009), p. 6).

Summarizing, various time-related variables are used as dependent variable, such as the dwell time, the glance duration and the fixation duration.

Frequency-related variables

The number of glances per time unit (glance rate) is used in combination with predefined AOIs to assess how frequently operators are looking at a certain AOI (Eisma et al. (2017), Manske and Schier (2015), and Robinski and Stein (2013). Robinski and Stein and Trouvain and Schlick (2007) counted the total number of fixations and Valerie et al. (2005) used the proportion of glances to the GNC displays. Jamson, Merat, Carsten, and Lai (2013) used the proportion of glances within the centre of the road. Lin, Zhang, and Watson, (2003) and Fidopiastis et al. (2009) used the percentage of the number of fixations per AOI. Bjorklund et al. (2006) used percentage of fixations to AOI during a mode transition. Wang et al. (2015) used the number of AOIs looked at.

Scan path analysis

The order in which AOIs are glanced at can be determined with a scan path analysis. A scan path can be visualized by accumulating the scanning data as a 2D plot with either the raw data or the fixations and saccades (BeGaze user manual, 2014). There are different ways to analyse the scan path, for example with Markov models. Markov models use probability distributions for sequences of AOI transitions to model the scan path (Salvucci and Goldberg, 2000), showing the overall transition probabilities among AOIs. Koffskey et al. (2014) used the scan path to compare participants eye movements in their research. Imants and de Greef (2014) used the scan path length (summed distance between all fixations) in an air traffic control simulation task as an indicator of task demands.

Other

In four supervisory control studies dependent variables have been found which could not be classified in one of the above categories. In order to give a complete overview, these variables are mentioned in this paragraph. Saccadic duration (not per AOI but in general) has been used as a dependent variable by Das et al. (2018). Wang et al. (2015) used the saccadic velocity as a dependent variable. Imants and de Greef (2014) used the nearest neighbour index and the spatial density as a dependent variable. De Greef et al. (2009) used saccade distance and saccade speed but did not find a significant effect on task demands with these dependent variables. So, there are other dependent variables used in the empirical research about supervisory control, however, in the 23 studies that are compared in this review they did not occur more than once.

Overview

Table 2 shows an overview of the number of studies in which each independent variable was used. The total number is higher than 23 since some studies used multiple dependent variables. It can be concluded that in practice the glance rate and dwell time are the most used dependent variables for supervisory control tasks in the included studies. The saccadic duration and peak saccadic velocity, as distinguished as dependent variables by Sharma et al. (2016) were not often used as dependent variables. These dependent variables were probably not relevant for most applications, because, during saccades, only little visual processing is achieved (Rayner and Keith, 1998).

Table 2: Overview table of the dependent variables

Independent variable	**Used in # studies**	**References**
Time related • Dwell time • Fixation duration • Glance duration • Total time	13	Bjorklund, Alfredson and Dekker (2006), Thomas and Wickens (2004), Kodappully et al. (2016), Eisma et al. (2017), Haslbeck (2012),Moray and Rotenberg (1989), Wang et al. (2015), Das et al. (2018), Eisma et al. (2017), de Greef et al.(2009), Flemisch and Onken (2000), Diez et al. (2001) and, Sharma et al. (2016)
Frequency related • Glance rate • Total no of fixations • Proportion of glances • Number of AOIs looked at	9	Eisma et al. (2017) Lin, Zhang and Watson (2003), Ikuma, Robinski and Stein (2013),Fidopiastis et al. (2009), Manske and Schier (2015), Wang et al. (2015), Bjorklund et al. (2006) Trouvain and Schlick (2007), Valerie et al. (2005) and, Jamson, Merat, Carsten and Lai (2013)
Scan path • Scan path length • Scan path visualisation	2	Koffskey et al. (2014), and, Imants and de Greef (2014)
Other • Saccadic duration • Saccadic velocity • Nearest neighbour index • Spatical density • Saccadic distance • Saccadic speed	4	Das et al. (2018), Wang et al. (2015), Imants and de Greef (2014) and, De Greef et al. (2009)

CONCLUSION

This study aimed to combine knowledge about eye movements during supervisory control by examining to which extent eye-movement analyses in empirical supervisory control studies reflect the SEEV model and the four supervisory control characteristics of eye movement (models) as defined by Wickens et al. (2001); i.e., (1) the process is dynamic, (2) operators look for critical events, (3) the dependent variable is the attention distribution, (4) the operator has to look at the right critical event at the right time.

Independent variables

It is found that in 23 empirical studies on supervisory control, there are six independent variables for which eye movements are used to express the response of operators on different conditions. In decreasing amount of occurrence, they are: tasks demands (9x), learning/experience (6x), task context (5x) , level of automation (4x), error occurrence (4x) and bandwidth (2x). The studies on the level of automation suggest that operators looked less to the main control task when the level of automation or technical support was higher. Flemisch and Onken found that the highlighting of signals resulted in slightly fewer glances to those signals instead of more, probably because it was easier to comprehend the data. Furthermore, a higher bandwidth signal was found to be looked at more often, which is in correspondence with the expectancy variable in the SEEV model. Moreover, higher cognitive task demands were found to lead to a higher fixation duration. More experience, on the other hand, leads to a more efficient division of visual attention.

Except from the expectancy, the other independent variables are not directly relatable to the variables in the SEEV model, however, also no contradictions to the model are found in the results of the empirical studies. The variables bandwidth and error occurrence reflect the dynamic property of supervisory control as mentioned by Wickens et al. (2001).

Dependent variables

Almost all supervisory control studies used a time or frequency related dependent variable. A time-related measure like dwell time, fixation duration or glance duration is used in 13 of the 23 studies. The number of glances or fixations is used 9 times, 2 studies used a scan path to analyse the data, and 4 studies used other dependent variables.

The dependent variable for eye movements in supervisory control according to Wickens is the attention distribution. This was also the case in almost all studies by measuring the time per AOI or the glance rate per AOI. We found that the attention distribution is often reported per AOI. This relates to the finding that the frequency related dependent variables are important. The frequency of fixations need to be expressed towards certain areas. The importance of AOIs in eye tracking studies should be considered when designing an empirical study.

The dynamic element of supervisory control (as used in the definition of Wickens et al., 2001) was not found as a dependent variable in any of the reviewed studies, except in combination with time related events. For example, Sarter, Muman, and Wickens (2007) examined whether a fixation was conducted within 10 seconds of a transition or not. Similarly, Sharma et al. (2016) investigated ‘when an operator looked where’, to comprehend the data of one participant in relation to a system failure at a certain time. However, this analysis was not used as a measure to summarize the results of all participants. Instead, they used the dwell time distribution per group per AOI as a dependent variable, thereby ‘losing’ the moment in time the fixations occurred. Moray and Rotenberg (1989) reported at which time the gaze was at which location to see what the difference was before and after the fault. Here, the percentage time on AOIs was compared before and after the fault. Summarizing, the dynamics of eye movements were considered when the effect of an error occurrence was investigated; however, it was not used as a variable to describe eye movements.

To conclude, this review provides an overview of the empirical research done on eye movements for supervisory control tasks. The variable expectancy from the SEEV model, as well as the second and third characteristic as stated by Wickens, are directly found back in the variables in empirical studies. The dynamic element as stated by Wickens is found in some of the studies, but was not found to be crucial to describe eye movements during supervisory control. Furthermore, also independent variables are found that are not used in the prediction of the SEEV model. This means there is potential to use and further develop the SEEV model with knowledge from empirical data. The refined models could be used to judge and predict the dynamic eye-movements of participants. Furthermore, this overview could provide insight for researchers who want to set up a supervisory control experiment using eye tracking. It provides an insight to which independent variables are often used in eye movement experiments and to which extent the theories and models about eye movements during supervisory control are in agreement with empirical studies.

Finally, this literature overview contained a comparison of only 23 selected studies; more research is needed to establish the effect of the different independent variables on eye movements during supervisory control. Those relations could then, together with the SEEV model, be used to predict eye movements across experiments and to get a better understanding of how an operator is performing based on eye movements.

REFERENCE LIST

Backs, R. W., & Walrath, L. C. (1992). Eye movement and pupillary response indices of mental workload during visual search of symbolic displays. *Applied Ergonomics*. https://doi.org/10.1016/0003-6870(92)90152-L

Bjorklund, C., Alfredson, J., & Dekker, S. (2006). Mode Monitoring and Call-Outs: An Eye-Tracking Study of Two-Crew Automated Flight Deck Operations. *The International Journal of Aviation Psychology*, *16*(3), 263–275. https://doi.org/10.1207/s15327108ijap1603_2

Boag, C., Neal, A., Loft, S., & Halford, G. S. (2006). An analysis of relational complexity in an air traffic control conflict detection task. *Ergonomics*, *49*(14), 1508–1526. https://doi.org/10.1080/00140130600779744

Braddick, O. (1974). A short-range process in apparent motion. *Vision Research*, *14*(7), 519–527. https://doi.org/10.1016/0042-6989(74)90041-8

Carbonell, J. R. (1966). A queueing model of many-instrument visual sampling. *IEEE Transactions on Human Factors in Electronics*, *HFE-7*(4), 157–164. https://doi.org/10.1109/THFE.1966.232984

Das, L., Iqbal, M. U., Bhavsar, P., Srinivasan, B., & Srinivasan, R. (2018). Toward Preventing Accidents in Process Industries by Inferring the Cognitive State of Control Room Operators through Eye Tracking. *ACS Sustainable Chemistry & Engineering*, *6*, 2517–2528. https://doi.org/10.1021/acssuschemeng.7b03971

de Greef, T., Lafeber, H., van Oostendorp, H., & Lindenberg, J. (2009). Eye Movements as Indicators of Mental Workload to Trigger Adaptive Automation. https://doi.org/10.1007/978-3-642-02812-0

DeLucia, P. R., & Liddell, G. W. (1998). Cognitive Motion Extrapolation and Cognitive Clocking in Prediction Motion Tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 901–914. https://doi.org/10.1037/0096-1523.24.3.901

Duerrschmid, K., & Danner, L. (2018). Eye Tracking in Consumer Research. *Methods in Consumer Research*, *2*, 279–318. https://doi.org/doi.org/10.1016/B978-0-08-101743-2.00012-1

Eisma, Y. B., Cabrall, C. D. D., & De Winter, J. C. F. (2017). Visual Sampling Processes Revisited: Replicating and Extending Senders (1983) Using Modern Eye-Tracking Equipment. *IEEE Transactions on Human-Machine Systems Manuscript*.

Ellerbroek, J. (2013). *Airborne conflict resolution in three dimensions*.

Eyferth, K., Niessen, C., & Spaeth, O. (2003). A model of air traffic controllers' conflict detection and conflict resolution. *Aerospace Science and Technology*, *7*(6), 409–416. https://doi.org/10.1016/S1270-9638(03)00064-6

Fidopiastis, C. M., Drexler, J., Barber, D., Cosenzo, K., Barnes, M., Chen, J. Y. C., & Nicholson, D. (2009). Impact of automation and task load on unmanned system operator's eye movement patterns. *Foundations of Augmented Cognition, Neuroergonomics and Operational Neuroscience*, 229–238. https://doi.org/10.1007/978-3-642-02812-0_27

Flemisch, F. O., & Onken, R. (2000). Detecting Usability Problems with Eye Tracking in Airborne Battle Management Support. *RTO MP*, *57*(23).

Hancock, P. A. (2014). Automation: how much is too much? *Ergonomics*. Taylor & Francis. https://doi.org/10.1080/00140139.2013.816375

Haslbeck, A. (2012). The relationship between pilots ' manual flying skills and their visual behavior : a flight simulator study using eye tracking, (November 2015), 3131–3138.

Hildreth, E. C., & Ullman, S. (1982). Ad-a128 398 the measurement of visual motion(u) massachusetts inst.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

Hunter, A. C., & Parush, A. (2009). Using Eye Movements to Uncover Conflict Detection Strategies. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, *53*, 1729–1733. https://doi.org/10.1518/107118109X12524444081872

Ikuma, L. H., Harvey, C., Taylor, C. F., & Handal, C. (2014). A guide for assessing control room operator performance using speed and accuracy, perceived workload, situation awareness, and eye tracking. *Journal of Loss Prevention in the Process Industries*, *32*, 454–465. https://doi.org/10.1016/j.jlp.2014.11.001

Imants, P., & de Greef, T. (2014). Eye Metrics for Task-Dependent Automation. In *Proceedings of the 2014 European Conference on Cognitive Ergonomics - ECCE '14*. https://doi.org/10.1145/2637248.2637274

ISO 15007-1. (2014). Road vehicles - Measurement of driver visual behavior with respect to transport information and control systems - Part 1: Definitions and parameters. Geneva, Switzerland: International Standards Organization.

Jamson, A. H., Merat, N., Carsten, O. M. J., & Lai, F. C. H.

(2013). Behavioral changes in drivers experiencing highly-automated vehicle control in varying traffic conditions. *Transportation Research Part C: Emerging Technologies*, *30*, 116–125. https://doi.org/10.1016/j.trc.2013.02.008

Kilingaru, K., Tweedale, J. W., Thatcher, S., & Jain, L. C. (2013). Monitoring pilot "Situation Awareness." *Journal of Intelligent and Fuzzy Systems*, *24*(3), 457–466. https://doi.org/10.3233/IFS-2012-0566

Kodappully, M., Srinivasan, B., & Srinivasan, R. (2016). Towards predicting human error: Eye gaze analysis for identification of cognitive steps performed by control room operators. *Journal of Loss Prevention in the Process Industries*, *42*, 35–46. https://doi.org/10.1016/j.jlp.2015.07.001

Koffskey, C., Ikuma, L. H., Harvey, C., & Aghazadeh, F. (2014). Using Eye-Tracking to Investigate Strategy and Performance of Expert and Novice Control Room Operators. In *Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting* (pp. 1667–1671). https://doi.org/10.1177/1541931214581348

Lin, Y., Zhang, W. J., & Watson, L. G. (2003). Using eye movement parameters for evaluating human-machine interface frameworks under normal control operation and fault detection situations. *International Journal of Human Computer Studies*, *59*(6), 837–873. https://doi.org/10.1016/S1071-5819(03)00122-8

Loft, S., Bolland, S., Humphreys, M. S., & Neal, A. (2009). A Theory and Model of Conflict Detection in Air Traffic Control: Incorporating Environmental Constraints. *Journal of Experimental Psychology: Applied*, *15*(2), 106–124. https://doi.org/10.1037/a0016118

Mackintosh, M.-A., Dunbar, M., Lozito, S., Cashion, P., Mcgann, A., Dulchinos, V., … Gent, R. Van. (1998). Self-Separation from the Air and Ground Perspective. *USA/Europe Air Traffic Management R&D Seminar*, (December), 1–16.

Manske, P. G., & Schier, S. L. (2015). Visual Scanning in an Air Traffic Control Tower - A Simulation Study. *Procedia Manufacturing*, *3*, 3274–3279. https://doi.org/10.1016/j.promfg.2015.07.397

Moray, N. (1986). Monitoring Behavior and Supervisory Control. In *Handbook of perception and human performance, Vol. 2* (pp. 40-01, 40–46).

Moray, N., & Rotenberg, I. (1989). Fault management in process control: eye movements and action. *Ergonomics*, *32*(11), 1319–1342. https://doi.org/10.1080/00140138908966910

Neal, A., & Kwantes, P. (2009). An Evidence Accumulation Model for Conflict Detection Performance in a Simulated Air Traffic Control Task. *Human Factors*, *51*(2), 164–180. https://doi.org/10.1177/0018720809335071.

Oprins, E., Burggraaff, E., & Weerdenburg, H. Van. (2006). Design of a Competence-Based Assessment System for Air Traffic Control Training. *International Journal of Aviation*, *8414*(3), 297–320. https://doi.org/10.1207/s15327108ijap1603

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans.*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

Rantanen, E. M., & Nunes, A. (2005). Hierarchical Conflict Detection in Air Traffic Control. *The International Journal of Aviation Psychology*, *15*(4), 339–362. https://doi.org/10.1207/s15327108ijap1504

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Remington, R. W., Johnston, J. C., Ruthruff, E., Gold, M., & Romera, M. (2000). Visual search in complex displays: factors affecting conflict detection by air traffic controllers. *Human Factors*, *42*(3), 349–366. https://doi.org/10.1518/001872000779698105

Robinski, M., & Stein, M. (2013). Tracking Visual Scanning Techniques in Training Simulation for Helicopter Landing. *Journal of Eye Movement Research*, *6*(2:3), 1–17. https://doi.org/10.16910/jemr.6.2.3

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the Symposium on Eye Tracking Research & Applications - ETRA '00*, 71–78. https://doi.org/10.1145/355017.355028

San Agustin, J. (2010). Off-the-Shelf Gaze Interaction. *Communication*, (September), 179.

Sarter, N. B., Mumaw, R. J., & Wickens, C. D. (2007). Pilots' monitoring strategies and performance on automated flight decks: an empirical study combining behavioral and eye-tracking data. *Human Factors*, *49*(3), 347–357. https://doi.org/10.1518/001872007X196685

Senders, J. W. (1964). The Human Operator as a Monitor and Controller of Multidegree of Freedom Systems. *IEEE Transactions of Human Factors in Electronics*, *HFE-5*(1).

Senders, J. W. (1983). *Visual sampling processes*.

Sharma, C., Bhavsar, P., Srinivasan, B., & Srinivasan, R. (2016). Eye gaze movement studies of control room operators: A novel approach to improve process safety. *Computers and Chemical Engineering*, *85*, 43–57. https://doi.org/10.1016/j.compchemeng.2015.09.012

Sheridan, T. (1984). Supervisory contol of remote manipulators, vehicles and dynamic process: Experiments in command and display aiding. In *MASSACHUSETS INST OF TECH CAMBRIDGE MAN/MACHINE SYSTEMS LAB*.

Sheridan, T. (1986). Human supervisory control of robot systems. *Proceedings 1986 IEEE International Conference on Robotics and Automation*, *3*, 808–812. https://doi.org/10.1109/ROBOT.1986.1087506

SMI cooperation. (2014). BeGaze User Manual.

Stankovic, S., Raufaste, E., & Averty, P. (2008). Determinants of conflict detection: a model of risk judgments in air traffic control. *Human Factors*, *50*(1), 121–34. https://doi.org/10.1518/001872008X250584.

Thomas, L. C., & Wickens, C. D. (2004). Eye-tracking and Individual Differences in off-Normal Event Detection when Flying with a Synthetic Vision System Display. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *48*(1), 223–227. https://doi.org/10.1177/154193120404800148

Thomas, L. C., & Wickens, C. D. (2006). Display Dimensionality, Conflict Geometry, and Time Pressure on Conflict Detection and Resolution Performace Using a Cockpit Display of Traffic Information. *The International Journal of Aviation Psychology*, *16*(3), 321–342.

Trouvain, B., & Schlick, C. M. (2007). A Comparative Study of Multimodal Displays for Multirobot Supervisory Control. *Engin. Psychol. and Cog. Ergonomics, HCII*, 184–193.

Valerie, A., Huemer, M. S., Hayashi, M., Renema, F., Elkins, S., McCandless, J. W., & McCann, R. S. (2005). Characterizing Scan Patterns in a Spacecraft Cockpit Simulator: Expert Vs. Novice Performance. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 49, pp. 83–87). https://doi.org/10.1177/154193120504900119

Wang, Y., Cong, W., Dong, B., Wu, F., & Hu, M. (2015). Statistical analysis of air traffic controllers ' eye movements. In *11th USA/Europe Air Traffic Management Research and Development Seminar* (pp. 1–9).

Wickens, C. D., Goh, J., Helleberg, J., Horrey, W. J., & Talleur, D. A. (2003). Attentional Models of Multitask Pilot Performance Using Advanced Display Technology. *Human Factors, 45*(3), 360–380. https://doi.org/10.1518/hfes.45.3.360.27250

Wickens, C. D., Helleberg, J., Goh, J., Xu, X., & Horrey, W. J. (2001). *Pilot Task Management: Testing an Attentional Expected Value Model of Visual Scanning*.

Wickens, C. D., McCarley, J. S., Alexander, A. L., Thomas, L. C., Ambinder, M., & Zheng, S. (2008). Attention-Situation Awareness (ASA) Model of Pilot Error. In *Human performance modeling in aviation* (pp. 213–239). https://doi.org/10.1201/9781420062984.ch9

Xing, J. (2006). Color and visual factors in ATC displays, (June), 1–15.

Xu, X., & Rantanen, E. (2003). Conflict detection In air traffic control: A task analysis, a literature review, and a need for further research. *Proceedings of the International Symposium on Aviation Psychology*, (March).

APPENDIX: OVERVIEW TABLE WITH THE 23 ARTICLES AND THEIR DEPENDENT AND INDEPENDENT VARIABLES

Table 3: Overview of the articles in which eye movements are measured during a supervisory control task

Authors	**Year**	**Type**	**Nr of pp.**	**Dependent variables**	**Independent variables**
Bjorklund, Alfredson and Dekker	2006	Aviation	12	dwell time, scanpath	level of experience
Das et al.	2018	Control room	4	dwell time, saccadic duration	task demand
De Greef et al.	2009	Combat management	18	dwell time	task demand
Diez et al.	2014	Aviation	5	glance rate per AOI, dwell time per AOI	operational context
Eisma, Cabrall and de Winter	2017	Dials	86	glance rate per AOI, glance freq per AOI per sec	bandwidth, task demand, level of experience, operational context
Fidopiastis et al.	2009	robot vehicle	9	glance rate per AOI	level of automation
Flemisch, Onken	2000	Aviation	6	glance rate per AOI	level of automation
Haslbeck	2012	Aviation	57	percent dwell time per AOI	operational context
Ikuma, Harvey, Taylor et al.	2014	Control room	12	glance rate per AOI	task demand
Imants, de Greef	2014	ATC	8	scan path length, spatial density, nearest neighbour index	task demand
Jamson, Merat, Carsten and Lai	2013	Driving	49	percent glances to road center	level of automation, task demand
Kodapully, Srinivasan, Shrinivasan	2016	control room	11	percent dwell time per AOI	operational context
Koffskey et al.	2014	Control room	9	scanpath	level of experience
Lin, Zhang and Watson	2003	Control room	20	glance rate per AOI,	task demand, operational context (interface type)
Manske, Schier	2015	ATC	6	scanpath	task demand
Moray, Rotenberg	1989	control of hydraulic system	12	dwell time, proportion time to AOI	error occurrence
Robinski, Stein	2013	Aviation	33	glance rate per AOI	level of experience
Sarter, Mumaw and Wickens	2007	Aviation	20	glance rate per AOI	bandwidth
Sharma et al.	2016	Control Room	72	glance rate per AOI, mean dwell time per AOI	error occurrence
Thomas, Wickens	2004	Aviation	8	percent dwell time per AOI	error occurrence
Trouvain, Schlick	2007	Robots	14	scanpath	level of automation
Valerie et al.	2005	Spacecraft	11	glance to one AOI	error occurrence, level of experience, operational context
Wang, Cong, Dong et al.	2015	ATC	25	mean dwell time, saccadic velocity	level of experience