CRANFIELD UNIVERSITY

TÂM LE MINH

# THE INVESTIGATION OF TEXT, NUMBERS AND SYMBOLS IN HUMAN-COMPUTER INTERACTION BY ANALYSING VISUAL BEHAVIOUR

CRANFIELD UNIVERSITY
SAFETY AND HUMAN FACTORS IN AVIATION

MSc. Safety and Human Factors in Aviation
Academic Year: 2015 - 2016

Supervisor: Dr. Wen-Chin Li
August 2016

CRANFIELD UNIVERSITY

CRANFIELD UNIVERSITY

SAFETY AND HUMAN FACTORS IN AVIATION

MSc. Safety and Human Factors in Aviation

Academic Year 2015 - 2016

TÂM LE MINH

# The investigation of text, numbers and symbols in human-computer interaction by analysing visual behaviour

Supervisor: Dr. Wen-Chin Li

August 2016

This paper is submitted in partial fulfilment of the requirements for the degree of MSc. Safety and Human Factors in Aviation

# The investigation of text, numbers and symbols in human-computer interaction by analysing visual behaviour


Student: Tâm Le Minh

Supervisor: Dr. Wen-Chin Li

*MSc. Safety and Human Factors in Aviation. Academic Year 2015-2016*
*Cranfield Safety and Accident Investigation Centre*
*Cranfield University – UK*



ABSTRACT: The modern commercial aircraft flight decks are particularly sophisticated human-machine interfaces. The number of functions and the information available to the crew is huge. In order to limit the likelihood of human error, the different displays have to give the crew the best situation awareness possible requiring a minimal mental workload. There are different ways to represent visual information and nothing indicates that they lead to similar performances. For example, numbers are mainly used to represent quantitative information such as the airspeed or the altitude, text is notably used for messages and symbols are signs that have to be interpreted. This paper's aim is to investigate the differences among these three forms of information representation on the operator's mental workload. Twenty-four aviation professionals were faced with videos recorded on a flight simulator. The cockpit included the primary flight display, the navigation display and the EICAS display, each of which design are respectively more focused on numbers, symbols and text. The participants' eye-tracking data on the three displays were collected and analysed with particular metrics. The mean fixation duration, the number of fixations and the fixation time on each AOIs are three rather commonly used metrics. This paper provides an extension of the pupil dilation theories, the mean pupil size on each AOI, used as a metric. However, the data analysis showed that two of the other metrics, the fixation number and the fixation time, are too affected by the tasks to be interpreted in this experiment. On the other hand, the mean fixation duration and the mean pupil size were significantly higher for the EICAS display than the two other displays. No conclusive differences were found between the PFD and the ND. These results suggest that the cognitive processing of text required longer time and led to higher mental workload than the processing of numbers or symbols. This can be explained by the fact that words consist in a combination of letters which are symbols themselves, but in contrast to figures or other symbols, they do not carry any meaning alone. Complete words, then sentences have to be processed to make sense. This confirms the design of most of the HMIs in aviation and other domains, which tries to minimise the amount of text in their displays.


## 1 INTRODUCTION

The human-machine interfaces (HMI) are a focus in human-computer interaction studies as they are central to the functioning of whole system. Such a system consists of human agents and the machine used which can also be considered as one agent. The performance of the system depends on the different agents' ability to communicate between each other: not only between the different members of the team operating the machine but also between the machine and the team. The design of human-machine interfaces may consider several functions, such as giving information on the state of the system, assisting the team in their tasks or even supporting them in their decision making process. This paper is focused on the commercial aviation domain, especially the cockpit displays.

In a highly automated environment such as a modern commercial aircraft, the HMIs play an even more important part in the interactions as it is necessary to keep the team informed of the actions taken by the automation. The primary flight display (PFD) contains all the basic flight information on the state of the aircraft and the autopilot. The airspeed, the altitude, the attitude, the heading of the aircraft can be seized immediately, just like the autopilot modes on the flight mode annunciator (FMA). The navigation display (ND) acts like a radar gathering information on the relative position of the aircraft to the flight path, beacons, other aircrafts or waypoints. The engine indication and crew alerting system (EICAS) display shows information on the engines and display alert messages if needed. Although the necessary information is given to the pilots, the displays also help them in the piloting and the decision making process with messages. Thus, a "PULL UP" message will appear on the PFD if the aircraft is too low on altitude or in overspeed. Also, in the approach phase, when it comes to aligning with the localizer or the glide slope, pink diamonds or dots are displayed in the bottom of and next to the attitude indicator to show in which direction the aircraft has to correct its trajectory to help the pilot to decide of their inputs on the control column.

These functions are strongly related to the three levels of Endsley's model of situation awareness (1995). Indeed, the three level are not exclusive but each level needs its lower ones. The level 1 of situation awareness is the perception of the elements in the current situation, the level 2 is the comprehension of the current situation and the level 3 is the projection of the future status. The perception of the elements is necessary to the comprehension of the situation. Similarly, the projection of the future status can only be done if the situation is comprehended. The layout of the instruments is designed to facilitate the good perception of the different elements needed at every moment (Newman and Greeley, 2001). One example is the PFD T-shape layout, which is a standard in aviation and used in both commercial and general aviation. The pilot will always know where to look when they need to know the airspeed or the altitude. Also, the same scan T-pattern can be used to capture all the relevant information when monitoring the aircraft for example. The alert messages will help the pilots in the understanding of the situation and the projection of the future state.

However, to reduce the risks of human error and therefore improve the safety, not only the display should be designed to give the pilots a good situation awareness, but also to minimise their mental workload. Lin and Lu (2016) found a strong relationship between helicopter pilots' mental workload and their situation awareness. An excessive mental workload would lead to a poor situation awareness due to the pilots' inability to perceive the relevant information. In parallel, the more information is made available to them in order to improve their situation awareness, the more mental workload will be needed to perceive all of it. A typical commercial aircraft cockpit (fig. 1) is greatly sophisticated and finding the necessary information at the right time is sometimes challenging (Dornheim, 1995). There is a balance that needs to be found, between displaying conveniently all the relevant information depending on the situation and removing the information that is not needed. The information must be easy to find (or noticed in the case of an alert). The second step after localising the information is to interpret it.

One piece of information can be represented in multiple ways. In the case of a numerical value like the airspeed, there are several dimensions of representation. First, the value itself can be displayed numerically (e.g.: "airspeed = 150 kt"), but it can be removed in favour of a circular gauge as it is analogically done for cars or general aviation aircrafts, or a digital gauge like in modern commercial aircraft cockpits (Lindo et al., 2012). Though it is highly improbable, the numerical value could have been displayed in words (fig. 2). Also, the colour, shape, size and police font are other dimensions to consider in the interpretation (Whitfield and Whiltshire, 1990). This paper is focusing on the first dimension, in identifying the differences between text, numbers and symbols and their influence on the operators' mental workload from eye-tracking data.

To achieve this objective, the three displays described above are considered. The ND consists mainly in symbols. The EICAS consists mostly in symbols of engine indications in a normal situation, but it has been modified so when crew alerting messages appear, the checklist instructions are also displayed (figure). This makes the EICAS more based on text information. Finally, the PFD also contains a lot of symbols, but when monitored, the most important information are numbers such as the airspeed or the altitude (fig. 3).


Figure 1. Modern commercial aircraft cockpit. There are many places where the relevant information could be.

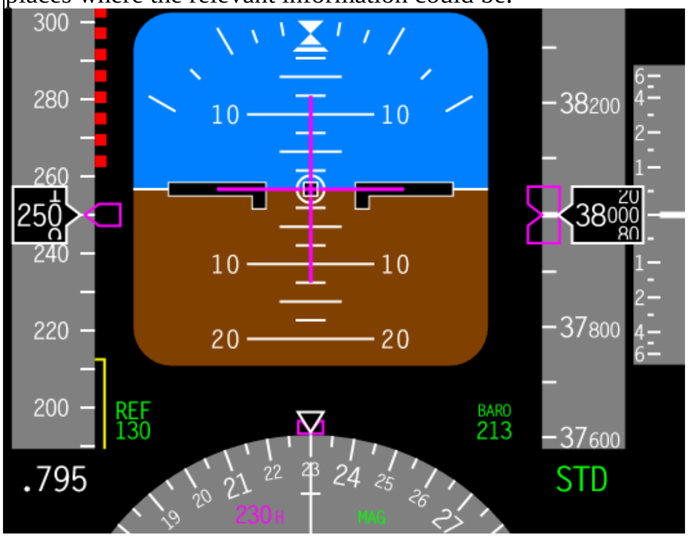


Figure 2. Possible PFD design. The target airspeed and altitude are displayed with words. Common sense makes this kind of design highly unlikely.

Figure 3. Main cockpit displays. From left to right: the PFD, the ND and the EICAS display.

## 2 LITERATURE REVIEW

### 2.1 *Introduction to eye-tracking*

Although its technology was not as developed as now, eye-tracking was already used for over than 100 years (Poole and Ball, 2006). Starting with direct observation of the subject's eyes, the methods have evolved with the creation of diverse eye-tracking devices using electro or magneto-oculographic techniques. These techniques were highly invasive for a long time but the current devices are not as restrictive anymore, mostly based on head-mounted or fixed cameras using the "corneal-reflection/pupil-center" method (Goldberg and Wichansky, 2003). In order to assess where the user is looking, the eye-tracker uses an infrared light directed into the eye. As the light enters the retina, it is mainly reflected, making the pupil look brighter to the infrared camera. In addition to that, a part of the light is reflected on the cornea, and its reflection image can also be detected. With both data, a computer software is able to compute the gaze direction from the centre of the pupil and the corneal reflection image.

Knowing where an individual is looking, it is possible to determine where they direct their attention. Just and Carpenter (1976) assumed that the object being fixated is the reflection of "what is at the top of the stack" of the information processing. As a consequence of this, they proposed the eye-mind assumption, stating that "there is no appreciable lag between what is being fixated and what is being processed". Although this hypothesis could be argued because objects can be perceived with the peripheral vision, Posner (1980) highlighted how the peripheral vision is used to direct the attention, thus the point of gaze, more than to simply process the information. Hence, the observation of the gaze point can not only indicate if an individual has paid attention to an object but how quickly they directed their attention to it and how long it stayed on it.

### 2.2 *Eye movement classification*

To facilitate the analysis of the eye-tracking data, the eye movements are separated into several categories. Indeed, the eye does not sweep smoothly from object to object like it could be assumed for reading, but its movement consists in several stops on the objects called fixations. In between each of them, there is a quick phase of transition to move the point of gaze called saccades. However, there is no general consensus on the criteria to distinguish the fixations from the saccades. In fact, it depends on the type of study and how the fixations or saccades are meant to be interpreted.

In psychology, fixations are the reflections of information processing. They are defined by a stability of the gaze position. Thus their two main characteristics used to identify them are the low velocity of the gaze point or the low spatial variation. Therefore, two of the most largely used algorithms to distinguish them are velocity-based or dispersion-based. Salvucci and Goldberg (2006) reviewed them alongside with other more sophisticated algorithms such as the Minimum Span Tree identification or the Hidden Markov Model fixation identification algorithms and concluded that although the latter ones were more sophisticated, they do not always lead to significantly better results. In this study, the identification is dispersion-based.

There are two threshold to determine in order to implement the algorithm: a dispersion threshold and a duration threshold. That means that in order to be considered as a fixation, the gaze point has to stay within the dispersion threshold for longer than the duration threshold. Marshall (2000), as well as Goldberg and Kotval (1999), stated that the minimum duration of a fixation to be considered is 100 ms. Although Kilingaru et al. (2013) acknowledged that a fixation time of 200 ms is required to actually perceive information, they called the fixations below 200 ms as "glances", which represent brief looks with not enough time for recognition. They also distinguished the fixations with longer duration than 600 ms as "stares" which are signs of misperception.

For the dispersion threshold, the areas chosen are more diverse in the literature. It is possible to use a circle with a radius of 0.5° (Hoffman and Subramaniam, 1995) or 1° (Liang et al., 2007). Marshall (2000) used a square of 1° by 1°, Regis et al. (2012) a square of 2° by 2°. Jacob and Karn (2003) mentioned a threshold of "typically ~2°". Goldberg and Kotval (1999) tried several values between 2° and 4°

for the threshold and adjusted it when there are too few or too many fixations. Salvucci and Goldberg (2000) described a new way to calculate the dispersion as the sum of the vertical and horizontal dispersion ($D = [\max(x) - \min(y)] + [\max(y) - \min(y)]$) and recommended to use a threshold between 0.5° and 1°, though it can also be adjusted during the data analysis.

This paper does not make the distinction between sub-categories of fixation and uses a 100 ms duration threshold to identify the fixations. The dispersion area chosen is a circle of 1° radius around the centroid of the different points of gaze contained in the fixation. Also, it only considers fixations and saccades by not taking into account other eye movement types as glissades/dynamic overshoots (Nyström and Holmqvist, 2010), smooth pursuits (Liang et al., 2007) or blinks (Regis et al., 2012).

### 2.3 *Limitations of eye-tracking*

Even with a very simple model of eye movements only consisting of fixations and saccades, an eye-tracking study has to face several limitations. Jacob and Karn (2003) listed some of them in all the phases of the study: the technical experiment, the data extraction and the data analysis.

First, there is not any eye-tracking device that can suit every participant. Even for good quality devices, there are at least 10% of the participants, even with a good sight, whose eyes cannot be tracked reliably. The diversity of eye shapes and the range of pupil size make it difficult to physically design a universal device. In addition, glasses or contact lenses add refractions that are detrimental to the reliability of the data. For a head-mounted eye-tracker, the shape of the face can also be a problem for some participants. For a fixed eye-tracker, the camera can easily lose track of the participant's eyes.

The eye-tracker, in order to correctly collect the data, has also physical limitations. Some eye-trackers require the participant to not move their head and their body, which does not fit to some usability studies because it would make the tasks different to real life. Similarly, with head-mounted eye-trackers, the individual has to be at some distance from the target objects. These limitations are to be taken into account when questioning the validity of the experiment. Besides, Poole and Ball (2006) insisted on how eye-tracking data is likely to be contaminated if the experiment environment control is poor. Visual distractions must be avoided and the lighting must be constant, especially if the pupil size is to be analysed (Di Nocera et al., 2007).

Another limitations are the technical limitations of the cameras themselves. It is necessary to ensure that the cameras specifications are sufficient for the study. In some study fields such as reading and when analysing saccades, the sample rate needed is higher than a normal usability study. For these situations, Collewijn (1999) recommended a sample rate of 500 Hz although Jacob and Karn (2003) argued that 250 Hz is sufficient. However, the study described in this paper is focused on fixation analysis, where 25 Hz would be enough (Regis et al., 2012).

### 2.4 *Common metrics*

In order to be interpreted, metrics should be used to analyse the eye-tracking data. These metrics need be relevant to the tasks and the topic. In this study, the main interest revolves around mental workload and cognitive processing. A lot of them can be found in the literature.

One of most used is the average fixation duration. As fixations are assumed to be the reflection of the processing of information, the more time is spent looking at an element, the more time is spent understanding this information (Just and Carpenter, 1976). Goldberg and Kotval (1999) used it to indicate how difficult an element in a display was to interpret. Callan (1998) found that longer fixations are related to an increasing workload in a flight simulator research task. Indeed, as this metric is directly related to the cognitive processes, it is among the most relevant ones.

The number of fixations (on an AOI or overall) is also frequently used. As an element which is more often fixated means that it is more important, this metric is task-dependent when used to compare different AOIs. Goldberg and Kotval (1999) used it overall to measure the search efficiency as a larger number of fixations may reveal a poorly designed layout.

The fixation time or rate (on each AOI) can be used to compare the importance of each AOI. The fixation time is strongly related to each of the two metrics described above because it is a combination of them (Jacob and Karn, 2003). The fixation time is obtained by adding the fixation durations up, but it is the same to multiply the average fixation time with the number of fixations. The fixation rate is the percentage of time spent fixating elements within the AOI, but when the span used to analyse the data is constant, it is the same thing, obtained by just dividing the fixation time with the total time of the experiment.

The fixation/saccade ratio is defined by Goldberg and Kotval (1999) as a comparison between the time spent processing and the time spent searching. It is the raw ratio between the total fixation time and the total saccade time. Thus a higher ratio can either mean more difficulty to interpret the elements or faster search for the relevant information, but additional metric is needed to confirm. For this reason, this metric is not as meaningful as the fixation or saccadic rate because it confounds the both of them, but it can be applied to the assessment of reading performances. Indeed, when the text is more conceptually difficult, the average saccade length decreases while the fixation duration increases (Rayner, 1998).

There are also other metrics less commonly found in the literature but that need to be highlighted. Harris et al. (1986) suggested to use the spatial distribution of the fixations to measure the mental workload. They defined the visual entropy similarly as the entropy used in thermodynamics to assess the randomness of a visual scan, which they showed to be negatively related to the mental workload: more mental workload means less randomness. However, Di Nocera et al. (2007) discussed this measure approach based on randomness as they used another metric measuring randomness. The nearest neighbour index (NNI) is meant to measure how clustered or sparse a scatter plot is. Applied to aviation, they found using the NNI that a more random distribution is related with higher mental workload requiring flight phases such as the take-off or the landing phases, in contrast with the cruise phase.

Some studies also used saccades analyses to measure the efficiency of displays, although this is rarer due to the necessity of using a higher speed eye-tracker than for fixation analysis. Wang et al. (2014) evaluated three different designs of a military aircraft head mounted display (HMD) instrument using different symbologies. They used the amplitude and the length of the saccades to assess the efficiency and acceptability of the displays, alongside with a qualitative approach analysing the trajectory of the saccades. Di Stasi et al. (2011a) found that there is a consistent relation between the saccade (spatial) amplitude, the saccade length (in time) and their peak of velocity (PV, maximum velocity during the saccade). This relation called Main Sequence is determined experimentally and varies accordingly to the mental workload. They also insisted on the fact that saccades not voluntarily controlled, so they may be more accurate reflections of the variations in mental state than the fixations, although they are also influenced by external factors such as natural vigilance fluctuations or sleep deprivation (Di Stasi et al., 2011b).

Finally, the pupil size is one most reliable eye parameter used to measure mental workload. Beatty (1982) showed that pupillary response can be considered as an effective measure of cognitive processing load, because it suits the three criteria stated by Kahneman (1973) for physiological indicators of processing load: within-task, between-task and between-individual variations can be exploited. However, several papers (Poole and Ball, 2006, Di Nocera et al., 2007) emphasises the difficulty to collect valid pupil size data, a good experimental control is mandatory as it is very sensitive to lighting and can easily be contaminated. Poole and Ball (2006) also pointed out the large individual differences which have to be considered when analysing the data.

### 2.5 *Application to aviation psychology*

The assessment of mental workload and the measure of cognitive processing are not the only applications of eye-tracking and some others are useful to this study. Eye-tracking data are relevant to perception studies (Wilder et al., 2009). Duchowski (2002) reviewed the main research domains where eye-tracking was used as a research technology such as reading (McConkie and Raner, 1975, Rayner, 1998, Reichle et al., 1998), driving (Chapman and Underwood, 1998, Dishart and Land, 1998, Ho et al., 2001), marketing/advertising (Lohse, 1997, Wedel and Pieters, 2000, Rayner et al., 2001) and computer software displays (Byrne et al., 1993, Goldberg and Kotval, 1999, Goldberg and Wichansky, 2002). Even in aviation psychology, eye-tracking data have been used to study the usability of some instruments or scan patterns to raise the design or training needs.

As an example, Kasarkis et al. (2001) determined the differences between expert and novice pilots scanning strategies in relation with their performance. Sarter et al. (2007) extended usability theories to the cockpit panel by studying the FMA. While Goldberg and Kotval (1999) suggested that the eye “naturally fixate upon surprising, salient or important through experience areas” and addressed the influence of factors like the density or the colour of the information on the displays, Sarter et al. (2007) identified the factors that made pilots effectively notice and understand changes in the FMA, a crucial point in aviation situation awareness.

With a different approach, Thomas and Wickens (2004) balanced the qualities of including too many compelling symbols, no matter how important the represented information was, because they are also causing attentional tunnelling. With an eye to the implementation of real-time analysis during the flight, Kilingaru et al. (2013) and Regis et al. (2012) proposed real-time methods to detect situations where the pilots’ situation awareness is threatened by studying their attention.

These studies opened more possibilities up for eye-tracking studies to improve the flight deck design. As a result of this literature review, this study's hypotheses incorporate the test of the different metrics in relation with its objectives, with the exception for the pupil size, which has to be evaluated as a mental workload measurement first because of its sensitiveness. The null hypotheses are:

- There is no correlation between the pupil size and the perceived mental workload.
- There is no difference in the average fixation duration between the processing of symbols, numbers and text.
- There is no difference of pupil size between the processing of symbols, numbers and text.
- There is no difference of fixation numbers between the processing of symbols, numbers and text.
- There is no difference in fixation time between the processing of symbols, numbers and text.

# 3 METHOD

## 3.1 *Participants*

24 aviation professionals (12 males and 12 females) were contacted to conduct the experiment. They were aeronautical engineers (12) or academics (12) at Cranfield University. Their age range was from 21 to 50 years old (M = 27.5 year, SD = 7.0). As the experiment did not require an extensive knowledge of piloting, no flight experience was needed. However, they must know the basics of aviation. Therefore, 20 of them declared to have flight experience, including license holders (PPL, glider pilots and helicopter pilots). Their average flight experience was 154 hours (minimum = 0, maximum = 3000 h), even though 4 of them almost did not have any.

They were recruited by local advertisement within Cranfield University. Before the start of the experience, they were asked to give their informed consent after being given a description of the experiment. They were not given any financial compensation nor incentive for their participation.

## 3.2 *Ethics approval*

The project had been approved by a commission of ethics (CURES/1524/2016) with a level 2b of risk: "Risks to the researcher/participant are no greater than those typically encountered in ordinary life". As stated in the consent form filled by the participants, the data collected by the eye-tracker and the post-experiment interviews were confidential. If they wished so, they could freely refuse to answer any question, they could end the experiment at any time, and they could ask to withdraw their data at any moment even after the data collection.

## 3.3 *Apparatus*

### 3.3.1 *Flight simulator*

The flight simulator used during the experiment and to make the videos relied on the software Microsoft Flight Simulator X (FSX). The computer itself was mounted on a Saitek Pro Flight Yoke with a Three-lever Throttle, although only one of the levers was actually used (the black one as the thrust lever). Both the flight yoke and the levers were cautiously fixed on a table to avoid any kind of incident involving excessive human inputs.

The aircraft used is a Boeing 777-200 LR/F from the PMDG 777-200 package. The aircraft was developed in collaboration with real-life 777 crews and maintenance teams, making its cockpit accurate and reproducing nearly all the functions of a real cockpit with a high fidelity.

### 3.3.2 *Eye tracking device*

The eye tracker used was designed by Pupil Labs. It consisted in a plastic frame destined to be worn like glasses. Two cameras were attached to it: a world camera and an eye camera. The world camera was fixed on the outside part of the frame, in front of the forehead. It captured the field of view of the participant, following their head movements as the frame moved with it. Its resolution was 1280 x 720 pixels for a sample rate of 60 frames per second. The eye camera was fixed on the side of the frame, below the participant's right eye. It was meant to capture the participant's pupil and its movements. To detect the pupil, it was equipped with an infrared illumination and an infrared detector. An infrared light was projected on the pupil and the infrared detector captured the pupil and the corneal reflection in accordance with the "corneal-reflection/pupil-center" method (Goldberg and Wichansky, 2003) described in the literature review section.

The software used consists in two open-source programs created by Pupil Labs: Pupil Capture and Pupil Player. Pupil Capture was used to record the world camera video and the pupil position and therefore necessary to the data acquisition. Pupil Player computed the gaze position from the data collected by Pupil Capture, hence it allowed the visualisation of the video captured by the world camera with a cursor representing the gaze position. It could also be used to calculate the fixations and to define and track the

AOIs. Only Pupil Capture was used during the experiment, only Pupil Player was used during the data exportation and analysis.

## 3.4 *Research design*

### 3.4.1 *Scenarios*

.The participants were asked to watch 4 videos of 90 seconds recorded from a flight from London Heathrow Airport (London, United Kingdom) runway 27R to Paris-Charles-de-Gaulle Airport (Roissy-en-France, France) in the flight simulator. These 4 videos consisted of different scenarios of emergent situation in which the EICAS display had been modified. In each video, the emergent event occurred at a specific but different time. Normally, the Boeing 777 pilot had to notice the CAS message on the EICAS display first, and then look at the instructions on another display. But for the experiment, the instructions had been added directly under the CAS message when needed, so the pilots did not have to look elsewhere (fig. 4). This new design was referred as “integrated EICAS design”. Between the different videos, the view inside the cockpit remained unchanged, so the position and the size of the displays were actually the same through all the videos (fig. 5).

*Scenario 1: Left engine fire*

The aircraft had just taken-off from London Heathrow Airport (2,500 ft) and was still climbing. At the beginning of the scenario, the aircraft was turning to the left according to the flight plan (runway heading was 272, the route to follow is at 137) but the turn was finished 10 seconds before the emergent event, which occurs at 0:45. Two main signals were given to the crew: an audio alert which was repeated every 11 seconds and a red master warning light. The message displayed on the EICAS is FIRE ENG L in the red colour. The single engine fire would ultimately result in a dissymmetry between the engines exhaust gas temperature which grows as the scenario continues. (fig. 5)

*Scenario 2: Cabin depressurization*

Like the previous scenario, the aircraft was still in the climbing phase but in a later moment. The aircraft was already at 12,000 ft, had finished his turn and was climbing straight forward. At 1:10, the message DOOR ENTRY 1L appeared in amber, and was replaced 1 second after by CABIN ALTITUDE in red (fig. 6). It went along with a permanent audio alert, the red master warning light and the cabin pressure information appearing at the bottom of the EICAS display.

*Scenario 3: Right engine failure*

In this scenario, the aircraft was already in the cruise phase. It was flying straight forward, levelled at a 10,000 ft altitude. At 0:20, the right engine rotation speed and exhaust gas temperature decreased dramatically. At 0:25, the amber EICAS message appeared: ENG FAIL R. (fig. 7). At the same time, the amber master caution light and a brief audio alert could be perceived. Unlike the two previous scenarios, the audio alert only rang once and did not stay until the end of the scenario. Also, it was possible to detect the failure thanks to the sound of the engines, one of which was shutting down.

Figure 4. EICAS modification: on the left, the normal EICAS, on the right, the “integrated” design.

Figure 5. Cockpit view used in the videos.

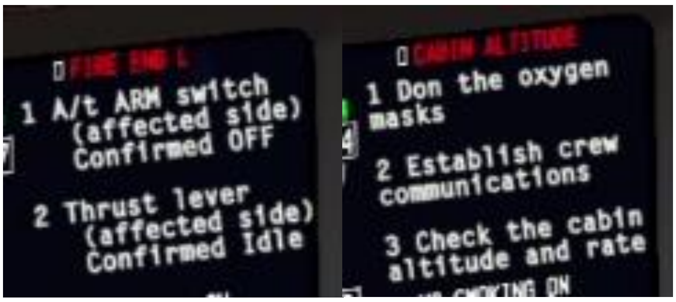


Figure 6. Checklist instructions for scenarios 1 and 2.

Figure 8. The experiment environment.

*Scenario 4: Right generator failure*

This scenario was taken in the same conditions than the first one: the aircraft had just taken-off, it was climbing and turning to the left. At 00:40, the CAS message ELEC GEN OFF R appeared (fig. 7). However, there was no other indication of the failure: no audio alert, no master light and no other signs.

### 3.4.2 *Experiment environment*

The experiment took place in an isolated room with constant lighting, especially important when it comes to analyse the pupil size because of its high sensitiveness to light variations. The room had two parts, one where the main experiment took place, one where a trial took place. In the experiment part, the participant was sat in front of a desk. On this desk, there was a 23 inch monitor with a resolution of 1080p on which the videos were projected. The participant wore the eye tracker and sat approximatively 1 meter away from the screen. The eye-tracker was plugged to a MAC computer for the data acquisition. For the data analysis and the areas of interest identification, specific screen markers (models downloaded from Pupil Labs website) were stuck on the corners of the screen (fig. 8).

In the other part of the room used for the trial, there was a table with the flight simulator fixed (the computer, the flight yoke and the levers as mentionned above). Figure 9 shows all the hardware used during the experiment and the trial.

### 3.4.3 *Procedure*

The experiment was divided in 6 major parts.

It started with an introduction to the experiment or "pre-briefing". The participant was explained the experiment procedure as a 6 part-experiment, the 6 parts being those described in this section. The participant was asked to read thoroughly and to sign the consent form. Also, they were asked if they were short-sighted, in which case they were not qualified to perform as the experiment needed them to sit away from a monitor and still be able to read relatively small instructions.

Then the trial session could start, where the participant was asked to perform a take-off in FSX with a PMDG Boeing 777-200, in the same configuration as in the scenarios. The main display elements of the flight deck were presented: the primary flight display containing the flight mode annunciator, the airspeed tape, the attitude indicator, the altitude tape, the vertical speed gauge and the heading indicator, the navigation display containing the heading, the aircraft relative position to planned flight path, the traffic, beacons and terrain indications, the EICAS display containing the engine rotation speed, the engine exhaust gas temperature and the CAS messages. The aircraft was on runway 27R and the participant had to push the thrust lever and then pull the yoke when the airspeed reaches 181 kt which was the rotation speed. After taking off, the participant was asked to follow the flight director which was programmed to perform the climbing phase to 10,000 ft. The aim of this trial was not only to let them be familiar with the 777-200 cockpit instruments but also to demonstrate how CAS alerts work. To ensure that the participant would face at least one message, a failure "HYD PRESS PRI L" was programmed and armed but some participants also saw other messages depending on how they performed the take-off. The most common ones were "TAIL STRIKE", "LOW AIRSPEED" and "OVERSPEED".

The third part of the experiment was the briefing. The participant was explained what they have to do while watching the videos. They would have to monitor the PFD and the ND until a CAS message appeared, then they would have to read out the message. In case of multiple messages, they should only take the first line into account. After that, they should read the checklist instructions on the EICAS screen si-

Figure 9. Hardware used for the experiment.

Figure 10. The calibration process.

lently. When finished, they had to go back to monitoring the PFD until the end of the video. A demonstration was also performed to make it clear, especially for the participants with commercial aircraft experience (real-life and simulator) who might be used to a normal EICAS design with a separated checklist.

The fourth part was the adjusting of the eye-tracker device to fit the participant's eye and the calibration of the cameras. The participant should sit at a comfortable distance to both read the instructions on the screen and have a sufficient field of view to capture all the screen with the eye-tracker world camera (which was the reason why the participants could not be short-sighted). The calibration was done manually by placing a paper target in several strategic locations of the monitor such as the corners, and the participant was asked to look at it (fig. 10). The calibration was considered acceptable if the fixations on the main components of the displays were in the right AOI. These components were the airspeed, the altitude and the FMA for the PFD, the aircraft symbol and the two top corners for the ND, the top message and the top left engine indicator for the EICAS. After checking, if it was not acceptable, the calibration was performed again until satisfaction.

After that, the 4 videos were played successively in an order pre-defined by a randomised table and the participant would do what he was instructed to during the briefing. The capture was monitored in case where errors of offsets appeared. In these situations, the video would not be replayed, but the video would have to be analysed later with a special care. After each video, there was a 10 second-pause to ensure that the participant noticed the video had ended and that another one was about to start. Sometimes, if the participant had moved too much, another calibration was needed to minimize the error.

Once all the videos had been played, a post-experiment interview would take place where the participant would be asked two questions. They had to choose the display from which they perceived the needed information the most easily, so they had to write down its name (PFD, ND or EICAS). Then they had to rate the mental workload they perceived during the experiment when they were monitoring the PFD, then when the emergent scenario occurred on a scale from 1 to 10, 1 being the minimum workload and 10 the maximum. After these questions, the participant was asked to fill a demographic form to gather information about their gender, age, occupation, nationality and flight experience. Finally, they were free to ask questions or for some information about the study before leaving.

### 3.5 *Data analysis*

During the data extraction with Pupil Player, three AOIs were defined, corresponding to the three displays considered: the PFD, the ND and the EICAS display (fig. 11).

The program was able to identify fixations as long as the criteria were correctly defined. Thus, the duration threshold of 100 ms and the dispersion threshold of 1° were specified. Combined with the AOIs, it was possible to extract all the fixations on each AOI with the following characteristics: the ID indicating the ordinal position of the fixation compared to all the fixation overall, the timestamp which is a time reference associated with the start of the fixation, the duration of the fixation in seconds, the start and end frame numbers of the fixation, the normalised vertical and horizontal positions of the fixation, the average pupil size during the fixation.

Once extracted, these data were processed with Microsoft Excel 2013 so the relevant metrics could be applied. Hence, the final data included the average fixation duration in each AOI, the fixation number in each AOI, the total fixation time in each AOI and the average pupil size in each AOI. In each video, two extractions were made: one containing the 15 seconds preceding the emergence of the event, and one containing the 15 seconds following it. For each video, the process was repeated, so that means that there were 8 extractions for each participant.

Finally, for each participant, the set of eye-tracking data contained 4 eye-tracking parameters (average fixation time, number of fixations, fixation time, average pupil size) for each of the 3 AOIs (PFD, ND, EICAS), for each scenario (engine fire, cabin depressurisation, engine failure, electrical generator shutdown), before and after the emergent event. This set was completed with the data collected with the post-experiment interview data (easiest display to retrieve the information from, perceive mental workloads before and after the emergent event) and the demographical data (age, flight experience).

To test the hypotheses, the statistical analysis software IBM SPSS Statistics 23 was used. In order to test the relationship hypotheses, bivariate correlation tests were used. For the testing of the difference hypotheses, one-way ANOVA tests were used.

## 4 RESULTS

### 4.1 *Pupil size*

#### 4.1.1 *Relation with mental workload*

With data on the fixations on 3 AOIs, in 4 scenarios for each one of the 24 participants, each scenario separated in 2 (before and after the emergent scenario), there were 576 values of mean pupil size. However, some of these values were null because they had not fixated a certain AOI during one of the 8 semi-scenarios, so these values were removed, which led to 462 values. The pupil size was found to be correlated to the perceived mental workload ($r = 0.210$, $n = 462$, $p < 0.001$).

#### 4.1.2 *Comparison between the AOIs*

An analysis of variance test showed significant differences of pupil size between the different AOIs ($F(2, 459) = 3.5$, $p = 0.030$). However, a post hoc analysis using Tukey tests found that the only significant difference was between the PFD and the EICAS ($p = 0.030$). No other significant differences in pupil size could be found between the ND and the PFD ($p = 0.882$) or the ND and the EICAS ($p = 0.160$).

### 4.2 *Fixation duration*

The analysis of variance on the three levels PFD, ND and EICAS found significant differences in the mean fixation duration between the three AOIs ($F(2, 459) = 8.48$, $p < 0.001$). The post hoc analysis with Tukey tests showed that the significant difference was between the ND and the EICAS ($p < 0.001$). The differences between the PFD and the ND ($p = 0.073$) and between the EICAS and the PFD ($p = 0.091$) were not significant at the 0.05 level.


Figure 11. Definition of the three AOIs.

When separating the sequences before and after the emergent event, more interesting results could be found. Thus, each AOI (PFD, ND, EICAS) was separated in two sections (PFD before/PFDb, PFD after/PFDa, ND before/NDb, ND after/NDa, EICAS before/EICASb, EICAS after/EICASa), and two another analysis of variance was conducted, with the six levels PFDb, PFDa, NDb, NDa, EICASb and EICASa. Significant differences were found ($F(5, 456) = 10.3$, $p < 0.001$). The post hoc analysis did not found significant differences between the three AOIs before the emergent events ($p > 0.3$ for comparisons between PFDb, NDb, EICASb). After the emergent events, no differences were found between PFDa and NDa ($p = 0.321$). However significant differences between EICASa and PFDa ($p < 0.001$) and between EICASa and NDa ($p < 0.001$) were found.

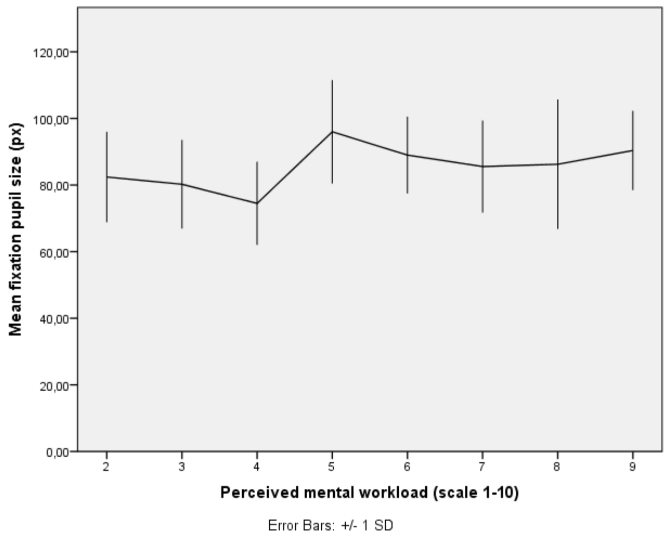


Figure 12. Variations of the mean pupil size for different perceived mental workload levels

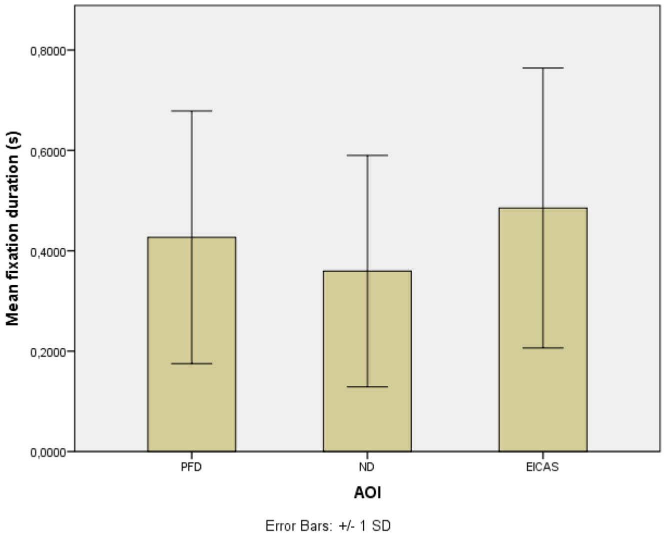


Figure 14. Mean fixation duration on each AOI

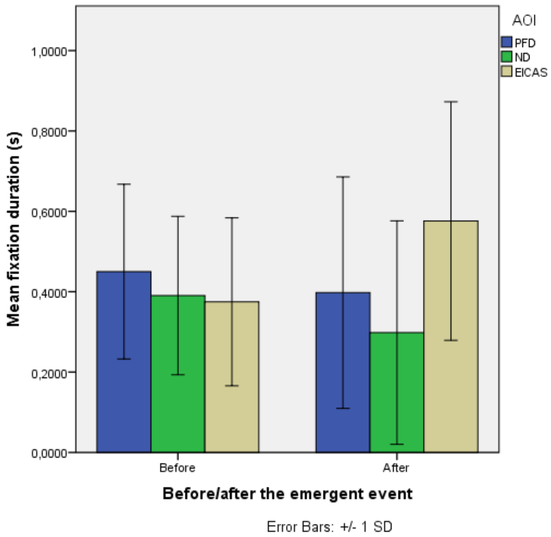


Figure 15. Mean fixation duration on each AOI before and after the event.

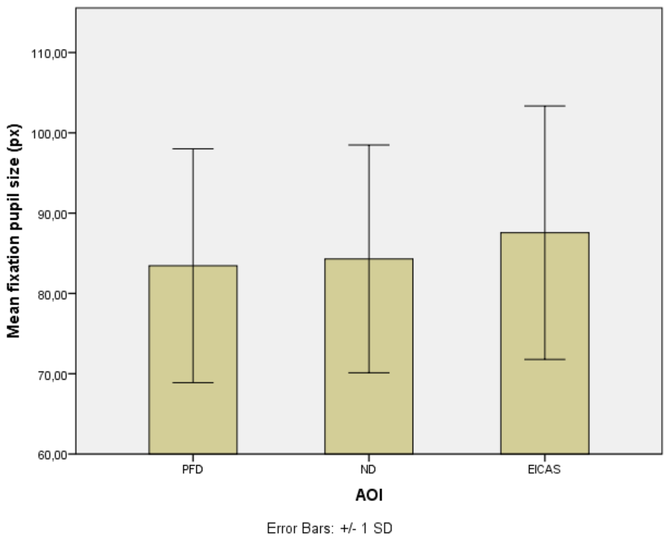


Figure 13. Mean pupil size on each AOI

### 4.3 *Number of fixation*

The analysis of variance on three levels PFD, ND and EICAS found significant differences in the number of fixation on each AOI ($F(2, 567) = 65.4$, $p < 0.001$). The post hoc analysis with Tukey tests suggested differences between the ND and the PFD ($p < 0.001$) and between the ND and the EICAS ($p < 0.001$). However, no significant differences were found between the PFD and the EICAS ($p = 0.864$) (fig. 16).

When distinguishing before and after the emergent event (fig. 17), the analysis of variance on the six levels PFDb, PFDa, NDb, NDa, EICASb and EICASa also found significant differences ($F(5, 564) = 151.3$, $p < 0.001$). Before the emergent event, the post hoc analysis revealed significant differences between the PFDb and the NDb ($p < 0.001$) and between the PFDb and the EICASb ($p < 0.001$), but not between the NDb and the EICASb ($p = 0.877$). After the emergent event, all three AOIs (PFDa, NDa, EICASa) were significantly different from the two others ($p < 0.001$ for all of the comparisons).

### 4.4 *Fixation time*

The analysis of variance found significant differences in the time of fixation on each AOI ($F(2, 569) = 72.1$, $p < 0.001$). The post hoc analysis with Tukey tests suggested differences between the ND and the PFD ($p < 0.001$) and between the ND and the EICAS ($p < 0.001$). No significant differences were found between the PFD and the EICAS ($p = 0.280$) (fig. 18).

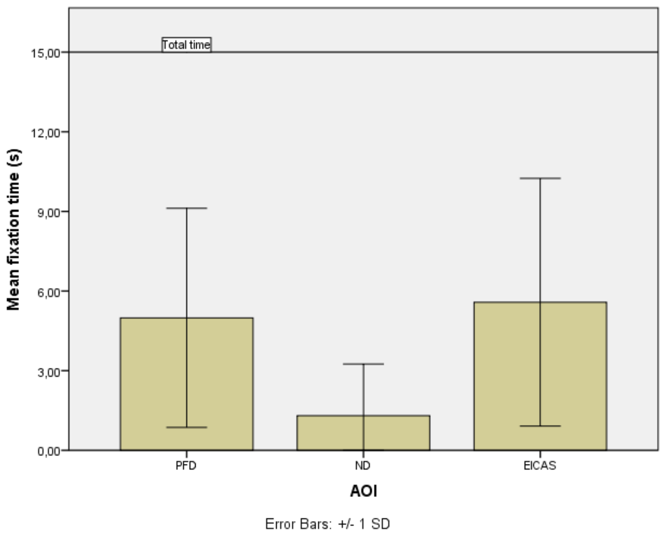


Figure 18. Mean fixation time on each AOI.

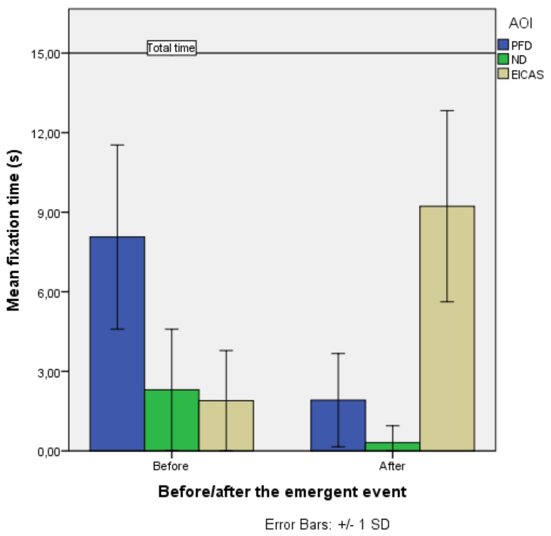


Figure 19. Mean fixation time on each AOI before and after the event.

When separating before and after the emergent event (fig. 19), the analysis of variance also found significant differences ($F(5, 566) = 212.2$, $p < 0.001$). Before the emergent event, the post hoc analysis revealed significant differences between the PFDb and the NDb ($p < 0.001$) and between the PFDb and the EICASb ($p < 0.001$), but not between the NDb and the EICASb ($p = 0.861$). After the emergent event, all three AOIs were significantly different from the two others ($p < 0.001$ for all of the comparisons).

## 5 DISCUSSION

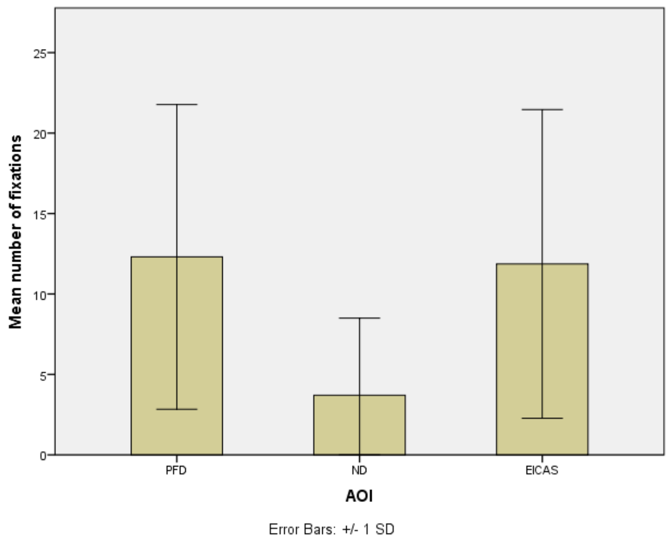


Figure 16. Mean number of fixations on each AOI.

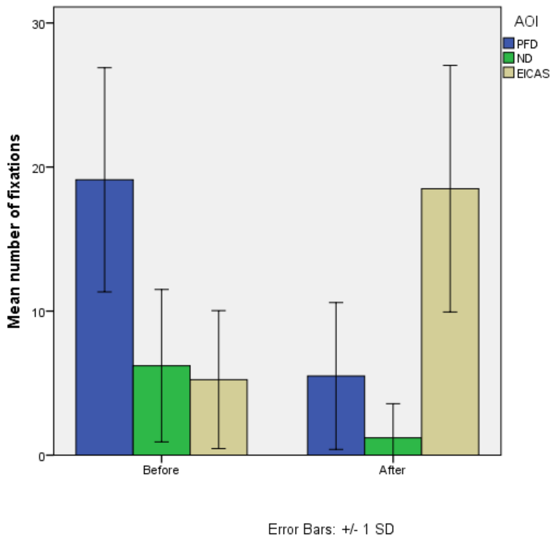


Figure 17. Mean number of fixations on each AOI before and after the emergent event.

### 5.1 *Validation of pupil size*

The significant correlation between the pupil size and the perceived mental workload opens up the possibility to interpret pupil size data in relation with mental workload. Indeed, the flaws of this metric such as the individual variations rule out the possibilities of between-participant analysis. However, within-task and within-individual analyses are reasonable. To make pupil size relevant for between-participant analyses, one wants to study the pupil dilation, that means the variations of pupil size due to mental activity rather than the pupil size itself. Also, as figure 12 suggests, the pupil size seems not enough to detect small differences in perceived mental workload. However it can make the difference between large variations. On this graph, two levels of mental workload are distinguishable. The pupil size is markedly higher for a perceived mental workload of 5 and more, in comparison with a perceived mental workload of 4 or less. Then, the pupil size discriminated the EICAS from the PFD and the ND, meaning that the mean pupil size of the fixations on the EICAS is significantly higher, also hinting that the perceived mental workload needed to read the text (EICAS) is higher than the numbers (PFD) and the symbols (ND).

### 5.2 *Comparison of AOIs*

Even without making the distinction between the situation before and after the emergent event, the four metrics (mean fixation duration, fixation number, total fixation time, mean pupil size) provided different results to discriminate the AOIs.

The mean fixation duration is the highest for the EICAS and the two other AOIs could not be significantly differentiated, implying the participants spent more time to interpret a symbol when they are looking at it. The number of fixations was significantly lower for the ND than for the PFD and the EICAS. These differences could indicate that the information was very clear and there was no need for many fixations to extract all the needed information, but that could also mean that the information displayed on the ND were less needed. The fixation time confirms these possibilities as it also pointed out the ND for being the less fixated in terms of time.

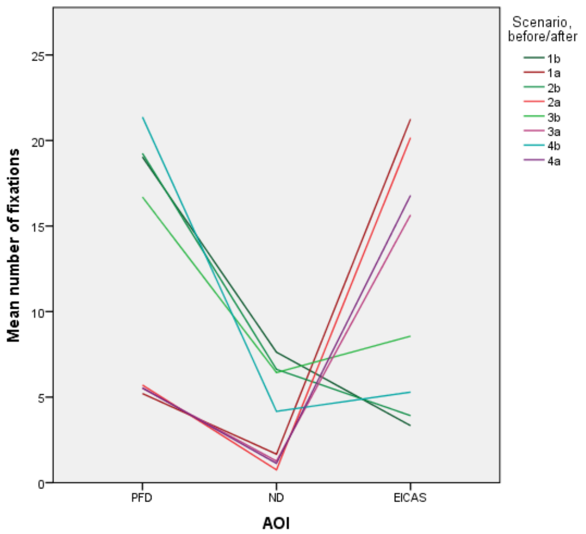


Figure 20. Mean number of fixations on each AOI during each scenario before (green/blue) and after (red/purple) the event.

In favour of the first hypothesis stating that the ND had less fixations because the information was easier to understand, the mean fixation duration time was the lowest for the ND, but this argument is ruled out because a vast majority of participants identified the PFD as the easiest display to get the information from over the ND (17 for PFD against 2 for ND). Also, as most of the participants have less than 10 hours of flight experience (15, against 9 with over than 10 hours), it is likely that they do not know all the symbols and indications displayed on it, while the PFD consists mainly in numbers. On the other hand, when monitoring in a normal flight, all the information needed is in the PFD (Duchowski, 2002), so the participants would only check the ND occasionally. To highlight this point, it is relevant to distinguish the situations before and after the emergent scenario because they consist in different tasks.

Before, the participant had to monitor the PFD. Most of the fixations are predicted to be directed at it. After, the participant had to read the instructions on the EICAS. The number of fixations on the EICAS is predicted to rise dramatically. Figure 18 illustrates this perfectly. The fixation number in the PFD before (PFDb) and in the EICAS after (EICASa) the event is significantly higher with all the other numbers. The fixation time analysis (fig. 19) fully support this. It has to be precised that scanning with fixations is not enough to monitor or read properly, but is necessary (Sarter et al., 2007). Thus, the main factor in the variations of fixation time and number of fixations is the type of task and not the cognitive processing. The graph (fig. 20) distinguishing all the scenarios before/after clearly shows patterns that are task-dependent, underlining the difference before/after. Although the scenarios are different in several information representations (different text colour and different symbols), the tasks demanded are the same through all of them and it is well reflected in this graph. That task-dependency explains all the significant differences found for the fixation number and the fixation time.

The before/after separation analysis of the fixation duration also led to interesting results. As there are no CAS messages on the EICAS in normal situations, the fixations on EICASb were directed to the engine indicators, which are mainly symbols (gauges). Therefore, the fixations on the EICAS were not only for text, but for symbols in EICASa and text in EICASb. To actually compare text with symbols or numbers, it is necessary to isolate EICASa. The mean fixation duration on the EICASa is significantly higher than the others (fig. 14). Although there seems to be a trend where the fixation duration on the PFD is higher than the ND, the lack of statistical significance does not allow the comparison between both of them. Hence, as the time spent extracting the information represented by words is the highest, it can be safely assumed that it is the most cognitively demanding form of information representation when processed. This confirms the conclusion of the pupil size analysis, where it is suggested that the elements on the EICAS required more workload to be processed.

The process of reading text seems to be more demanding than reading numbers or symbols. Symbols are used to encode one specific meaning. For instance, on the ND, there is one symbol for VOR and one symbol for airports. When the pilots look at one of these symbols, they know directly if it is an airport or a VOR, the relative position to the aircraft and the flight path and if it is tuned on the NAV radio (with the colour of the symbol). This is a large amount of information all contained in one symbol. Numbers consist in several figures which are related to the concept of quantity (Calude, 2011) so the relation is more direct. Although numbers are an encoded form of text, Brysbaert (1995) showed that number are processed differently. However, when reading text, the user has to process the letters to group them as a word, then retrieve its meaning (lexical access, discussed by Onifer and Swinney, 1981), then read the other words, then retrieve the sense of the sentence in relation of the context, while resolving ambiguities (Miyake et al., 1994). More steps are required to interpret text than numbers or symbols. One may argue that the ND symbols representing airports or VORs are also associated with letters, but there is actually no need to interpret these letters because they are just the name of the station.

In a text, different letters will refer to different words and different words refer to different concepts, but in the ND letters, different letters represent different names, but always refer to the concept of a name, just like different figures represent different numbers, but within the only concept of quantity. Although all this text processing is done within milliseconds, this is sufficient to make the difference in fixation duration and mental workload compared with numbers and symbols.

## 6 CONCLUSIONS

This paper reviewed the main eye-tracking methods of analysing cognitive processing. The most critical points in an eye-tracking study are the choice of the metrics and the experimental design. Both must be defined with care. The pupil size as a metric was introduced, while the classic approach is to use the pupil dilation. Although its analysis is more limited, as between-participants comparisons are not possible, this study showed that it is still a reasonable mean to identify large workload variations. To summarise, it is not mandatory to normalize the pupil size, as long as there is no attempt to compare between individuals.

The fixation number on each AOI gave interesting results. There are several ways to interpret it in relation with the mental workload. It increases if the user need several fixations to understand or when they are searching for the relevant information. However, this metric is very task dependant as the user fixates logically the areas that are more important for the specific task he is asked to complete. To correctly use it, this study recommends to define a specific task for all the participants to do, and then to compare them for this task. One example of application would be a comparison between novice and expert pilots doing a same task. As the pupil size is lacking between-participants possibilities, the fixation number on each AOI can be a great metric to complete the analysis.

The fixation duration was the most promising metric as it is directly linked to the psychological process of information processing. But this was only true under a strong hypothesis, according to which the fixations are the reflections of the attention direction. As showed in the literature review, a lot of studies support this hypothesis. However, in some cases such as extreme workload, stress or fatigue, there are some cases where the operator fixates something without extracting the information from it. Phenomena such as these are similar to the inattentional deafness (applied to sight instead of hearing) described by Dehais et al. (2014) and have led to a poor SA and sometimes to accidents as Kilingaru et al. (2013) rewiewed some of them. Further researches should pay attention to very high levels of workload and their influence on pupil duration.

The total fixation time in each AOI is logically very correlated and related to the fixation duration and the number of fixations in each AOI, because it is confounding the both of them. Although it made sense physically as the total amount of time spent to process the information in one AOI during a chosen period, it seems to have inherited of the drawbacks of its super metrics: the task-dependency of the fixation number and the possible errors for very high workload of the fixation duration. A good approach would be to always use it in complement to the number of fixations and the fixation duration and never alone, because a variation of fixation time could infer a change in either the fixation duration, the number of fixations or both, as Jacob and Karn (2003) advised.

Dismissing the total fixation time and the fixation number because the tasks asked before and after the event were not the same, the two relevant eye movement parameters the pupil size and the mean fixation duration both pointed at the text as the most difficult to comprehend, confirming that text processing is more demanding than number and symbol processing in flight deck operations. However, this study was not able to make a significant difference between number and symbol processing but one must take into account that its participants are not commercial pilots, therefore they do not know the perfect meaning of all the symbols despite the trial session.

As the results were unequivocal, they were not surprising. Flight deck designers seems to already think that the amount of text should be minimal on the displays (Newman and Greeley, 2001). This is why the PFD and the ND are essentially consisting in symbols and numbers. However, all the information cannot be converted into numbers or symbols. The numbers and symbols can only carry a limited amount of information, so elements like CAS messages and checklist instructions cannot be reduced to symbols. Also, the limitation with symbols is that they have to be balanced between being simple and specific. While a complex symbol can carry more information, it is more difficult to recognise and to process. On the other hand, if the symbol is too simple, it is less indicative and it is even possible that the user give another meaning to it. Hence, symbology design is a relevant field of research that can also benefit a large scope of other industrial fields, notably the ones where the safety and the performance of HMIs are related and both critical such as nuclear power plants, automobile or aerospace systems.

## 7 ACKNOWLEDGEMENTS

First and foremost, I have to thank my research supervisor, Dr. Wen-Chin Li, for his expert advice and priceless assistance through all the project despite his terribly busy timetable. He has been an inexhaustible source of motivation and enthusiasm. I do not think I could have a better supervisor for this research project.

Also, I would like to give special thanks to my experiment partner, Jiaqi Cao, for his incredible patience and his immeasurable mastery of the eye-tracker. His determination after all these hours of data collection was truly inspiring. In this respect, I am really grateful to all the participants, who accepted to give a few moments of their precious time to make this study possible.

But there were more than academic support. Many thanks to my housemates Alex, Summer and Zach for their inestimable support and, of course, the amazing burgers. Finally, I will never be able to thank my family enough, as their support has been unconditional during all these years.

## 8 REFERENCES

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91(2), 276-292.

Brysbaert, M. (1995). Arabic number reading: On the nature of the numerical scale and the origin of phonological reading. Journal of experimental psychology, 124(4), 434-452.

Byrne, M., Anderson, J., Douglas, S. and Matessa, M. (1999). Eye tracking the visual search of click-down menus. In: *Human factors in computing systems: CHI '99 conference proceedings*, 402-409. New York: ACM Press.

Callan, D. (1998). Eye movement relationships to excessive performance error in aviation. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 42(15), 1132-1136. SAGE Publications.

Calude, A. (2011). The science of numbers: does language help or hinder, *Language Sciences*, 33(4), 562-568.

Chapman, P. and Underwood, G. (1998). Visual search of dynamic scenes: Event types and the role of experience in viewing driving situations. In: Underwood, G. (Ed.), *Eye guidance in reading and scene perception*, 369-394. Amsterdam: Elsevier.

Collewijn, H. (1999). Eye movement recording. In: Carpenter, R. and Robson, J. (Eds.), *Vision Research: A Practical Guide to Laboratory Methods*, 245–285. Oxford: Oxford University Press.

Dehais, F., Causse, M., Vachon, F., Régis, N., Menant, E. and Tremblay, S. (2014). Failure to detect critical auditory alerts in the cockpit evidence for inattentional deafness. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 56(4), 631-644.

Di Stasi, L., Antolí, A. and Cañas, J. (2011). Main sequence: An index for detecting mental workload variation in complex tasks. *Applied Ergonomics*, 42(6), 807-813.

Di Stasi, L., Renner, R., Catena, A., Cañas, J., Velichkovsky, B. and Pannasch, S. (2011). Towards a driver fatigue test based on the saccadic main sequence: A partial validation by subjective report data, *Transportation Research Part C*, 21, 122-133.

Dishart, D. and Land, M. (1998). The developpement of the eye movement strategies of learner drivers. In: Underwood, G. (Ed.), *Eye guidance in reading and scene perception*, 419-430. Amsterdam: Elsevier.

Dornheim, M. (1995). Modern cockpit complexity challenges pilot interfaces. *Aviation Week and Space Technology (New York)*, 142(5), 60-63.

Duchowski, A. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers*, 34(4), 455-470.

Endsley, M. (1995). Toward a theory of situation awareness. *Human Factors*, 37(2), 32-64.

Goldberg, J. and Kotval, X. (1999). Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics*, 24(6), 631-645.

Goldberg, J. and Wichansky, A. (2002). Eye tracking in usability evaluation: A practitioner's guide. In: Hyönä, J., Radach, R. and Deubel, H. (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research*, 493-516.

Harris, R., Glover, P. and Spady, A. (1986). Analytical techniques of pilot scanning behavior and their application. *NASA Technical Report 2525*.

Ho, G., Scialfa, C., Caird, J. and Graw, T. (2001). Visual search for traffic signs: The effect of clutter, luminance and aging. *Human Factors*, 43, 194-207.

Hoffman, J. and Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Attention, Perception, & Psychophysics*, 57 (6), 787–795.

Jacob, R. and Karn, K. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind*, 2(3), 573-605.

Just, M. and Carpenter, P. (1976). Eye fixations and cognitive processes. *Cognitive psychology*, 8(4), 441-480.

Just, M. and Carpenter, P. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review*, 87(4), 329.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kasarskis, P., Stehwien, J., Hickox, J., Aretz, A. and Wickens, C. (2001). Comparison of expert and novice scan behaviors during VFR flight. In: *Proceedings of the 11th International Symposium on Aviation Psychology*, 1-6.

Kilingaru, K., Tweedale, J., Thatcher, S. and Jain, L. (2013). Monitoring pilot 'Situation Awareness'. *Journal of Intelligent & Fuzzy Systems*, 24(3), 457-466.

Liang, Y., Reyes, M. and Lee, J. (2007). « Real-Time Detection of Driver Cognitive Distraction Using Support Vector Machines ». *IEEE Transactions on Intelligent Transportation Systems*, 8, 340–350

Lin, L. and Lu, M. (2016). Empirical research on the relationship between helicopter pilots' mental workloads and situation awareness levels. *Journal of the American Helicopter Society*, 61(3), 1-8.

Lindo, R., Deaton, J., Cain, J. and Lang, C. (2012). Methods of instrument training and effects on pilots' performance with different types of flight instrument displays. *Aviation Psychology and Applied Human Factors*, 2(2), 62-71.

Lohse, G. (1997). Consumer eye movement patterns on Yellow Pages advertising. *Journal of Advertising*, 26, 61-73.

Marshall, S. (2000). *Method and apparatus for eye tracking and monitoring pupil dilation to evaluate cognitive activity*. Google Patents.

McConkie, G. and Rayner, K. (1998). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, 17, 578-586.

Miyake, A., Just, M. and Carpenter, P. (1994). Working memory constraints on the resolution of lexical ambiguity: Maintaining multiple interpretations in neutral contexts. *Journal of memory and language*, 33(2), 175-202.

Newman, R. and Greeley, K. (2001). *Cockpit Displays*. Aldershot, Hants, England: Ashgate.

Nyström, M. and Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior research methods*, 42(1), 188-204.

Onifer, W. and Swinney, D. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition*, 9(3), 225-236.

Poole, A. and Ball, L. (2006). Eye tracking in HCI and usability research. *Encyclopedia of human computer interaction*, 1, 211-219.

Posner, M. (1980). Orienting of attention. *Quarterly journal of experimental psychology*, 32(1), 3-25.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3), 372.

Rayner, K., Rotello, C., Stewart, A., Keir, J. and Duffy, S. (2001). Integrating text and pictoral information: Eye movements when looking at print advertisements. *Journal of Experimental Psychology: Applied*, 4, 219-226.

Reichle, E., Pollatsek, A., Fisher, D. and Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological review*, 105, 125-157.

Salvucci, D. and Goldberg, J. (2000). Identifying fixations and saccades in eye-tracking protocols. In: *ETRA 2000 - Proceedings of the Eye Tracking Research and Application Symposium*. New York: ACM, 71-78.

Sarter, N., Mumaw, R. and Wickens, C. (2007). Pilots' monitoring strategies and performance on automated flight decks: An empirical study combining behavioral and eye-tracking data. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *49*(3), 347-357.

Thomas, L. and Wickens, C. (2004). Eye-tracking and individual differences in off-normal event detection when flying with a synthetic vision system display. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 48(1), 223-227. SAGE Publications.

Wang, L., Wen, F., Ma, C., Zhao, S. and Liu, X. (2014). Evaluation of helmet-mounted display targeting symbology based on eye tracking technology. In: *SPIE Defense+ Security (pp. 90860P-90860P)*. International Society for Optics and Photonics.

Wedel, M. and Pieters, R. (2000). Eye fixations on advertisements and memory for brands: A model and findings. *Marketing Science*, 19, 297-312.

Whitfield, T. and Whiltshire, T. (1990). Color psychology: a critical review. *Genetic, social, and general psychology monographs*, 116(4), 385-411.

Wilder, J., Kowler, E., Schnitzer, B., Gersch, T. and Dosher, B. (2009). Attention during active visual tasks: Counting, pointing, or simply looking. *Vision research*, 49(9), 1017-1031.