THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

Présentée et soutenue le *28 octobre 2016* **par :**

Vsevolod PEYSAKHOVICH

Study of pupil diameter and eye movements to enhance flight safety

Étude de diamètre pupillaire et de mouvements oculaires pour la sécurité aérienne

---

JURY

Thierry BACCINO	Rapporteur
Frédéric DEHAIS	Directeur
José J. CAÑAS DELGADO	Rapporteur
Andrew DUCHOWSKI	Rapporteur
Jean-Pierre JESSEL	Examinateur
François VACHON	Co-Directeur

École doctorale et spécialité :
EDSYS : Informatique 4200018
Unité de Recherche :
Équipe d'accueil doctoral CSDV - Commande des systèmes et dynamique du vol
Directeur(s) de Thèse :
Frédéric DEHAIS et *François VACHON*
Rapporteurs :
Thierry BACCINO, *José J. CAÑAS DELGADO* et *Andrew DUCHOWSKI*

STUDY OF PUPIL DIAMETER AND EYE MOVEMENTS TO ENHANCE FLIGHT SAFETY
Most aviation accidents include failures in monitoring or decision-making which are hampered by arousal, stress or high workload. One promising avenue to further enhance the flight safety is looking into the pilots' eyes. The pupil is a good indicator of cognitive/attentional states while eye movements reveal monitoring strategies. This thesis reflected upon the application of eye tracking in aviation with following contributions: 1-2) The two pupil experiments revealed that the luminance impacts the cognitive pupil reaction. Depending on the nature of the cognitive load – sustained or transient – the corresponding pupillary component would be impacted. The same amount of cognitive load under dimmer luminance condition would elicit larger tonic pupil diameter in a sustained load paradigm and larger phasic pupil response in a transient load paradigm. 3) We designed a novel mathematical framework and method that provide comprehensive illustrations of scanpaths for qualitative analysis. This framework also makes a lane for new methods of scanpaths comparison. 4) The developed technique of analysis of fixations and construction of "explore-exploit" ratio is presented and verified on the data from two experiments in flight simulators. 5) Eventually, we proposed a framework of eye tracking integration into the cockpits. It contains four stages presented in both chronological order of its integration and technical complexity.

ÉTUDE DE DIAMÈTRE PUPILLAIRE ET DE MOUVEMENTS DES YEUX
POUR AMÉLIORER LA SÉCURITÉ AÉRIENNE
L'analyse d'événements aériens révèle que la plupart des accidents aéronautiques ont pour origine une surveillance inadaptée de paramètres de vol induite par une vigilance réduite, le stress ou une charge de travail importante. Une solution prometteuse pour améliorer la sécurité aérienne est d'étudier le regard des pilotes. La pupille est un bon indicateur de l'état attentionnel/cognitif tandis que les mouvements oculaires révèlent des stratégies de prises d'information. La question posée dans ce manuscrit est d'évaluer l'apport de l'oculométrie pour la sécurité aérienne par les contributions suivantes : 1-2) Les deux premières études de ce doctorat ont démontré que les effets d'interaction entre la luminance et la charge cognitive sur la réaction pupillaire. La composante pupillaire impactée dépend de la nature de la charge – soutenue ou transitoire. 3) Un cadre mathématique développé fournit un moyen d'illustration de schémas visuels pour l'analyse qualitative. Ce cadre ouvre également la voie à de nouvelles méthodes pour comparer quantitativement ces schémas visuels. 4) Une technique originale d'analyse de fixations et de construction d'un ratio "exploration-exploitation" est proposée et est appliquée dans deux cas d'études en simulateur de vol. 5) Enfin, on propose un cadre théorique d'intégration de l'oculométrie dans les cockpits. Ce cadre comporte quatre étapes présentées dans, à la fois, l'ordre chronologique de l'intégration et la complexité technique de réalisation.

ACKNOWLEDGEMENTS

I would like to thank all the people without whom this project would not come true, all those people who gave me scientific and personal support. First of all, I am grateful to my adviser Dr. Frédéric Dehais for introducing me to the domain almost four years ago. I remember how I was impressed with the works of the Neuroergonomics and Human Factors team he presented to me and how I dreamed of understanding all these complicated things from computer science to neuroscience through human factors. Frédéric gave me the opportunity to join the team, an interesting subject to work on and followed my project for the last three years. I gratefully thank him for the endless support he provided me. I am also grateful to my co-adviser Dr. François Vachon for supervising my work from across the ocean and giving helpful advice during the project.

I am grateful to all members of the Human Factors and Neuroergonomics research group at ISAE-SUPAERO and the rest of the department staff who participated in the project, it was (and it is) a real pleasure to work with you. I want particularly thank Dr. Mickaël Causse for his friendship and precious scientific advice. I am also grateful to Dr. Sébastien Scannella for his time and help. I also thank my colleagues from Laval University, and individually Dr. Sébastien Tremblay, and from ENAC, in particular, Dr. Christophe Hurter and Dr. Jean-Paul Imbert. The list of wonderful persons with whom I have a chance to work is long, and you know that I am grateful to you all. I also want to thank reviewers and committee members for their time and fruitful discussions around the dissertation.

The last and very special thanks go to my family. Спасибо большое! ¡Y muchísimas gracias a ti, mi amor!

To my family

PREFACE

The commercial aviation is one of the safest modes of transportation nowadays and it's getting safer every day. The extremely low rates of accidents, that were 20–30 times higher in the 60s, are mostly due to the technological progress of aeronautical systems, but also to improvements in pilot training, flight crew, and air traffic control procedures. Technical failure today is the cause of about only 10% of accidents, leaving a big percentage to human implication. The exact values vary over years and sources, but approximately from 60 to 80 percent of aviation accidents involve human error.

About half of these accidents include failures in monitoring or faulty decision-making. Such factors as arousal, workload and fatigue can hamper an adequate monitoring of flight instruments and good decision-making. Therefore, one promising avenue to further enhance the flight safety is looking into the pilots' eyes. At one hand, pupil diameter brings a light on our state of arousal, working memory load, emotional state etc. All these factors jeopardize decision-making and monitoring and in order to mitigate their adverse effects there is a need for a robust measure of human attentional and cognitive behavior and the pupil diameter can be the one. At another, eye movements tell us about visual scan patterns and monitoring strategies, as well as give us a hint about flight crew's internal state.

This Ph.D. work focuses on eye tracking around two axes to enhance the flight safety — pupil diameter and eye movements — and is organized in the following way. Chapter 1 opens the thesis by presenting the human error in aviation. It explains why it is crucial to study it and why eye tracking is a possible solution. Chapters 2 and 3 provide readers with necessary notions on pupil diameter and eye movements. In particular, the chapters explain how these things relate to internal states and, therefore, can be a useful tool to estimate it. These three introductory chapters are followed by five contributions. Each contributory chapter, fully or partly, has either been published or submitted to an international peer-reviewed journal or a conference. Chapters 4 and 5 present the two pupillometry experiments revealing that the luminance impacts the cognitive pupil reaction. The first study showed that during transient load task the phasic pupil response varied according to luminance, the response being more pronounced under dimmer condition. In the second experiment, we studied the impact of luminance on tonic and phasic components of pupillary response during sustained processing. We found an impact of luminance on tonic pupil diameter, reaction also being larger under dimmer light. Taken into account the findings from both experiments, we concluded that depending on the nature of the cognitive load – sustained or transient – the corresponding component of the pupillary response would be impacted. We postulated that the same

amount of cognitive load under dimmer luminance condition would elicit larger tonic pupil diameter in a sustained load paradigm and larger phasic pupil response in a transient load paradigm. Chapter 6 presents a novel mathematical framework and method of Attribute-Driven Edge Bundling that provide comprehensive illustrations of scanpaths for qualitative analysis of eye data. It also makes a lane for new quantitative methods of scanpaths comparison by providing 2D maps that can be compared across conditions and participants. Chapter 7 presents the developed technique of analysis of fixational eye movements and the construction of “explore-exploit" or “search-processing" ratio that was verified on the data from two experiments in full flight simulators. The presented method also highlights the importance of separating shorter fixations from longer and studying the full distribution of fixation durations. Eventually, in Chapter 8, we proposed a framework of eye-tracking integration into the cockpits. The framework contains four stages presented in both chronological order of its integration and technical complexity. These four stages are Pilot Training, On-board Gaze Recording, Gaze-Based Flight Deck and Aircraft Adaptations. To summarize, this Ph.D. thesis explains how pupil diameter and eye movements can enhance flight safety. Different contributions promote eye tracking application in aviation and ease the process of integration into the modern cockpits.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

LIST OF TABLES

Part I

EYES AT THE ORIGIN OF AIRCRAFT ACCIDENTS & INCIDENTS: PUPIL DIAMETER AND EYE MOVEMENTS OF THE PILOTS

1

HUMAN ERRORS IN AVIATION

> Je réféchissais, je me disais que ce qui manquait surtout pour le bon fonctionnement du système, c'était l'erreur humaine, et que celle-ci devait intervenir d'urgence.
>
> *Gros-Câlin* by ROMAIN GARY

> I was thinking. I was telling myself that what was particularly missing for the good system performance, it was the human error, and that it has to urgently get involved.
>
> my loose translation of *Gros-Câlin* by ROMAIN GARY

1.1 HUMANS ERR

1.1.1 *Anything that can go wrong will go wrong*

Everyone heard of the Murphy's Law. It states that *"Anything that can go wrong will go wrong"*. What you might *not* know about the Murphy's law is that its namesake, Mr. Edward Murphy, was an aerospace safety engineer. And the aerospace domain is a perfect example of the practical faithfulness of the famous law. For instance, can a severe turbulence hurt flight attendants? Yes, it can. Well, then it will (Air Accidents Investigation Unit (AAIU), 2014). Can an aircraft's cabin lose its pressure in-flight? Yes, it can. Well, then it will (Transport Accident and Incident Investigation Bureau (TAIIB), 2012). If the captain has, let say, a prosthetic arm, can it be accidently detached during the landing? Yes, it can. Well, then it surely will one day (Air Accidents Investigation Branch (AAIB), 2014). The latter example being quite anecdotical, it shows that everything can happen in the sky and on the ground that might reduce safety margins of a successful flight. And if there are meteorological forecasts over the regions the aircraft flyover, no such thing exist for a possible system error. And, according to the

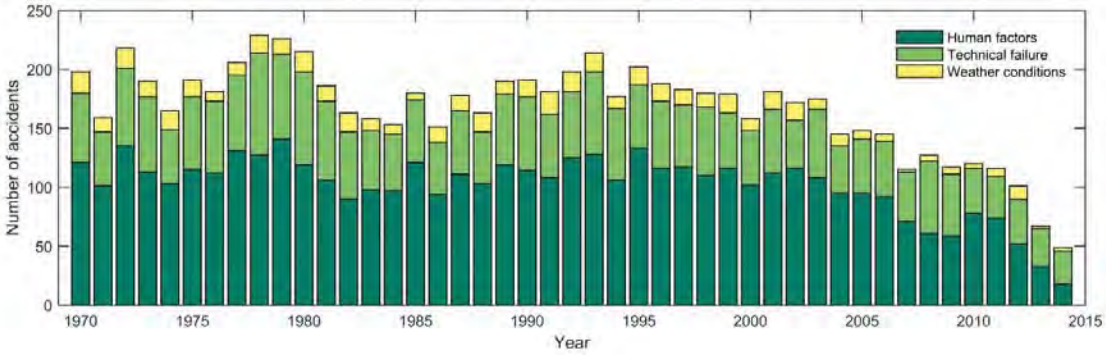


Figure 1.: Aeronautical accidents from 1970 to 2014 with human factors, technical failure and weather conditions as contributory factors. Data retrieved from Bureau of Aircraft Accidents Archives (www.baaa-acro.com)

Murphy's Law, if an aeronautical system allows for a malfunctioning, then one day or another it will rise to the surface. And at that moment of time, only the flight crew will be able to cope with it. But, first, they will need to perceive it, understand it, and then, if they know how, fix it. However, the things are not that simple and under the effect of many factors sometimes incidents occur. *Errare est humanum*, flight crew members are humans, therefore, sometimes they commit errors too. But not that often.

1.1.2 *Safest transport with murderous annals*

The commercial aviation is one of the safest modes of transportation nowadays and it's getting safer every day. Between 2005 and 2014 the accident rates were as low as 0.58 for hull loss[1] accidents and 0.28 for fatal accidents[2] per million departures[3] for scheduled commercial passenger operations (Boeing, 2015). Although it's safe nowadays, needless to say, the aviation history is unfortunately littered with all kinds of accidents. The drama is often redoubled by the media who like to savor the simultaneous death of many persons at the same time. One crashed aircraft carrying 230 passengers is perceived as something much more dramatic than 230 car crashes that happened all over the world at the same day. Nevertheless, since the late 50s, the aeronautical domain attained the unprecedented levels of safety so that it is now statistically safer to fly in an airliner than to walk across a busy city street. The extremely low rates of accidents, that were 20–30 times higher in the 60s, are mostly due to the technological progress of aero-

1 A hull loss accident is an accident after which the airplane is totally destroyed or damaged and not repaired. It also includes events in which the aircraft is missing.

2 A fatal accident is an accident that results in fatal injury, an injury being fatal when it results in deaths within 30 days of the accident.

3 With the total of 192 million departures.

nautical systems, but also to improvements in pilot training, flight crew and air traffic control procedures. The progressive introduction of modern “computerized" cockpits and retrofitting of conventional cockpits have considerably improved the flight safety compared to the previous generation of aircraft. Technical failure today is the cause of about only 10% of accidents, leaving a big percentage to human implication. The exact values vary over years and sources, but approximately from 60 to 80 percent of aviation accidents involve human error (Shappell et al., 2006). Figure 1 traces the number of accidents for civil and military aircraft from 1970 to 2014 according to one of three major contributing factors: human factors, technical failure, and bad weather conditions. Most notably, the number of accidents diminishes over years, but the percentage of accidents implying a human error still often dominates purely technical failures. But what is a human error in aviation?

1.1.3 *Types of human errors in aviation*

With some tragic exceptions such as the Germanwings’ flight where the ground collision was fully due to the deliberate and planned action of the co-pilot (Bureau Enquêtes-Accidents (BEA), 2016), the aviation accidents generally are the result of a chain of unfortunate events. It can be thought as a series of error dominos, each domino effecting the toppling of the next, like in Bird’s Domino Theory (Bird, 1984); or as a more popular “Swiss cheese" model of James Reason (1990) where the “cheese holes" need to be aligned so that a series of unsafe events or acts culminate in an accident. An important aspect of these models is that some failures are latent but do influence the following element of the events’ chain. And it is important not only to focus the accident investigations upon the very last level of human error but also consider the latent levels. It also worth noting that the reconstruction of the right sequence of events is an elaborate process and some precautions are to be taken to avoid the traps of hindsight (Dekker, 2002). The Reason’s work, being theoretical at first[4], eventually acquired an applied aeronautical body – HFACS (Shappell and Wiegmann, 2000, 2001; Wiegmann and Shappell, 2001)– a human error framework for aviation accident investigators.

The HFACS classifies human error in aviation at four different levels according to Reasons’ model: (a) organizational influences, (b) unsafe supervision, (c) preconditions for unsafe acts, and (d) the unsafe acts of operators. It is useful to note, removing a possible misunderstanding, that 1) the level (b) includes middle management supervision and not the supervision of flight instruments; 2) the operators mentioned in the levels (d) include not only aircrew but also maintainers and air traffic controllers[5] There is a total of 19 causal categories subdivided into these four levels. The mention of the HFACS

4 The original work of James Reason reflects upon the nuclear plants and not the aeronautical accidents.

5 Often in this manuscript, I will refer to the flight crew and piloting but all being said will be still applicable for the air traffic control and other professions that include monitoring such as security surveillance.

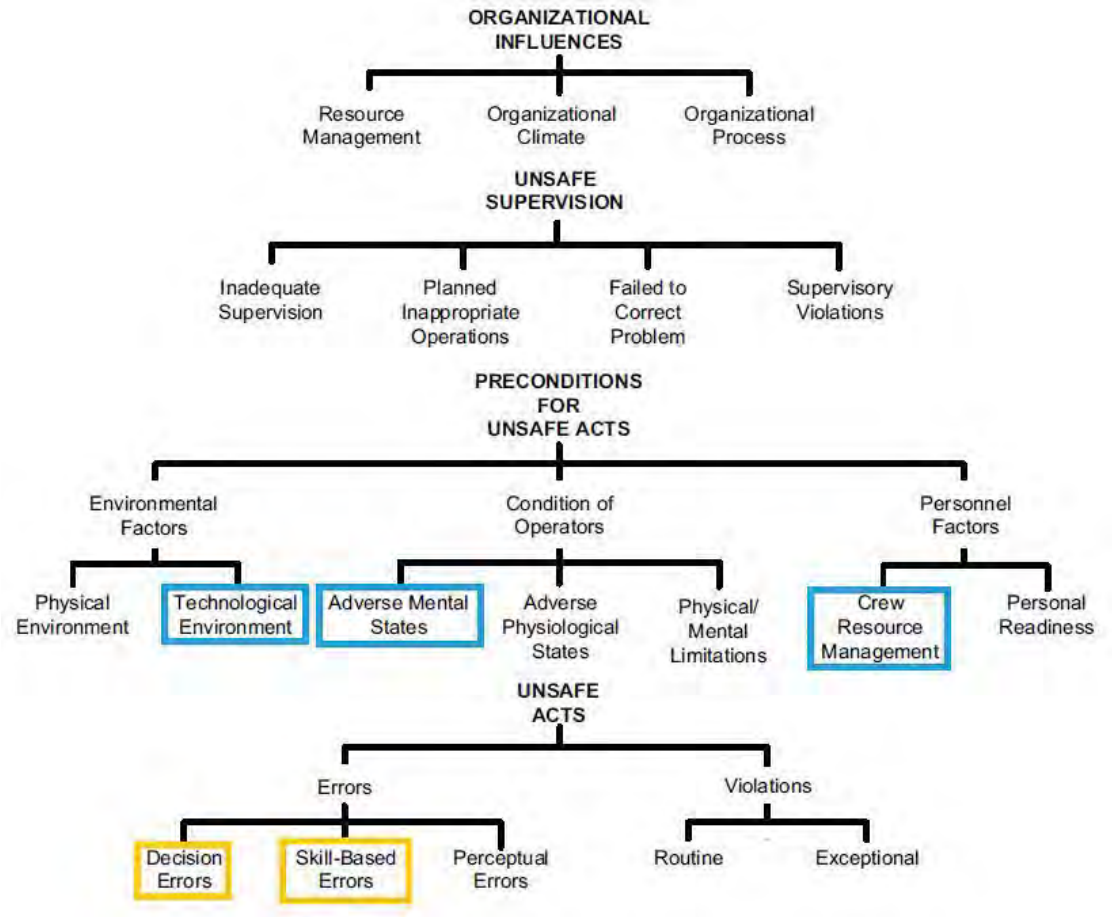


Figure 2.: The HFACS framework with highlighted categories relevant to the present thesis. Adapted from Shappell et al. (2007)

system is helpful in this chapter to understand that there are many different types of human errors that have contributed to aviation incidents and accidents. Therefore, talking about human error in aviation is in no case equivalent to talking about piloting errors.

The figure 2 represents the 19 causal categories of the HFACS. I highlighted five categories as being most relevant to the present thesis. Two of these highlighted categories (Decision Errors, Skill-Based Errors) belong to the last active level of *unsafe acts*, and three others (Technological environment, Adverse Mental States, Crew Resource Management) are situated at the level just before – the first latent level of *preconditions for unsafe acts*:

- Technological environment especially encompasses the design of the interface (displays, controls);

- Adverse mental states include psychological and mental factors such as fatigue, attentional lapses etc.;
- Crew resource management comprehends communication, coordination and teamwork issues;
- Decision errors represent poorly executed procedures and wrong choices; it also encompasses the misuse or misinterpretation of relevant information;
- Skill-based errors include, in particular, breakdowns in visual scan patterns.

All these five categories include the vision as an important ingredient. Displays design can enhance or negatively affect the visual attention. Adverse mental states affect the efficiency of the visual scan. Crew resource management includes the cross-monitoring between the pilots and efficient task sharing (such as not to observe both the same aspect of the situation). The processing of visual information contributes to efficient decision-making. Eventually, adequate visual scan patterns are an essential skill for the aircraft piloting and air traffic control.

Shappell et al. (2007) used the HFACS to classify 1020 commercial aviation accidents that occurred over a 13-year period. The authors found that technological environment was involved in about 1.5% of the accidents, adverse mental states — in about 6.5%, crew resource management – 10.7%, decision errors – 36.7% and skill-based errors in 56.5%. The two latter categories are the largest and, thus, deserve a particular attention. Also, combining these numbers with the percentage of the human error in aviation, we can conclude that the latter category contributed to almost a half of aeronautical accidents. In the next sections, after providing an illustrative case, I describe in more details some aspects of decision-making in aeronautics and why the monitoring skills are of such crucial importance.

1.2 CASE STUDY OF THE FLIGHT AF72

A concrete example of the accident is the Air France Boeing 747-428 B registered F-GITA on September 13, 1993, at Tahiti Faaa. We will first review the accident, then give some explanations about the contributive factors, and, eventually, reflect upon the usefulness of knowing the gaze position during such situations to at least facilitate the investigation or even prevent the accident. All the details of the accident are reported as described in the official BEA report (BEA, 1999).

1.2.1 *Summary of the approach ending in the lagoon*

A flight from Paris to Tahiti via Los Angeles was coming to end. It was night, about 8.40 p.m., and meteorological conditions were favorable. The early stages of the approach

did not bode anything unusual. The first officer was at the controls, being Pilot Flying (PF), while the captain took the role of Pilot Monitoring. At 10 Nm from the runway at about 3000 feet, the aircraft was configured for landing, the landing gear being extended, the flaps raised to 30°, and holding a stable approach speed maintained by the auto-throttle at 149 knots (Vref+5 kt)[6], the autopilot being disconnected. The Flight Mode Annunciator (FMA) displayed SPD VNAV PTH[7] at that moment.

Later, when the radio altimeter announced 500 feet, the captain read out, according to the recordings of the Cockpit Voice Recorder, the instructions of the FMA THR REF VNAV SPD[8]. At that moment, the throttle has increased, and then the speed increased. Six seconds later, the synthesized voice announced 400 feet and then the captain recognized that the airplane was above the descent path. Few seconds afterwards he also noted and reported that they were going too fast (189 kt that is Vref+35 kt[9]). In compliance with airline directives, low altitude combined with such excessive speed (30 kt above the normal) should have implied an immediate go-around. Later, the captain explained that he persisted in approach because of good meteorological conditions (good visibility and low wind), and dry and long runway. After recognizing the thrust increase, the first officer, was holding the throttles back at idle position (see ① in Figure 4). The crew had about sixteen seconds before the

Pilot Flying (PF)

PF is a pilot role during task sharing. He/she is responsible for controlling the vertical and horizontal flight path and for energy management. He fulfills these responsibilities by either supervising the auto-pilot and the auto-thrust when engaged, or hand-flying the aircraft. PF also can give orders to PM (see below).

Pilot Monitoring (PM)

PM is a pilot role during task sharing. PM is responsible for systems-related tasks and monitoring of flight panel. He/she also performs the actions requested by the PF such as radio communications, systems configurations, and automatic mode selections etc. Eventually, the PM is responsible for monitoring the PF to cross-check his/her actions and backup if necessary.

6 Vref is the landing reference speed that is not less than 1.3 times the stall speed in the normal landing configuration.

7 SPD is an auto-throttle mode that maintains speed selected in Mode Control Panel (MCP). VNAV PTH is a pitch mode that maintains altitude or vertical descent path commanded by Flight Management Computer (FMC) to follow route vertical profile.

8 THR REF is an auto-throttle mode that maintains reference thrust which is shown on Engine Indication and Crew Alerting System (EICAS) display. VNAV SPD is a pitch mode that maintains FMC commanded airspeed by pitching the aircraft up or down.

9 When the speed during the approach (with flaps set on 25° or 30°) is situated outside the [Vref-5 kt, Vref+30 kt] interval, the airline's flight analysis service starts a specific investigation procedure.

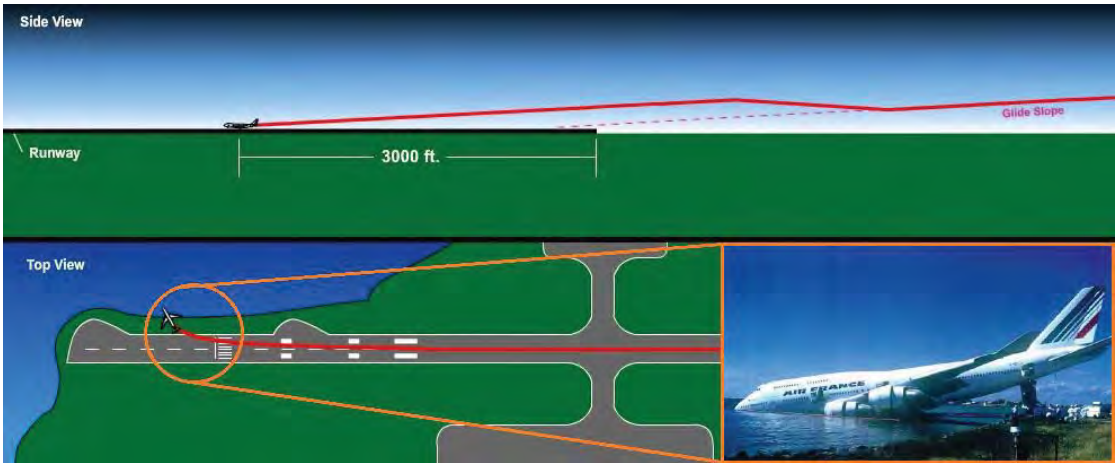


Figure 3.: Slide and top views of the flight trajectory above the descent slope ending in the water of the lagoon

touchdown. Maintaining the throttles at idle reduced the airspeed decreased to 168 knots but that was still almost 20 knots faster than the recommended approach speed. Two seconds after touchdown, the power to engine n°1 began to increase stabilizing in forward thrust that disabled the extension of the spoilers and disarmed the automatic braking system. During taxiing, the crew noted the dissymmetry of thrust and attempted to control the trajectory, but, eventually, the aircraft went off the runway laterally and ended in the water of the lagoon (Figure 3). There were no serious injuries to persons, and the aircraft retrieved the wind under the wings one year later. But how this aircraft, perfectly functional and metrological conditions being good, could end in the water of the lagoon?

1.2.2 *Contributive elements: investigation*

The accident investigation showed that the automatic flight system initiated a go-around maneuver at the missed approach point (at about 500 feet). The autopilot was disconnected, and the aircraft's pitch attitude did not change, but the auto-throttle control was active and commanded full-forward thrust. The philosophy of the selected approach mode supposes that at the End of Descent point at the decision height, either the crew has enough visual references and continues the approach or initiates a go-around maneuver. By design, the automatic flight system considered that the visual approach was not decided by the flight crew because the auto-throttle control was still active. Therefore, the system applied the missed approach procedure. Few seconds before the touchdown, probably at the moment when the pilot had moved his hand to grasp the reverser con-

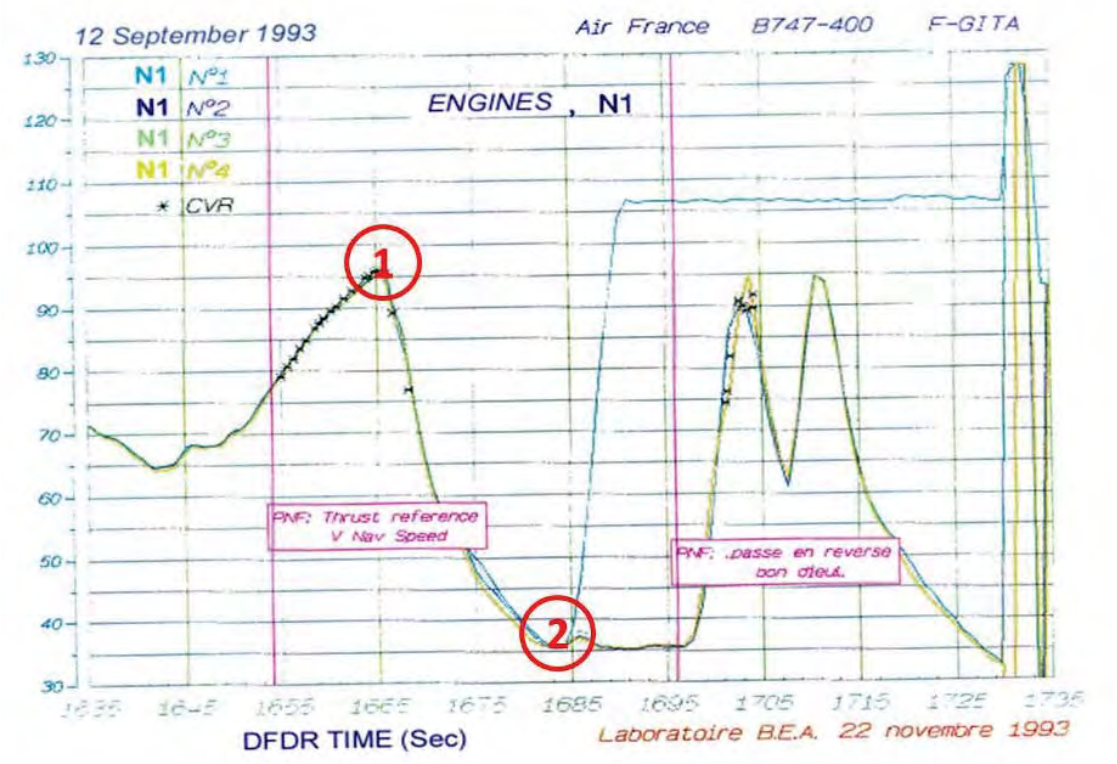


Figure 4.: Extract of recordings of engines' power from the Flight Data Recorder

trols, the throttle control n°1 slipped from the PF (see ② in Figure 4) and engine n°1 went into full forward thrust, thus destabilizing the aircraft.

The main reason of this accident is that for a long time neither of the two crew members was aware of the automatically initiated go-around. Yet, the captain announced the THR REF VNAV SPD, an FMA message that indicated the mode changes. Unfortunately, he did nothing but announce the message, without analyzing its meaning. Likewise, the first officer did not confirm the message. The report also states that "*the crew, having moved into visual flight sometime before and having the runway [...] in sight, were not looking at the flight director; their attention being focused on maintaining the visual descent path and not the monitoring of the instruments and the automatic flight controls*". Analyzing the task sharing between two pilots, the BEA also underlined that the flight crew had a significant difference in flight experience and that the captain got involved more and more in managing the descent path. Once the captain noticed that the aircraft was above the descent path, instead of paying attention to instruments monitoring, as his role of the Pilot Monitoring (PM) implied, he focused the attention outside, making sure the aircraft returned to the correct descent path. The consequence is that the crew was focused outside, on one aspect of the situation, and forgot to su-

pervise the instruments (engines parameters, thrust reversers, spoilers etc.), violating, therefore, their respective roles and responsibilities. In summary, after analyzing the important contributing elements of the accident, we note particularly:

- Mechanical reading of the FMA message by the pilot monitoring without analyzing its content and the absence of confirmation by the PF;
- Both crew members focused on one aspect of the situation;
  pilots' gazes were at the same place;
- The absence of instrumental check of engines, reversers, spoilers etc.

1.2.3 *Would knowledge of the gaze position prevent this accident?*

As we saw, the important contributive elements of the AF72 accident on Tahiti are relative to the visual attention. Given this, would knowledge of the gaze position help in any way to prevent the accident? The technique of measuring eye movements is called *eye tracking* (see Chapter 3). So, would eye tracking provide a supplementary technological safeguard? My answer is positive. But how an integrated eye tracking system could have help to prevent the Tahiti accident?

First of all, by processing the ocular information and having the knowledge of the geometry of the cockpit, we can precisely deduce the position of the gaze of each pilot at each moment of time. The raw space coordinates are not particularly interesting, but rather an instrument of a part of an instrument on which a pilot is looking. Thus, such system can automatically generate messages like "Pilot 1 – Out of the window", "Pilot 2 – Altimeter", "Pilot 2 – FMA" etc. These data combined with the information about the flight phase and flight parameters can be used to generate visual or auditory notifications/alerts that remind the crew their respective roles, thus preventing the inadequate monitoring of the instruments.

In the case of the AF72 flight, for instance, a preventive ocular-based system would detect that the two pilots both focus on the same aspect of the situation, i.e. visual control of the trajectory by looking outside and ignoring the other information (such as thrust) important for a safe landing. The lack of monitoring of engine parameters or thrust reversers could also be detected because for a long time no fixation would have been detected neither from PM nor PF. An audio message such as "check thrust", for instance, could remind the pilots their responsibilities and prevent the accident. Nevertheless, such approach would require the conception of a knowledge database of the visual scanning rules with recommended or admissible durations and frequency of instrument consultations.

Don't we already know something about crew's gaze?

According to the voice recordings of the case study flight AF72, we can guess few visual checks that the crew performed. For example, when PF asks to read the indications of the vertical speed indicator, and then PM reports "Two point eight", we can fairly suggest that the pilot is looking at the variometer or have just looked at it. For this flight, we can manually identify about 40-50 fixations from the Cockpit Voice Recorder (CVR) transcript (BEA, 1999, Appendix 4), when pilots report instrument values by reading them out loud. Just 50 fixations during the half an hour recording being kept by the CVR! Suppose the pilots perform an average of 3 fixations per minute. Then, there should be around 10800 fixations corresponding to the period recorded by the CVR. A rapid calculation tells us that on the sole base of voice recordings we can deduce less than 0.5% of the crew's eye movements. That's not a lot! And for sure is not enough to initiate an investigation of the flight crew's visual scanning.

Note also that eye tracking has its limits and a fixation on an object does not necessarily mean the processing of the displayed information as in the considered case of AF72 flight and the FMA message that was read by the captain without being properly analyzed. However, all the information inside the cockpit is intertwined by flight dynamics and the aeronautic systems. Therefore, a non-analyzed bit of information will be compensated by another source of information that would give a hint or a complete explanation of the situation. In the considered case, the captain could monitor the engine power instead of looking outside to be aware of automatic go-around initiation.

1.2.4 *Conclusion on the lagoon landing accident*

In this section, we reviewed the accident involving the flight AF72. The case is not dramatic — nobody was hurt and the aircraft was not damaged that much — but it is quite instructive. The studied accident enters into two most popular categories of the HFACS classification that we described previously: decision errors and skill-based errors. A wrong decision to continue the approach, inadequate monitoring of the flight parameters — all these could be prevented using eye tracking and pupillometry and this thesis will answer how. In the next sections I will consider the fatal importance of monitoring skills in aviation and decision-making that can be hampered by crew's internal state.

Figure 5.: Inside the Antonov An-225 Mriya cockpit. Photo credit: Slava Stepanov

1.3 FATAL IMPORTANCE OF MONITORING SKILLS

1.3.1 *Visual environment of the cockpit*

It is enough to give a rapid glance inside an aircraft cockpit to understand the paramount importance that the vision has when flying an aircraft. Figure 5 immediately dazzles a viewer with the complexity of flight deck and countless dials, buttons and controls. Although real, this example is a bit grotesque because the illustrated cockpit belongs to the Antonov An-225 called Mriya (Ukrainian for *dream*) — the biggest ever build and still operated aircraft in the unique exemplar. Operated by a crew of 6, this aircraft with six engines requires for sure such a complex control. But while being quite anecdotical, this example of a cockpit is quite indicative of the complexity and the amount of visual information that airlines' crew have daily to cope with. In more modern designs[10] traditional infinite knobs and gauges were replaced by glass cockpits policy with multi-function digital displays. Yet, the beautiful layout of 32 gauges to inform the Boeing

10 The An-225 Mriya was constructed in 1988.

B-52 Stratofortress'[11] crew of its eight engines' parameters can still be admired by Stanley Kubrick's and ergonomics' lovers in Dr. Strangelove.

1.3.2 *Monitoring skills and failures*

The development of the automation systems impacted the role of the crew members which from direct (manual) controllers of an aircraft they rather became the system supervisors. However, the automation is not always fully understood nor surveilled correctly (Mumaw et al., 2001). Thus, the adequate active visual monitoring of the flight parameters in the cockpit is an essential piloting skill and becomes one of the most critical issues for flight safety (Sumwalt et al., 2015). As the Loss of Control Action Group supported by British Airways and other airlines put in their Guidance on the Development of Pilot Monitoring Skills (Civil Aviation Authority, 2013): "*Monitoring is an essential ingredient in achieving synergy with the highly automated and complex aircraft systems and effective crew co-ordination*".

It is no mere chance, that the term monitoring come from the latin verb *monere* — to warn, admonish, remind. The flight crew has to monitor the flight deck to be mindful of the current situation. The monitoring requires the observation (perception) and the interpretation (comparison against the expected values) of the flight path (including energy status), automation modes and all onboard systems. Naturally, the monitoring strategy differs according to the flight phase because the different flight systems are not of the same importance during the taxiing out or the descent, for instance. The monitoring skills should also include the cross-checking, the monitoring of the other crew member. One of the two crew members is explicitly designated as PM (quod vide side note on previous pages).

But maintaining an excellent and adequate monitoring performance is difficult to achieve. As Casner and Schooler (2015) showed in their recent study of airline pilots in a high-fidelity simulator, when pilots are free to monitor, monitoring lapses often occur. "*Whether we are diligent, distracted, or daydreaming, is monitoring doomed to fail?*" — the authors asked themselves. Thus, the National Transportation Safety Board (NTSB) in their accident report of the UPS Flight 1354 (NTSB, 2014a) underlined that flight crew monitoring skills are of critical importance in accident prevention; and that inadequate monitoring and/or cross-checking had occurred during more than 80% of major flight-crew-involved accidents between 1978 and 1990 (NTSB, 1994). But even if there is the comprehension of the fatal importance of an adequate monitoring, the accidents with monitoring lapses as contributing factors still occur.

In 2007, during an approach to Bournemouth (AAIB, 2009), the auto-throttle system disengaged, but it went unnoticed by the flight crew in spite of the red warning of disengagement flickering for one minute. The airspeed was not monitored either that would

11 B-52 has been operated by the Unites States Air Force since the 50s and is still in service.

have gave a hint. Eventually, the pilots had to initiate the go-around and performed a second stabilized approach.

In 2009, Colgan Air flight 3407, after an aerodynamic stall, crashed into a residence about 5 nautical miles of the Buffalo-Niagara International airport, New York (NTSB, 2010). All 49 persons on board were killed, one person on the ground was fatally injured; the airplane was destroyed. The report underlined that the major event contributing to the accident was the flight crew's failure to monitor airspeed. The report insists on prevention of flight crew monitoring failures and the need for enhanced training of PM skills.

In 2013, Asiana Airlines flight 214, after an unstabilized approach with the airplane below acceptable glide path and excessive descent rate, struck a seawall at San Francisco International Airport (NTSB, 2014b). Three passengers were fatally injured; 49 other people received serious injuries; the airplane was destroyed by the impact and the subsequent fire. The report concluded that one of the causes of the accident was the flight crew's inadequate monitoring of airspeed. Also, the NTSB underlined that the contributing factors were the PF's inadequate training on the visual approaches and the PM's inadequate supervision of the PF.

One of the proposed solutions to prevent the failure of monitoring is the pilot training. Recently, the Federal Aviation Administration (FAA) published a final training rule that requires enhanced pilot monitoring training to be incorporated into existing air careers training programs (FAA, 2013), the compliance date being March 2019.

1.3.3 *Factors that hamper monitoring*

Why pilots sometimes turn out to be unable to properly monitor the flight deck? Besides external factors such as not ergonomic design, the biggest jeopardy for an adequate monitoring are internal factors that relate to vulnerabilities of the brain and mind. Among such factors are lapses of attention, inavailable attentional resources, low arousal, anxiety or other stressors. The table 1 gives few examples of the accidents where monitoring lapses were considered as a contributory factor. It is also detailed what factors hampered adequate monitoring.

Importantly, inadequate monitoring often implies decision-making errors. Decisions from the simplest motor tasks to complex life-impacting judgments are ubiquitous in our lives. But a decision is never isolated by itself and is being made in some context. Arousal, alertness, the load on memory, emotions — lots of ingredients contribute to the decision-making process Bechara et al. (2000); Dunn et al. (2006); Harrison and Horne (2000); Miu et al. (2008). Situation uncertainty, information overload, multitasking, time pressure and fatigue may all impose a heavy demand on cognition and impair decision-making. The process of decision-making itself is out of the scope of this thesis, but we will be interested in detecting the dangerous internal states (such as high working

Table 1.: Aeronautical accidents with monitoring lapses as a contributory factor. Adapted from (Civil Aviation Authority, 2013)

Where & When	**Failure to Monitor**	**Dominant Factor**	**Other Factors**	**Fatalities**
San Francisco, 2013 (NTSB)	Low Speed	Inattention	Fatigue	3
North Atlantic Ocean, 2011 (BEA)	Flight Path parameters, AP selection	Startle	Distraction	-
Atlantic Ocean, 2009 (BEA)	Speed Inconsistencies, Flight path parameters	Startle	Distraction, Inattention	228
Buffalo, 2009 (NTSB)	Low Speed Indication	Distraction	Fatigue	50
Schipol, 2009 (Dutch Safety Board)	Low Speed Indication, Autothrottle Disconnect	High Workload	Distraction	9
Bournemouth,2007 (AAIB)	Autothrottle disconnect	Inattention		-
Indonesia, 2007 (NTSC)	Flight path	Disorientation	Tunnel Vision	102
Cali, 1995 (Aeronautica Civil)	Flight Path	High Workload	Inattention	159
Palmerston North NZ, 1995 (TAIC)	Flight Path	Lack of attentional resource	Distraction, Tunnel Vision	4
Everglades, 1972 (NTSB)	Flight Path	Distraction	Workload	99

memory load) that can compromise the decision-making. The decision-making plays indeed as crucial role as monitoring skills. In their analysis of major flight-crew-involved accidents from 1978 to 1990 (NTSB, 1994), the NTSB identified 51 tactical decision errors occurred in 25 out of 37 accidents. The majority (49) of the tactical decision errors were made by the captains while serving as PF (44). A half (26) of these errors were the omissions, i.e. the failure to undertake an action when the situation demanded change. Between the omission errors, 16 involved the failure to initiate a go-around procedure.

1.4 LOOKING INTO THE EYES — A POSSIBLE SOLUTION?

The great challenge for research institutes and civil aviation safety organizations is to further improve the industry's safety by considering the complex attentional and cognitive processes underlying pilots' decision-making and monitoring skills. The understanding of how the brain operates in complex realistic situations is necessary because simply observing the behavior is not enough. The approach known as Neuroergonomics (Mehta and Parasuraman, 2013) aims to design ergonomic systems for safer and more efficient operations using the knowledge of the functioning of the human brain. One of the challenges of the Neuroergonomics is to design more efficient systems and to find valid and robust measures of human attentional and cognitive behavior in order to mitigate the adverse effects of attentional failure, stress or fatigue. While electroencephalography, electrocardiography (Borghini et al., 2014) or functional infra-red spectroscopy (Gateau et al., 2015) are often used in flight/drive simulators to study the cognitive and attentional states, none of these techniques can properly take measures remotely without direct contact with human body. Moreover, these techniques are prone to several signal noise challenges and inter/intra-individual variability issues. Meanwhile, this is one of the essential criteria to measure human attentional and cognitive behavior inside the cockpit, especially during emergency situations. It is indeed difficult to imagine the captain of the tragically known flight AF-447 (BEA, 2012), returning to the cockpit after the in-flight rest, to calmly put on some wearable device and to proceed to a calibration while there are invalid speed indicators and stall warning.

1.5 CONCLUSION

As discussed in this Chapter, the aviation being today the safest transportation mode is still not lacking in some number of incidents and accidents. About half of these accidents include failures in monitoring or faulty decision-making. Such factors as arousal, workload and fatigue can hamper an adequate monitoring of flight instruments and good decision-making. Therefore, one promising avenue to further enhance the flight safety is looking into the pilots' eyes. At one hand, eye movements tell us about visual scan patterns and monitoring strategies, as well as give us a hint about flight

crew's internal state (see Chapter 3). At another, pupil diameter brings a light on our state of arousal, working memory load, emotional state etc. (see Chapter 2). All these factors jeopardize decision-making and monitoring and in order to mitigate their adverse effects there is a need for a robust measure of human attentional and cognitive behavior and the pupil diameter can be the one.

2

WHAT CAN PUPIL DIAMETER REVEAL ABOUT PILOTS?

О, глаза — значительная вещь. Вроде барометра. Всё видно.

Собачье сердце, Михаил Булгаков

Oh, eyes are very important. Something like a barometer. You can see everything...

The Heart of a Dog by MIKHAIL BULGAKOV
Translated by Avril Pyman

In Chapter 1 we saw that many accidents are due to excessive arousal, high working memory load or emotional state that can jeopardize adequate monitoring and decision-making. All these internal factors can be monitored using pupillometry — the study of pupil diameter.s. In this chapter, after a brief historical background, we first present readers with the pupil physiology. Next, we discuss the relationship of the pupil diameter changes with the level of cognitive load and the light conditions. We also report recent findings of the relationship between Locus-Coeruleus Norepinephrine (LC-NE) system and the pupil diameter, tending to explain the cognitive pupil reaction. Eventually, we present two models of pupillary response and different techniques of measuring pupillary changes.

2.1 HISTORICAL BACKGROUND

How often did you looked into the lovely eyes of the love of your life and said to yourself: “Ain’t she wonderful?" What happened before it? The light reflected from your love’s eyes passes through the pupil and fell onto your retina to form the image. However, the pupil is far more useful than just an aperture to pass the light. The pupil dilation which occurs in the absence of environmental changes have been noted as early as in the mid-1700s, see the Loewenfeld (1958) for a complete historical review. Also,

the observation that larger pupils are associated with sexual interest was used by Italian women in the XVII century long ago before its scientific report (Hess, 1965). The cunning consisted of putting drops of *Atropa belladonna* into the eyes to dilate the pupils. Thus, men being more attracted to women with larger pupils, it gave its Italian name to this toxic plant its name, *bella donna* — beautiful woman. A few reports of pupil dilation to non-visual stimuli were done by German neurologists in the XIX century. However, the first systematic scientific study using pupillometry was conducted by Hess and Polt in 1960. Since then, there was a great influx of interest to the pupillometry, especially after the work done by Daniel Kahneman (1973) and Jackson Beatty (1982b).

Today, pupil diameter is a widespread indirect measure of neural activity. It is often referred to as pupillometry. While in the 60th, researchers had to measure pupil size using a ruler and a film/photos projection, nowadays pupil size is available with the most standard eye-tracking systems (see the next Chapter 3). Therefore, it is the least invasive measure of neural activity because there is no direct contact with the skin or body. In this chapter we present basic notions of pupillary physiology, highlight the relationship between the pupil diameter and the locus-coeruleus norepinephrine system. We also give a brief review of works concerning the relationship between pupil diameter and cognitive load.

An interesting fact

Hess and Polt started their pioneering work 1960 with a quote of lines of Guillaume de Salluste: *"These lovely lamps, these windows of the soul"*. While the phrase that "The eyes are the windows of the soul" became ubiquitous nowadays, the interesting fact is that Guillaume de Salluste du Bartas was an Occitan poet born in the XVI century in Montfort, 60 kilometers away from Toulouse, where I performed the research presented in this dissertation.

2.2 PUPIL PHYSIOLOGY

The pupil is an opening in the center of the iris — the most beautiful part of the eye that looks like an unknown galaxy. Therefore, the pupil by itself is not a structure at all, but an aperture in the iris which contains the muscles that control the pupil size — *radial dilator pupillae* and *circular sphincter pupillae* (Figure 6). These two opposing groups of smooth muscles are controlled by, respectively, sympathetic and parasympathetic nervous systems. Receiving input from both sympathetic and parasympathetic nervous systems, the aperture in the iris provides interesting information about autonomic nervous system activity. The parasympathetic excitation and/or sympathetic inhibition result in pupil constriction, while sympathetic excitation and/or parasympathetic inhibition re-

sult in pupillary dilation (Beatty and Lucero-Wagoner, 2000; Steinhauer et al., 2004). Thus, the pupil behavior is hypothesized to reflect the balance of relative sympathetic and parasympathetic activation. The main function of the iris aperture is to regulate the amount of the light falling onto the retina to keep the image of high quality. Pupil size varies on average from 1 mm to 8-9 mm. The pupil can react to stimuli in just 200 ms with response peaking in 500-1000 ms (Andreassi, 2000; Lowenstein and Loewenfeld, 1962).

The study of pupil reflexes (direct light reflex, indirect or consensual light reflex, near reflex) is a useful tool in ophthalmology to determine whether there is a disorder due to ocular or cerebral causes. However, besides the clinical application, pupillometry is a wide-used tool to study the human cognition.

2.3 PUPIL DILATION, COGNITIVE LOAD AND LIGHT CONDITIONS

Different correlations between pupil activity and attentional effort or cognitive processing have been established from early sixties (Hess and Polt, 1960, 1964; Kahneman and Beatty, 1966) to nowadays (Ariel and Castel, 2014; Causse et al., 2010; Kang et al., 2014; Naber and Nakayama, 2013). See Beatty (1982b), Beatty and Lucero-Wagoner (2000), Goldwater (1972), Andreassi (2000) and Laeng et al. (2012) for exhaustive reviews of different established relationships between cognition and pupil diameter.

As Daniel Kahneman noted in his work on attention and effort (1973), a good measure of mental activity must satisfy three criteria: it should be sensitive to between-task and within-task variations as well as between-subject differences. The measure of pupil diameter seems to satisfy all three of these criteria (Andreassi, 2000; Beatty and Lucero-

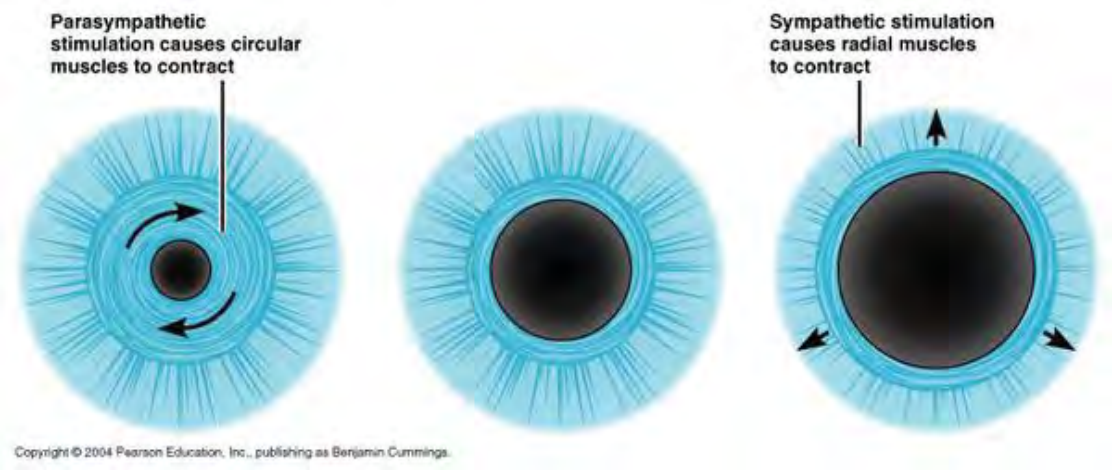


Figure 6.: The action of dilator pupillae and sphincter pupillae — two opposing group of smooth muscles that control pupil size.

Wagoner, 2000; Kahneman, 1973). Nevertheless, the pupil is sensitive to many other factors besides the mental effort. Tryon (1975) listed 23 factors that influence pupil size among which luminosity variation is most commonly cited. Thus, in spite of the successful application of pupillometry under controlled laboratory settings, this psychophysiological proxy is still not available to infer cognitive states in realistic environments (e.g. flight deck, air traffic control radar room).

Comparing results of studies from different laboratories, Beatty in his review (1982b) concluded that, TEPR's amplitude was a consistent index of cognitive activity despite various illumination conditions across studies. However, the interaction between luminance and the extent of pupillary cognitive dilation has remained unclear. Bradshaw (1969) conducted a simple reaction task under two different light conditions and found no difference in dilation peaks. More recently, Pomplun et al. (2003) and Xu et al. (2011) measured pupillary changes during a visual task experiment and an arithmetic task. No interaction between cognitive load and luminance conditions were found in these studies.

More than that, whereas it has been known that pupil reacts in response to luminance changes in the environment (Loewenfeld and Lowenstein, 1993), it should be also noted that even when investigators control light conditions (ambient, computer screen etc.) and maintain them constant, the pupil muscles also respond to the luminance of the fixation point neighborhood (Pereverzeva et al., 2012). For example, while observing an image of a white cup filled with black coffee, a fixation of our gaze on the black beverage will cause the iris aperture to dilate in comparison to a fixation on the cup. Palinko and Kun (2012) described an example of an experiment taking this pupillary light response into account. Three targets of different shades of gray were displayed on a static image. The participants performed a vigilance task while focusing on one of three presented targets. The cognitive component reflecting the level of vigilance was identified by subtracting the pupillary reaction due to the change of focus between two targets of different luminance. The results of this experimentation showed that the cognitive component of pupil reaction was still present despite the occurrence of the light reflex. Therefore, it was possible to detect the cognitive effort even when it interfered with a pupillary light response. Nevertheless, this method is condition-locked since the luminance conditions were controlled and an averaged pupillary light response was obtained for each condition. In a real environment, such transitions from one level of luminance to another are limitless and it is impossible to obtain an averaged pupillary light response for each condition. Another method to identify the cognitive pupillary component by calculating an independent component analysis was proposed by Jainta and Baccino (2010). Although such methods could be successfully applied for post-analysis, they are inconvenient for on-line estimation of mental effort. Another method, convenient for on-line workload estimation, called "Index of Cognitive Activity" (Marshall, 2002) identifies fast pupillary dilations using wavelet analysis and counts their occurrences per

second. It was reported to be sensitive to workload but not light conditions (Marshall, 2002). Even though no experimental study was conducted, Marshall manipulated both screen luminance and room illumination to obtain dark and light conditions and reported that the index of cognitive activity was insensitive to luminance change. Furthermore, a recent study on nonhuman primates (Hampson et al., 2010) confirmed that fast pupillary dilations correlated with neurons firing in the frontal cortex, hence suggesting a neural basis for the index of cognitive activity.

Nevertheless, very few studies have addressed the interaction issues between the TEPR and the light conditions (e.g. screen luminance, room illumination). Steinhauer and colleagues (2004) conducted an experiment under moderate room light and in darkness. They found an influence of task difficulty on overall pupil diameter during recording in light. Benedetto and colleagues (2014) manipulated both screen luminance and ambient illuminance during digital reading. They found that the screen luminance had a more significant effect on pupil diameter compared to ambient illuminance. Authors proposed to explain it by proximity and concentration on the computer screen. Hence, the challenge is to define a pupil-based measure of cognitive load that would be independent of light conditions. Indeed, frequency analysis of pupillary response could bring to light its hidden behavior that reflects autonomic nervous system's reaction to light and cognition (see Chapter 4).

2.4 PUPIL DIAMETER AND LOCUS COERULEUS

While for a long period of time, observation of cognitive pupil dilation was simply empirical, recent evidence of a relationship between the pupil diameter and the LC-NE system pour new light to the understanding of cognitive pupillometry. The Locus-Coeruleus (LC) is a small nucleus in the brainstem that provides the majority of norepinephrine to the brain (Aston-Jones and Waterhouse, 2016; Samuels and Szabadi, 2008) and plays a crucial role in cognition (Sara and Bouret, 2012). The firing rate of the LC-NE system was found to be closely correlated to the pupil diameter in primates (Rajkowski et al., 1993) by direct recordings of LC neurons firing rate (Figure 7). More recently, this relationship was established in humans (Murphy et al., 2014) in an fMRI study.

The most recent reviews of the complex relationship between pupil size and LC-NE system can be found in Joshi et al. (2016) and Costa and Rudebeck (2016) where the authors discussed the direct and indirect evidence supporting this relationship. Interestingly, Joshi et al. (2016) found that the correlations found previously are not certainly specific to the LC-NE but also present in the inferior and superior colliculi and anterior and posterior cingulate cortex. However, because these brain regions are highly interconnected with the LC-NE system, the results nevertheless suggest that the pupil diameter reflects LC-mediated coordination of neuronal activity.

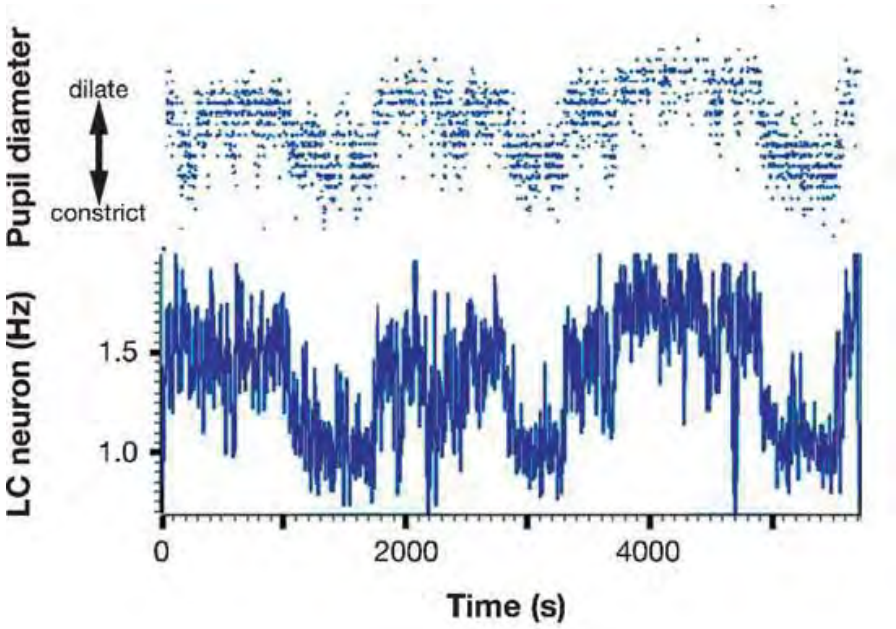


Figure 7.: Relationship between tonic pupil diameter and baseline firing rate of an LC neuron in monkey. Image from Aston-Jones and Cohen (2005)

2.5 MODELS OF PUPIL BEHAVIOR

In this section, we present two very different models of pupillary response. The first one, proposed by Usui and Stark, associates the neural input (sympathetic and parasympathetic nervous activities) and the pupil diameter. The second one, proposed by Hoeks and Levelt, associates the attentional effort, modeled as a series of impulses of different amplitudes, and the pupil diameter. The nonlinear Usui-Stark model is mechanical and based on anatomical and physiological evidence while the linear Hoeks-Levelt model is purely empirical. While the Usui-Stark model is suitable for estimation of neural activity based on the pupillary flash response, the Hoeks-Levelt model is suitable for estimation of attentional impulses based on the continuous pupillary signal. None of these models takes into account the complex interaction between light conditions and pupil reaction (see Chapters 4 and 5).

2.5.1 *Model of Usui and Stark*

Usui and Stark proposed in 1982 a parametric model of iris muscles controlling the pupil size. This forward nonlinear dynamic model of pupillary muscle plant connects the activity of the autonomic nervous system (two inputs: sympathetic and parasympathetic nervous activities) and the pupil diameter. Both the sphincter and the dilator are

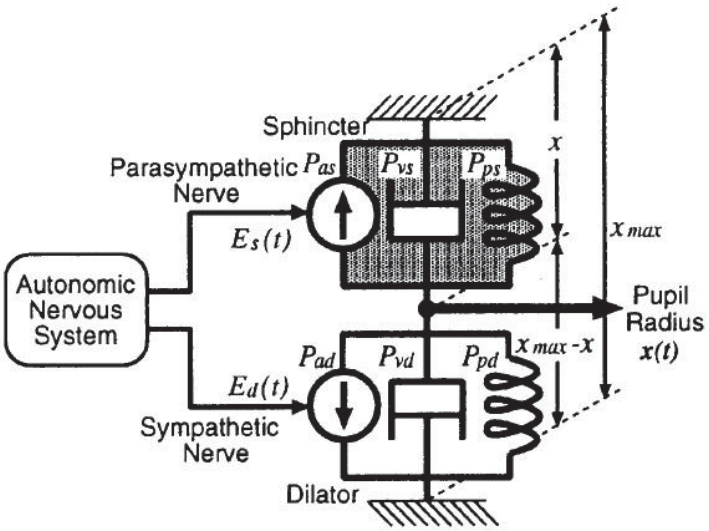


FIGURE 1. Nonlinear dynamical model for the human pupillary muscle plant. $E_s(t)$ and $E_d(t)$ represent the parasympathetic and sympathetic nervous activities to the sphincter and the dilator, respectively. x is the pupil radius and x_{max} is the maximum pupil radius.

Figure 8.: Illustration of the Usui-Stark model from (Usui and Hirata, 1995).

presented in terms of viscoelastic elements and a tension generator and are described by the same function with different parameters. The parameters were estimated by fitting the model response to experimental data. Note that the Usui-Stark model is complex (described by a system of 18 equations), and needs the estimation of numerous parameters. Usui and Hirata (Usui and Hirata, 1995) described how to use the inverse model of the pupil muscle plant and to estimate the neural input based on the pupillary signal.

To the best of our knowledge, the model of Usui-Start was always used on the pupillary flash response. Furthermore, the Usui-Stark model simulated the human pupillary muscle plant and is not suitable to separate visual from non-visual and cognitive from non-cognitive pupillary effects.

2.5.2 *Model of Hoeks and Levelt*

Hoeks and Levelt proposed in 1993 a model of pupillary dilation as a consequence of attentional effort. In their methodological paper, the authors proposed to model the attentional input as a string of attentional pulses. The authors supposed that the system is linear, i.e. if for an attentional input x_1 and x_2, the corresponding output is y_1 and y_2 respectively, then the output for $\alpha_1 x_1 + \alpha_2 x_2$ equals to $\alpha_1 y_1 + \alpha_2 y_2$, where α_i are some

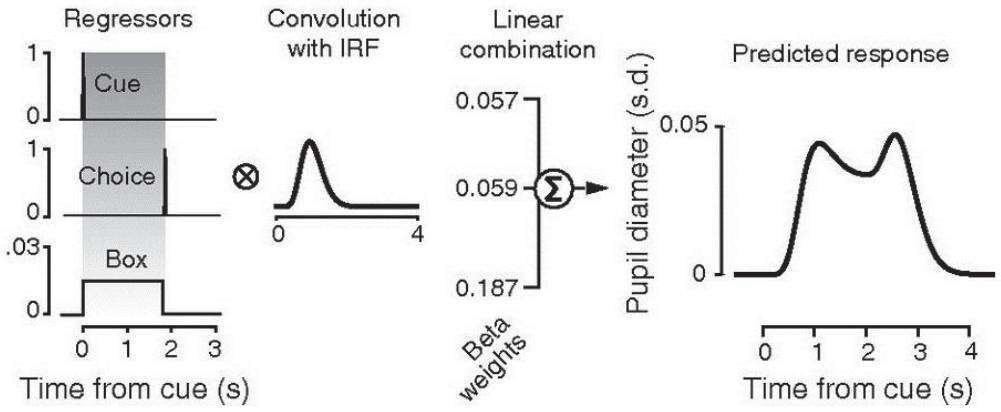


Figure 9.: Modeling the pupil response (example trial). Image from (de Gee et al., 2014).

constants. The goal was to find the impulse response of the pupillary system so that the output $y(t)$ and the input $x(t)$ will be related as follows:

$$y(t) = h(t) * x(t),$$

where $*$ is the convolution operator. Hoeks and Levelt proposed to model the pupil system as a Erlang gamma function

$$h(t) = t^n e^{-nt/t_{max}} \mathbb{1}_{\mathbb{R}_+}.$$

The authors estimated the two free parameters n and t_{max} on the base of recording of pupillary responses of 8 subjects to 1000-Hz auditory tones. The resulted mean values are $n = 10.1$ and t_{max}=930. Using this model, one can compute the string of attentional input given some pupillary response (using the deconvolution method).

There are two notable examples of the use of this model by Hoeks and Levelt (de Gee et al., 2014; Wierda et al., 2012). First, Wierda and colleagues (2012) demonstrated in an attentional blink paradigm that the pupil dilation deconvolution can be used to track the attentional processes at high temporal resolution. The authors also discussed that a bimodel version of Hoeks-Levelt model can be used as proposed by O'Neill and Zimmermann (2000). In this more complex model, both parasympathetic and sympathetic components are taken into account. However, this significantly increases the number of parameters to estimate. More recently, de Gee and colleagues (2014) used the Hoeks-Levelt model to study the decision-making in a visual task. To test their hypothesis, the authors fitted the model with different regressors (Figure 9), i.e. different supposed attentional inputs. By statistically comparing the beta weights, the could conclude on the origins of the differences in pupillary response. This elegant application of the Hoeks-Levelts model showed the interest of such approach.

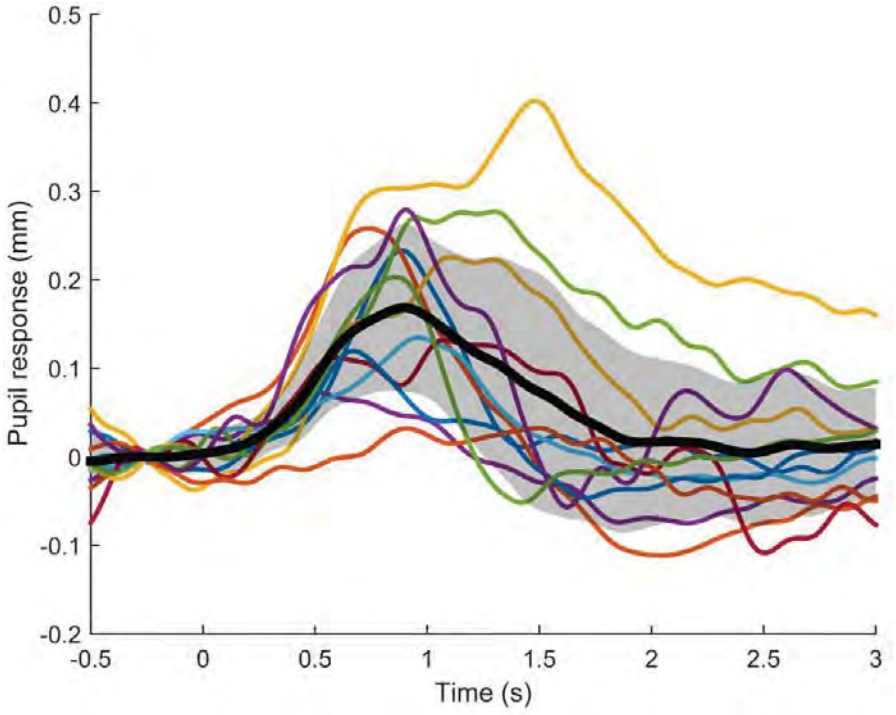


Figure 10.: The averaging of pupil responses: individual trials represented in colored thin lines, their point-by-point average is represented in black thick line. The shadow represents the point-by-point standard deviation.

2.6 ANALYSIS OF THE PUPILLARY RESPONSE

2.6.1 *Pupil response in the time domain*

The pupillary signal is commonly analyzed both in time and frequency domains. In the time domain, one of the most current techniques of pupillary signal analysis consists in comparing the TEPR, a notion generalized by Beatty (1982b). The TEPR is an averaged stimuli-locked pupil reaction reported to a short pre-stimulus baseline. It could be compared to an event-related potential — electrophysiological brain response — where pupillary "potential" is measured in pixels or millimeters (Figure 24). As event-related potentials, the pupillary response is extremely variable from one trial to another and there is a need of repeating the stimulus and averaging the response to it. Whereas for motor tasks, 5-10 repetitions of a stimulus might be enough, for complex cognitive paradigms, an average of 15-20 repetitions is a better idea.

The notion of the TEPR proposed by Jackson Beatty should be revised nowadays. As an analogy of the ERP — event-related potential, where the value of the baseline is not of interest, — the notion of TEPR has it value at the time when researchers thought that

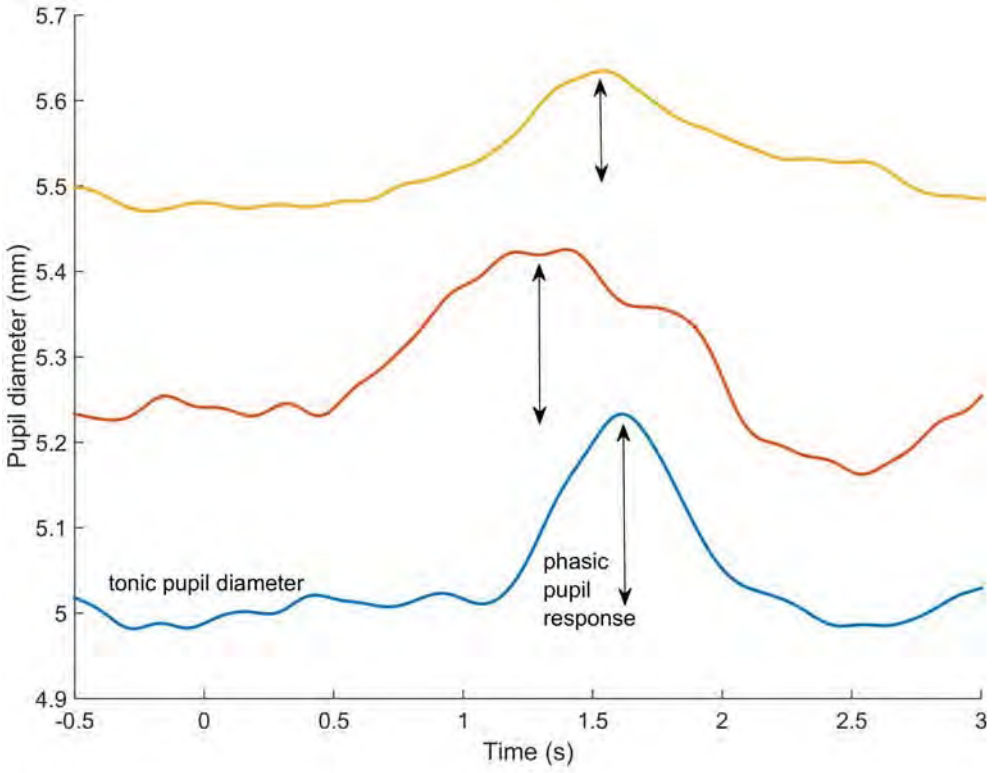


Figure 11.: Examples of the two components of pupillary reaction: tonic pupil diameter and phasic pupil response. The trial in blue has the smallest tonic pupil diameter and the largest phasic pupil response. On contrary, the trial in yellow has the largest tonic pupil diameter and the smallest phasic pupil response.

the pupillary response is independent of pupillary baseline level. It is now established that it is not the case anymore (see Chapter 5), therefore, we should adopt the notions of *tonic pupil diameter* and *phasic pupil response*. These notions were also proposed by Jackson Beatty (1982a). In most cases, phasic pupil response will correspond to what is called TEPR, and tonic pupil diameter to what, in studies using the notion of TEPR, is called “baseline diameter" (see also Introduction in Chapter 5).

The pupillary reaction can be divided into two components: tonic pupil diameter and phasic pupil response (Beatty, 1982a). Tonic pupil diameter reflects a sustained component of the pupillary response and is expressed as an absolute pupil diameter. Often, tonic pupil diameter is used as basal pupillary diameter. In turn, phasic pupil response refers to a transient component of the pupillary response and is expressed as dilation relative to some basal pupil diameter. While the typical order of magnitude of the tonic pupil diameter is 1 mm, that of phasic pupil response is 0.1 mm. Figure 11 shows an

example containing three trials with pupillary reactions of different tonic pupil diameters and phasic pupil response. The trial in blue has the smallest tonic pupil diameter and the largest phasic pupil response. On the contrary, the trial in yellow has the largest tonic pupil diameter and the smallest phasic pupil response. The trial in orange has medium tonic pupil diameter and phasic pupil response similar to those of the blue trial. Many authors stated that the magnitude of phasic pupil response to a given task was independent of tonic pupil diameter (Beatty, 1982b; Bradshaw, 1969; Hahnemann and Beatty, 1967). Thus, given the presumption of the independence of these two pupillary components, Beatty (1982b) concluded that it is possible to compare the phasic pupil responses issued from various set-ups and reported by different laboratories. Notably, in the review, he presented a table of quantitative comparison of qualitatively different cognitive tasks (memory, language, reasoning and perception). The table confronted the results obtained by different researchers and permitted to see that, for example, the storage in memory of four words makes pupil dilate more than that of a multiplicand, which is roughly equivalent to retaining in memory two digits. According to the corresponding pupillary reactions, it also put an easy multiplication problem higher (phasic pupil response about 0.1 mm larger) than a hard auditory discrimination task. However, one may call in question such ordering assuming that multiplication of two digits is sometimes easier than detection of a deviant sound. Such task classification, using the magnitude of phasic pupil response as a marker of difficulty, would prevail but on one condition; if tonic pupil diameter does not impact phasic pupil response. Suppose, indeed, that tonic pupil diameter varies as a function of the experimental setup at one hand, and phasic pupil response depends on tonic pupil diameter at another. In this case, in order to compare results issued from different experimental setups one should first make sure that the conditions were the same or at least similar. The investigation of these questions is of an importance when using pupil reaction as a marker of stress or workload in ecological conditions where such factors as light are difficult to control. If there exists a strong relationship between tonic pupil diameter and phasic pupil response, the transportability of laboratory results into real life conditions for applications such as human factors in aviation needs a whole reflection apart.

The dependence of the extent of a physiological reaction to an event on the prestimulation basal level was named *law of initial value* in the fifties (Lacey, 1956; Wilder, 1967). Lacey (1956) postulated that a high autonomic excitation before a stimulus would affect the reactivity and diminish the response but did not refer to the pupil, talking rather about skin resistance, heart rate, blood pressure, muscle potentials, etc. Recently, few mentions of law appeared in pupillometric studies (Gilzenrat et al., 2010; Höfle et al., 2008; Van Gerven et al., 2004). Formulated in terms of tonic and phasic components of pupillary response, the law of initial value would postulate that a large tonic pupil diameter would imply a smaller phasic pupil response. Yet, Jin (1992) argued that the law of initial value should be revisited as follows: *The higher the initial value, the greater the*

organism's following reactivity, although a tendency to reversed responses may occur when the initial value reaches its upper extremity. Therefore, Jin proposed to consider the law of initial value as a restriction of pupillary dynamic range, i.e. when the pupil is already large, it cannot dilate further. Thus, the direction of the law is still questionable.

The tonic pupil diameter has numerous sources of variation (Tryon, 1975). For instance, it is modulated by general organism's arousal, sustained cognitive load, or light conditions, both ambient illumination and focal luminance (see Section X). When tonic pupil diameter is modulated by vigilance state, an inverted relationship between tonic and phasic pupil diameters was found by Gilzernat et al. (2010) in an auditory oddball task. The authors discussed this finding with regard to the law of initial value but considered it as exclusively mechanical. Therefore, the authors verified if the inverse relationship between tonic pupil diameter and phasic pupil response held true when tonic pupil diameter was modulated by light conditions and proved it false in that case. This finding was afterward confirmed by Murphy et al. (2011) also in an auditory oddball task and, more recently, by de Gee et al. (2014) in a perceptual decision-making paradigm and Knapen et al. (2016) in an auditory vigilance task. As for cognitive tasks implying working memory, Steinhauer et al. (2004) found that the phasic pupil diameter was modulated by ambient illuminance when engaged in sustained processing. More recently, Peysakhovich et al. (2015a) found that the phasic pupil diameter was modulated by the screen luminance in a short-term memory task. Most recently, Pfleging et al. (2016) also studied pupillary response, manipulating illuminance and luminance during a cognitive task. However, the authors used a one-factor-at-a-time method that does not enable the investigation of the illuminance-luminance interaction and reported exclusively the absolute pupil diameter values making impossible to compare tonic and phasic pupil responses. Altogether, in order to be able to compare pupil reactions issued from different studies that maintain different light conditions, and to transport the laboratory results into real-life applications, it is important to further investigate the relationship between the tonic and phasic components of the pupillary response and the factors that modulate these components. The pupillometry literature still has not given a clear answer on these questions and a further investigation is needed. To the best of our knowledge, no studies investigated the impact of luminance on the tonic and phasic components of the pupillary response during sustained cognitive load.

2.6.2 *Pupil response in the frequency domain*

In the frequency domain, pupillary oscillations drew researchers' attention starting with Lowenstein's work (1963) describing pupillary fatigue waves that were quantified later by Lüdtke (1998). Since then, spectral analysis is often applied to analyze pupillary data. Pupillary spectrum is used either alone or together with other psychophysiological markers, for example, as an input for an artificial neural network (Ren et al., 2013). A

few studies reported an increase of Power Spectrum Density (PSD) of the pupillary signal under mental workload conditions compared to the rest or a control condition. Thus, Nakayama and Shimizu (2004) found a significant increase of PSD in the frequency bands of 0.1–0.5 Hz and 1.6–3.5 Hz within calculation tasks. Fourier analysis of pupillary response is often applied in psychopathology (Grünberger et al., 1999; Grünberger, 2003) and sleep research (Lüdtke et al., 1998; Wilhelm et al., 1998), where researchers often have a long pupillary record that permits a precise frequency analysis. According to these studies, the amplitudes of pupillary oscillations are proportional to cognitive activity. Eventually, irregularities of pupillary oscillations can give insights into the activation of sympathetic/parasympathetic nervous systems and the balance between them. For instance, Lew and colleagues (Lew et al., 2008) discussed a possible application of short-time Fourier transforms for pupillary data analysis in order to investigate the changes of frequency components across time.

Along with pupil diameter, heart activity is another widespread psycho-physiological proxy. As the pupil, the heart is influenced by both parasympathetic and sympathetic activity, decreasing and increasing its rhythm, respectively. The cardiac activity can be described by the heart rate variability method. The analysis of power spectrum of R-R intervals reveals two particular frequency components of cardiac activity: blood pressure variation located within a Low Frequency (LF) band from 0.04 up to 0.15 Hz, as defined by European Society of Cardiology and the North American Society of Pacing and Electrophysiology (1996); and respiratory sinus arrhythmia located within a High Frequency (HF) band from 0.15 up to 0.40 Hz. Variability of the HF component is associated with the parasympathetic activity, while the LF band is thought to be under both sympathetic and parasympathetic control with dominant sympathetic influence (Billman, 2011). Since the proposition of Pagani and colleagues (Pagani et al., 1986), the ratio of cardiac signal powers within LF and HF bands, referred as LF/HF or LH ratio, although controversial (Billman, 2013), is often used to measure the sympatho-vagal balance. Some researchers reported an increase in LF/HF ratio induced by mental effort (Durantin et al., 2014; Mizuno et al., 2011; Mukherjee et al., 2011). Since LF/HF ratios in these studies were greater than 1, this increase could be interpreted as a greater increase in the LF band compared with the HF band.

Calcagnini and colleagues (1997) simultaneously recorded pupillary and cardiac activity and reported that cardiovascular rhythms were contained in the pupillary signal. Since then, there have been a few applications of LF/HF ratio technique for pupillary analysis. For example, Murata and Iwase (1998; 2000) reported increasing LF/HF ratio for pupil oscillations with increasing mental workload, using mental arithmetic and Sternberg short-term memory tasks. To compute the LF and HF activity, they used bands from 0.05 up to 0.15 Hz and from 0.3 up to 0.5 Hz, respectively. More recently, the ratio was used in a motor repetitive task (Reiner and Gelfeld, 2014). Authors found

that the LF/HF ratio decreased with the number of repetitions, indicating a decrease of mental workload (probably due to a habituation effect).

2.7 CONCLUSION

In this Chapter, we revised the physiology of the pupil, its relationship with the LC-NE system, and different findings and the impact of cognitive load on pupil diameter. We underlined that the effect of light conditions is not enough investigated and should be taken into account when using models of pupil behavior (such as those presented in this chapter). We also presented different techniques of pupillary signal analysis and find out that the pupillary signal is analyzed more often in time domain compared with the frequency domain. In next Chapter, we will review the eye-tracking methods and will discover why and how we move our eyes.

3

WHAT CAN EYE MOVEMENTS REVEAL ABOUT PILOTS?

> They can't tell so much about you if you got your eyes closed.
>
> *One flew over the cuckoo's nest* by KEN KESEY

In Chapter 1 we also saw that vision plays an important role in piloting and understood that many accidents resulted from poor monitoring issues or lack of visual circuits abilities. These observations naturally bring into light the eye-tracking technology that has two facets: pupillometry (reviewed in previous Chapter 2) and eye movements. In this Chapter, we will first give a brief historical background and a presentation of existed and existing eye-tracking methods. Then, we will explain why and how do we move our eyes. Eventually, we will present a reader with an investigator game during which we will reconstruct events during a period of flight based on the eye movements.

3.1 HISTORICAL BACKGROUND

Eye tracking — the technology of recording eye movements — is barely new. Studies on eye movements are being published for over 60 years. But the market of commercial eye trackers has been considerably democratized during the last decade, and such devices are today non-invasive (can be remote without direct contact with the pilot), relatively low cost, reliable enough, and, importantly, eye tracking data can be available in near real time. Eye tracking is used nowadays in Neuroscience, Psychology, Industrial Engineering and Human Factors, Marketing/Advertising, and Computer Science (Duchowski, 2002). All these advantages make the idea of the integration of eye trackers as a tool to monitor and to enhance pilots' performance in modern flight decks quite realistic.

In the pioneer studies by Fitts and colleagues in the 50th (Fitts et al., 1950; Jones et al., 1949; Milton et al., 1949), researchers had to be innovative and to record eye movements using a system of mirrors and film recorders that were treated manually. Such analyses were fastidious and fatiguing. By contrast, today even a barely experienced student can easily record eye movements and analyze the data with a few mouse clicks. We are

persuaded that the eye tracking technology will be ubiquitous before too long in the automotive and aeronautical industries. An example of a successful integration is Seeing Machines' Driver Safety Solution[1] that tracks truck driver's eyes and face direction to detect when he is about to fall asleep. This solution is already functional for many years and is used by such giants of industry as Caterpillar[2]. In the aeronautical domain, let me cite Lufthansa Systems that announced[3] an integration of eye tracking for their NetLine/Ops++ solution used by Lufthansa operations controllers (and 69 other airlines around the worlds) that are responsible for monitoring and managing flight schedule. The real-time analyses of eye movements will provide operators, which use up six monitors, with additional visual or auditory alerts when they did not notice an important message.

3.2 EYE TRACKING METHODS

In this section, we briefly present the existed and modern eye tracking methods. All the essential information about eye tracking technology can be found in the two "Bibles" of the domain wrote by Duchowski (2007) and Holmqvist et al. (2011).

3.2.1 *Visual observation*

The oldest technique for the study of eye movements is a simple visual observation. It is, of course, qualitative and not quantitative, and small eye movements such as 1° rotation will go unnoticed. Yarbus (Yarbus, 1967) wrote in his work: "*Some authors (Javal, 1879) used a mirror for this purpose. Observations were made on the image of the eye in a mirror. The experimenter stood behind the subject and did not distract him during the experiment.*" Later, optical instruments were used to magnify the eye image for more precise observations.

3.2.2 *Mechanical recording*

Another outmoded method is a mechanical recording of eye movements, i.e. the connection between the eye and the recording device, in this case, was mechanical (Delabarre, 1898). Participants wear a cup with an aperture in the center; a lever or thread was at-

1 https://www.seeingmachines.com/solutions/fleet/

2 Caterpillar Global Mining forms alliance with Seeing Machines to deliver operator fatigue monitoring technology (Release Number: 228PR13). http://www.cat.com/en_MX/news/machine-press-releases/caterpillar-globalminingformsalliancewithseeingmachinestodeliver.html

3 Eye Tracking – Support for operations controllers (Press release). https://www.lhsystems.com/article/eye-tracking-support-operations-controllers

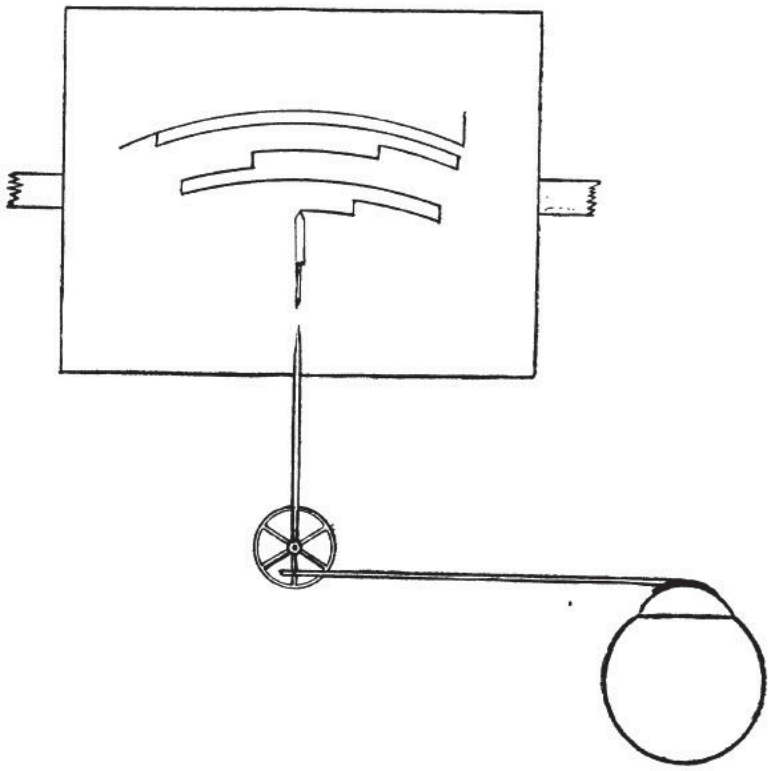


Figure 12.: The mechanical apparatus for eye movements recording.
Image from Huey (1898).

tached to the cup, transmitting the eye movements to the recording system (Figure 12). The eye was anesthetized using the cocaine (Huey, 1898).

3.2.2.1 *Scleral search coil*

Another technique of measuring eye movements is scleral search coil (Robinson, 1963). The subject is exposed to an alternating magnetic field; the eye position is calculated from the changes in the voltage of a coil of wire embedded in a scleral contact lens worn by the subject. This is an electrical version of the mechanical recording, a contact lens being put on the eyeball. While being precise, the method is highly intrusive and is surely outdated.

3.2.3 *Electrooculography*

The electro-oculographic (EOG) is the oldest among modern techniques to monitor ocular activity. Schott in 1922 and, later, Mowrer, Ruch and Miller in 1935 found that

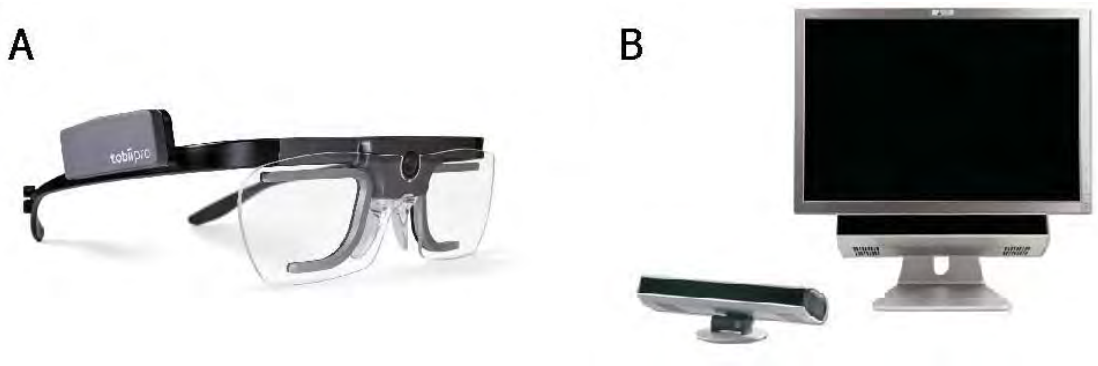
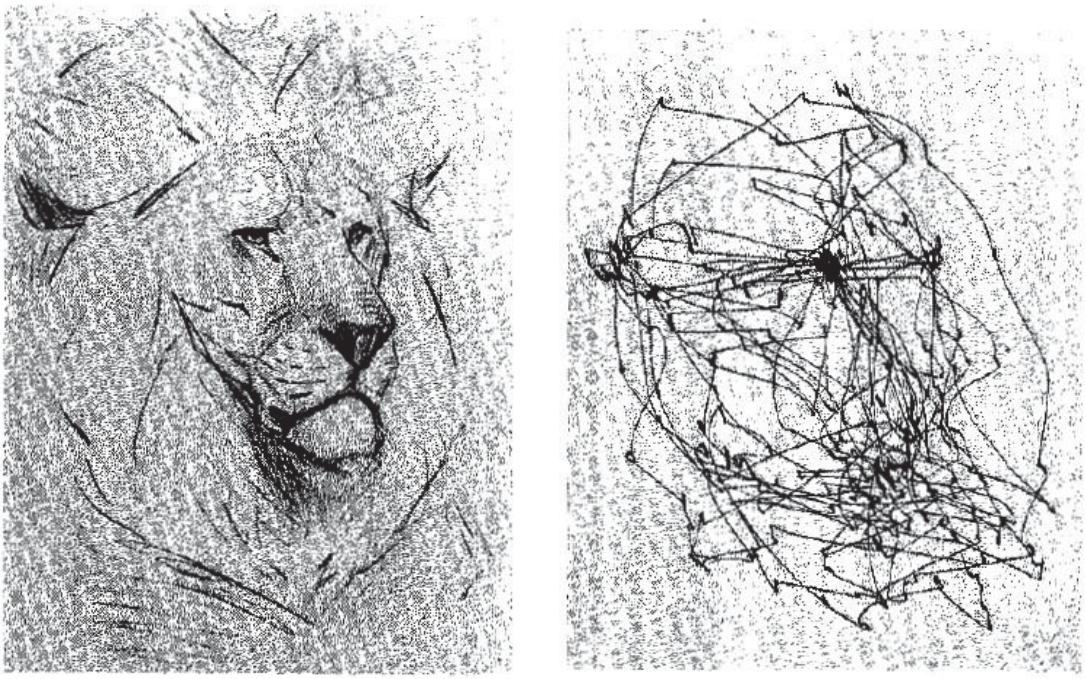


Figure 13.: Modern eye tracking systems. A) Head-mounted Tobii Glasses v2 and B) remote SMI RED500 on the right.

potential differences recorded at skin electrodes near the eye allow monitoring eye movements. The source of these differences is the corneoretinal potential — the cornea remains positive with respect to the retina — that depend on the angular movements of the eye. However, it appeared (North, 1965) that the corneoretinal potential depended also on illumination. Therefore, it is not recommended to use the EOG in the environment with changing light conditions. The EOG cannot determine the point of gaze in the environment. However, the signal can be used to detect different eye movements. Today, the EOG is most often used to correct the ocular artifact from electroencephalographic recording (Croft and Barry, 2000). Nevertheless, some researchers even proposed to construct a wearable EOG-glasses (Bulling et al., 2009) to detect eye movements in ecological environments.

3.2.4 *Video-based eye tracking*

Recording of eye movements using corneal reflection started its history as early as in the beginning of the XIX century (Dodge and Cline, 1901). However, the development of video recording techniques in the 70th allowed a huge progress in eye tracking. Nowadays, high-resolution video cameras record the eye movements. Interesting features (such as corneal reflection of an infra-red light source and the pupil center) are tracked in real-time (Duchowski, 2007). A calibration procedure allows determining the point of gaze in the environment. The modern eye tracking systems can be head-mounted (for example, Figure 13A) or remote (for example, Figure 13B). The video-based eye tracking using corneal/pupil reflection is the most used eye tracking technique in the world. It offers a non-intrusive monitoring of eye movements with the computation of the point of regard in real time with very high temporal resolution.



(a) Stimulus — a drawing of lion by V. A. Vatagin

(b) Record of eye movements during free examination for 2 minutes

Figure 14.: An example of eye movements recording. From (Yarbus, 1967).

3.3 THE EYE AND ITS MOVEMENTS

The notion of the world that we have in our minds is mostly due to our eyes. The incredible human visual system supplies us with information we rely on in everyday activity. The anatomy of the eye and the process of vision in it whole are out of the scope of this chapter (see (Purves et al., 2001; Snell and Lemp, 2013)). However, a few notions are necessary to understand why we move our eyes.

3.3.1 *Why do we move our eyes*

The light, passing through the pupil, reaches the retina to form an image. The retina contains two types of photoreceptors of different functions — rods and cones (Purves et al., 2001). The rod system is specialized for sensing light because it has low spatial resolution and high sensitivity to light. On the contrary, the cone system is specialized for acuity because it has high spatial resolution and is quite insensitive to light. These two types of photoreceptor cells are not uniformly distributed across the surface of the retina and it has important consequences for vision. The density of rods is important

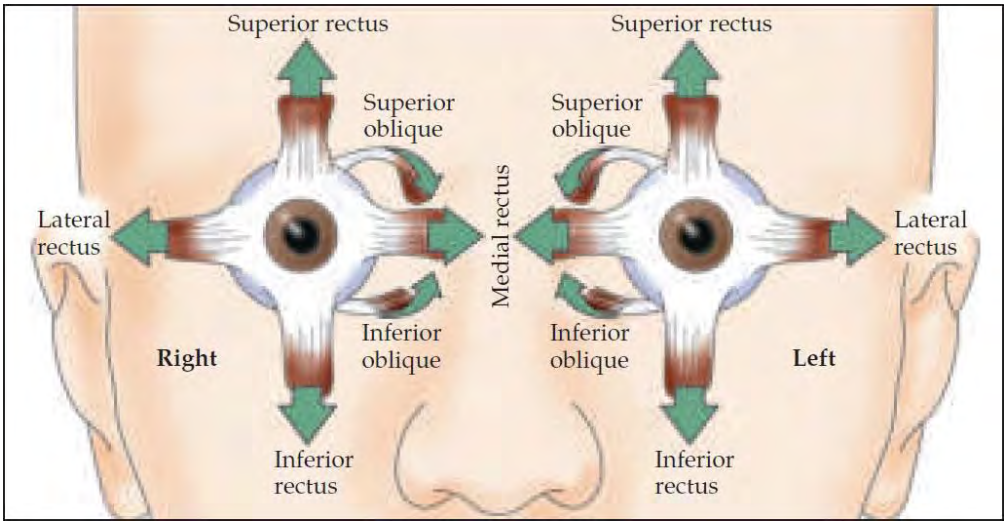


Figure 15.: The contributions of the six extraocular muscles to vertical and horizontal eye movements. Horizontal movements are mediated by the medial and lateral rectus muscles, while vertical movements are mediated by the superior and inferior rectus and the superior and inferior oblique muscle groups. From (Purves et al., 2001).

almost everywhere except the fovea — a small region of the central retina that measures only about 1.2 millimeters in diameter. On the contrary, cone density is highest in this region. In a central micro-part of the fovea, called foveola (that measures about 300 μm in diameter), there is no rod at all. Therefore, the fovea is the region of the highest acuity of the vision. We often have the impression that we process the entire visual field in a single fixation, but it is obviously not true. The distribution of rods and cones explains why humans move the eyes all the time: we direct our foveas on the objects that interest us.

3.3.2 *How we move our eyes*

Alfred Yarbus, a Russian physiologist, showed in the middle of the last century the importance of eye movements in vision and demonstrated that different patterns of eye movements are used depending on the nature of the task (Yarbus, 1967). Figure 14 shows an example from the Yarbus' work — a subject examined freely a drawing of a lion, while his eye movements were recorded. We distinguish saccades — ballistic rapid eye movements — and fixations — where the lines are dense and the gaze changes its direction afterward. From these lines of eye movements on the right, we can clearly imagine lion's head. It happens because we do not fixate random points in space just to acquire the whole image but rather interesting features of the image. Thus, we can see

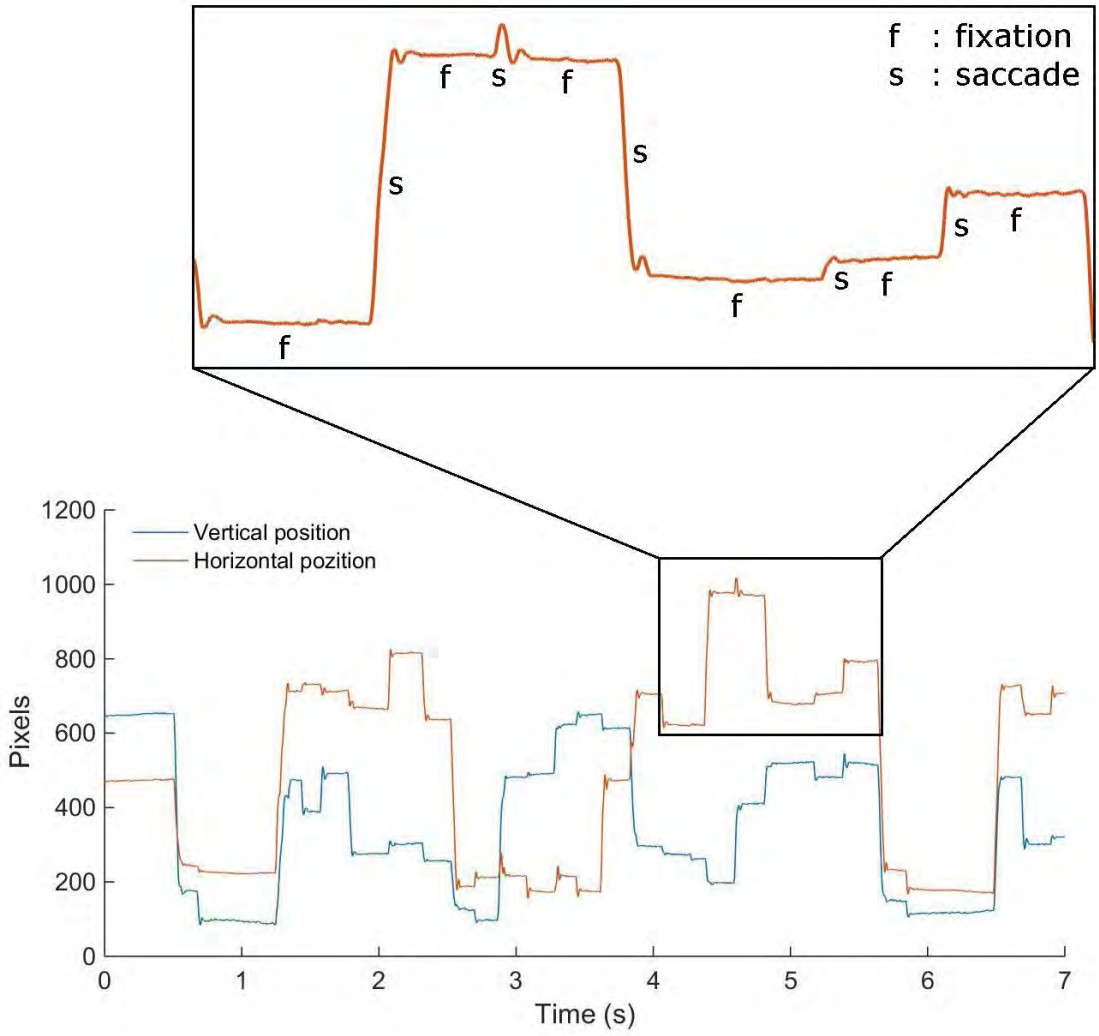


Figure 16.: An example of recording of eye movements at 1000 Hz.

that the subject fixated a lot lion's eyes, nose, mouse and year — important features of lion's face.

There exist six extraocular muscles to adjust the eye position. These muscles divided into three antagonistic pairs: the lateral and medial rectus muscles, the superior and inferior rectus muscles, and the superior and inferior oblique muscles (Figure 15). These muscles are responsible for horizontal, vertical and torsional movements of the eyes.

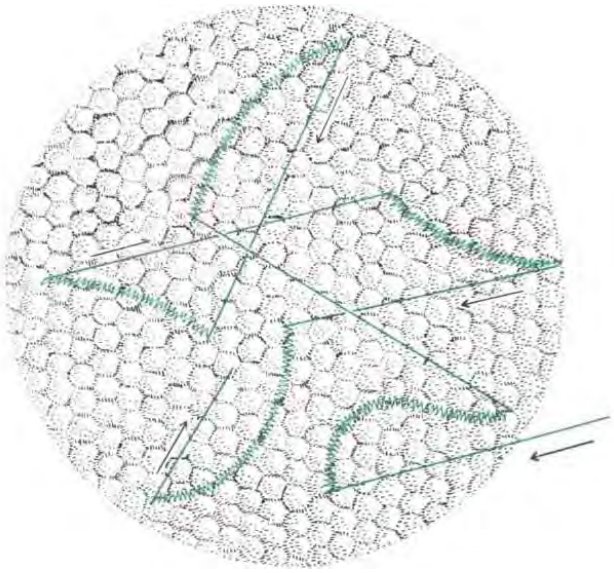


Figure 17.: Eye movements during a fixation: drift (curved lines) that carries the image away from the center of vision, microsaccades (straight lines) that bring it back toward the center and high-frequency tremor superimposed on the drift. The magnitude of all these movements is very small. The diameter of the patch of the fovea shown above is only .05 millimeter. From (Pritchard, 1961).

3.3.3 *Saccades and fixations*

There are four basic types of eye movements including saccades, smooth pursuit, vergence movements and vestibulo-ocular movements. However, in cognitive research, the last three are rarely used. The two basic eye movements used in the research are saccades and fixational movements.

Saccades are ballistic rapid movements that change the point of fixation. They differ in amplitudes and speed. Saccades are called ballistic because it cannot be corrected once started, and if the target moves during the saccade, a new saccade should be programmed and performed. Moreover, the vision is suppressed during a saccade (Matin, 1974).

Fixations are periods of time when the eye is relatively immobile and the visual information can be treated. Actually, the eyes are never motionless (Figure 17) and we need to perform constant eye movements not to be blind. If the image across the retina is stabilized, it will progressively fade away (Pritchard, 1961). You can experience it right now by holding one eyeball with a finger for few seconds — the image will disappear.

Studies showed that masking or making the fixated word disappear for 50-60 ms did not perturb the text reading (Liversedge et al., 2004; Rayner et al., 1981, 2003, 2006).

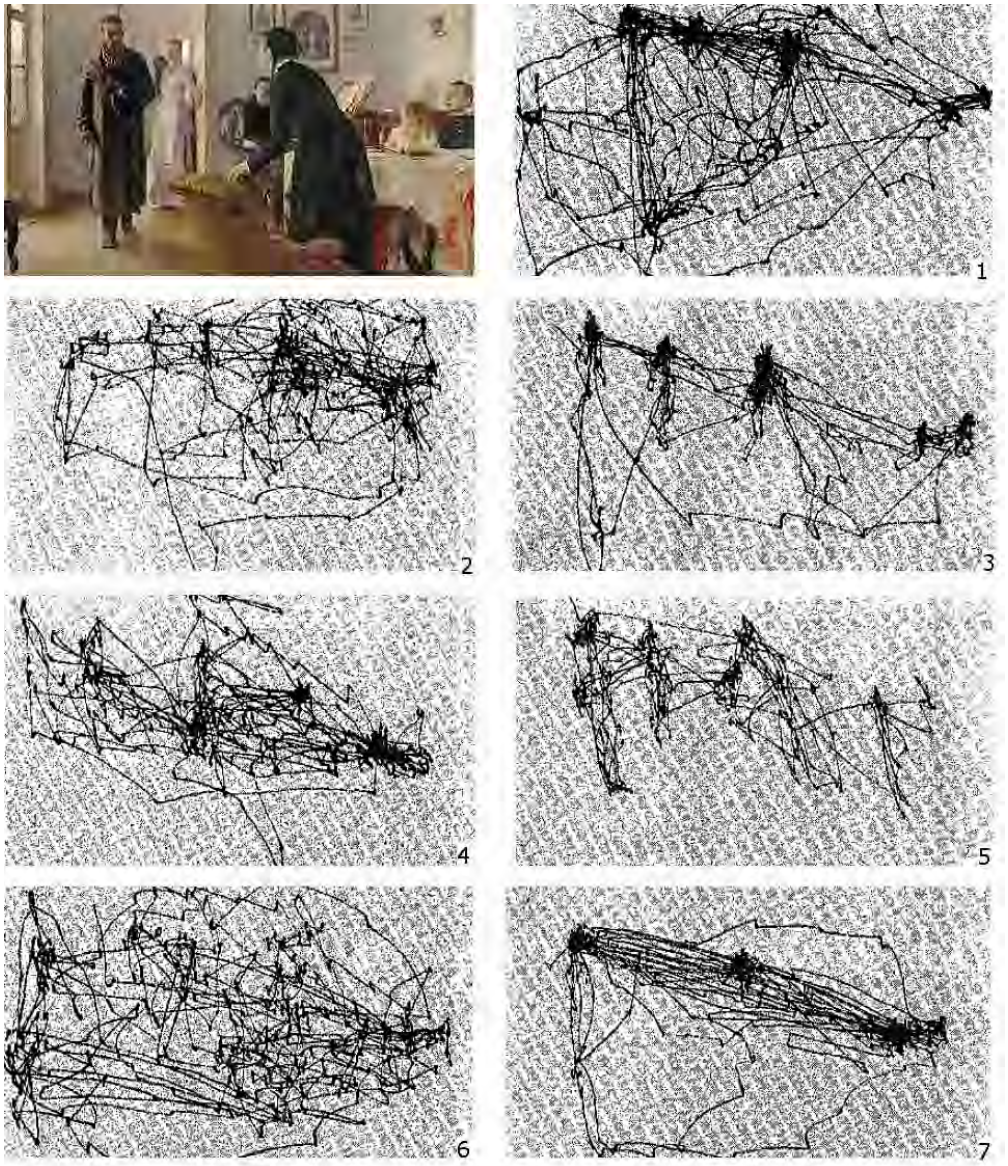


Figure 18.: Eye movements of 3 minutes recording by the same subject who was asked to: 1) examine freely the picture; 2) estimate the material circumstances of the family in the picture; 3) give the ages of the people; 4) surmise what the family had been doing before the arrival of the “unexpected visitor"; 5) remember the clothes worn by the people; 6) remember the position of the people and objects in the room; 7) estimate how long the “unexpected visitor" had been away from the family. From (Yarbus, 1967)

In complex scene perception, viewers need at least 150 ms to process the scene properly (Rayner, 2009). However, it has also been found that viewers can get a scene gist as fast as after 40 ms exposure (Castelhano and Henderson, 2008). Therefore, we can postulate that the fixation duration depends on the task, its complexity, the observer's knowledge of the environment, and the level of scene analysis to perform.

3.4 EYE MOVEMENTS REVEAL FLIGHT CREW'S ACTIVITY

3.4.1 *Eye movements reveal observer's task*

The fundamental work of Alfred Yarbus (1967) showed that "*depending on the task in which a person is engaged, [...] the distribution of the points of fixation will vary correspondingly*" (Figure 18). Although with some exceptions (Greene et al., 2012), the observations of Yarbus were supported by recent findings (Borji and Itti, 2014; Henderson et al., 2013; Iqbal and Bailey, 2004). However, without complex computations, as used in these papers, simple metrics are often used to characterize the ocular behavior. Among these are dwell time, duration and frequency of fixations, saccade length, transition matrices, microsaccades etc. An exhaustive review of the eye movements measures can be found in the "Eye tracking: A comprehensive guide to methods and measures" (Holmqvist et al., 2011) and "Eye tracking methodology: Theory and practice" (Duchowski, 2007). An excellent review of these eye movements measures applied in the aviation domain can be found in a scientific report by Glaholt (2014). We will not provide another review of eye movements metrics, the two cited works being rather complete. Instead, we will provide an example of investigation using eye tracking data and some real flight analysis.

3.4.2 *Gaze-based investigation of flight crew's activity*

Let us consider an example of an inquiry of a small portion of a flight based on the flight crew's gaze analysis. The presented data is a part of a study led by the BEA together with ISAE-SUPAERO. The questions set by the study are out of the scope of this section. Let us play instead the *Investigator* game: suppose that we have in our hands only a small piece of the gaze data (2 minutes and 25 seconds) recorded from the two flight crew members. We know only that the aircraft was landing at this point, nothing more. The gaze data contains angular horizontal and vertical positions of the gaze point, and also an integrated onboard device, knowing the position of the flight instruments, recorded the number of the current Area Of Interest (AOI). The Figure 19 gives an example of what pilots looked at during the flight, but let us do not digress from the *Investigator* game. Let us explore the data and discover who was doing what and what happened during this landing.

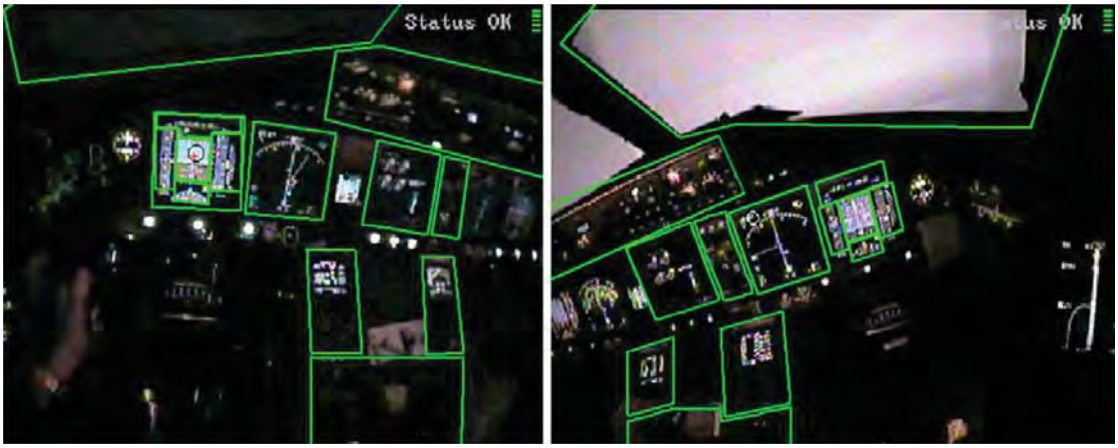


Figure 19.: An example of the inside view of the flight simulator. Some areas of interest are highlighted in green.

3.4.3 *Raw gaze coordinates*

First, we plot and analyze the angular positions of the crew's gaze (see Figure 20). A few things pop out immediately. The most evident of them is that the ocular behavior drastically changes somewhere after 75 seconds — the both pilots stopped doing large vertical movements. Also, there is a clear difference between the two pilots. Whereas the ocular activity of the pilot B seems more or less regular, that of the pilot A seems more intense, he does more transitions, especially in the seconds part of the flight portion. The third observation is that the space coordinates apparently correspond to some precise zones — both pilots recurrently return to some X-Y coordinates in space. Thus, the pilot B often consults a zone (-10,30) during the first part of the period, and then a zone around (-20, 0) during the second part. These coordinates correspond indeed to particular zones within the cockpit and were computed by the "onboard device".

3.4.4 *Areas of interest*

When the gaze of a pilot is inside one of these zones — AOIs — we can detect it thanks to the calibration procedure. Then, we can study not simply the raw coordinates, but something that is much more understandable, i.e. the history of when each AOI was looked at. The figure 21 shows the distribution of the gaze of the two flight crew members between different AOIs. We can now see that for the pilot B the zone (-10,30) corresponds to the outside (out of the window) and (-20,0) to the attitude indicator. We also discover that the pilot A looked a lot outside, on the speed and the attitude indicators. He entered some information on the Flight Control Unit (FCU) at around 100 seconds. Moreover, he controlled the flaps a lot compared to the pilot B. The second pilot also

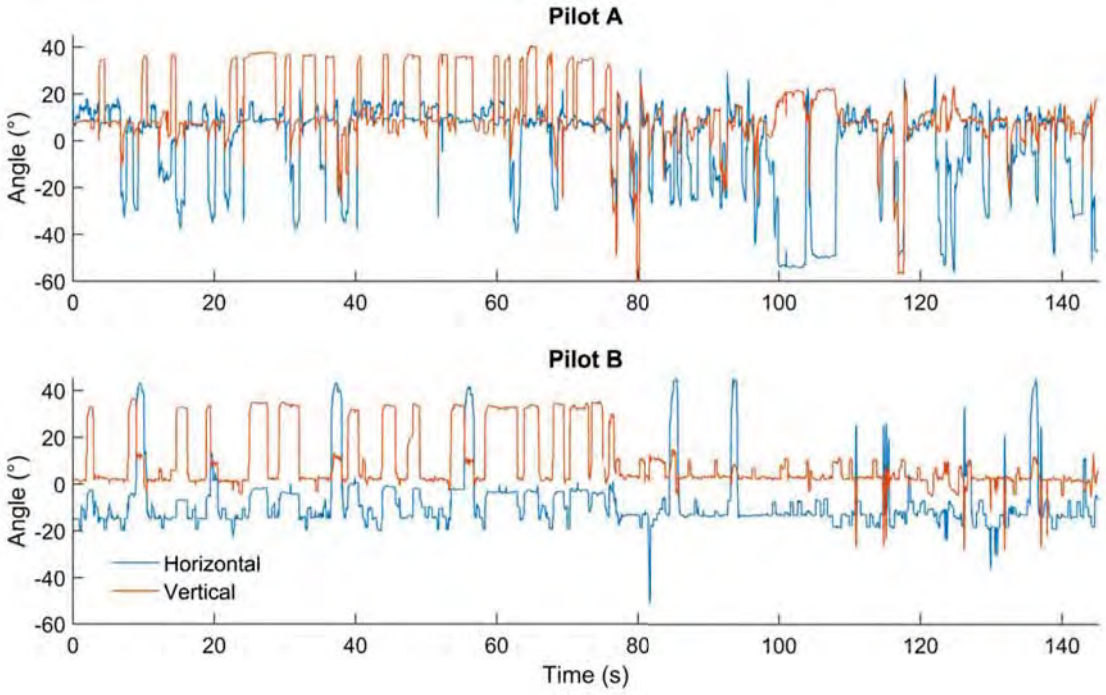


Figure 20.: Horizontal and vertical angular positions of the two flight crew members.

looked a lot outside but checked the Navigational Display (ND) less often compared to the pilot A. Most of his time, he passed on the attitude indicator. Based on all said above, we can deduce that the pilot A took the role of Pilot Monitoring, whereas the pilot B was Pilot Flying (see Chapter 1). We can also note that the behavior of the two pilots changed (see the previous subsection). The flight crew stopped gazing outside — that is the most obvious change of their ocular behavior. However, what exactly happened?

3.4.5 *Unexpected event*

Based on the data displayed in Figure 21 we can compute the dwell percentages on different flight instruments. Figure 22 depicts such distributions for both pilots and the two parts of the flight — before and after an unexpected event. Let us first analyze the gaze of the Pilot Monitoring — the pilot A. During the second part, he

- stopped looking outside,
- looked at FCU,
- looked more at ND,
- looked more at the flaps.

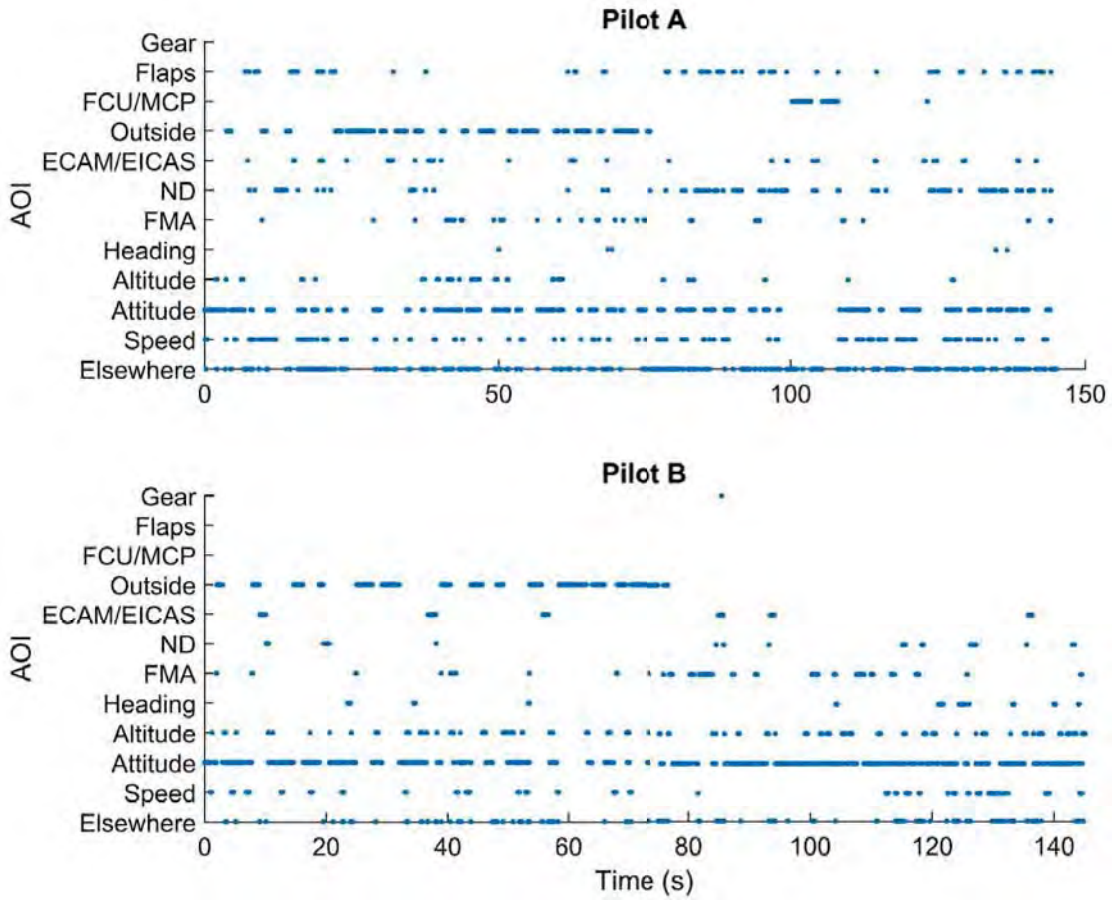


Figure 21.: Gaze position according to different AOIs for the two flight crew members.

In turn, the Pilot Flying — the pilot B — during the second part,

- also stopped looking outside,
- checked the FMA
- spend more time controlling the attitude

Overall, the flight crew's ocular behavior shows that the pilots, most probably, initiated a go-around procedure during the recorded approach. The crew stopped looking outside because it became irrelevant. The pilot monitoring set the flaps and the autopilot configuration on the FCU, while the pilot flying was supervising the horizontal and vertical trajectory of the aircraft, also verifying that the appropriate auto-thrust mode was selected (displayed on the FMA).

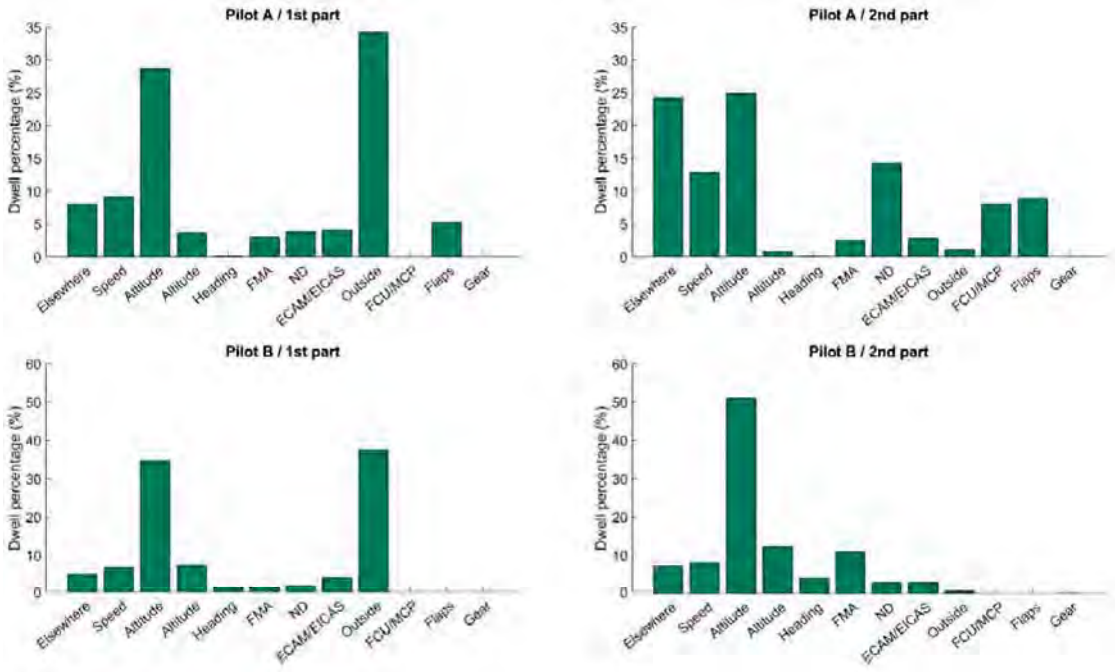


Figure 22.: Dwell percentages for the two parts of the flight for the two flight crew members.

3.4.6 *Investigator's conclusion*

To summarize this section, we demonstrated that the analysis of the pilots' ocular behavior is informative enough. It adds another dimension to the investigation procedure[4] — not only we can analyze the crew's voices and actions such as button presses and thrust levers displacement but also their psychomotor behavior, i.e. eye movements. Based exclusively on the gaze data, we deduced the respective role of each pilot in task sharing, understood what they were doing at each moment and what happened during the recording. Therefore, eye tracking is a powerful tool for accidents investigations.

3.5 CONCLUSION

In this Chapter, we revised that the eyes are in continuous movement to update the visual information on the retina and to place the fovea on the point of interest. We also reviewed the existed and existing methods to track these movements. Eventually, we showed how basic observations on eye movements can be useful to investigate flight crew's activity.

4 See also Chapter 1 for an illustration of an investigation procedure.

Part II

CONTRIBUTIONS

4

THE IMPACT OF LUMINANCE ON PUPILLARY RESPONSES TO TRANSCIENT COGNITIVE LOAD

Засвети же свечу
на краю темноты.
Я увидеть хочу
то, что чувствуешь ты.

Иосиф Бродский

Then light a candle
on the edge of darkness.
I want to see
what you feel.

JOSEPH BRODSKY
my loose word-by-word translation

4.1 INTRODUCTION

This chapter was published in Peysakhovich et al. (Peysakhovich et al., 2015a). In this paper we did not aim to separate explicitely tonic pupil diameter and phasic pupil response and to compare transcient versus sustained cognitive loads. Therefore, we keep here the terms used in the paper, i.e. Task-Evoked Pupillary Response — to designate the phasic pupil response, and “baseline diameter" — to designate tonic pupil diameter. We keep these terms throughout the chapter to keep the integrity of the paper, and also because these are the terms often used in the literature. Note, however, that the equality “task-evoked pupillary response = phasic response" is not 100% correct. Indeed, the changes in tonic pupil diameter can also be task-evoked or task-induced as in the next chapter. However, because most of protocols are using tasks with transcient load, task-evoked pupillary response is almost always equals phasic pupil response. For the global conclusion on both experiments see the following chapter.

Pupil diameter, despite its convenience for non-intrusive operator state monitoring in complex environments, is still not available for in situ measurements because of numerous methodological limitations. The most important of these limitations is the influence of pupillary light reflex. Hence, there is the need of providing a pupil-based cognitive load measure that is independent of light conditions.

The objective of the present study was to explore the interaction between luminance and cognition on pupillary dilation to deduce a pupil-based measure of load on memory that would be independent of luminance conditions. To that end, we manipulated computer screen luminance and load on memory. We investigated the influence of the luminance conditions on the task-evoked pupillary response and on the power spectrum of the pupillary signal under different levels of load on memory. To the authors' knowledge, no studies have performed an analysis of mental effort via the power spectrum of the pupillary signal under different light conditions. In this study, we assessed the following questions: a) How do luminance conditions affect the pupillary PSD? b) Is there an interaction effect between luminance and load on memory conditions on the TEPRs and power spectrum of the pupillary signal? c) If so, can we extract a frequency-based feature of the pupillary signal that would be solely dependent on the load on memory?

In this chapter, we present a promising technique of pupillary signal analysis resulting in luminance-independent measure of mental effort that could be used in real-time without a priori on luminous conditions. Twenty-two participants performed a short-term memory task under different screen luminance conditions. Our results showed that the amplitude of pupillary dilation due to load on memory was luminance-dependent with higher amplitude corresponding to lower-luminance condition. Furthermore, our experimentation showed that load on memory and luminance factors express themselves differently according to frequency. Therefore, as our statistical analysis revealed, the ratio between low (0–1.6 Hz) and high frequency (1.6–4 Hz) bands (LF/HF ratio) of power spectral densities of pupillary signal is sensitive to the cognitive load but not to luminance. Our results are promising for the measurement of load on memory in ecological settings.

4.2 MATERIALS AND METHODS

4.2.1 *Subjects*

22 healthy volunteers (4 females, 3 left-handed, age 24.5±2.8, education 15.1±1.0), students of ISAE (French Aerospace Engineering School), all native French speakers, participated in the experiment after they gave their informed written consent. All reported normal auditory acuity and normal or corrected-to-normal vision.

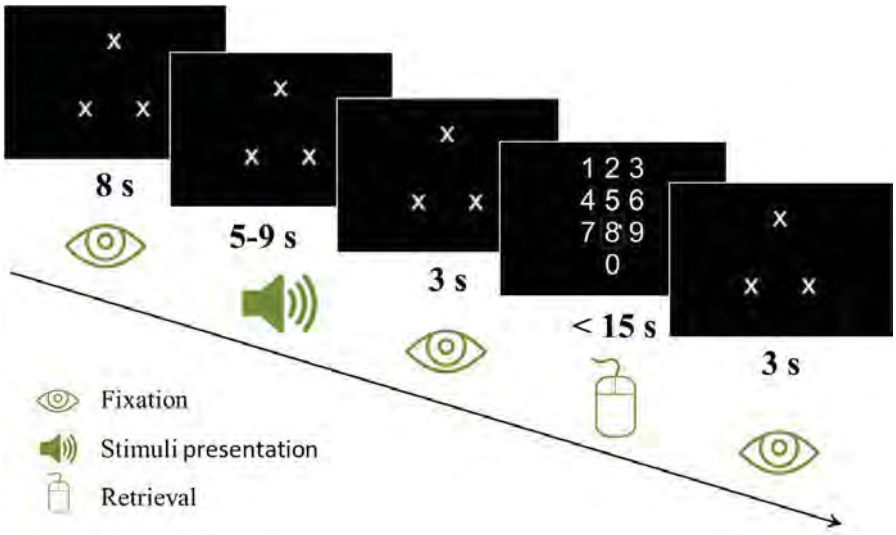


Figure 23.: Time course of a task trial

4.2.2 *Experimental design*

Participants performed a short-term memory task where they were asked to recall a paced sequence of digits. Stimuli were presented in the auditory modality via two stereo speakers, positioned at each side of a computer monitor. Mean sound level of stimuli was measured at 75.3±0.9 dB. Three levels of difficulty were produced with stimuli of 5, 7 or 9 digits, generated pseudo randomly (never two same successive digits) while the screen luminance was changing from trial to trial. The screen background was black (3.4 cd/m^2), gray (24 cd/m^2) or white (54.8 cd/m^2). In half of cases the participants were asked to retain the series of digits and to report it back (load on memory condition), while in others they were only asked to listen passively (control condition). The time course of an example trial with black background and load on memory conditions is represented on Figure 23.

Except for the retrieval phase (when a numeric keypad was showed) three fixation crosses were forming a fixation triangle in the center of the computer screen. The subject was informed about the condition (whether or not to retain the series) by means of the direction of the fixation triangle during the whole trial. If the triangle pointed at the top (as on Figure 23), subject was informed to recall the series. If the triangle pointed at the bottom (inverted compared to that on Figure 23: two fixation crosses at the top and one at the bottom), subject was informed to listen to the stimulus passively (i.e. not to recall it). Furthermore, for control trials the on-screen numeric keypad was replaced by sharp signs placed in the same configuration as 10 pad digits. Prior to the start of the main experiment, participants were visually provided with instructions and performed a practice block of 18 trials to familiarize with each condition. Then

participants completed two blocks of 45 trials each (about 20 minutes long) with a 5 minutes pause in-between. The experiment lasted for a total of 45 minutes and contained 5 trials per each condition.

4.2.3 *Stimuli and apparatus*

A pseudo-random list of stimuli (used for all participants) was generated using the following four rules: 1) the screen background changed each trial, 2) no more than two successive control trials were administered, 3) no more than two successive series of the same size (5, 7 or 9 digits) were presented, 4) each block began with at least two task trials. The first rule helped to inform participants about the start of the next trial. The second rule averted the excessive distraction (that one could expect during three or more successive control trials that would last approximately for one minute and a half) and prevented the rhythm that one could expect in the case of alternation of control and load on memory conditions on every trial. The third rule made it impossible to think out a strategy of series retention because of the randomness of series size (i.e. the participants were never aware of the number of digits they would have to memorize until the end of the sequence). Finally, the fourth rule helped the participants to get into gear in the beginning of each block.

Each trial started with an 8-second pre-stimulus accommodation period, during which participants' eyes were adapting to the new screen luminance condition. Then a sequence of digits was presented at a rate of one digit per second. After a 3-second retention pause, participants reported the sequence back (for the load on memory condition only) through the on-screen numeric keypad using the mouse. For the control condition they were asked to click on sharp signs randomly. Participants had 12 seconds to enter the series but were not informed about the exact maximum. Once the response was given, each trial ended with an additional 3-second pause. No additional stimulus onset asynchrony was used as each trial length varied randomly as a function of response duration.

The experiment was conducted in a darkened sound-proof chamber. Ambient illuminance was measured at 10 lux. Participants were seated at a distance of approximately 65 cm from the 22-inch monitor (1680×1250). During the whole experiment, participants' gaze position and pupil diameter were recorded with a remote SMI RED eye-tracker (SensoMotoric Instruments GmbH, Germany) at sampling rate of 120 Hz. This device tracks the pupil diameter with precision despite the head is not fixed with a chin-rest. Before each block the 5-points calibration of the eye-tracker was validated with four additional fixation points.

The stimuli presentation and data acquisition routines were implemented in Python programming language using PsychoPy software (Peirce, 2007). The data acquisition

routine used iViewX SDK to communicate with the eye-tracker. The data analysis was performed in Matlab (Mathworks) and Statistica (StatSoft) software.

4.2.4 *Pupil signal pre-processing*

As suggested by Siegle (2011) the raw pupillary data were smoothed with a simple procedure described by Glaser and Ruchkin (1976), namely with a "two pass" 9-point filter (low-pass with cutoff frequency of 5.9 Hz). The attenuation in the band up to 4 Hz of the applied filter was less than 0.1 dB. Identified blinks were then replaced using linear interpolation. In order to remove the eyelid-closure effect as well as the consequences of filtering in the neighborhood of signal discontinuities caused by blinks, 12 adjoined points (100 ms) from each side were replaced as well. The pupillary signal was then down-sampled to 50 Hz. This was done not only to reduce the data size but also to be able to average point-by-point stimuli-locked pupillary responses. Indeed, small irregular inter-sampling lags can interfere with the averaging procedure. The down-sampling technique generates a regular time grid and allows therefore computing correctly the TEPRs.

4.2.5 *Data processing for statistical analyses*

For the time domain analysis, baseline pupil diameter was calculated as the median value of the 500 ms preceding the stimulus onset. Trials were segregated according to experimental conditions and averaged point-by-point giving a TEPR per condition per participant. For the statistical analysis, the mean value of pupil diameter during the retention pause (3 seconds following the stimuli) was used.

The frequency analysis was performed on 8-second periods consisting of the 5 last seconds of stimuli presentation and 3 seconds of the pause before retrieval. The linear trend was removed using detrend MATLAB function that removes the best straight-line fit from the signal. Then, each period of 400 samples (8 sec $\times$ 50 Hz) was zero-padded to 512 samples so that the frequency resolution was less than 0.1 Hz. The PSD was estimated using the Welch's method (pwelch function from Matlab Signal Processing Toolbox) with segments of length 50 with 50% overlap. Each segment was windowed with a Hamming window. Finally, the power was converted to dB. Most of the pupillary activity appears within the frequency band up to 4 Hz according to Nakayama and Shimizu (2004). Thereby, the TEPRs frequency components were analyzed within the frequency band from 0 up to 4 Hz.

4.2.6 *Statistical analyses*

A trial was rejected if the overall time spent to blink exceeded 50% of the period of interest (i.e. exceeded 4 seconds) and if the longest blink exceeded 1 second in length. After data rejection, a 2×3×3 three-way ANOVA — load (load on memory vs. control) × luminance (black vs. gray vs. white) × size (5 digits vs. 7 digits vs. 9 digits) — was performed to examine the number of valid trials for each condition. The analysis revealed a significant data loss depending on luminous condition, $F(2,42) = 17.7, p < .001$, partial $\eta^2 = 0.46$. Tukey's HSD (Honestly Significant Difference) post-hoc comparison showed that the average number of valid trials was significantly lower ($p < .001$) for the black luminous condition (1.9 ± 0.3 out of 5) compared with gray (3.5 ± 0.2 trials) and white (3.5 ± 0.2 trials) background conditions. During the darkest condition participants probably gave their eyes a rest after an exposure to a bright screen in darkened room that implied higher blink rate (Benedetto et al., 2014). Therefore, this condition was not used for the further analysis because of insufficient number of average valid trials to have a valid mean value. Furthermore, as stated by Nakayama (2006), loss of pupil signal caused by blinks has an impact on frequency analysis adding extra power in the spectrum. More importantly, as suggested by Siegle (2011), at least 5-10 trials have to be averaged in order to obtain reliable results. However, we retained at most 3.9 ± 0.3 (out of 5) valid trials per condition. In order to further increase the number of observations and to focus the analysis of the interaction between load and luminous conditions, we aggregated the three different levels of difficulties for all the subsequent analysis. Eventually, we had at least 10.2 ± 0.9 valid trials (out of 15) within each condition.

Statistical two-sided t-test was performed on the number of correctly recalled sequences for load on memory condition to investigate the influence of the luminance condition on performance.

Four 2×2 two-way repeated measures ANOVAs with within subject factors load (load on memory vs. control) and luminance (gray background vs. white background) were carried out on the absolute pupil diameters (during baseline and retention pause), TEPRs and total PSD. Tukey's HSD was used for post hoc comparisons.

To compare the effect of two levels within each of two factors (load and luminance) across different frequencies, we computed Cohen's D value for each frequency point. As consistent with Cohen (*A medium effect size is conceived as one large enough to be visible to the naked eye.*) we chose the threshold value of 0.4 (Cohen, 1988) to select the data points significantly impacted by the load on memory effect across the frequencies. Low Frequency band and High Frequency band were identified accordingly. Three 2×2 two-way repeated measures ANOVAs with within subject factors load (load on memory vs. control) and luminance (gray background vs. white background) were carried out

on the mean PSD within LF and HF bands, as well as on LF/HF ratio. Tukey's HSD was used for post-hoc comparisons.

4.3 RESULTS

4.3.1 *Behavioral results*

Statistical two-sided t-test comparison showed that screen luminance had no influence on participants' performance, $t(22) = 0.34, p = 0.74$.

4.3.2 *Absolute pupil diameter*

During 500 ms pre-stimulus baseline period: The two-way ANOVA (load $\times$ luminance) showed no effect of task, $F(1,21) < 1, p = 0.78$, but a strong main effect of luminance (smaller pupil for brighter screen), $F(1,21) = 128.0, p < .001$, partial $\eta^2 = 0.86$, on absolute pupil diameter during the baseline period. No interaction was found, $F(1,21) < 1, p = 0.33$.

During retention pause: Two-way ANOVA (load $\times$ luminance) revealed significant main effects of both load, $F(1,21) = 78.8, p < .001$, partial $\eta^2 = 0.79$, and luminance factors, $F(1,21) = 180.5, p < .001$, partial $\eta^2 = 0.90$, on absolute pupil diameter. Pupil was greater for the load on memory condition compared to the control condition. And it was greater for the gray background compared to the white background. The interaction was also significant, $F(1,21) = 24.35, p < .001$, partial $\eta^2 = 0.54$. As shown by the Tukey's HSD test, pupil was significantly smaller within the load on memory condition with white background ($p < .001$) compared to the control condition with gray background. In addition, the post-hoc test showed that load on memory vs. control condition implied significantly greater ($p < .001$) pupil diameter for both luminance conditions.

4.3.3 *TEPRs in time domain*

The two-way ANOVA (load $\times$ luminance) revealed a significant main effect of load on pupil diameter that was larger during the load on memory condition compared to the control condition, $F(1,21) = 69.5, p < .001$, partial $\eta^2 = 0.77$. In addition, we found a main effect of luminance, $F(1,21) = 29.9, p < .001$, partial $\eta^2 = 0.59$, corresponding to a smaller dilation amplitude within the white background condition (see Figure 24)). The interaction of the two factors was also significant, $F(1,21) = 5.96, p < .05$, partial $\eta^2 = 0.22$. The Tukey's HSD post-hoc tests showed that the load effect (load on memory vs. control) was significant for both light conditions ($p < .001$) whereas

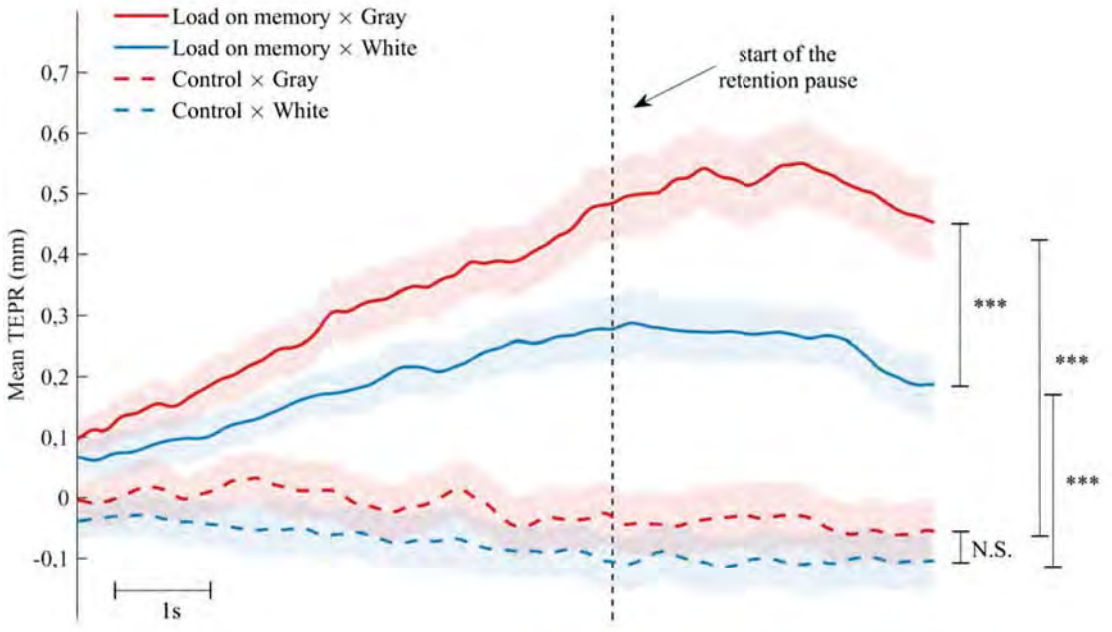


Figure 24.: Grand-average response-aligned TEPR (with shaped error-type) per condition. Timeline corresponds to the last 4 seconds of stimulus presentation and 3 seconds of retention pause. The vertical dashed line depicts the start of the retention pause.

the influence of luminance was significant only during the load on memory condition ($p_{load} < .001, p_{control} = 0.42$; Figure 24).

4.3.4 *TEPRs in frequency domain*

The two-way ANOVA (load × luminance) for the total spectrum (0–4 Hz) showed that the PSD was significantly higher under load on memory condition than under control condition, $F(1,21) = 9.6, p < .01, \eta^2 = 0.31$. The gray background condition yielded significantly higher PSD compared to the white background, $F(1,21) = 38.5, p < .001, \eta^2 = 0.65$. No significant interaction was found, $F(1,21) = 0.01, p = 0.91$.

Figure 25 shows the mean spectrum values across participants for the main effect of both load on memory (25.a) and luminance conditions (25.b). While the effect of the luminance factor was high for the total spectrum (25.d), the effect of the load factor faded away as a function of frequency (25.c). After the comparison of the effect of two levels within each of two factors (load and luminance) across different frequencies using Cohen's D values, we determined a frequency threshold of 1.6 Hz (Low Frequency band

= 0–1.6 Hz; High Frequency band = 1.6–4 Hz). We averaged PSD in these two bands in order to assess load on memory (Figure 25.c) and luminance effects (Figure 25.d) with classical analysis of variance.

In the Low Frequency band (0–1.6 Hz) the mean PSD was significantly higher under load on memory vs. control condition, $F(1,21) = 9.41, p < .01, \eta^2 = 0.31$. In addition, the white background caused significantly smaller mean PSD than the gray background, $F(1,21) = 14.9, p < .001, \eta^2 = 0.42$. No interaction was found, $F(1,21) = 0.3, p = 0.59$. Regarding the High Frequency band (1.6–4 Hz), no significant effect of load on memory on the mean PSD was found, $F(1,21) = 4.1, p = .06$, while luminance had a significant effect, $F(1,21) = 32.4, p < .001, \eta^2 = 0.61$. No interaction was found, $F(1,21) = 0.2, p = 0.64$. In summary, the effect of load was significant only for the low frequency band whereas the effect of luminance was significant for both frequency bands.

We then computed LF/HF ratio of mean PSD. The results of the ANOVA showed a significantly lower ratio under load on memory vs. control condition, $F(1,21) = 6.83, p < .05, \eta^2 = 0.25$, while no effect of the luminance was found, $F(1,21) = 1.67, p = .21$. No interaction was found, $F(1,21) = 0.5, p = 0.47$. Table 2 summarizes the mean and standard deviation values for the main results.

4.4 DISCUSSION

We investigated the effect of load on memory under different luminance conditions on the pupillary task-evoked response and its different components of power spectrum. Our main results showed that the TEPR amplitude was smaller under bright vs. dark condition and that the load on memory impacted specifically the low frequency component of pupillary spectrum, whereas the luminance affected both low and high frequency bands. These findings allowed identifying a frequency-based metric of pupillary activity — the LF/HF ratio — that is sensitive to load on memory but not to luminance.

4.4.1 *Effects of load on memory and luminance factors on the absolute pupil diameter*

There was no effect of load on memory on the baseline (500 ms pre-stimulus) pupil diameter. This indicates the absence of the anticipation effect, i.e. no mental preparation effect to load on memory condition vs. control condition. Pupil baseline diameter was solely impacted by luminance. Thus, the effects of load on memory on TEPR are not due to any confounding factors on the baseline.

The absolute pupil diameter was significantly higher under load on memory vs. control condition, which is coherent with Beatty and Lucero-Wagoner (2000). Pupil diameter was also significantly smaller under white background condition. However, the pupil size for the load on memory condition with white background was smaller than for

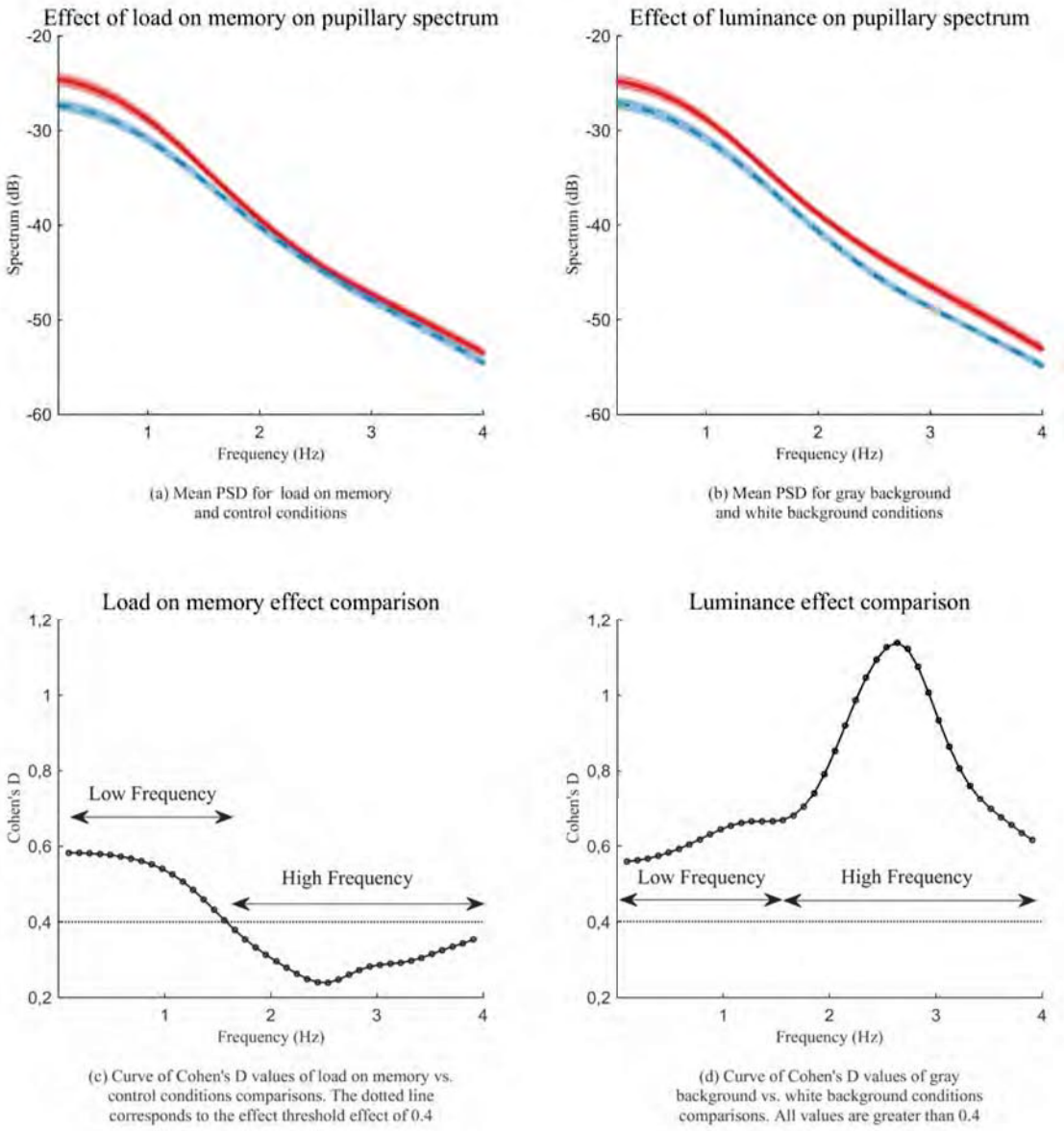


Figure 25.: Curves of mean values (with shaped error-type) for (a) load on memory vs. control conditions and (b) gray background vs. white background conditions; and (c, d) corresponding curves of Cohen's D values of point-by-point comparisons.

the control condition under gray background condition. These results confirm the findings of Xu et al. (2011); the light reaction is predominant over the cognitive pupillary

a)

Dependent variable	**Load on memory factor**	
	Control	Load on memory
Absolute diameter during baseline period (mm)	4.04 (0.17)	4.02 (0.16)
Absolute diameter during retention pause (mm) ***	3.96 (0.16)	4.36 (0.18)
TEPR amplitude (mm) ***	-0.07 (0.02)	0.33 (0.04)
Total spectrum (dB) **	-39.80 (0.59)	-38.48 (0.67)
Mean PSD in LF (dB) **	-30.46 (0.67)	-28.27 (0.78)
Mean PSD in HF (dB)	-46.02 (0.60)	-45.29 (0.67)
LF/HF ratio of mean PSD *	0.66 (0.01)	0.62 (0.01)

b)

Dependent variable	**Luminance factor**	
	Gray	White
Absolute diameter during baseline period (mm) ***	4.42 (0.19)	3.64 (0.14)
Absolute diameter during retention pause (mm) ***	4.62 (0.20)	3.71 (0.14)
TEPR amplitude (mm) ***	0.20 (0.03)	0.06 (0.02)
Total spectrum (dB) ***	-38.10 (0.64)	-40.18 (0.59)
Mean PSD in LF (dB) ***	-28.30 (0.70)	-30.43 (0.68)
Mean PSD in HF (dB) ***	-44.63 (0.70)	-46.68 (0.57)
LF/HF ratio of mean PSD	0.63 (0.01)	0.65 (0.01)

Table 2.: a) Effect of load on memory condition on the dependent variables. b) Effect of background luminance condition on the dependent variables. Values are mean $\pm$ (SD), (n = 22 for all variables). $* = p < .05, ** = p < .01, *** = p < .001$.

component. This major light influence motivates the use of relative pupillary dilations (TEPRs).

4.4.2 *Effects of load on memory and luminance factors on the TEPR amplitude*

The TEPR amplitude under load on memory condition was not the same according to luminance condition; the same amount of load on memory induced higher relative pupillary dilation within darker background condition. Thus, to accurately measure the mental workload based on the pupil diameter, it might be important to take the point of fixation and its luminance into account.

This finding is coherent with previous findings. Namely, pupil diameter was found to have a close relationship with the firing rate of the LC, a nucleus in the brainstem that

is involved in the neural circuitry regulating arousal and autonomic function (Gilzenrat et al., 2010; Rajkowski et al., 1993; Samuels and Szabadi, 2008). Increased LC activity correlates with higher sympathetic tone (Elam et al., 1986; Gilzenrat et al., 2010) and leads to parasympathetic inhibition of Edinger-Westphal complex (Samuels and Szabadi, 2008), that both result in pupillary dilation. Decreased parasympathetic tone under dim light (Steinhauer et al., 2004) should therefore result in higher sympathetic influence on the peak pupil diameter (i.e. larger aperture). It should be noted that the sympathetic component of pupillary dilation has greater latency; the inhibition of parasympathetic pathways results in an earlier peak, whereas sympathetic activity is responsible for pupillary peak dilation after about 1200 ms of stimulus presentation (Steinhauer and Hakerem, 1992). Consequently, when studying sustained processing measuring mean pupil diameter for a long period of time (e.g. (Steinhauer et al., 2004)), the average pupil size is slightly greater under bright light conditions (contribution of early parasympathetic dilations). Conversely, while focusing on mean peak dilations as in our study, the brighter light diminished the TEPR's peak amplitude as a consequence of increased parasympathetic tone.

Pomplun and Sunkara (2003) proposed to perform a calibration procedure in order to determine a pupil baseline diameter as a function of display brightness for accurate cognitive load measurement in ecological situations. During the experiment, the authors suggested to subtract the calibration value from the pupillary signal according the current display brightness. This extends the standard baseline subtraction procedure to the cases when the display brightness changes during stimuli presentation, as in an ecological flight or drive stimulator. Furthermore, our results encourage the extension of this calibration procedure by specifying the relationship between the amplitude of pupillary dilation for a given amount of cognitive load and current display luminance. Interestingly, as no such model presently exists, an extra "cognitive" calibration could be performed simultaneously by asking participants to perform a mental calculation by varying display luminance and measuring the corresponding TEPR amplitude.

4.4.3 *Effects of load on memory and luminance factors on the spectrum of the pupillary signal*

The analysis of the total considered spectrum (0–4 Hz) showed that the PSD of the pupillary signal was higher under load on memory compared to the control condition, and under the gray background condition compared to the white background condition. While the absolute pupil diameter revealed the overall effort due to some cognitive process, the power of spectral density of pupillary signal characterizes the activity of the nervous system during this process. In the context of our experiment, the higher pupillary activity indicated higher overall effort (cf. results on absolute pupil diameter) because of the cumulative nature of this short-memory task. This result echoes the findings of

Nakayama and Shinizu (2004) who observed an increase in pupillary power spectrum during mental computations.

Besides, the statistical analysis showed that load on memory and luminance factors did not have the same effect on the pupillary signal spectrum. While under brighter condition the PSD was higher throughout the whole considered frequency band (0–4 Hz), the effect of the load on memory factor was only present in the low frequency band (up to 1.6 Hz). Indeed, the adaptation of the pupil to the luminous level (light/dark reflexes) and the cognitive component of the pupillary reaction have different neural pathways.

4.4.4 *Effects of load on memory and luminance factors on the LF/HF ratio*

The analysis of the effect of load on memory and luminance on the spectrum of pupillary signal allowed us to dissociate these two effects. The ratio of PSD of the pupillary signal within Low and High frequency bands was significantly lower under load on memory condition compared to the control condition; but it did not differ in the luminance conditions. This result is in line with previous studies on the influence of mental workload on heart rate variability showing a greater increase of LF activity compared to HF activity (Durantin et al., 2014; Mizuno et al., 2011; Mukherjee et al., 2011). Nevertheless, note the di fference between the frequency bands used in this study compared to the standardized bands used in heart rate variability studies. As for cardiac activity analysis (Billman, 2013), the physiological basis of these changes in the low-high frequency ratio is difficult to discern.

Compared with the index of cognitive activity (Marshall, 2002), that tracks the fast punctual pupillary dilations and may indicate increases of neuron firing (Hampson et al., 2010), the presented LF/HF ratio of pupillary power spectrum indicates the overall tonic state during mental activity. Thereby, the LF/HF ratio measures the cognitive load with lower temporal resolution, but requires lower sampling frequency compared with the index of cognitive activity.

4.5 CONCLUSION

This study showed that the amplitude of task-evoked pupillary response depends on luminance conditions. As you have noticed, color of this page is grayish. Therefore, when you started to read it, you pupils have slightly contracted, once adapted after the previous white page, more bright. Moreover, if you would try to compute, let say, 37×13 in mind, while reading this paragraph, the change of your pupil size would be more important if you would do the same calculation while reading the previous bright page. Therefore, the interpretation of pupillary data in complex ecological settings, where it is difficult to constantly control for display luminosity, might be done carefully. For

example, a stimuli-locked dilation of 0.5 mm on a bright vs. dark fixation area would not have the same interpretation in terms of mental effort. Thus, an extra calibration procedure could be performed prior to the experiment, in order to deduce a relationship between the extent of cognitive pupillary dilation and the current luminance level. It would be also interesting to investigate the interaction effects of the ambient illuminance and the fixation area luminance on the pupillary cognitive dilation. Such experiments could lead to a model linking the amplitude of TEPR and light conditions to facilitate calibration procedure.

Furthermore, the frequency analysis turns to be an efficient tool for pupillary data investigation. Different factors impact differently pupillary frequency components. While the luminance impacts both high and low frequency components, the load on memory factor manifests itself only within low frequency band. Therefore, we can construct some useful features based on signal frequency components, for example the LF/HF ratio of pupillary power spectrum. The presented pupillary LF/HF ratio could potentially be an efficient objective measure of mental effort based on pupil diameter that does not depend on luminance conditions. Besides its off-line applications for pupillary signal analysis, this finding could be helpful in creating a near real-time luminance-independent metric for mental workload estimation. Thus, there will be no need of measuring the point of fixation and its neighborhood luminance to correct the pupillary measurements. Mental effort could be detected with a simple camera measuring pupil diameter.

Future studies could validate the current findings by recording simultaneously the cardiac, respiratory and pupillary activities under different luminance conditions with longer periods of interest. The confrontation of cardiac and respiratory data (on one hand) and pupillary data (on the other) would give extra physiological sense to pupillary frequency components.

5

THE IMPACT OF LUMINANCE ON PUPILLARY RESPONSES TO SUSTAINED COGNITIVE LOAD

> Plus claire la lumière, plus sombre l'obscurité...
> Il est impossible d'apprécier correctement la lumière sans connaître les ténèbres.
>
> JEAN-PAUL SARTRE

> Brighter is the light, darker is the obscurity...
> It is impossible to appraise correctly the light without knowing the darkness.
>
> *my loose translation of* JEAN-PAUL SARTRE

5.1 INTRODUCTION

This chapter was submitted to the International Journal of Psychophysiology and currently is under review (Peysakhovich et al., 2016b). Previous chapter showed that the light conditions can impact the cognitive pupillary reaction. However, it is still unclear how luminance modulates the sustained and transient components of pupillary reaction — tonic pupil diameter and phasic pupil response. In the present study, we investigated the impact of the luminance on these two components under sustained cognitive load. Fourteen participants performed a novel working memory task combining mathematical computations with a classical n-back task. We studied both tonic pupil diameter and phasic pupil response under low (1-back) and high (2-back) working memory load and two luminance levels (gray and white). We found that the impact of working memory load on the tonic pupil diameter was modulated by the level of luminance, the increase in tonic pupil diameter with the load being larger under lower luminance. In contrast, the smaller phasic pupil response found under high load remained unaffected by luminance. These results showed that the cognitive pupillary reaction is impacted by luminance - tonic pupil diameter (phasic pupil response) being modulated under sustained

(respectively, transient) cognitive load. These findings also support the relationship between the locus-coeruleus system, presumably functioning in two firing modes — tonic and phasic — and the pupil diameter. We suggest that the tonic pupil diameter tracks the tonic activity of the locus-coeruleus while phasic pupil response reflects its phasic activity. Besides, the designed novel cognitive paradigm allows the simultaneous manipulation of sustained and transient components of the cognitive load and is useful for dissociating the effects on the tonic pupil diameter and phasic pupil response.

In the present study, we manipulated the sustained cognitive load and the screen luminance. In order to study both tonic and phasic pupil response and so that both components would reflect cognitive processing, we used the Toulouse N-back Task – a novel working-memory task that couples n-back task with mathematical problems solving. This paradigm has the particularity of to combine sustained memory load during a block and transient stimulus processing during each trial. We did not manipulate the transient load, and the stimulus processing was equal for all conditions. The objective of the study was to investigate the impact of luminance on the tonic and phasic pupil response during various levels of sustained cognitive load. We assessed the following questions: a) How the luminance impacts both tonic and phasic pupillary components under different sustained cognitive load conditions? b) What is the relationship between tonic and phasic pupil response during sustained cognitive load?

5.2 MATERIALS AND METHODS

5.2.1 *Subjects*

The subjects were 14 healthy volunteers (4 females, 2 left-handed, age 26.6±5.0, educational level 15.9±2.4), students and staff of ISAE-SUPAERO (French Aerospace Engineering School). All reported normal auditory acuity and normal or corrected-to-normal vision, had no history of neurological diseases and were free of the regular use of medication. All participants gave their informed written consent in accordance with local ethical board requirements prior to the experiment.

5.2.2 *Experimental design and procedure*

The experiment was conducted in a dimly lit sound-attenuated room with one indirect light source behind the participants' back. The ambient illuminance was about 10 lux at the site of participants' eyes. Participants were seated at a viewing distance of approximately 65 cm from the 22-inch LCD monitor (1680 × 1250 pixels screen resolution) with a refresh rate of 60 Hz. Stimulus display and behavioral data acquisition were conducted using Psychophysiological Toolbox V3 for Matlab.

Participants performed the Toulouse N-back Task (Causse et al., 2017; Mandrick et al., 2016) — an N-back task coupled with mathematical calculation —- where participants have to solve a simple mathematical formula in order to perform the n-back task. Two levels of working memory load were produced with 1-back and 2-back tasks. Two levels of luminance were produced by changing the screen background from block to block that was either gray ($\sim$11 cd/m^2) or white ($\sim$28 cd/m^2). As illustrated in Figure 26, each block began with the announcement of the working memory load ("1-BACK" or "2-BACK"; $1.76° \times 7.88°$ in the center of a screen) for 15 s. It allowed participants to calm down between blocks but primarily served as an accommodation period to the screen luminosity. Each bloc was comprised of 25 trials that began with the presentation of a mathematical problem ($1.76° \times 6.15°$ in the center of a screen) for 3,000 ms, followed by a 1,000-ms blank screen. Participants had to resolve the current problem and then to match the result with the previous (1-back) or with the result of the problem two presentations earlier in the sequence (2-back). Subjects were instructed to respond as quickly and accurately as possible for each trial. They had to answer via a response Cedrus Pad placed under their right and left index fingers and containing a green "yes" key and a red "no" key. Participants were told to press "no" key for the first trial in the 1-back task and for the first two trials in the 2-back task. Each block contained 8 matches.

Prior to the main experiment, to familiarize themselves with the Toulouse N-back Task, participants performed two blocks of each load level with both gray and white backgrounds containing 10 practice trials. The practice phase contained a feedback for each trial indicating whether the response was correct or not. Then participants completed 12 blocks (3 blocks for each of 4 conditions) with a short pause after each four blocks. The experiment lasted for a total of about 30 minutes.

5.2.3 *Pupillary recording and processing*

Participants' gaze position and pupil diameter were recorded at 120 Hz with a remote SMI RED eye-tracker (SensoMotoric Instruments GmbH, Germany). This device allows tracking the pupil diameter with precision despite the absence of a chinrest and small head movements. At the begging of the experiment and after each four blocks, a 5-point calibration was validated with four additional fixation points, until a precision of gaze position was inferior to 1°.

The data analyzes were performed using Matlab R2014b (Mathworks, USA). The data losses (including eye blinks) were replaced using linear interpolation. To minimize the eyelid closure effect in the neighborhood of blinks (considered as such if the data loss exceeded 30 ms), 12 adjoined samples (100 ms) from each side of a blink were replaced as well. Then the data were smoothed using a "two-pass" 9-point filter. Trials were segregated according to experimental conditions and averaged point-by-point,

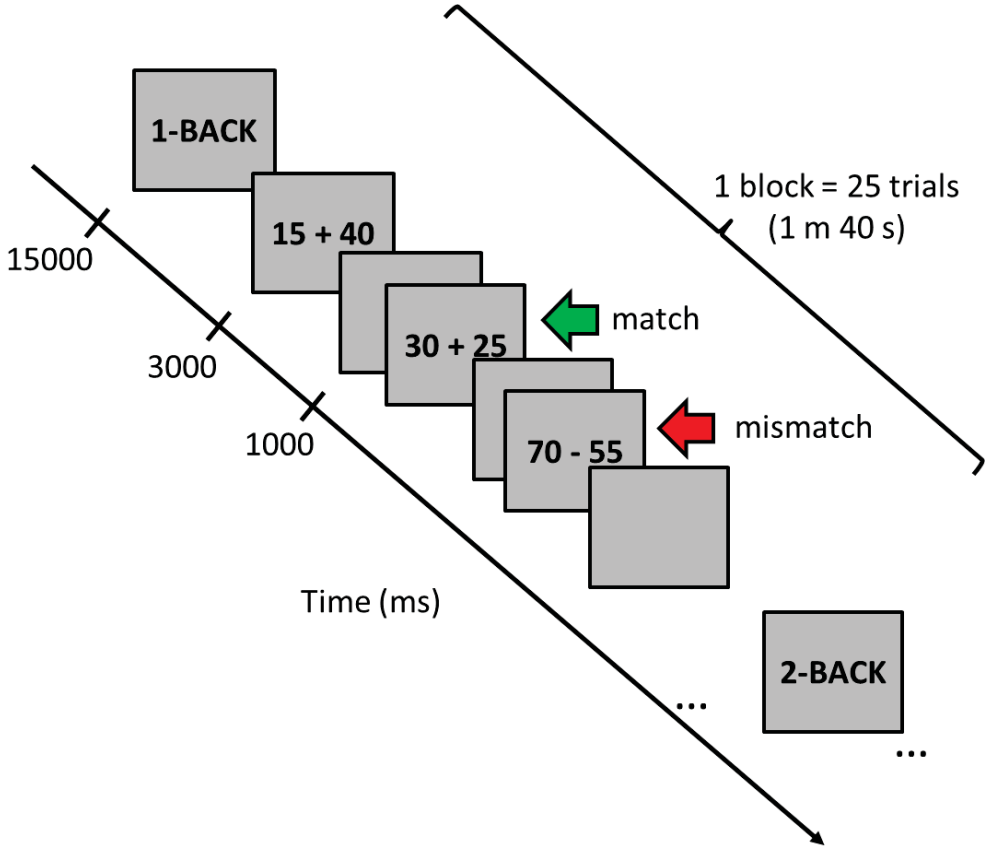


Figure 26.: Illustration of the Toulouse N-back Task paradigm.

giving a pupillary per condition per participant. A trial was validated for the statistical analysis if 1) all the gaze points during the trial and the baseline were within 400×400 pixels square in the screen center, and 2) there were at least 50% of trial data without any original blink or data loss. With these criteria, a mean of 52±18 trials was available for the statistical analyzes. A two-way ANOVA showed that a number of valid trials per analyzes were condition-independent ($ps > .28$). For statistical analyzes, the mean value between 1 and 3 s post-stimulus was used as the tonic pupil diameter. For the phasic pupil response, a maximum value within the same interval was used, after subtraction of a baseline defined as a median of 500 ms pre-stimulus.

5.2.4 *Statistical analyzes*

Statistical analyzes were performed using Statistica 10 (StatSoft, USA) software. Descriptive data were presented as a mean ± standard error. The significance level for all statistical tests was set at 0.01. Four two-way repeated-measures ANOVA were per-

formed on mean RT, accuracy rate, tonic pupil diameter, and pupil phasic response with working memory load (1-back vs. 2-back) and luminance (gray vs. white) as the two within-subject factors. Tukey's Honestly Significant Difference (HSD) was used for post hoc testing. Partial eta-squared (partial η^2) was reported to demonstrate the effect size in ANOVA tests when the effect was significant.

5.3 RESULTS

5.3.1 *Behavioral performance*

The behavioral performance measures are presented in Figure 27. The main effect of working memory load on accuracy was significant, $F(1,13) = 19.2, p < .001$, partial $\eta^2 = .60$, participants performing better at the 1-back task compared with the 2-back task. There were neither luminance, $F(1,13) = 1.24, p = .28$, nor interaction effect on accuracy, $F(1,13) = 0.05, p = .83$. The main effect of working memory load on RT was also significant, $F(1,13) = 48.8, p < .0001$, partial $\eta^2 = .79$, participants being faster to answer in the 1-back task than in the 2-back task. There were neither a luminance, $F(1,13) = 0.73, p = .41$, nor an interaction effect on RT, $F(1,13) = 1.05, p = .33$.

5.3.2 *Tonic pupil diameter*

The measures of tonic pupil diameter are presented in Figure 28A. The main effect of working memory load on the tonic pupil diameter was significant, $F(1,13) = 64.0, p < .00001$, partial $\eta^2 = .84$, corresponding to larger pupils under high working memory load. The main effect of luminance was also significant, $F(1,13) = 122.7, p < .000001$, partial $\eta^2 = .90$, corresponding to larger pupils under dimmer (gray) condition. In addition, there was a significant interaction between working memory load and luminance, $F(1,13) = 38.8, p < .0001$, partial $\eta^2 = .75$. The posthoc analysis of the interaction showed that the differences between the 1-back and the 2-back tasks were larger under the gray background condition ($HSD < .001$) compared to the white background condition ($HSD = .002$). This interaction effect was present for all the participants (see Figure 29).

5.3.3 *Phasic pupil response*

The measures of phasic pupil response are presented in Figure 28B. The phasic pupil response is represented in Figure 30. The main effect of working memory load on phasic pupil response was significant, $F(1,13) = 9.74, p = .008$, partial $\eta^2 = .43$. The phasic pupil response was larger for the 1-back task compared with the 2-back task. There were neither a luminance, $F(1,13) = 0.81, p = .38$, nor an interaction effect,

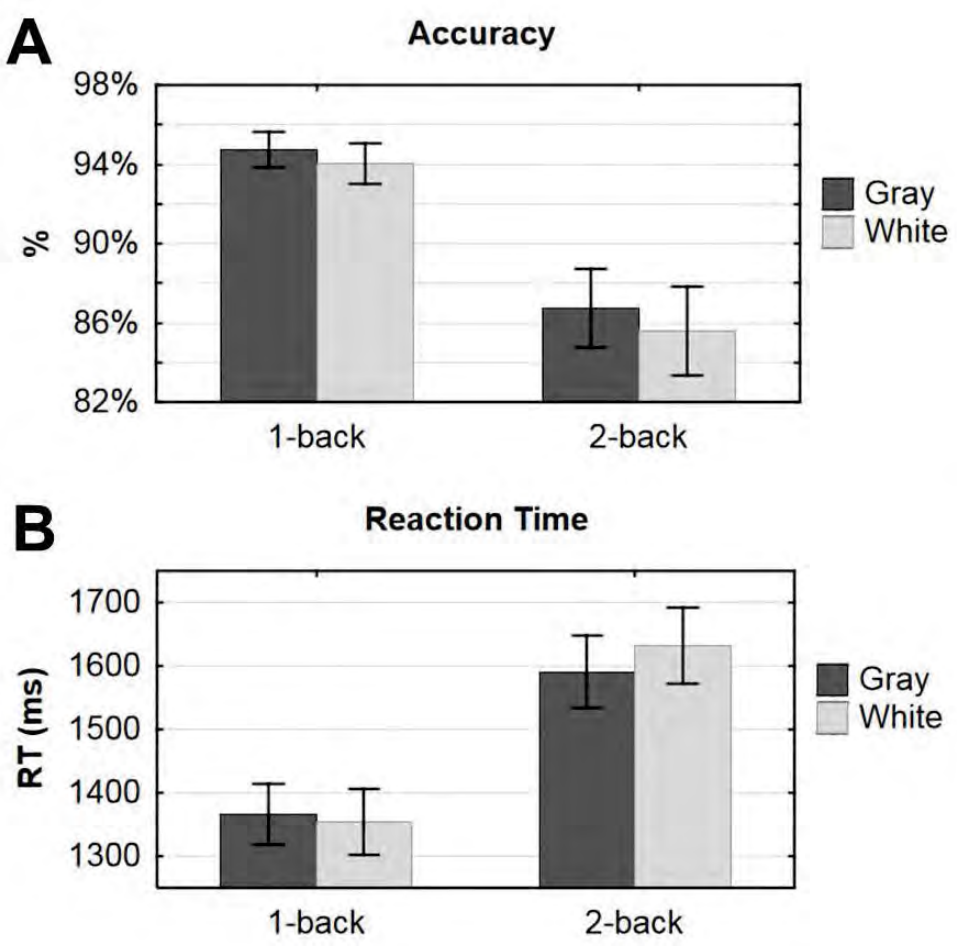


Figure 27.: Mean accuracy (A) and RT (B) for low (1-back) and high (2-back) working memory load in the gray and white background conditions. Error bars denote the standard error of the mean.

$F(1,13) = 0.19, p = .67$. The main effect of working memory load on phasic pupil response was individually present for 11 subjects out of 14.

5.4 DISCUSSION

The goal of this study was to examine the impact of luminance on sustained and transient components of cognitive pupillary response – tonic pupil diameter and phasic pupil response — during sustained cognitive load. To that end, we designed a novel paradigm — Toulouse N-back Task —- that allows a simultaneous study of both sustained and transient components of cognitive pupillary response. This working-memory task couples

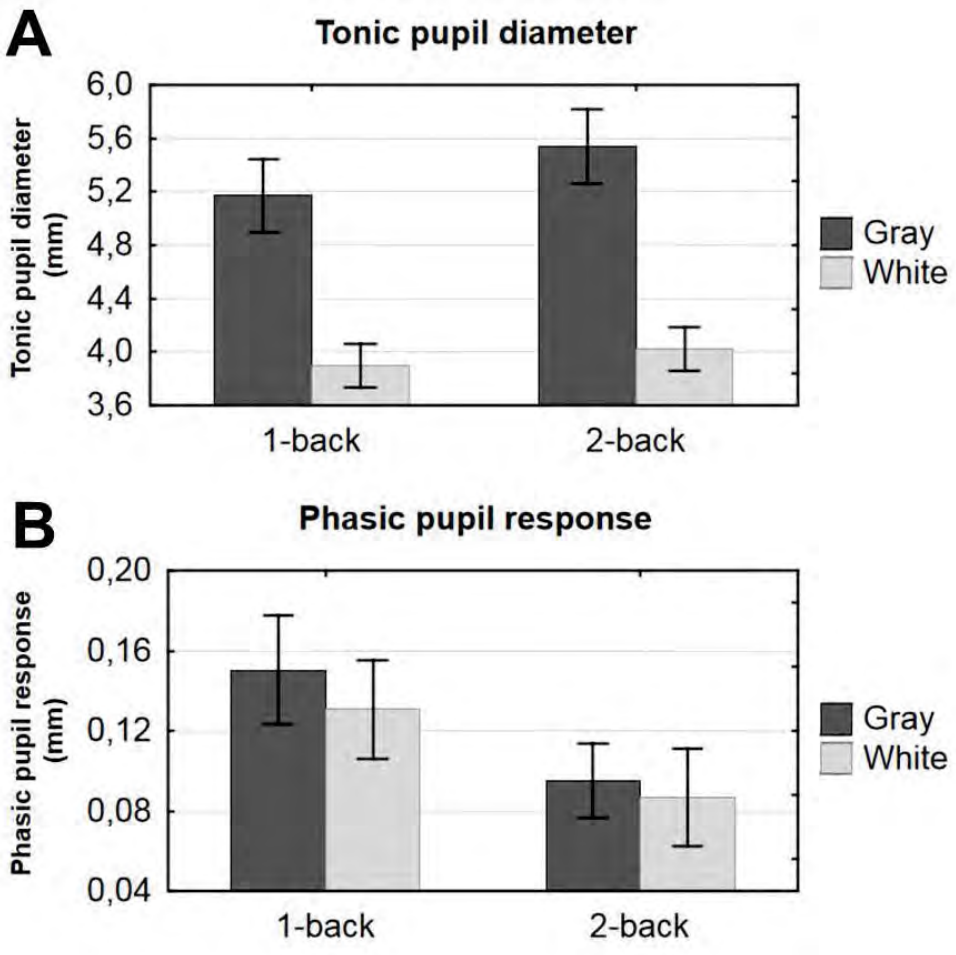


Figure 28.: Mean tonic pupil diameter (A) and phasic pupil response (B) for low (1-back) and high (2-back) working memory load in the gray and white background conditions. Error bars denote the standard error of the mean.

classical n-back task with mathematical problem-solving. The task induces sustained cognitive load, maintained during the whole block. The mathematical operations, on the other hand, need to be resolved punctually for each trial and are of equal difficulty for both 2-back and 1-back conditions. We obtained pupillary data for low and high working memory load conditions (1-back and 2-back respectively), the screen background being gray or white, modifying thus its luminance. We aimed to investigate the impact of the luminance on both tonic and phasic components under different sustained cognitive load and the relationship between these two components.

The behavioral results showed that the 2-back condition was more difficult compared to the 1-back condition. Participants were less accurate and took more time to give

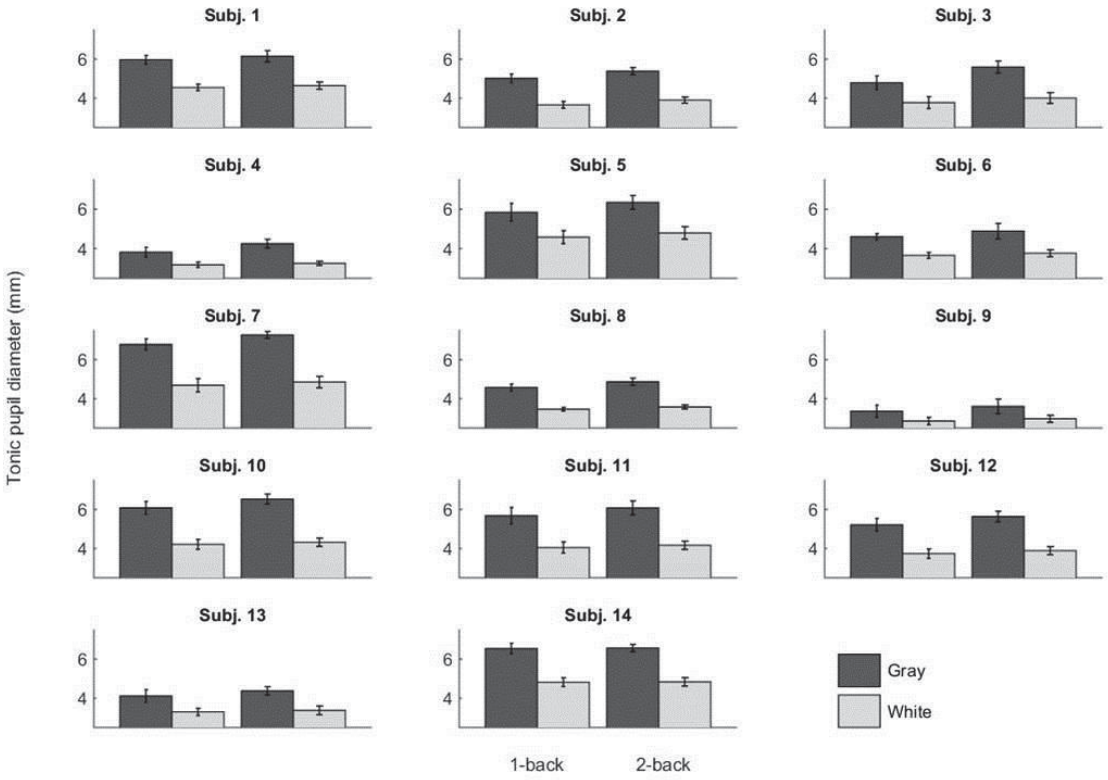


Figure 29.: Tonic pupil diameter for the low (1-back) and high (2-back) working memory load in the gray and white background conditions from all subjects (N=14). Vertical bars denote standard deviation across trials for each subject.

an answer in the 2-back condition. As expected, the screen luminance did not affect behavioral performances. The analysis of the tonic pupil diameter showed that it was modulated by both working memory load, being greater under high working memory load condition, and luminance, being larger under dimmer condition. We also found a significant interaction, the difference between two working memory load conditions being larger under gray luminance condition. In turn, the analysis of the phasic pupil response showed that it was modulated exclusively by working memory load but not luminance. The following discussion is divided into two parts – first, we discuss the impact of the luminance on the two components of cognitive pupillary response; second, we reflect upon the relationship between the phasic pupil response and tonic pupil diameter. Also, before discussing the results of the experiment, a few words are worth being said about the study limitations. The first limitation concerns the population. The

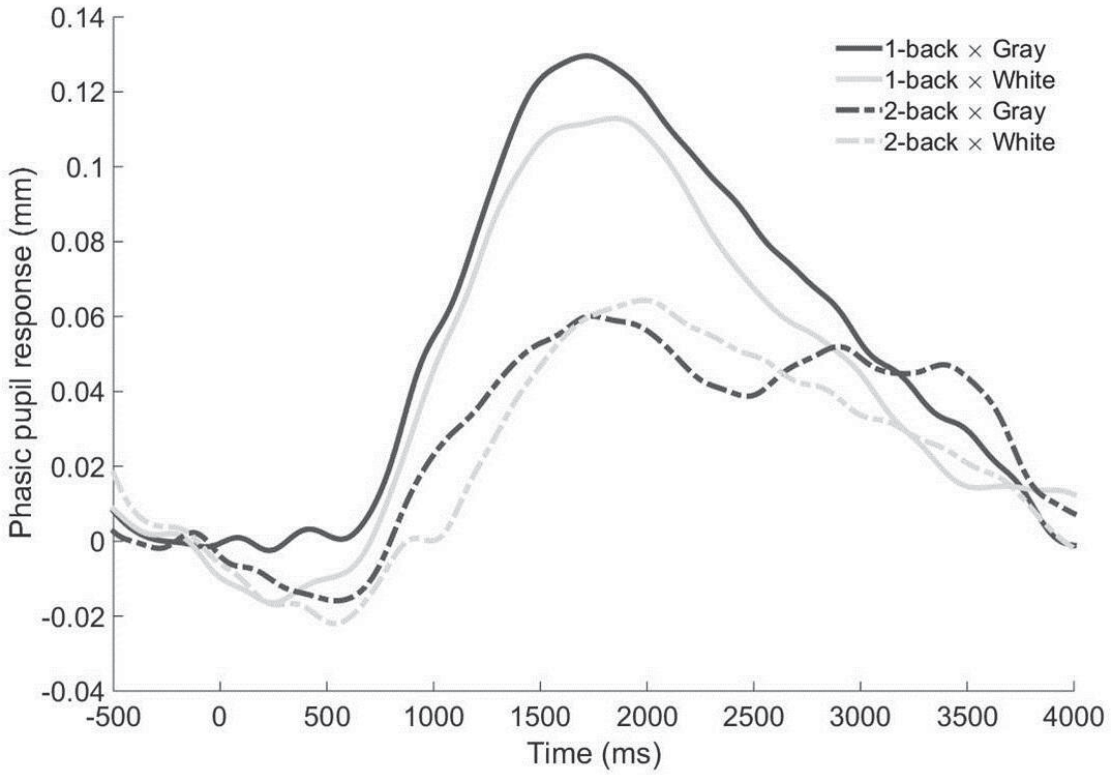


Figure 30.: Mean phasic pupil response for the low (1-back; solid lines) and high (2-back; dashed lines) working memory load in the gray and white background conditions.

relatively small number of participants appears to be sufficient, given the size effects values and the presence of effects individually subject by subject (see further discussion). Further, we did not forbid caffeine intake before the experiment, and more than half of participants took coffee that might alter the pupillary behavior. However, given the within-subject design, we believe that the caffeine did not 19 significantly influence our results in any way. The second limitation concerns the 2×2 experimental design and the number of factor levels. As mentioned in the introduction, the pupil size range might be the problem in pupillary studies, and therefore we chose the two intermediate levels of luminance as used in our previous study (Peysakhovich et al., 2015a). We believe that our results are valid because their direction "more dilation under dimmer background" in the case of a range problem would be rather "less dilation under dimmer background." Therefore, we used a robust 2×2 design to keep the time-on-task reasonable. However, it would be interesting in further 1 studies to explore more levels of luminance to verify the validity of our results. Note that the inverse relationship between the tonic pupil di-

ameter and phasic pupil response (see below), consistent with our findings and using the same cognitive task and three levels of working memory load, was recently published by Mandrick et al. (2016).

5.4.1 *Impact of luminance on tonic and phasic pupil responses*

As showed the analysis of tonic pupil diameter – the pupils were larger in the 2-back condition than in the 1-back condition, and larger under lower luminance condition. Both main effects are well-established and known from the literature (Brouwer et al., 2014; Winn et al., 1994). But most interestingly, there was an interaction between the two factors. The difference between 1-back and 2-back conditions was larger under lower luminance. Moreover, this group effect was individually present for all the participants. Phasic pupil response was, in turn, found to be modulated by the working memory load only, neither luminance nor interaction of the two factors being found. Phasic pupil response was smaller in the 2-back condition compared to the 1-back condition, the effect being individually present in 11 out of 14 participants.

Because the used paradigm elicits sustained cognitive load, tonic pupil diameter is of particular interest because it reflects the sustained component of the pupillary response. The mathematical computations being of the same difficulty across working memory load condition, the transient load is roughly the same. Therefore, the modulation of the phasic pupil response by working memory load condition is a consequence of the relationship between tonic pupil diameter and phasic pupil response that is further discussed in the following subsection.

To the best of the authors' knowledge, this study was the first to report the impact of luminance on the cognitive pupillary response to sustained cognitive load. Previously, Steinhauer et al. (2004) studied the impact of the ambient illuminance on the pupillary response to sustained load. Nevertheless, they did not separate pupillary recordings into tonic and phasic responses but rather studies the 5-s period before the task and the difference of an average of 60-s response and the baseline period. Such a methodological issue does not allow comparison between our findings with their results. At the other hand, Peysakhovich et al. (2015a) recently reported their findings on impacts of the luminance on the cognitive pupillary response to transient cognitive load (in a short-term memory task). They found that the phasic pupil response was modulated by the luminance for a given amount of working memory load the response was larger under dimmer luminance. Our results complete their findings and we can summarize the impact of the luminance on tonic pupil diameter and phasic pupil response in the following way. The luminance impacts the cognitive pupillary reaction with the response to the same stimulus being larger under lower luminance condition. In the sustained cognitive load paradigms, the changes of tonic pupil diameter are affected; in turn, in the tran-

sient cognitive load paradigms, it is the phasic pupil response which is modulated by the luminance.

Peysakhovich et al. (2015a) speculated that the differences in their findings with the results of Steinhauer et al. (2004) were due to the nature of the cognitive task: sustained versus transient. But the present chapter showed that even in response to sustained cognitive load, the pupillary response is greater under dimmer luminance. Given the recent findings by Benedetto et al. (2014) that the pupil diameter is more impacted by screen luminance rather than the ambient illuminance, it would be interesting though to perform an experiment containing both sustained and transient cognitive load varying both screen luminance and ambient illuminance to further investigate the interaction between these factors.

The present study also shows the importance of dissociating tonic pupil diameter and phasic pupil response. This proposition is supported by the theories about working memory functioning and the existence of the two mechanisms of activation – sustained and transient (Cohen et al., 1997; Reynolds et al., 2009). As the results showed, depending on the sustained versus transient nature of the cognitive task, the luminance modulates the corresponding component of the pupillary response. Note also that the designed Toulouse N-back Task allows the simultaneous manipulation of the sustained and the transient cognitive load, which is useful for the future studies of the relationship of these two pupillary components.

It is also worth noting that the modulation of tonic pupil diameter and phasic pupil response are probably linked to the cognitive aspects of the tasks. Thus, in studies by Bradshaw (1969) and Gilzenrat et al. (2010[Experiment 1B]), the authors used the reaction time task and auditory oddball discrimination tasks but did not report any significant differences under different light conditions. That may indicate that the effects of luminance are stronger when performing a more demanding cognitive task such as the Toulouse N-back Task.

Additionally, the fact that under a constant transient load the phasic pupil response was not modulated can have an interesting application to design a luminance-independent measure of cognitive load. Indeed, given an existing relationship of phasic pupil response and tonic pupil diameter (see section below) and the luminance-independent phasic response, it would be possible to induce the level of the load by measuring the pupil response to some probe of a constant load (for instance, an auditory probe).

5.4.2 *Relationship of phasic pupil response and tonic pupil diameter*

The analysis of the peak dilation showed that the phasic pupil response was smaller for the 2-back condition compared to the 1-back condition. The luminance had no influence on the phasic pupil response. The same pattern of smaller phasic pupil response for the 2-back condition held true for both luminance conditions. Comparing these results with

the findings on tonic pupil diameter, we conclude that a more important load on memory evokes larger tonic pupil diameter and is associated with the smaller phasic response. The absence of any luminance impact on pupil phasic response demonstrates that the relationship between large tonic pupil diameter and smaller phasic pupil response is not an issue of pupillary dynamic range but cognitively driven. These findings bring new light on the law of initial value (in its physiological and not mechanical interpretation) concerning the pupil diameter (see (Van Gerven et al., 2004)). It shows that the law of "greater baseline – smaller reactivity" stands true contrary to the Jin's interpretation (1992). Note also that the smaller reactivity is not an issue of disengagement from the task, the accuracy rates for the high working memory load condition being elevated (around 86%).

This is the first study to report such a pattern of larger tonic pupil diameter corresponding to lower phasic pupil response in a cognitive task inducing working memory load. Earlier, such relationship was found in auditory vigilance paradigms (Gilzenrat et al., 2010; Knapen et al., 2016; Murphy et al., 2011) and a perceptual decision-making task (de Gee et al., 2014). But while in these studies, the tonic pupil diameter was modulated by arousal and vigilance of participants, in the present study, tonic pupil diameter was modulated by working memory load. These findings are in line with theories about the relationship between pupil size and LC-NE system. LC is a small brainstem nucleus that provides the majority of norepinephrine to the brain (Aston-Jones and Waterhouse, 2016; Samuels and Szabadi, 2008). The LC-NE system was found to be closely related to the pupil diameter, first in primates (Rajkowski et al., 1993), then in humans (Murphy et al., 2014). Recent reviews by Joshi et al. (2016) and Costa and Rudebeck (2016) discussed the complex relationship between pupil size and LC-NE system and the direct and indirect evidence supporting this relationship. Note, in particular, that Joshi et al. (2016) found that the relationship is not specific to the LC-NE and also present in the inferior and superior colliculi and anterior and posterior cingulate cortex. But because these brain regions are interconnected with the LC, the results nevertheless suggest that the pupil diameter reflects LC-mediated coordination of neuronal activity. Although the pupil diameter is often interpreted as a biomarker of LC activity, the exact neural substrates linking cognitive state and pupil changes remain unclear. Thus, the recent results of Wang and Munoz (2015) and Lehmann and Corneil (2016) on primates suggested the existence, in addition to LC-mediated arousal circuits, of a parallel pathway from the frontal cortex to pupil diameter through the Superior Colliculus (SC). This SC pathway should assume primacy for pupil orienting response such as in auditory vigilance paradigms used in previous studies establishing the inverse relationship between tonic pupil diameter and phasic pupil response. Given that in our experiment, the working memory load was highly involved, we suppose that the dominant pathway responsible for pupil changes in our case was LC-mediated arousal circuit. Hence, together with our results published in Mandrick et al. (2016), to our best knowledge, this is the

first report of an inverse tonic-phasic relationship in a paradigm using working memory load. Furthermore, the findings of the present study are, in particular, supported by the theory of two existing modes of LC-NE firing. The LC-NE system supposedly functions according to a continuum of states between tonic firing mode and phasic firing mode (Aston-Jones and Cohen, 2005; Aston-Jones et al., 1999). High tonic firing mode is associated with high overall arousal. Phasic firing mode allows for selective responses to a particular target. The relationship supposedly resembles an inverted-U shape. A moderate level of tonic activity is needed to be aroused enough to perform the task, but an excess of such activity should be harmful to the performance because the phasic firing would be drowned in the tonic arousal. Such understanding of the LC-NE functioning also supports the hypothesis of a greater tonic pupil diameter implying smaller phasic pupil response. Together with previous findings in attentional and perceptual tasks, our results using a cognitive paradigm, give another strong evidence of the relationship between LC-NE and pupil size and, in particular, the association between tonic pupil diameter and tonic LC firing mode and phasic pupil response and phasic LC firing mode.

5.5 CONCLUSION

In the present study, we showed that the screen luminance has an impact on the cognitive pupillary reaction. It is an important issue, together with the impact of ambient illuminance, when performing pupillometric experiments in ecological conditions, and also when comparing results issued from different laboratories or setups. We dissociated tonic pupil diameter and phasic pupil response and showed that depending on the nature of the cognitive load —- sustained or transient —- the corresponding component of the pupillary response would be impacted. We can postulate that the same amount of cognitive load under dimmer luminance condition would elicit larger tonic pupil diameter in a sustained load paradigm and larger phasic pupil response in a transient load paradigm.

Furthermore, we found a smaller phasic pupil response for a larger tonic pupil diameter. This finding supports the relationship between the LC-NE system, presumably functioning in two firing modes — tonic and phasic — and the pupil diameter. We suggested that the tonic pupil diameter tracks the tonic activity of the LC-NE system while phasic pupil response reflects the phasic activity of the LC.

Finally, we designed a novel cognitive paradigm — Toulouse N-back Task — that allows the simultaneous manipulation of sustained and transient components of the cognitive load. Therefore, it allows the further investigation of the complex relationship between the tonic pupil diameter and phasic pupil response, and, especially, the interaction of screen luminance and ambient illumination.

In the future, it will be interesting to test the frequency approach like in the previous chapter in order to discover whether there are some parts of frequency band that are more sensitive to the working memory or luminance. We've tested the same frequency bands and the results were not significant. The differences in task nature and period length might be an explanation. A further investigation is needed.

6

GAZE VISUALIZATION: QUALITATIVE EXPLORATION

> 'What is the use of a book', thought Alice, 'without pictures or conversations?'
>
> *Alice's Adventures in Wonderland* by LEWIS CARROLL

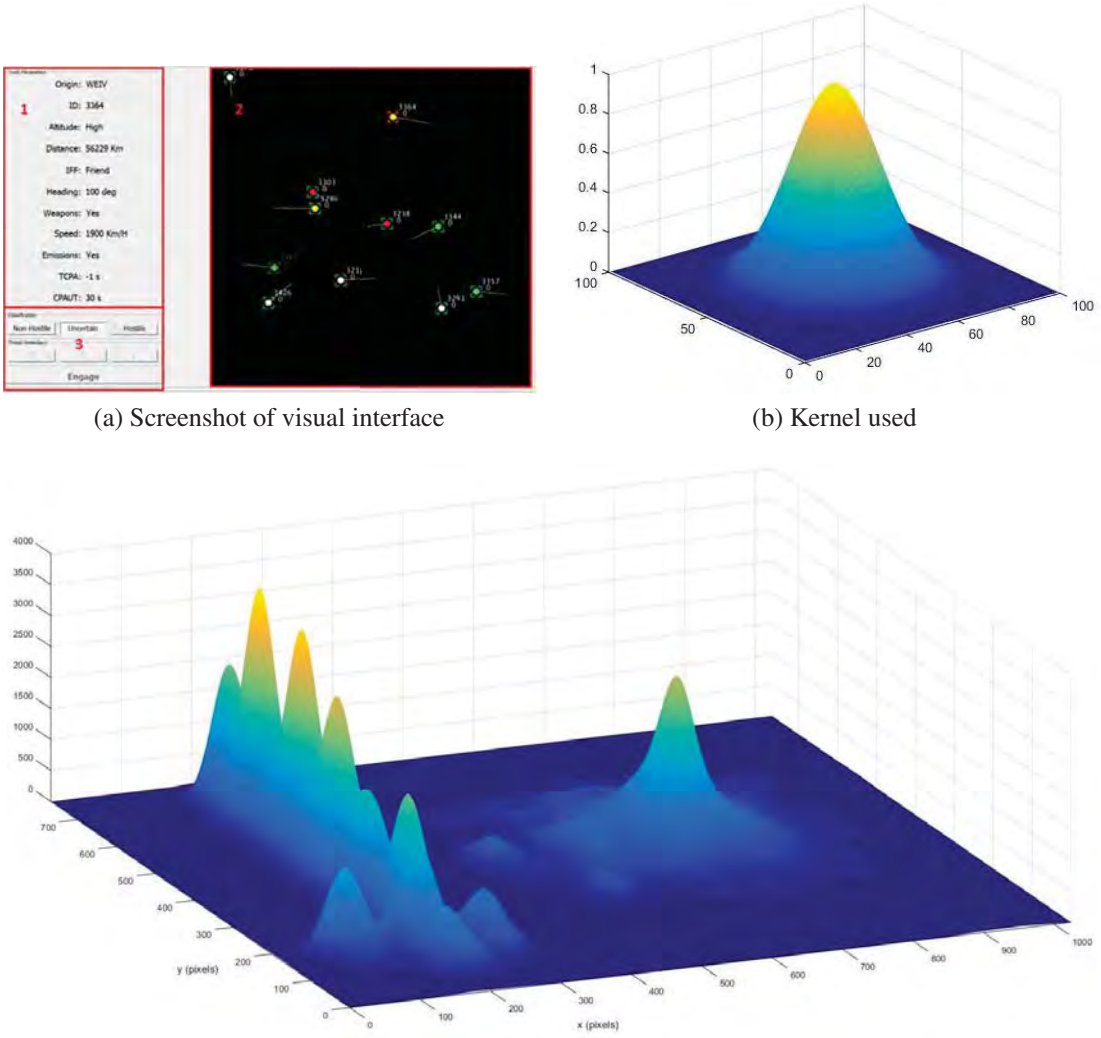
6.1 HEAT MAPS AND SCAN PATHS VISUALIZATIONS

6.1.1 *Attentional and heat maps*

The most popular tool of all eye-tracking practitioners (not only academic) is attention map or heat map (Wooding, 2002). If you search for "eye-tracking" in Google Images, heat maps will represent about 70% of results. These maps are intuitive and easy to understand, even for naive users. That is why they are so popular and are a part of every eye-tracking software. Figure 31 shows an example of two heat maps of a user

(a) Heat map with kernel of 60 px (b) Heat map with kernel of 100 px

Figure 31.: Two heat maps width different settings of observing *The Garden of Earthly Delights* by *Hieronymus Bosch*



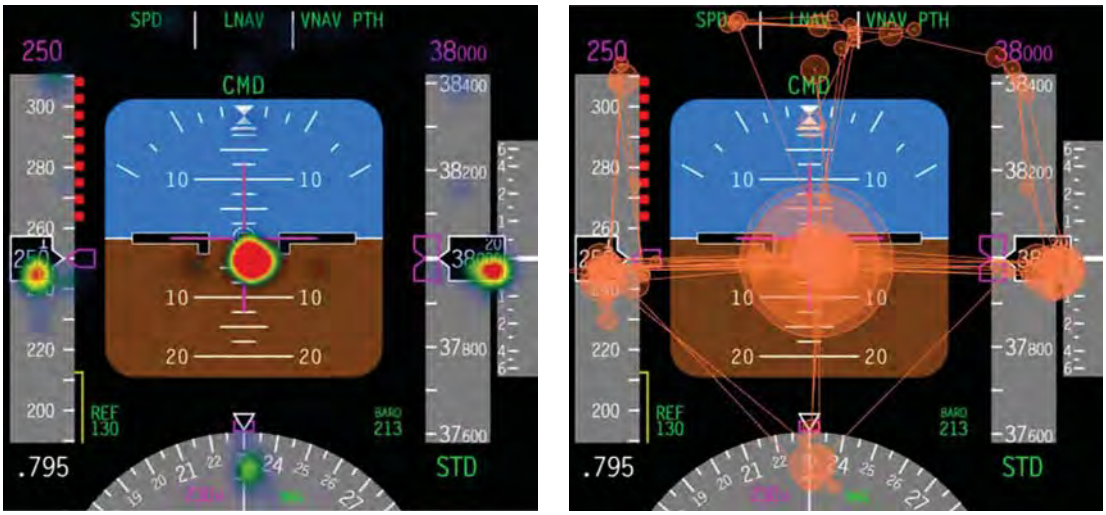
(a) Screenshot of visual interface

(b) Kernel used

(c) Attention map

Figure 32.: An example of attention map of the radar interface.

that observed the wonderful triptych of Hieronymus Bosch for a minute. The “hottest" parts of the images depicted in red are those that were looked the most. The difference between the two maps is the width of the kernel used for visualization, the data being the same. Figure 32 helps to understand better how such images are generated. In this example an operator performs a visuocognitive task with the interface shown in Figure 32a (Peysakhovich et al., 2015c). By putting a kernel around each gaze point, we can construct an attention map (Figure 32c). Its reading is quite intuitive — accumulated kernels generate “mountains" the height of which is proportional to the time spent



(a) Heat map

(b) Scan path

Figure 33.: Heat map and scan path of the same data over the PFD

around the base of the mountain. By looking at this attention map from the top, we obtain a heat map.

6.1.2 *Scan paths and transitions*

A heat map allows us to instantly understand what elements of the scene were looked at. However, this is mostly useful when the elements are not connected between them. On contrary, when different parts are interconnected and the processing of the information depends on the order in which the elements are consulted, the scan paths are more informative. A simple example in Figure 33 with the Primary Flight Display (PFD) — a flight instrument display — depicts both heat map and scan path from the same dataset. The visualizations are generated using SMI BeGaze software, but are quite representative of all software of the market. Figure 33b represents each fixation as a small round the size of which is proportional to the fixation duration. This visualization is useful because we can deduce the existing connections between different elements of the PFD and their importance. We see that there are connections between the attitude indicator (center), heading (bottom), FMA (top), speed (left) and altitude (right). We could guess from the heat map 33a that the most dense connections are between the attitude indicator and those of speed and altitude. Yet, it is intuitive and visual on the scan paths illustration 33b. Several problems of such visualization are however need to be solved so that it would be informative for a good qualitative analysis of the data. Notably, such images become rapidly cluttered when the length of the recording

is significant. In the presented example, less than one minute of data is depicted, and the image begins already to be cluttered. When the duration of the recording is about 30 minutes, such visualization becomes absolutely unreadable. Another problem is the absence of any information about the transition direction. For example, we have no idea in which direction was performed the transition (one line) between the heading and the altitude indicators. Moreover, the east-west and west-east transitions between the speed and the attitude indicators are mixed up. Also, no information about the time of eye movements is represented. Did the FMA was consulted in the beginning of the recording or in the end, for instance? All these issues are need to be solved to have an effective and intuitive illustration of scan paths for good qualitative analysis of eye-tracking data. For a state-of-the-art of eye-tracking visualization techniques before the publication of Peysakhovich et al. (2015b) please consult Blascheck et al. (2014).

This chapter presents a visualization technique called Attribute-Driven Edge Bundling (ADEB). It is one of many so-called edge bundling methods that reduce visual clutter of dense and occluded graphs by aggregating the spatially closed edges (or lines) together. For example, an edge bundling methods would aggregate together all the transitions between the speed and the attitude indicators in Figure 33b in a one thick line. ADEB is the general edge bundling method that can be used for all kind of graphs but was developed especially for eye-tracking data taking into account all the issues mentioned in the previous paragraph.

Existing bundling techniques either ignore edge properties such as direction and data attributes, or are otherwise computationally not scalable, which makes them unsuitable for tasks such as exploration of large trajectory datasets. We present a new framework to generate bundled graph layouts according to any numerical edge attributes such as directions, timestamps or weights. We propose a GPU-based implementation linear in number of edges, which makes our algorithm applicable to large datasets. We demonstrate the method with eye-tracking applications. Another application to aircraft trajectories can be found in Peysakhovich et al. (2015b) where this chapter was partially published.

Idea! Pupillary Reaction Map

Just as like heat maps for gaze distribution, we can compute “pupillary reaction maps". At each gaze point we can associate a Gaussian proportional to the value of pupil diameter at this point. Thus, we will obtain a map of preferences — the hottest places will correspond to the moments of the greatest dilation of the pupil. Nevertheless, such map can be generated only for pictures without dramatic changes in luminance so that the pupillary reaction maps correspond to workload/emotions and not the light conditions. Such maps could be used especially for the evaluations of human-computer interfaces indicating by “hot" areas the parts of the design that induced a lot of information processing or an emotional moment.

6.2 ATTRIBUTE-DRIVEN EDGE BUNDLING (ADEB) FRAMEWORK

Large attributed graphs are ubiquitous in many application domains, such as traffic analysis and planning, network analysis, and bioinformatics. They represent a special subclass of general graphs where edge directions encode essential information for the analysis task at hand. For such graphs, classical visual metaphors such as node-link diagrams produce too much clutter to generate useful pictures.

Graph bundling methods attempt to reduce clutter by grouping edges found to be compatible into so-called bundles (see “previous work" section in Peysakhovich et al. (2015b). Yet, few such methods handle attributed graphs – edge compatibility is mainly based on spatial position and does not use attributes such as edge direction or data values. Reasoning about such attributes (seeing their values reflected in the bundling) is essential in many applications. Bundling methods for attributed graphs exist, but are computationally not scalable, and thus not usable for large-scale data exploration (Lhuillier and Hurter, 2015).

We present *ADEB*, a generalized edge bundling technique based on edge advection in compatible vector subspaces, which extends the recent kernel density estimation edge-bundling (Kernel Density Estimation Edge Bundling (KDEEB)) (Hurter et al., 2012) with the following contributions:

1. generation of bundled layouts where edge compatibility is defined by one or several numerical edge attributes, *e.g.*, direction, time, or weight;
2. a simple and scalable GPU implementation of our method;
3. application of ADEB for exploring eye-tracking datasets demonstrating the method’s added value as opposed to classical bundling of undirected graphs.

6.2.1 *Trail visualization*

Trail datasets consist of sequences of points recording the motion of several objects, *e.g.* vehicles, or tracked trajectories of object parts, *e.g.* eyes or human silhouettes. Such datasets have three main features. First, they can be described as *attributed* graphs, where nodes represent trajectory endpoints and (curved) edge control points represent actual motion paths. Attributes include motion direction, speed, and timestamps. Secondly, they are typically *large*: For instance, 15 minutes of eye-tracking data recorded at 50 Hz already gives 45K sample points. Finally, they are *complex*, *e.g.* contain many intersecting trails covering large spatial domains. All these factors make trail exploration a challenging proposal with respect to attribute depiction, clutter reduction, and computational scalability. We discuss below several methods related to this task.

Eye-tracking exploration: Density maps tackle the visual scalability problem by aggregating spatially close information for eye-tracking analysis (Andrienko et al., 2011,

2012; Marzuoli et al., 2012). Separately, Burch *et al.* transform eye-tracking data into a dynamic graph to help exploration (Burch et al., 2014). Areas of interest (AOI) analysis provides information about dwell times, transitions, fixation points, and AOI hits, but require accurate a priori knowledge to define such areas (Duchowski, 2007; Holmqvist et al., 2011). A few methods exist that are able to group eye-movement paths; however, they are neither visual nor intuitive to use (Le Meur and Baccino, 2013). An overview of eye movement pattern analysis is given by Laube (Laube, 2009).

Given the size and clutter of trail datasets, bundling methods have emerged as an effective way to show the coarse-scale connectivity structure of these datasets (Cui et al., 2008; Ersoy et al., 2011; Holten and van Wijk, 2009a; Hurter et al., 2012, 2014; Lambert et al., 2010). However, as analyzing trail datasets requires reasoning about trail attributes (*e.g.* time-stamps, speed, and direction), undirected bundling techniques are of limited use. For instance, a bundle connecting two node-groups A and B will only tell that A and B are connected, but not in which direction(s) and at which time instant(s) objects moved from A to B. As a side effect, mixing edges of different attribute values in a bundle makes the attribute-based shading-and-blending schemes proposed by several authors (Holten and van Wijk, 2009a; Hurter et al., 2014; Lambert et al., 2010) of limited effectiveness. Conversely, using bundling techniques for attributed graphs (Holten and van Wijk, 2009a; Selassie et al., 2011) is not practical, as these are far too slow for large datasets such as trail data.

We next describe our ADEB method, a framework that bundles large graphs using an edge-compatibility scheme that can be flexibly controlled to include (mixes of) edge attributes. We first outline the main elements of the KDEEB algorithm from which ADEB is inspired (Sec. 6.2.2). Section 6.2.3 presents ADEB's mathematical details. Sections 6.2.4 and 6.2.5 explain how to use attributes to create various edge compatibility metrics. Implementation details are given in Sec. 6.2.6.

6.2.2 *Kernel density bundling*

We first review the main principles of the kernel density estimation edge bundling (KDEEB) technique. Let $G = \{e_i\}_{1 \le i \le N} \subset \mathbb{R}^2$ be the drawing of a graph of N edges, where each edge e_i is represented by a straight line or arbitrary 2D curve. First, an edge-density map $\rho : \mathbb{R}^2 \to \mathbb{R}^+$ is computed using kernel density estimation

$$\rho(\mathbf{x} \in \mathbb{R}^2) = \frac{1}{h^2} \sum_{i=1}^{N} \int_{\mathbf{y} \in e_i} K\left(\frac{\mathbf{x}-\mathbf{y}}{h}\right), \tag{1}$$

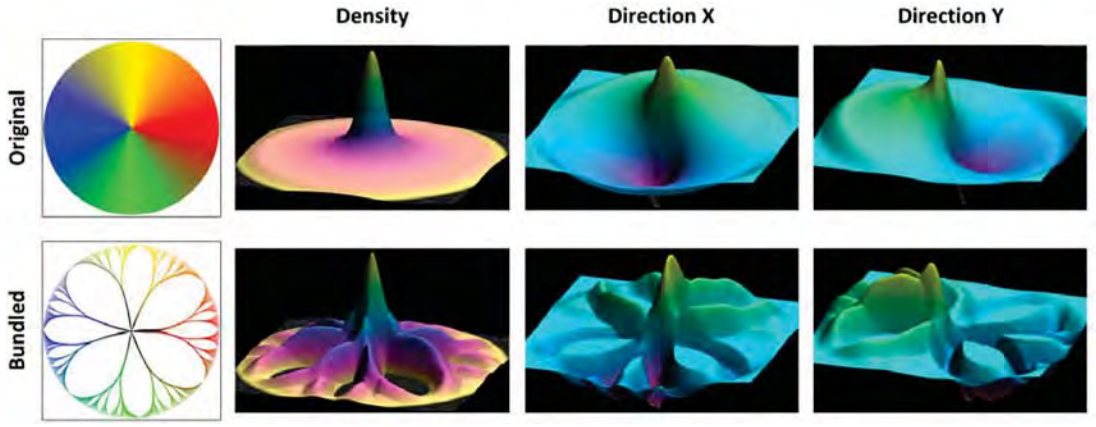


Figure 34.: Trajectories forming a circle with the start point in common in the center. Original and bundled layouts with visualization of density and two directional maps.

where $K : \mathbb{R}^2 \to \mathbb{R}^+$ is a bivariate radial kernel obtained from a symmetric univariate non-negative kernel function of bandwidth $h > 0$, *e.g.* Gaussian or Epanechnikov. The bundling of G is given by the fixed point of the ordinary differential equation

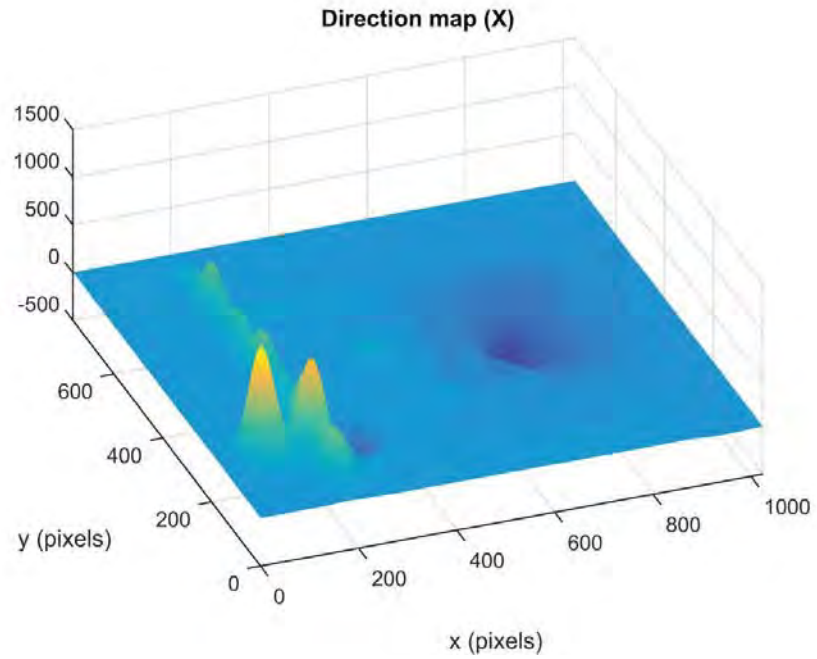
$$\frac{d\mathbf{x}}{dt} = \frac{h(t)\nabla\rho(\mathbf{x},t)}{\max(\|\nabla\rho(\mathbf{x},t)\|,\varepsilon)}, \quad \forall \mathbf{x} \in G, \tag{2}$$

with initial conditions given by the input graph drawing G. The regularization constant $\varepsilon \ll 1$ takes care of zero gradients. After a few Euler iterations used to solve Eqn. 2, during which one resamples edges, recomputes ρ, decreases h, and makes a 1D Laplacian smoothing pass over G's edges, KDEEB converges and aggregates edges into bundles. Note that KDEEB is nothing else but the well-known mean-shift algorithm (Comaniciu and Meer, 2002) applied to the graph drawing G. Hence, KDEEB inherits smoothness, noise robustness, and convergence results proven for mean-shift. For details, we refer to (Hurter et al., 2012).

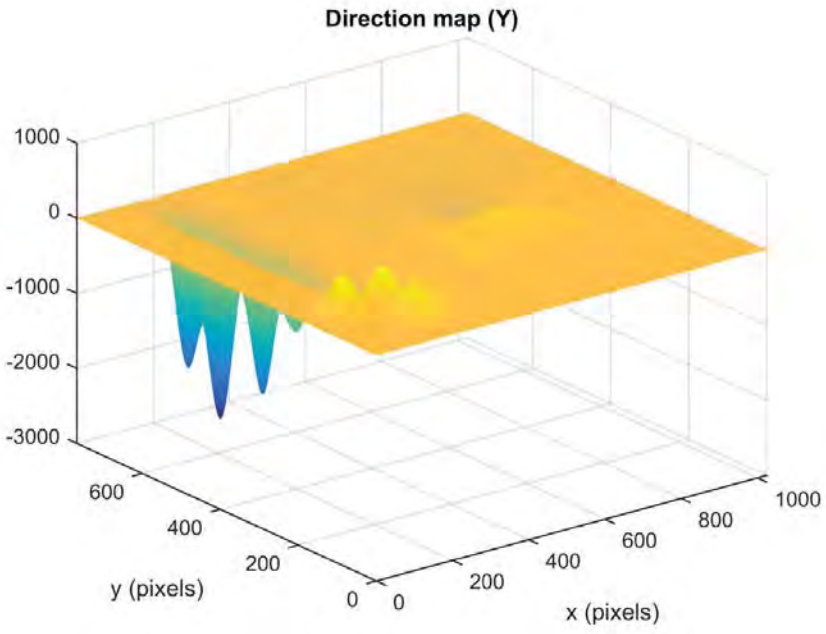
6.2.3 *Mathematical model for ADEB*

We proceed by first introducing a few necessary definitions.

Edge directions: Consider a given drawing $G = \{e_i\}_{1 \leq i \leq N}$ of a directed graph or trail dataset. For each point $\mathbf{x} \in e_i \subset G$, we denote by $\mathbf{d}(\mathbf{x})$ the unit vector tangent to the curve representing e_i, oriented in the direction given by walking on e_i from its start to its end point. For our trail datasets (Sec. 6.2.1), e_i will be typically a 2D curve rather than straight line (as in a classical node-link graph drawing), so $\mathbf{d}$ changes along e_i. Secondly,



(a)



(b)

Figure 35.: A visualization of two directional maps.

note that **d** encodes the local directions of the *initial* graph drawing, given as input to the bundling process.

Flow direction map: Given a graph drawing G as above, the *flow direction map* $\theta : \mathbb{R}^2 \to \mathbb{R}^2$ is defined as

$$\theta(\mathbf{x} \in \mathbb{R}^2) = \frac{1}{h^2} \sum_{i=1}^{N} \int_{\mathbf{y} \in e_i} \mathbf{d}(\mathbf{y}) K\left(\frac{\mathbf{x}-\mathbf{y}}{h}\right), \tag{3}$$

where K is the same kernel as in Eqn. 1. We can think of θ as being a 2D viscous fluid flow whose boundary conditions are given by the **d** values on G. The bandwidth h in this analogy is the fluid viscosity. Note that several points have $\theta(\mathbf{x}) = 0$, *e.g.*, points equidistant from two parallel opposite-direction edges of. In our analogy, these would be flow separation lines. Points **x** close to an edge have values $\theta(\mathbf{x})$ close to the edge direction **d**. The flow magnitude $\|\theta\|$ decreases monotonically and smoothly away from the flow boundary G. The figure 34 shows a basic example of two directional maps. Note greater values towards the center, the most dense area. Another example is provided in Figure 35 and corresponds to two directional maps of dataset from Figure 32. From these maps we can deduce the general scan path. The values around the radar zone are negative along x-axis and small but positive along y-axis. Therefore, the participants moved their gaze from the radar zone to the west (and a bit to the north), that corresponds to the list of parameters in the upper left part of the screen (see Figure 32a). Around this list, the values along x-axis are small and positive and along y-axis are negative. Thus, participants moved their gaze from the upper element of the list to the lowest. Around the buttons, values along x-axis are positive for the two first buttons and negative for the third button, and positive along y-axis for all three buttons. The third button corresponding to the classification of an aircraft as hostile (see Peysakhovich et al., 2015c), participants might look back at the parameter list in order to verify once again whether the aircraft is really hostile and to prevent the friendly-fire. The participants' gaze from the two other buttons moved north-east, to the radar screen in order to choose another aircraft.

Directional compatibility: At each point $\mathbf{x} \in G$, we define a subspace $\Omega_{\mathbf{x},c}$ of compatible directions as

$$\Omega_{\mathbf{x},c} = \left\{ \mathbf{y} \in \mathbb{R}^2 \setminus \ker(\theta) \,\middle|\, \frac{\mathbf{d}(\mathbf{x}) \cdot \theta(\mathbf{y})}{\|\theta(\mathbf{y})\|} \geq c \right\} \subset \mathbb{R}^2, \tag{4}$$

where $\ker(\theta) = \{\mathbf{x} \in \mathbb{R}^2 | \|\theta(\mathbf{x})\| = 0\}$ and $c \in [-1,1]$ is a compatibility factor. This factor represents the cosine of the maximum allowed angle between the edge direction and flow direction at **y**. Thus, $\Omega_{\mathbf{x},0}$ covers all points where the flow direction deviates from $\mathbf{d}(\mathbf{x})$ by $\pi/2$ at most, and $\Omega_{\mathbf{x},-1} \equiv \mathbb{R}^2$. In general, given a point **x** of the initial

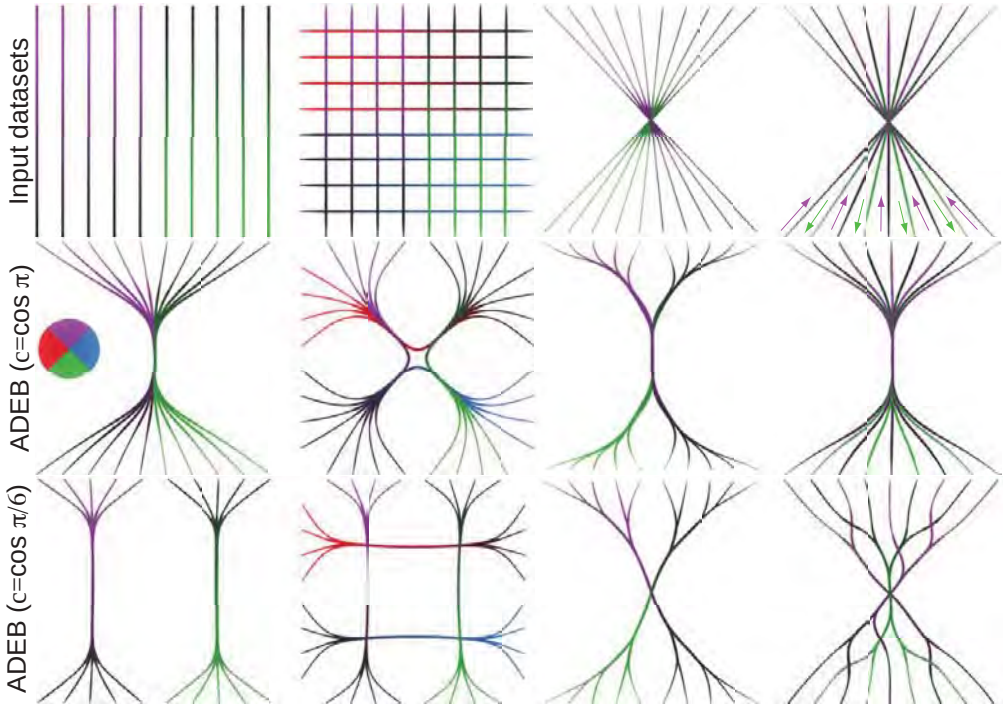


Figure 36.: Influence of compatibility factor on bundling result.

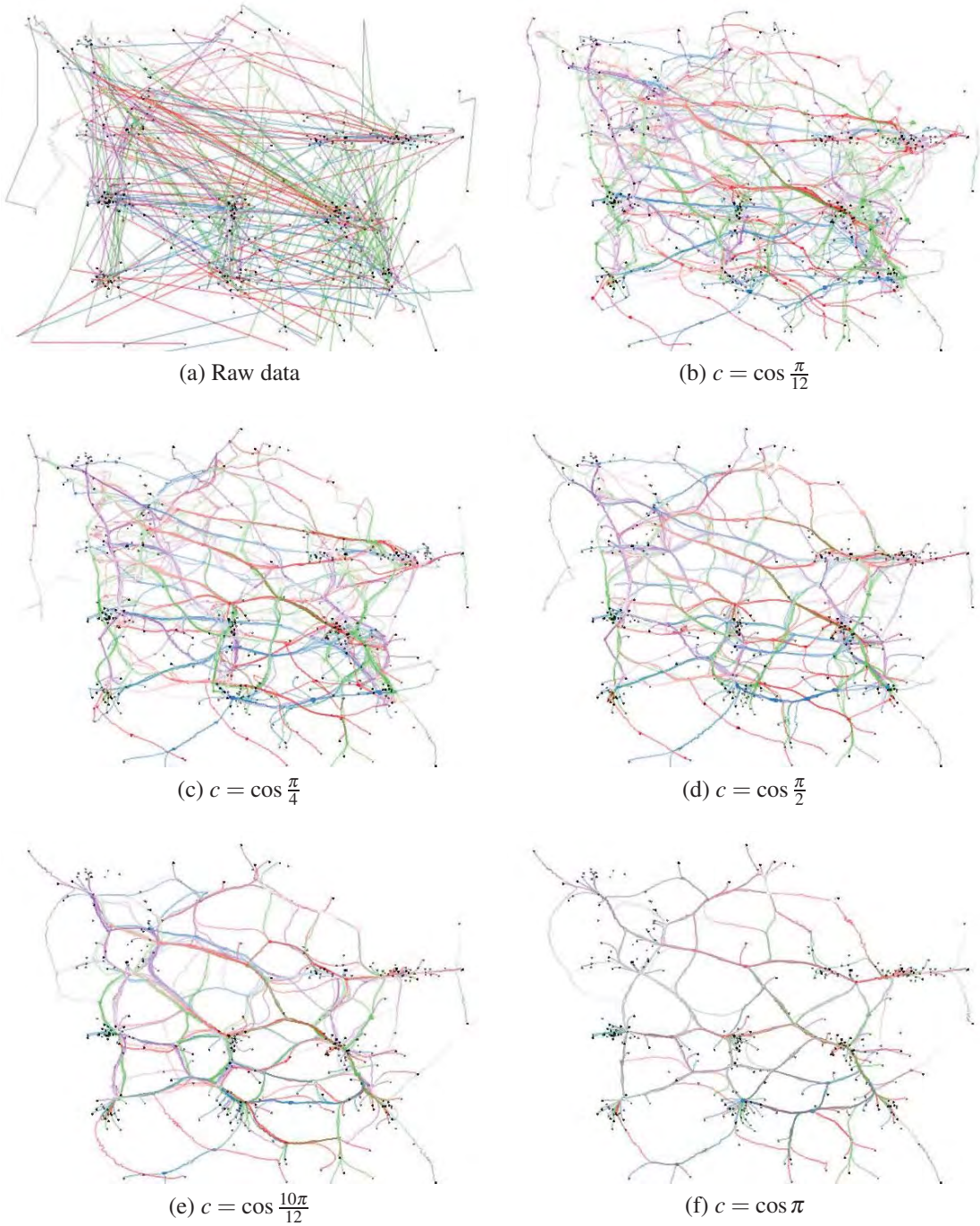
graph drawing G, $\Omega_{\mathbf{x},c}$ gives all points in $\mathbb{R}^2$ where the flow map θ has roughly the same direction as $\mathbf{x}$, subject to the factor c.

Given the ρ and θ maps, the subspace $\Omega_{\mathbf{x},c}$ is defined at each point $\mathbf{x} \in G$. We can now define the ADEB bundling of G as the fixed point of the ordinary differential equation

$$\frac{d\mathbf{x}}{dt} = \frac{h(t)\nabla_{\Omega_{\mathbf{x},c}}\rho(\mathbf{x},t)}{\max(\|\nabla_{\Omega_{\mathbf{x},c}}\rho(\mathbf{x},t)\|,\varepsilon)}, \quad \forall \mathbf{x} \in G \tag{5}$$

In contrast to KDEEB (Eqn. 2), the gradient $\nabla_{\Omega_{\mathbf{x},c}}\rho$ is estimated now *only* over the subspace of compatible directions $\Omega_{\mathbf{x},c}$. In analogy with KDEEB, we solve Eqn. 5 by a few Euler iterations, during which we resample edges, recompute ρ and θ, decrease h, and make a 1D Laplacian smoothing pass to remove small wiggles.

In summary, while KDEEB advects points of a given drawing towards local maxima of its edge density, ADEB advects points towards local density maxima of *compatible* directions. Figure 36 shows the influence of the compatibility factor c on four simple graphs. Edges are colored using a directional colormap (see color wheel in figure: edges going west are red, edges going north are purple, edges going south are green, and edges going east are blue). Luminance is linearly increased along edges, so their direction is also shown by following the dark-to-saturated color gradient, in line with (Holten and van Wijk, 2009b). The top row shows the input graphs, which are simple for illustration



(a) Raw data

(b) $c = \cos\frac{\pi}{12}$

(c) $c = \cos\frac{\pi}{4}$

(d) $c = \cos\frac{\pi}{2}$

(e) $c = \cos\frac{10\pi}{12}$

(f) $c = \cos\pi$

Figure 37.: A few iterations with different values of compatibility coefficient on eye movement multitasking dataset.

purposes. The middle row uses $c = \cos(\pi)$, in which case ADEB produces identical results with KDEEB. As shown by the colors, bundles contain edges running in different directions. This creates bundles which arguably do not reflect well the connectivity pattern in the input graph, see e.g. the second row from left.

The bottom row shows ADEB for $c = \cos(\pi/6)$, which creates bundles containing only edges having similar directions. As argued also by Selassie *et al.*, this is useful to separate a graph into several bundled 'layers', each representing the bundling of a subset of edges having locally similar directions (Selassie et al., 2011). The bottom-right image in Fig. 36 shows this: The input graph (top-right image), which contains edges of alternating directions as indicated by the arrows, has been effectively bundled to yield two similar red and purple bundle structures, one for each input direction (compare this example with Fig. 7 in (Selassie et al., 2011)).

See also Figure 37 that shows few iterations with different values of the compatibility factor c on the same eye-tracking dataset.

6.2.4 *Attribute-based bundling*

Apart from direction, we can easily use other edge attributes to define edge compatibility. Consider for instance that edge e_i has, at point $\mathbf{x}$, an attribute value $t_i(\mathbf{x}) \in \mathbb{R}$. We next transform t_i to a vector $\mathbf{t}_i = (\mathbf{t}_i^x, \mathbf{t}_i^y) \in \mathbb{R}^2$ by using polar coordinates, *i.e.* set

$$\mathbf{t}_i^x(\mathbf{x}) = \cos\left(\frac{t_i(\mathbf{x}) - t_{min}}{t_{max} - t_{min}} \cdot \pi\right), \tag{6}$$

$$\mathbf{t}_i^y(\mathbf{x}) = \sin\left(\frac{t_i(\mathbf{x}) - t_{min}}{t_{max} - t_{min}} \cdot \pi\right) \tag{7}$$

with $t_i \in [t_{min}, t_{max}]$ over the entire G. Next, we use ADEB (Sec. 6.2.3) unchanged. The resulting flow direction map encodes the compatibility of our edge-attribute t_i. The compatibility factor c is to be set accordingly: For instance, if edges are compatible in time when their timestamp difference is less than σ seconds and the maximum timestamp value is t_{max}, then c should be set to $\cos(\frac{\sigma\pi}{t_{max}})$.

6.2.5 *Multi-criteria bundling*

We can use the ADEB framework to bundle graphs using compatibility criteria defined by several attributes. Suppose that we calculated two flow direction maps corresponding to edge direction and edge timestamps respectively. We can next compute two subspaces of compatible directions and compatible timestamps. Next, by computing the gradient $\nabla\rho$ over the *intersection* of these two subspaces, the result given by Eqn. 5 is a multi-criteria bundled layout.

Figure 38 shows this for a dataset recording the gaze for a one-minute exploration of Ilya Repin's painting *Unexpected Visitor*. Similar to Yarbus (Yarbus, 1967), a subject was asked to look at the painting and first assess and verbalize the age of the main characters, and next the age of secondary ones. Using time-based compatibility, ADEB successfully bundles together the trails separated in time (Fig. 38 b, trails colored by time via a blue-green-red colormap). We find that there are two main 'visual scans' in this scenario – first, a scan between the faces of foreground characters A and B (green bundle), followed by a second scan between background characters C and D (red bundle). However, since we don't use direction for bundling, we cannot tell if in these scans the eye moved from left to right, right to left, or both directions. Figure 38 b shows the result of directional bundling, colored by trail directions like in Fig. 36. We now see that both scans A-B and C-D include left-to-right (blue) and right-to-left (red) bundles of similar thickness. Hence, the eye moved in both cases in *both* directions, and it did so in a balanced way (comparable amount of times in both directions). However, close to character C, the ends of the C-D bundles are pulled towards and mixed with the A-B bundles, even though the two scans are temporally separated. Using time-and-direction based compatibility, we can obtain both desirable effects shown before (Fig. 38 d): Each scan is separated in a left-to-right and right-to-left bundle, and temporally different scans are not mixed.

6.2.6 *Implementation details*

We evaluate the two components θ_x and θ_y of the flow direction θ by using Eqn. 3 twice, for the two corresponding components $\mathbf{d}_x$ and $\mathbf{d}_y$ of the edge direction vector $\mathbf{d}$ respectively. We implement this on the GPU by splatting the kernel K at all edge sample points using additive blending, and accumulating the resulting θ_x and θ_y values in two floating-point textures. Next, we compute the density gradient $\nabla_{\Omega_{\mathbf{x},c}}\rho$ by finite differences, on a compatible subspace $\Omega_{x,c} \cap \nu_\varepsilon(\mathbf{x})$, where $\nu_\varepsilon(\mathbf{x})$ is a neighborhood of $\mathbf{x}$ of radius $\varepsilon \geq h > 0$. While covering the neighborhood $\nu_\varepsilon(\mathbf{x})$, each point $\mathbf{y}$ of it is tested for directional compatibility. The finite differences are computed only on points that pass the compatibility test in Eqn. 4. Finally, while doing the bundling iterations that solve Eqn. 5, we relax the user-given compatibility value c so it reaches $\cos\pi$ at the last iteration. This way, the first iterations force edges to aggregate into compatible bundles, while the last iterations 'compact' edges already found to be compatible into smoother final bundles.

Our method has similar complexity and computational performance with KDEEB. Specifically, computing the flow direction map θ takes twice the time required for KDEEB's density map computation. Estimating the directional compatibility criterion adds only a simple logical condition for the density gradient computation. Memory-wise, our method requires two floating-point buffers of the size of the graph drawing G,

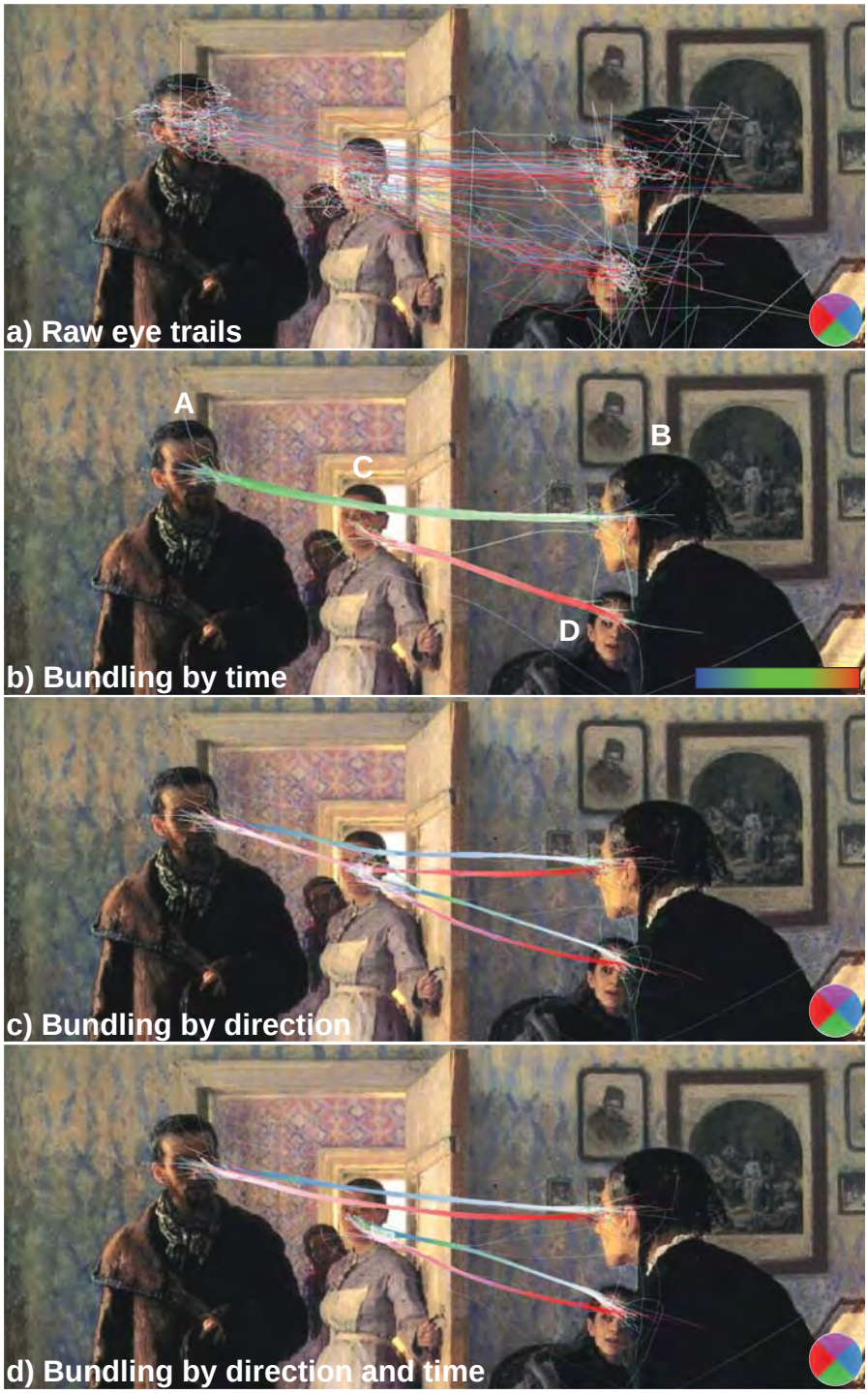
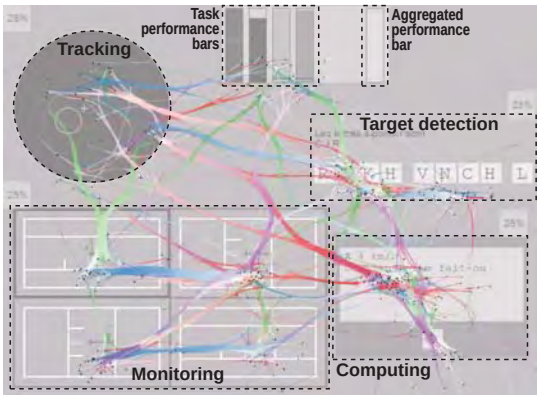


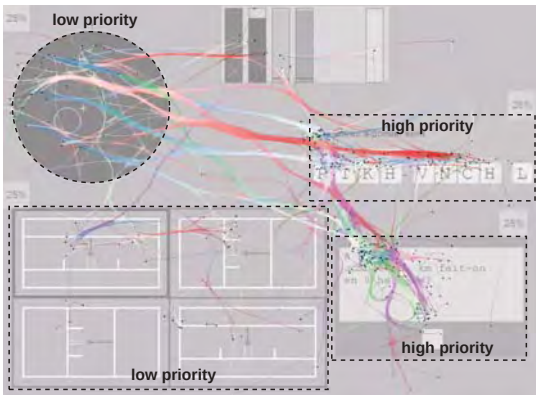
Figure 38.: Multi-criteria bundling of an eye-tracking dataset.

which is a negligible cost increase as compared to KDEEB. Compared to FDEB (Holten and van Wijk, 2009a) and divided edges (Selassie et al., 2011), two other prominent directional bundling techniques, our method is over one order of magnitude faster (see also Sec. 6.4).

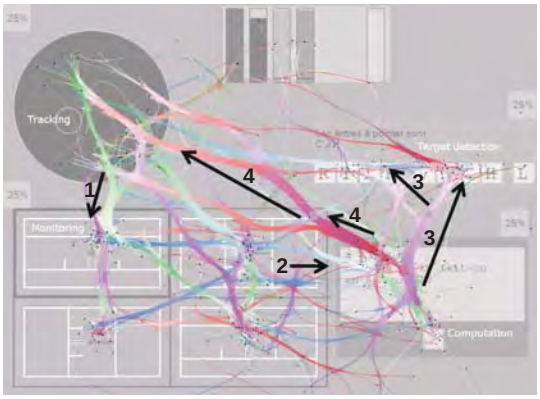
We render the bundled edges as curves colored by the value of one attribute of interest, *e.g.* time or direction. We also scale the curve thickness at each point $\mathbf{x}$ with the local edge-density value $\rho(\mathbf{x})$ (Eqn. 1). Bundles (or bundle fragments) thus appear locally thicker in areas where they contain many edges, thereby enabling users to estimate the importance of the respective connection patterns.



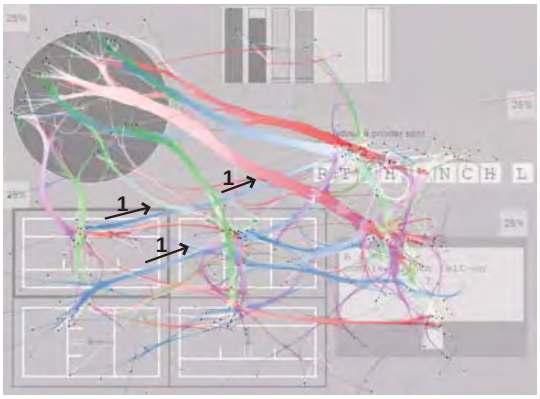
(a) Novice user performance with even priorities



(b) Novice user performance with uneven priorities



(c) Expert performance with even priorities



(d) Expert performance with uneven priorities

Figure 39.: Bundled eye-tracking trails of novice and expert subjects for a multitask experiment done with two priority conditions.

6.3 EYE-TRACKING EXPLORATION USING ADEB

We next present the application of ADEB for the qualitative analysis of trail datasets (here eye-tracking data) represented by directed attributed graphs. First, we outline how we extracted an attributed graph from raw eye-tracking data. Next, we detail two experiments where our bundling proved to be an effective aid for getting insight into the recorded data. We wish to thank Nadine Matton (ENAC), Frédéric Dehais (ISAE) and Sébastien Scannella (ISAE) for provided eye-tracking datasets.

Preprocessing: Raw gaze data, recorded at 50 Hz, was preprocessed to extract fixations and saccades as follows. Consecutive sample points located in a square of 20×20 pixels and separated by at least 200 ms were considered to be a fixation event. Points marked as a one fixation event were replaced by their average. This reduces small-scale noise (scattered close fixation points) generated by micro-saccades and eye-tracking device imprecision during a fixation. Clustering algorithms could be used (Comaniciu and Meer, 2002) to further reduce clutter and aggregate adjacent fixations corresponding to one object of interest but separated spatially due to the imprecision of human visual system. Next, trails formed by consecutive sample points were cut between fixations, yielding the saccades. Finally, a graph was created using fixations as nodes and saccades as curved edges. Note that since human eye does not encode any information during the saccadic movement, curved edges of our graph do not have direct attentional interpretation and thus could be distorted.

6.3.1 *Multitask experiment*

Eye-tracking technology can be used to study repartition of visual attention. Consider, for example, how a pilot watches the cockpit instruments during a certain maneuvre (task). Such tasks, *e.g.* take-off or landing, are typically split into multiple subtasks, each involving several instruments between which the pilot has to allocate attention. For each subtask, instruments have to be watched in appropriate order and with a given frequency. For instance, the priority of the PFD instrument with respect to the FCU will be different for the final landing approach as compared to the cruise phase (Matton et al., 2014).

Data and tasks: This dataset is part of the priority management testing of the pilot selection at French Civil Aviation School (Matton et al., 2014). The subject has to perform a main task containing four concurrent subtasks of different nature (tracking, monitoring, target detection, mental computing), with two priority conditions: (a) all subtasks have even priorities, and (b) two high-priority and two low-priority subtasks (uneven priorities). Tasks are performed using a dashboard-like multitask interface on which several instruments show dynamically-changing data. The interface and location of the

various instruments used for each task is shown in the background of Fig. 39a. The top bar widgets let subjects monitor their performance (four bars for each task, and one aggregated-performance bar). For the uneven condition, subtask priorities are shown in Fig. 39b. A detailed description of this experiment and the multitask design are given in (Matton et al., 2014). Two subject groups, one formed by experienced pilots, and the other one formed by novices, performed the task under both priority conditions, while their gaze was continuously recorded. For each such run, which took 4 minutes, we used the described above preprocessing to obtain a graph having 234 saccades of 24 samples each on average. Using this data, we next try to find similarities and differences in task execution across users and priority conditions.

Results: Figure 39 shows the result of applying directional bundling on four sample test runs, involving both priority conditions and user groups, colored by direction as in earlier figures. Directional bundling allows us to identify the main scanpaths (characteristic sequences of saccades) involved in each run, and also to compare these scanpaths across users and conditions. Data is normalized in the sense that all runs are of the same length (in time and number of sample points), and also contain similar numbers of fixation points. As such, differences between such bundled images reveal both inter-subject and intra-subject similarities and differences.

Let us first analyze the influence of expertise level on the visual attention distribution. The expert has more pronounced transitions between the four subtasks as compared to the novice (Figs. 39a, 39b), as shown by the thicker bundles in Figs. 39c, 39d. Given that the average saccade speed is similar over humans, and the experiment duration is identical for both user types, we deduce that the expert needs less time to perform a subtask and spends more time to switch between the four screen areas to update his knowledge about subtasks. He also largely ignores the performance bar in his 'attentional walk' as compared to a novice. This way of scanning an image, also called *ocular strategy*, leads to better performance (Matton et al., 2014).

The ocular strategy changes when the priorities of the four subtasks are uneven, as shown by the different structure of the bundles in Figs. 39b,39d as compared to Figs. 39a,39c. Yet, the priority condition affects novices and experts differently. The expert shifts his attention to the more important subtasks, as shown by the increased bundle coverage in the right part of Fig. 39d. In contrast, the novice almost abandons the monitoring task and spends most of the time on the two high-priority subtasks. For the uneven priority case, the tracking task induces significant attention for both the expert and novice. This can be due to the fact that, in this experiment, the user's right hand always rests on the joystick that controls the tracking subtask, so they unconsciously dedicate significant attention to this task, even when they are instructed that tracking is a low-priority subtask.

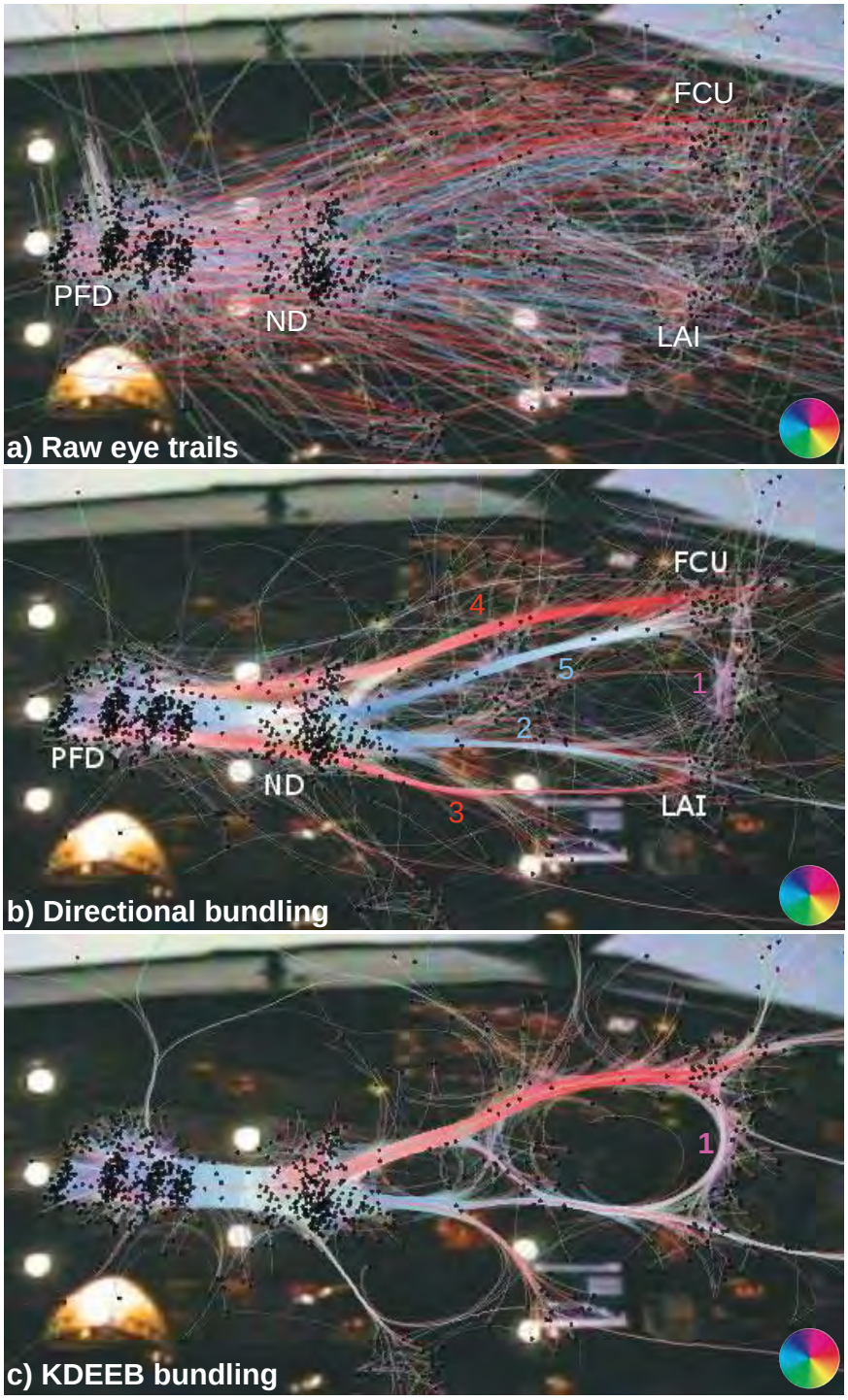


Figure 40.: Raw and bundled eye trails, landing scenario (Sec. 6.3.2).

Directional bundling also shows us the sequence (order) of attendance of different instruments. Consider the expert participant with even priorities (Fig. 39c). The main transitions between subtasks, covering about 80% of the entire set of saccades, are captured by the following thick bundles: tracking to monitoring (green bundle 1), monitoring to computing (two merging blue bundles 2), computing to target detection (purple bundle 3), computing to tracking (red bundle 4). Hence, a special order of subtask monitoring can be deduced. First, we deduce that the expert monitored information in anticlockwise order (tracking, then monitoring, then computing, then tracking again). We also see, for the even priority case, that neither the expert nor the novice do have significant transitions between target detection and monitoring. For the uneven priority case, however, the expert has many transitions from monitoring to target detection (Fig. 39d, blue bundles 1). This shows that the expert remembered that target detection is of high priority, and periodically switches attention to it from the less important monitoring.

6.3.2 *Landing experiment*

Data and tasks: This dataset is part of a study conducted at the French Aerospace Engineering School. During the experiment, a pilot performed a landing manoeuvre in a flight simulator. The experiment lasted 15 minutes. The aim was to test a new cockpit instrument providing landing assistance, which we next call the Landing Aid Instrument (LAI). More specifically, we wanted to understand whether (and how) the new LAI is used along the other instruments. Eye-tracking recording resulted in a graph with 1194 saccades of 8 samples each on average.

Results: Figure 40 shows the raw trails from this experiment, the results of our ADEB bundling, and those of undirected (KDEEB) bundling. Bundles are directionally colored, as in earlier figures. The background image shows the cockpit dashboard. Carefully inspecting the raw trails (Fig. 40 a) reveals several saccades connecting the PFD (left), ND (center), FCU (top-right), and the new LAI (bottom-right). However, the raw trails display is too cluttered. We cannot be sure that we found all relevant instrument-to-instrument trails, and we cannot see high-level scanpath patterns. When using ADEB bundling (Fig. 40 b), the following connections become obvious: faint purple bundle 1 LAI→FCU, blue bundle 2 PFD→LAI passing by ND, red bundle 3 in the opposite direction, red bundle 4 FCU→PFD, and blue bundle 5 in the opposite direction. The bundlePFD–ND is denser than the PFD–FCU one, and much denser than the PFD–LAI one. Thus, the main pattern in the pilot's gaze can be deduced: PFD→ND→FCU→PFD. The LAI is rarely consulted because of its novelty — the pilot is used to land the aircraft without this instrument. In more detail, we see that there is no dense connection FCU→LAI. The pilots explained that, after entering flight parameters on the FCU, they are used to check them immediately on the PFD (as shown by the red bundle 4). Thus, to better integrate the new LAI in this maneuvre, a new training protocol should be de-

signed in which pilots are specifically told when to add the visual consultation of the LAI in their action sequence.

Figure 40 c shows the results of undirected bundling. We see that undirected bundling highlights only coarse-level connection patterns and smooths our finer-grained ones. For example, the red LAI→PFD and FCU→PFD bundles visible with ADEB (Fig. 40 b) are now hard to see, as they appear to end at ND. More problematically, we notice a quite thick bundle between FCU and LAI (Fig. 40 c, bundle 1), which suggests that the subject *did* in fact frequently use the LAI in conjunction with the FCU, which we know is not true. ADEB bundling does not show this bundle, as there are not enough saccades coherent in *both* space and time between FCU and LAI, and thus conveys us the correct insight (LAI rarely used).

6.4 TECHNICAL ASPECTS

We discuss next several technical aspects of our method.

Generality: Most earlier bundling methods use only edge length, absolute angles, and relative distances (the latter which we also use). So far, only (Selassie et al., 2011) explicitly shows directional bundling; (Holten and van Wijk, 2009a) mentions this possibility, but does not actually demonstrate it. Separately, attribute-based bundling is noted as a possibility in (Telea and Ersoy, 2010), but not actually demonstrated. In contrast to all above, our ADEB demonstrates how both local edge direction and edge attributes can be added to earlier geometric compatibility metrics.

Scalability: The most prominent bundling techniques using edge compatibility measures (beyond edge distance) are (Holten and van Wijk, 2009a; Selassie et al., 2011). Both have a complexity of $O(E^2C)$ where E is the number of edges in the input graph, and C the number of edge sampling-points. In contrast, our method, just as KDEEB (Hurter et al., 2012), is $O(C)$. This makes our method considerably more scalable than (Holten and van Wijk, 2009a; Selassie et al., 2011), and of about the same scalability as (Hurter et al., 2012). In detail, FDEB takes 19 seconds on a graph of 2101 edges and 205 nodes (the well-known *US airlines* dataset), with 12 sample-points/edge on average (thus, roughly 25K sample points in total); the method in (Selassie et al., 2011) takes 24 seconds for the same graph, for a sampling of up to 25 points/edge (up to 75K sample points), excluding preprocessing costs which authors note to be 10% of the total cost. For the same graph, KDEEB takes 0.5 seconds, at 86K sample points. ADEB takes 1.3 seconds for the same amount of sampling points. Overall, ADEB is roughly twice as slow as the undirected KDEEB, but about 18 times faster than the directional method in (Selassie et al., 2011).

Interaction: Edge bundling aims to create simplified *overviews* of complex graphs. To solve specific tasks, extra tools such *e.g.* local levels-of-detail (Hurter et al., 2011), attribute-based selection (Hurter et al., 2009), or interactive aggregation (van den Elzen

and van Wijk, 2014), are obviously beneficial. Such tools can be easily added to ADEB. In line with earlier bundling papers (Cui et al., 2008; Ersoy et al., 2011; Holten and van Wijk, 2009a; Hurter et al., 2012; Lambert et al., 2010; Selassie et al., 2011), we do not detail such extensions here, so that our contribution on directional bundling is more focused and easier to separate from such additional tooling mechanisms.

Limitations: Our attribute-based bundling implies that we can map distances between attribute-values (in their original space) to angle distances (Eqn. 7). This is immediate for quantitative and ordinal attributes, but is not evident for categorical attributes. For the latter case, such a mapping would require an algebra definition on the category space. Hence, for categorical attributes, we suggest to cluster trails according to their category and perform separate bundling computations for each such cluster (much like (Telea and Ersoy, 2010)). Separately, we showed the added-value of ADEB *vs* a single undirected bundling method (KDEEB). This leaves the open question whether *other* undirected bundling methods would be comparably better. While such comparisons are yet to be done, we argue that, for tasks that require seeing and analyzing bundles of different directions, all undirected-bundling methods would exhibit similar limitations as KDEEB. Finally, we note that our use-cases of ADEB for analyzing eye-tracking data do not imply that ADEB is the ultimate technique to gain all types of insights from such data, as compared to *e.g.* AOI, dwell-time-per-AOI, or transition matrices. Rather, we employ this use-case to show how ADEB can be of added value as opposed to a raw or undirected-bundling display of trails (Fig. 40).

6.5 CONCLUSION

In this chapter, we proposed ADEB, a method that creates edge bundles from attributed graphs where edge compatibility can be defined by one or several attributes. ADEB is fast, simple to implement, scalable, generic, and could be easily extended to handle dynamic graphs. We demonstrated ADEB, and implicitly the added-value of directional bundling, by showing two applications in the analysis of eye movement recordings. It is the first edge bundling method suitable for eye-tracking visualization because it separated saccades in space, direction and time. It offers a good alternative to heat maps and provides a first intuitive and informative illustration of scan paths. Moreover, the ADEB framework provides mathematical objects — directional maps, i.e. a function $\theta : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ — that can be compared between across subjects and conditions. This technique makes a whole new lane for novel methods of scanpaths comparisons.

7

EYE MOVEMENTS IN FLIGHT SIMULATOR STUDIES: "EXPLORE/EXPLOIT" RATIO

> You could see her thoughts swimming around in her eyes, like fish — some bright, some dark, some fast, quick, some slow and easy, and sometimes, like when she looked up where Earth was, being nothing but color and nothing else.
>
> *The Martian Chronicles* by RAY BRADBURY

7.1 INTRODUCTION

Just like in this passage from The Martian Chronicles by Ray Bradbury, the eye movements can reflect your thoughts, swimming around your eyes. The eye movements vary from one activity to another. Thus, for instance, the duration of fixational movements differs from one fixation to another and varies considerably depending on the goal of the task. In this section we present two experiments where we used eye movements to characterize the behavior of the flight crew[1].

In the first study, pilots experienced an *automation surprise* situation when they did not understand the behavior of the auto-pilot. Eye movements unearth their surprise expressed as intensive visual search aiming to help them to assess the situation.

In the second study, the flight crews were performing a landing exercise when suddenly, an ATC operator asked them to perform a go-around maneuver. We observed the changes in the ocular behavior of the both pilots of flight crews according to their role in task sharing, i.e. PF and PM (see Chapter 1). Eye movements revealed that the PF became more concentrated during the go-around, whereas the PM started to monitor further, as their roles prescribed.

1 In this section, I present only analyses which I have performed or in which I have participated, the both studies being a part of Frédéric Dehais' work. The complete results of the studies are presented in Dehais et al. (2015) and Dehais et al. (2016)

In these two studies, we developed a technique allowing to isolate eye movements that were significantly different across conditions. After studying the number of occurrences of these movements according to different factors, we can aggregate them into a sole "explore/exploit" or "search/processing" ratio. The presented works included two different experimental factorial designs: one-way design in the first experiment and two-way design (2×2) in the second one.

Yet, before presenting the two aviation studies, let me present a small preliminary experiment that gives some insights about fixation durations and different visual tasks.

7.2 VISUAL TASKS AND FIXATION DURATION. PRELIMINARY EXPERIMENT

The main objective of this preliminary study[2] was to compare the duration of fixations during various ocular activities differing in the degree of exploration/exploitation. We hypothesized that fixation would be shorter during exploration compared to exploitation activity.

7.2.1 *Materials and methods*

7.2.1.1 *Subjects*

The subjects were 4 healthy volunteers (3 females, 1 left-handed, age 23.3 ± 2.6), myself and the three students. All had a normal or corrected-to-normal vision, no history of neurological diseases and were free of the regular use of medication.

7.2.1.2 *Experimental design*

Participants performed six different visual tasks. Each task was performed three times.

MAP SEARCH Subjects were asked to find three cities on a map with time limited to 25 seconds. The three stimuli were: a map of France with Dijon, Annecy and Niort to find; a map of Brazil with Belém, Campo Grande and Pelotas to find; and a map of Italy with Bari, Livorno, and Gela to find.

READING Subjects were asked to read a small text. The three stimuli were fables of Jean de La Fontaine: *La Cigale et la Fourmi*, *La Grenouille qui veut se faire plus grosse que le Boeuf*, and *Le Corbeau et le Renard*.

2 This preliminary study was a part of my teaching activity and was performed as a Research Initiation Project at Université Toulouse III — Paul Sabatier with master program students Belazouz Damia, El Youssoufi Kawtar et Mornas Diane.

Figure 41.: An example of a stimulus for the *find differences* task.

LETTER COUNTING Subjects were asked to count all the occurrences of the letter *o* in the text. The stimuli were the same as in the reading task.

FIND DIFFERENCES Subjects were asked to find as many differences as they could find between two images displayed side by side with time limited to 25 seconds. The images were found on the Internet, an example of a stimulus is shown in Figure 41.

FIND WINDOWS Subjects were asked to find as many windows as possible on a Pieter Brueghel painting with time limited to 25 seconds. The three paintings were *Children's Game* (1560), *The Gloomy Day* (1565) and *The Census at Bethlehem* (1566).

FIND RED-ORANGE OBJECTS Subjects were asked to find as many objects that were red or orange as they could on the painting with time limited to 25 seconds. The paintings were the same as in the "find windows" task.

7.2.1.3 *Eye movements recording and processing*

Participants were seated at a viewing distance of approximately 65 cm from the 22-inch LCD monitor (1680 $\times$ 1250 pixels screen resolution). Participants' gaze position was recorded at 500 Hz with a remote SMI RED eye-tracker (SensoMotoric Instruments GmbH, Germany). This device allows tracking eyes with precision despite the absence of a chinrest and small head movements. At the begging of the experiment, a 5-point

calibration was validated with four additional fixation points, until a precision of gaze position was inferior to 1°.

Stimulus display was conducted using SMI Experiment Center software. The fixations were detected using Event Detector of SMI that uses an algorithm with a dispersion threshold. If gaze point was situated for at least 50 ms within the box with a maximum dispersion of 100 px, it was considered as a fixation. The obtained data were imported to Matlab and segregated by activity.

7.2.2 *Results*

We aggregated fixations for each task across three repetitions of each one and calculated the mean values that are given in Table 3. The rows of the table are color-coded with green cells corresponding to smallest values and red cells to the biggest ones.

Figure 42 shows the distributions of fixation durations for the six visuals tasks in percentage. Each bar is averaged across 4 participants, and vertical black lines show standard deviation. The colored bars shows the mean and median duration.

Table 3.: Mean fixation duration per task and participant. The rows are color-coded with green cells corresponding to the smallest value and the red cells corresponding to the biggest value.

Subject No	**Map search**	**Reading**	**Letter Counting**	**Find differences**	**Find windows**	**Find red objects**
1	217.5	195.9	228.8	136.9	226.4	190.3
2	210.8	183.1	217.4	148.3	176.5	182.4
3	247.4	200.2	209.8	165.2	194.1	179.4
4	113.1	119.6	95.4	91.8	82.7	72.0
Mean	197.2	174.7	187.9	135.6	169.9	156.0
Std	58.3	37.5	62.1	31.4	61.7	56.2

7.2.3 *Discussion and conclusion*

The experiment showed that even for a small sample size there were the differences in fixation duration. One group level, the shortest fixations were observed for the *find differences* task. Next, a bit longer fixations for the two visual tasks with Pieter Brueghel paintings — *find windows* and *find red-orange objects*. Finally, the longest fixations were observed for letter counting, reading and map search tasks. The Figure 42 gives us a better understanding of fixation duration distributions. Each bar of this histogram shows the percentage of a fixation of a given length. The more a distribution is on the

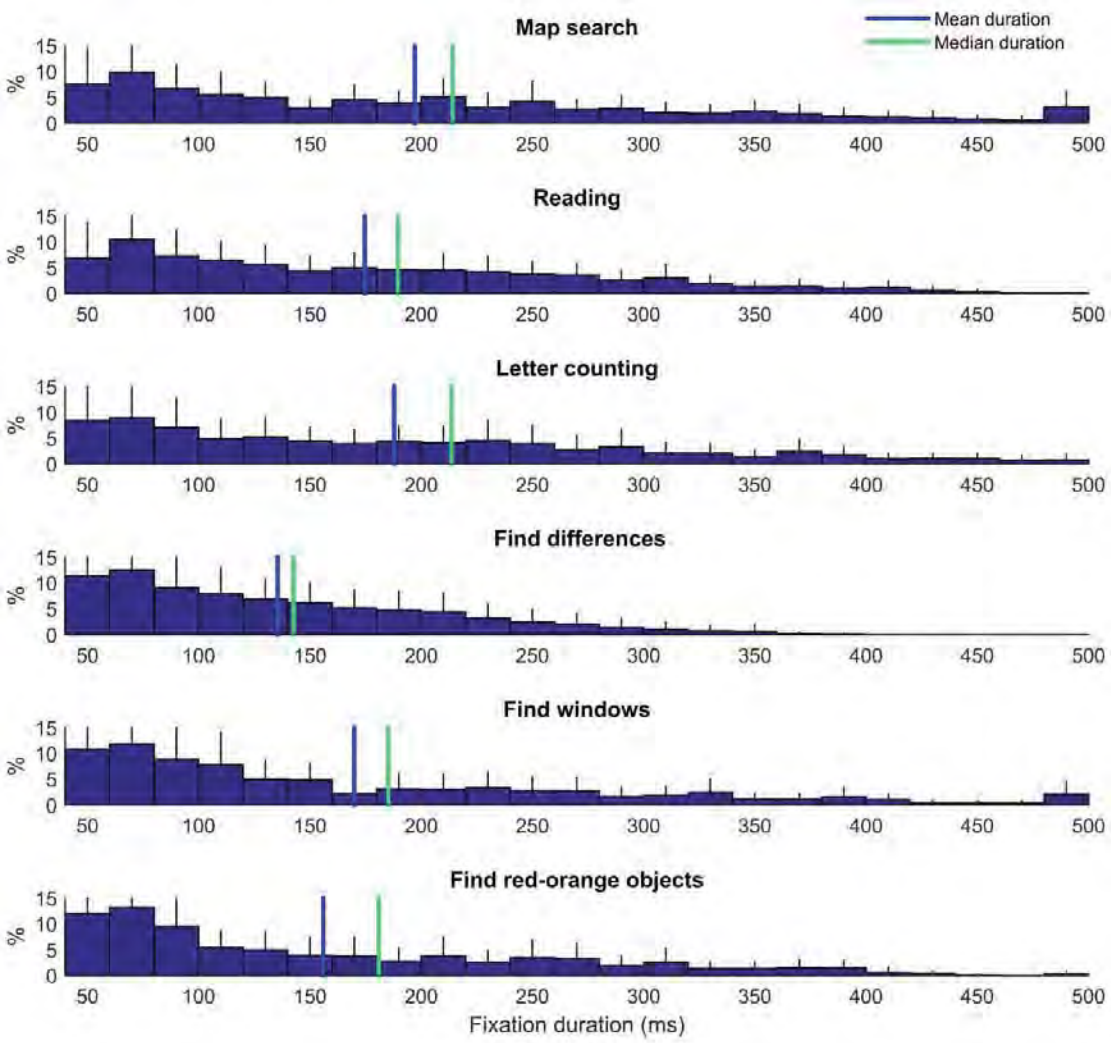


Figure 42.: Distributions of fixation durations for the six visual task. Vertical black lines denote standard deviation.

left, the shorter fixations were on average. Thus, for instance, map search, reading and letter counting tasks have much flatter distributions with many fixations about 200–400 ms. Find differences has the shortest tale and has almost no fixations longer than 300 ms. The two Pieter Brueghel tasks is a mix of two — they have a lot of short fixations and then a flat tail of medium fixations.

The obtained results can be explained by the nature of the tasks and the associated degree of exploration/exploitation. Thus, in the *find differences* tasks, participants have to explore the scene rapidly by making many saccades comparing two pictures. In Pieter Brueghel tasks, subjects have to explore the painting rapidly by also to compare each

fixated object either with a window or to identify whether it is red. Reading needs a comprehension of the text and thus induce much longer fixations. Letter counting must have induced quite long fixations because participants continued to "read" the text without actually analyzing its sense but doing so-called *z-reading*, following the lines as if they were reading. Eventually, the map search task turned out not to be more of reading than visual search. Subjects indeed were reading city names one after another comparing them with the target and were not switching chaotically between the city names.

Note also that all subjects slightly differed in their fixation durations. The most obvious one is the Subject No 4 that had much shorter fixations compared to others[3]. She also had the shortest fixation for the Pieter Brueghel tasks and not the *find differences* one as the three others. This example shows the considerable inter-subject variability of eye movements and visual strategies.

In summary, this introductory experiment highlighted the variability of fixation duration and its relationship with the task nature. When the task implies active visual search, comparison and exploration — the fixations are short. When the visual task requires deep visual processing and information exploitation — the fixations are long. These issues have to be considered when analyzing eye movements under realistic settings such as operating an aircraft.

7.3 EYE MOVEMENTS DURING AN AUTOMATION SURPRISE IN AVIATION

New technologies introduced into the cockpits added new capabilities but also additional cognitive demands on flight crews. Moreover, loss of mode awareness became a significant problem in aviation (Sarter and Woods, 1995). The study case of Chapter 1 is an example of an insufficient mode awareness. Recall, the flight crew of an AF72 flight to Tahiti (BEA, 1999) was unaware of an automatically initiated go-around, and the aircraft finished the nose into the lagoon. However, it is not an isolated case. Whereas in the case of AF72 flight everything went well in the end, in some cases it ended much more dramatically. Thus, the same year on 25th February a Boeing 737-800 of Turkish Airlines crashed during approach near Amsterdam Schiphol Airport (Dutch Safety Board, 2010). Nine people died, 120 more were injured. As the accident report showed, due to high workload while intercepting the glide scope, the crew did not notice an autothrottle mode change. The aircraft ended in a stall situation. As stated in the report: *Shortly after the accident, the initial investigation results indicated that the left radio altimeter system had passed on an erroneous altitude reading of -8 feet to the automatic throttle control system (the autothrottle).* Therefore, the mode change was due to the problems with radio altimeter systems in the Boeing 737-800. However, an appropriate monitoring would allow the pilots to understand and to recover from the dangerous

3 The subject also had eyeglasses that could distort a bit the eye-tracking

situation. As showed a survey of 186 airline pilots, *automation surprise is a relatively common phenomenon which occurs about once every month for the average pilot* (Hurts and de Boer, 2014). Therefore, we insist once again on the fatal importance of monitoring skills for pilots (see Chapter 1). Here we present the experimented conducted in a 3-axis motion flight simulator with 16 pilots equipped with an eye-tracker to analyze their behavior and eye movements during the occurrence of an “automation surprise" situation.

7.3.1 *Materials and methods*

As mentioned before, this experiment was designed and conducted by Frédéric Dehais and colleagues. I owe only eye movements processing, other parts of materials and methods are necessary to understand the protocol, but are taken from Dehais et al. (2015) with some modifications.

7.3.1.1 *Subjects*

Sixteen healthy volunteers (two females; mean age = 30.8, SD = 14; mean flight experience = 1391 hours, range 55–9000; mean automated flight desk experience 200 hours, range 15–700), all pilots from the French Aerospace Engineering School (ISAE-SUPAERO) participated after giving written informed consent. All had good auditory acuity and normal or corrected-to-normal vision according to their pilot medical certificate.

7.3.1.2 *Flight simulator*

The experiment was conducted in the PEGASE flight simulator (Figure 43) situated at the ISAE-SUPAERO. The flight deck of the stimulator is composed of the standard electronic flight instrument displays including the PFD and the ND. The participants had a side-stick, rudder pedals, thrust levers and a FCU to control the flight guidance. The FCU was dedicated to interacting with the auto-pilot via four knobs (speed, heading, altitude, and vertical speed). The auto-pilot (vertical and horizontal profile management) was (dis)engaged via a push button AP on the FCU. Autopilot disengagement was accompanied by an auditory warning (*cavalry charge*). The auto-thrust (speed management) was (dis)engaged via a push button ATHR on the FCU.

For the purpose of the experiment, a realistic auto-flight system was designed that replicated poor automation designs from different civil aircraft (see Pizziol et al. (2014) for a formal description of this autopilot).

- The autopilot had one lateral mode (heading) and three vertical modes (positive/negative/null vertical speed). Note that the vertical speed was null (i.e. 0

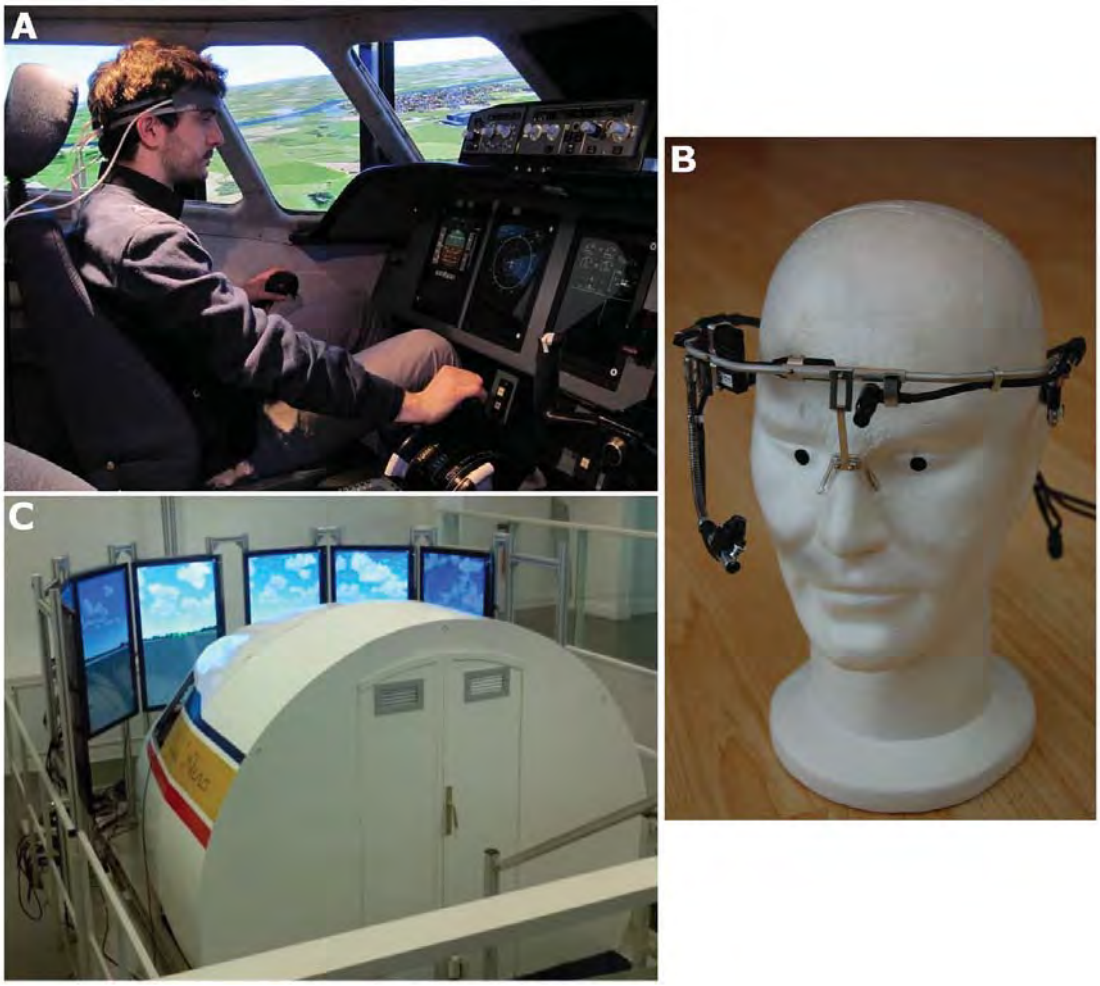


Figure 43.: (a)-(c) Inside and outside views of the PEGASE 3-axis motion flight simulator situated at ISAE-SUPAERO (b) Head-mounted eye tracker Pertech

ft/min) when the autopilot reached the target altitude or when the pilot pushed the vertical speed knob to level off. These different flight modes were displayed on the FMA (the upper part of the PFD);

- The autopilot automatically disengaged during overspeed or low speed/stall events. The overspeed limit value depends on three possible flaps configurations and equals 180 knots, 220 knots or 330 knots. When the aircraft was in overspeed, an auditory warning was triggered (triple chime) that inhibited the autopilot disengagement auditory warning;

- The aircraft leveled off in case of inconsistent programming of the vertical speed with regards to the target altitude (i.e. it was not possible to climb with a negative vertical speed or to descend with a positive vertical speed);
- The autopilot reversed from level off or negative vertical speed to positive vertical speed when the aircraft speed was 5 knots below overspeed. This mode reversion was dedicated to avoiding a possible overspeed.

7.3.1.3 *Experimental scenario*

The scenario, which included the occurrence of one conflict with automation, lasted 10 minutes. The simulated ATC cleared the pilots for take-off from Blagnac airport (Toulouse, France) and the airplane was vectored regularly according to a flight plan that was identical for each participant. At 9000 ft, the ATC required the participant to "*accelerate 325 knots, descend 5000 feet with a -1000 ft/min vertical speed*" to avoid an incoming aircraft flying at the same flight level. As the speed reached 325 knots, the autopilot reversed to positive vertical speed mode (+1000 ft/min) to anticipate potential overspeed (330 knots — see the previous section for the autopilot logic). This situation eventually led the airplane to level off instead of descending as the selected target altitude (i.e. 5000 feet) could not be reached with a positive vertical speed (Figure 44).

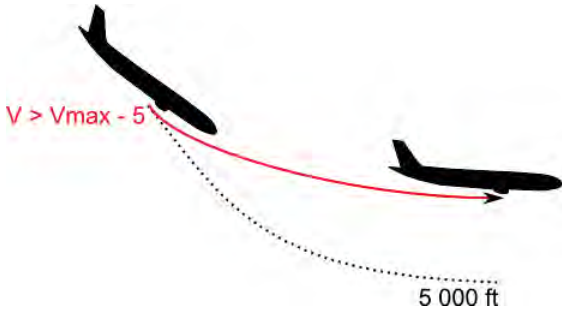


Figure 44.: The "impossible descent" automation surprise: the ATC required the participants to descend but with an excessive speed. The combination of two autopilot behaviors led the aircraft to level off to prevent overspeed.

7.3.1.4 *Procedure*

The participants had a 20-minute tutorial presentation that detailed the flight simulator user interfaces (PFD, FCU, etc.) with a special focus on the different nominal and off-nominal autopilot behaviors. In order to check the understanding of the tutorial, the participants had to comment on each slide of the tutorial and to recall the autopilot behavior. The volunteers completed a one-hour training session in the flight simulator that included basic manual flying, landings, take-offs, stall and overspeed recovery.

They were then trained to interacting with the autopilot and had to set different parameters according to ATC instructions (e.g. *Supaero32, steer 200 degrees, climb 6000 feet, vertical speed +1000 ft/min*). Eventually, off-nominal situations such as autopilot automatic disengagement due to overspeed/stall events, inconsistent FCU programming, level off to vertical speed mode reversion were induced and the participants were told to recover from them. After the training, the participants had to repeat once the interaction with the FCU and the autopilot behavior as well as the different auditory warnings. The experimental scenario was employed after this training.

7.3.1.5 *Eye movements recording and processing*

Participant eye movements were recorded with a head-mounted Pertech eye tracker (Figure 43b). This lightweight non-intrusive device has 0.25° of accuracy and a 50 Hz sampling rate. After a 9-point calibration, a video capturing the visual scene (eye tracker frontal camera) was available with the eye gaze mapped on it. The vertical and horizontal movements in degrees were also computed. We detected ocular events (fixation, saccade) with a dispersion-velocity based algorithm. Parts of recording with gaze moving under 30°/s with a dispersion threshold of 1° were considered as fixations. Other samples (if the eye was not closed) were considered as saccades.

The number of occurrences of these ocular events can be used to characterize the visual processing behaviors. Thus, saccades of small amplitudes characterize local information search whereas those of large amplitudes characterize global information search. As for fixations, we distinguish two types of these ocular events: short ones, the insufficient length of which does not allow the participant to extract and process the associated information and are therefore associated with information search; and long ones, linked to sustained visual and cognitive processing of fixated information. For more detailed discussion on eye fixation durations see Salthouse et al.(1980). As the minimum time one needs to process basic interface information depends on the interface/environment, the exact separating threshold is questionable. Short fixations are rarely considered as information processing in most of the studies where authors prefer to choose a threshold of 200 ms to detect a meaningful fixation (Manor and Gordon, 2003). The distribution of fixation durations is affected by the visual task and the type of the stimulus (see the preliminary experiment), and Manor and Gordon argued (Manor and Gordon, 2003) that a temporal threshold of 100 ms would be more accurate to separate fixations from other oculomotor activity and in some cases fixations from 100 to 200 ms represent about 20-25% of all fixations. In this study, we also considered even smaller fixations, because *most of the visual information necessary for reading can be acquired during the 1st 50 msec that information is available during an eye fixation* (Rayner et al., 1981). And also as our preliminary study showed, fixations can be shorter than 200 ms in simple visual tasks.

To express the balance between search and information processing, Goldberg and Kotval (1999) proposed to aggregate the fixation and saccade rates in a sole ratio in their study on the computer interface evaluation. However, the ratio did not significantly differ between the two interfaces in their study. Regis et al. (2012) in their study on attentional tunneling (i.e. excessive focus) also used the fixation/saccade ratio and did not find either the differences between the different conditions. However, Regis et al. studied the occurrences of fixations of different length. Regis further developed these ideas in his unpublished thesis and proposed a modified ratio where he considered short fixation to serve for exploration. Therefore, we decided to follow this approach and to search for a particular ocular behavior that would characterize the moment of conflict in comparison to a nominal autopilot monitoring sequence. The autopilot mode change (i.e. “level off" to “positive climb") triggered the onset of the conflict. The “baseline" window was chosen as the onset of conflict window, minus two minutes of flight —- that brings it to a nominal autopilot monitoring activity (i.e. basic autopilot supervision and interaction with FCU knobs). We chose a 15-second window of basic piloting activity and compared it to the equivalent 15-second window following the rise of the conflict. We ran non-parametric Wilcoxon tests on fixations and saccades of different lengths (multiples of 20 ms – the inter-sampling period) to determine what events best characterize the conflict occurrence. We selected then all significant events and merged adjacent length values in the same class. We obtained thus three classes: saccades of 120–160 ms duration, short fixations of 80–120 ms duration and long fixations of 240–260 ms. According to the previously mentioned above literature, we argue that the short fixations are not long enough to extract the complex information displayed in the cockpit, and were therefore associated with information search. Keeping these three classes, one can modify the Goldberg and Kotval’s idea and compose a new adapted “explore/exploit" ratio defined as

$$R = \frac{\text{saccades} + \text{short fixations}}{\text{long fixations}}.$$

On one hand, when participants explore the interface, the number of saccades and short fixations increases while the number of long fixations decreases; leading to an increase of the ratio. On the other hand, when participants exploit the interface, long fixations dominate, and the ratio decreases. Figure 45 illustrates the evolution of the ratio on three most representative participants.

7.3.2 *Results*

The debriefing session revealed that 14 out of 16 pilots perceived the abnormal behavior of the autopilot. Among them, we found that only 7 out of 16 pilots understood the situation and initiated the correct recovery procedure within the critical time window

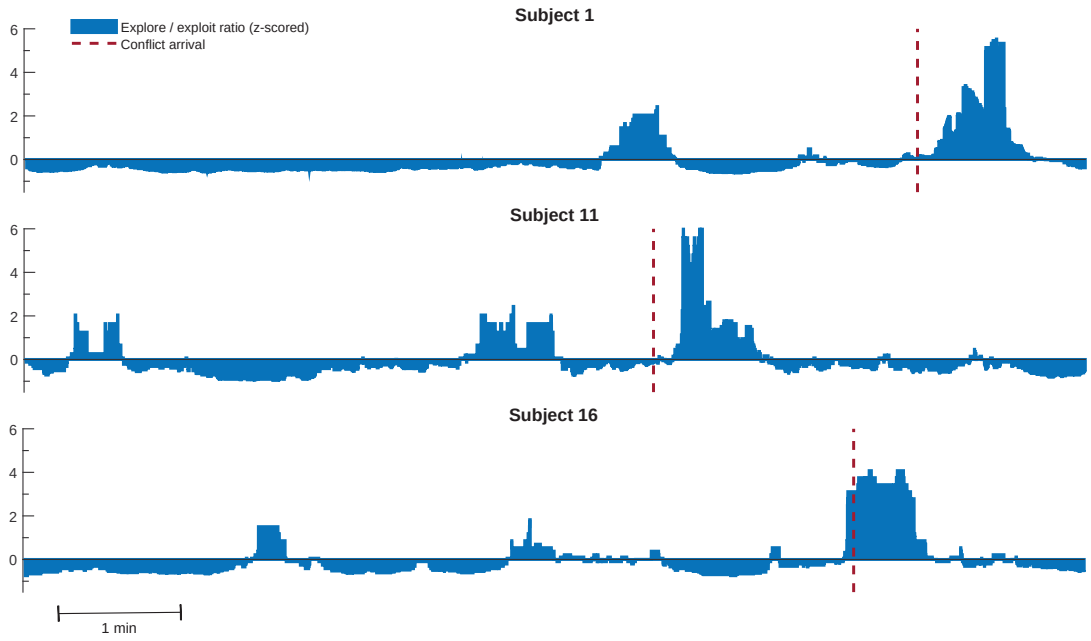


Figure 45.: An example of z-scored explore/exploit ratio for three representative participants. The red line shows the formal arrival of conflict.

of 30 s to avoid the incoming traffic (mean reaction time 18 *s* after the onset of the automation surprise, SD = 9.3). Regarding the Wilcoxon Matched-pairs tests on the ocular events, we found significant effects of the conflict corresponding to an increase of the saccade rate, $p < 0.05; Z^4 = 2.04$, and short fixations, $p < 0.05; Z = 2.36$, and a decrease in the long fixation rate, $p < 0.001; Z = 3.40$; (Figure 46). In addition, the “explore/exploit" ratio revealed a significant effect of the conflict with a higher ratio for the conflict situation comparing to the “baseline", $p < 0.001; Z = 3.29$ confirming that the conflict implies higher exploration activity.

7.3.3 *Discussion and conclusion*

The objective of this experiment was to study the ocular behavior of pilots when faced with a conflict with automation that would provoke “automation surprise". Fourteen pilots out of 16 declared that they perceived a conflicting situation as the automation prevented them from descending to avoid a collision with an incoming aircraft. Whereas conflict solving was “straightforward" (i.e. reducing the selected speed with the dedicated FCU knob), most of the pilots were stuck and failed to deal with the situation

4 Please note that Z is a converted "W" of the Wilcoxon test.

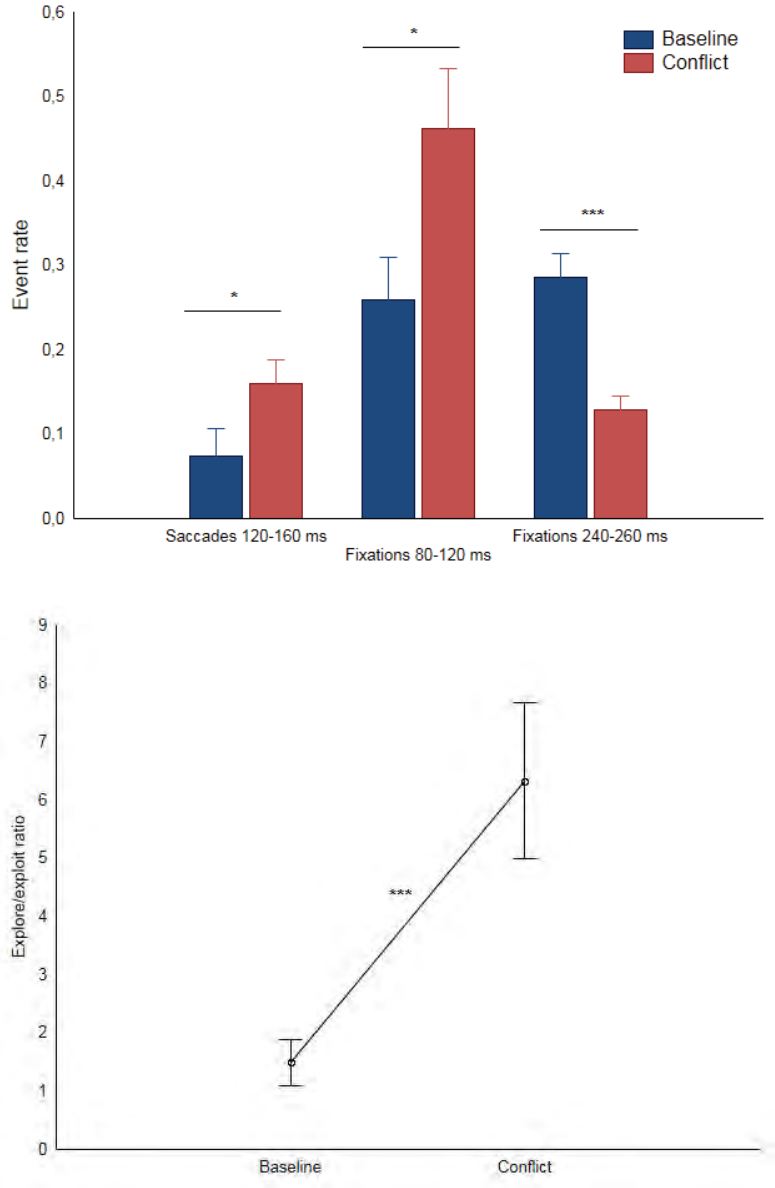


Figure 46.: Group means ± standard errors (N=16) of ocular events rate and the "explore/exploit" ratio for baseline and conflict periods. *:$p < 0.05$; ***:$p < 0.001$.

immediately. Only 7 pilots managed to solve the conflict but with a long mean reaction time (18 s) considering that a collision with another aircraft was possible. Overall, the experiment succeeded to induce the automation surprise when the pilots did not understand the automation behavior for some time and were looking for hints. The analysis

of ocular events revealed that the pilots exhibited higher visual search (more short fixations and saccades) to the detriment of information processing (less long fixations) during conflict compared to the baseline period. It is worth noting that the "explore/-exploit" ratio (i.e search vs. process) increased and was about five times higher during conflict than baseline.

7.4 EYE MOVEMENTS DURING A GO-AROUND MANEUVER IN AVIATION

Final approach and landing are probably the most important phases of flight. Even if an aircraft is exposed to these phases only for 4% of the time for a 1.5-hour flight, these two phases represent represented 48% of fatal accidents from 2005 trough 2014 (Boeing, 2015). When an approach is not stabilized at a defined height, i.e. is not performed with a constant angle and constant rate of descent approach profile, the flight crew should decide to initiate a go-around. In order to better understand why trained pilots sometimes fail to properly execute this safety maneuver, the BEA initiated the Aeroplane State Awareness during Go-Around (ASAGA) study. The ASAGA study consisted of two main components: a survey (questionnaires fulfilled by 831 pilots) and an experiment in a full flight simulator (Boeing-777 and Airbus-330), the data from which are analyzed in this section. Here we present this experiment with 12 flight crews (24 pilots) equipped with two eye-trackers to analyze their behavior and eye movements before and after the go-around initiation.

7.4.1 *Materials and methods*

As mentioned before, this experiment was designed and conducted by Frédéric Dehais, Mickaël Causse and colleagues as a part of the BEA study. I owe only eye movements processing, other parts of materials and methods are necessary to understand the protocol, but are taken from Dehais et al. (2016) with some modifications.

7.4.1.1 *Subjects*

Twenty-four French airline pilots (2 females — 1 PF and 1 PM) participated voluntarily in the experiment (PF's mean age = 49.8 years, SD = 4.9, mean flight experience = 14540 hours, SD = 2573.51; PM's mean age = 41 years, SD = 4.7, mean flight experience = 8743.5 hours, SD = 2733.3). Each of the twelve flight crews was composed of a captain and a first officer. As in airline reality, the flight crews were paired based on their scheduling availability. The captain took the role of the PF and the co-pilot took the role of the PM. All participants were informed about the experiment and gave written consent before participating.

7.4.1.2 *Experimental scenario*

Two different twin-engines simulators were used for the experiment: seven crews flew on a Boeing-777 full flight simulator and five crews flew on an Airbus-330 full flight simulator. Aircraft type was not of fundamental interest in the study. The crews were instructed that the simulator sessions aimed to study the human visual system during a standard flight. Importantly, we did not inform the flight crews about the forthcoming go-around situation. The general scenario was as follows: a flight lasting for about 40 minutes, taking off from Bordeaux and heading to Lyon. No particular meteorological phenomenon occurred. The cruising level was FL 260. As the crew was descending, the ATC asked the pilots to perform a manual night Instrument Landing System (ILS) approach, under radar vectoring, to runway 36 L (flight director “on", no head-up display). The landing ground became perfectly visible when the aircraft descended below a height of 1000 feet. Without advance notice, a go-around was ordered by the ATC, at a height below 200 feet, caused by traffic on the runway. Unlike the published missed approach, that requires the crew to climb in the runway’s axis on a magnetic heading of 350° to an altitude of 5000 feet, the ATC instructed the crew to turn left on a heading of 340° and to climb initially to an altitude of 2500 feet. This change induced a go-around performed at low altitude with a disruption caused by an altitude limitation imposed by ATC. It was expected that this unexpected maneuver combining a simultaneous climb and turn would leave the crew unprepared to adequately face the situation.

7.4.1.3 *Eye movements recording and processing*

The participants’ eye movements were recorded using two synchronized 50 Hz head-mounted eye trackers Pertech (Figure 43). These devices have 0.25° of accuracy and allow natural head movements. To perform analyses of eye movements, we detected fixations and saccades using a dispersion-velocity threshold. To enhance the precision of movements detection and to attenuate the influence of simulator vibrations, that induced slight vibrations of the eye tracker, the vertical and horizontal components of gaze coordinates, as well as estimated gaze velocity, were smoothed using a moving-average filter with a window of size 3 samples. Then samples corresponding to the eyes moving under 30°/s not exceeding 1° in dispersion were considered as fixations. Other samples (except the blinks) were considered as saccadic movements. We computed the number of occurrences of these ocular events per second for each participant for each of two following phases: 1) final approach (1 minute 15 seconds) and 2) go-around (until the stabilization of the aircraft at 2500 ft).

As in previous experiment[5], to select the eye movements that were characteristic of the go-around procedure, we also compared the rates of ocular events of a given length (a multiple of the inter-sampling interval of 20 ms) for both PF and PM between final ap-

5 See the discussion about the fixation duration in section 7.3.1.5

proach and go-around phases. We selected all events that resulted in a Cohen's D value greater than a value of 0.8, considered as a large effect size (Cohen, 1988). Eventually, we defined 4 categories of eye events consisting of at least 3 adjacent event lengths that were merged: 1) short fixations 40–80 ms, 2) medium fixations of 180–220 ms, 3) long fixations 340–500 ms and 4) saccades of 140–260 ms. To express the balance between search and extraction/treatment of information, as some authors proposed (Dehais et al., 2015; Goldberg and Kotval, 1999; Regis et al., 2012), we aggregated the fixation and saccade rates in a sole ratio. Similar to previous experiment (Dehais et al., 2015), we constructed a modified ratio defined as

$$R = \frac{\text{saccades} + \text{short fixations}}{\text{long fixations} + \text{medium fixations}}.$$

It expresses the ratio between the eye movements of more explorative rather than exploitative nature, or between visual search vs information processing. We conducted 5 2×2 ANOVAs to examine the interaction between the pilot type (PF vs. PM) and the flight phase (final approach vs. go-around) on the ratio and on each of 4 eye movement types. Turkey's HSD test was used for posthoc comparisons.

Table 4.: Statistical results on each type of eye movements

	Pilot type factor	**Phase factor**	**Interaction**
Long fixations	$F = 2.71, p = .13,$ $\eta^2 = 0.20$	$F = 0.72, p = .41,$ $\eta^2 = 0.06$	$F = 12.03, p < .01, \eta^2 = 0.52$ More during Go-around for the PF compared to the PM ($HSD < .05$)
Medium fixations	$F = 1.47, p = .25,$ $\eta^2 = 0.12$	$F = 10.08, p < .01,$ $\eta^2 = 0.48$ More during Go-around	$F = 5.54, p < .05, \eta^2 = 0.33$ More during Go-around only for the PF ($HSD < .05$)
Short fixations	$F^* = 14.78, p < .01,$ $\eta^2 = 0.57$ More for PM	$F = 10.27, p < .01,$ $\eta^2 = 0.48$ More during Go-around	$F = 8.19, p < .05, \eta^2 = 0.43$ More during Go-around only for the PM ($HSD < .01$)
Saccades	$F = 7.12, p < .05,$ $\eta^2 = 0.39$ More for PM	$F = 20.7, p < .001,$ $\eta^2 = 0.65$ Less during Go-around	$F = 3.17, p = .10, \eta^2 = 0.22$

* All F stand for $F(1, 11)$

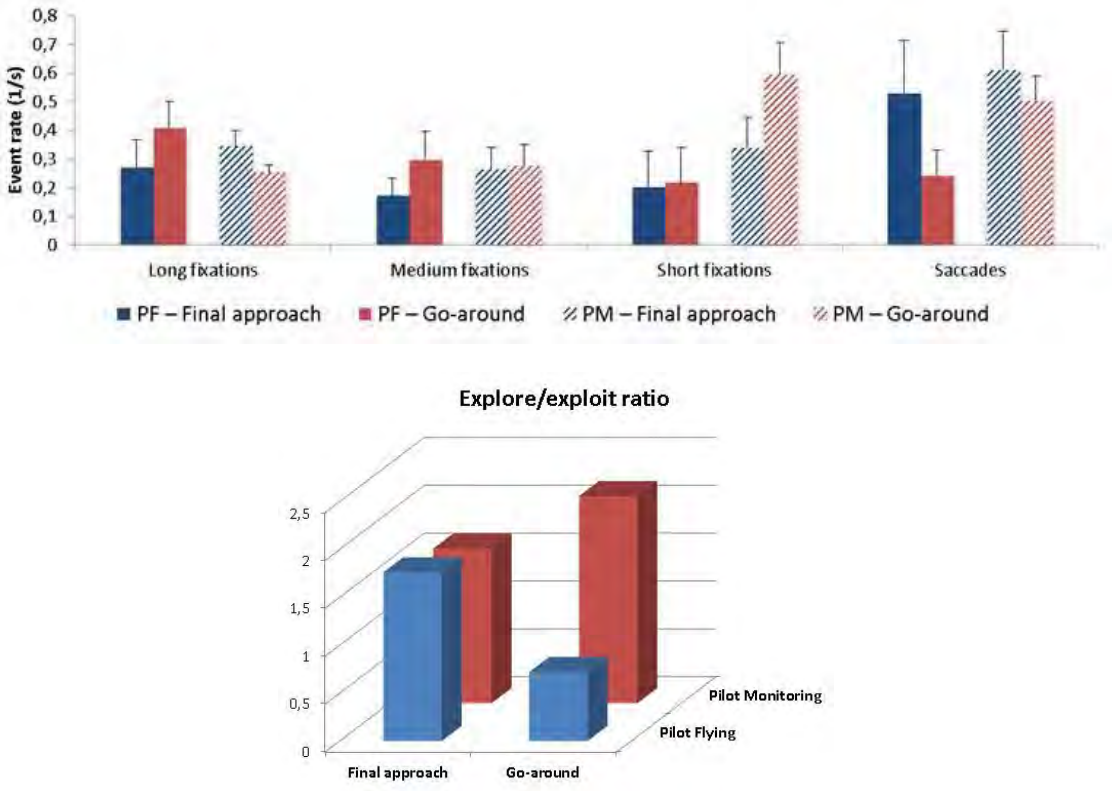


Figure 47.: Group means for ocular events rate and “explore/exploit ratio" for pilot flying and pilot monitoring during final approach and during the go-around.

7.4.2 *Results*

Table 4 resumes the results of ANOVAs performed on basic eye movements used for the “explore/exploit" ratio (Figure 47).

As for the constructed “explore/exploit" ratio, the ANOVA revealed the main effect of the Pilot type, $F(1,11) = 10.09, p < .01, \eta^2 = 0.48$, corresponding higher exploration activity for PM (see Figure 47). The interaction between two factors was significant, $F(1,11) = 18.03, p < .01, \eta^2 = 0.62$. The decrease of the explore/exploit ratio between Final approach and Go-around phases was significant for the PF ($HSD < .05$), that was not true for the increase of ratio for PM. Eventually, the mean ratio of PF during the Go-around procedure, was significantly lower than the ratio for the PM during Final approach ($HSD < .05$) and Go-around ($HSD < .01$) phases.

Figure 48.: Attentional maps during go-around for pilot flying (on the left) and pilot monitoring (on the right).

7.4.3 *Discussion and conclusion*

The objective of this study was to investigate the crew's ocular behavior during the go-around maneuver. Twelve flight crews, equipped with portable eye trackers, were placed in realistic simulators and faced an unexpected go-around scenario. The analysis of the behavioral performance of the participants revealed that the go-around maneuver was initiated by all PFs at the request of ATC. In addition, all flight crews managed to stabilize the aircraft in accordance with ATC clearances.

The analysis of the eye movements aggregated into a sole "explore/exploit" ratio revealed that the PF and PM exhibited different ocular behaviors during the go-around. The PF was mainly in an exploiting/focusing mode whereas the PM was in an exploring/monitoring mode. Such differences are showed in the typical attentional maps in Figure 48. These findings are coherent with the task sharing between the pilots. It shows that both pilots changed their ocular behavior after the initiating of go-around procedure, but according to their respective role.

7.5 GENERAL DISCUSSION AND CONCLUSION

In this chapter, we presented two experiments conducted in full flight simulators with experimented pilots. During the analysis of eye movements, we reflected particularly upon the distribution of fixation durations. Two points are important to retain:

- Fixations vary in duration according to the visuocognitive task and the stimuli
- Short fixations (about 40–120 ms) and long fixations ($>$ 180 ms) should not be mixed together because represent different ocular activities, i.e. visual search and information processing

Therefore, we recommend to all researchers to be extremely careful while analyzing the fixation experimental data. It might be not a good idea to take a mean or median value of all fixations during some period of interest. It should be more meaningful to study the distribution of fixation durations and to report fixational histograms (as in the preliminary experiment). Moreover, throughout the studies in flight simulators (also presented in chronological order in this chapter) we developed the following technique of eye movements analysis:

1. Detect fixations. The possible duration is a multiple of the inter-sampling interval, for instance, 20 ms for 50 Hz recording;
2. Compute the number of fixation occurrences of each length (up to 600 ms or higher). If the sampling frequency is high (500 Hz, for example), regroup the fixations so that the duration would be multiple of 10 or 20;
3. Compute the effect size value (Cohen's D) for each factor;
4. Select all fixation groups that results in a value greater than 0.8 (large effect size);
5. Define categories of eye movements by merging adjacent selected groups (of at least 3 groups to eliminate statistical noise);
6. Perform a statistical analysis on each category of eye movements;
7. Construct a ratio "explore/exploit" or search/processing by putting saccades and short fixations in the numerator and longer fixations in the denominator.
8. Perform a statistical analysis on the ratio

This method should allow constructing a ratio and extract eye movements that differed across conditions. Such study will tell which eye movements were affected by activity and how: were the participants mostly searching for an information of processing it. One can criticize this approach saying that there will be always some differences between conditions. Yet, one should accurately inspect the selected groups of eye movements and the direction in which they vary according to conditions. As in the presented studies, the results should be coherent with behavioral observations and the logic of the experimental scenario. In the both cases, this method resulted in 3-4 groups of eye movements that we assembled in the ratio. In the first study, the ratio increased with the occurrence of the automation surprise. In the second study, while being equal during the final approach, after the go-around initiating, the flight crew had different ocular activity according to their role in task sharing: the PF had a decreased ratio signaling his focus on profile controlling, while the PM had an increased ratio indicating his engine/flaps/FCU monitoring activity.

Note, however, that the delicate point of this approach, as of all studies in eye-tracking, is the event detection. There are many technical and methodological issues concerning the data quality (Holmqvist et al., 2012), calibration procedures (Nyström et al., 2013) and the event detection algorithm itself (Andersson et al., 2016).

8

NEUROERGONOMICS OF THE AIRCRAFT COCKPITS: THE FOUR STAGES OF EYE-TRACKING INTEGRATION TO ENHANCE FLIGHT SAFETY

L'imaginaire est ce qui tend à devenir réel.

ANDRÉ BRETON

The imaginary is what tends to become real.

ANDRÉ BRETON

8.1 INTRODUCTION

In previous Chapter we presented findings concerning the pupil diameter (Chapters 4 and 5) and eye movements (Chapters 6 and 7). These contributions provide additional necessary knowledge and new tools to measure flight crew's behavior. However, even if we know exactly how to measure the pilots' internal state — it is not enough. It is necessary to reflect upon the integration of all these measures in the aviation domain. At what moment do we need to measure exactly something? There is a need for conceptualizing its integration in the cockpit. In the present Chapter, we identify and define a framework of 4 stages of eye-tracking integration in modern cockpits. These four stages are Pilot Training, On-board Gaze Recording, Gaze-Based Flight Deck and Aircraft Adaptations. We support each stage with a description of relevant occurred incidents or accidents and explain in what way the eye-tracking integration can enhance the flight safety. An estimated milestone of the integration of each stage is also proposed accompanied with a list of some implementation limitations. This chapter is under review in the Safety Science journal (Peysakhovich et al., 2016a).

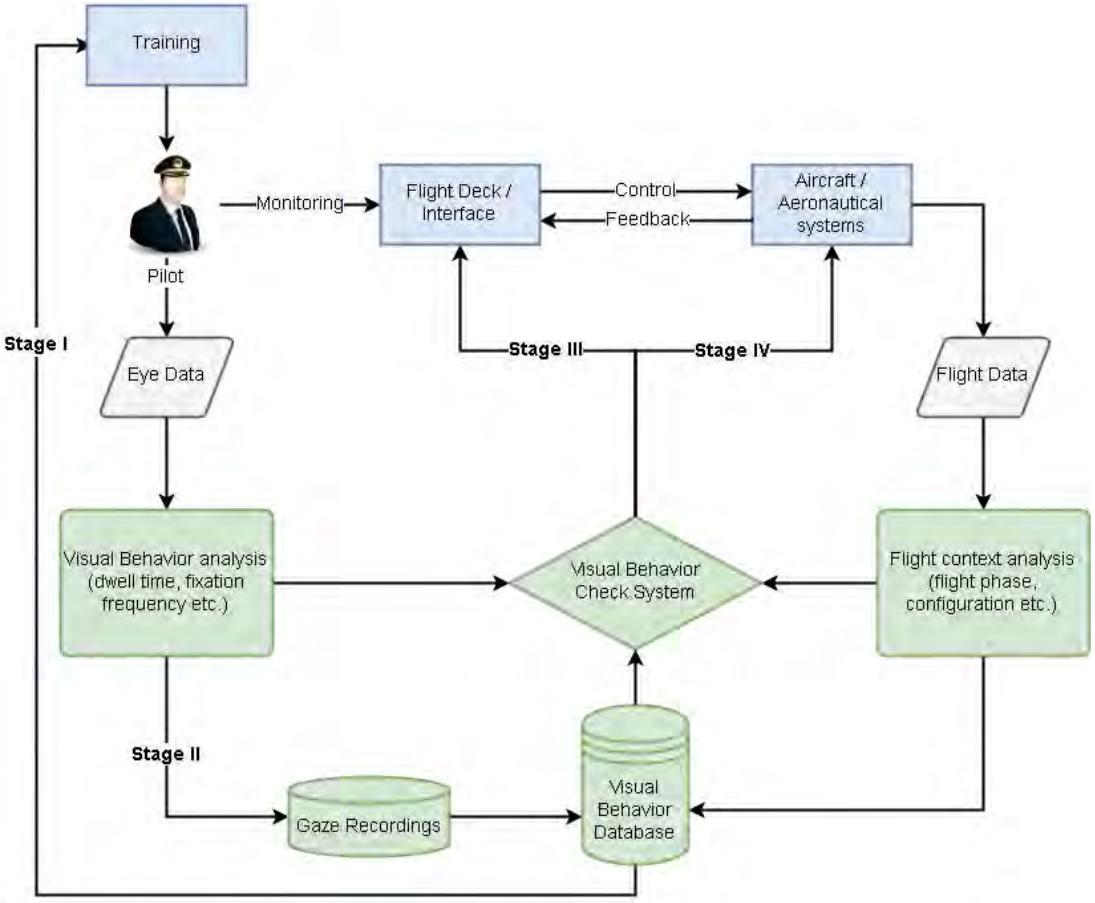


Figure 49.: Flowchart of the eye-tracking integration into the modern cockpits, including all four stages of the framework. The eye and flight data are recorded (Stage II) and proceeded to form a visual behavior database. This database can be used to enhance pilots' training (Stage I) and to check the consistency of the visual behavior according to the flight context. If an inconsistency is detected, we can adapt flight deck (Stage III) or aircraft systems (Stage IV).

8.2 FOUR STAGES OF EYE-TRACKING INTEGRATION

Hereby, we identify and define 4 stages of eye-tracking technology integration in the piloting activity. The first stage relates to pilot training on-ground, whereas the other three stages correspond to operational flight situations. We highlight the interest of the eye-tracking in each of the 4 stages (with examples of aeronautical incidents and accidents for stages III and IV), give an estimated integration date milestone in years, and consider some implementation and use limitations. Figure 50 presents a flowchart of this eye-tracking integration, showing the positioning of each of the four stages.

8.2.1 *Stage I: Pilot Training / Flight Performance Analysis*

In line with the recent training rule published by FAA (FAA, 2013) and that requires airlines to include a specific training program to improve monitoring by March 2019, one important axis of flight safety improvement is pilot training. On this stage I of pilot training and flight performance analysis, eye-tracking data can be exploited in three different ways: as a flight expertise estimate, for incident debriefing, and for example-based learning.

Eye-tracking data can be exploited as an estimate of pilots' monitoring (flight) skills. Gegenfurtner and colleagues (2011) concluded in their review of eye-tracking research in various domains (such as radiology, car driving, sports, chess etc.) that experts (compared to less experienced professionals) have in general shorter fixation durations during the comprehension of visualizations. This echoes with previous Fitts' conclusion about experienced pilots and their fixation durations (Fitts et al., 1950). In the helicopter domain, during a simulated overland navigation task, Sullivan et al. (2011) found an estimate of expertise cost (or rather benefit) on scan management skill. Their model predicted that, on average, for every additional 1000 flight hours the median dwell time will decrease of 28 ms and the number of transitions between zones of interest will increase. In their review of the literature on eye-movements in medicine and chess, Reingold and Sheridan (2011) called it "superior perceptual encoding" of domain-related patterns. But this perceptual advantage is not the only thing gained with the expertise. It is closely related to the global processing advantage and the knowledge of relevant visual patterns. Thus, Gegenfurtner et al. (2011) found, for example, that experts fixate more on task-relevant areas and spend less time to first fixate relevant information. This visual strategy of "information-reduction" (Haider and Frensch, 1999), that helps to optimize the visual information processing by separation of task-relevant from task-irrelevant information, was also found between expert and novice pilots in aviation (Bellenkes et al., 1997; Kasarskis et al., 2001; Schriver et al., 2008). For example, Schriver and colleagues (2008) compared the distribution of visual attention using eye-tracking between experienced and novice pilots during problem diagnosis in a flight simulator. They found, in particular, that faster correct decisions by experts during failure diagnosis were accompanied by correct attending of cues relevant to that failure. Van Meeuwen et al. (2014) compared visual strategies of air traffic controllers with a different level of expertise and found that experts had more effective strategies, and that, furthermore, the scan paths were more similar across expert controllers compared with novices. This suggests that there is an optimal visual scan path for a given visual problem that can be acquired with years of expertise but also can be learned with eye-movement educational examples.

Secondly, the recorded movie clips where the eye gaze position is superimposed on the visual scene could be used in training programs as example-based learning. Such

eye-movement modeling examples were already successfully applied during cooperative problem-solving (Velichkovsky, 1995) and clinical visual observation (Jarodzka et al., 2012; Litchfield et al., 2010). The gaze following is a natural and innate tool of the learning (Flom et al., 2007), including the causal one. It allows the learner to discover what information is relevant and how to visually guide the attention for a given goal by following (expert's) gaze. Quite recently, Jarodzka and colleagues (2013) argued that the attentional guidance by gaze following was not only effective to improve the current performance but also fosters learning. Earlier, Nalanagula and colleagues (2006) have also suggested that the gaze following improves training for visual inspection task. More recently, Leff and colleagues (2015) showed that trainees' performance is enhanced when their visual attention is guided by experts' gaze in surgery context.

Eventually, the movie clips of replayed pilots' eye movements can be used for ulterior analysis and debriefing. Pilots do often request the recordings of flight parameters after a non-routine event in order to better understand what happened during an incident and how to prevent it in the future. It is impossible (and unnecessary) to remember all the eye movements we perform. Moreover, the majority of fixation patterns are automatic and unconscious. Pilots do not say to themselves "*Ok, I'll first check the vertical speed, then a rapid glance at the cap, and then a long fixation outside*", they just do it. Therefore, a replay of the attentional behavior during a flight simulator session, or a real flight (in the case of later integration of eye-tracking systems), would allow pointing out attentional errors of pilots such as an excessive focus on a particular flight parameter or, on the contrary, its disregard.

To conclude this stage I of pilot training / flight performance analysis, because 1) we have enough meta-cognition to analyze our gaze patterns, 2) visual scan paths can be learned by gaze following and 3) monitoring skills can be estimated using eye-tracking data, we reckon that eye-tracking technology is fairly useful for pilots' training on ground. The present stage can be already integrated and adopted by airline companies around the world using relatively inexpensive head-mounted eye trackers available today on the market or by integrating a remote eye tracker system in their training flight simulator.

8.2.2 *Stage II: On-Board Gaze Recording*

The eye-tracking has its potential not only on-ground for pilot training and flight skills analysis, but also in operational settings. The first step of an operational use is a recording of crew's points of gaze to facilitate incident/accident investigations. All modern aircraft are equipped with a black box comprising a cockpit voice recorder and a flight data recorder for recording information required for crash investigation purposes. Pilots do not verbalize everything they do. Therefore, flight parameters and voice recordings are sometimes simply not enough to accurately reconstruct the course of events. Eye-

tracking technology can enhance these devices by also recording the gaze data. First, tracking of pilots' gaze is useful for accident aftermath to verify what flight instruments were looked by the pilots. Secondly, pilots don't look exclusively outside and on the cockpit instrument panel, they also monitor one another's activity. In particular, they observe one another's hands when they press a button, move the engine lever or point at a display (Hutchins and Palen, 1997). Pointing in the airline cockpit is an important part of piloting; it guides own and other's visual attention and helps to establish shared situation awareness (Nevile, 2007). Thus, a recording of pilots' gaze could help investigation agencies to infer the joint awareness and crew coordination. Moreover, these data would be beneficial for cockpit manufacturers, revealing the effectiveness of their user interfaces and yielding to design improvement. Similar benefits for the design of the operational procedures are expected.

However, although technologically feasible, we recognize that the process of introduction of on-board gaze recording might take some time. For example, Transportation Safety Board of Canada indicates in a recent reassessment of the responses to aviation safety recommendation concerning cockpit image-recording systems (Transportation Safety Board of Canada, 2014) that *since 1999, the NTSB has issued 14 recommendations to the FAA related to the installation of cockpit image-recording systems. To all but a single recommendation, the NTSB has assessed the FAA responses as unacceptable.* The most recent safety recommendation of the NTSB (2015) requires notably that each aircraft should be retrofitted with a crash-protected cockpit image recording system that should *be capable of recording, in color, a view of the entire cockpit including each control position and each action (such as display selections or system activations) taken by people in the cockpit.* Let us also note that contrary to the cockpit image-recording systems, this stage does not necessarily interfere with the pilots' privacy. The gaze recordings are impersonal, and should record exclusively a gaze point within an established 3D-model of a given cockpit, or the name of the corresponded area of interest chosen from an exhaustive list of all the flight deck instruments and visual displays. Thus, the eye-tracking recording would not entail any privacy infringement (such as recording of pilots' faces).

Eventually, the cockpit voice recorder should be promoted to human data recorder by also recording gaze data. These data should comprise two channels to keep points of the gaze of both pilot flying and pilot monitoring. Such application is already possible today because it would not have any safety effects and, therefore, does not have to meet any failure probability requirement. An important limitation of this stage is that the recordings should provide an accurate testimony of flight events. Therefore, it is mandatory to have an eye-tracking system of an excellent quality, high precision, and robustness. Such on-board gaze recordings would provide a valuable documentation for safety boards to improve incidents and accidents investigations. Altogether, we speculate that the early adaptation of such gaze recording will emerge within 5 years, the

main requirement being high reliability, precision, and accuracy. If such device would necessitate a calibration procedure, the setup should be quick and simple (for example, during the pre-flight checklist). It is also required that the system takes into account all 360° because pilots' often turn their heads and the cockpits visual environment is large (see Chapter 1).

8.2.3 *Stage III: Gaze-Based Flight Deck Adaptation*

The gaze recording is the simplest but not the only possible use of eye-tracking technology in the operational settings. Numerous study successfully used eye-tracking technique to detect different degraded cognitive states, for example, spatial disorientation in the flight simulator (Cheung and Hofer, 2003), fatigue of pilots (Caldwell et al., 2009; McKinley et al., 2011), attentional tunneling in robot operators (Régis et al., 2014), or automation surprise in the flight simulator (Dehais et al., 2015). Except the detection of undesirable cognitive states, eye-tracking can be also used to figure out one's intentions. For example, Peysakhovich et al. (2015c, see Appendix C) showed that eye movements and pupil size can be predictive of upcoming decision-making in a simulated maritime environment. In the automobile domain, Zhou et al. (2009), proposed a method to infer the truck driver's intent to change the lane based on eye-movements. Ha et al. (2015) showed that it is possible to infer the thoughts of a nuclear power plant operator using his or her eye movements. This lets us imagine an operational support system that assists flight crews using human eye movements and maybe even pupil diameter as an input. Such an enhancement of the flight desk would take the human psychophysiological state into account using the eye-tracking data collected in real-time. An aircraft would thus adapt itself explicitly using the information derived from the crew gaze. Note, nevertheless, that if such support tool would be considered as a part of the aircraft system, safety objectives might be applied. And if such system's failure would have a minor safety effect (minor safety condition may include a slight reduction in functional capabilities and/or a slight increase in crew workload), then the aviation certification imposes at least 10^{-3} allowable failure probability (FAA, 2011). As long as modern video-based eye-tracking systems cannot guarantee even the eye capture for 99.9% of the time (because of luminosity and geometry problems), it will take some time before eye-tracking technology and the corresponding market evolve and achieve sufficient reliability. Moreover, a lot of high-quality research is to be done to properly detect critical and complex cognitive states. However, the eye-tracking can be used in a simpler manner to detect the presence or the absence of an eye fixation on relevant information in the cockpit.

Now, in order to better understand how useful an eye-tracking system can be in operational flight situations, let us consider the following incident. An Air France Boeing 777-200 performing flight AF-471 was about to land at Paris Charles de Gaulle on

November 16th, 2011 (BEA, 2014b). During the approach, at the altitude of 490 ft, while the aircraft was stabilized on the descent path, a Master Warning alarm indicated that the automatic landing mode changed, increasing decision height from 20 to 50 ft. According to the operator's instructions, in this case, a go-around should be called out and initiated, that was done by the PM. Then, according to the protocol, the PF should handle in particular the thrust and the pitch, while the PM should handle the configuration change of flaps. The PF erroneously and unintentionally pushed the auto-throttle disconnection switch instead of Takeoff/Go-around switch that engages the go-around modes. Therefore, his nose-up inputs were contradictory with the auto-pilot system that was trying to keep the airplane on the descent trajectory. Meanwhile, the PM was fully concentrated on monitoring the retraction of flaps, a process which takes about ten seconds. The PM failure to properly monitor aircraft's attitude and energy state has resulted in a brief loss of flight path and late adaptation of the go-around pitch. Hopefully, a relief pilot was on board on the observers' seat and made two deviation callouts of "pitch attitude" that made the pilots apply the nose-up input. In the report's conclusion, BEA underlined that *the serious incident was due to the inadequate monitoring of flight parameters by the flight crew* (BEA, 2014b). Indeed, without the relief pilot on board who called out the pitch attitude, such a behavior in the ground proximity could have resulted in a controlled flight into terrain, the second most recurrent type of accidents (Boeing, 2015).

One of the reasons for temporal loss of control of the flight parameters was the excessive focus of PM resources on the flaps position indicator rather than on aircraft's attitude and engine parameters. Both the lack of pitch monitoring and the excessive focus on flaps configuration could be detected with an eye-tracking system. Thus, the first would be expressed as an absence of fixation on the attitude indicator, while the latter would induce an excessively long fixation on flaps control. Once such an event detected, and, in particular, when normal flight envelope is about to being exceeded, a counter-measure could be applied to the crew. For instance, a brief retraction of the display of flaps configuration on the flight deck could disengage the PM from flaps and to make him allocate some attention to energy state and pitch attitude. A discrete but a visible highlight of these two latter parameters to attract crew's attention is also feasible.

Skippers Aviation de Havilland Dash 8-300 performing a charter flight to Laverton on May 17th, 2012, is another example of an incident with degraded situation awareness as a contributing factor (Australia's Transportation Safety Board, 2013). While conducting a visual circling approach to the runway, a high descent rate triggered the aircraft's warning system. Despite the alarm, the crew continued the approach while the operator's procedures stipulate a go-around initiation in this situation. As noted in the report, *the crew's monitoring of the aircraft's rate of descent and altitude relative to the minimum stabilization height was secondary to their monitoring of their position in relation to the ground. This external focus degraded the crew's overall situation aware-*

ness. Thus, the crew was focused on descending through a break in the cloud and did not respond correctly to the sink rate alert and continued the approach. As in the previous situation, this unstable approach incident could have been avoided by re-engaging the captain and the first officer's attention to the rate of descent and altitude monitoring. The first necessary step to do so is to detect the pilots' excessive fixation on the runway environment. Next, a brief counter-measure could be applied to disengage the captain and the first officer's attention from the outside. This would reorient their attention to the rate of descent and the altitude, essential parameters to determine whether or not the approach is stabilized.

When eye-tracking systems will be reliable enough to capture the crew's gazes in all conditions (different visual angles, head movements, turbulences, luminance), the stage III of gaze-based flight deck adaptation will be relatively easy to implement, at least at some basic level of logic. For example, for each instrument within the cockpit with the knowledge of the system and according to the flight phase (determined by real-time analysis of relevant flight parameters such as position, altitude, speed) one may define: 1) maximum allowed dwell time and 2) maximum allowed time without monitoring. Then, if we detect that one of the rules is broken, a counter-measure could be applied. For instance, if the descent rate was not monitored (with a fixation) for some time by neither pilot flying nor pilot monitoring during the descent phase, an auditory alert may be played back. Alternatively, if we detect an extremely long fixation on flaps configuration, for instance, the information can be retracted for a short period to re-engage pilot's attention. Figure 3 illustrates this simple logic in the form of a flowchart of stage III. Taken all this into account, we speculate that the early adaptation of such gaze-based support system may be within a decade.

8.2.4 *Stage IV: Gaze-Based Aircraft Adaptation*

Eventually, the ultimate and the most far-off (because of immature technology) step of the eye-tracking integration in the cockpit, is the gaze-based aircraft adaptation. In order to palliate humans' limited attentional and cognitive resources, different counter-measures and automation changes can be used and triggered using the pilot's estimated attentional state. This stage can be combined with the previous stage III to provide the crew some feedback about aircraft system changes. On stage IV, we suppose to allow the automation to change aircraft's configuration or to take authority based on the crews' gaze, and not only the visualization enhancement and sonification as on stage III. For example, the automation could temporarily take authority over crew in case of incapacity of the human operator (Inagaki, 2006; Inagaki and Sheridan, 2012; Tessier and Dehais, 2012). This will be, for instance, the case in the next standard version of Dassault Rafale fighters that will automatically perform an auto-pull maneuver if the pilot does not properly react in order to avoid an upcoming controlled flight into terrain.

In this example, however, the automatic maneuver is triggered at the last moment to leave more freedom for the pilot. Yet, eye-tracking can help to anticipate erroneous visual circuits that may precede an important degradation of the situation (Peysakhovich et al., 2015c). However, important safety issues of such pilot-aircraft interaction imply important constraints on the probability of a failure condition. If such a system change would lead to a hull loss and fatal injury of the flight crew, then the superior limit of failure condition (classified as "catastrophic" in this case) probability should be between 10^{-6} and 10^{-9}, depending on the aircraft class (FAA, 2011).

Let us consider the two following examples to figure out how such gaze-based aircraft adaptation could improve flight safety. A Flybe de Havilland Dash 8-400 was performing flight BE-1794 to Exeter on September 11th, 2010 (AAIB, 2012). During the approach, the crew, being distracted by an unfamiliar avionics failure, forget to monitor the flight path and descended below its cleared altitude. As states the AAIB's report, *both pilots became distracted from the primary roles of flying and monitoring the aircraft and did not notice that ALTITUDE SELECT and VERTICAL SPEED modes were no longer engaged.* A terrain proximity warning system alerted the crew more than 700 ft below the selected altitude. After a difficult recovery, the aircraft continued to a safe landing. Another similar incident occurred with the Flybe flight BEE247S on approach to Edinburgh Airport on December 23th, 2008 (AAIB, 2010). The aircraft descended below the glide slope, as the appropriate mode of the flight director was not selected. The flight path deviation was noticed by an approach controller. As contributory factors, AAIB cited *an absence of appropriate monitoring of the flight path and the FMA*. As already mentioned earlier, this lack of flight parameters surveillance can be detected by an eye-tracking system. It would be expressed as an absence of fixations on the FMA, the descent rate, and the altitude for some time. In order to palliate it, within the integration stage III, an auditory verbal alarm can remind the crew of the required visual controls during the approach. In the case of stage IV, the difference is that the automation can take the initiative. For example, in the case of BE-1794 flight, the captain encountered a failure of an Input-Output Processor and lost some of the primary flight display elements. He tried to regain the indications of the primary flight display by switching the air data computer that disengaged the ALTITUDE SELECT and VERTICAL SPEED modes. That did not help to regain the indications, and as the pilot did not verify the FMA status, he was unaware of the disengagement of these modes. One could imagine, that the aircraft re-engages the previously selected modes automatically (maybe with some auditory feedback) without the proper FMA verification neither from the pilot flying nor the pilot monitoring. This is an example of stage IV eye-tracking integration in the cockpit when the aircraft takes over the pilot input if the degraded situational awareness was detected. Taking into account both the insufficient present reliability of eye-tracking system and important consequences of false detection of degraded situational awareness, we speculate that the early adaptation of this stage IV is to expect not

before 20 next years. Table 5 crowns this section by providing the few examples of possible eye-tracking usage within each of four integration stages.

Note on the use of pupil diameter. While discussing the use of eye-tracking data for the modern cockpit integration, the present chapter focused particularly on eye movements and flight crew point of gaze. This is no hazard and there is a reason for such neglect of pupil diameter as a measure. The reason is the reliability of this measure that is simply not high enough to use it with high confidence and change the aircraft attitude according this measure. Also, we found in Chapters 4 and 5 that the cognitive pupillary reaction interacts with light conditions. Therefore, the integration of the pupil diameter as a psycho-physiological proxy on-board requires further investigation and is to be considered very carefully.

Nevertheless, pupil diameter is still a great measure of attentional and cognitive states. Therefor, it can be used on all four stages of eye-tracking integration as a measure of stress and workload. On the stage I it would track whether the levels of stress and workload diminish over years of expertise. On the stage II it would serve as a measure of difficulty of each flight segment. And, eventually, on stage III and IV we could adapt the flight deck or the aircraft if, according the pupil diameter, the flight crew is overloaded or incapable of taking a proper decision.

8.3 CONCLUSION

The aviation domain can benefit from the growing and maturing eye-tracking technology in order to further reduce incidents rate due to human implication. The paper based on this Chapter is the first one of the genre to define and to retrace the framework of the integration of eye-tracking systems into the cockpits consisting of four main stages. These four steps are ordered in the chronological order of its possible implementation, thus retracing the vector of the eye-tracking integration process into the cockpits. We believe that the eye-tracking technology if used wisely, can facilitate and accelerate pilot training, facilitate the investigation of in-flight incidents and considerably enhance the flight safety. Whereas this Chapter reflects upon the aviation domain, we also believe that this work will be useful for practitioners in other domains such as automobile, air traffic control and command and control centers.

Table 5.: Gaze data usage examples within each of four integration stages.

Integration stage	**Eye-tracking data usage**
Stage I: Pilot training / Flight Performance Analysis	• Movie clips with eye-movement modeling for example-based learning • Statistics of dwell percentages on different instruments after a training session • Comparing of scan paths and fixation durations to evaluate the progress of pilot trainees / estimate pilots' skills
Stage II: On-Board Gaze Recording	• Reconstruction of pilots' visual attentional distribution during a flight for incidents/accidents investigation • Analyze of crew's joint attention and shared situational awareness
Stage III: Gaze-Based Flight Deck Adaptation	• Display a notification at the point of pilot's gaze to ensure its visual processing Counter-measure an excessive focus on a parameter (expressed as an extremely long fixation) by retracting the information for a short period of time or highlighting the relevant ones to attract attention • An auditory notification if a critical parameter (according to the flight phase) is left without monitoring for a long period of time
Stage IV: Gaze-Based Aircraft Adaptation	• Turn on an automation mode if the crew considered to be unable to pursue the operation • Perform an automatic manoeuvre

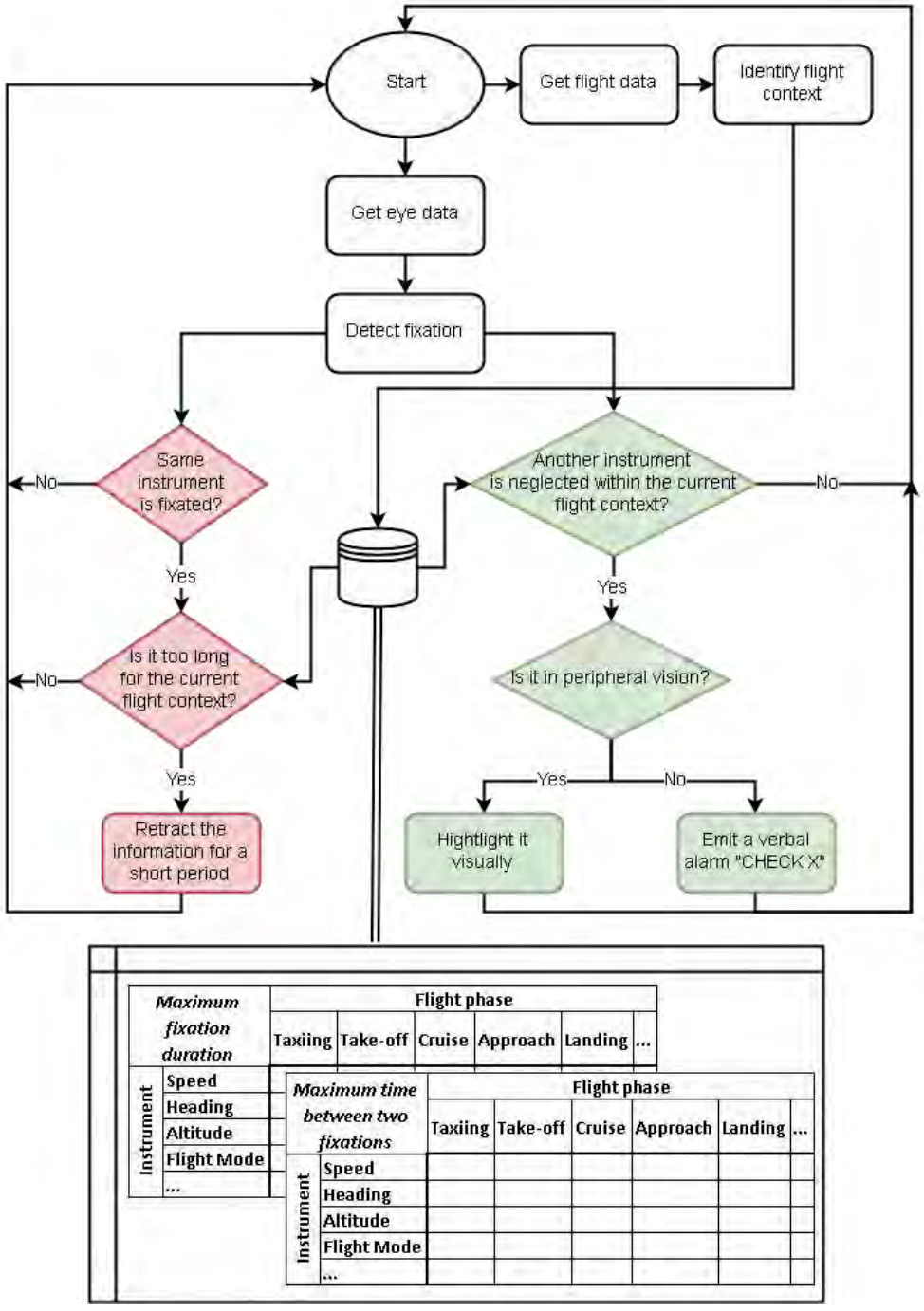


Figure 50.: An example of a possible flowchart of stage III "Gaze-Based Flight Deck Adaptation". The red pathway treats an excessive focus on a particular instrument. The green pathway treats a lack of instruments' monitoring. In this example, the visual behavior database consists of a set of rules concerning maximum allowed values (a fixation duration, a time between two fixations) as a function of both instrument and flight context.

Part III

DISCUSSION AND CONCLUSION. HOW PUPIL DIAMETER AND EYE MOVEMENTS CAN ENHANCE FLIGHT SAFETY

9

DISCUSSION AND CONCLUSION

The goal of this Ph.D. thesis was to investigate how eye-tracking technology can further enhance flight safety. It could seem surprising that cutting-edge domain such as aviation has not yet adopted the eye-tracking technology. However, the complexity of good quality recording and the numerous interpretations of the data explain partly this lack of operational eye-tracking presence in the aviation domain. In research, most eye-tracking studies analyze data a posteriori and off-line, often with some percentage of data loss and uncertainty. In Chapter 8 of the present thesis, we humbly reflect upon the integration of the eye-tracking technology in civil (and military, with some possible modifications) aviation. The reflection led us to a framework of four stages of eye-tracking integration that ordered both by chronological order of supposed integration and technical complexity. This chapter is an attempt of a “well-grounded" reflection minimizing the percentage of science-fiction. However, before discussing this closing chapter, let us review and discuss the contributions of this thesis.

The thesis presents different findings and developed methods and tools regrouped following two logical axes: pupil diameter (Chapters 2, 4 and 5) and eye movements (Chapters 3, 6 and 7). When talking about eye-tracking, one needs indeed to consider both types of measure for the following two reasons. First, the two types of measure are available because the computation of gaze point requires the detection of the pupil (therefore, pupil diameter is available). Second, the two types of measure are complementary when studying the behavior of an individual. Pupil diameter reflects the internal state of an individual (see Chapter 2) and shows whether a participant is, for instance, stressed, cognitively charged, or tired. Eye movements reveal monitoring strategies and are a great complement of behavioral measures (see Chapter 3). Eye movements depend on the nature of a given task and, moreover, one can deduce the task from eye movements. Hence, in the present doctoral work we considered the importance of both pupil diameter and eye movements and investigated both measures, keeping in mind, that the goal of the application is to enhance flight safety.

In the Chapter 1 we showed that the great number of accidents in aviation are due to lack of adequate monitoring. Moreover, cited accidents and incidents have shown that an effective monitoring and decision-making can be hampered by excessive arousal,

high load on working memory or negative emotions. Consequently, we can enhance flight safety and reduce the number of incidents and accidents in two manners. Firstly, we have to monitor the internal state of the flight crew to be aware of potentially hazardous moments. This information is crucial and can be used both for better system design and better operative pilot-aircraft interaction. Secondly, we need to be able to measure and to characterize the monitoring strategy. Naturally, the two contributions on pupil diameter (Chapters 4 and 5) facilitate the use of this psycho-physiological proxy in ecological settings and, thus, are meant to enhance flight safety in a first manner. The two contributions on eye movements (Chapters 6 and 7) provide tools of both qualitative and quantitative characterization of monitoring strategies and, therefore, are meant to enhance flight safety in a second manner.

9.1 PUPIL DIAMETER

Chapter 4 showed that amplitude of phasic pupil response was modulated by the luminance condition, the amplitude being larger under dimmer condition for the same amount of load on memory. This result is an important issue for ecological use of pupillometry. In the aviation domain, for instance, there are numerous ambient light conditions (day/sunset/night) as well as different levels of illumination as a function of flight deck instruments. Also, even when the ambient light conditions do not change, the amount of light illuminating the eye changes depending on the point of gaze. Therefore, it is necessary to take this information into account when elaborating models of pupillary response such as the model of Hoeks and Levelt described in Chapter 2. Thus, we recommend that the transfer function $h(t)$ include some coefficient depending on light conditions that would equalize to extent of pupillary reactions for the same amount of cognitive load. This coefficient should be estimated in an experiment similar to that presented in the original paper by Hoeks and Levelt (1993).

While Chapter 4 investigated the influence of luminance conditions using transient cognitive load paradigm, Chapter 5 explored the relationship of luminance and pupillary response using sustained cognitive load paradigm. It is indeed important to distinguish these two kinds of cognitive load. Sustained load (for instance, manually maintaining the descent slope) persist in time, while transient load (for instance, remember an ATC instruction) lasts only for a few seconds. The experiment described in Chapter 5 demonstrated that it is tonic pupil diameter that is impacted by the load on memory during sustained load. The differences between two levels of load on memory were larger under dimmer luminance condition. Thus, to summarize the findings of the two experiments, we can postulate that the same amount of cognitive load under dimmer luminance condition would elicit a) larger tonic pupil diameter in a sustained load paradigm and b) larger phasic pupil response in a transient load paradigm. Moreover, in Chapter 5 we found an inverse relationship between tonic pupil diameter and phasic pupil response

(see also Annex B), that can be partly explained by the relationship between pupil size and LC-NE system.

These pupillary findings show that the model of pupillary response proposed by Hoeks and Levelt should be revised for taking into account 1) the relationship between light conditions and pupillary response 2) the relationship between tonic pupil diameter and phasic pupil response. However, further investigation is necessary to estimate additional coefficients for the model. Hoeks and Levelt (1993) used simple attentional pulses (reaction to brief auditory stimuli) to estimate the model parameters. However, a more elaborate protocol is needed to include both sustained and transient load concepts in the model. The novel cognitive paradigm described in Chapter 5 allows simultaneous manipulation of both sustained and transient components of the cognitive load. It can be used in the future for further investigation of the complex relationship between the pupil tonic diameter and phasic pupil response and also for estimation of model parameters.

These results on the impact of luminance on tonic and phasic components of pupillary response underline the importance of taking the luminance into account in ecological conditions when using pupillometry to estimate pilots' internal state. However, using temporal models (corrected for luminance effects, of course) is not the only way to estimate the cognitive effort based on the pupillary signal. The frequency analysis also turned out to be an efficient tool. Chapter 4 showed that the pupillary power spectrum was impacted by both luminance and load on memory factors. However, the two factors manifested themselves differently. Thus, while the luminance impacted both high and low-frequency components (defined in Chapter 4), the load on memory impacted only low-frequency band. Thus, an LF/HF ratio, constructed as a ratio of mean power within LF and HF band, turned out to significantly differ between load on memory conditions but not luminance conditions. Therefore, the presented pupillary LF/HF ratio could potentially be an efficient objective measure of mental effort based on pupil diameter that does not depend on luminance conditions. This finding was published in Peysakhovich et al. (2015a). However, it requires further investigation. The pupillary data from the second experiment described in Chapter 5 should be also analyzed in future in the frequency domain to validate or extend the previous findings concerning this luminance-independent measure.

9.2 EYE MOVEMENTS

While pupil diameter is a good indicator of pilots' internal state, eye movements reveal visual strategies. Eye movements are similar to behavioral measures in the sense that they represent observable physical activity. The question of whether some zone of interest was looked at or not is indeed quite similar to the one of whether some button was pressed or not. In Chapters 6 and 7 we focused on two most weak, in our opinion,

points of eye movements analysis — scanpath analysis (Chapter 6) and fixation duration analysis (Chapter 7).

In Chapter 6 we presented a novel method of gaze visualization for qualitative exploration of eye data. The presented technique is a part of a so-called edge bundling techniques when compatible edges of a graph are grouped into bundles. For example, two lines close one to each other will be bundled together to form a thicker line etc. However, the existed bundling methods were unable to handle eye-tracking data because they did not take into account the directional information that is essential in gaze analysis. The ADEB framework is a mathematical framework that renders possible the handling of attribute graphs (where, in particular, the edge direction is defined). It is, of course, a general edge bundling method, but, in particular, suitable for eye-tracking data. Taking into account both spatial (density and direction) and temporal information, ADEB generates graphical representations of eye movements that are as intuitive for scanpaths comparisons as heat maps for fixation analysis. Larger bundles represent denser connections between parts of interface and color code allows to understand the direction of these bundles. The ADEB method needs a fine tuning (like all edge bundling methods) and further investigations on best parameters for eye-tracking applications.

Using the pictures generated by ADEB, human factor specialists should be able in bare seconds to get an idea of the visual strategy used by an operator (pilot or air traffic controller). Moreover, ADEB is not just a fancy tool for graph visualization. The mathematical framework behind (the directional maps to represent direction or any other numerical attribute such as a timestamp or some weight) opens a lane to novel methods of quantitative comparison of scanpaths. As one can compare two heat maps of two different participants or two different conditions, we can compare these directional maps between subjects or conditions. Although it is more complicated computationally (direction map is a vector field while heat map is a scalar field), there is a large number of comparison possibilities to investigate in further research.

In chapter 7 we focused on another aspect of eye movements: fixation duration. First, we verified that fixation duration depends indeed on the nature of the task. As results showed, tasks, implying more exploration than exploitation, induced shorter fixations on average. This observation means that it is important to separate short fixations from the longer ones and never mix them together. Therefore, we proposed a method of fixation duration analysis. Different fixation durations are compared for each possible value (a multiple of the inter-sampling interval). Then, groups of fixations of different duration are assembled according to the effect size value (Cohen's D) between different conditions. The result groups are mixed in an "explore/exploit" or "search-processing" ratio by putting saccades and short fixations in the numerator and longer fixations in the denominator. This technique was tested in two different experiments in full flight simulators. In the first study, we showed that the technique resulted in a ratio that reveals an automation surprise during an approach. In the second study, an aggregated

ratio was indicative of respective roles of flight crew members – PF and PM – and the changes in their ocular strategies after a go-around initiation. Altogether, the Chapter 7 underlined the importance of a thorough investigation of fixational temporal distribution and proposed a novel method of eye movements analysis.

9.3 EYE-TRACKING INTEGRATION

The manuscript is crowned by Chapter 8 that presents the developed framework of four stages of eye-tracking integration in modern cockpits. This chapter traces the future of eye-tracking in the aeronautical domain in our view. The stage I include the pilot training programs where eye-tracking can give insights on training progress and flight performance analysis. Stage II requires onboard gaze recording that can be used for both pedagogical and accident/incident investigation purposes. The stages III and IV takes into account the estimated state of the pilot to make changes (if necessary) of the flight deck (stage III) or the aircraft behavior (stage IV). Fixation duration distribution, scanpaths, and pupil diameter changes all provide necessary information about pilots' visual strategies and internal state. This information can be used on all four stages either to analyze eye data or provide input for the adaptive cockpit system. Therefore, Chapters 4, 5, 6 and 7 contribute to flight safety by providing necessary findings and tools for the efficient use of eye-tracking information in aviation.

9.4 CONCLUSION

In this Ph.D. thesis, we considered eye-tracking as a way to enhance the flight safety. Throughout the chapters, we investigated why considering eye-tracking is useful to enhance flight safety and reduce the number of accidents/incidents in aviation. We focused at the same time on the two axes of eye-tracking data, i.e. pupil diameter and eye movements. We pushed forward the knowledge on cognitive pupil reaction and its components and their interaction with light conditions, useful for pupillometry application in ecological conditions. We designed novel methods of eye movement analysis, both qualitative and quantitative. All these findings together will be used to enhance flight safety by integration eye-tracking into modern cockpits according to the developed framework of four chronological stages. The results presented in this manuscript promote eye-tracking application in aviation and should ease the process of integration into the modern cockpits. Moreover, each of the findings has an important application in its own field (cognitive pupillometry, graph visualization, eye-movement analysis) and are useful for researchers and practitioners beyond the domain of aviation.

Part IV

APPENDIX

A

SCIENTIFIC PRODUCTION

A.1 LIST OF PUBLICATIONS USED IN THE MANUSCRIPT

- **Peysakhovich, V.**, Causse, M., Scannella, S., & Dehais, F. (2015). Frequency analysis of a task-evoked pupillary response: Luminance-independent measure of mental effort. *International Journal of Psychophysiology*, 97(1), 30-37.

- **Peysakhovich, V.**, Vachon F., & Dehais, F. The impact of luminance on tonic and phasic pupillary responses to sustained cognitive load. Accepted in *International Journal of Psychophysiology*

- **Peysakhovich, V.**, Hurter, C., & Telea, A. (2015, April). Attribute-driven edge bundling for general graphs with applications in trail analysis. In 2015 IEEE Pacific Visualization Symposium (PacificVis) (pp. 39-46). IEEE.

- **Peysakhovich, V.**, Lefrançois, O., Dehais, F., & Causse, M. Neuroergonomics of the aircraft cockpits: the four stages of eye tracking integration to enhance flight safety. Under review in *Safety Science*

- Dehais, F., **Peysakhovich, V.**, Scannella, S., Fongue, J., & Gateau, T. (2015, April). Automation surprise in aviation: Real-time solutions. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 2525-2534). ACM.

- Dehais, F., Behrend, J., **Peysakhovich, V.**, Causse, M., & Wickens, C. State Awareness during Go-Around Execution in Aviation: A Behavioral and Eye-Tracking Study. Submitted to *International Journal of Aviation Psychology*

A.2 LIST OF OTHER PUBLICATIONS

- **Peysakhovich, V.**, Vachon, F., Vallières, B. R., Dehais, F., & Tremblay, S. (2015, September). Pupil Dilation and Eye Movements Can Reveal Upcoming Choice

in Dynamic Decision-Making. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 59, No. 1, pp. 210-214). SAGE Publications.

- Mandrick, K., **Peysakhovich, V.**, Rémy, F., Lepron, E., & Causse, M. (2016). Neural and psychophysiological correlates of human performance under stress and high mental workload. *Biological psychology*, 121, 62-73.

- Imbert, J. P., Hurter, C., **Peysakhovich, V.**, Blättler, C., Dehais, F., & Camachon, C. (2015). Design Requirements to Integrate Eye Trackers in Simulation Environments: Aeronautical Use Case. In Intelligent Decision Technologies (pp. 231-241). Springer International Publishing.

- Causse, M., **Peysakhovich, V.**, & Fabre, E. F. (2016). High Working Memory Load Impairs Language Processing during a Simulated Piloting Task: An ERP and Pupillometry Study. *Frontiers in human neuroscience*, 10.

- Causse, M., **Peysakhovich, V.**, & Mandrick, K. (2017). Eliciting Sustained Mental Effort Using the Toulouse N-Back Task: Prefrontal Cortex and Pupillary Responses. In Advances in Neuroergonomics and Cognitive Engineering (pp. 185-193). Springer International Publishing.

- Valéry, B., **Peysakhovich, V.**, & Causse, M. (2015). Hear me flying! Does visual impairment improve auditory display usability during a simulated flight?. Procedia Manufacturing, 3, 5237-5244.

- **Peysakhovich, V.**, Dehais, F., & Causse, M. (2015). Pupil Diameter as a Measure of Cognitive Load during Auditory-visual Interference in a Simple Piloting Task. Procedia Manufacturing, 3, 5199-5205.

- Causse, M., Fabre, E., Giraudet, L., Gonzalez, M., & **Peysakhovich, V.** (2015). EEG/ERP as a measure of mental workload in a simple piloting task. Procedia Manufacturing, 3, 5230-5236.

B

TONIC AND PHASIC PUPILLARY RESPONSES TO COGNITIVE LOAD

In this chapter we present a part of a study on the physiological cost of mental effort and stress in the prefrontal cortex that was published in Biological Psychology (Mandrick et al., 2016). Here we aimed to investigate how a novel mathematical N-back task affects tonic and phasic pupillary responses.

B.1 MATERIALS AND METHODS

B.1.1 *Subjects*

Fourteen healthy individuals (5 women, age 25.8 $\pm$ 3.8 years) volunteered to participate in this experiment. A total of 13 subjects were included for pupillary analyses (one subject was discarded due to technical issues during the recording). All were students at ENAC (French National Civil Aviation School) and ISAE-SUPAERO (French Aerospace Engineering School) in Toulouse, France. None was under medication that might affect the brain or autonomic functions. All volunteers reported normal or corrected-to-normal vision. All participants gave written informed consent in accordance with local ethical board committee.

B.1.2 *Experimental design and procedure*

Participants performed the Toulouse N-back Task described in details in Chapter 5. In this experiment there were three levels of difficulty. In each trial, volunteers were required to compute the result and compare it with either a fixed number (0-back) or the result they obtained 1 (1-back) or 2 (2-back) trials before. There was also a rest period during which “00+00" trivial operations were presented and did not required any response. All operations were displayed in the center of a gray background.

The experiment included 30 blocks in total, with 10 blocks of 0-, 1-, and 2-back tasks presented in a pseudo-randomized order. One N-back block lasted 36 s, and blocks were interleaved with 18 s rest periods. Stimuli were presented for 2 s with an inter-

stimulus interval of 1 s. Each block contained 12 stimuli, with 4 targets in random positions. During some blocks auditory stressors were used to induce anxiety. However, the analysis of the stress effect is out of the scope of this Chapter. In this Chapter we analyze only blocks during which no sounds were presented and participants were aware of it. Therefor, we focus exclusively on the working memory load factor. For such analyses, 4 blocks for each N-back task difficulty (0-, 1-, and 2-back) were available.

B.1.3 *Pupillary recording and processing*

Participants were seated at approximately 70 cm from a 22" computer screen (1680 $\times$ 1250). Ambient luminance was of 10 lx. During the whole experiment, participants' gaze position and pupil diameter were tracked using a remote SMI RED500 eye-tracker (SensoMotoric Instruments GmbH, Germany) at a sampling rate of 120 Hz. This device allows tracking the pupil despite small head movements. Before each run, participants performed a 5-point calibration procedure validated with 4 additional fixation points. The data acquisition routine used iViewX SDK to communicate with Matlab software.

Identified blinks and short periods of signal loss were linearly interpolated. Then the signal was filtered with a two-pass 9-point filter (low-pass with a cutoff frequency of 5.9 Hz). The pupillary recordings were segregated in 3-s segments corresponding to each trial. For all subjects, an average pupillary response was obtained for each condition. A trial was excluded if the total blinking time exceeded 50% (1.5 s) or median gaze position during the trial deviated from the screen center of more than 150 px. Median values calculated on the 500-ms period preceding the trial onset were used as baselines. For statistical analyzes of phasic pupil response, we used the maximal pupil diameter value in the interval between 1 and 2 s post-stimulus. For tonic pupil diameter analyzes, we used the median pupil diameter value for each block.

B.1.4 *Statistical analyzes*

Statistical analyzes were performed using Statistica 10 (StatSoft, USA) software. Descriptive data were presented as a mean $\pm$ standard error. The significance level for all statistical tests was set at 0.05. Four one-way repeated-measures ANOVA were performed on mean reaction time (RT), accuracy rate, tonic pupil diameter, and pupil phasic response with working memory load (0-back vs. 1-back vs. 2-back) as the within-subject factor. Tukey's Honestly Significant Difference (HSD) was used for post hoc testing. Partial eta-squared (partial η^2) was reported to demonstrate the effect size in ANOVA tests when the effect was significant.

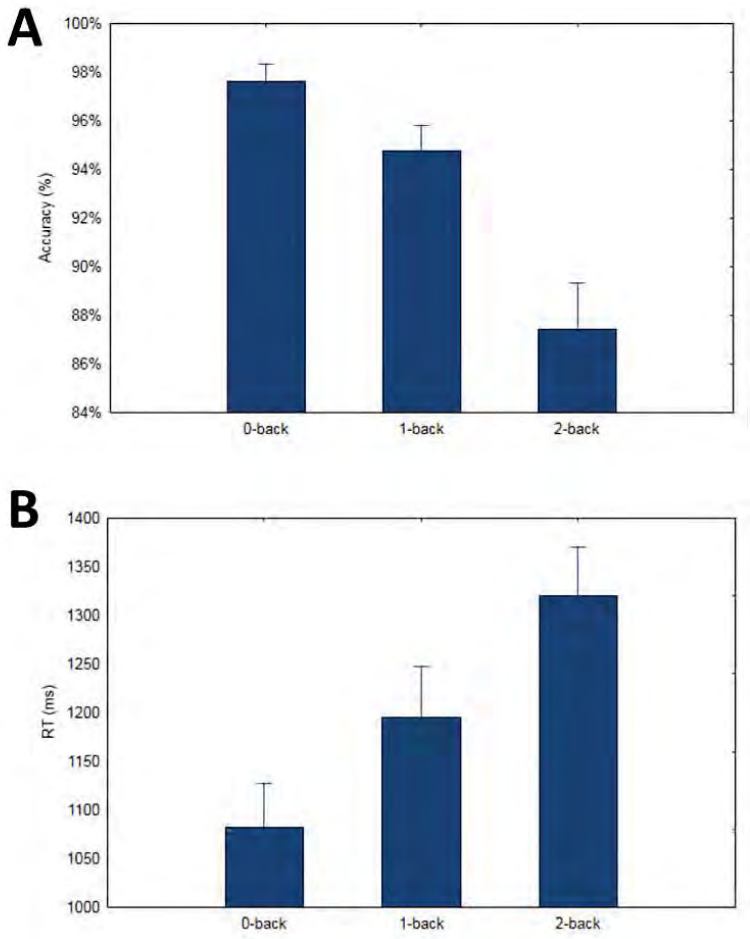


Figure 51.: Mean accuracy and reaction times for different levels of working memory load. Error bars denote the standard error of the mean.

B.2 RESULTS

B.2.1 *Behavioral results*

The working memory load had a significant effect on accuracy, $F(1,12) = 22.3, p < .001, \eta^2 = 0.65$. Post-hoc showed that 2-back significantly decreased the accuracy compared to the other two levels ($HSD < .001$), see Figure 51A.

The working memory load had a significant effect on reaction time, $F(1,12) = 19.4, p < .001, \eta^2 = 0.62$. Post-hoc showed that 2-back significantly increased the reaction time compared to 1-back ($HSD < .01$) and to 0-back ($HSD < .001$). Also, the reaction times in 1-back was longer compared to 0-back ($HSD < .05$), see Figure 51B.

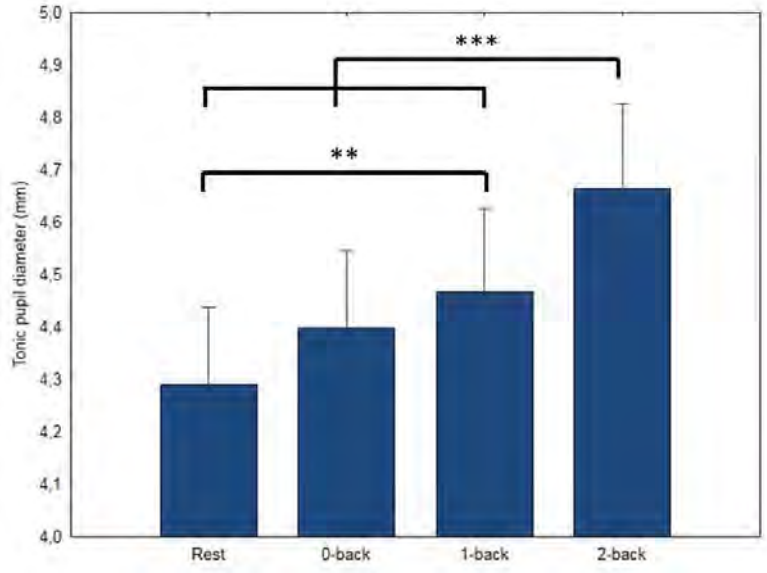


Figure 52.: Mean tonic pupil diameter for different levels of working memory load. Error bars denote the standard error of the mean.

B.2.2 *Tonic pupil diameter*

The working memory load had a significant effect on tonic pupil diameter, $F(1,12) = 22.6, p < .001, \eta^2 = 0.65$. Post-hoc showed that 2-back elicited significantly higher tonic pupil diameter than other three levels ($HSD < .001$ for all), Figure 52. Also, 1-back elicited significantly higher tonic pupil diameter compared to the rest period ($HSD < .01$).

B.2.3 *Phasic pupil response*

The working memory load had a significant effect on phasic pupil response, $F(1,12) = 12.8, p < .001, \eta^2 = 0.52$. Post-hoc showed that there was a significant difference between the rest period and the three levels of N-back ($HSD < .01$), and also that the 0-back was different from the 2-back ($HSD < .05$).

B.3 DISCUSSION AND CONCLUSION

In this experiment we manipulated the level of working memory load and studies its influence on both tonic and phasic pupillary responses. As showed by the behavioral results (i.e. accuracy rate and reaction times), the difficulty effectively increased from 0-back to 2-back. The tonic pupil diameter followed the increasing of the N-back task

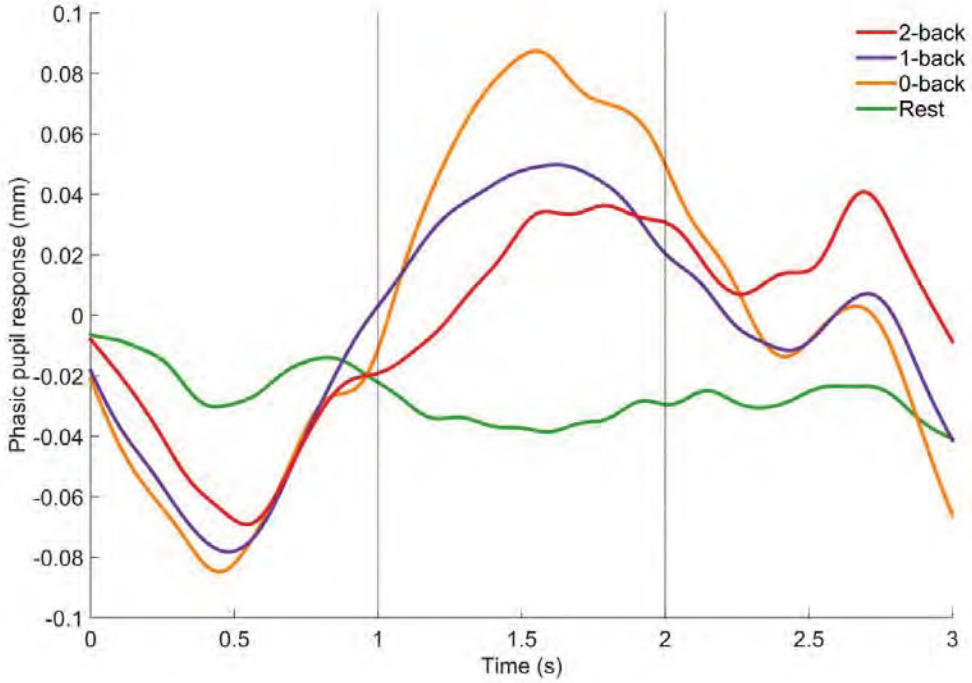


Figure 53.: Average phasic pupil response for different levels of working memory load. Error bars denote the standard error of the mean.

difficulty. It was also larger during any task (0-, 1-, and 2-back) compared to the resting period.T he phasic pupillary response also followed the changes in N-back task difficulty. Greatest for the easiest level of difficulty, pupillary reaction diminished with increasing difficulty. It is almost non-existing for the resting period where neither computational effort nor working memory were employed. This pattern of high tonic diameter corresponding to lower computation-evoked pupillary response is in line with our previous findings in Chapter 5. The present study confirms the inverted pattern of larger tonic pupil diameter corresponding to lower phasic pupil response in a cognitive task inducing working memory load. Please, consult the Chapter 5 for a complete discussion on this relationship.

C

PUPIL DILATION AND EYE MOVEMENTS CAN REVEAL UPCOMING CHOICE IN DYNAMIC DECISION-MAKING

In dynamic environments such as air-traffic control, emergency response and security surveillance, there are severe constraints to information processing and decision-making. Human operators must constantly monitor, assess, and integrate incoming information in order to make optimal decisions in such complex environments. In order to maximize operators' performance, there is a need for effective technological support for dynamic decision-making. Eye tracking is one promising avenue that can provide online, non-obtrusive indices of cognitive functioning. Using a simulated maritime decision-making environment, we evaluated whether oculometry may be exploited to foretell the decision made by the operator beforehand. Our results showed that pupil dilation and fixation transitions can reveal the upcoming judgment of the human operator by about half a second before the decision. This finding can be useful to design adaptive support tools for dynamic decision-making by integrating the operator's cognitive state.

C.1 INTRODUCTION

Dynamic Decision-Making (DDM) involves a sequence of multiple interdependent decisions made in real-time in a continuously evolving environment. In dynamic environments such as air-traffic control, emergency response and security surveillance, there are severe constraints to information processing and DDM. Situation uncertainty, information overload, multitasking, time pressure and fatigue may all impose a heavy demand on cognition and impair decision-making (Gonzalez, 2005). Human operators in this context must perform several tasks simultaneously, each one with a different priority. It is well known, however, that humans are cognitively bounded, insofar as human mental capacities are fundamentally limited. Consequently, allocating more resources to one task will inevitably limit the amount of resources available for processing other information and lead to erroneous DDM (Wickens, 2002). There is a need to understand how human operators process data on the user interface to make decisions.

The study of eye tracking data could serve to enhance the design of user interfaces but also the implementation of adaptive systems (Wilson and Russell, 2007). These

systems derive cognitive performance from different psychophysiological data to optimize human-system interaction by reallocating tasks or applying cognitive countermeasures (Dehais et al., 2015). Following these perspectives, eye tracking is a promising measurement as it provides behavioral metrics as well as psychophysiological ones. For instance, several behavioral ocular metrics have been shown to reflect cognitive activity, such as the number of fixations and their duration, the scanpath direction and length, or the switching rate between areas of interest (see Jacob and Karn, 2003, for a review). Pupil diameter is also a well-documented measure of arousal and cognitive load (Andreassi, 2000; Beatty and Lucero-Wagoner, 2000). Modulated through LC-NE system (Gilzenrat et al., 2010), the iris aperture tracks information processing and attention dynamics according to the function of LC-NE system (Nieuwenhuis et al., 2005). Thus, an eye tracking technique appears to be an interesting means to link cognitive effort and DDM. Furthermore, these devices are nowadays affordable and non-obtrusive (remote without direct contact with human).

The objective of the present study is to evaluate whether oculometry may be exploited to inform an abstract Decision-Support System (DSS) about the decision made by the operator before the actual choice is actioned. For this goal, we used the Simulated Combat Control System (S-CCS) microworld, an emulation of the essential subset of cognitive activities per-formed by tactical coordinators aboard Canadian navy frigates. In this simulation, participants monitor a radar screen representing the airspace around the ship and must assess air-craft threat level and immediacy and take appropriate defensive measures. This microworld has been used as a means to test DSSs (Tremblay et al., 2012), to investigate the sources of change blindness (Vachon et al., 2012), and to assess how auditory distraction (Hodgetts et al., 2014) and task interruption (Hodgetts et al., 2015) may impair DDM. In the present study, we evaluated different eye tracking metrics to reveal upcoming judgments made during the threat-evaluation task. We focused our work on the pupil dilation — shown to be positively correlated to the quality of DDM (Vachon and Tremblay, 2014) and to measure the norepinephrine release during consolidating behavioral decisions (Aston-Jones and Cohen, 2005) — and the number of transitions — linked to information search efficiency (Goldberg and Kotval, 1999). The objective is to understand how the upcoming choice affects these metrics during DDM, and whether or not it is possible to foretell the decision before an actual judgment is made. This approach is the first necessary step towards the implementation of intelligent DSSs.

Figure 54.: Screenshot of S-CCS visual interface divided into three areas: 1) a list providing information on a number of parameters about the selected aircraft; 2) a radar display depicting in real-time all aircraft moving at various speeds and trajectories around the ship (central point) and 3) a set of classification buttons allowing the participants to allocate threat level and immediacy to an aircraft and to engage with missile fire a hostile candidate.

C.2 METHOD

C.2.1 *Participants*

Twenty-one students from Université Laval (11 men; mean age 22.5 years) took part in the experiment. All participants reported normal or corrected-to-normal vision.

C.2.2 *Experimental Design and Procedure*

S-CCS is a low-level computer-controlled simulation of a naval air-defense environment. Multiple aircraft, potentially dangerous, moving in the vicinity of the ship, required the participant to be vigilant and to make decisions based on the dynamic continuously evolving environment. Fig. 54 illustrates the various parts of the visual interface.

Participants assessed the level of threat posed by an aircraft by classifying it as non-hostile, uncertain, or hostile. They had to take into account 5 of the 11 parameters

displayed in the list (Origin, Altitude, IFF, Weapons, and Emissions), none of which were intrinsically more important than the others. Each relevant parameter could take either a threatening or a non-threatening value. Participants were asked to employ the following classification rule based on the number of threatening cues to classify aircraft appearing on their radar: an aircraft was “non-hostile" when it showed 0-1 threatening cue, “uncertain" when it manifested 2-3 threatening cues, and “hostile" if it exhibited 4-5 threatening cues. When a decision was made, participants had to click on the corresponding classification button. The white dot representing the selected aircraft changed color according to the level of threat assigned to it: green (non-hostile), yellow (uncertain), or red (hostile). Given that threat level could change over time, participants had to check regularly the parameters of previously classified aircraft to determine whether they needed reassessment.

Following a tutorial describing the context of the simulation and the task, participants undertook the threat-assessment task from static screenshots to verify their understanding. Familiarity with the simulation was established in two training sessions, each comprised of four 3-min scenarios. After calibrating the Tobii T1750 eye tracker (sampling rate of 50 Hz), participants performed four randomized experimental blocks comprised of four 4-min scenarios of similar difficulty. Each scenario involved 27 aircraft (11 “non-hostile", 8 “uncertain", 8 “hostile") varying in speed and trajectory. A maximum of 10 aircraft could appear on the radar screen at the same time.

C.2.3 *Data Processing*

The raw pupillary data was low-pass filtered (ten order Butterworth, passband: 0–4 Hz). As a baseline procedure, mean pupil diameter was subtracted from the signal for each scenario. Pupillary recordings were divided into decision cycles starting with an aircraft selection on the radar screen and ending with a mouse click on a classification button. All trials were then segregated into three groups according to the selected threat level, yielding 215$\pm$42 “non-hostile", 161$\pm$21 “uncertain", and 133$\pm$5 “hostile" classifications per parti-cipant, on average. The mean task-evoked pupillary response (TEPR) per participant was also calculated for each condition.

Fixation events were detected using a dispersion-based algorithm with a temporal threshold of 100 ms and spatial threshold of 30 pixels. We identified fixations on relevant parameters and calculated the average number of transitions between these parameters (a saccade starting at one parameter and ending at another) per condition for every participant.

C.2.4 *Statistical Analyses*

To examine the effect of threat evaluation on ocular metrics (TEPR and average number of transitions across relevant parameters), one-way repeated measures ANOVA with the factor threat evaluation (non-hostile vs. uncertain vs. hostile) was performed. As these metrics are known to depend on time, we also performed ANOVA on classification time as dependent variable. The significant results were further examined using Tukey's HSD post hoc tests.

C.3 RESULTS

C.3.1 *Classification Time*

The ANOVA revealed a significant main effect of threat evaluation on classification time, $F(2,40) = 44.18, p < .001, \eta^2 = 0.69$. As revealed by the post-hoc tests, participants were faster to classify an aircraft as "hostile" (2349±123 ms) compared to "uncertain" (3047±147 ms) and "non-hostile" (2925±160 ms) aircraft ($HSD < .001$).

C.3.2 *TEPR*

The impact of threat evaluation on pupillary response is shown in Fig. 55. The ANOVA showed a significant main effect of threat evaluation on TEPR, $F(2,40) = 40.96, p < .001, \eta^2 = 0.67$. Post-hoc tests revealed increased pupil diameter for "hostile" classifications compared to "uncertain" and "nonhostile" classifications ($HSD < .001$) but no difference between "non-hostile" and "uncertain" classifications.

C.3.3 *Transitions between Relevant Parameters*

The effect of threat evaluation on the number of transitions between relevant parameters is shown in Fig. 56. The ANOVA showed a significant main effect of threat evaluation, $F(2,40) = 20.24, p < .001, \eta^2 = 0.50$. Post-hoc tests revealed fewer transitions for "hostile" classifications compared to "un-certain" and "non-hostile" classifications ($HSD < .001$) but no difference between "non-hostile" and "uncertain" decisions.

C.3.4 *Temporal Evolution of TEPR and Number of Transitions*

To validate the results described above and related to the whole decision cycles we examined the relationship of TEPRs and also the number of transitions with the time course of decision cycles. We computed the average curves from the first fixation on relevant

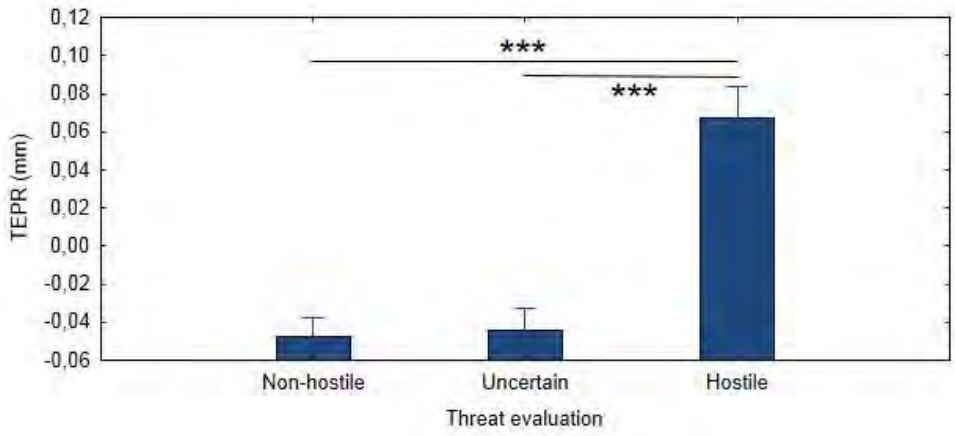


Figure 55.: Average TERP as a function of threat evaluation and post-hoc comparisons. Error bars indicate S.E.M.

parameters for the trials of at least 1,500 ms after the first fixation. This threshold is sufficiently high to observe the TEPR (Andreassi, 2000) and sufficiently low to consider most of the trials (93±4% of “non-hostile", 92±3% of “uncertain", and 86±8% of “hostile" classifications). We chose the first fixation as a time origin in order to observe the pupillary reaction on the background of constant luminosity. Fig. 57 and 5 depict the temporal evolution of TEPR and number of transitions correspondingly. It is noteworthy, that the smoothness of the curve on Fig. 58, is the result of averaging of many step lines (an increment at each transition).

To quantify the moment when upcoming choice resulted in significantly different values of metrics, we computed Cohen’s d value for each time point for “uncertain" and “hostile" classifications. We fixed the threshold on subjective level of 0.6 to select the separation point, which is between the moderate and large size effects (Cohen, 1988). As showed the Cohen’s d comparison, the difference started to be pronounced at 1,060 ms for TEPR and at 1,320 for number of transitions.

In order to better understand the slight difference between “hostile" and the other two assessments at time origin (see Fig. 57), an additional ANOVA was performed on Hook-to-Check time — the time interval between an aircraft selection and first fixation on relevant parameters.

C.3.5 *Hook-to-Check Time*

The effect of threat evaluation on the Hook-to-Check time is presented in Fig. 59. The ANOVA showed a significant main effect of threat evaluation, $F(2,40) = 12.83, p < .001, \eta^2 = 0.39$. Post-hoc tests revealed smaller delay before first fixation for “hostile"

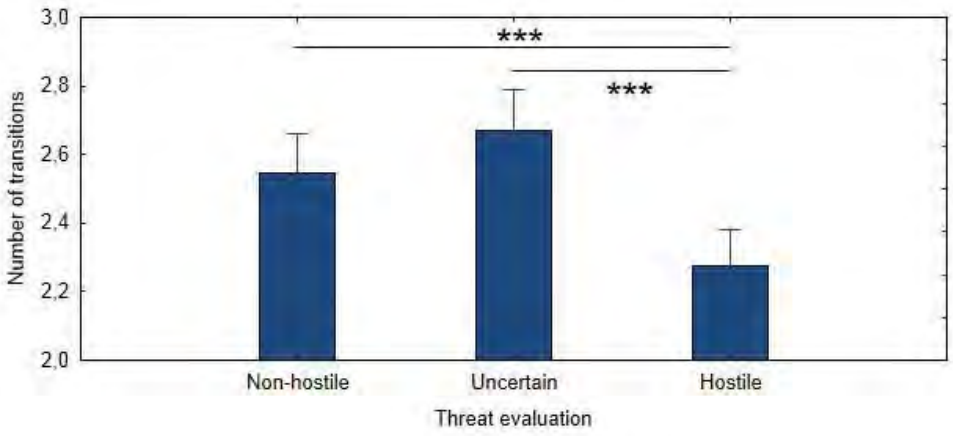


Figure 56.: Mean number of transitions between relevant parameters as a function of threat evaluation and post-hoc comparisons. Error bars indicate S.E.M.

classifications compared to “uncertain" and “non-hostile" classifications ($HSD < .001$) but no difference between “non-hostile" and “uncertain" decisions.

C.4 DISCUSSION

The objective of the present study was to understand how making a decision affects eye-tracking metrics during DDM and whether it is possible to foresee the upcoming choice before the decision is made. The results showed that decision outcome in the threat-assessment task — regardless of the accuracy of that judgment — induced different arousal levels and exploration strategies as revealed by pupil diameter and number of transitions on the parameters list, respectively. Participants exhibited larger TEPR and performed fewer transitions for “hostile" classifications. Vachon and Tremblay (2014), in their study conducted within the same microworld, showed that time-related ocular metrics (e.g., scanpath length, fixation duration) could predict decision-making efficiency and, to a lesser extent, that pupil size variation was related to decision accuracy. In the present paper, we extend these findings by showing that eye movements together with pupillometry can be indicative of the decision outcome even prior to the response. Since the visual processing was the same across all conditions, differences in scanning pattern were associated with differences in top-down decision-related mechanisms of overt attention (Henderson et al., 2007). In turn, dynamic pupil size changes likely reflect an alteration in the level of overall arousal, possibly in reaction to a new perceived threat. Altogether, eye-tracking metrics provide reliable indices of cognitive functioning during DDM.

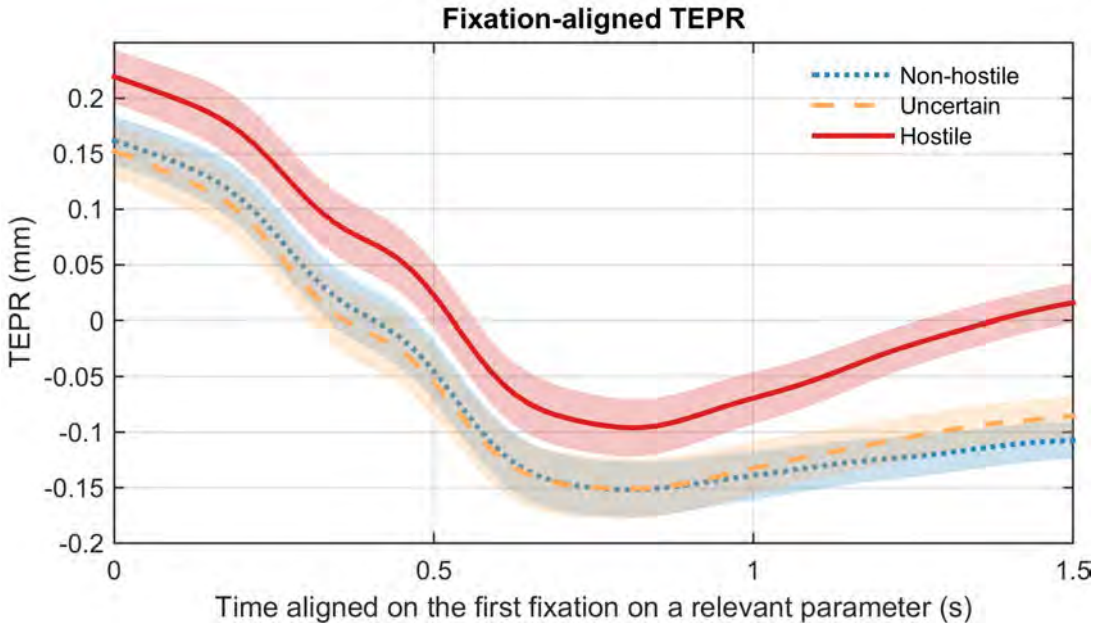


Figure 57.: Average TEPR locked to the first fixation on relevant parameters. Shaded regions indicate S.E.M.

It is however noteworthy that one should be cautious while interpreting results of an eye tracking study conducted in an ecological environment such as S-CCS microworld. Decisions made in such context are complex and include multiple actions engaging working memory performed sequentially during the whole decision-cycle. Therefore, given the latency and time course of TEPR (Beatty and Lucero-Wagoner, 2000), a decision cycle does not necessarily embrace the end of the pupillary reaction, especially in ecological protocols without any inter-stimulus interval. Examining only a part of reaction might lead to erroneous conclusions. Consequently, investigators using oculometry should thoroughly control for the relation between eye-tracking metrics and the time course of the trials. The following discussion clarifies this relation to justify the validity of our statistical results on the whole decision cycles. We also demonstrate that one can get an indication of a decision outcome before the operator clicks on the threat-assessment button. Finally, we discuss what the findings imply in terms of realistic applications of DSSs.

C.4.1 *TEPR*

As shown by our results, participants had larger pupils overall during “hostile" classifications. Taking into account the decision-time results combined with the visual design of the task (dark radar screen vs. bright parameter list), one could argue that this re-

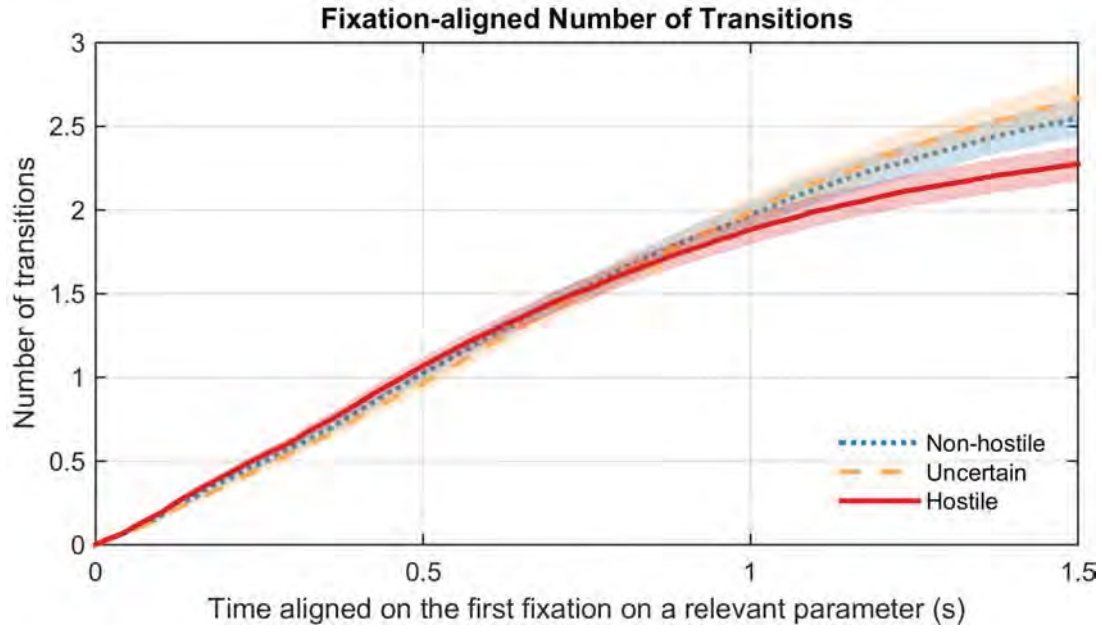


Figure 58.: Average temporal evolution of number of transitions. Shaded regions indicate S.E.M.

sult is due to pupil adaptation. Indeed, longer trials could allow the opportunity for further constriction. However, as shown in Fig. 57, the pupil did have time to adapt to new luminosity conditions for all threat conditions in-dependently of duration. Thus, after fixating the first relevant parameter, subjects showed a decrease in pupil diameter corresponding to pupillary light reflex — a constriction due to the difference of luminosities between the dark radar screen and the bright parameters list. The subsequent pupillary response over the course of the classification was cognition-driven (since participants' eyes had already adapted to the light background) and depended on evidence accumulation and corresponding decision-making processes. Participants' pupil size was indicative of their arousal/cognitive effort induced by a "hostile" aircraft classification. TEPR, mediated by sympathetic innervation, usually has 200 ms latency with a peak reaction occurring 1 s post-stimulus onset (Andreassi, 2000). As illustrated by Fig. 4 and Cohen's d analysis, the process of evidence accumulation, requiring participants to classify an aircraft as hostile, resulted in a growing difference in pupil diameter starting from 1,060 ms after the first fixation on relevant parameters. Also, in accordance with the literature, part of pupillary dilation might be attributed to affective processing. Hostile aircraft are threatening stimuli that are known to cause emotional response and higher sympathetic activation (Bradley et al., 2008).

Besides, the slightly larger pupil diameter at the moment of the first fixation (see Fig. 57) is noteworthy. The difference at the time origin may seem peculiar, since par-

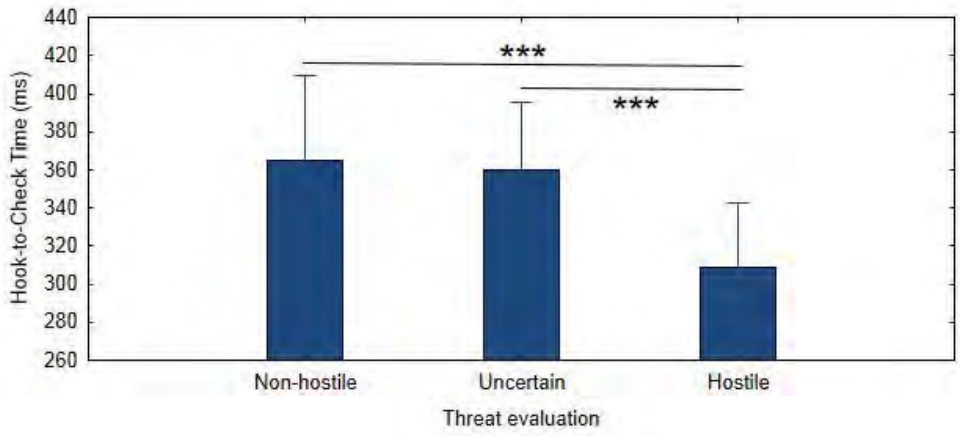


Figure 59.: Mean Hook-to-Check time as a function of threat evaluation and post-hoc comparisons. Error bars indicate S.E.M.

ticipants at that moment had not processed enough information to judge an aircraft as hostile. Indeed, they could not have used any parameter to make a decision at that time, suggesting that any advance suspicion about the hostile status of an aircraft would not be decision-related. Given that the pupil diameter can reflect the state of the autonomous nervous system and its sympathetic activation (Bradley et al., 2008), it is possible that this early pupil dilation for “hostile" decisions was associated with subjects’ feeling about the situation. Aircraft for which the status may switch from “non-hostile" to “hostile" altered their trajectory towards the ship, which may have signaled the occurrence of a new threat before checking the parameters, hence provoking sympathetic activation. In line with this hypothesis is the fact that subjects were faster to fixate the parameter list for “hostile" decisions, probably because they had identified an “unofficial" sign of the threat. Overall, at the moment of checking the parameters, subjects had an a priori opinion of the threat level of the aircraft, and the indicated difference in pupil diameter at the time origin reflects the premature arousal vis-à-vis the subsequent decision cycle.

C.4.2 *Number of Transitions*

After selecting an aircraft to classify, subjects switched to the parameters list to encode displayed values of parameters. If we assume that the number of total transitions between relevant parameters expresses the efficiency of visual search (or attentional cost; (Goldberg and Kotval, 1999)), fewer transitions would imply fewer cross-checks (and retro-saccades) between parameters. As mentioned earlier, investigators should thoroughly control the relationship between eye-tracking metrics and the time course of trials. If the conditions were of similar transition rate, longer decision-cycles would indeed

involve more gaze transitions across parameters. Therefore, we explored the temporal evolution of number of transitions and found that the conditions had different transition rates (i.e. the slope of curves on Fig. 58). The Cohen's d analysis confirmed this finding by indicating an approximate threshold (1,320 ms after the first fixation on relevant parameters) where the significant difference (fewer transitions for "hostile" classifications) emerged independently of decision-cycle length. We reported d values for the "uncertain" vs. "hostile" comparison as being most critical in terms of decision-making and possible implications for DSS, yet the "non-hostile" vs. "hostile" comparison gave similar results. Thus, the pre-decision ocular pattern over relevant information can be revealing of what that decision is likely to be.

C.4.3 *Implications for Intelligent DSSs*

These findings have some implications for the implementation of adaptive aiding systems. If we add to the maximum separation time (as shown by Cohen's d analysis) an average Hook-to-Check time, it means that one can have an indication of the upcoming decision within about 1.5 s after an aircraft selection. That leaves slightly more than half a second (for shortest cycles on average) before the actual judgment when the operator clicks on the threat-classification button. Therefore, an adaptive DSS could provide an estimation of upcoming aircraft assessment and undertake some preventive actions if, for example, the system has some reason to "think" that the operator is going to be wrong. That might, for example, reduce the possibility of Friendly Fire in military environments. Moreover, knowing that some ocular patterns (as revealed by Hook-to-Check time and pupil diameter) are characteristic of an upcoming decision but are purely subjective and therefore biased — since they are not due to the application of decision rules to the parameter values — the system might alert the operator of his/her behavior and suggest that the decision rules should be thoroughly applied. Note, nevertheless, that if such a support tool is to be implemented in air-traffic control, emergency response or security surveillance, then safety objectives might be applied. Since ocular metrics, and especially pupillary response, are known to be largely sensitive to environmental (Peysakhovich et al., 2015a) and psycho-physiological noise, to meet certification requirements and to further increase the discriminability, such a DSS must take into account more features (Chen and Epps, 2013) and combine them with the input of other sensors such as cardiopulmonary measures (e.g., (Dehais et al., 2011)).

C.5 CONCLUSION

The present study suggests that oculometry can provide reliable cues on the nature of the upcoming judgment made during a threat-evaluation task in an ecological DDM environment. The "hostile" threat assessment was associated with higher sympathetic

activation — as reflected by larger pupil dilation and faster gaze displacement — and a different visual scanning pattern during information extraction from the list of parameters. These findings have important implications for the development of adaptive DSSs. One could envision systems that would register likely upcoming decisions (e.g., threatening vs. non-threatening) slightly more than half a second before the actual choice is made (i.e. response entered on the user interface).

BIBLIOGRAPHY

AAIU. Accident: Boeing 737-8AS, EI-DWF; Near Perugia, Italy; 18 July 2013 (Report No. 2014-010). Technical report, July 2014. Retrieved from http://www.aaiu.ie/sites/default/files/report-attachments/REPORT%202014-010_0.pdf.

AAIB. Report on the serious incident to Boeing 737-3Q8, registration G-THOF, on approach to Runway 26, Bournemouth Airport, Hampshire on 23 September 2007. Formal Report AAR 3/2009. Technical report, 2009. Retrieved from https://www.gov.uk/aaib-reports/aar-3-2009-boeing-737-3q8-g-thof-23-september-2007.

AAIB. AAIB investigation to DHC-8-402 Dash 8, G-JECI (Report No. EW/C2008/12/05). Technical report, 2010. Retrieved from https://assets.digital.cabinet-office.gov.uk/media/5423028e40f0b61342000a93/DHC-8-402_Dash_8__G-JECI_03-10.pdf.

AAIB. AAIB investigation to DHC-8-402 Dash 8, G-JECF (Report No. EW/C2010/09/04). Technical report, 2012. Retrieved from https://assets.digital.cabinet-office.gov.uk/media/5422fc25ed915d1371000899/DHC-8-402_Dash_8_G-JECF_06-12.pdf.

AAIB. AAIB investigation to DHC-8-402 Dash 8, G-JECJ. Bulletin 8/2014, EW/G2014/02/03. Technical report, 2014. Retrieved from https://www.gov.uk/aaib-reports/dhc-8-402-dash-8-g-jecj-12-february-2014.

Aeronautica Civil. Aeronautica Civil of the Republic of Columbia Santafe de Bogata DC Columbia, Controlled Flight into Terrain AA 965 Boeing 757-223 N651AA near Cali, Columbia, 20 December 1995. Technical report, 1996. Retrieved from http://sunnyday.mit.edu/accidents/calirep.html.

Richard Andersson, Linnea Larsson, Kenneth Holmqvist, Martin Stridh, and Marcus Nyström. One algorithm to rule them all? an evaluation and discussion of ten eye movement event-detection algorithms. *Behavior research methods*, pages 1–22, 2016.

John L Andreassi. *Psychophysiology: Human behavior & physiological response*, chapter Pupillary response and behavior, pages 218–233. Psychology Press, 2000.

Gennady Andrienko, Natalia Andrienko, and Marco Heurich. An event-based conceptual model for context-aware movement analysis. *Int J Geogr Info Sci*, 25(9):1347–1370, 2011.

Gennady Andrienko, Natalia Andrienko, Michael Burch, and Daniel Weiskopf. Visual analytics methodology for eye movement studies. *IEEE TVCG*, 18(12):2889–2898, 2012.

Robert Ariel and Alan D Castel. Eyes wide open: Enhanced pupil dilation when selectively studying important information. *Experimental brain research*, 232(1):337–344, 2014.

G Aston-Jones and B Waterhouse. Locus coeruleus: From global projection system to adaptive regulation of behavior. *Brain research*, 2016.

Gary Aston-Jones and Jonathan D Cohen. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.*, 28:403–450, 2005.

Gary Aston-Jones, Janusz Rajkowski, and Jonathan Cohen. Role of locus coeruleus in attention and behavioral flexibility. *Biological psychiatry*, 46(9):1309–1320, 1999.

Australia's Transportation Safety Board. Unstable approach involving de Havilland Canada Dash 8 VH-XFZ (Report No. AO-2012-070). Technical report, 2013. Retrieved from http://www.atsb.gov.au/media/4471254/ao-2012-070_final.pdf.

BEA. Report No. f-ta930913 on the accident to the Air France Boeing 747-428 B registered F-GITA on September 13 1993 at Tahiti Faaa. Technical report, 1999. Retrieved from http://lessonslearned.faa.gov/AirFrance747/accident_report.pdf. Original French version is available at http://www.bea.aero/docspa/1993/f-ta930913/pdf/f-ta930913.pdf.

BEA. Final Report on the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro – Paris. (Report No. f-cp090601.en). Technical report, 2012. Retrieved from http://www.bea.aero/docspa/2011/f-pp111116.en/pdf/f-cp090601.en.pdf. Original French version is available at https://www.bea.aero/fileadmin/documents/docspa/2009/f-cp090601/pdf/f-cp090601.pdf.

BEA. Report No. f-cg120217. Serious Incident Report. Strong turbulence in cruise, momentary loss of control of the flight path by the crew. Technical report, 2014a. Retrieved from https://www.bea.aero/fileadmin/documents/docspa/2012/f-cg120227.en/pdf/f-cg120227.en.pdf. Original French version is available at https://www.bea.aero/fileadmin/documents/docspa/2012/f-cg120227/pdf/f-cg120227.pdf.

BEA. Momentary Loss of Control of the Flight Path during a Go-around (Report No. f-pp111116.en). Technical report, 2014b. Retrieved from http://www.bea.aero/docspa/2009/f-cp090601.en/pdf/f-pp111116.en.pdf. Original French version is available at https://www.bea.aero/fileadmin/documents/docspa/2011/f-pp111116/pdf/f-pp111116.pdf.

BEA. Report No. BEA2015-0125.en on the accident to the Airbus A320-211, registered D-AIPX and operated by Germanwings, flight GWI18G, on 03/24/15 at Prads-Haute-Bléone. Technical report, 2016. Retrieved from https://www.bea.aero/uploads/tx_elydbrapports/BEA2015-0125.en-LR.pdf. Original French version is available at https://www.bea.aero/uploads/tx_elydbrapports/BEA2015-0125-LR.pdf.

Jackson Beatty. Phasic not tonic pupillary responses vary with auditory vigilance performance. *Psychophysiology*, 19(2):167–172, 1982a.

Jackson Beatty. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin*, 91(2):276, 1982b.

Jackson Beatty and Brennis Lucero-Wagoner. The pupillary system. *Handbook of psychophysiology*, 2: 142–162, 2000.

Antoine Bechara, Hanna Damasio, and Antonio R Damasio. Emotion, decision making and the orbitofrontal cortex. *Cerebral cortex*, 10(3):295–307, 2000.

Andrew H Bellenkes, Christopher D Wickens, and Arthur F Kramer. Visual scanning and pilot expertise: the role of attentional flexibility and mental model development. *Aviation, Space, and Environmental Medicine*, 1997.

Simone Benedetto, Andrea Carbone, Véronique Drai-Zerbib, Marco Pedrotti, and Thierry Baccino. Effects of luminance and illuminance on visual fatigue and arousal during digital reading. *Computers in Human Behavior*, 41:112–119, 2014.

George E Billman. Heart rate variability - a historical perspective. *Frontiers in Physiology*, 2(86), 2011. ISSN 1664-042X. doi: 10.3389/fphys.2011.00086. URL http://www.frontiersin.org/clinical_and_translational_physiology/10.3389/fphys.2011.00086/abstract.

George E Billman. The lf/hf ratio does not accurately measure cardiac sympatho-vagal balance. *Frontiers in Physiology*, 4(26), 2013. ISSN 1664-042X. doi: 10.3389/fphys.2013.00026. URL http://www.frontiersin.org/clinical_and_translational_physiology/10.3389/fphys.2013.00026/full.

Frank E Bird. *Management guide to loss control*. Industrial Accident Prevention Association, Ontario, 1984.

T Blascheck, K Kurzhals, M Raschke, M Burch, D Weiskopf, and T Ertl. State-of-the-art of visualization for eye tracking data. In *Proceedings of EuroVis*, volume 2014, 2014.

Boeing. Statistical Summary of Commercial Jet Airplane Accidents Worldwide Operations, 1959 – 2014. Seattle, WA: Aviation Safety, Boeing Commercial Airplanes. Technical report, 2015. Retrieved from http://www.boeing.com/news/techissues/pdf/statsum.pdf.

Gianluca Borghini, Laura Astolfi, Giovanni Vecchiato, Donatella Mattia, and Fabio Babiloni. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews*, 44:58–75, 2014.

Ali Borji and Laurent Itti. Defending yarbus: Eye movements reveal observers' task. *Journal of vision*, 14(3):29–29, 2014.

Margaret M Bradley, Laura Miccoli, Miguel A Escrig, and Peter J Lang. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45(4):602–607, 2008.

John L Bradshaw. Background light intensity and the pupillary response in a reaction time task. *Psychonomic Science*, 14(6):271–272, 1969.

Anne-Marie Brouwer, Maarten A Hogervorst, Michael Holewijn, and Jan BF van Erp. Evidence for effects of task difficulty but not learning on neurophysiological variables associated with effort. *International Journal of Psychophysiology*, 93(2):242–252, 2014.

Andreas Bulling, Daniel Roggen, and Gerhard Tröster. Wearable eog goggles: Seamless sensing and context-awareness in everyday environments. *Journal of Ambient Intelligence and Smart Environments*, 1(2):157–171, 2009.

Michael Burch, Fabian Beck, Michael Raschke, Tanja Blascheck, and Daniel Weiskopf. A dynamic graph visualization perspective on eye movement data. In *Proc. ETRA*, pages 151–158. ACM, 2014.

Giovanni Calcagnini, S Lino, F Censi, and S Cerutti. Cardiovascular autonomic rhythms in spontaneous pupil fluctuations. In *Computers in Cardiology 1997*, pages 133–136. IEEE, 1997.

John A Caldwell, Melissa M Mallis, J Lynn Caldwell, Michel A Paul, James C Miller, and David F Neri. Fatigue countermeasures in aviation. *Aviation, space, and environmental medicine*, 80(1):29–59, 2009.

Stephen M Casner and Jonathan W Schooler. Vigilance impossible: Diligence, distraction, and daydreaming all lead to failures in a practical monitoring task. *Consciousness and cognition*, 35:33–41, 2015.

Monica S Castelhano and John M Henderson. The influence of color on the perception of scene gist. *Journal of Experimental Psychology: Human perception and performance*, 34(3):660, 2008.

Mickaël Causse, Jean-Michel Sénard, Jean François Démonet, and Josette Pastor. Monitoring cognitive and emotional processes through pupil and cardiac response during dynamic versus logical task. *Applied psychophysiology and biofeedback*, 35(2):115–123, 2010.

Mickaël Causse, Vsevolod Peysakhovich, and Kevin Mandrick. Eliciting sustained mental effort using the toulouse n-back task: Prefrontal cortex and pupillary responses. *Advances in Neuroergonomics and Cognitive Engineering*, page 185, 2017.

Siyuan Chen and Julien Epps. Automatic classification of eye activity for cognitive load measurement with emotion interference. *Computer methods and programs in biomedicine*, 110(2):111–124, 2013.

Bob Cheung and Kevin Hofer. Eye tracking, point of gaze, and performance degradation during disorientation. *Aviation, space, and environmental medicine*, 74(1):11–20, 2003.

Civil Aviation Authority. Monitoring Matters. Guidance on the Development of Pilot Monitoring Skills. CAA Paper 2013/02. Technical report, 2013. Retrieved from https://publicapps.caa.co.uk/docs/33/9323-CAA-Monitoring%20Matters%202nd%20Edition%20April%202013.pdf.

Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences. 2nd edn. Hillsdale, New Jersey: L.* Erlbaum, 1988.

Jonathan D Cohen, William M Perlstein, Todd S Braver, Leigh E Nystrom, Douglas C Noll, John Jonides, and Edward E Smith. Temporal dynamics of brain activation during a working memory task. *Nature*, 386(6625):604–608, 1997.

D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE TPAMI*, 24(5):603–619, 2002.

Vincent D Costa and Peter H Rudebeck. More than meets the eye: the relationship between pupil size and locus coeruleus activity. *Neuron*, 89(1):8–10, 2016.

Rodney J Croft and Robert J Barry. Removal of ocular artifact from the eeg: a review. *Neurophysiologie Clinique/Clinical Neurophysiology*, 30(1):5–19, 2000.

W. Cui, H. Zhou, H. Qu, P. Wong, and X. Li. Geometry-based edge clustering for graph visualization. *IEEE TVCG*, 14(6):1277–1284, 2008.

Jan Willem de Gee, Tomas Knapen, and Tobias H Donner. Decision-related pupil dilation reflects upcoming choice and individual bias. *Proceedings of the National Academy of Sciences*, 111(5):E618–E625, 2014.

Frédéric Dehais, Mickaël Causse, and Sébastien Tremblay. Mitigation of conflicts with automation use of cognitive countermeasures. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 53(5):448–460, 2011.

Frederic Dehais, Vsevolod Peysakhovich, Sebastien Scannella, Jennifer Fongue, and Thibault Gateau. Automation surprise in aviation: Real-time solutions. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 2525–2534. ACM, 2015.

Frédéric Dehais, Julia Behrend, Vsevolod Peysakhovich, Mickaël Causse, and Christopher Wickens. Pilot fying and pilot monitoring's state awareness during go-around execution in aviation: A behavioral and eye-tracking study. In *The International Journal of Aviation Psychology*, 2016. under review.

Sidney WA Dekker. Reconstructing human contributions to accidents: the new view on error and performance. *Journal of Safety Research*, 33(3):371–385, 2002.

Edmund B Delabarre. A method of recording eye-movements. *The American Journal of Psychology*, 9 (4):572–574, 1898.

Raymond Dodge and Thomas Sparks Cline. The angle velocity of eye movements. *Psychological Review*, 8(2):145, 1901.

Andrew Duchowski. *Eye tracking methodology: Theory and practice*, volume 373. Springer Science & Business Media, 2007.

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

Barnaby D Dunn, Tim Dalgleish, and Andrew D Lawrence. The somatic marker hypothesis: A critical evaluation. *Neuroscience & Biobehavioral Reviews*, 30(2):239–271, 2006.

Gautier Durantin, J-F Gagnon, Sébastien Tremblay, and Frédéric Dehais. Using near infrared spectroscopy and heart rate variability to detect mental overload. *Behavioural brain research*, 259:16–23, 2014.

Dutch Safety Board. Crashed during approach, Boeing 737-800, near Amsterdam Schiphol airport, 25 February 2009. Technical report, 2010. Retrieved from http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-63j-system-safety-fall-2012/related-resources/MIT16_63JF12_B737.pdf.

Mikael Elam, Torgny H Svensson, and Peter Thorén. Locus coeruleus neurons and sympathetic nerves: activation by cutaneous sensory afferents. *Brain research*, 366(1):254–261, 1986.

O. Ersoy, C. Hurter, F. Paulovich, G. Cantareira, and A. Telea. Skeleton-based edge bundles for graph visualization. *IEEE TVCG*, 17(2):2364 – 2373, 2011.

FAA. Advisory Circular No 23.1309-1E "System Safety Analysis and Assessment for Part 23 Airplanes". Technical report, 2011. Retrieved from http://www.faa.gov/documentLibrary/media/Advisory_Circular/AC%2023.1309-1E.pdf.

FAA. 14 CFR Part 121 Qualification, service, and use of crewmembers and aircraft dispatchers; final rule. Federal Register (78), 218. Technical report, 2013. Retrieved from http://www.gpo.gov/fdsys/granule/FR-2013-11-12/2013-26845.

Paul M Fitts, Richard E Jones, and John L Milton. Eye movements of aircraft pilots during instrument-landing approaches. *Aeronautical Engineering Review*, 9(2):24–29, 1950.

Ross Ed Flom, Kang Ed Lee, and Darwin Ed Muir. *Gaze-following: Its development and significance.* Lawrence Erlbaum Associates Publishers, 2007.

Thibault Gateau, Gautier Durantin, Francois Lancelot, Sebastien Scannella, and Frederic Dehais. Real-time state estimation in a flight simulator using fnirs. *PloS one*, 10(3):e0121279, 2015.

Andreas Gegenfurtner, Erno Lehtinen, and Roger Säljö. Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, 23(4):523–552, 2011.

Mark S Gilzenrat, Sander Nieuwenhuis, Marieke Jepma, and Jonathan D Cohen. Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective, & Behavioral Neuroscience*, 10(2):252–269, 2010.

Mackenzie G Glaholt. Eye tracking in the cockpit: a review of the relationships between eye movements and the aviator's cognitive state. 2014. URL http://cradpdf.drdc-rddc.gc.ca/PDFS/unc198/p801398_A1b.pdf.

Edmund Glaser and D Ruchkin. *Principles of neurobiological signal analysis*. Academic Press, 1976.

Joseph H Goldberg and Xerxes P Kotval. Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics*, 24(6):631–645, 1999.

Bram C Goldwater. Psychological significance of pupillary movements. *Psychological bulletin*, 77(5): 340, 1972.

Cleotilde Gonzalez. Decision support for real-time, dynamic decision-making tasks. *Organizational Behavior and Human Decision Processes*, 96(2):142–154, 2005.

Michelle R Greene, Tommy Liu, and Jeremy M Wolfe. Reconsidering yarbus: A failure to predict observers' task from eye movement patterns. *Vision research*, 62:1–8, 2012.

J Grünberger, L Linzmayer, H Walter, M Rainer, A Masching, L Pezawas, G Saletu-Zyhlarz, H Stöhr, and M Grünberger. Receptor test (pupillary dilatation after application of 0.01% tropicamide solution) and determination of central nervous activation (fourier analysis of pupillary oscillations) in patients with alzheimer's disease. *Neuropsychobiology*, 40(1):40–46, 1999.

Josef Grünberger. *Pupillometrie in der klinisch-psychophysiologischen Diagnostik*. Springer Vienna, 2003. doi: 10.1007/978-3-7091-6032-9. URL http://link.springer.com/book/10.1007%2F978-3-7091-6032-9.

Jun Su Ha, Young-Ji Byon, Joonsang Baek, and Poong Hyun Seong. Method for inference of operators' thoughts from eye movement data in nuclear power plants. *Nuclear Engineering and Technology*, 2015.

Daniel Hahnemann and Jackson Beatty. Pupillary responses in a pitch-discrimination task. *Perception & Psychophysics*, 2(3):101–105, 1967.

Hilde Haider and Peter A Frensch. Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(1):172, 1999.

RE Hampson, Ioan Opris, and SA Deadwyler. Neural correlates of fast pupil dilation in nonhuman primates: relation to behavioral performance and cognitive workload. *Behavioural brain research*, 212(1):1–11, 2010.

Yvonne Harrison and James A Horne. The impact of sleep deprivation on decision making: a review. *Journal of Experimental Psychology: Applied*, 6(3):236, 2000.

John M Henderson, James R Brockmole, Monica S Castelhano, and Michael Mack. Visual saliency does not account for eye movements during visual search in real-world scenes. *Eye movements: A window on mind and brain*, pages 537–562, 2007.

John M Henderson, Svetlana V Shinkareva, Jing Wang, Steven G Luke, and Jenn Olejarczyk. Predicting cognitive state from eye movements. *PloS one*, 8(5):e64937, 2013.

Eckhard H Hess. Attitude and pupil size. *Scientific american*, 1965.

Eckhard H Hess and James M Polt. Pupil size as related to interest value of visual stimuli. *Science*, 132 (3423):349–350, 1960.

Eckhard H Hess and James M Polt. Pupil size in relation to mental activity during simple problem-solving. *Science*, 143(3611):1190–1192, 1964.

Helen M Hodgetts, François Vachon, and Sébastien Tremblay. Background sound impairs interruption recovery in dynamic task situations: Procedural conflict? *Applied Cognitive Psychology*, 28(1):10–21, 2014.

Helen M Hodgetts, Sébastien Tremblay, Benoît R Vallières, and François Vachon. Decision support and vulnerability to interruption in a dynamic multitasking environment. *International Journal of Human-Computer Studies*, 79:106–117, 2015.

Bert Hoeks and Willem JM Levelt. Pupillary dilation as a measure of attention: A quantitative system analysis. *Behavior Research Methods, Instruments, & Computers*, 25(1):16–26, 1993.

Marion Höfle, Ramona Kenntner-Mabiala, Paul Pauli, and Georg W Alpers. You can see pain in the eye: pupillometry as an index of pain intensity under different luminance conditions. *International journal of psychophysiology*, 70(3):171–175, 2008.

Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van de Weijer. *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford, 2011.

Kenneth Holmqvist, Marcus Nyström, and Fiona Mulvey. Eye tracker data quality: what it is and how to measure it. In *Proceedings of the symposium on eye tracking research and applications*, pages 45–52. ACM, 2012.

D. Holten and J. J. van Wijk. Force-directed edge bundling for graph visualization. *CGF*, 28(3):670–677, 2009a.

Danny Holten and Jarke J van Wijk. A user study on visualizing directed edges in graphs. In *Proc. ACM CHI*, pages 2299–2308, 2009b.

Edmund B Huey. Preliminary experiments in the physiology and psychology of reading. *The American Journal of Psychology*, 9(4):575–586, 1898.

C. Hurter, B. Tissoires, and S. Conversy. FromDaDy: Spreading data across views to support iterative exploration of aircraft trajectories. *IEEE TVCG*, 15(6):1017–1024, 2009.

C. Hurter, O. Ersoy, and A. Telea. Graph bundling by kernel density estimation. *CGF*, 31(3):435–443, 2012.

C. Hurter, O. Ersoy, S. Fabrikant, T. Klein, and A. Telea. Bundled visualization of dynamic graph and trail data. *IEEE TVCG*, 20(8):1141–1157, 2014.

Christophe Hurter, Alexandru Telea, and Ozan Ersoy. MoleView: An attribute and structure-based semantic lens for large element-based plots. *IEEE TVCG*, 17(12):2600–2609, 2011.

Karel Hurts and Robert J de Boer. "what is it doing now?" results of a survey into automation surprise. In *Proceedings of 31st EAAP Conference*, pages 197–210, 2014.

Edwin Hutchins and Leysia Palen. Constructing meaning from space, gesture, and speech. In *Discourse, Tools and Reasoning*, pages 23–40. Springer, 1997.

Toshiyuki Inagaki. Design of human–machine interactions in light of domain-dependence of human-centered automation. *Cognition, Technology & Work*, 8(3):161–167, 2006.

Toshiyuki Inagaki and Thomas B Sheridan. Authority and responsibility in human–machine systems: probability theoretic validation of machine-initiated trading of authority. *Cognition, technology & work*, 14(1):29–37, 2012.

Shamsi T Iqbal and Brian P Bailey. Using eye gaze patterns to identify user tasks. In *The Grace Hopper Celebration of Women in Computing*, pages 5–10, 2004.

RJ Jacob and Keith S Karn. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind*, 2(3):4, 2003.

S Jainta and Thierry Baccino. Analyzing the pupil response due to increased cognitive demand: An independent component analysis study. *International Journal of Psychophysiology*, 77(1):1–7, 2010.

Halszka Jarodzka, Thomas Balslev, Kenneth Holmqvist, Marcus Nyström, Katharina Scheiter, Peter Gerjets, and Berit Eika. Conveying clinical reasoning based on visual observation via eye-movement modelling examples. *Instructional Science*, 40(5):813–827, 2012.

Halszka Jarodzka, Tamara van Gog, Michael Dorr, Katharina Scheiter, and Peter Gerjets. Learning to see: Guiding students' attention via a model's eye movements fosters learning. *Learning and Instruction*, 25:62–70, 2013.

Emile Javal. Essai sur la physiologie de la lecture. In *Annales D\'Oculistique*, 1879.

Putai Jin. Toward a reconceptualization of the law of initial value. *Psychological bulletin*, 111(1):176, 1992.

Richard E Jones, John L Milton, and Paul M Fitts. Eye fixations of aircraft pilots, i. a review of prior eye-movement studies and a description of a technique for recording the frequency, duration and sequences of eye-fixations during instrument flight. *Wright Patterson AFB, OH, USAF Tech. Rep*, 5837, 1949.

Siddhartha Joshi, Yin Li, Rishi M Kalwani, and Joshua I Gold. Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron*, 89(1):221–234, 2016.

Daniel Kahneman. *Attention and effort*. Prentice-Hall, 1973.

Daniel Kahneman and Jackson Beatty. Pupil diameter and load on memory. *Science*, 154(3756):1583–1585, 1966.

Olivia E Kang, Katherine E Huffer, and Thalia P Wheatley. Pupil dilation dynamics track attention to high-level information. *PloS one*, 9(8):e102463, 2014.

Peter Kasarskis, Jennifer Stehwien, Joey Hickox, Anthony Aretz, and Chris Wickens. Comparison of expert and novice scan behaviors during vfr flight. In *Proceedings of the 11th International Symposium on Aviation Psychology*, pages 1–6. Citeseer, 2001.

Tomas Knapen, Jan Willem de Gee, Jan Brascamp, Stijn Nuiten, Sylco Hoppenbrouwers, and Jan Theeuwes. Cognitive and ocular factors jointly determine pupil responses under equiluminance. *PloS one*, 11(5):e0155574, 2016.

John I Lacey. The evaluation of autonomic responses: Toward a general solution. *Annals of the New York Academy of Sciences*, 67(5):125–163, 1956.

Bruno Laeng, Sylvain Sirois, and Gustaf Gredebäck. Pupillometry a window to the preconscious? *Perspectives on psychological science*, 7(1):18–27, 2012.

A. Lambert, R. Bourqui, and D. Auber. Winding roads: Routing edges into bundles. *CGF*, 29(3):432–439, 2010.

P. Laube. *Progress in Movement Pattern Analysis*. BMI Books, 2009.

Olivier Le Meur and Thierry Baccino. Methods for comparing scanpaths and saliency maps: strengths and weaknesses. *Behavior research methods*, 45(1):251–266, 2013.

Daniel R Leff, David RC James, Felipe Orihuela-Espina, Ka-Wai Kwok, Loi Wah Sun, George Mylonas, Thanos Athanasiou, Ara W Darzi, and Guang-Zhong Yang. The impact of expert visual guidance on trainee visual search strategy, visual attention and motor skills. *Frontiers in human neuroscience*, 9, 2015.

Sebastian J Lehmann and Brian D Corneil. Transient pupil dilation after subsaccadic microstimulation of primate frontal eye fields. *The Journal of Neuroscience*, 36(13):3765–3776, 2016.

Roger Lew, Brian P Dyre, Steffen Werner, Brian Wotring, and Tuan Tran. Exploring the potential of short-time fourier transforms for analyzing skin conductance and pupillometry in real-time applications. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 52, pages 1536–1540. Sage Publications, 2008.

Antoine Lhuillier and Christophe Hurter. Bundling, graph simplification trough visual aggregation: existing techniques and challenges. In *Proceedings of the 27th Conference on l'Interaction Homme-Machine*, page 10. ACM, 2015.

Damien Litchfield, Linden J Ball, Tim Donovan, David J Manning, and Trevor Crawford. Viewing another person's eye movements improves identification of pulmonary nodules in chest x-ray inspection. *Journal of Experimental Psychology: Applied*, 16(3):251, 2010.

Simon P Liversedge, Keith Rayner, Sarah J White, Dorine Vergilino-Perez, John M Findlay, and Robert W Kentridge. Eye movements when reading disappearing text: is there a gap effect in reading? *Vision research*, 44(10):1013–1024, 2004.

Irene E Loewenfeld. Mechanisms of reflex dilatation of the pupil. *Documenta Ophthalmologica*, 12(1): 185–448, 1958.

Irene E Loewenfeld and Otto Lowenstein. *The pupil: anatomy, physiology, and clinical applications*, volume 2. Wiley-Blackwell, 1993.

Otto Lowenstein and Irene E Loewenfeld. The pupil. *The eye*, 3:231–267, 1962.

Otto Lowenstein, Richard Feinberg, and Irene E Loewenfeld. Pupillary movements during acute and chronic fatigue a new test for the objective evaluation of tiredness. *Investigative Ophthalmology & Visual Science*, 2(2):138–157, 1963.

Holger Lüdtke, Barbara Wilhelm, Martin Adler, Frank Schaeffel, and Helmut Wilhelm. Mathematical procedures in data recording and processing of pupillary fatigue waves. *Vision research*, 38(19):2889–2896, 1998.

Kevin Mandrick, Vsevolod Peysakhovich, Florence Rémy, Evelyne Lepron, and Mickaël Causse. Neural and psychophysiological correlates of human performance under stress and high mental workload. *Biological psychology*, 121:62–73, 2016.

Barry R Manor and Evian Gordon. Defining the temporal threshold for ocular fixation in free-viewing visuocognitive tasks. *Journal of neuroscience methods*, 128(1):85–93, 2003.

Sandra P Marshall. The index of cognitive activity: Measuring cognitive workload. In *Human factors and power plants, 2002. proceedings of the 2002 IEEE 7th conference on*, pages 7–5. IEEE, 2002.

Aude Marzuoli, Christophe Hurter, and Eric Feron. Data visualization techniques for airspace flow modeling. In *Proc. CIDU*, pages 79–86. IEEE, 2012.

Ethel Matin. Saccadic suppression: a review and an analysis. *Psychological bulletin*, 81(12):899, 1974.

Nadine Matton, Pierre Paubel, Julien Cegarra, and Eric Raufaste. Resource allocation strategies in multitasking after switch in task priorities. *Advances in Cognitive Engineering and Neuroergonomics*, 11: 187, 2014.

R Andy McKinley, Lindsey K McIntire, Regina Schmidt, Daniel W Repperger, and John A Caldwell. Evaluation of eye metrics as a detector of fatigue. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 53(4):403–414, 2011.

Ranjana K Mehta and Raja Parasuraman. Neuroergonomics: a review of applications to physical and cognitive work. *Front. Hum. Neurosci*, 7:889, 2013.

JL Milton, RE Jones, and PM Fitts. Eye movements of aircraft pilots: Ii. frequency, duration, and sequence of fixation when flying the usaf instrument low approach system (ILAS). *Wright-Patterson Air Force Base, Dayton, OH, USAF Tech. Rep*, (5839), 1949.

Andrei C Miu, Renata M Heilman, and Daniel Houser. Anxiety impairs decision-making: psychophysiological evidence from an iowa gambling task. *Biological psychology*, 77(3):353–358, 2008.

Kei Mizuno, Masaaki Tanaka, Kouzi Yamaguti, Osami Kajimoto, Hirohiko Kuratsune, and Yasuyoshi Watanabe. Mental fatigue caused by prolonged cognitive load associated with sympathetic hyperactivity. *Behavioral and brain functions*, 7(1):1, 2011.

OH Mowrer, Theodore C Ruch, and NE Miller. The corneo-retinal potential difference as the basis of the galvanometric method of recording eye movements. *American Journal of Physiology–Legacy Content*, 114(2):423–428, 1935.

Shalini Mukherjee, Rajeev Yadav, Iris Yung, Daniel P Zajdel, and Barry S Oken. Sensitivity to mental effort and test–retest reliability of heart rate variability measures in healthy seniors. *Clinical Neurophysiology*, 122(10):2059–2066, 2011.

Randall J Mumaw, N Sarter, and Christopher D Wickens. Analysis of pilots' monitoring and performance on an automated flight deck. In *11th International Symposium on Aviation Psychology, Columbus, OH*, 2001.

Atsuo Murata and Hirokazu Iwase. Evaluation of mental workload by fluctuation analysis of pupil area. In *Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE*, volume 6, pages 3094–3097. IEEE, 1998.

Atsuo Murata and Hirokazu Iwase. Evaluation of mental workload by variability of pupil area. *IEICE TRANSACTIONS on Information and Systems*, 83(5):1187–1190, 2000.

Peter R Murphy, Ian H Robertson, Joshua H Balsters, and Redmond G O'connell. Pupillometry and p3 index the locus coeruleus–noradrenergic arousal function in humans. *Psychophysiology*, 48(11): 1532–1543, 2011.

Peter R Murphy, Redmond G O'Connell, Michael O'Sullivan, Ian H Robertson, and Joshua H Balsters. Pupil diameter covaries with bold activity in human locus coeruleus. *Human Brain Mapping*, 35(8): 4140–4154, 2014.

Marnix Naber and Ken Nakayama. Pupil responses to high-level image content. *Journal of vision*, 13(6): 7–7, 2013.

Minoru Nakayama. Influence of blink on pupillary indices. In *Biomedical Circuits and Systems Conference, 2006. BioCAS 2006. IEEE*, pages 29–32. IEEE, 2006.

Minoru Nakayama and Yasutaka Shimizu. Frequency analysis of task evoked pupillary response and eye-movement. In *Proceedings of the 2004 symposium on Eye tracking research & applications*, pages 71–76. ACM, 2004.

Deepthi Nalanagula, Joel S Greenstein, and Anand K Gramopadhye. Evaluation of the effect of feed-forward training displays of search strategy on visual search performance. *International Journal of Industrial Ergonomics*, 36(4):289–300, 2006.

Maurice Nevile. Seeing the point: attention and participation in the airline cockpit. In *Interacting Bodies. Online Proceedings of the 2nd International Conference of the International Society for Gesture Studies. Lyon: ENS LSH and ICAR Research Lab, http://gesture-lyon2005. enslsh. fr/article. php3*, 2007.

Sander Nieuwenhuis, Gary Aston-Jones, and Jonathan D Cohen. Decision making, the p3, and the locus coeruleus–norepinephrine system. *Psychological bulletin*, 131(4):510, 2005.

Alvin W North. Accuracy and precision of electro-oculographic recording. *Investigative Ophthalmology & Visual Science*, 4(3):343–348, 1965.

NTSB. Aircraft Accidet Report, Eastern Air Lines, Inc., L-1011, N310EA, Miami, Florida, December 29, 1972 (Report No. AAR-73-14). Technical report, 1973. Retrieved from http://www.ntsb.gov/investigations/AccidentReports/Reports/AAR7314.pdf.

NTSB. A Review of Flightcrew-Involved, Major Accidents of U.S. Carriers, 1978 through 1990, Safety Study NTSB/SS-94/01. Technical report, 1994. Retrieved from http://libraryonline.erau.edu/online-full-text/ntsb/safety-studies/SS94-01.pdf.

NTSB. Loss of Control on Approach, Colgan Air, Inc., Operating as Continental Connection Flight 3407, Bombardier DHC-8-400, N200WQ, Clarence Center, New York, February 12, 2009, NTSB/AAR-10/01. Technical report, 2010. Retrieved from www.ntsb.gov/investigations/AccidentReports/Reports/AAR1001.pdf.

NTSB. Crash During a Nighttime Nonprecision Instrument Approach to Landing, UPS Flight 1354 (Report No. AAR1402). Technical report, 2014a. Retrieved from http://www.ntsb.gov/investigations/AccidentReports/Reports/AAR1402.pdf.

NTSB. Descent Below Visual Glidepath and Impact With Seawall, Asiana Airlines Flight 214, Boeing 777-200ER, HL7742, San Francisco, California, July 6, 2013, NTSB/AAR-14/01. Technical report, 2014b. Retrieved from http://www.ntsb.gov/investigations/AccidentReports/Reports/AAR1401.pdf.

NTSB. Safety Recommendation Ref. A-15-1 through -8. Technical report, 2015. Retrieved from http://www.ntsb.gov/safety/safety-recs/recletters/a-15-001-008.pdf.

NTSC. National Transportation Safety Committee, Indonesia, KNKT/07.01/08.01/36, Boeing 737-4Q8, Makassar Strait Sulawesi Republic of Indonesia, 1 January 2007. Technical report, 2007. Retrieved from https://aviation-safety.net/get.php?http://asndata.aviation-safety.net/reports/2007/20070101-0_B734_PK-KKW.pdf.

Marcus Nyström, Richard Andersson, Kenneth Holmqvist, and Joost Van De Weijer. The influence of calibration method and eye physiology on eyetracking data quality. *Behavior research methods*, 45(1): 272–288, 2013.

William D O'Neill and S Zimmermann. Neurologial interpretations and the information in the cognitive pupillary response. *Methods of information in medicine*, 39(2):122–124, 2000.

Massimo Pagani, Federico Lombardi, Stefano Guzzetti, Ornella Rimoldi, RAFFAELLO Furlan, PAOLO Pizzinelli, Giulia Sandrone, Gabriella Malfatto, Simonetta Dell'Orto, and Emanuela Piccaluga. Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympatho-vagal interaction in man and conscious dog. *Circulation research*, 59(2):178–193, 1986.

Oskar Palinko and Andrew L Kun. Exploring the effects of visual cognitive load and illumination on pupil diameter in driving simulators. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 413–416. ACM, 2012.

Jonathan W Peirce. Psychopy—psychophysics software in python. *Journal of neuroscience methods*, 162(1):8–13, 2007.

Maria Pereverzeva, Paola Binda, and Scott O Murray. Covert attention to bright and dark surfaces drives pupillary responses. *Journal of Vision*, 12(9):661–661, 2012.

Vsevolod Peysakhovich, Mickaël Causse, Sébastien Scannella, and Frédéric Dehais. Frequency analysis of a task-evoked pupillary response: Luminance-independent measure of mental effort. *International Journal of Psychophysiology*, 97(1):30–37, 2015a.

Vsevolod Peysakhovich, Christophe Hurter, and Alexandru Telea. Attribute-driven edge bundling for general graphs with applications in trail analysis. In *2015 IEEE Pacific Visualization Symposium (PacificVis)*, pages 39–46. IEEE, 2015b.

Vsevolod Peysakhovich, François Vachon, Benoît R Vallières, Frédéric Dehais, and Sébastien Tremblay. Pupil dilation and eye movements can reveal upcoming choice in dynamic decision-making. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 59, pages 210–214. SAGE Publications, 2015c.

Vsevolod Peysakhovich, Olivier Lefrançois, Frédéric Dehais, and Mickaël Causse. Neuroergonomics of the aircraft cockpits: the four stages of eye-tracking integration to enhance flight safety. *Safety Science*, 2016a. under review.

Vsevolod Peysakhovich, François Vachon, and Frédéric Dehais. The impact of luminance on tonic and phasic pupillary responses to sustained cognitive load. *International Journal of Psychophysiology*, 2016b. accepted.

Bastian Pfleging, Drea K Fekety, Albrecht Schmidt, and Andrew L Kun. A model relating pupil diameter to mental workload and lighting conditions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 5776–5788. ACM, 2016.

Sergio Pizziol, Catherine Tessier, and Frédéric Dehais. Petri net-based modelling of human–automation conflicts in aviation. *Ergonomics*, 57(3):319–331, 2014.

Marc Pomplun and Sindhura Sunkara. Pupil dilation as an indicator of cognitive workload in human-computer interaction. In *Proceedings of the International Conference on HCI*, pages 542–546. Citeseer, 2003.

Roy M Pritchard. *Stabilized images on the retina*, volume 511. WH Freeman Company, 1961.

Dale Purves, George J Augustine, David Fitzpatrick, LC Katz, Anthony-Samuel LaMantia, JO McNamara, and SM Williams. *Neuroscience*. Sinauer, Sunderland, 2001.

J Rajkowski, P Kubiak, and G Aston-Jones. Correlations between locus coeruleus (lc) neural activity, pupil diameter and behavior in monkey support a role of lc in attention. In *Society for neuroscience abstracts*, volume 19, page 974, 1993.

Keith Rayner. Eye movements and attention in reading, scene perception, and visual search. *The quarterly journal of experimental psychology*, 62(8):1457–1506, 2009.

Keith Rayner, Albrecht Werner Inhoff, Robert E Morrison, Maria L Slowiaczek, and James H Bertera. Masking of foveal and parafoveal vision during eye fixations in reading. *Journal of Experimental Psychology: Human perception and performance*, 7(1):167, 1981.

Keith Rayner, Simon P Liversedge, Sarah J White, and Dorine Vergilino-Perez. Reading disappearing text cognitive control of eye movements. *Psychological science*, 14(4):385–388, 2003.

Keith Rayner, Simon P Liversedge, and Sarah J White. Eye movements when reading disappearing text: The importance of the word to the right of fixation. *Vision research*, 46(3):310–323, 2006.

James Reason. *Human error*. Cambridge university press, 1990.

Nicolas Regis, Frédéric Dehais, Catherine Tessier, and Jean-François Gagnon. Ocular metrics for detecting attentional tunnelling. *Human Factors and Ergonomics Society Europe Chapter Annual Meeting*, 2012.

Nicolas Régis, Frédéric Dehais, Emmanuel Rachelson, Charles Thooris, Sergio Pizziol, Mickaël Causse, and Cedric Tessier. Formal detection of attentional tunneling in human operator–automation interactions. volume 44, pages 326–336. IEEE, 2014.

Miriam Reiner and Tatiana M Gelfeld. Estimating mental workload through event-related fluctuations of pupil area during a task in a virtual world. *International Journal of Psychophysiology*, 93(1):38–44, 2014.

Eyal M Reingold and Heather Sheridan. Eye movements and visual expertise in chess and medicine. *Oxford handbook on eye movements*, pages 528–550, 2011.

Peng Ren, Ana Barreto, Ying Gao, and Malek Adjouadi. Affective assessment by digital processing of the pupil diameter. *Affective Computing, IEEE Transactions on*, 4(1):2–14, 2013.

Jeremy R Reynolds, Robert West, and Todd Braver. Distinct neural circuits support transient and sustained processes in prospective memory and working memory. *Cerebral Cortex*, 19(5):1208–1221, 2009.

David A Robinson. A method of measuring eye movemnent using a scieral search coil in a magnetic field. *IEEE Transactions on bio-medical electronics*, 10(4):137–145, 1963.

Timothy A Salthouse and Cecil L Ellis. Determinants of eye-fixation duration. *The American journal of psychology*, pages 207–234, 1980.

ER Samuels and E Szabadi. Functional neuroanatomy of the noradrenergic locus coeruleus: its roles in the regulation of arousal and autonomic function part i: principles of functional organisation. *Current neuropharmacology*, 6(3):235–253, 2008.

Susan J Sara and Sebastien Bouret. Orienting and reorienting: the locus coeruleus mediates cognition through arousal. *Neuron*, 76(1):130–141, 2012.

Nadine B Sarter and David D Woods. How in the world did we ever get into that mode? mode error and awareness in supervisory control. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1):5–19, 1995.

Eduard Schott. Über die registrierung des nystagmus und anderer augenbewegungen vermittels des saitengalvanometers. *Dtsch Arch klin med*, 140:79–90, 1922.

Angela T Schriver, Daniel G Morrow, Christopher D Wickens, and Donald A Talleur. Expertise differences in attentional strategies related to pilot decision making. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(6):864–878, 2008.

D. Selassie, B. Heller, and J. Heer. Divided edge bundling for directional network data. *IEEE TVCG*, 19 (12):754–763, 2011.

Scott Shappell, Cristy Detwiler, Kali Holcomb, Carla Hackworth, Albert Boquet, and Douglas A Wiegmann. Human error and commercial aviation accidents: an analysis using the human factors analysis and classification system. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 49(2):227–242, 2007.

Scott A Shappell and Douglas A Wiegmann. The human factors analysis and classification system – HFACS. Technical report, US Federal Aviation Administration, Office of Aviation Medicine, 2000.

Scott A Shappell and Douglas A Wiegmann. Applying reason: The human factors analysis and classification system (HFACS). *Human Factors and Aerospace Safety*, 2001.

Scott A Shappell, Cristy A Detwiler, Kali A Holcomb, Carla A Hackworth, Albert J Boquet, and Douglas A Wiegmann. Human error and commercial aviation accidents: A comprehensive, fine-grained analysis using HFACS. Technical report, DTIC Document, 2006.

G Siegle. Analysis of pupillary data. In *Fundamentals of Pupillary Measures and Eye Movements Workshop*. Society for Psychophysiological Research, 2011.

Richard S Snell and Michael A Lemp. *Clinical anatomy of the eye*. John Wiley & Sons, 2013.

Stuart R Steinhauer and Gad Hakerem. The pupillary response in cognitive psychophysiology and schizophreniaa. *Annals of the New York Academy of Sciences*, 658(1):182–204, 1992.

Stuart R Steinhauer, Greg J Siegle, Ruth Condray, and Misha Pless. Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. *International journal of psychophysiology*, 52(1):77–86, 2004.

Joseph Sullivan, Ji Hyun Yang, Michael Day, and Quinn Kennedy. Training simulation for helicopter navigation by characterizing visual scan patterns. *Aviation, space, and environmental medicine*, 82(9): 871–878, 2011.

Robert Sumwalt, David Cross, and Dennis Lessard. Examining how breakdowns in pilot monitoring of the aircraft flight path. *International Journal of Aviation, Aeronautics, and Aerospace*, 2(3):8, 2015.

TAIC. Transport Accident Investigation Commission, New Zealand, Report 95-011, de Havilland DHC-8, ZK-NEY, controlled flight into terrain near Palmerston North, 9 June 1995. Technical report, 1995. Retrieved from https://aviation-safety.net/get.php?http://asndata.aviation-safety.net/reports/2007/20070101-0_B734_PK-KKW.pdf.

TAIIB. Cabin de-pressurization of Bombardier DHC-8-400 aircraft, registration YL-BAH on september 9, 2012. Final report Nr.4-02/7-12/(6-13) of the aircraft serious incident. Technical report, 2012. Retrieved from http://www.taiib.gov.lv/uploads/FINAL%20REPORT%20Nr.4-027-12_6-13_.pdf.

Task Force of the European Society of Cardiology the North American Society of Pacing. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *Circulation*, 93(5):1043–1065, 1996. doi: 10.1161/01.CIR.93.5.1043. URL http://circ.ahajournals.org/content/93/5/1043.short.

A. Telea and O. Ersoy. Image-based edge bundles: Simplified visualization of large graphs. *CGF*, 29(3): 543–551, 2010.

Catherine Tessier and Frédéric Dehais. Authority management and conflict solving in human-machine systems. *AerospaceLab*, (4):p–1, 2012.

Transportation Safety Board of Canada. Reassessment of the responses to aviation safety recommendation A03-08. Technical report, 2014. Retrieved from http://www.bst-tsb.gc.ca/eng/recommandations-recommendations/aviation/2003/rec_a0308.pdf.

Sébastien Tremblay, François Vachon, Robert Rousseau, and Richard Breton. Promoting temporal awareness for dynamic decision making in command and control. *Advances in cognitive engineering and neuroergonomics*, pages 3422–3431, 2012.

Warren W Tryon. Pupillometry: A survey of sources of variation. *Psychophysiology*, 12(1):90–93, 1975.

Shiro Usui and Yutaka Hirata. Estimation of autonomic nervous activity using the inverse dynamic model of the pupil muscle plant. *Annals of biomedical engineering*, 23(4):375–387, 1995.

Shiro Usui and Lawrence Stark. A model for nonlinear stochastic behavior of the pupil. *Biological cybernetics*, 45(1):13–21, 1982.

François Vachon and Sébastien Tremblay. What eye tracking can reveal about dynamic decision-making. *Advances in cognitive engineering and neuroergonomics*, 11:157, 2014.

François Vachon, Benoît R Vallières, Dylan M Jones, and Sébastien Tremblay. Nonexplicit change detection in complex dynamic settings what eye movements reveal. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 54(6):996–1007, 2012.

S. van den Elzen and J.J. van Wijk. Multivariate network exploration and presentation: from detail to overview via selections and aggregations. *IEEE TVCG*, 20(12):2310–2319, 2014.

Pascal WM Van Gerven, Fred Paas, Jeroen JG Van Merriënboer, and Henk G Schmidt. Memory load and the cognitive pupillary response in aging. *Psychophysiology*, 41(2):167–174, 2004.

Ludo W van Meeuwen, Halszka Jarodzka, Saskia Brand-Gruwel, Paul A Kirschner, Jeano JPR de Bock, and Jeroen JG van Merriënboer. Identification of effective visual problem solving strategies in a complex visual domain. *Learning and Instruction*, 32:10–21, 2014.

Boris M Velichkovsky. Communicating attention: Gaze position transfer in cooperative problem solving. *Pragmatics & Cognition*, 3(2):199–223, 1995.

Chin-An Wang and Douglas P Munoz. A circuit for pupil orienting responses: implications for cognitive modulation of pupil size. *Current opinion in neurobiology*, 33:134–140, 2015.

Christopher D Wickens. Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, 3(2):159–177, 2002.

Douglas A Wiegmann and Scott A Shappell. Applying the human factors analysis and classification system (HFACS) to the analysis of commercial aviation accident data. In *11th international symposium on aviation psychology*, 2001.

Stefan M Wierda, Hedderik van Rijn, Niels A Taatgen, and Sander Martens. Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proceedings of the National Academy of Sciences*, 109(22):8456–8460, 2012.

Joseph Wilder. *Stimulus and response: The law of initial value*. Wright, Bristol, England, 1967.

Barbara Wilhelm, Helmut Wilhelm, H Lüdtke, Peter Streicher, and Martin Adler. Pupillographic assessment of sleepiness in sleep-deprived healthy subjects. *Sleep*, 21(3):258–265, 1998.

Glenn F Wilson and Christopher A Russell. Performance enhancement in an uninhabited air vehicle task using psychophysiologically determined adaptive aiding. *Human factors: the journal of the human factors and ergonomics society*, 49(6):1005–1018, 2007.

Barry Winn, David Whitaker, David B Elliott, and Nicholas J Phillips. Factors affecting light-adapted pupil size in normal human subjects. *Investigative Ophthalmology & Visual Science*, 35(3):1132–1137, 1994.

David S Wooding. Fixation maps: quantifying eye-movement traces. In *Proceedings of the 2002 symposium on Eye tracking research & applications*, pages 31–36. ACM, 2002.

Jie Xu, Yang Wang, Fang Chen, and Eric Choi. Pupillary response based cognitive workload measurement under luminance changes. In *Human-Computer Interaction–INTERACT 2011*, pages 178–185. Springer, 2011.

Alfred L Yarbus. *Eye movements during perception of complex objects*. Springer, 1967.

Huiping Zhou, Makoto Itoh, and Toshiyuki Inagaki. Eye movement-based inference of truck driver's intent of changing lanes. *SICE Journal of Control, Measurement, and System Integration*, 2(5):291–298, 2009.