FACULTY OF MEDICINE AND PSYCHOLOGY
PhD Program: PSYCHOLOGY AND COGNITIVE SCIENCE

PhD thesis:

Sensitivity of the spatial distribution of fixations to variations in the type of task demand and its effectiveness as a trigger for adaptive automation.

PhD student:
Piero Maggi

Supervisor:
Prof. Francesco Di Nocera

Co-supervisor:
Prof. Fabio Ferlazzo

Abstract

High-risk environments such as healthcare, transport and air traffic control are characterised by highly dynamic, unpredictable, and uncertain events. A human operator's presence is necessary to monitor and control the system when critical events occur in these contexts. At the same time, the system should monitor the operators' functional status and support him when necessary. The proposed research activity investigates the spatial distribution of eye fixations as a real-time measure of mental workload. Ocular activity is known to be sensitive to variations in mental workload. Many attempts have been made to derive a stable measure of the cognitive resources allocated to a task using eye-trackers' information.

Recent studies have successfully related the distribution of eye fixations to the mental load. The scope of this research project is to devise a set of experiments for separating the contribution of three types of tasks demands (i.e., temporal, mental, and physical) and, to determine which of these (and when) should be considered for using an index of spatial distribution as a trigger in ocular-based adaptive systems.

The project has three different objectives: 1) assessing the sensitivity of the proposed measure to different types of tasks demands with a large sample and a within-subject design; 2) evaluate the effectiveness of the proposed measure as a trigger for adaptive automation and 3) using more complex algorithms to provide a more stable measurement over time and investigate variations in the frequency domain.

The first chapter provides a review of the theoretical models proposed in the literature about automation, highlighting the relationship between machine and operator, and the cognitive processes involved. The second chapter describes the physiological indicators of mental workload present in the literature, focuses on measures derived from ocular parameters such as pupillary diameter, saccades, fixations and scanpath analysis. In the last two chapters, four experimental studies are described and discussed. The aim was to evaluate how the visual exploration strategy changes with different mental workload levels and task demands. The index used to analyse the visual strategy, the nearest neighbour index, was then investigated as a trigger in an adaptive automation system. The results indicated the high diagnostic power of the measure and provided the background for future applications.

Index

Introduction: High-Risk Environments

In high-risk environments characterized by highly dynamic, unpredictable, and uncertain events, many visual elements displayed in the complex control interfaces (e.g., monitoring sensors and warning indicators) need operator's attention causing cognitive overload. The organisations should aim to minimise the risk of accidents by creating procedures that the operator must perform correctly: the "Standard Operating Procedures" (Degani & Wiener, 1994). In the context of commercial aviation, the pilot's skill to make "good decisions" is acquired through an intensive training period in a high-procedural environment. Errors caused by the operator usually occur in terms of deviation from the procedure. The deviations are not visible for operational management and are left unresolved, but these become evident following an accident. Lautman and Gallimore (1987) conducted a study of aircraft accident reports to understand the causes better. They analysed 93 fuel loss accidents that occurred between 1977 and 1984. Their study’s leading crew-caused factor was "pilot deviation from basic operational procedures" (Figure 1). A model for making operations safe emerges based on the data discussed above and on the history of industrial disasters (where the lack of procedures or rules to follow has played an important role) (Furuta et al., 2000; Dekker, 2003):

- The procedures should be the best way to perform the job.
- The procedures are based on a simple "if-then" logic.
- The security of the system is based on the following procedures.
- Organisations must invest in the knowledge of these and ensure that they are performed correctly.

However, despite rigorous procedures the ability to cope with unpredictable situations is required (which may be caused by environmental factors external to the system).

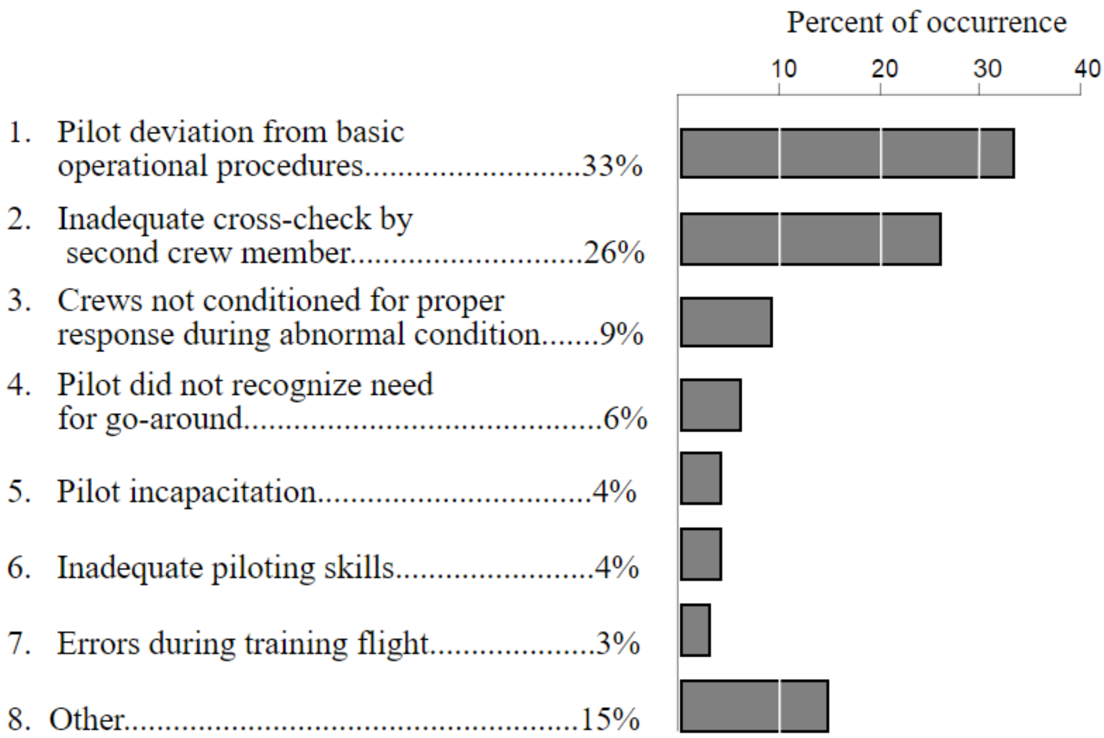

Figure 1: Significant factors caused by the crew in 93 hull damage incidents. Source: Lautman and Gallimore (1987)

The operators' ability to monitor the system is a key element. Monitoring is defined as a repeated assessment of the system status and is essential to detect, evaluate and recognise unexpected changes in many areas with high safety standards (Brookhuis et al., 2003; Metzger & Parasuraman, 2001). In recent years, inadequate system analysis has been identified as one of the main causes of loss-of-control events (Dutch Safety Board, 2010; NTSB, 2013a, 2013b). This activity is directly related to a well-known concept in literature called "Situation awareness", defined by Endsley (1988) as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future". This implies the detection of critical factors, understanding their meaning about the objectives to be achieved, and the projection of future state changes. Situation Awareness plays a crucial role in those systems where an error can led to severe consequences and, where only a human operator can perform control tasks interpreting and managing unexpected events correctly.

The restricted attentional capacity of the humans constitutes a well-known "bottleneck" that has been the object of many studies on human information processing (Tombu et. al., 2011; Marois & Ivanoff, 2005; Wolfe, Reinecke & Brawn, 2006) and on operator functional state, that is “the intrinsic relationship between human task performance and the background state of the individual.” (Hockey, Gaillard & Burov, 2003).

The high complexity of these organisations involves a multitude of aspects. In 1967, Thompson argued that "uncertainty appears to be the fundamental problem for complex organisations, and managing it is at the basis of the administrative process". Galbraith (1973) defines uncertainty as to the absence of information: "the difference between the amount of information needed to perform a task and the amount of information already in the organisation's hands". In a “decision making” context, uncertainty can be caused by incomplete information, misinterpretation of the available information and the presence of indistinct alternatives (i.e. alternatives that are equally attractive or not; Lipshitz & Strauss, 1997). It may concern the probability of an event (uncertainty of the state), lack of information about its outcome and underlying cause-effect relationships (uncertainty of the effect), or lack of knowledge about response options and their likely consequences (uncertainty of the response) (Milliken, 1987). At an individual level, the system must communicate effectively and efficiently the information required by the operator.

The main objective of studies on human performance and information processing is to reduce the possibility of cognitive overload as much as possible. In the early 90s, Wickens (1992) highlighted the importance of achieving the highest compatibility between the operator's capabilities and the characteristics of the surrounding environment. A mismatch between the machine and the operator can lead to a deterioration of performance and an increase in the workload (Gaillard & Wientjes, 1994; Hockey, 1986). To avoid this, finding real-time indexes of the operator functional state has become crucial. Also, this should be accomplished by using effective and non-intrusive tools to be applied in high-risk environments.

The use of modern technologies has caused an increase in cognitive demand and a decrease in physical demand. Tasks of monitoring and information processing have led to the need to understand how the mental workload affects the performance of the operator, a relevant issue in the study of human factors (Flemisch & Onken 2002; Loft et al., 2007; Parasuraman & Hancock 2001; Tsang & Vidulich 2006; Wickens 2008). The mental workload (MWL) is a multidimensional construct that considers different factors such as the person, the task to be performed and its environment. Young and Stanton (2005) suggest that MWL reflects the level of attentional resources required to meet performance standards, which could be affected by the task difficulty, external support, and operator experience. In this definition, attentional resources have a limited capacity over which performance degradation occurs. However, the amount of resources invested depends on the user’s willingness, who manages them to maintain an optimal performance level. The goal is to establish a relationship between MWL and performance. Probably, the workload is not optimal when the performance level is slightly below a minimum value standard. The words "Not optimal" may mean overload or underload (Brookhuis & De Waard, 2000). The first can happen when too many stimuli are present, and the operator's attention is "diverted" from the primary information.

On the contrary, the second can be caused by the absence of stimuli, leading to boredom and sleepiness. Both cause a deviation between the operator's abilities and the characteristics of the system. One of the most widespread solutions proposed to decrease a mismatch's probability is to increase computer control or automate more tasks.

Chapter 1- Automation

Historical notes

Today, there are decision-support systems, communication tools available for automation in marketing and sales processes, and suppliers' relations. All these attempts aim to automate the industry in the production and distribution processes (supply-chain), ranging from the acquisition of raw material to delivering the final product to the consumer, having as main objective to offer the customer a more excellent experience. Technology is the foundation of all these developments, causing disturbances in the industry and generating revolutions (Viswanadham, 2002).

Initially, people saw automation with scepticism. In the early 1800s, "Luddite movement" was the first significant demonstration of negative social impact caused by labour-saving machinery (Brain, 1998). During this period, British workers tried to destroy textile machinery to prevent their use in industries. In the 1950s, Diebold coined the term automation as a new word that denotes both automatic operation and the process of doing things automatically (Diebold, 1983). He wrote in 1952 the text entitled *Automation: the advent of the Automatic Factory*. The author presented his vision of the use of programmable electronic systems in the economic field. Automation became part of the manual work context and, the machines acquired the functions and operations previously performed by man. This change is supposed to benefit from eliminating many manual jobs and developing productivity in terms of material and energy savings, increasing quality and accuracy (Vagia, Transeth & Fjerdingen 2015). Sheridan and Verplank (1978) state that automation shifts the operator's role from manual work to a supervisor.

Definition

Parasuraman and Riley (1997) define “automation” as: "A machine that performs the functions previously performed by one or more humans". Underlining that automation will change and evolve. This does not replace individuals, but it changes the nature of their activities. The tasks in which the operator is engaged become primarily of monitoring. The individual has become a detector of system status, at least until unexpected events, emergencies, alarms, malfunctions, or failures occur. "Automation is proposed to reduce mental workload, increase productivity and reduce errors. Consequently, generating new types of cognitive workload and errors" (Di Nocera, 2011).

Therefore, Automation can be described as a machine agent capable of performing the functions usually performed by a person (Parasuraman & Riley, 1997). For example, a car's automatic gearbox manages to press the clutch and scaling or increasing the gear in use.

Levels of automation

The most common automation systems are called "static": the task is permanently assigned to the machine or operator without the possibility of switch roles during activity. Human-machine interaction can be expressed through one or more levels of automation (LOA). On the one hand, we have the operator who independently manages all activities (Level one: “absence of automation”). On the other hand, it is the machine that controls all phases of the production, limiting the operator to the monitoring of the system. The automation level refers to the organisation of tasks and the level of performance achieved between a human operator and a computer in controlling a complex system (Billings 1991, Kaber 1997). Sheridan (1997) discussed the various degrees of automation that have been defined in terms of system autonomy in information detection and task execution.

Sheridan and Verplank (1978) proposed a taxonomy in the context of teleoperators in submarine monitoring. This hierarchy includes various tasks to determine options and their implementation. The different LOA have been differentiated in terms of decisional functions and action selection. These have been built on the amount of information that the system must provide to the operator and the transmission mode. Sheridan and Verplank aimed to define "who" (between the operator and the computer) has the greater responsibility in the control of the system, without explicitly describing how to share the information processing between the components (Table 1).

Level of autonomy	Description	Explanation
Level 1	Fully manual control	The computer offers no assistance.
Level 2	The computer offers a complete set of decision/action alternatives	Several options are provided to the human who decides.
Level 3	The computer narrows the selection down to a few	Human still has to decide.
Level 4	The computer suggests one alternative	Human decides amongst suggestions.
Level 5	The computer executes that suggestion if the human approves	Human approval needed for execution.
Level 6	The computer allows the human a restricted time to veto before automatic execution	Limited time for veto given to the human.
Level 7	The computer executes automatically, then necessarily informs the human	No human interference, just information at the end.
Level 8	The computer informs the human only if asked	Human gets information only if asks.
Level 9	The computer informs the human only if it decides to	Computer decides whether to give information.
Level 10	Fully autonomous Control	The computer decides everything and acts autonomously, ignoring the human.

Table 1: Taxonomy proposed by Sheridan and Verplank (1978)

Subsequently, in 1987, Endsley developed a taxonomy relating to support systems for human decision-making processes. This hierarchy established that a task could be performed using:

1. **manual control**, without assistance.
2. **decision support** provides by system through recommendations.
3. **consensual artificial intelligence** (AI), where the system selects the action to perform after authorisation by the operator.
4. **monitored artificial intelligence**, where the system automatically acts unless the operator blocks it.
5. **complete automation** without operator interaction.

This list applies to cognitive tasks, where the operator's ability to make decisions is critical to overall performance. Subsequently, Ntuen and Park (1988) developed a five-level taxonomy in the context of teleoperations. Both these taxonomies can be considered like the hierarchy levels provided by Sheridan and Verplank. Based on this work, Endsley and Kaber (1997, 1999) propose a 10-levels taxonomy to provide broader applicability with tasks that require real-time control within several domains, including air traffic control, aircraft piloting, advanced production and teleoperations. These domains have many common features: multiple competing goals; multiple tasks competing for an operator's attention, each with different relevance to the system's goals; high demands on jobs with limited time resources. Four functions related to these domains at the basis of the taxonomy in question have therefore been identified:

- **monitoring** - acquire all relevant information to perceive the state of the system correctly.
- **generation** - formulation of options and strategies for achieving the objective.
- **selection** - choice of action or strategy.
- **execution** - implementation of the selected operation through the control interface.

So, ten levels of automation have been built to assign these functions to human or computer or a combination of the two:

1. **manual** - The operator performs all the tasks of monitoring the system's status, defining possible options, selecting and achieving them.
2. **action support** - The operator monitors the system, which assists him in acting.
3. **block processing** - The operator selects several options that will then be performed automatically by the machine. Automation consists of the physical execution of the tasks.
4. **shared control** - Both operator and machine generate several choices. However, the operator chooses the final option that will be carried out shared with the device.
5. **decision support** - The computer generates a list of options from which the operator can choose or develop his options. Once the human has selected an alternative, the computer performs it. This level represents many expert systems or decision support systems that guide the possibilities that the operator can use or ignore when performing a task. This level is indicative of a decision support system that is also capable of performing tasks, while the previous level (*shared control*) is not.
6. **mixed decision making** - The computer generates a list of options, which it selects and executes after the operator's consent. Besides, the operator can choose a different option that will be completed by the machine.
7. **rigid system** - The computer provides a limited set of actions from which the operator must choose, without the possibility of generating others. The machine will then perform the selected action.
8. **automated decision making** - The system selects and performs the best option from a list of alternatives that it generates (increased by the possibilities suggested by the operator).
9. **supervisory control** - The system generates, selects and implements the action. The human operator monitors the three steps and can intervene by developing and selecting a different option.
10. **complete automation** - The system performs all the actions. The man is entirely out of the control circuit and cannot intervene. This level is representative of a fully automated system in which human processing is not considered necessary.

This taxonomy provides some advantages over the previous ones. It offers numerous combinations of the four functions mentioned above and on which it is based. For each level, the "who does what" is well defined and is applicable in several contexts (since they are not built for a specific sector).

Choosing the right LOA is a complex process that considers several factors such as human performance, reliability, and automation costs (Parasuraman et al. 2000). In the context of Human-centred design, the effect of this choice on information processing and performance is relevant, regardless of the application context. The human operator's ability to detect and take over after a failure is the crucial point of the out-of-the-loop (OOTL) error problem. Therefore, it should be central in decisions regarding LOA choices. OOTL issues are characterised by a decrease in the operator's ability to intervene in the system control loop and take over manual control if necessary (Louw & Merat, 2017; Endsley & Kiris, 1995). The concept of adaptive automation (AA) has been introduced to overcome limits due to applying a single LOA. AA is also defined as the dynamic allocation of functions. If the user is going to overload, a controlled mechanism to dynamically balance the work between user and machine should reduce the attentional demand imposed by the task on the operator and optimise the performance. In other words, the "division of labour" is dynamically assigned according to the task requests and the user's capabilities to achieve an optimal level of performance (Byrne & Parasuraman, 1996). This system is designed to be the best solution to the problem of allocation of functions and tasks, overcoming the limits linked to static automation. Dynamic assignment necessarily requires studies to be evaluated by trying to answer four questions:

1. What is automated?
2. How is it automated?
3. What task can be shared between the machine and the operator?
4. When do these parameters have to change?

The system must measure performance values, the user's capabilities, and mental state to implement these features. In the literature, it is possible to trace different activation criteria, also defined as "triggers," to adapt the system based on the theoretical model used.

Adaptive automation

It is possible to highlight three main models (Parasuraman et al., 1992; Rouse, 1988):

1. **Critical event model**: If an "x" event occurs, the system should perform the corresponding "x" action. An example of adaptive automation based on critical events is conflict detection aids for air traffic controllers. The load imposed by the task and/or

the traffic complexity can be predicted partially (e.g., based on the number of aircraft in the analysed area and other "dynamic density" measures; Smith et al., 1998). So, the system could be adaptively provided such aids only in case of high traffic. However, this model is limited because it cannot always anticipate critical events. Another disadvantage of this method is its possible insensitivity to the human operator's real task resolution performance. The critical event model will invoke automation regardless of whether the operator requires help (e.g., due to high workload) when the critical event occurs (Hancock & Szalma, 2008).

2. **Programmed model**: automation is programmed a priori based on a model relating to optimal operator performance (Inagaki, 2003; Parasuraman, Bahri, Deaton, Morrison, & Barnes, 1992; Prinzel, 2003).
3. **Continuously measured model**: automation is adjusted based on continuous monitoring of the operator's mental and physical state (Byrne & Parasuraman, 1996), based on the performance level (Kaber & Endsley, 2004), or based on a combination of both (Wilson & Russell, 2003).

Conceptually the three models may have aspects in common and a combination of these, through hybrid logic, has been recommended as the best way to create an adaptive system (Parasuraman et al., 1992).

Subsequently, Scerbo (1996) in the review "Theoretical Perspectives on Adaptive Automation" describes several adaptability strategies proposed by the literature.

- **Performance**: the operator's interactions with the system can be monitored and evaluated into a standard to determine when to change the automation level. The main issue concerns the definition of a reference standard against which to assess operator performance. To apply this, it is necessary to build a performance model. Several authors (Geddes, 1985; Rouse, Geddes, & Curry, 1987) propose a model that considers the system's current state, external events, and expected operator actions.
- **Workload**: Hancock and Chignell (1988) provide a model where tasks are assigned to the operator or system based on present and future workload levels. Again, current workload levels are determined in part by deviations from an ideal model.
- **Biometric measurements**: In this model proposed by Morrison and Gluckman (1994), the idea is that physiological signals reflecting the activity of the autonomic nervous system (e.g., heartbeat, galvanic response, pupil diameter) or the brain presumably change concerning the workload level and could be used as automation activators.

Several important factors need to be considered to develop the models, including the system's reactivity. The system's response time is a determining element in usability and user acceptance (Bailey, 1982). This becomes more critical in adaptive automation. Successful automation will require adequate timing, as well as the right amount of automated processes. Under high-risk conditions, instant system response is needed. Several authors argue that such speed is only possible by predicting the system's future state in terms of performance level and workload (Greenstein & Revesman, 1986; Rouse, Geddes, & Curry, 1987).

Chapter 2 - Eye movement

Physiological workload indicators

Several psychophysiological measures have been studied for identifying new real-time indicators of operator functional state. Mehler, Reimer, Coughlin, and Dusek (2009) conducted a research study on a sample of 121 participants with the aim of examining the sensitivity of parameters such as heart rate variability, skin conductance and respiratory rate as continuous measures of workload in a simulated driving environment. The analysis showed a significant effect of the difficulty level in all the psychophysiological parameters. A more recent study (Pakarinen, Korpela, Torniainen, Laarni, & Karvonen, 2018) examined the relationship between the mental workload and the response to the physiological stress of nuclear power plants operators, who were assigned the task of managing the simulation of a large-scale accident through the control room. Records of heart rate and heart rate variability (respectively) were used to measure stress on a sample of 22 volunteer operators. The results showed a relation between the psychophysiological measures and the increase in workload experienced during a high accident risk scenario. In addition, these findings confirmed the data from self-report measures (NASA-TLX) and corroborated previous research results (Bernardi et. al., 2000; Reimer & Mehler, 2011; Hwang et al., 2008). More specifically, many studies in the literature have investigated the relationship between neurophysiological measures and mental workload, such as electroencephalography (EEG) (Brookings, Wilson, & Swain, 1996; Gevins & Smith, 2003; Borghini et al., 2014), functional Near-InfraRed Spectroscopy (fNIRS), and functional Magnetic Resonance Imaging (fMRI) (Gabbard, Fendley, Dar, Warren, & Kashou, 2017; Liu, Ayaz & Shewokis, 2017; Ranchet, Morgan, Akinwuntan & Devos, 2017). In a study by Aricò et. al. (2016), a workload index based on EEG measurements was used as a trigger in an adaptive automation system implemented in a realistic Air Traffic Control Environment.

Eye movements as a behavioural measure

With the continuous development of technology and automation, human information processing has changed. Pilots, for example, have different tools available to monitor the environment. The task of monitoring and interpreting these information has become primary compared to the past. In the absence of these devices, the operator had to independently acquire this information by placing the attention outside the cockpit. In the aviation sector, visual data processing remains one of the critical elements of the system's safety and effectiveness.

The oculomotor system provides a lot of information about the person's cognitive processes and mental effort during a task's execution. Eye trackers allow the measurement of the ocular parameters: the direction of gaze, pupil diameter changes, and the eyeblinks. Analyses of this data provide various numerical indices (e.g., number of fixations and amplitude of the saccades) and allow us to obtain graphical representations (Figure 2) that can be used to interpret a person's behaviour. Moreover, it is essential to emphasise that research in this field has continued to evolve thanks to continuous technological advances, which have led to increasingly advanced, less intrusive, and more accurate ocular activity monitoring tools (Marshall, 2007).

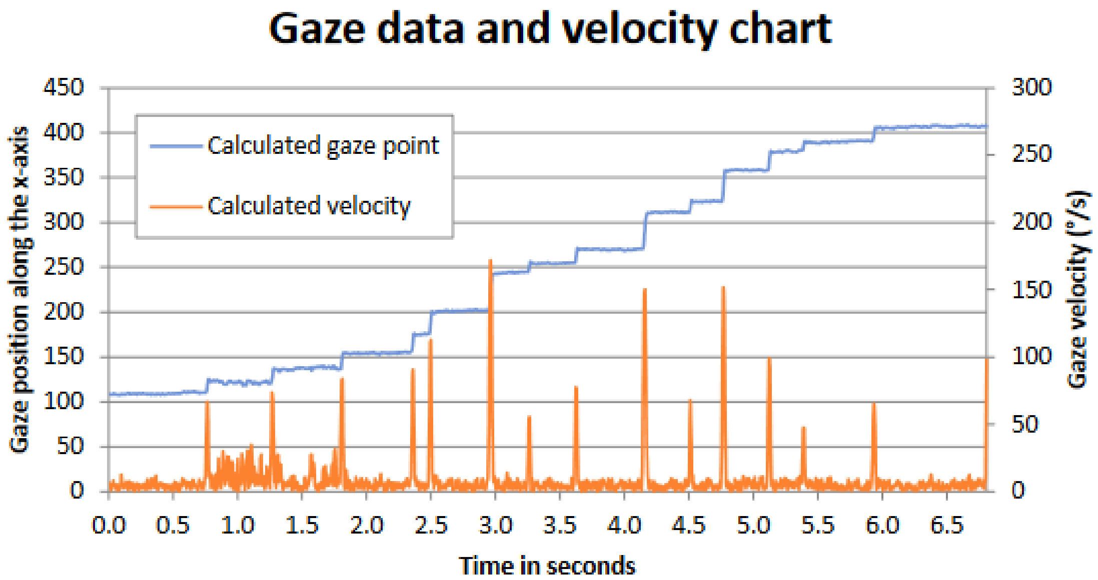

Figure 2: Speed and position of gaze-points in relation to the x-axis. Peaks in the calculated speed can be used to identify the saccades between one fixation and the next.

Several attempts have been made to find an ocular index as a continuous measure of the operator's mental state. Below, the most common among these has been analyses, highlighting the strengths and weaknesses of each.

Pupillary diameter and workload

Several authors have shown that pupil dilation may be related to cognitive processing and to the mental effort required to perform a given task (Othman & Romli, 2016; van der Wel & van Steenbergen, 2018; Kosch, 2018). This relation has been analysed in various tasks including short-term memory (Peavler, 1974) and visual search tasks (Porter, Troscianko, & Gilchrist, 2007), but also in air traffic control (Hilburn et al., 1997) and driving (Razaei and Klette, 2011). Just and Carpenter (1993) identified changes in pupil diameter when understanding single sentences with different degrees of difficulty. Iqbal, Adamczyk, Zheng, and Bailey (2004) confirmed a correlation between pupil variation and mental workload, when participants were asked to perform various tasks including text comprehension, mathematical reasoning, target stimulus research and object manipulation. Already in 1966, Beatty and Kahneman had detected a 10% dilation of the pupil diameter during the first half-second after the presentation of a stimulus (corresponding to a familiar name) that required the subject to issue a verbal response (the corresponding phone number). The two authors performed an accurate analysis of pupil diameter variation. Beatty defined it: "pupillometric analysis of task-evoked pupillary response in a short-term memory task" (1982, p. 277). The analysis was conducted by asking subjects to write sequences of numbers of different length (from 3 to 7 digits) presented a few seconds earlier by auditory means. The pupil diameter graph clearly showed an increase during the presentation of the stimuli and a decrease during the writing of digiting. Also, the pupil diameter variation occurred as a function of the stimulus's length (greater when they had to rewrite seven digits; Kahneman & Beatty, 1966). Later Just and Carpenter (1993) identified pupil diameter variations during the understanding of single sentences with different degrees of difficulty. Iqbal and colleagues (2004) confirmed a correlation between pupil variation and mental workload: subjects were required to perform various tasks, including understanding a text, mathematical reasoning, finding a target stimulus and manipulating objects. In an application context, it was observed that this measure increases as a function of the density of targets present in the visual scene (Van Orden et al., 2001). In 2010, Palinko and colleagues associated the variation in pupil diameter with driving activity, finding an increase in pupil diameter when subjects were required to drive with a simulator compared to when they sat on the passenger seat.

The pupil diameter has been reported to be very promising as a measure of mental workload. However, an important limitation of this measure is the difficulty of keeping constant the brightness of the environment in which the task is performed. The amount of light reaching the eye causes rapid changes in pupil diameter and this can limit the benefits of this metric in working environments. In fact, unlike controlled laboratory settings, the brightness of the working environment (e.g. the brightness of the displays or the room) is variable. Changes in pupil diameter may be due partly to physical components in the environment and partly to workload, making it difficult to isolate a valid and reliable measure of cognitive effort.

Eyeblink and workload

Eyeblink describes the rapid closure of the eyelids and provides three metrics: frequency, duration and latency. In 1988 Stern described the blink with the term "mental punctuation". Specifying that is necessary at the end of a cognitive process, for the analysis of a stimulus. He was arguing how this parameter is influenced by the complexity of the information presented. Starting from these arguments Boehm-Davis, Gray and Schoelles (2000) examine the relationship between the difficulty of the task and the frequency of eyelid closure, using a sample of 64 university students. The researchers hypothesise that the eyeblink frequency is reduced during the information processing and increases subsequently. The subjects had to monitor a series of elements present within a radar and acquire specific information from each one with the mouse cursor. The participant had to correctly classify the elements by assigning each one a risk value from 1 to 7. The number of elements was manipulated to create two conditions with different difficulty: 9 targets in the first and 18 in the second. The results confirm the hypothesis, suggesting that the eyeblink is suppressed when the task has a large amount of information to analyse. It is also noted that the frequency of eyeblink increases after the end of the cognitive processing phase compared to baseline values. The literature describes three types of eyeblink: corneal, voluntary and endogenous reflex. The endogenous blink is spontaneous and is not caused by environmental stimuli (Neumann and Lipp, 2002). It is possible to measure the frequency and duration of this type of eyeblink. Generally, the former decreases in conditions with high visual and/or cognitive load, such as driving a car in the city rather than on the highway (Pfaff, Fruhstorfer, & Peter, 1976). Considering that eyeblink inevitably causes the interruption of visual inputs, the reduced frequency of eyeblink allows for the better and constant analysis of the information present in a specific area of space, especially in high mental workload conditions. For the same reason, the duration of the blink also decreases as the visual load increases. In 1987 Wilson and colleagues observed that the duration is shorter during a real flight execution rather than simulated. In 1988, Sirevaag et al. found lower blink duration values in the double task condition than in the single-task condition

(Wilson, Purvis, Skelly Fullenkamp, & Davis, 1987; Sirevaag, Kramer, DeJong, & Mecklinger, 1988). Conversely, blink frequency increases when subjects need to move the gaze from one instrument to another in rapid sequence. The closing of the eyelids helps to point out the end of acquiring an information set (Fogarty & Stern, 1989). Wierwille, Rahimi, & Casali (1985) observe an increase in frequency with an increase in navigation demand during a flight mission simulation. Orchard and Stern (1991) argue that an endogenous blink is directly related to perception and information processing. If the task requires a high level of attention, the blink frequency will decrease. Also, if a motor action has to be performed to a specific stimulus, the blink may be inhibited until it is completed. Stern et al. (1984) observe that an increase in latency in the blink corresponds to increased mental workload. The delay in the corneal reflex, following the presentation of a relevant stimulus, increases according to the volume of information to remember (Bauer, Goldstein, & Stern, 1987). Also, the delay increases when the subject has to respond to a sound discrimination task (Goldstein, Walrath, Stern, & Strock, 1985). However, blink also occurs as a pupillary reflection, involuntarily, following a sudden and sufficiently intense stimulus such as loud noises or a flash. Pupillary diameter and blink are measurements directly provided by the most common eye-tracker models. Their use in an adaptive system does not require raw data processing algorithms. However, it is not possible to do the same when we talk about Fixations and Saccades, since these measurements are obtained by analysing the distribution of individual gaze-points defined as the projection on the visual scene, having coordinate 'x' and 'y', given by the direction of the gaze.

Fixations and Saccades

The movements of the ocular muscles bring specific portions of the image to the fovea. The foveal area offers a high visual acuity and allows to capture image details very accurately. Therefore, it is possible to distinguish between periods in which the eyes’ position is kept stable, defined by the term "Fixation", and periods in which there are rapid eye movements that direct the gaze toward a new area, defined as "Saccade". The saccade is considered one of the fastest movements of the human body, that varies in amplitude, duration and maximum angular velocity (defined as peak velocity). The relationship of these three parameters has been defined as "the main sequence", indicating that the peak velocity value and the saccade duration increase systematically with the same amplitude (Di Stasi et al. 2010). Looking at a graphical representation (Figure 2) of an eye-tracker’s raw data, it is relatively easy to identify and distinguish these two events. However, the high sampling rates and the huge amount of raw data provided by eye-trackers make it necessary to use specific software to analyse and apply automatic classification algorithms. These consist of search and categorisation rules. The first aims to split the fastest "periods" (saccades) from the slowest "periods" (fixations). The second aims to accept or reject, divide and/or merge possible events (saccades and fixations) based on a set of criteria (e.g. minimum time of fixation or maximum duration of a saccade) (Hessels, R. S., Niehorster, D. C., Kemner, C., & Hooge, I. T., 2017). For example, speed-based algorithms emphasise this value, assuming that fixations are positively correlated with a slowing down of movements (Salvucci, 2000). The criteria can be chosen based on the algorithm used (Holmqvist et al., 2011) or directly by the experimenter based on the task, such as reading a text or a visual research task (Rayner, 2009). The analysis and, consequently, the raw data reduction is useful for the following reasons: during a saccadic movement, it is not possible to process visual information. For this reason, the image portions where saccades are performed are typically irrelevant in the application of many searches. In addition, micro-movements recorded during a fixation such as a tremor or rapid movements away from the focus (flicks), often count for little during high-level analysis (Salvucci, 2000). The mathematical definition of Fixation and Saccade allowed researchers to study new indexes and then observe how they vary according to mental workload.

Visual search strategy and workload

Fixations' duration, frequency, amplitude, and velocity of the saccades constitute the entire visual research strategy adopted to complete a specific task. In 2009, De Greef et al. explored how these parameters vary according to the system's mental workload. The study involved 18 subjects, who were asked to identify themselves in an operator's role, responsible for managing a combat workstation, on board a military ship. The interface included a schematic visual representation of the ship's area, built with real data from a radar system. Each subject had to monitor, classify and identify the different elements present in the scene (aeroplanes and/or ships). Each time an element was heading towards "its" position (corresponding to the centre of the radar), thus presenting a hostile behaviour, the operator was required to perform three tasks: Acquisition of information of the single element (Track), partly provided by the radar and partly obtained through the communication tools that allowed the subject to get in contact with the "Track", the air control or the coast guard. With this information, the subject had to finally classify the element as a friend, enemy or neutral. If it was confirmed as "enemy", it was necessary to perform the last action of engaging the weapons and, consequently, its elimination. To create three mental workload conditions, the researchers manipulated the number of elements present in the radar and the percentage of these that had a special behaviour, forcing the user to perform more operations. The authors confirm a significant difference in the subjects' mental workload in all three conditions (under load, normal and overload). Consequently, the results show a positive correlation of the workload with the fixations duration, as confirmed by previous researches (Tole, J.R., Harris, R.L., Stephens, A.T., Ephrath, A.R., 1982; Callan, 1998). More recently, Di Stasi et al. (2016) investigated the effects of simulated flights' duration on the speed of saccadic movements made by pilots. The objective was to investigate how the saccades' speed varies according to the fatigue experienced by the subjects during long periods of flight, and therefore the potential application of this parameter to evaluate the physical and mental state of the operator before and during flight operations. To verify this hypothesis, a sample of 26 pilots was used, divided into two experimental groups who were asked to perform a simulated flight with an average duration of 54 minutes for the first group, and 109 minutes for the second group. The results show that the time spent on the task increases together to the subject's perceived fatigue and so, the speed of the saccadic movements decreases. This suggests that ocular metric can be a behavioural index of fatigue (Di Stasi et al., 2016). The study takes into consideration separately the parameters derived from eye movements. However, some research shows a systematic relationship between the fixations duration and the saccadic amplitude. In 1974, Antes was one of the first researchers to analyze how the duration of fixations and saccades varies during the free exploration of an image. He noted that the fixations duration increases and the saccade amplitude decreased in the first seconds of observation. Karpov, Luria and Yarbus (1968) attribute this effect to initial and final phases of global visual exploration in the same period. These results are also explained as an adaptation in the exploration strategy used to perform the task. Similar results were reported by Galpin and Underwood (2005) in a comparative visual research task. The authors proposed that the initial model of short fixations and long saccadic functions find an optimal point in the visual scene to start the search task. The relationship between fixations duration and saccadic amplitude is of particular interest for its possible diagnostic value related to the two main modes of visual processing (Unema et al., 2005; Velichkovsky, Dornhoefer, Pannasch, & Unema, 2000; Velichkovsky, Rothert, Kopf, Dornhoefer, & Joos, 2002). The experiments carried out by Velichkovsky in 2005 show that the differences in the mode of recognition of visual stimuli are manifested in the fixations duration and the amplitude of the saccades adjacent to them. Therefore, the results show that a higher performance was correlated with longer fixation times and saccadic with amplitude < 5°, rather than short fixations and large saccades.

Scanpath analysis

Regarding scanpath, fewer researchers have investigated the ocular activity about other factors, such as mental workload. The topography of the visual scanning, as well as its dynamics, was quantitatively approached in two studies by Tole, Stephens, Vivaudou, Ephrath & Young (1983) and Harris, Glover & Spady (1986) who suggested using the entropy rate of the visual scanning for discriminating between different levels of mental workload. Their results suggested that scanpath tends to be cluttered and random when the workload is low. Instead, it would become regular and predictable as the demand increases. Although very appealing, entropy has seldom been used as a measure of workload and therefore its properties have not been properly tested.

Moreover, entropy is limited by the need to rely on predefined Areas Of Interest (AOIs) to compute transitions between them: in many operational settings visual scanning happens outside specific AOIs, or the AOIs change dynamically. To overcome this limit, Di Nocera, Camilli and Terenzi (2007), introduced an alternative approach based on the spatial distribution analysis of fixations through the Nearest neighbour Index (NNI). The spatial distribution appeared sensitivity to changes in mental workload. Earlier studies on its functional significance suggest that scanpath appears to be more scattered when the temporal demand increases (as the time pressure). On the other hand, a higher concentration of fixations in specific areas depends on the visuospatial demand (Camilli, Nacchia, Terenzi & Di Nocera, 2008). It is similar to the concept of an ordered scanpath provided by the entropy rate measure. Entropy is based on transitions between AOIs and has been applied in scenarios with high visuospatial demand. However, the two metrics have never been compared. The analytical details of the two approaches are described in the following sections.

Entropy rate

Entropy can be defined as a measure of the disorder found in any physical system and this concept was then applied by Tole et al. (1983) to eye movements. When the individual looks at all the quadrants in the scene and crosses all the potential combinations of stimuli with a stable frequency, the entropy will increase. Instead, the entropy value will be lower when the individual focuses attention on a narrower range of possible areas of interest. That happens because the frequency of transitions from one area to another decreases. A regular and systematic visual exploration strategy is shown in a condition of low entropy, which corresponds to a more orderly passage to other areas. The principal benefit of this analysis is the possibility to "summarize" the visual strategy using a single value. The first step in estimating the amount of entropy is to identify the areas of interest in the visual field, and then computing the proportion of time taken by the participant to look at each of these areas:

$$\boldsymbol{Entropy\,rate} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{D}} [(\boldsymbol{E}/\boldsymbol{E\_max})/\boldsymbol{DT_i}]$$

$$\boldsymbol{E} = -\sum_{\boldsymbol{i=1}}^{\boldsymbol{D}} \boldsymbol{P_i log_2 P_i}$$

where E represents the value of the observed average entropy, E _max is the maximum entropy value computed from the total number of AOIs in the scene (it constitutes the entropy value when all AOIs are accessed with the same probability), Pi represents the probability that the sequence i occurs, DTi is the average duration of fixation for the i-th sequence when the individual is visually exploring the scene, and D expresses the number of the distinct sequences in the scanpath. The index is indicated in bits/second.

NNI

The Nearest Neighbour Index (NNI) provides data on the distribution of points in space. The average distance between the fixations collected during the execution of a task and the average distance between the fixations expected in a random distribution are taken into account in the application of the NNI to eye movements. The result is represented by a single value where 1 indicates that the empirical and the random distribution are not different; values above 1 indicate dispersion, while values below 1 show clustering. The index can be computed for small epochs if sufficient fixations are available (about 50 as a rule of thumb) and then analyzed as a time series, therefore offering information on the temporal variations of distribution of fixation points. A methodological study (Camilli, Terenzi & Di Nocera, 2007) supports the validity of this algorithm as a measure of mental workload, highlighting the consistency of the index with subjective and psychophysiological measures. To estimate the index it is first necessary to calculate the Nearest Neighbour distance or d(NN):

$$d(NN) = \sum_{i=1}^{N} [min \frac{(dij)}{N}], \qquad 1 \le j \le N, \qquad j \ne i$$

where min (dij) represents the distance existing between each point i and the nearest point j (with the j value between 1 and N and different from i), and N corresponds to the number of points in the distribution. The next step is calculating the average random distance or d(ran) to obtain the second element of the equation; this value would represent the value of d(NN), supposing that the distribution of the points were totally random:

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

where A indicates the polygon, area delineated by the most extreme fixations and N represents the number of points. The NNI value is then calculated by dividing the Nearest Neighbour Index distance, d(NN), by the average random distance, d(ran):

$$NNI = \frac{d(NN)}{d(ran)}$$

Chapter 3 - Experimental studies

Study one

The objective of this experiment was to test the sensitivity and diagnosticity (see Wierwille & Eggemeier, 1993) of the NNI, that is how changes in the visual exploration strategy due to different workload levels and different types of demand imposed by the task are captured by the distribution of fixations. This aspect has been previously approached by Camilli, Terenzi and Di Nocera (2008) in a between-subject design, comparing the effects of the mental and temporal demands on the distribution of fixations.

Tools and software

Experimental software development

The Tetris game used in this study was coded using JavaScript and GoogleScript. The gaming area consisted of 300 cells deployed on a grid of 15 columns by 20 rows. Each tetromino (piece) was randomly extracted from a pool composed of 7 different tetrominos types and it descended at a constant speed. With the aim of creating three experimental conditions, specific variables have been modified to induce a different type of task demand. In Condition 1, speed of falling pieces has been manipulated to generate time pressure (temporal demand); in Condition 2, direction of pieces has been reversed to increase mental demand (each piece appears in the lower part of the game area and then rises to the top); in Condition 3, the interaction with pieces was occasionally blocked, therefore forcing the user to press the control keys several times to move the pieces (physical demand).

The manipulations were coherent with the NASA-TLX definition of mental, temporal and physical demand. Mental demand: "How much mental and perceptual activity was required? Was the task easy or demanding, simple or complex?"; Temporal demand: "How much time pressure did you feel due to the pace at which the tasks or task elements occurred? Was the pace slow or rapid?"; Physical demand: "How much physical activity was required? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?". Therefore, we consider the visuo-spatial demand imposed here as an expression of the mental demand.

In our version of Tetris, the "game-over" consisted of the exhaustion of the playing area given by the excessive accumulation of pieces but did not represent the end of the game. When the event occurs, the program automatically resets the entire area deleting all the accumulated pieces and allowing the user to continue the game until the end of the experiment. The number of pieces accommodated and the number of completed lines were used as performance measures. The number and shape of the pieces, the size of the playing area, and the difficulty between levels were based on the original version of the Tetris.

Ocular activity recordings

The Gazepoint GP3HD eye-tracking system was used to record ocular activity. This system allows the researcher to collect ocular data without using invasive and/or uncomfortable head-mounted instruments. Gazepoint, the eye tracker manufacturer, claims accuracy within 0.5 to 1.0 degrees and reads data at a rate of 150Hz. The eye tracker was calibrated using the default 9-point calibration test using Gazepoint's included software.

Participants

Thirty university students (19 women and 11 males, M = 25 years old, SD = 3.6) volunteered and participated in the experiment. All participants had a normal or corrected-to-normal vision and were naïve as to the aims of the experiment. This research study was completed with the tenets of the Declaration of Helsinki and was approved by the Institutional Review Board of the Department of Psychology, Sapienza University of Rome. Informed consent was obtained from each participant. Participants received a € 20.00 worth bookstore gift card. One subject was excluded from the data analysis due to the low quality of recorded eye movements.

Procedure

Participants were tested in a within-subject design in which the same task was manipulated -in three different sessions- for manipulating the mental, the temporal, and the physical demand. Participants played a custom coded version of the Tetris game, a commonly known tile-matching puzzle videogame successfully used in a variety of studies (e.g., Trimmel & Huber, 1998). For experimental purposes, the game restarted from a blank screen each time the stack of Tetrominoes reached the top of the gaming area and no new Tetrominoes were able to enter. This condition commonly denotes the end of the game, whereas in this experiment it was scored as a loss (performance measure). Participants were instructed to gain as many points as possible (i.e., complete lines and avoid losses).

Training session

Before the experimental session started, each participant performed a training session, whose scope was to familiarize the participants with the experimental setting. To this aim, each participant played the Tetris game starting from a low difficulty level and moving on to Baseline, Temporal demand (TD), Mental demand (MD) and Physical demand (PD) conditions. The training had a 5-minute duration and did not include the evaluation of the participants' performance level in this phase. The scheme of the training session is reported below:

- One minute of gameplay at Level 1 (drop speed: 1250 ms per block), with the aim of verifying the correct understanding of the game rules and allowing the participant to familiarize with the use of directional keys.
- Baseline condition: one minute, configured at level 6 (drop speed: 208 ms per block). It was used to acquire the baseline for the experimental session.
- TD condition: one minute, set at level 8 (drop speed: 156 ms per block).
- MD condition: one minute, during which the entire playing area was rotated by 180° and each Tetromino appeared on the bottom side and went up, accumulating on the top of the gaming area.
- PD condition: one minute, in which the participant needed to press the directional keys repeatedly to move the piece quickly in the chosen direction (instead of keeping the key pressed).

The difficulty level of the conditions, as determined by the drop speed of the pieces, was defined on the basis of previous studies (Camilli, Terenzi & Di Nocera, 2008; Camilli, Terenzi & Di Nocera, 2007).

Experimental session

After the calibration of the eye-tracker, participants were instructed to play the game earning as many points as possible (i.e., complete lines and avoid losses). Each condition lasted 10 minutes and the order of presentation was randomized across participants. After completing each condition (Baseline vs. TD vs. MD vs. PD), participants were requested to fill in the NASA-TLX (Hart & Staveland, 1988).

Data analysis and results

Performance data

A performance index was computed based on the number of lines completed in relation to the maximum number of lines that could be completed. The maximum value is obtained by the total number of Tetrominoes that the participant managed in each condition (For example, with 60 pieces it was possible to complete a total of 16 lines if managed in an optimal way). The index goes from 0 to 1, where 1 means that the player has obtained the maximum achievable score. The performance index was used as dependent variable in a repeated measures ANOVA design, using Condition (Baseline vs. TD vs. MD vs. PD) as repeated factor. Results showed a main effect of the condition [$F_{3, 84} = 16.88$, $p < .001$] (Figure 3). The faster (TD) and Reversal conditions (MD) were associated with the worse performance with respect to the baseline (Table 2). Overall, performance is low in all conditions. This could indicate either a limited ability of the subjects in the Tetris game or a wrong selection of difficulty levels (described in the previous paragraph). However, the baseline is easier than the experimental conditions (excluding the PD condition), which allows us to read the results obtained from the subjective and ocular measures in line with the starting hypotheses.

	TD	MD	PD
Baseline	.001	.011	.285
TD		.01	.001
MD			.001

*Table 2: Post hoc analysis carried out through the Duncan test. Pairwise comparison among PI scores and conditions (*p < .05).*

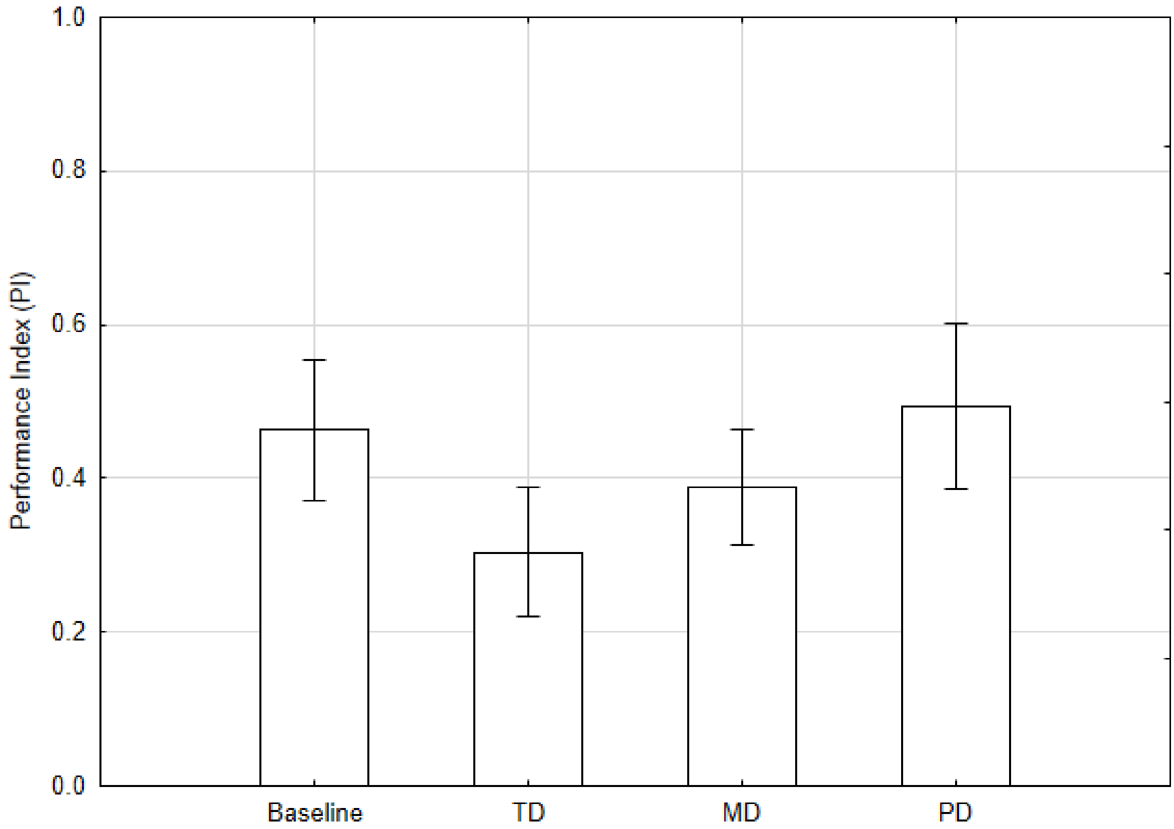

Figure 3: The performance index shows the overall strategy used by participants. It is the ratio of the number of completed lines to the number of pieces that appeared during the game. Values close to 1 means an optimal game with a high number of lines completed.

Subjective measure

NASA-TLX weighted ratings were used as dependent variables in a repeated measures ANOVA design using Condition as repeated factor. Results showed a main effect of Condition [$F_{3, 84} = 11.11$, $p < .001$] (Figure 4, Table 3), consistent with those obtained for the performance index. Although analyses on the single items are questionable from a statistical standpoint, it is worth noting that TD, MD, and PD conditions showed higher values for temporal, mental and physical demand scales respectively (Figure 5, Table 4-6). These results show that the manipulations made with the Tetris have indeed taxed specific aspects or resources.

	TD	MD	PD
Baseline	.001	.001	.51
TD		.153	.001
MD			.01

*Table 3: Post hoc analysis carried out through the Duncan test. Pairwise comparison among NASA-tlx scores and conditions (*p < .05).*

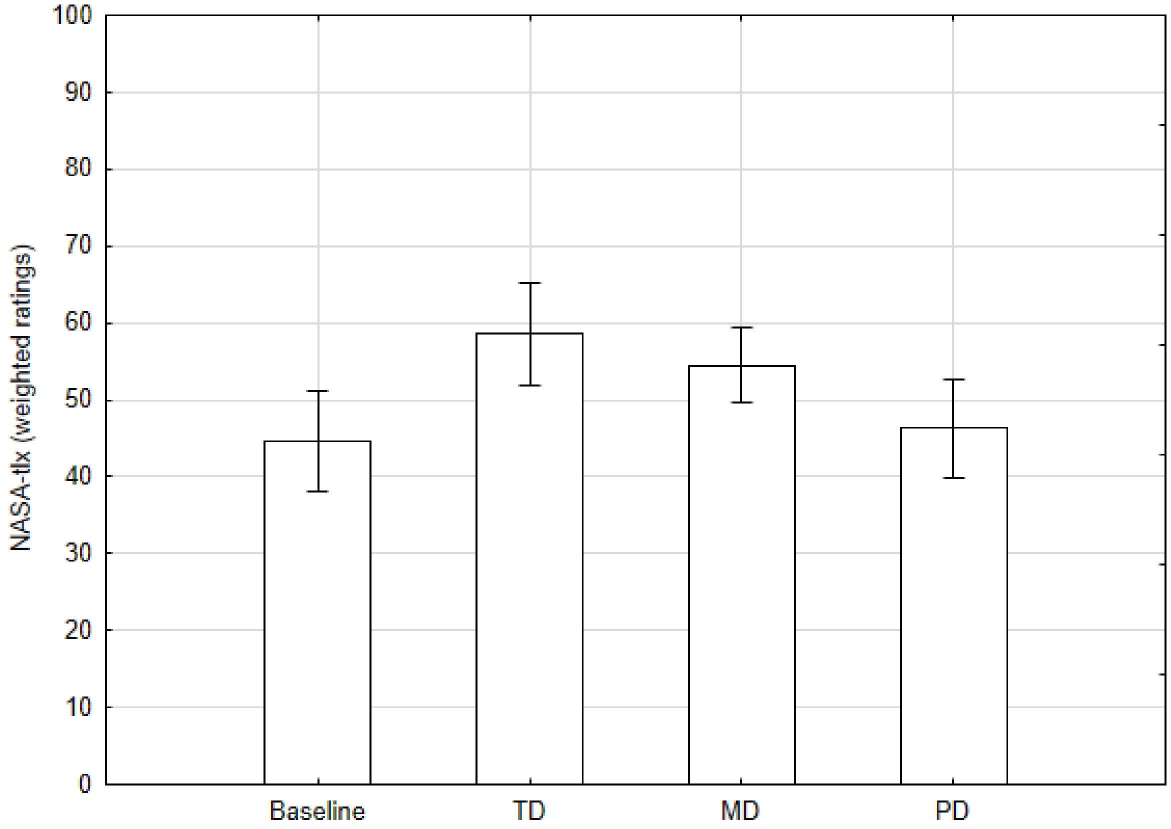

Figure 4: NASA-tlx values (weighted scores) separately for the conditions. Error bars denote .95 confidence intervals.

	TD	MD	PD
Baseline	.001	.829	.917
TD		.001	.001
MD			.765

*Table 4: Post hoc analysis carried out through the Duncan test. Pairwise comparison among Temporal demand scale scores of NASA-tlx and conditions (*p < .05).*

	TD	MD	PD
Baseline	.05	.001	.612
TD		.01	.016
MD			.001

*Table 5: Post hoc analysis carried out through the Duncan test. Pairwise comparison among Mental demand scale scores of NASA-tlx and conditions (*p < .05).*

	TD	MD	PD
Baseline	.882	.66	.001
TD		.583	.001
MD			.001

*Table 6: Post hoc analysis carried out through the Duncan test. Pairwise comparison among Physical demand scale scores of NASA-tlx and conditions (*p < .05).*

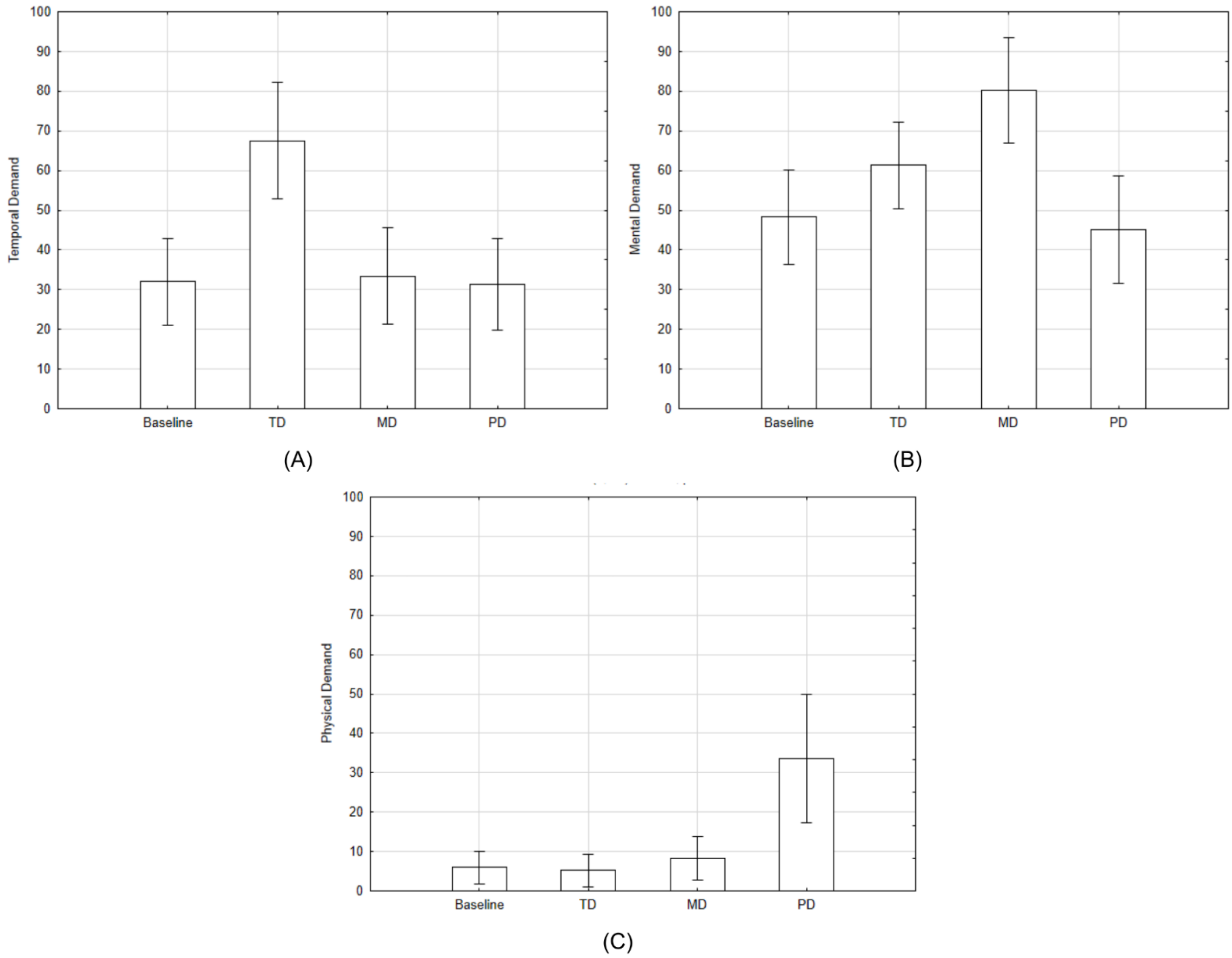

Figure 5: NASA-TLX subscales values (Temporal (A)[$F_{3, 84}$ = 14.36, p < .001], Mental (B)[$F_{3, 84}$ = 12.685, p < .001] and Physical (C) [$F_{3, 84}$=13.27, p < .001] demand) separately for the conditions. Error bars denote .95 confidence intervals.

Number and duration of fixations

The number and duration of fixations were computed on epochs of 1 minute for each participant and then averaged. Averaged number and duration of fixations were used as dependent variables in a repeated measures ANOVA design using Condition as the repeated factor. No significant differences between conditions were found (Figure 6-A and 6-B; Table 7-8) [Fixations number: $F_{3, 84}$ =.365, p > .05] [Fixations duration: $F_{3, 84}$ = .798, p > .05].

	TD	MD	PD
Baseline	.36	.78	.61
TD		.49	.64
MD			.79

*Table 7: Post hoc analysis carried out through the Duncan test. Pairwise comparison among fixations number and conditions (*p < .05).*

	TD	MD	PD
Baseline	.48	.46	.71
TD		.18	.69
MD			.30

*Table 8: Post hoc analysis carried out through the Duncan test. Pairwise comparison among fixations duration and conditions (*p < .05).*

Amplitude of saccades

The amplitude of saccades was computed on epochs of 1 minute for each participant and then averaged. The averaged amplitude of saccades was used as dependent variable in a repeated measures ANOVA design using conditions as repeated factor (Figure 6-C). Results showed a main effect of condition [$F_{3, 84} = 12.84$, $p < .001$].

	TD	MD	PD
Baseline	.60	.001	.19
TD		.001	.08
MD			.001

*Table 9: Post hoc analysis carried out through the Duncan test. Pairwise comparison among amplitude of saccades and conditions (*p < .05).*

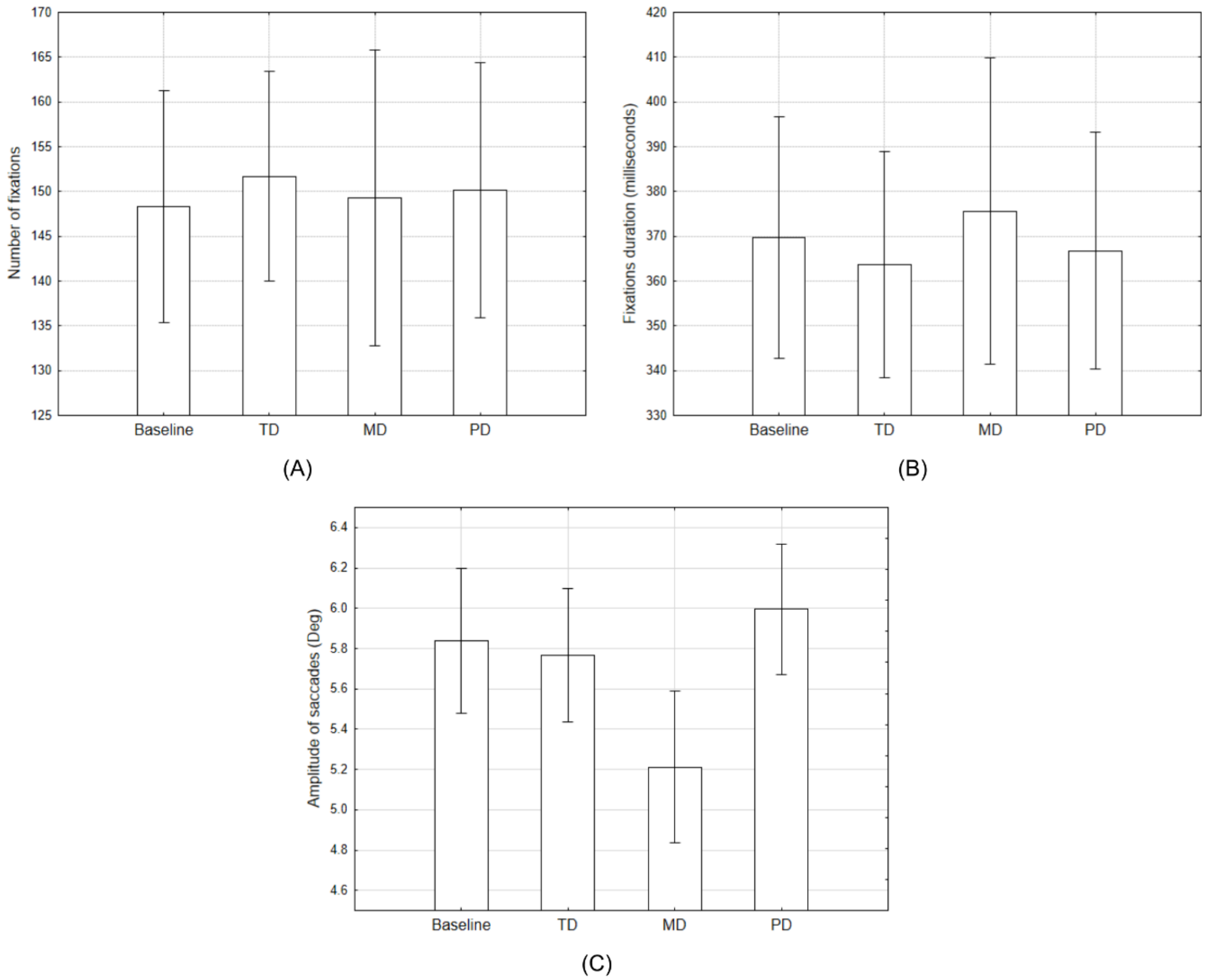

Figure 6: Averaged number (A) and duration (B) of fixations, and amplitude of saccades (C), for the conditions compared with the baseline. Error bars denote .95 confidence intervals.

Nearest neighbour Index (NNI)

The NNI was computed on epochs of 1 minute (Di Nocera, Ranvaud & Pasquali, 2015) for each participant and then averaged . Averaged NNI values were used as the dependent variable in repeated measures ANOVA using conditions as the repeated factor. Results showed a main effect of Condition [$F_{3, 84} = 12.31$, $p < .001$]. TD condition showed higher NNI values (i.e., a more dispersed distribution of fixations) than the baseline (Figure 7-A), while in the MD condition NNI values were the lowest (Figure 7-B) (Table 10).

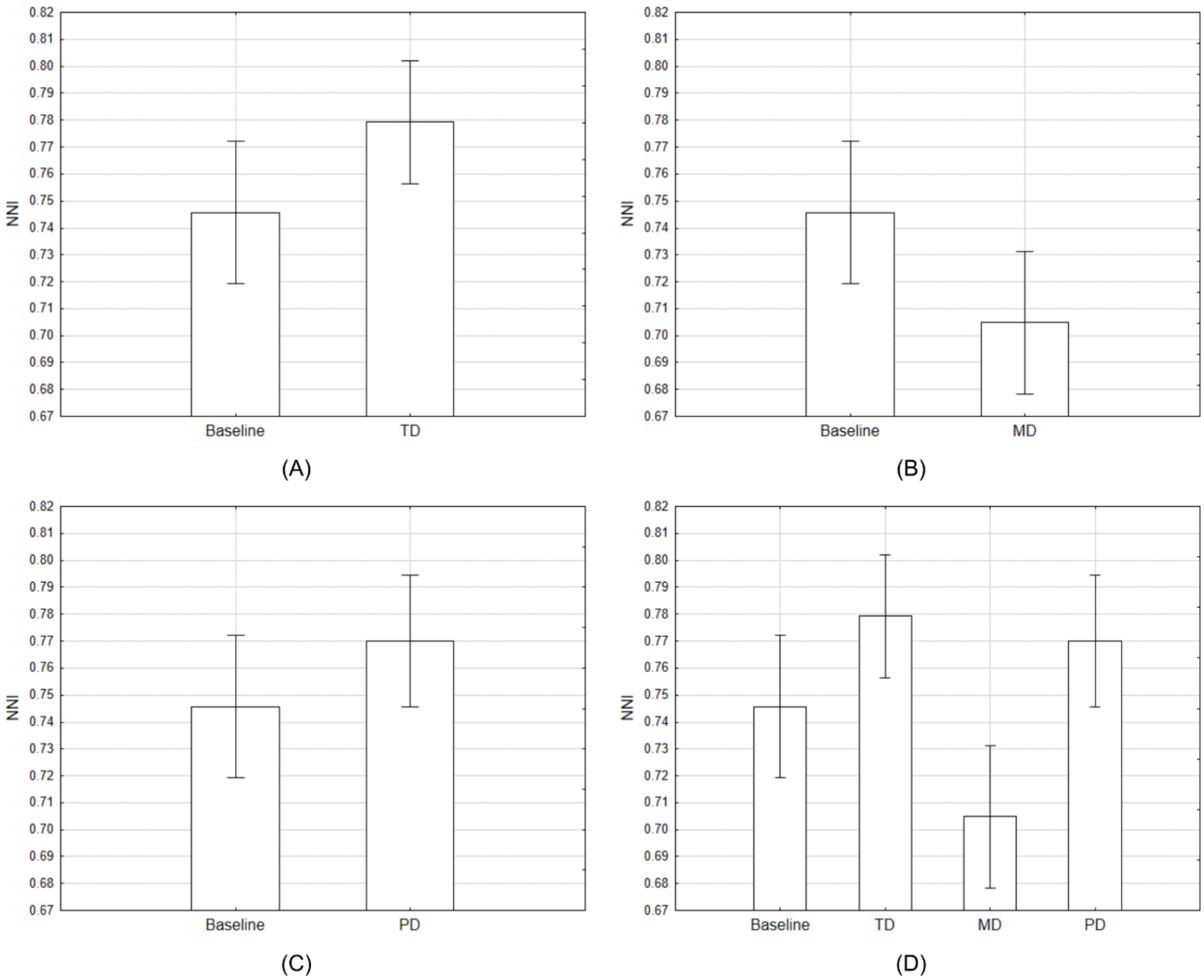

Figure 7: Average NNI for the conditions compared with the baseline separately. Error bars denote .95 confidence intervals. Baseline Vs TD (A); Baseline Vs MD (B); Baseline Vs PD (C); all condition (D).

	TD	MD	PD
Baseline	.018	.01	.072
TD		.001	.49
MD			.001

*Table 10: Post hoc analysis carried out through the Duncan test. Pairwise comparison among NNI and conditions (*p < .05).*

Discussion

This first study aimed at investigating how the visual exploration strategy changes both along with the task-load and with the type of task demand. The results showed an increase in the

NASA-TLX values of the single subscales (mental demand, physical demand, and temporal demand) matching the respective manipulation. Overall, we observed a greater workload in the MD and TD conditions compared to the control and PD conditions. The latter has shown higher values in the corresponding NASA-TLX scale, but the manipulation of the physical demand did not affect the overall self-reported workload. Finally, and more important to our aims, the analysis of the fixations pattern showed high clustering when the task-load increment was obtained by changing the mental (visuo-spatial) demand, and low clustering when it was obtained by changing the temporal demand.

Study two

The entropy-based analysis of the scanpath and the spatial distribution of fixations points are reported to be good indices of mental workload. However, they have never been directly compared. The aim of this second study is to perform such a comparison.

Tools and software

Stimuli

To induce high visual-spatial demand and to assess how that affects visual search, a single pair of black and white pictures (figure 8 and 9). Pictures were rich in details so that the numerous elements would engage participants in a long visual exploration session. The size of each picture was 9.8 x 5.5 inches, and both featured thirty-five subtle differences but were otherwise identical. The two images were aligned horizontally and in full-screen mode on a 27” display.

Ocular activity recordings

Prior recording, participants performed a nine-point calibration and then their eye movements were recorded through the Pupil Labs system with binocular 120 Hz Eye Tracking Camera (Pupil Labs GmbH, Germany) claims accuracy of 0.6 degree.

Participants

The experiment involved fourteen university students (9 women and 6 males, mean age = 24 years, S.D. = 2.6) who participated on a voluntary basis. All participants had a normal or corrected-to-normal vision and were naïve as to the aims of the experiment. This study was compliant with the principles of the Declaration of Helsinki and was the protocol approved by the Institutional Review Board of the Department of Psychology, Sapienza University of Rome. Each participant provided informed consent.

Procedure

The experiment was conducted in a dark room and participants were seated at approximately 2 ft. from a computer screen. During the task, they had to find as many differences as they could between the two images in a 24-minute session. They were requested to click with the mouse on each difference they identified. The differences found were highlighted with a circle throughout the session. Participants were also asked to provide a subjective evaluation of mental workload on a 2-minute schedule (Instantaneous Self-Assessment (ISA): Tattersall & Foord, 1996).

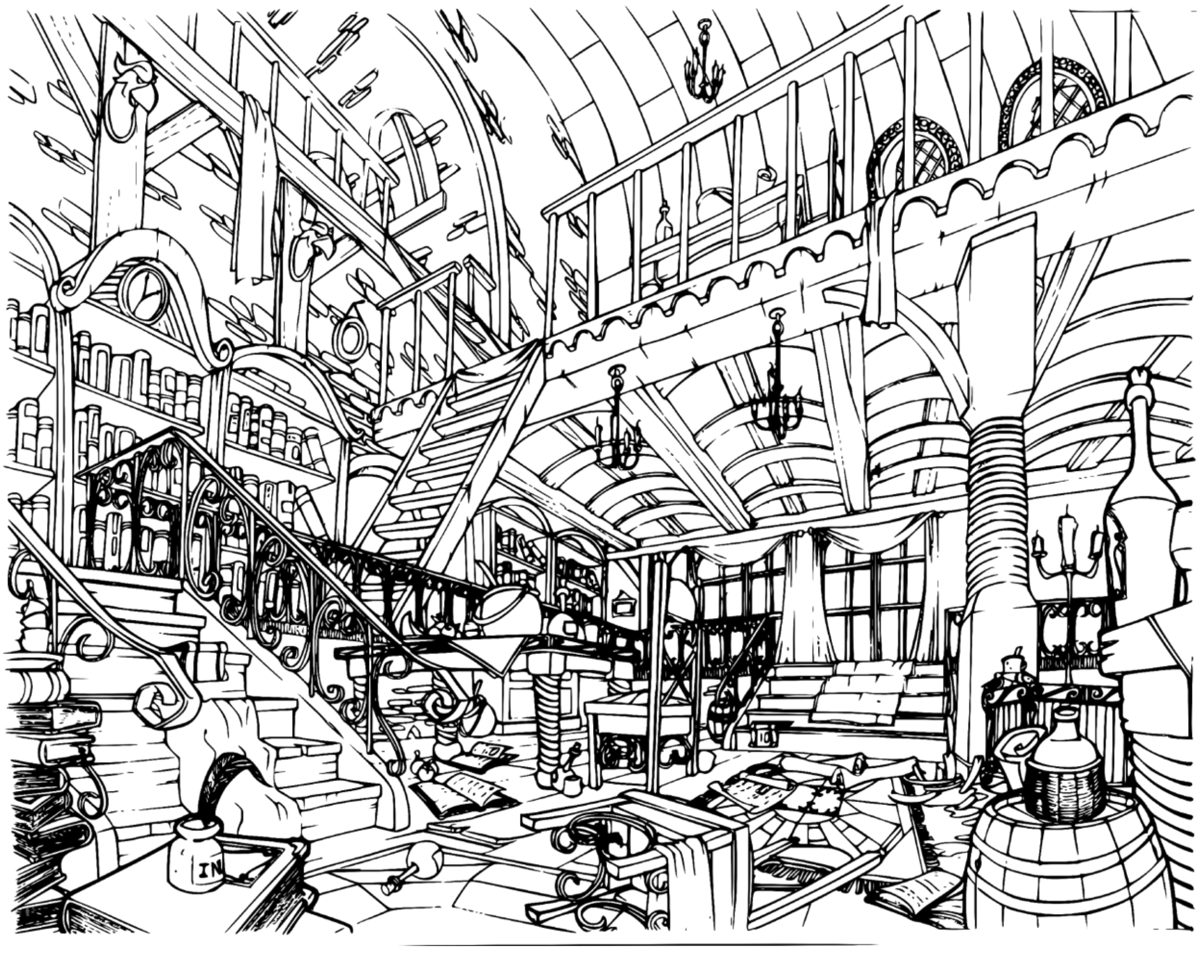

Figure 8: Left panel. Artwork by Benoit Tranchet (reproduced with permission).

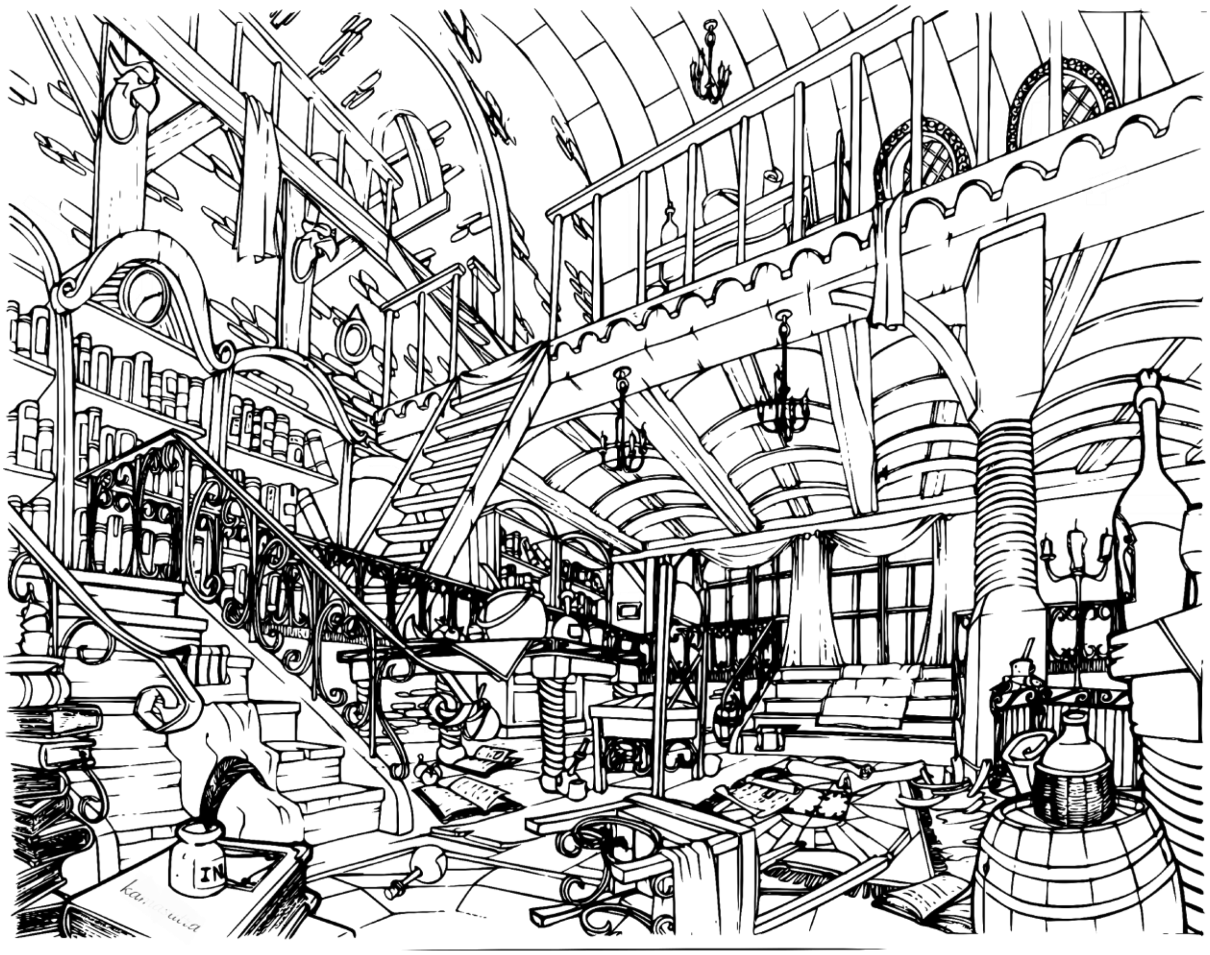

Figure 9: Right panel. Modified version of the original artwork with 35 differences.

Data analysis and results

Performance and self-reporting measures

The whole activity was split into 12 periods of two minutes each, to match performance and subjective evaluations. The number of differences identified by each subject in each epoch was used as a performance indicator. The number of differences identified, and the ISA scores were used as dependent variables in two repeated measures ANOVA designs using Epoch as repeated factor. A main effect of Epoch was found both on the number of differences [$F_{11, 143} = 16.52$, $p < .001$] (figure 10) and the ISA scores [$F_{11, 143} = 15.50$, $p < .001$] (figure 11). Plots reveal Duncan post-hoc testing revealed asymptotic pattern for both the performance measure and the workload estimates starting from the twelfth minute (Table 11; Table 12).

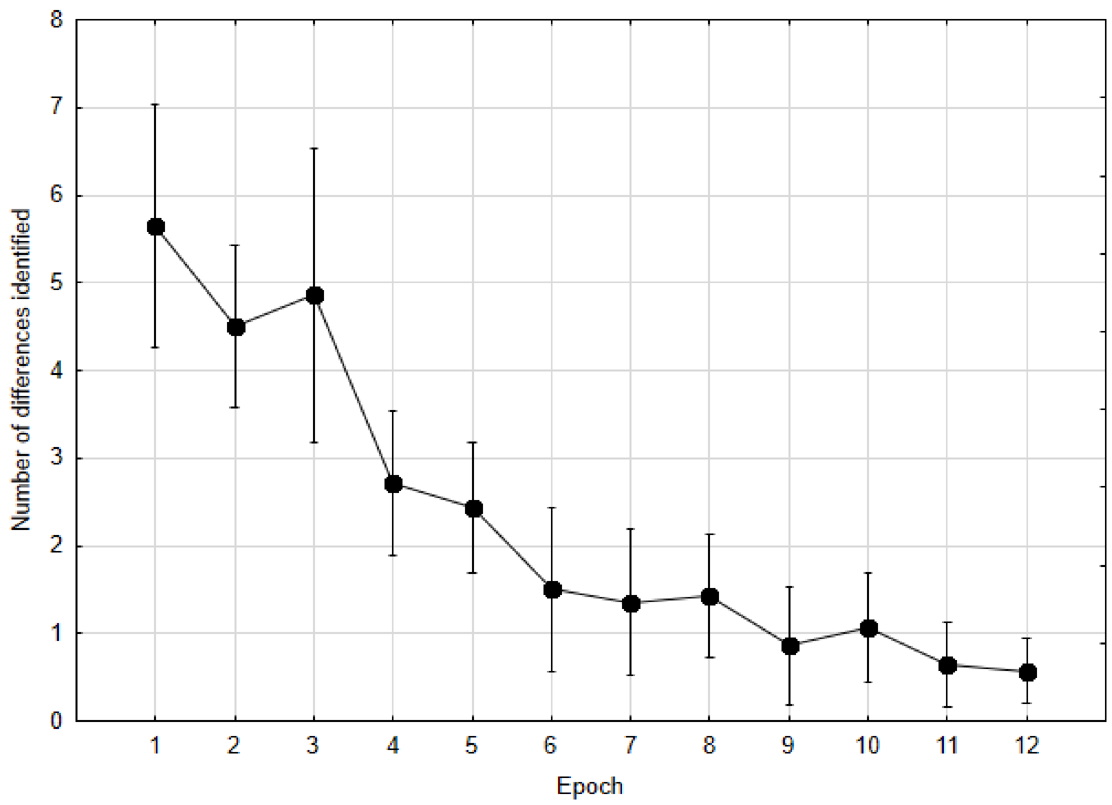

Figure 4: Task performance (number of differences found) along time. Error bars denote .95 confidence intervals.

Epoches	2	3	4	5	6	7	8	9	10	11	12
1	.079	.201	.001	.001	.001	.001	.001	.001	.001	.001	.001
2		.561	.004	.001	.001	.001	.001	.001	.001	.001	.001
3			.001	.001	.001	.001	.001	.001	.001	.001	.001
4				.642	.061	.048	.055	.007	.017	.002	.002
5					.131	.113	.125	.023	.048	.010	.008
6						.829	.908	.362	.533	.231	.201
7							.908	.448	.642	.296	.263
8								.405	.588	.263	.231
9									.728	.728	.665
10										.516	.467
11											.908

*Table 11: Post hoc analysis carried out through the Duncan test. Pairwise comparison among Number of differences identified and Epoches (*p < .05).*

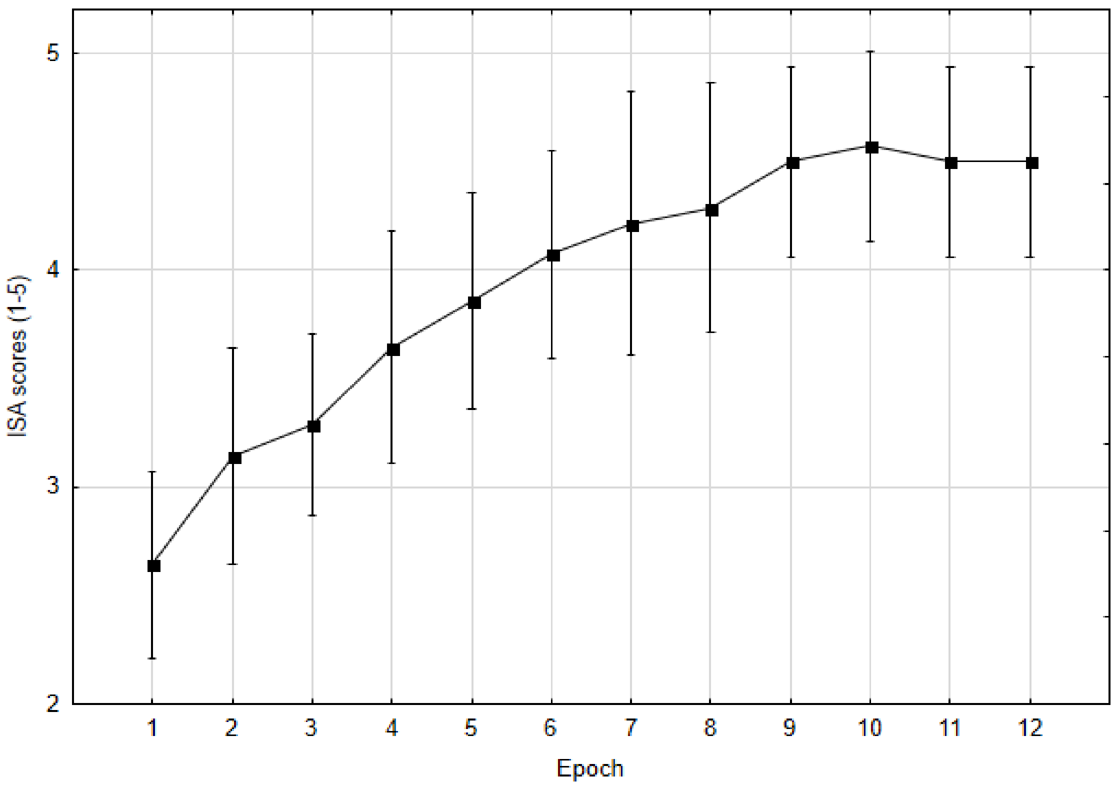

Figure 11: Subjective workload (ratings from 1 to 5) along time. Error bars denote .95 confidence intervals.

Epoches	2	3	4	5	6	7	8	9	10	11	12
1	.028	.006	.001	.001	.001	.001	.001	.001	.001	.001	.001
2		.529	.036	.003	.001	.001	.001	.001	.001	.001	.001
3			.116	.016	.001	.001	.001	.001	.001	.001	.001
4				.345	.074	.019	.009	.001	.001	.001	.001
5					.345	.138	.085	.012	.005	.011	.009
6						.529	.377	.100	.056	.093	.085
7							.753	.270	.174	.257	.238
8								.397	.270	.377	.345
9									.753	1.000	1.000
10										.770	.779
11											1.000

*Table 12: Post hoc analysis carried out through the Duncan test. Pairwise comparison among ISA scores and Epoches (*p < .05).*

Nearest neighbour Index (NNI)

For each participant, the NNI was calculated taking into account 1-minute epochs (Di Nocera, Ranvaud & Pasquali, 2015). Average NNI values were used as dependent variables in a repeated measures ANOVA design using Epoch as repeated factor. A main effect of the Epoch was found [$F_{11, 143} = 4.41$, $p < .001$] (Figure 12). Duncan post-hoc testing showed that the visual strategy applied in the first two minutes significantly differs from all other periods (Table 13).

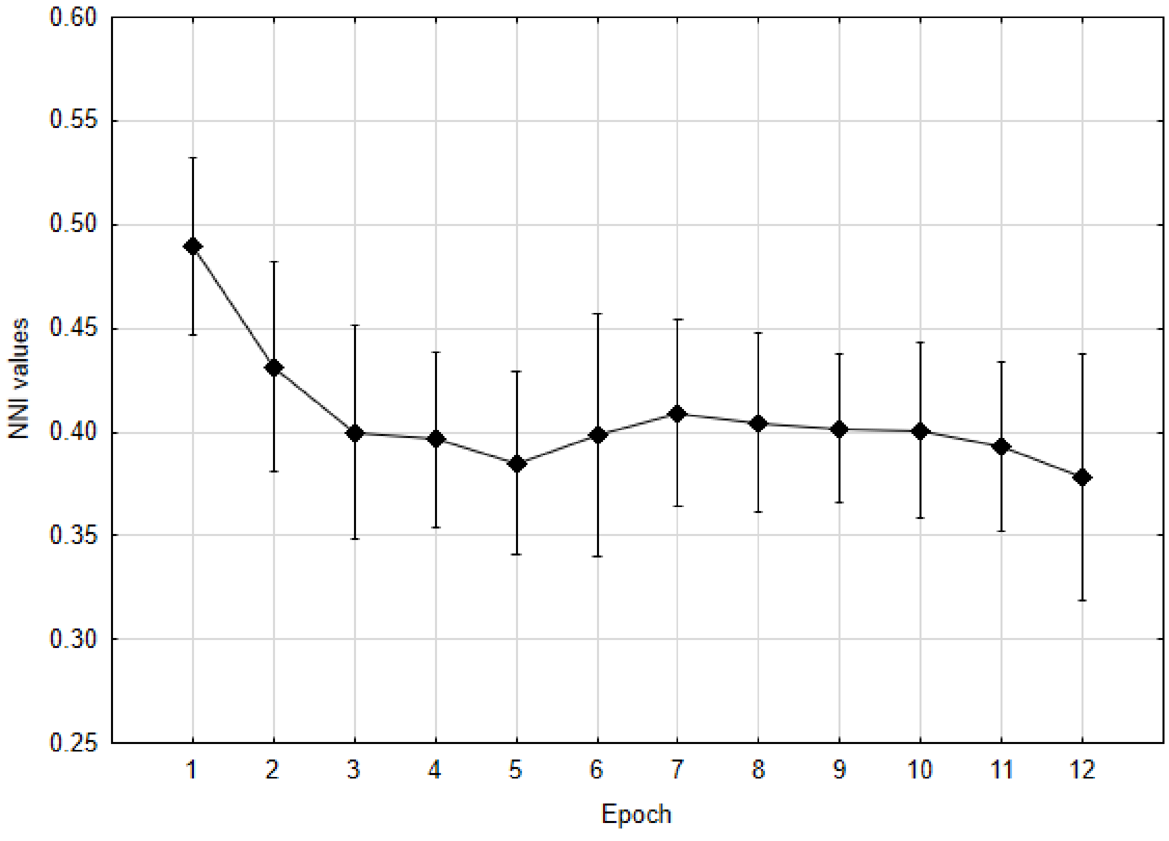

Figure 12: NNI values along time. Error bars denote .95 confidence intervals.

Epoches	2	3	4	5	6	7	8	9	10	11	12
1	.003	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001
2		.163	.131	.046	.154	.253	.197	.172	.165	.099	.021
3			.869	.513	.953	.683	.829	.923	.969	.751	.347
4				.591	.907	.589	.723	.808	.847	.857	.408
5					.536	.315	.411	.474	.501	.692	.728
6						.653	.794	.884	.927	.782	.364
7							.819	.732	.700	.495	.197
8								.890	.850	.618	.269
9									.948	.696	.317
10										.732	.338
11											.487

*Table 13: Post hoc analysis carried out through the Duncan test. Pairwise comparison among NNI scores and Epoches (*p < .05).*

Entropy rate

The whole visual area has been divided into two Areas of Interest (AOI), namely the two images displayed. For each minute, the maximum number and duration of fixations made on each AOI were assessed. For these AOIs, the entropy rate has been adopted as a measure of scan randomness (Tole et. Al., 1983). The entropy rate (H-rate) is expressed in units of bit / s (i.e. the information given by each observation, assessed in bits over seconds). A random pattern is represented by a high H-rate. In this study, all the scanpaths performed by the participants were used to compute the entropy rate. The entropy rates (H_rate) of the sequences of one length for the two images used were computed as a measure of the randomness of the scan. Average H_rate values were used as dependent variables in a repeated measures ANOVA design using the epoch as repeated factor. A main effect of time [$F_{11.143} = 3.69$, $p < .001$] was found (Figure 13). Duncan post-hoc testing showed a steady pattern in the first two minutes of visual exploration, consistently with that obtained with the NNI (Table 14).

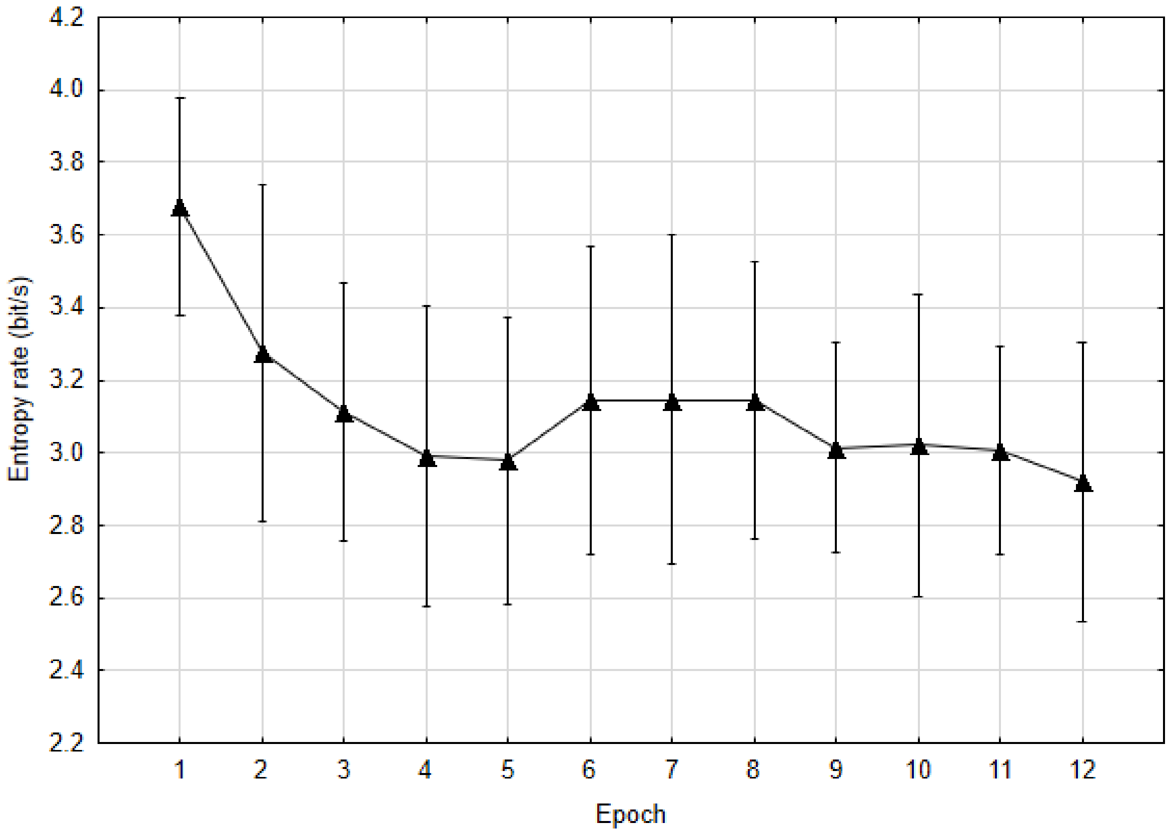

Figure 13: Entropy rate values along time. Error bars denote .95 confidence intervals.

Epoches	2	3	4	5	6	7	8	9	10	11	12
1	.007	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001
2		.337	.110	.097	.432	.412	.381	.132	.137	.129	.046
3			.481	.444	.825	.837	.843	.534	.537	.530	.280
4				.929	.379	.387	.393	.894	.863	.914	.660
5					.346	.353	.358	.836	.807	.854	.702
6						1.000	.999	.426	.434	.422	.209
7							.999	.440	.453	.432	.214
8								.449	.466	.439	.218
9									.960	.972	.593
10										.937	.572
11											.604

*Table 14: Post hoc analysis carried out through the Duncan test. Pairwise comparison among H-rate scores and Epoches (*p < .05).*

Discussion

This second study aimed at comparing two scanpath analysis methods that have been previously reported to be sensitive to changes in the task-load: Entropy rate and Nearest Neighbour Index. Results showed an overall increase of difficulty after the first few minutes of the task. The entropy rate confirms the presence of a less random and more stereotyped pattern starting from the second minute of recording. A similar trend was found for the NNI. The average NNI values in the first two minutes of activity were significantly higher than in the following epochs, therefore showing a change towards fixations grouping as the task-load increased. This study was designed to evaluate the potential of these two measures under the effect of increasing visual-spatial demand. The results showed the same trend, therefore confirming that the two indices are sensitive to changes in the visuo-spatial demand. However, unlike the entropy rate, the NNI is also suitable for estimating changes due to the temporal demand (see Study 1). This is an aspect that could not be accommodated by the entropy rate, which is based on the transitions between AOIs, hence it is completely based on the visuo-spatial performance.

Study three

The third studio aims to investigate the possibility of using the Nearest neighbour Index as a trigger within an adaptive automation system, through two steps: i) identifying the right modality of automation, verifying if it is helpful for the individual; ii) observing if NNI values return to the baseline after the implementation of the automation support.

An automation system was embedded in the Tetris version described in the first study, as it was a necessary condition to carry out the next steps of the research project. An "autopilot", able to take total control of the system, was designed as the best solution to avoid game-over in critical situations. Therefore, a function has been added to detect alignment errors and calculate the best possible combination. The automatic positioning is done by simulating in the background, all combinations between the piece played and those previously (accumulated in the lower part of the area). The system selects the optimal combination, calculated considering the maximum size of the piece surface that touches the bottom of the play (an example is shown in figure 14).

Figure 14: an example of the different combinations, calculated in the background, for the autopilot. The second image represents an incorrect positioning, while in the third image, we see the best combination compared to the one shown in image 1.

Experimental design

The design involved the use of two independent variables at two levels, including four experimental conditions. The "Difficulty" variable (i.e., Easy, Hard) and the "Manual Automation" variable (i.e. Present, Absent). All subjects performed the 4 conditions in random order (Table 15). For each condition, physiological, objective and subjective measures were acquired: NNI, performance and NASA-TLX respectively. Automation consists of the activation of the autopilot that takes the game control until the user turns it off. In order to keep the user engaged, during the automation mode, it changed tasks and reported each time the piece (controlled by the PC) turned white for 200 milliseconds. This can be considered a blink, the interval between blinks varies between 3 and 6 seconds. The study aimed to determine if the NNI can be used as a trigger in an adaptive automation system and whether the autopilot is an optimal automation level for the purpose. Therefore, we have tried to verify this by the following assumptions: i) Subjects will perceive the hard condition as more complex than the easy one. This simply confirms that the two conditions have a different mental workload, both in terms of NNI, NASA-TLX and performance and ii) The easy condition will not be different from conditions with manual automation, both in terms of NNI, NASA-TLX and performance.

		Manual automation	
		Present	Absent
Difficulty	Easy	Condition 1 (EMA)	Condition 2 (EN)
	Hard	Condition 3 (HMA)	Condition 4 (HN)

Table 15: Two independent variables at two levels, including four experimental conditions.

Tools and software

The Gazepoint GP3HD eye-tracking system was used to record ocular activity. This system allows the researcher to collect ocular data without using invasive and/or uncomfortable head-mounted instruments. Gazepoint, the eye tracker manufacturer, claims accuracy within 0.5 to 1.0 degrees and reads data at a rate of 150Hz. The eye tracker was calibrated using the default 9-point calibration test with Gazepoint's included software.

Participants

Eighteen university students (10 women and 8 males, mean age = 27.3 years, St. dev. = 3.5) volunteered in the experiment. All participants had a normal or corrected-to-normal vision, and they were naïve as to the aims of the experiment, its expected outcomes, and its methodology. This research complied with the tenets of the Declaration of Helsinki and was approved by the Institutional Review Board of the Department of Psychology, Sapienza University of Rome, Italy. Informed consent was obtained from each participant.

Procedure

Four conditions have been created (Table 15):

1. EN: Easy level of difficulty, without automation
2. HN: Hard level of difficulty, without automation
3. EMA: Easy level of difficulty, with manual activation automation
4. HMA: Hard level of difficulty, with manual activation automation

After the eye-tracker's calibration, the subjects were explained that their task was to play the game earning as many points as possible (i.e., complete lines and avoid losses). Each condition lasted 10 minutes, and the order of presentation was randomized across participants. In conditions where manual automation was present, subjects received instruction to activate it by pressing the “CTRL key”, any time they perceived a too high difficulty level. Then, the autopilot took control of the game until the subject deactivated it using the same input (i.e., CTRL key). To keep the subject engaged during the autopilot execution, a secondary detection task was asked to be performed: in this phase, the piece, controlled by the computer, turned white for 200ms at intervals between 3 and 6s. The task consisted of pressing the spacebar, as soon as possible, every time this happened.

After completing each condition (i.e., EN vs HN vs EMA vs HMA), participants were requested to fill in the NASA-TLX (Hart & Staveland, 1988).

Data analysis and results

Performance data

To better analyze the performance between different conditions, we computed a Performance Index (PI). The PI was based on the number of lines completed in relation to the maximum number of lines that can be performed. The maximum value is obtained by the total number of pieces that the subject managed in each condition (For example, with 60 pieces it is possible to complete a total of 16 lines if managed in an optimal way). The index goes from 0 to 1, where 1 means that the player has obtained the maximum achievable score. The performance values were used as dependent variables in a two-factor repeated-measures ANOVA design, using difficulty level (easy or hard) and present/absent automation as factors. The interaction effect was not significant [$F_{1, 17}$ =.389, $p > .05$]. However, the results showed a main effect of the difficulty [$F_{1,17}$ = 25.37, $p < .001$] and automation factors [$F_{1,17}$ = 15.86, $p < .001$]. The conditions with the presence of automation (EMA and HMA) were associated with high performance, with respect to the absence of automation (EN and HN) (Figure 15). Moreover, according to the assumptions, better performance is observed under easy conditions (EN and EMA) than under the difficult ones (HN and HMA) (Figure 16).

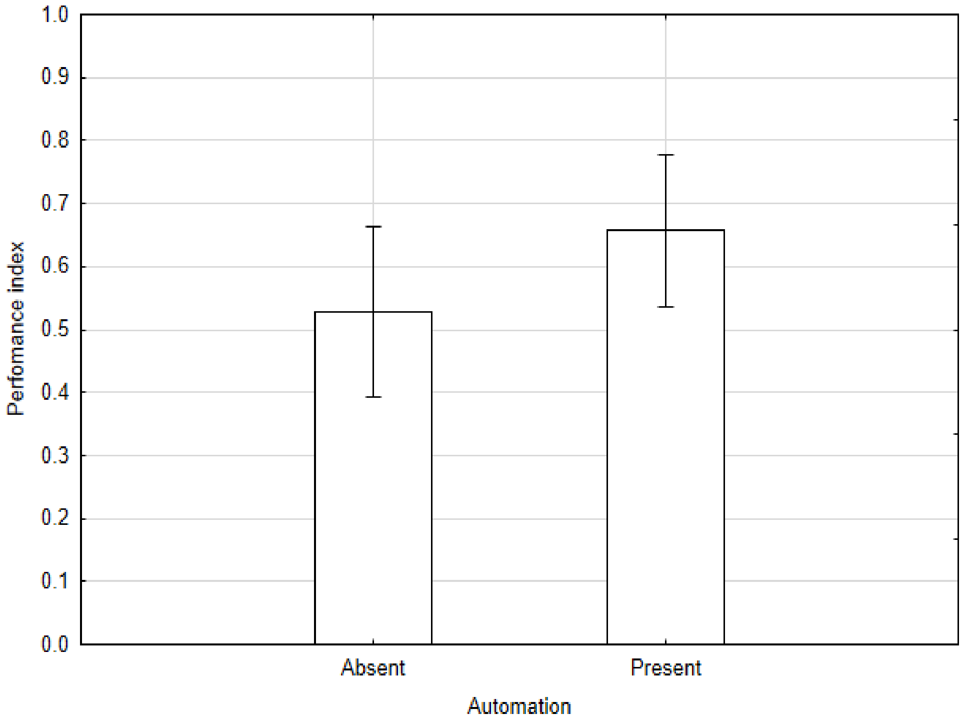

Figure 15: Performance index, the main effect of automation factor. Error bars denote .95 confidence intervals.

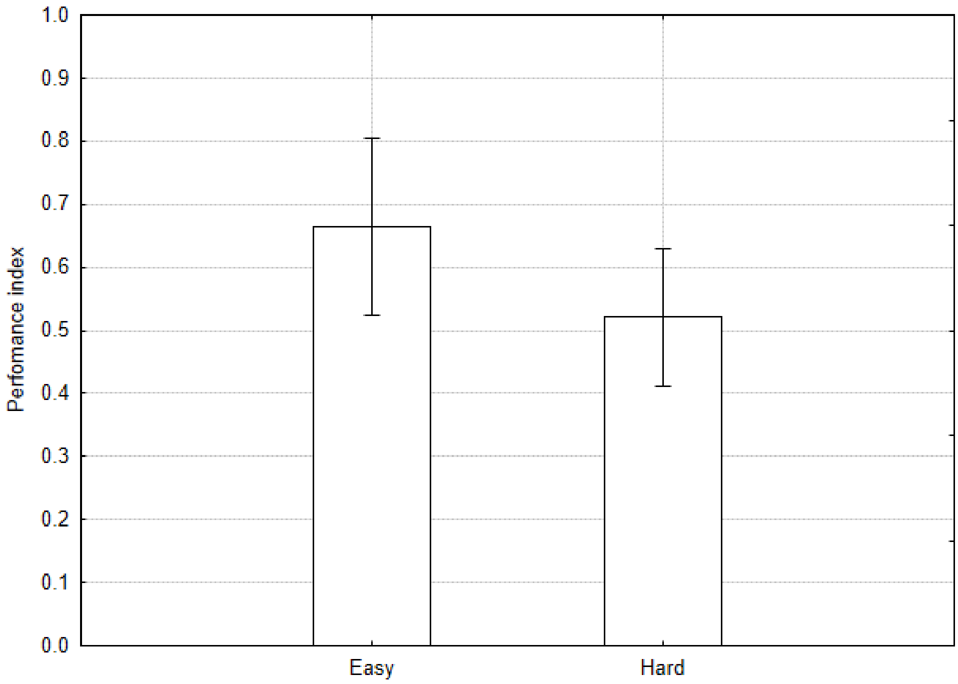

Figure 16: Performance index, the main effect of the difficulty factor. Error bars denote .95 confidence intervals.

Subjective measure

NASA-TLX weighted ratings were used as dependent variables in a two-factor repeated-measures ANOVA design, using difficulty level (easy or hard) and present/absent automation as factors. The results did not show an effect of interaction [$F_{1, 17}$ =.68, p > .05]. We observe a significant main effect of the difficulty factor [$F_{1,17}$ = 22.73, p < .001] (Figure 17), consistent with the performance results. However, there are no differences between conditions with and without automation [Automation factor: $F_{1, 17}$ = .152, p > .05]. The latter suggests that automation has not changed the perception of difficulty provided by the subjects, although the objective performance values are consistent with the assumptions made.

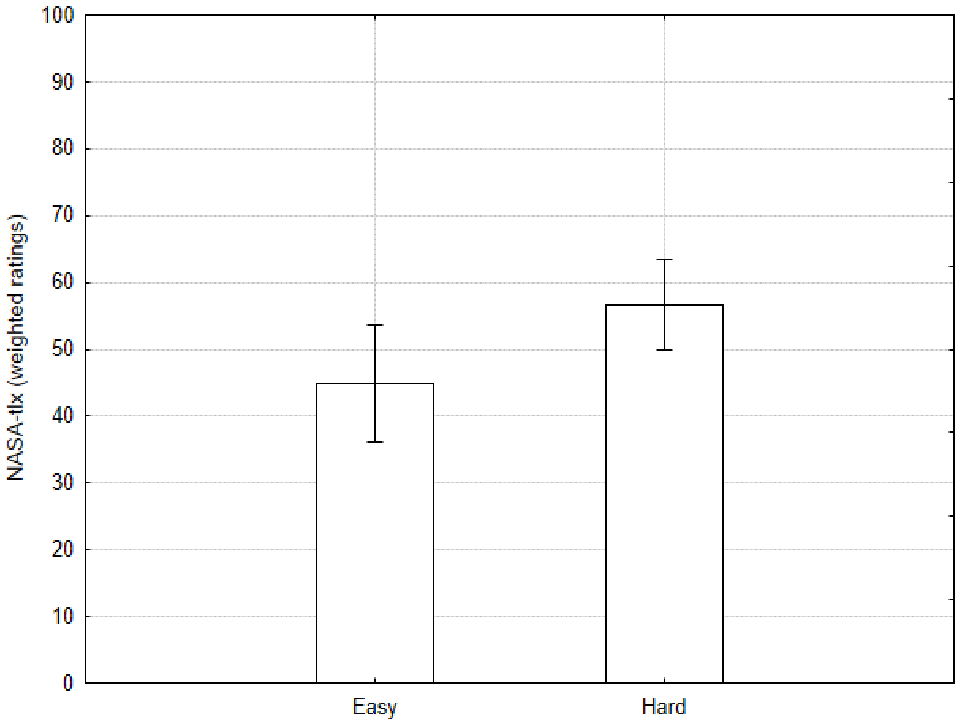

Figure 17: NASA tlx score, the main effect of the difficulty factor. Error bars denote .95 confidence intervals.

Ocular metrics: Nearest neighbour Index analysis

The NNI was computed on epochs of 1 minute for each participant. Averaged NNI values were used as dependent variables in a two-factor repeated-measures ANOVA design, using difficulty level (easy or hard) and present/absent automation as factors. The results showed a significant interaction effect between difficulty and automation factors [$F_{1,17}$ = 14.78, p < .01] (Figure 18). In conditions where automation (EMA and HMA), NNI values are not different between difficulty levels. Moreover, these are very similar to EN conditions. The hard condition, without automation, results in the highest NNI values (Table 16).

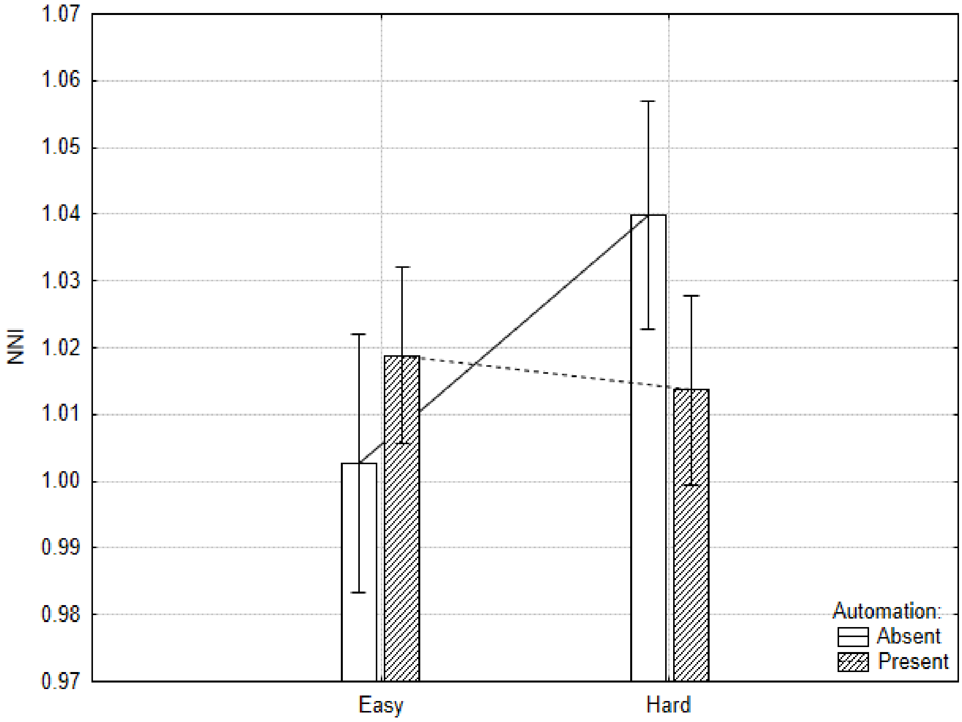

Figure 18: NNI values, the interaction effect between difficulty and automation factors. Error bars denote .95 confidence intervals.

	Automation	Absent	Present	Present
Automation	**Difficulty**	Hard	Easy	Hard
Absent	Easy	.001	.064	.177
Absent	Hard		.015	.004
Present	Easy			.515

*Table 16: Post hoc analysis carried out through the Duncan test. Pairwise comparison among NNI scores and Conditions (*p < .05).*

Ocular metrics: Number of triggers, classifying of NNI values.

The NNI points were classified by a threshold value obtained from the EN condition. This threshold was calculated per subject, and it considered the average plus one standard deviation of the NNI scores in the "Easy-Without Automation" condition. Subsequently, for all conditions, the total number of values that exceeded this limit was calculated. This analysis is based on the idea that the NNI points set up a range of "Normality" in the optimal condition. Therefore, in the non-optimal conditions, the scores can be expected outside this range, indicating a change in the subject's visual exploration strategy during a critical situation. NNI scores over the threshold, which is defined in a version of the task with an optimal difficulty level, could be used as powerful triggers in an automation system. The total number of NNI values over the threshold were used as dependent variables in a two-factor repeated-measures

ANOVA design, using difficulty level (easy or hard) and present/absent automation as factors. The results showed a significant interaction effect between difficulty and automation factors [$F_{1,17} = 9.07$, $p < .01$] (Figure 19). In difficult condition, without automation (HN), the results show the highest number of trigger values than EN and HMA conditions. However, HN was not different from EMA condition (Table 17).

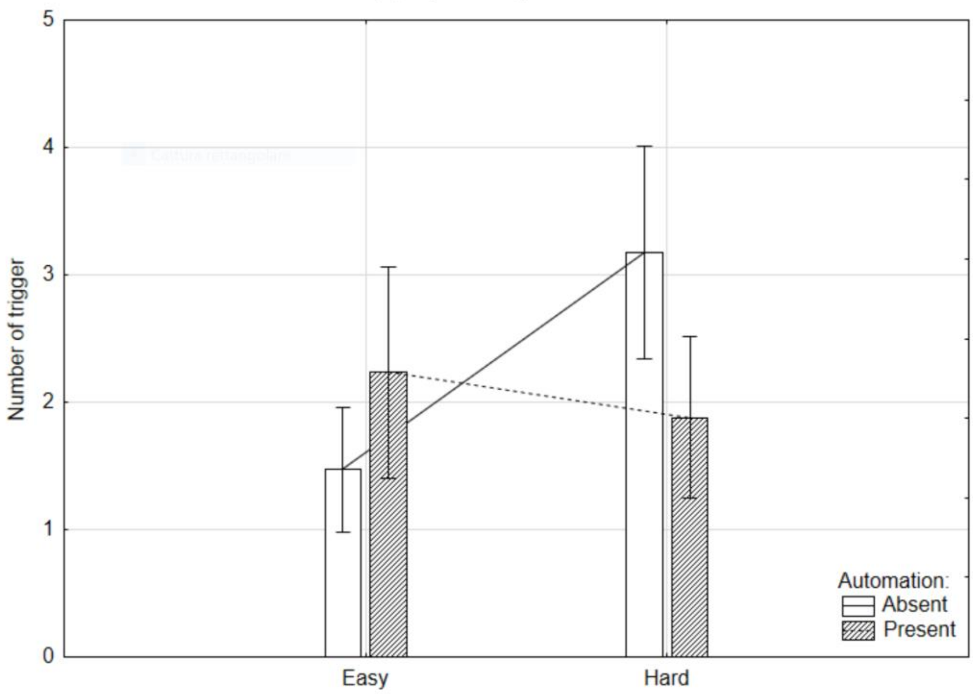

Figure 19: The total number of NNI values over the threshold (Number of triggers), the interaction effect between difficulty and automation factors. Error bars denote .95 confidence intervals.

	Automation	Absent	Present	Present
Automation	**Difficulty**	Hard	Easy	Hard
Absent	Easy	.001	.151	.406
Absent	Hard		.068	.02
Present	Easy			.475

*Table 17: Post hoc analysis carried out through the Duncan test. Pairwise comparison among Number of triggers and Conditions (*p < .05).*

Discussion

This study aimed to test the effectiveness of the autopilot as an automation system. So, objective (i.e., performance at the task), subjective (i.e., questionnaires on mental workload) and physiological (i.e., ocular metrics) measures were compared among different conditions. The first hypothesis was based on previous studies. All measurements confirm a higher level of difficulty in the HN condition than in the EN. Subsequently, the analyses on NNI and performance values showed a trend in line with the second hypothesis. Regardless of the difficulty level, automation conditions are not harder than the easy one without automation (EN), set to be an optimal game condition. However, the subjective measurement of NASA-TLX is not consistent with this result. Here, the hard conditions (with and without automation) do not show significant differences.

On the one hand, automation seems to be an effective aid in terms of performance, on the other hand, it does not seem to affect the perception of the overall difficulty. This can be explained by the frequent switching between manual and autopilot driving. As several studies in the literature suggest, frequent switching between automation levels could increase mental workload.

Study four

The study aims to validate NNI as a trigger in an adaptive automation system. The experiment did not provide a real "adaptivity", rather the automation was activated/deactivated according to a schedule defined a priori by the investigator, and that varies for each subject. Therefore, a series of “3-min units” was defined, allowing us to observe what happens, in terms of ocular strategies, before and after the present/absence of automation. The effectiveness of the automation itself was verified, which consists of a real autopilot, through subjective measurements (i.e., NASA-TLX), performance (i.e., the number of lines performed) and ocular metrics (NNI).

Participant

Thirty university students (i.e., 17 women and 13 males, mean age = 26.3 years, St. dev. = 3.2) volunteered in the experiment. All participants had a normal or corrected-to-normal vision, and they were naïve as to the aims of the experiment, its expected outcomes, and its methodology. This research complied with the tenets of the Declaration of Helsinki and was approved by the Institutional Review Board of the Department of Psychology, Sapienza University of Rome, Italy. Informed consent was obtained from each participant.

Apparatus

The Gazepoint GP3HD eye-tracking system was used to record ocular activity. This system allows the researcher to collect ocular data without using invasive and/or uncomfortable head-mounted instruments. Gazepoint, the eye tracker manufacturer, claims accuracy within 0.5 to 1.0 degrees and reads data at a rate of 150Hz. The eye tracker was calibrated using the default 9-point calibration test with Gazepoint’s included software.

Procedure

The experiment included two game sessions of 10 and 31 minutes, respectively. In the first 10-minute session, a baseline of eye movements was calculated, thus defining the threshold given by the average NNI values ±1 standard deviation. Subsequently, it was used to identify the periods of high complexity (when the NNI values exceeded this threshold). The 10 minutes were divided into two 5-minute units, one with automation and the other without random order. As in previous studies, the entire session was set to an easy level of play, level 6: with a drop speed of 208 ms per block. In the second 31-minute phase, automation was activated/deactivated based on a schedule set by the experimenter and customized for each subject. The schedule was created using a partial randomization method: each 31-min session contained a total of fifteen “3-minute trials” (defined as “series”), where the third minute corresponds to the first of the next series. Only half of the series included the autopilot's use, provided only in the second minute of each one (figure 20). This session was set to a hard level of play (level 8: with a drop speed of 156 ms per block) to increase the mental workload. At the end of each phase (the first and last 5 minutes of the training session and the 31-minute session), NASA-TLX was administered.

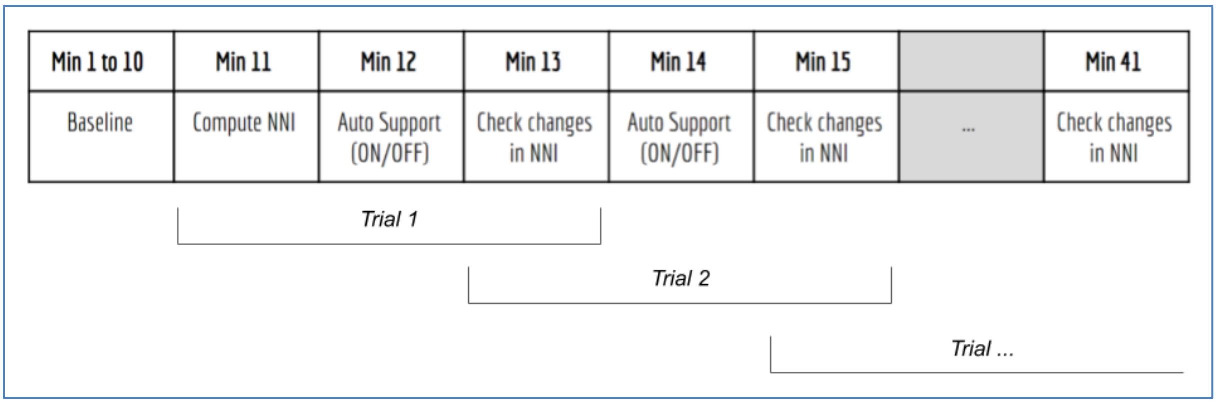

Figure 20: Graphic representation of the setting.

Data analysis and results

Performance data: Tetris score

The Performance Index (PI) was calculated as in previous studies. The performance values were used as dependent variables in a two-factor repeated-measures ANOVA design, using sessions (training and “31 minutes” sessions) and present/absent automation as factors. The interaction effect was significant [$F_{1, 29} = 40.85$, $p < .001$]. The results showed that subjects achieve lower performance when automation is absent. In addition, there is a difference between the training session and the experimental session (i.e., 31 minutes: "31-min"), again when automation is absent (Table 18). Subjects play better during the training phase, which is set to be easier, according to the assumptions above (better performance is observed under easy conditions (Figure 21).

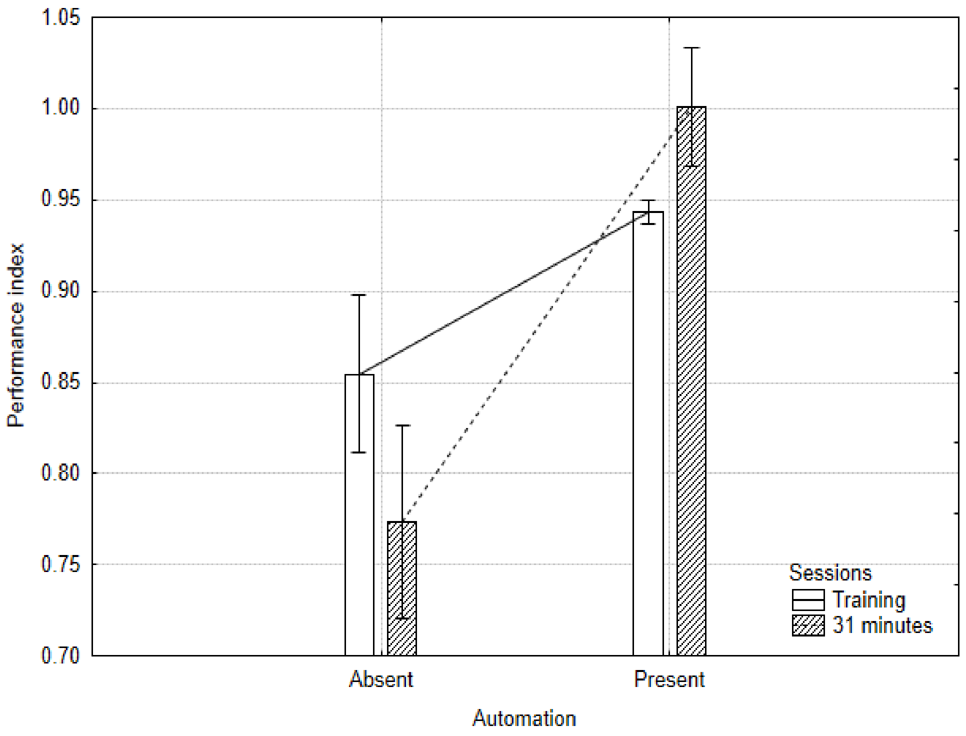

Figure 21: Performance index, the interaction effect between sessions and automation factors. Error bars denote .95 confidence intervals

	Automation	Present	Absent	Present
Automation	**Session**	Training	31-min	31-min
Absent	Training	.001	.001	.001
Present	Training		.001	.001
Absent	31-min			.001

*Table 18: Post hoc analysis carried out through the Duncan test. Pairwise comparison among PI score and Conditions (*p < .05).*

Subjective measure

NASA-TLX weighted ratings were used as dependent variables in repeated measures ANOVA design using three conditions as the repeated factor: First and last 5-minutes of the training session and the 31-min session. Results showed a main effect of the condition [$F_{2, 58} = 104.78$, $p < .001$] (Figure 22; Table 19), consistent with the performance index and previous studies.

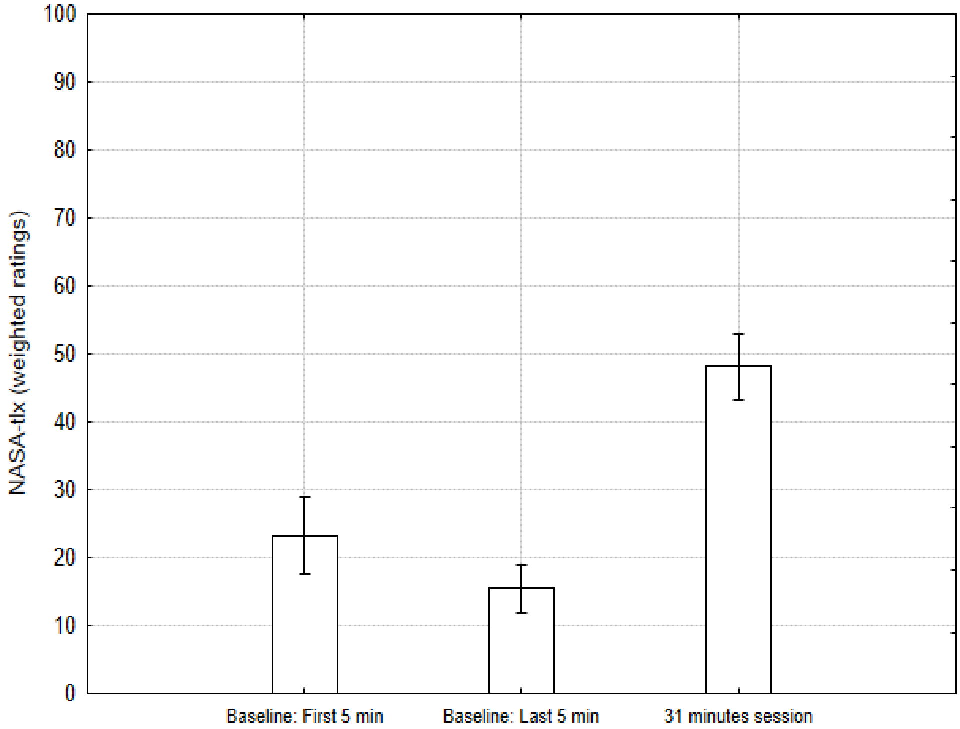

Figure 22: NASA-TLX values (weighted scores) separately for the conditions. "Baseline: First 5 min" refers to the first 5 minutes of the training session, while "Baseline: Last 5 min" refers to last minutes of the same. Error bars denote .95 confidence intervals.

Session	Training (With Automation)	31-min
Training (no Automation)	.001	.001
Training (With Automation)		.001

*Table 19: Post hoc analysis carried out through the Duncan test. Pairwise comparison among NASA-TLX score and Conditions (*p < .05).*

Performance data: blinks reaction time

To keep the subject engaged during the autopilot execution, a secondary detection task was asked to be performed: in this phase, the piece, controlled by the computer, turned white for 200ms at intervals between 3 and 6 seconds. The task consisted of pressing the spacebar, as soon as possible, every time this happened. The values of average reaction times were used as dependent variables in a repeated-measures ANOVA design, using conditions with automation (last 5-minutes of training and "31 minutes" sessions) as repeated factors. Results showed a main effect of the condition [$F_{1, 29} = 93.49$, $p < .001$], the reaction times increased in the 31-min session with respect to the training session with automation (last 5-minutes) (Figure 23).

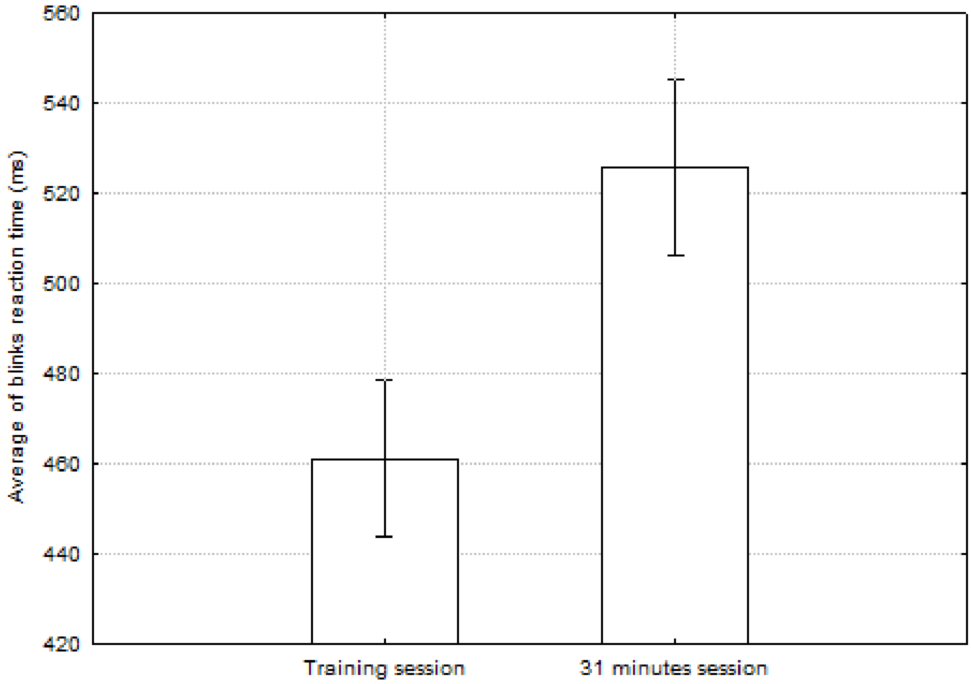

Figure 23: Comparison of average reaction times between the training phase and the 31-minute experimental session. Error bars denote .95 confidence intervals.

Ocular metrics: Proportion of NNI values within the range

Classification of ocular data

The presence or absence of automation support was classified with four categories:

- Adaptively **(NP)**: in the first minute of series, the NNI was out of range and, in the next minute, the automation was present.
- Invalidly **(nNP)**: in the first minute of series, the NNI was in the range and, in the next minute, the automation was present.
- Not provided when needed **(NA)**: In the first minute of series, the NNI was out of range and, in the next minute, the automation was absent.
- Not provided when not needed **(nNA)**: in the first minute of series, the NNI was in the range and, in the next minute, the automation was absent.

The proportion of within-range values is expected to be significantly higher in the valid conditions (NP and nNA) rather than in the invalid conditions (NA and nNP).

Results

The NNI values obtained in the 31-minute session were used to catalogue the single series in NP, nNP, NA, nNA categories. Those four conditions were used as factors in a repeated measure ANOVA design. The results did not show significant differences [$F_{3,81} = 1.35$, $p >$

.05] (Figure 24). Furthermore, the comparison between valid and invalid conditions, as two different groups, did not show significant differences [$F_{1, 29}$ =3.56, p = .069] (Figure 25). This result is not in line with the initial assumptions but shows a compatible trend, and it is possible that more trials may decrease data variability and provide stronger results.

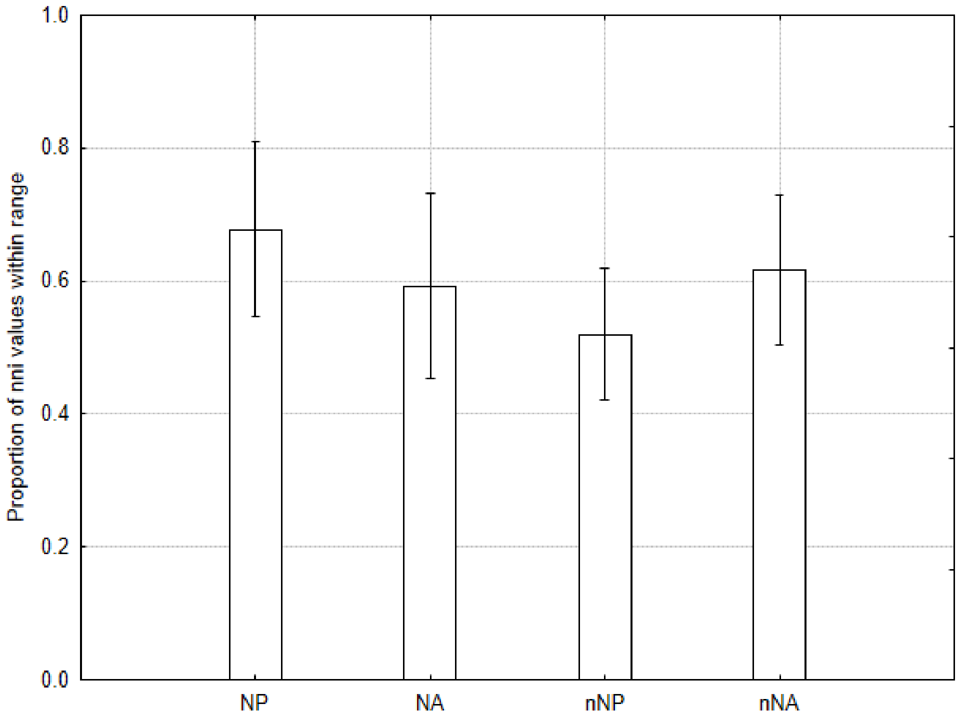

Figure 24: Comparison of NP, NA, nNP, nNA conditions.

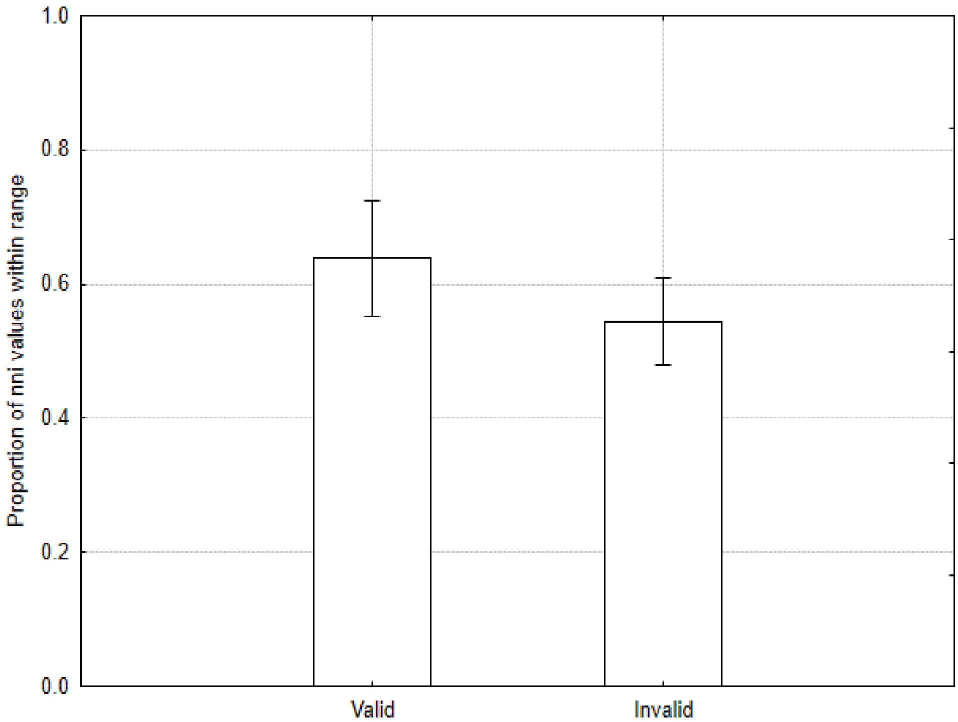

Figure 25: Comparison of conditions Valid (NP, nNA) and Invalid (NA, nNP).

Discussion

In this study, the autopilot was applied to compare its effect on eye behaviour among different trials. In particular, the 31 minutes were divided into 15 trials, each composed of 3 minutes of detection: in the first minute, the autopilot was absent. In the second minute, this could be present or absent. The presence of automation in the second minute of each trial was random, and therefore, it was presented after a game session in which the subject had shown a high mental workload, operating as adaptive automation. In other cases, it was presented after a game session in which the subject had not shown a high mental workload (NNI within the baseline range), not operating as adaptive automation. The results showed high variability, probably caused by the low number of occurrences for each category (NP, nNP, NA, nNA) and the high dynamism of the task, where a single error can lead to game-over. However, it should be noted that this experiment was based on the calculation of a 10-minute baseline that allowed us to calculate threshold values. Future studies should consider a more durable baseline to provide more accurate values for each subject.

New methods to analyse NNI and scanpath.

Introduction

Algorithms like Nearest neighbour index, rarely consider the temporal dimension and sequentiality of points in a trajectory. Linking spatial variation to eye movements over time has been done by determining the distribution of fixations separately in each temporal period (on minute) (Di Nocera, Ranvaud & Pasquali, 2015). Also, the sequence of eye movements can be analyzed by comparing graphical representations of scanpaths. Di Nocera & Bolia (2007) analyzed pilots' scanpaths using stochastic PERT networks to gather detailed information on the processes underlying the ocular activity. One of the goals of this thesis is to link human performance to Spatio-temporal patterns in eye-movement data. The number of statistics that could be used (and have never been used) in the spatiotemporal analysis of the scanpath is very large (Stark & Young, 1981; Smith, 1998). For example, in 1975, Pinder and Witherick, proposed an adaptation of the NNI algorithm, for linear one-dimensional situations. Unlike Clark & Evans (1954) original study, the authors do not consider the area occupied by the points in space, but the line that connects them. In this way, they try to meet some sectors' needs, including the study of archaeological sites or cities arranged, for example, along a waterway.

Research questions

The first goal was to obtain a more stable NNI measure over time that can filter out the peaks due to fixations far from the interaction area, often caused by sampling errors or elements outside the task's area that catch the subject's attention. The previous analyses have always considered the average of more values, showing how this varies with the mental workload. In an application context, each minute's value should be the result of the NNI net of the "noise" mentioned above.

Next, the NNI was analysed in the frequency domain. This second goal focused on examining the NNI from a new perspective, trying to detect "outlier" frequencies in comparison with frequencies generated by an optimal condition. To realize it, the data from the first study were reused as the basis for the new analyses.

Data analysis and results

K-Nearest neighbour algorithm

The k-nearest neighbours (k-NN) is an algorithm used to classify objects, based on the characteristics of items near the targeted one (Figure 26). An item is classified based on the majority vote of its k neighbours. K is a positive integer, typically not very high. If "K = 1", then the object is assigned to the class of its neighbour one. In a binary context where there are only two classes, it is appropriate to choose k odd to avoid ending up in a position of equality. The K-NN classification algorithm decides the output based on the most represented class among the K neighbours. If the output is continuous, the decision to the majority does not have more sense (the values can be all different). In this case, the K neighbours' average can be assumed as the output value (Imandoust & Bolandraftar, 2013). Considering only the votes of K neighbouring objects, there is the drawback due to the predominance of classes with more objects. In this case, it may be useful to weigh the neighbours' contributions to give, in the calculation of the average, greater importance according to the distance from the object considered. The choice of K depends on the characteristics of the data. Generally, as K increases, the noise that compromises the classification is reduced, but the criterion of choice for the class becomes more labile. The choice can be made through heuristic techniques (Manning & Schuetze, 1999).

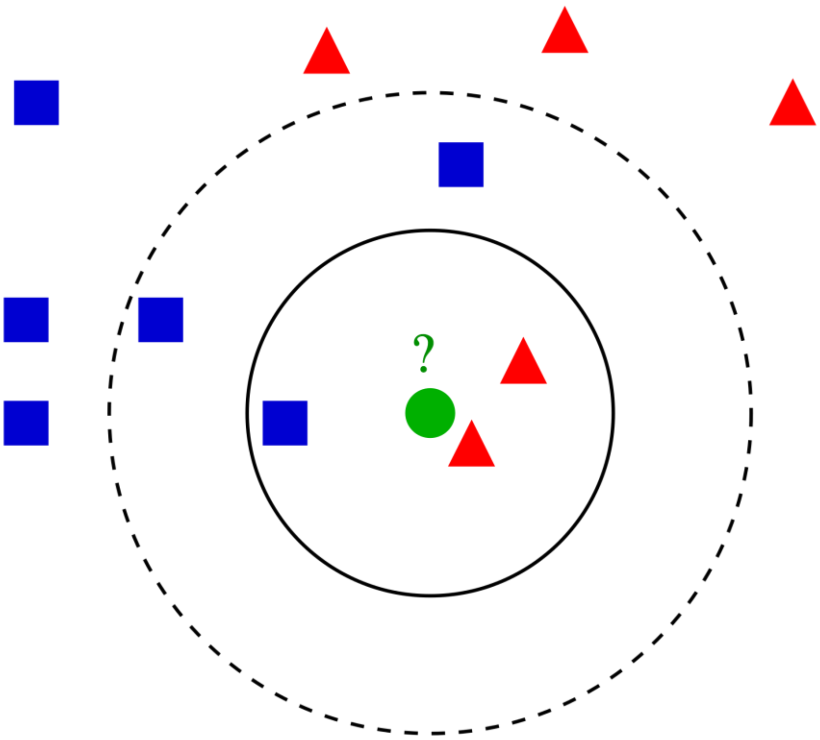

Figure 26: If k = 3 (i.e., the 3 closest objects are considered), then the green dot is placed in the same class as the red triangles because 2 triangles and 1 square are present. If k = 5 then it is placed in the same class as the blue squares because 3 squares and 2 triangles are present.

The two algorithms, NNI and K-NN, have two different purposes: the former describes the distribution of points as clustered or dispersed, the latter is a classification algorithm. There is no use in the literature of the variable K within the NNI algorithm as described in the first study of this thesis. Given that the NNI is a continuous measure, K was integrated into the algorithm as follows: The ratio of 1) the minimum average distance of a point to its K nearest points and 2) the minimum average distance between points if they were perfectly distributed within the area. It should be kept in mind that by using the K, the NNI can greatly exceed the expected limit value of 2.15. This latter effect is caused by the denominator of the equation, where the K is not applicable. In this sense, using a K greater than 1 can reduce the "noise" of the instrument given by the calculation of fixations (e.g., fixations that are too short and too close together).

The K-NN was computed in epochs of 1 minute for each participant with a second (K = 2) and third-order k (K = 3). One subject was excluded from the data analysis due to the low quality of recorded eye movements. Averaged K-NN values were used as the dependent variable in repeated measures ANOVA using conditions as the repeated factor. With K = 2, results showed a main effect of condition [$F_{3, 87} = 13.09$, $p < .001$] (Figure 27; Table 20). TD condition showed higher K-NN values (i.e., a more dispersed distribution of fixations) than the baseline, while in the MD condition, we obtained lower values.

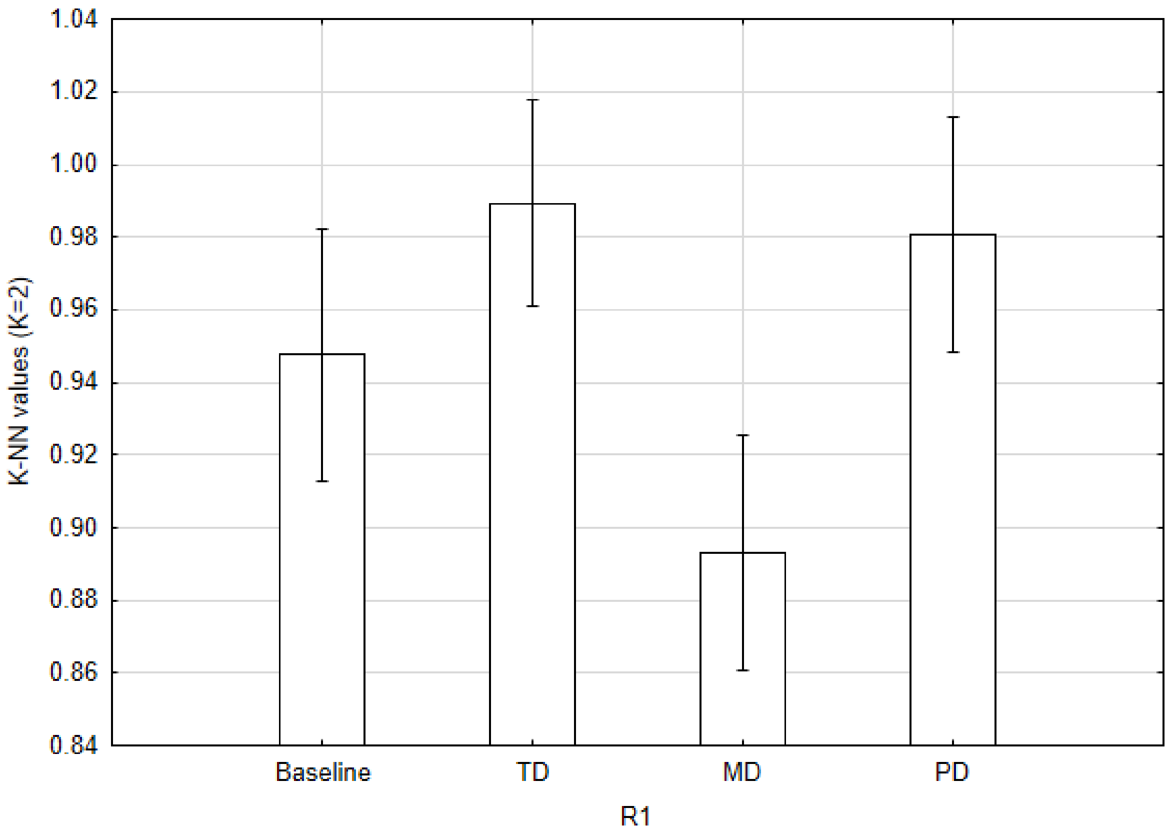

Figure 27: Average K-NN value (K=2) for the conditions compared with the baseline separately. Error bars denote .95 confidence intervals.

	TD	MD	PD
Baseline	.021	.002	.056
TD		.001	.609
MD			.001

*Table 20: Post hoc analysis carried out through the Duncan test. Pairwise comparison among K-NN (K=2) scores and conditions (*p < .05).*

With K = 3, results showed a main effect of condition [$F_{3, 87} = 13.09$, $p < .001$] (Figure 28; Table 21). TD condition showed higher K-NN values (i.e., a more dispersed distribution of fixations) than the baseline, while in the MD condition, we obtained lower values.

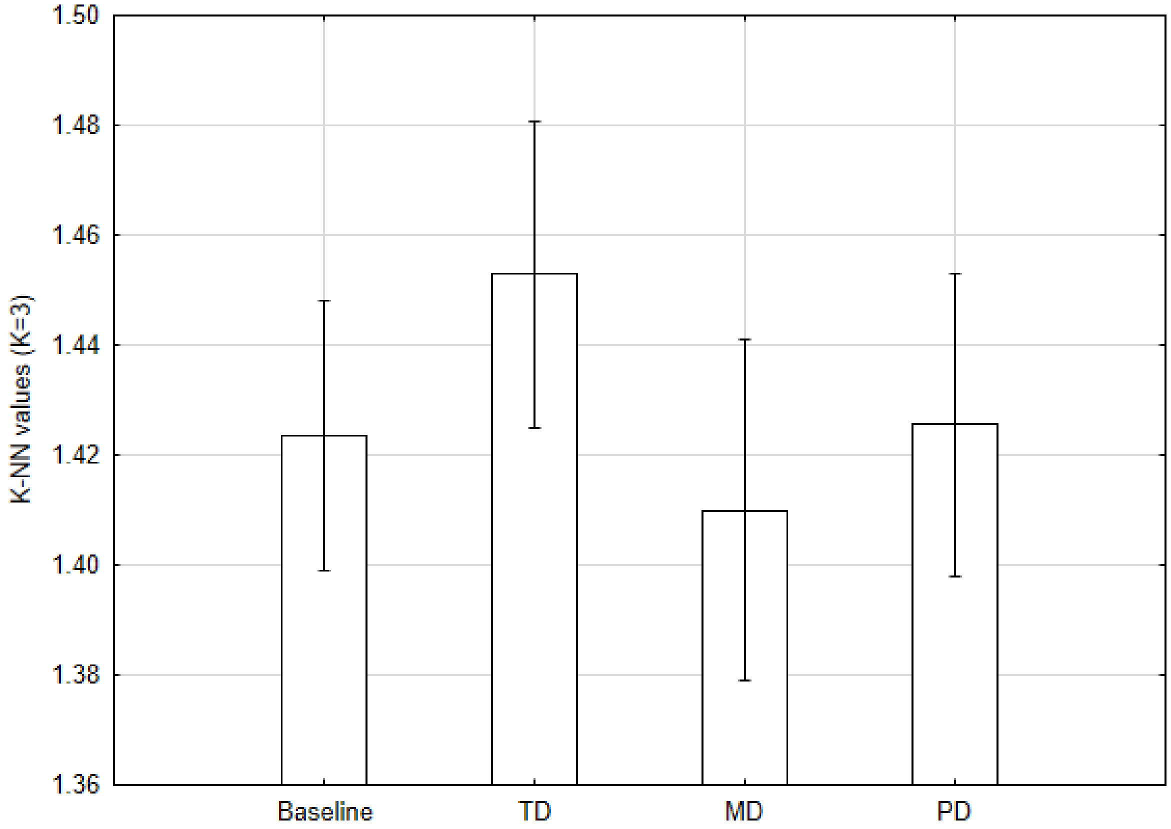

Figure 28: Average K-NN value (K=3) for the conditions compared with the baseline separately. Error bars denote .95 confidence intervals.

	TD	MD	PD
Baseline	.002	.119	.824
TD		.001	.002
MD			.092

*Table 21: Post hoc analysis carried out through the Duncan test. Pairwise comparison among K-NN (K=3) scores and conditions (*p < .05).*

Figure 29 compares the results obtained with second-and third-order K-NN versus the classical NNI algorithm used in the first study. The plot shows 3 very similar trends, especially in relation to NNI and second-order K-NN, where the significance values are equivalent. However, in the third-order K-NN plot, a more pronounced change and loss of significance are observed than in the previous ones.

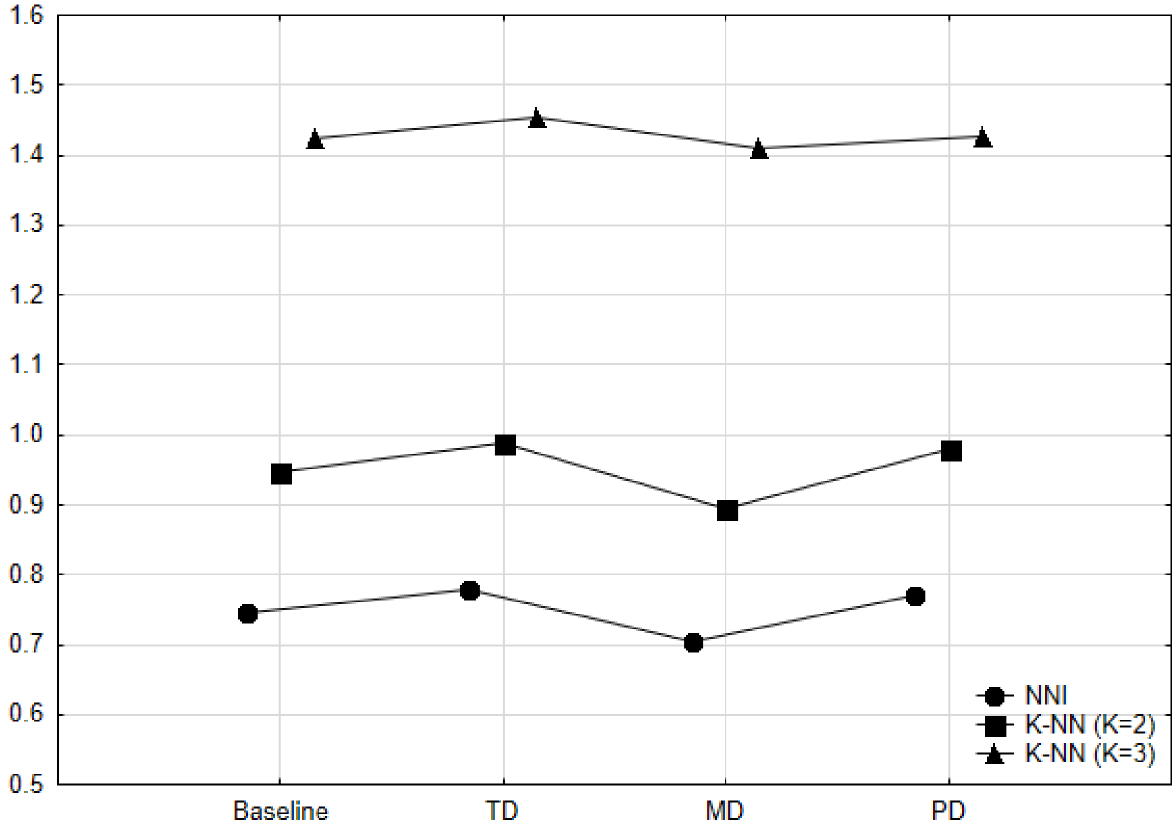

Figure 29: Comparison of values obtained by varying the K factor: NNI (K = 1) vs K = 2 vs K = 1.

Spectral analysis

The focus of this analysis was the quantitative time evolution of NNI as a task is carried out. To this purpose, data from the first study were further analysed using spectral analysis, which is appropriately and commonly used in studying measurements collected at regularly spaced intervals of time. As described previously, in the first study, subjects performed 4 sessions of 10 minutes each. Therefore, A total of 40 NNI points were calculated for each subject. To obtain a more detailed plot for each condition, NNI values were recalculated using a 60-second moving window with 1-second steps. In this way, we were able to obtain a total of 542 NNI points for each condition. Figure 31 shows the average power spectra for each condition. Visual inspection of individual spectrograms showed that the 4 conditions provide almost identical spectral power.

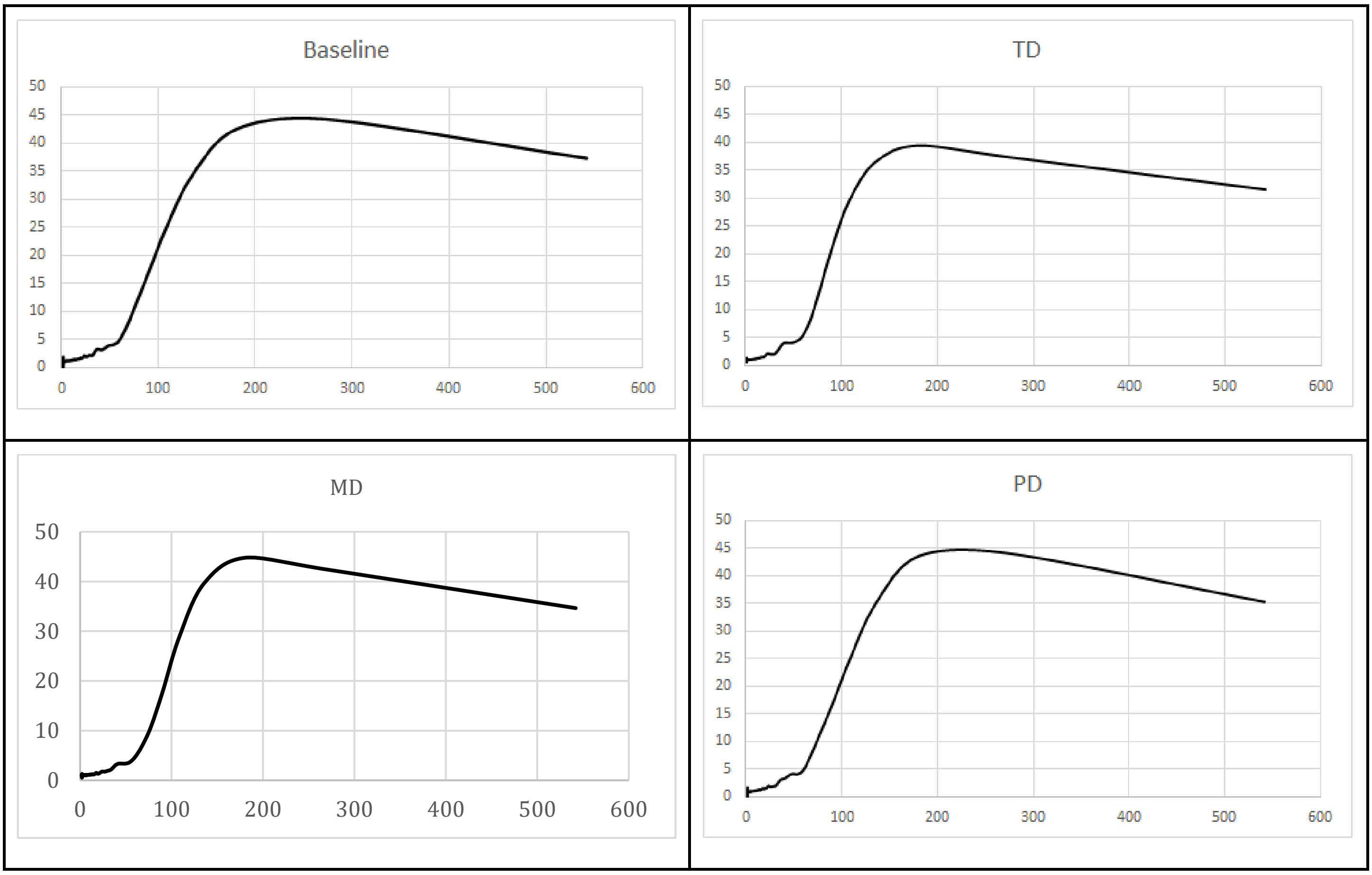

Figure 30: Power spectra representation of each condition

This result suggests that there are no observable differences in the frequency domain. However, it should be mentioned that experiments designed to study ultradian rhythms usually last for hours and make use of much longer series than those described here. The main limitation of the present account is the fact that the sessions lasted only 10 minutes.

Discussion

This latest study aimed to analyse the NNI from new perspectives, obtain new information for its interpretation concerning mental workload, and improve it by providing a more stable measure over time. In the first step, the parameter K was added to the basic algorithm. The second-and third-order K-NN was calculated to obtain a more defined outcome. The results with K = 2 showed a very similar trend compared to the classical NNI. However, it should be noted that the second-order K-NN provides data relative to the minimum average distance between a point and its k nearest points. Thus, it may be assumed that this result is more filtered (less affected by data noise) than the classical minimum average distance (Distance between a point and its nearest point). This data is certainly less sensitive to outside single fixations or data considered "dirty". In some contexts, this adjustment (K=2) may have benefits. Next, we observe that as K increases, the differences between conditions tend to decrease. The choice of the parameter K is crucial to obtain a valid result. The K should be based on the area's size, the type of task that may affect the fixation parameters and their sampling. Validation studies will be needed to define this process better.

Regarding spectral analysis, previous studies (Di Nocera, Ranvaud & Pasquali, 2015) report differences between low, mid, and high frequencies during flight operations. Different spectra have been observed about different flight phases, on total recordings of 38 minutes each. However, as already highlighted by the authors, usually spectral analysis is performed using a much longer time series, ranging from a few hours to several recordings. The result obtained in the latter analysis thus suggests that NNI does not vary in terms of frequency between conditions with the low or high mental workload. Future studies should be designed to specifically approach the oscillatory pattern examined here and compare it with that observed in prolonged vigilance tasks. This could be accomplished, for example, by adding a secondary reaction time task to understand whether or not the cyclic patterns in eye movements and performance data are comparable.

Chapter 4 - Discussions & Conclusions

One of the most important challenges for research in human-computer interaction is the creation of systems capable of understanding human behaviour because, in high-risk environments such as those mentioned above, unexpected events can only be dealt with by an operator. Therefore, it is essential to have continuous monitoring of the latter's mental and physical state, with the aim of the system to act where necessary.

The parallel measurement of all ocular indices is particularly relevant if we consider the ocular activity like similar of multidimensional nature of the mental workload (Neumann and Lipp, 2002). The application of ocular indices in real-world settings (e.g., in air traffic control) is minimal. Studies in the literature, in most cases, have focused on exploring the effects of the task on the visual search strategy. Furthermore, very few studies have investigated these effects outside of simulated environments.

Much research has been done to find a relationship between operator psychophysiology and perceived mental workload. As mentioned earlier, all of the indices investigated cannot be directly related to the workload. Therefore, to obtain an objective and reliable indicator, it is necessary to work towards a model that integrates several psychophysiological measures. However, such a model's construction is complex because the reliability and relative importance of the different measures are difficult to define. Eye-tracking can provide a valuable addition to the determination of the level of automation. However, when it concerns the practical application of the idea, several problems arise because many factors influence eye movements' workload and properties. In fact, during an experiment, these aspects can be kept as constant as possible. In a real environment, we cannot expect the operator to avoid drinking caffeinated beverages during a navigation operation. These aspects affected the workload indices, such as pupillary diameter.

This thesis reported a set of four studies designed to shed light on the relationship between mental workload and ocular scanning. This topic has been covered in the Human Factors / Ergonomics literature by using different approaches, but a complete understanding of that relationship is still a long way off. Previous studies of our laboratory have explored the opportunity to use the distribution of eye fixations as an indicator of mental workload. The Nearest Neighbour Index, a spatial statistic providing information about the distribution of points into a 2-dimensional space, was found to be sensitive to variations in mental workload. However, results obtained using the NNI were apparently different from those obtained in accredited studies using scanning randomness or entropy for summarizing the scanpath, therefore questioning the value of this approach. Di Nocera and Bolia (2007) have initially speculated that two processes respectively contribute to dispersion and grouping of the fixations: the temporal demand (that was manipulated in the NNI studies) and the visuo-spatial demand (that was manipulated in other studies, including those featuring entropy). That idea was partially tested by Camilli, Terenzi, and Di Nocera (2008) in a small between-subject study, but never deepened since then.

Indirect measures of mental workload (they all are) can be sensitive to variations in the task-load imposed on the individual. Many of them can provide only a coarse distinction between task-load levels, others have been reported to be more fine-grained. Nonetheless, sensitivity to task-load variation is not the only important property of a successful indicator: sensitivity to different types of task demands is also important. Indeed, what we call mental workload (independently of its conceptualization) may be generated in response to changes in the task-load that may be due to changes in the visuo-spatial component of the task (i.e. the task becomes more demanding because the individual need to look more, to find more, to discriminate more) or the task-load may be due to changes in the temporal component of the task (i.e. the task becomes faster, the interval between incoming stimuli becomes shorter, the time pressure for responding increases). The different types of demand are well represented by the NASA-TLX that features three scales named mental demand, temporal demand, physical demand (while the other three scales represent the individual reaction in terms of performance, effort, and frustration).

The first study reported here was designed to test the diagnosticity of the NNI, that is how the fixations distribution varied not only along with the task-load but also with the type of task demand. The results showed high clustering when the task-load increment was obtained by changing the mental (visuo-spatial) demand, and low clustering when it was obtained by changing the temporal demand. The physical demand, instead, did not affect the scanpath, possibly because our manipulation of this dimension was not appropriate or because the ocular behaviour is not sensitive to the manipulation of the physical component. Albeit results showed a significant increase in the subjective estimates of physical demand, the effect did not extend to the overall workload ratings nor to the analysis of the scanpath. Likely, the Tetris game involved minimal physical effort and the manipulation was not effective. To overcome this limitation, future studies could consider several options. One potential solution could be to manipulate the game controls producing frequent keypress failures in the high task-load condition. Alternatively, the keypress force could be manipulated in the high task-load condition to make the task more effortful.

In the second study, instead, the NNI was directly compared to the entropy approach that is considered one of the most prominent techniques for studying the scanpath in the HF/E domain. Results showed an overall increase of difficulty after the first few minutes of performance that reflected in both measures of mental workload. After two minutes, the search task generated both a stereotyped dwell pattern (consistent with the entropy prediction) and fixations grouping (consistent with the fixations distribution prediction). In other words, the two indices were found to be both sensitive to changes in the visuo-spatial demand and the plots were highly overlapping. Such a result sorts out the issue of the differences found between the two indicators, showing how that exclusively depends on the type of demand imposed. Also, results demonstrated that a dispersed fixation pattern (or moderately grouped) is not equivalent to high randomness in visual exploration. The two scanpath analysis algorithms show the same trend. From the post-hoc analyses, a cut-off of the values is observed starting in the fourth minute. However, compared to the performance and mental workload data, the cut-off occurs only after the sixth minute of activity. This difference suggests that the change in visual exploration strategy anticipated the decline in-game performance on the fourth minute. Further studies should be conducted to confirm this effect. In high-risk settings, the anticipation of critical events and the operator's mental state is essential.

In the third study, the autopilot function was introduced, to test the right level of automation and mental workload effects (calculated by performance measures, Nasa-TLX and NNI). The goals were 1) to determine whether the scanpath analysis via NNI could be used as a trigger for adaptive automation, and 2) to identify the optimal level of automation to be applied. An "autopilot", able to take total control of the system, was designed as the best solution to avoid game-over in critical situations. The study's design involved the use of two independent variables at two levels, including four experimental conditions. The “Difficulty” variable (i.e. Easy, Hard) and the “Manual Automation” variable (i.e. Present, Absent).

Following the first two studies, it is possible to observe a significant difference given by the Difficulty factor, in both performance and mental workload (obtained via Nasa-TLX). However, there is no difference in the scores of the Nasa-TLX between the conditions with and without automation. This result is partially in conflict with the hypotheses. In the conditions with autopilot, better performance and lower mental workload were expected. While in performance terms, the result matches the assumption, subjects do not perceive differences between the conditions with and without automation. However, it should be noted that while subjects could switch automation on/off at any time by pressing the CTRL key (only in the conditions where available, EMA and HMA), not all were able to use it correctly. The autopilot was often activated in highly critical situations, and this did not allow the automation to avoid game over. On the other hand, the NNI analysis results showed a significant interaction effect between the difficulty factor and automation. From the post-hoc analyses, we can observe that autopilot conditions (easy and hard) do not differ from the easy condition without automation (EN). This suggests that automation was effective in helping the subject in terms of performance and NNI.

The fourth study aimed to further verify the use of NNI as a possible trigger in adaptive systems. Unlike the previous one, single values were considered and not the average of the whole session. Therefore, a series of “3-min units” was defined, allowing us to observe what happens, in terms of ocular strategies, before and after the present/absence of automation. The presence or absence of automation support was classified with four categories: Provided when needed **(NP)**; Provided when not needed **(nNP)**; Not provided when needed **(NA)**; Not provided when not needed **(nNA)**. The proportion of within-range values is expected to be significantly higher in the valid conditions (NP and nNA) rather than in the invalid conditions (NA and nNP). The results showed high variability, probably caused by the low number of trials (only 15) and the high dynamism of the task, where a single error can lead to game-over. However, it should be noted that this experiment was based on the calculation of a 10-minute baseline that allowed us to calculate threshold values. Future studies should consider a more durable baseline to provide more accurate values for each subject.

In conclusion, two new methods of NNI analysis were tested. The first one considered the K-NN algorithm as a starting point. The K-factor was then introduced within the NNI algorithm to define the minimum average distance between points as "the minimum average distance between a point and its K closest points". This analysis aimed to optimize the calculation of NNI scores by limiting the effect caused by "wrong" fixations often caused by sampling errors. The analyses compared NNI scores with those measured by first- and second-order K-NN. The results with $K = 2$ showed a very similar trend compared to the classical NNI. Instead, with $K = 3$ the differences between conditions tend to decrease. The choice of the parameter K is crucial to obtain a valid and indicative result. This should be based on the area's size, the type of task that may affect the fixation parameters and their sampling. Subsequently, NNI points (provided by 60s moving windows with step of 1s) are being used in a spectral analysis. The result suggests that NNI does not vary in terms of frequency between conditions with the low or high mental workload. However, as already highlighted in a previous study (Di Nocera, Ranvaud & Pasquali, 2015), usually spectral analysis is performed using much longer time series ranging from a few hours to several days of recordings.

The NNI seems to be a good indicator of mental workload, but only under specific conditions:

1) First, it is necessary to estimate a single operator's baseline while performing a specific task (i.e., air traffic control). It should be considered that the interfaces have different visual characteristics. For example, during a driving task, the driver has to watch the fuel level, a navigation system, mirrors, and road. These elements require his attention and, therefore, the generated fixations will draw a specific visual area of that task. The baseline aims to identify the visual exploration strategy used during an optimal condition of the task (neither too easy nor too hard) and the subject's psychophysical state.
2) Subsequently, the scanpath is monitored and processed in real-time (minute by minute) using the Nearest neighbour Index. The values thus obtained are compared with the average value provided by the baseline. Considering the fourth study described here, the NNI values analyzed separately fluctuate within a wide range, causing a large variability without the possibility to distinguish the different conditions. However, when we analyze the average of 10 NNI values (provided by 10-minute sessions), a more stable result allows us to distinguish the difficult conditions (compared to the baseline) and the type of task demand imposed. In a realistic context, using more values generated every 5 or 10 minutes, it would provide a more accurate and stable data, less affected by sampling errors or extraneous visual elements. Supervision and control tasks that may last several hours are mentioned in many high-risk contexts, such as those mentioned at the beginning of this document. The NNI should be able to monitor the level of vigilance and mental workload perceived by the operator. To accomplish this, a psychophysiological measure generated every 10 minutes should be sufficient to reduce the risk of incidents in these contests.
3) The last recommendation concerns the proper level of automation to be applied when an anomaly in the NNI plot is detected. In the Tetris used here, the right automation would be slowing down the piece's descent (using the same speed as the easy level as set out in study 1, 3, and 4). However, this system does not seem realistic outside of a laboratory setting. In addition, any automation changes the nature of the task, and risks to invalidate the previously calculated baseline. A possible solution is to calculate a second baseline, to be used when the automation is active. Future studies should consider these aspects, to use the NNI as a trigger in an adaptive automation system.

References

Antes, J. R. (1974). The time course of picture viewing. Journal of experimental psychology, 103(1), 62.

Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Pozzi, S., and Babiloni, F. (2016). A passive Brain-Computer Interface (p-BCI) application for the mental workload assessment on professional Air Traffic Controllers (ATCOs) during realistic ATC tasks. Prog. Brain Res. 228, 295–328. doi: 10.1016/bs.pbr.2016.04.021

Bauer, L. O., Goldstein, R., & Stern, J. A. (1987). Effects of information-processing demands on physiological response patterns. Human Factors, 29(2), 213-234.

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

Beatty, J., & Kahneman, D. (1966). Pupillary changes in two memory tasks. Psychonomic Science, 5(10), 371-372.

Bernardi, L., Wdowczyk-Szulc, J., Valenti, C., Castoldi, S., Passino, C., Spadacini, G., & Sleight, P. (2000). Effects of controlled breathing, mental activity and mental stress with or without verbalization on heart rate variability. Journal of the American College of Cardiology, 35(6), 1462-1469.

Billings, C. E. (1991). Human-centered aircraft automation: A concept and guidelines (Vol. 103885). National Aeronautics and Space Administration, Ames Research Center.

Boehm-Davis, D. A., Gray, W. D., & Schoelles, M. J. (2000, July). The eye blink as a physiological indicator of cognitive workload. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 44, No. 33, pp. 6-116). Sage CA: Los Angeles, CA: SAGE Publications.

Borghini, Andrea & Gianicolo, Emilio & Picano, Eugenio & Andreassi, Maria Grazia. (2013). Ionizing radiation and atherosclerosis: Current knowledge and future challenges. Atherosclerosis. 230. 40-7. 10.1016/j.atherosclerosis.2013.06.010.

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. Neuroscience & Biobehavioral Reviews, 44, 58-75.

Brain, B., 1998. The Luddite Rebellion. NYU Press.

Brookhuis, K.A., and D. De Waard. 2000. "Assessment of drivers' workload: performance, subjective and physiological indices." In Stress, Workload and Fatigue, edited by P.A. Hancock and P.A. Desmond, 321-333. New Jersey: Lawrence Erlbaum.

Brookhuis, K.A., D. De Waard, and S. H. Fairclough. 2003. "Criteria for driver impairment." Ergonomics 46: 433-445.

Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. Biological psychology, 42(3), 361-377.

Byrne, E. A., & Parasuraman, R. (1996). Psychophysiology and adaptive automation. Biological psychology, 42(3), 249-268.

Camilli, M., Nacchia, R., Terenzi, M., & Di Nocera, F. (2008) ASTEF: A Simple Tool for Examining Fixations. Behaviour Research Methods, 40 (2), 373–382.

Camilli, M., Terenzi, M., & Di Nocera, F. (2007). Concurrent validity of an ocular measure of mental workload. In D. de Waard, G.R.J. Hockey, P. Nickel, and K.A. Brookhuis (Eds.), Human Factors Issues in Complex System Performance (pp. 117-129). Maastricht, the Netherlands: Shaker Publishing.

Camilli, M., Terenzi, M., & Di Nocera, F. (2008). Effects of temporal and spatial demands on the distribution of eye fixations. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 52(18), 1248-1251."

Chanel, G., Rebetez, C., Bétrancourt, M., & Pun, T. (2011). Emotion assessment from physiological signals for adaptation of game difficulty. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 41(6), 1052-1063.

Christopher D. Manning, Hinrich Schuetze - Foundations of Statistical Natural Language Processing (1999, The MIT Press)

Clark, PJ, Evans, RC, 1954 “Distance to nearest neighbour as a measure of spatial relationships in populations” Ecology 35 445-453

de Greef, T., Lafeber, H., van Oostendorp, H., & Lindenberg, J. (2009, July). Eye movement as indicators of mental workload to trigger adaptive automation. In International Conference on Foundations of Augmented Cognition (pp. 219-228). Springer, Berlin, Heidelberg.

Degani, A., & Wiener, E. L. (1994). On the design of flight-deck procedures.

Dekker, S. (2003). Failure to adapt or adaptations that fail: contrasting models on procedures and safety. Applied Ergonomics, 34(3), 233-238.

Delaitre, Antoine. "Nearest neighbour Index." IB Geography. N.p., 2011. Web, http://www.geoib.com/nearest-neighbour-index.html. Accessed 20 December 2020

Di Nocera (2011). Ergonomia Cognitiva. Roma, Carocci editore.

DI NOCERA, F., & BOLIA, R. S. (2007). PERT networks as a method for analyzing the visual scanning strategies of aircraft pilots. Proceedings of the 14th International Symposium on Aviation Psychology (pp. 165–169). Dayton, OH: Wright State University.

Di Nocera, F., Camilli, M., & Terenzi, M. (2007) A random glance at the flight deck: Pilots' scanning strategies and the real-time assessment of mental workload. Journal of Cognitive Engineering and Decision Making, 1(3), 271-285.

Di Nocera, F., Ranvaud, R., & Pasquali, V. (2015) Spatial pattern of eye fixations and evidence of ultradian rhythms in aircraft pilots. Aerospace Medicine and Human Performance, 86(7), 647-651.

Di Stasi, L. L., McCamy, M. B., Martinez-Conde, S., Gayles, E., Hoare, C., Foster, M., ... & Macknik, S. L. (2016). Effects of long and short simulated flights on the saccadic eye movement velocity of aviators. Physiology & behaviour, 153, 91-96.

Di Stasi, L. L., Renner, R., Staehr, P., Helmert, J. R., Velichkovsky, B. M., Cañas, J. J., ... & Pannasch, S. (2010). Saccadic peak velocity sensitivity to variations in mental workload. Aviation, Space, and Environmental Medicine, 81(4), 413-417.

Diebold, J. (1952). Automation. The Advent of the Automatic Factory. New York

Diebold, J. (1983). Automation. American Management Assoc., Inc..

Dillard, M. B., Warm, J. S., Funke, G. J., Funke, M. E., Finomore, V. S., Matthews, G., Parasuraman, R. (2014) The Sustained Attention to Response Task (SART) Does Not Promote Mindlessness During Vigilance Performance. Human Factors, 56(8), 1364–1379.

Ellis, K. K. E. (2009). Eye tracking metrics for workload estimation in flight deck operations. Theses and Dissertations, 288.

Endsley, M. R. (1987). The application of human factors to the development of expert systems for advanced cockpits. In Proceedings of the Human Factors Society 31st Annual Meeting (pp. 1388-1392). Santa Monica, CA: Human Factors and Ergonomics Society.

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. In Proceedings of the Human Factors Society 32nd Annual Meeting (pp. 97-101). Santa Monica, CA: Human Factors and Ergonomics Society.

Endsley, M. R., & Kaber, D. B. (1997). The use of level of automation as a means of alleviating out-of-the-loop performance problems: a taxonomy and empirical analysis. In P. Seppala, T. Luopajarvi, C. H. Nygard, & M. Mattila (eds.), 13th Triennial Congress of the International Ergonomics Association (Vol. 1, pp. 168-170). Helsinki: Finnish Institute of Occupational Health.

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop per-formance problem and level of control in automation. Human Factors, 37, 381–394. doi:10.1518/001872095779064555

Endsley, M.R., Kaber, D., 1999. Level of automation effects on performance, situation awareness and workload in a dynamic control task. Ergonomics 42, 462e492.

Fallahi, M., Motamedzade, M., Heidarimoghadam, R., Soltanian, A. R., & Miyake, S. (2016). Assessment of operators’ mental workload using physiological and subjective measures in cement, city traffic and power plant control centers. Health promotion perspectives, 6(2), 96.

Fidopiastis, C. M., Drexler, J., Barber, D., Cosenzo, K., Barnes, M., Chen, J. Y., & Nicholson, D. (2009, July) Impact of automation and task-load on unmanned system operator’s eye movement patterns. In International Conference on Foundations of Augmented Cognition, pp. 229-238, Springer, Berlin, Heidelberg.

Findlay, J. M., Findlay, J. M., & Gilchrist, I. D. (2003). Active vision: The psychology of looking and seeing (No. 37). Oxford University Press. Fischer, A. J., McGuire, J. J., Schaeffel, F., & Stell, W. K. (1999). Light-and focus-dependent expression of the transcription factor ZENK in the chick retina. Nature neuroscience, 2(8), 706.

Flemisch, Frank & Onken, R.. (2002). Open a Window to the Cognitive Work Process! Pointillist Analysis of Man–Machine Interaction. Cognition Technology and Work. 4. 160-170. 10.1007/s101110200015.

Fogarty, C., & Stern, J. A. (1989). Eye movements and blinks: their relationship to higher cognitive processes. International Journal of Psychophysiology, 8(1), 35-42.

Furuta, K., Sasou, K., Kubota, R., Ujita, H., Shuto, Y., Yagi, E., 2000. Human factor analysis of JCO criticality accident. Cognition Technol. Work 2 (4), 182–203.

G. Underwood, D. Crundall and P. Chapman. (2011) Driving simulator validation with hazard perception. Transportation Research Part F: Traffic Psychology and Behavior, 14, 435-446

Gabbard, R. D., Fendley, M., Dar, I. A., Warren, R., & Kashou, N. H. (2017). Utilizing functional near-infrared spectroscopy for prediction of cognitive workload in noisy work environments. Neurophotonics, 4(4), 041406.

Gaillard, A. W. K., & Wientjes, C. J. E. (1994). Mental load and work stress as two types of energy mobilization. Work & Stress, 8(2), 141-152.

Galbraith J (1973) Designing complex organizations. Addison-Wesley, Reading, MA

Galpin, A. J., & Underwood, G. (2005). Eye movements during search and detection in comparative visual search. Perception & Psychophysics, 67(8), 1313-1331.

Geddes, N. D. (1985). Intent inferencing using scripts and plans. In Proceedings of the First Annual Aerospace Applications of Artificial Intelligence Conference (pp. 160-172). Wright-Patterson Air Force Base, OH: U.S. Air Force.

Gevins, A., & Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. Theoretical Issues in Ergonomics Science, 4(1-2), 113-131.

Goldstein, R., Walrath, L. C., Stern, J. A., & Strock, B. D. (1985). Blink activity in a discrimination task as a function of stimulus modality and schedule of presentation. Psychophysiology, 22(6), 629-635.

Greenstein, J. S., & Revesman, M. E. (1986). Two simulation studies investigating means of human-computer communication for dynamic task allocation. IEEE transactions on systems, man, and cybernetics, 16(5), 726-730.

Guasti, L., Simoni, C., Mainardi, L., Crespi, C., Cimpanelli, M., Klersy, C., ... & Venco, A. (2005). Global link between heart rate and blood pressure oscillations at rest and during mental arousal in normotensive and hypertensive subjects. Autonomic Neuroscience, 120(1-2), 80-87.

Hancock, P. A., & Chignell, M. H. (1988). Mental workload dynamics in adaptive interface design. IEEE transactions on Systems, Man, and Cybernetics, 18(4), 647-658.

Hancock, P. A., & Szalma, J. L. (Eds.). (2008). Performance under stress. Ashgate Publishing, Ltd..

Harmat, L., de Manzano, Ö., Theorell, T., Högman, L., Fischer, H., & Ullén, F. (2015). Physiological correlates of the flow experience during computer game playing. International Journal of Psychophysiology, 97(1), 1-7

Harris, R.L., Glover, B.J., Spady, A.A. (1986) Analytical Techniques of Pilot Scanning Behavior and Their Application (Report No. NASA TP-2525). NASA Langley Research Center, Hampton, VA, US Retrieved from. https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19860018448.pdf

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), Advances in psychology, 52. Human mental workload 139–183. North-Holland. https://doi.org/10.1016/S0166-4115(08)62386-9

Hessels, R. S., Niehorster, D. C., Kemner, C., & Hooge, I. T. (2017). Noise-robust fixation detection in eye movement data: Identification by two-means clustering (I2MC). Behavior research methods, 49(5), 1802-1823.

Hilburn, B., Jorna, P. G., Byrne, E. A., & Parasuraman, R. (1997). The effect of adaptive air traffic control (ATC) decision aiding on controller mental workload. Human-automation interaction: Research and practice, 84-91.

Hockey, G. R. J. (1986). A state control theory of adaptation and individual differences in stress management. In Energetics and human information processing. Springer, Dordrecht, 285-298.

Hockey, G. R. J., Gaillard, A. W. K., & Burov, O. (Eds.) (2003). Operator functional state: the assessment and prediction of human performance : degradation in complex tasks. Amsterdam: IOS Press.

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

Hwang, S. L., Yau, Y. J., Lin, Y. T., Chen, J. H., Huang, T. H., Yenn, T. C., & Hsu, C. C. (2008). Predicting work performance in nuclear power plants. Safety science, 46(7), 1115-1124.

Imandoust, S. B., & Bolandraftar, M. (2013). Application of k-nearest neighbour (knn) approach for predicting economic events: Theoretical background. International Journal of Engineering Research and Applications, 3(5), 605-610.

Inagaki, T. (2003). Adaptive automation: Sharing and trading of control. In Handbook of cognitive task design (pp. 171-194). CRC Press.

Iqbal, S. T., Adamczyk, P. D., Zheng, X. S., & Bailey, B. P. (2004). Changes in mental workload during task execution. In Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology.

J. Callan, Daniel. (1998). Eye Movement Relationships to Excessive Performance Error in Aviation. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 42. 1132-1136.

Jorna, P. G. A. M. (1997, June). Human Machine interfaces for ATM: objective and subjective measurements on human interactions with future Flight deck and Air Traffic Control systems. In FAA/Eurocontrol ATM R&D Seminar, Paris, France.

Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: pupillometric indices of sentence processing. Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 47(2), 310.

Kaber, D. B. (1997). The Effect of Level of Automation and Adaptive Automation on Performance in Dynamic Control Environments (Tech. Work. Doc. No. ANRCP-NG-ITWD-97-01). Amarillo, TX: Amarillo National Resource Center for Plutonium.

Kaber, D. B., & Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. Theoretical Issues in Ergonomics Science, 5(2), 113-153.

Kahneman, D. (1973). Attention and effort (Vol. 1063). Englewood Cliffs, NJ: Prentice-Hall.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. Science, 154(3756), 1583-1585.

Karpov, B. A., Luria, A. R., & Yarbuss, A. L. (1968). Disturbances of the structure of active perception in lesions of the posterior and anterior regions of the brain. Neuropsychologia, 6(2), 157-166.

Kosch, T., Hassib, M., Buschek, D., & Schmidt, A. (2018, April). Look into my eyes: using pupil dilation to estimate mental workload for task complexity adaptation. In Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (pp. 1-6).

Lautman, L., and Gallimore, P. L. (1987). Control of the crew caused accident: Results of a 12-operator survey. Boeing Airliner, April-June, 1-6.

Léger, P. M., Davis, F. D., Cronan, T. P., & Perret, J. (2014). Neurophysiological correlates of cognitive absorption in an enactive training context. Computers in Human Behavior, 34, 273-283.

Lindstedt, J., & Gray, W. (2013, January). Extreme expertise: Exploring expert behavior in Tetris. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 35, No. 35).

Lipshitz R, Strauss O (1997) Coping with uncertainty: a naturalistic decision-making analysis. Organizat Behav Hum Decis Process 69:149–163

Liu, Y., Ayaz, H., & Shewokis, P. A. (2017). Mental workload classification with concurrent electroencephalography and functional near-infrared spectroscopy. Brain-Computer Interfaces, 4(3), 175-185.

Loft, S., Sanderson, P., Neal, A., Mooij, M., 2007. Modeling and predicting mental workload in en-route air traffic control: Critical review and broader implications. Human Factors, 49, 376–399.

Louw, T., & Merat, N. (2017). Are you in the loop? Using gaze dispersion to understand driver visual attention during vehicle automation. Transportation Research Part C: Emerging Tech-nologies, 76, 35–50. doi:10.1016/j.trc.2017.01.001

Maggi P., Ricciardi O., Di Nocera F. (2019) Ocular Indicators of Mental Workload: A Comparison of Scanpath Entropy and Fixations Clustering. In: Longo L., Leva M. (eds) Human Mental Workload: Models and Applications. H-WORKLOAD 2019. Communications in Computer and Information Science, vol 1107. Springer, Cham. https://doi.org/10.1007/978-3-030-32423-0_13

Marois, R., & Ivanoff, J. (2005). Capacity limits of information processing in the brain. Trends in cognitive sciences, 9(6), 296-305.

Marshall, S. P. (2007). Identifying cognitive state from eye metrics. Aviation, space, and environmental medicine, 78(5), B165-B175.

McIntire, L. K., McKinley, R. A., Goodyear, C., & McIntire, J. P. (2014). Detection of vigilance performance using eye blinks. Applied Ergonomics, 45(2), 354-362.

Mehler, B., Reimer, B., Coughlin, J. F., and Dusek, J. A. (2009). The impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. Transport. Res. Rec. 2138, 6–12. doi: 10.3141/2138-02

Metzger, U., and R. Parasuraman. 2005. “Automation in Future Air Traffic Management: Effects of Decision Aid Reliability on Controller Performance and Mental Workload.” Human Factors 47(1): 35-49.

Milliken FJ (1987) Three types of perceived uncertainty about the environment – state, effect, and response uncertainty. Acad Manag Rev 12:133–143

Morrison, J. G., Gluckman, J. P. (1994). Definitions and Prospective Guidelines for the Application of Adaptive Automation. In Mouloua, M., Parasuraman, R. (Eds.) Human performance in automated systems: Current research and trends. Hillsdale, NJ: Lawrence Erlbaum Associates.

National Transport Safety Board, 2013a. Crash During a Nighttime Nonprecision Instrument Approach to Landing UPS Flight 1354 Airbus A300–600, N155UP Birmingham, Alabama August 14, 2013 (NTSB/AAR-14/02 PB2014-107898). Author, Washington.

National Transport Safety Board, 2013b. Descent Below Visual Glidepath and Impact With Seawall, Asiana Airlines Flight 214 Boeing 777–200ER, HL7742, San Francisco, California July 6, 2013 (NTSB/AAR-1401 PB 2014–105984). Author, Washington.

Neumann, D. L., & Lipp, O. V. (2002). Spontaneous and reflexive eye activity measures of mental workload. Australian Journal of Psychology, 54, 174–179.

Ntuen, C. A., & Park, E. H. (1988, October). Human factor issues in teleoperated systems. In Proceedings of the First International Conference on Ergonomics of Hybrid Automated Systems I (pp. 203-210).

Orchard, L. N., & Stern, J. A. (1991). Blinks as an index of cognitive activity during reading. Integrative Physiological and Behavioral Science, 26(2), 108-116.

Othman, N., & Romli, F. I. (2016). Mental workload evaluation of pilots using pupil dilation. International Review of Aerospace Engineering, 9, 80-84.

Pakarinen, S., Korpela, J., Torniainen, J., Laarni, J., & Karvonen, H. (2018). Cardiac measures of nuclear power plant operator stress during simulated incident and accident scenarios. Psychophysiology, 55(7), 1-15.

Palinko, O., Kun, A. L., Shyrokov, A., & Heeman, P. (2010, March). Estimating cognitive load using remote eye tracking in a driving simulator. In Proceedings of the 2010 symposium on eye-tracking research & applications (pp. 141-144).parasuraman

Pannasch, S., Helmert, J. R., Roth, K., Herbold, A. K., & Walter, H. (2008). Visual fixation durations and saccade amplitudes: Shifting relationship in a variety of conditions. Journal of Eye Movement Research, 2(2). Peavler, W. S. (1974). Pupil size, information overload, and performance differences. Psychophysiology, 11(5), 559-566.

Parasuraman, R., & Hancock, P. A. (2001). Adaptive control of mental workload.

Parasuraman, R., Bahri, T., Deaton, J. E., Morrison, J. G., & Barnes, M. (1992). Theory and design of adaptive automation in aviation systems. CATHOLIC UNIV OF AMERICA WASHINGTON DC COGNITIVE SCIENCE LAB.

Parasuraman, R., Riley, V., (1997). Humans and automation: use, misuse, disuse, abuse. Hum. Factors 43, 250e253.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model of types and levels of human interaction with automation. IEEE Transactions on Systems, Man and Cybernetics, 30(3), 286-297.

Peavler, W. S. (1974). Pupil size, information overload, and performance differences. Psychophysiology, 11(5), 559-566.

Pfaff, U., Fruhstorfer, H., & Peter, H. H. (1976). Changes in eyeblink duration and frequency during car driving. Pfugers Archives, 363, R 21.

Pinder, D. A., & Witherick, M. E. (1975). A modification of nearest-neighbour analysis for use in linear situations. Geography, 16-23.

Porter, G., Troscianko, T., & Gilchrist, I. D. (2007). Effort during visual search and counting: Insights from pupillometry. The Quarterly Journal of Experimental Psychology, 60(2), 211-229.

Prinzel, L. J. (2003). Team-centered perspective for adaptive automation design. Washington, DC: National Aeronautics and Space Administration, Langley Research Center.

R.W. Bailey Human Performance Engineering: A Guide for System Designers, Prentice-Hall, Englewood Cliffs, NJ (1982)

Ranchet, M., Morgan, J. C., Akinwuntan, A. E., & Devos, H. (2017). Cognitive workload across the spectrum of cognitive impairments: A systematic review of physiological measures. Neuroscience & Biobehavioral Reviews, 80, 516-537.

Rayner, K. (2009). The Thirty-Fifth Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. The Quarterly Journal of Experimental Psychology, 62, 1457–1506.

Reimer, B., & Mehler, B. (2011). The impact of cognitive workload on physiological arousal in young adult drivers: a field study and simulation validation. Ergonomics, 54(10), 932-942.

Rezaei, M., & Klette, R. (2011). Simultaneous analysis of driver behaviour and road condition for driver distraction detection. International Journal of Image and Data Fusion, 2(3), 217-236.

Rouse, W. B., Geddes, N. D., and Curry, R. E. (1987). An architecture for intelligent interfaces: Outline of an approach to supporting operators of complex systems. Human-Computer Interaction, 3,87-122.

Salvucci, D. D., & Goldberg, J. H. (2000, November). Identifying fixations and saccades in eye-tracking protocols. In Proceedings of the 2000 symposium on Eye tracking research & applications. 71-78

Scerbo, M. W. (1996). Theoretical perspectives on adaptive automation.

Sheridan, T. B. (1997). Supervisory control. In G. Salvendy (ed.), Handbook of human factors (pp. 1295-1327). New York: John Wiley & Sons.

Sheridan, T.B., Verplank, W.L., (1978). Human and Computer Control of Undersea Teleoperators. Massachussetts Institute of Technology, Cambridge, Massachusetts: Man-Machine Systems Laboratory, Department of Mechanical Engineering.

Singh, H., & Singh, J. (2012). Human eye tracking and related issues: A review. International Journal of Scientific and Research Publications, 2(9), 1-9.

Sirevaag, E., Kramer, A., deJong, R., & Mecklinger, A., (1988). A psychophysiological analysis of multi-task processing demands. Psychophysiology, 25, 482.

Smith, C. E. (1998). Modeling high sinuosity meanders in a small flume. Geomorphology, 25(1-2), 19-30.

Smith, K., Scallen, S. F., Knecht, W., & Hancock, P. A. (1998). An Index of Dynamic Density. Human Factors, 40(1), 69–78. https://doi.org/10.1518/001872098779480604

Stark, Barbara & Young, Dennis. (1981). Linear Nearest neighbour Analysis. American Antiquity. 46. 284. 10.2307/280209.

Stern, J. A. (1988). Blink rate: A possible measure of fatigue. Human Factors, 36,285-297.

Stern, J. A., Walrath, L. C., & Goldstein, R. (1984). The endogenous eyeblink. Psychophysiology, 21(1), 22-33.

Tattersall, A. J., & Foord, P. S. (1996) An experimental evaluation of instantaneous self-assessment as a measure of workload. Ergonomics, 39(5), 740-748.

Thompson JD (1967) Organizations in action. McGraw-Hill, New York

Tole, J. R., Stephens, A. T., Vivaudou, M., Ephrath, A. R., & Young, L. R. (1983) Visual scanning behaviour and pilot workload.

Tole, J.R., Harris, R.L., Stephens, A.T., Ephrath, A.R.(1982). Visual scanning behavior and mental workload in aircraft pilots. Aviation, Space, and Environmental Medicine 53, 54-61

Tombu, Mike & Asplund, Christopher & Dux, Paul & Godwin, Douglass & Martin, Justin & Marois, Rene. (2011). A Unified attentional bottleneck in the human brain. Proceedings of the National Academy of Sciences of the United States of America. 108. 13426-31. 10.1073/pnas.1103583108.

Trimmel, M., & Huber, R. (1998). After-effects of human-computer interaction indicated by P300 of the event-related brain potential. Ergonomics, 41(5), 649-655.

Tsang, P. S., & Vidulich, M. A. (2006). Mental workload and situation awareness.

Unema, P. J. A., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. Visual Cognition, 12(3), 473-494.

Vagia, Marialena; Transeth, Aksel A.; Fjerdingen, Sigurd A. (2015). A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed?. Applied Ergonomics, Vol 53(Part A), pp. 190-202.

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. Psychonomic bulletin & review, 25(6), 2005-2015.

Van Orden, K. F., Limbert, W., Makeig, S., & Jung, T. P. (2001). Eye activity correlates of workload during a visuospatial memory task. Human Factors, 43(1), 111-121.

Velichkovsky, B. M., Dornhoefer, S. M., Pannasch, S., & Unema, P. J. A. (2000). Visual fixations and level of attentional processing. In A. Duhowski (Ed.), Proceedings of the Symposium on Eye Tracking Research and Applications (pp. 79-85). Palm Beach Gardens, NY: ACM Press.

Velichkovsky, B. M., Joos, M., Helmert, J. R., & Pannasch, S. (2005). Two visual systems and their eye movements: Evidence from static and dynamic scene perception. In B. G. Bara, L. Barsalou & M. Bucciarelli (Eds.), Proceedings of the XXVII Conference of the Cognitive Science Society (pp. 2283-2288). Mahwah, NJ: Lawrence Erlbaum.

Velichkovsky, B. M., Rothert, A., Kopf, M., Dornhoefer, S. M., & Joos, M. (2002). Towards an express diagnostics for level of processing and hazard perception. Transportation Research, Part F, 5(2), 145-156.

Viswanadham, N. (2002). The past, present, and future of supply-chain automation. IEEE Robotics & Automation Magazine, 9(2), 48-56.

Wang, D., Mulvey, F. B., Pelz, J. B., & Holmqvist, K. (2017). A study of artificial eyes for the measurement of precision in eye-trackers. Behavior Research Methods, 49(3), 947-959.

Wickens, C. D. (1992). Engineering psychology and human performance. (2nd ed.) New York, NY: Harper Collins.

Wickens, C. D. (2008). Multiple resources and mental workload. Human factors, 50(3), 449-455.

Wierwille, W. W., & Eggemeier, F. T. (1993). Recommendations for mental workload measurement in a test and evaluation environment. Human factors, 35(2), 263-281.

Wierwille, W.W., Rahimi, M., & Casali, J.G. (1985). Evaluation of 16 measures of mental workload using a simulated flight task emphaizing mediational activity. Human Factors, 25, 1-16.

Wilson, G. F., & Eggemeier, F. T. (1991). Physiological measures of workload in multi-task environments. In D. Damos (Ed.), Multiple-task performance (pp. 329–360). London: Taylor & Francis.

Wilson, G. F., & Russell, C. A. (2003). Operator functional state classification using multiple psychophysiological features in an air traffic control task. Human Factors, 45, 381–389.

Wilson, G. F., Purvis, B., Skelly, J., Fullenkamp, P., & Davis, I. (1987). Physiological data used to measure pilot workload in actual flight and simulator conditions. Proceedings of the Human Factors Society 31st Annual Meeting (pp. 779-783). Santa Monica, CA: Human Factors Society.

Wolfe, J. M., Reinecke, A., & Brawn, P. (2006). Why don't we see changes? The role of attentional bottlenecks and limited visual memory. Visual Cognition, 14(4-8), 749-780.

Yarbus A L. (1967). Eye Movements and Vision. New York: Plenum Press.

Young, M. S., and N. A. Stanton. 2005. “Mental workload.” In Handbook of Human Factors and Ergonomics Methods, edited by N. A. Stanton, A. Hedge, K. Brookhuis, E. Salas, and H. W. Hendrick, Ch. 39. London: Taylor & Francis.

Zeitlin, L. R. 1993. “Subsidiary Task Measures of Driver Mental Workload: A Long-Term Field Study.” In Driver Performance: Measurement and Modeling, IVHS, Information Systems, and Simulation, by the Transportation Research Board, National Research Council, Washington, D.C. National Academy Press, Washington, D.C., Transportation Research Record No.1403, 23-27.