REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
06-01-2007	Final Report	1 March 2005 - 29-Jan-07

4. TITLE AND SUBTITLE

Cognitive Aspects and Behavioral Effects of Transitions Between Levels of Automation

5a. CONTRACT NUMBER
FA8655-05-1-3021

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Dr. Francesco Di Nocera

5d. PROJECT NUMBER

5d. TASK NUMBER

5e. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Rome 'La Sapienza'
Via dei Marsi 78
Rome 00185
Italy

8. PERFORMING ORGANIZATION REPORT NUMBER

N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

EOARD
PSC 821 BOX 14
FPO AE 09421-0014

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
Grant 05-3021

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report results from a contract tasking University of Rome 'La Sapienza' as follows: The Grantee investigated the relation between scan path (ocular activity) and mental workload on the basis of the consideration that high workload should produce fixations grouping (because the operator needs to focus on some specific feature of the interface/task) whereas low workload should be associated with regular patterns, indicating a regular check of the interface space. According to this hypothesis, indexes providing information about the dispersion of point patterns should indicate regularity in the case of low workload and grouping in the case of high workload. The results suggest that nearest neighbor index used here is sensitive for investigating the processes underlying shifts in the level of automation, and their consequences on operator performance. On the costs of switching between levels of automation (LOA), a simple visuo-motor task employed in this study suggests that switching LOA affected individual's performance because of the cost associated with engagement/disengagement process. These findings suggest that when individuals perform a task, their cognitive systems are set to a particular level and no costs are observed until the level (or rule) is changed. Under some circumstances the results suggest that no shift can even lead to a better performance.

15. SUBJECT TERMS
EOARD, Human Factors, Operator-Machine Interface, Psychology, Eye Tracking

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18, NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE	UL		ROBERT N. KANG, Lt Col, USAF
UNCLAS	UNCLAS	UNCLAS		40	19b. TELEPHONE NUMBER *(Include area code)* +44 (0)20 7514 4437

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39-18

COGNITIVE ASPECTS AND BEHAVIORAL EFFECTS OF TRANSITIONS BETWEEN LEVELS OF AUTOMATION

Francesco Di Nocera, Marco Camilli, Michela Terenzi, Roberto Nacchia

Cognitive Ergonomics Laboratory, Department of Psychology

University of Rome "La Sapienza"

In accordance with Defense Federal Acquisition Regulation 252.227-7036, Declaration of Technical Data Conformity (Jan 1997), the Contractor, Francesco Di Nocera, hereby declares that, to the best of its knowledge and belief, the technical data delivered herewith under Contract No. FA8655-05-1-3021 is complete, accurate, and complies with all requirements of the contract.

In accordance with the requirements in Federal Acquisition Regulation 52.227-12, Patent Rights-Retention by the Contractor (Jan 1997), I certify that there were no subject inventions to declare as defined in FAR 52.227-12, during the performance of this contract.

DATE: Jan 1st 2007

Name and Title of Authorized Official: Francesco Di Nocera, Ph.D.

BACKGROUND

Automation is being introduced into various domains of work and everyday life. Automated subsystems now provide the human operator valuable support in such domains as air, ground, space, and maritime transportation, military command and control, health care, and other areas. These types of computer support can be considered to define different levels of automation (LOA) between the extremes of full manual and full automation control (Sheridan, 2002). Between these two extremes a variety of intermediate LOA can be identified, and each one could be conceptualized as a compromise between human and machine responsibilities.

A given system could be designed for a particular LOA on the basis of criteria such as system safety and efficiency, as well as human performance criteria such as the maintenance of situation awareness and balanced workload (Endsley & Kaber, 1999; Parasuraman, Sheridan, & Wickens, 2000). LOA may also be modified in real time during system operations, as in the so-called adaptive automation (Moray, Inagaki, & Itoh, 2000; Parasuraman, Molloy, & Singh, 1996; Scerbo, 1996; Scerbo et al., 2001). Indeed, Adaptive Automation (AA) can be defined as technology that can change dynamically its mode of operation, adjusting in real-time to the needs of the human operator. How such changes are accomplished may vary. For example, measures of human performance can be used to trigger automation, and operator models may be of use specifying in which conditions autonomous or semi-autonomous technology should take over. However, the use of physiological measures reflecting changes in the operator mental workload is considered one of the most promising methods (see Scerbo and colleagues 2001 for a review), since they provide real-time information on the state of the operator.

The idea of using psychophysiological measures in human factors (see Kramer & Weber 2000, for a recent review) is currently broadly accepted. Nonetheless, actual implementation of psychophysiology is limited to off-line contexts (i.e. assessment and training), and several factors, such as the high cost and the expertise needed to get them work, strongly discourage their use in real-world situations. However, new insights and motivations may come from research extending the use of psychophysiology for operator aiding and support. As introduced above, this emerging field is devoted to develop adaptive systems, which will be able to flexibly adapt to the operator needs. Real-world applications for adaptive technology are still uncommon, though existing. Systems able to detect changes in the operator alertness represent an example of such technology.

Alertness detection systems may be considered as the simplest existing form of adaptive technology. They provide binary decisions: either the operator is awake and performing or he/she is asleep. Some of these systems are already "up and running", but they do not represent the final answer in the domain of AA. In fact, detection and prediction of behavioral implications of variations in mental workload and attention are harder than assessing performance implications of lapses in alertness (Kramer & Weber, 2000).

The most important consideration here is that AA involves the assessment of graded changes in mental workload, which is more difficult and requires the use of different measures sensitive to different levels and types of processing demand.

Future technological changes will permit to overcome most of the technical limitations in this field. Current technology actually allows the use of psychophysiological indicators in simulated real-world tasks by means of "smart" flight helmets incorporating electrodes and preamplifiers for EEG and EOG recording (Gevins et al., 1995), 2D and 3D brain mapping (Heinonen, Lahtinen, & Häkkinen, 1999) for visualization of dynamic states induced by events, or even the use of neural network models as an EEG pattern recognition method to detect transient cognitive impairment (Gevins & Smith, 1999), whereas brain-computer interfaces (BCI) (see Wolpaw et al., 2002 for a recent review) provide evidence for communication between brain and technology using decoding algorithms. Nevertheless, many of these tools and procedures have some shortcomings. They do provide real time or near-real time information about the state of the operator, but they are not perfectly reliable, and a system whose accuracy is affected by unknown factors is simply unacceptable when safety is at risk. This is not only a matter of the amount of information we can get from the operator, and we are not going to work out this problem simply adding more indicators. For example, using an artificial neural network and a noteworthy collection of physiological data (EEG, ECG, EOG, and respiration inputs) recorded during task performance, Wilson and colleagues (Wilson, Lambert, & Russell, 2000) found mean correct OFS classifications across subjects ranging from 82% to 86%. This is a fairly successful result, but it is not enough for ensuring safety. Also, that was a laboratory study based on a simulated task (the NASA Multiple Attribute Task Battery), but moving from the off-line to the on-line context, additional issues have to be considered, such as rapid data collection, processing, artifact rejection, and interpretation (Kramer & Weber, 2000).

SHIFTING BETWEEN DIFFERENT LEVELS OF AUTOMATION

In adaptive systems, task allocation between the operator and the computer systems is flexible and context-dependent. Adaptive automation may reduce the human performance costs (unbalanced mental workload, reduced situation awareness, complacency, skill degradation, etc.) that are sometimes associated with high-level decision automation. Several investigators have looked at the effects of different Levels of Automation (LOA) on performance. According to Parasuraman et al. (2000) high LOA can be usefully implemented for information acquisition and analysis functions. Nevertheless, decision making functions are acknowledged to be best supported by moderate LOA. Studies by Crocoll & Coury (1990), Sarter & Schroeder (2001), and Rovira, McGarry, & Parasuraman (2002) support this view by showing that unreliable decision automation leads to greater costs than unreliable information automation.

Kaber, Onal, & Endsley (2000), Endsley & Kiris (1995), Endsley & Kaber (1999), and Kaber, Onal, & Endsley (1998) also provide support for a "moderate" LOA philosophy. The underlying rationale views moderate LOA as an optimal balance with respect to the performance trade-off resulting from the benefits of reduced workload associated with higher LOA on the one hand and with better maintenance of situation awareness associated with lower LOA on the other hand. These studies induced rare automation failure events that require operators to return to full manual control. Typically, the higher the LOA prior

to this event the poorer the return-to-manual performance or -in other words- the higher the out-of-the-loop performance cost. Lorenz, Di Nocera, Röttger, & Parasuraman (2002), however, have shown that a higher LOA in a complex fault-management task does not necessarily lead to poorer return-to-manual performance under automation failure in comparison to a moderate LOA, as long as the interface supports operator information sampling to maintain situation awareness. In fact, the moderate LOA was found to be linked to a higher disengagement of sampling fault-relevant information. Apparently this LOA directed the operator attention to lower-order manual implementation of fault recovery actions at the expense of monitoring the impact of these activities on higher-order system constraints. A mitigation of this effect could be achieved in a follow-up study that used an integrated display in support of fault state monitoring (Lorenz, Di Nocera & Parasuraman, 2004). According to these studies the LOA per se is not necessarily the crucial factor affecting the out-of-the-loop performance costs. In general, it appears that there are differential effects of LOA by stage of processing and interface type. Yet, the experimental procedure used in these studies involved LOA shifts in different blocks, making it difficult to generalize the effects found to the adaptive automation domain. Indeed, adaptive automation assumes changes in LOA within shorter time frames, e.g. even from trial to trial, and there is very little research on such dynamic shifts in LOA. Furthermore, it is unclear whether the direction of the shift (up or down the LOA continuum) affects performance. Di Nocera, Lorenz & Parasuraman (2005) carried out a study to verify whether distance and direction in LOA shifts can affect human performance when interacting with complex tasks. Results showed that specific costs were associated with the process of disengaging from one cognitive-behavioral set the operator was currently using to the engagement of another -more appropriate- set. Such effect was not only associated with variations in the difficulty of the task, but was also affected by the mental workload the operator was experiencing on the moment. Recent research results (Trafton et al., 2003) suggest that preparation may have an important role in resuming a task previously carried out, and one may wonder if "preparation lag" may have a role also in adjusting to the next level of automation.

REAL-TIME ASSESSMENT OF MENTAL WORKLOAD

Human Factors & Ergonomics (HF/E) research has abundantly demonstrated that extreme levels of mental workload increase the likelihood of human error, because they deteriorate human ability to adequately react to incoming information. Mental workload can be defined as the difference between the task demands on one hand and the operator's cognitive resources on the other (Gopher & Donchin, 1986; O'Donnell & Eggemeier, 1986). The nature of this construct is grounded in human physiology and can be related to a complex set of brain states mediating human performance in perceptual, cognitive and motor tasks (Parasuraman & Caggiano, 2002). Nearly all scholars currently agree that mental workload is a multi-dimensional construct (see Kramer, 1991) that would reflect the individual level of engagement and effort (Wickens & Hollands, 2000).

Notwithstanding the wide number of theoretical accounts on mental workload, the main objective of most studies is still its assessment. Indeed, mental workload cannot be measured directly, but it is rather estimated by measuring variables that are assumed to correlate with the operator's mental load:

- changes in performance due to the allocation of resources to multiple tasks;
- operator's self-reports (e.g., NASA-TLX, SWAT);
- changes in human physiology (e.g., heart rate variability, electrical brain activity) that are assumed to vary with mental load.

As stated above, psychophysiological indices of mental workload have been reported to be the most promising measures of mental workload, because they provide 1) information about covert processes, and 2) continuous information about the operator functional state (Hancock, Chignell, & Lowenthal, 1985; Morrison & Gluckman, 1994; Scerbo,1996; Byrne & Parasuraman, 1996) that may be eventually used to trigger adaptive systems.

Recently, the availability of less intrusive eye-tracking systems allowed researcher to effectively use indices of ocular activity as a measure of the operator mental workload (see Van Orden et al., 2001 for a recent account). For example, frequency and duration of eye-blinks have been found to be inversely correlated to mental load (Brookings, Wilson, & Swain, 1996; Hankins & Wilson, 1998). Additionally, some studies (Bunecke, 1987; Ephrath et al., 1980) have shown that workload affects the duration of fixations, whereas others (Bellenkes, Wickens, & Kramer, 1997; Miller, 1973) recorded shorter and more frequent fixations in expert operators (all these studies were run on aircraft pilots).

It is worth noting that different tasks can generate different patterns, depending on the type of index employed. Some indexes can be sensitive to visual demands, but they can be as well insensitive to cognitive demands. Wilson, Fullenkamp, & Davis (1994) have shown how the durations of eye-blinks decreased in a visual tracking task (which generates a minimum mental workload) respect to a more cognitively engaging task (a flight simulation).

Studies on driving behavior (see Recarte & Nunes, 2000; 2003) have shown that pupil diameter is affected by mental workload and, more important as for the aim of the present report, that increases in mental workload are associated with an increase in the concentration of eye-movements. The analysis of visual patterns is a technique often used in Human Factors research. For example, Diez et al. (2001) have used this technique for gathering information about the scanning strategies of pilots interacting with a Boeing 747 simulator. They divided the display in Areas Of Interest (AOI), each one including a tool inspected by pilots during a simulated flight. Although the scanpath is usually used to get qualitative information, it can also be used in association with advanced computing techniques. Visual scanning randomness, or entropy, has been proposed as a measure of mental workload (Tole et al., 1983; Harris, Glover & Spady, 1986). In thermodynamics the concept of entropy is related to the quantity of disorder in a system (in this case, the disorder in visual exploration). The rationale underlying this approach is that the exploration pattern becomes more stereotyped (that is, less random) as the workload increases. On the contrary, a decrease in mental workload should increase the randomness of the pattern. Hilburn et al. (1997), corroborated this hypothesis in a series of experiments run on air traffic controllers.

However, this research line seems to have been abandoned and no further studies are reported in the literature. One of the aims of the present research activiy is to investigate the relation between scanpath and workload on the basis of slightly different considerations. First, albeit it is quite straightforward that high workload may produce fixations grouping (because the operator needs to focus on some specific feature of the interface/task) there is no evidence (except for the studies reported above) that random patterns should be associated with low workload. Low workload may be associated with regular patterns as well, indicating a regular check of the interface space. According to this hypothesis, indexes providing information about the dispersion of point patterns should indicate regularity in the case of low workload and grouping in the case of high workload. The following section will briefly summarize information about one of these methods: the Complete Spatial Randomness (CSR) testing procedure.

THE NEAREST NEIGHBOR INDEX

The measurement and description of pattern distribution was first addressed in reference to plant and animal populations. In forestry, for example, the positions of trees in a forest form a point pattern in the plane. Information about the distribution of such points has been found relevant for investigating phenomena like plant infections or growing patterns. In the beginning, the basic assumption was that individuals of most populations (being them plants, animals, or fossils) were distributed at random, but it sooner became clear that the randomness assumption was not appropriate. The issue became then to establish the degree of variation from random expectation, as well as the significance of differences in the distribution of pattern of two or more populations. To this aim, Clark and Evans (1954) introduced the Nearest Neighbor Index (NNI), which is the ratio between 1) the average of the observed minimum distances between points and 2) the mean random distance that one would expect if the distribution were random. Fifty years later, this index is still one of the most used distance statistics in agriculture, paleontology, and analysis of crime (all of them deal with spatially arranged data).

As a first step, the nearest neighbor distance or d(NN) should be computed as follows:

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

where $\min(d_{ij})$ is the distance between each point and the point nearest to it, and N is the number of points in the distribution.

This index is nothing more than the average of the minimum distances. The second step is to compute the mean random distance or d(ran), that is the d(NN) one would expect if the distribution were random.

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

where A is the area of the region (the measurement unit of the index is related to the one used here), and N is the number of points.

The final step is the actual computation of the Nearest Neighbor Index as follows:

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

Of course, this ratio is equal to 1 when the distribution is random. Values lower than 1 suggest grouping, whereas values higher than 1 suggest regularity (i.e. the point pattern is dispersed in a non-random way). Theoretically, the NNI lies between 0 (maximum clustering) and 2.1491 (strictly regular hexagonal pattern).

Di Nocera, Terenzi, and Camilli (2006) applied this procedure to eye fixations (given that they are point patterns as well) and found this index to be sensitive to variation in mental workload, showing a tendency toward randomness in the high workload condition. This is the opposite of what the entropy-based method would predict. However, entropy studies have used ocular data within *specific* and *static* AOI, whereas that study used ocular data gathered from a *dynamic* scene (participants were requested to play the Asteroids PC game in two difficulty conditions) within a Convex Hull defined by the outermost fixations in the distribution. The high mental workload condition was obtained by preventing the use of the weapon to destroy the asteroids, whereas the low/moderate workload condition consisted of the regular game allowing the use of the weapon. Considering the dynamic nature of the Asteroids game (the ship moves around in the screen area), it is possible that the different distributions of fixations that have been found are strategy-driven rather than workload-driven. Indeed, even if the two versions of the game were geometrically equivalent (same exact number of asteroids either between conditions and throughout the game), avoiding the asteroids might favor a strategy aimed at spreading the fixations over a wide area, whereas the shooting condition might have been supported by a strategy based on focusing over the ship and target positions. In order to address the role of these differences, Di Nocera, Camilli, and Terenzi (in press) applied the same rationale to investigate ocular behavior during interaction with a somewhat "static" visual scene. To this aim a flight simulation task, comprising both high workload (Departure and Landing) and low-moderate workload (Climb, Cruise, and Descent) phases, was used. Of course, this was "static" in the sense that in a flight deck the locations of objects to monitor (namely, the instruments) did not change over time, even if the visual scene outside the cockpit changes. Albeit this study also showed the usefulness of the NNI as a workload measure, its validity was not specifically assessed.

The research activity reported in the present document was aimed at assessing the validity of this proposed index, and its sensitivity to changes in the level of automation. The following sections will describe:

1. the development of a software application for analyzing eye movements and computing the index;
2. the pilot study aimed at defining the taskload conditions to be used in the experimentation;

3. the first experiment aimed at assessing the concurrent validity of the measure;

4. the second experiment aimed at a) using the proposed index as a measure of workload in a LOA shifting paradigm, and b) studying the role of the time spent dealing with one specific LOA in adjusting to the next level of automation.

RESEARCH ACTIVITY: PHASE 1

DEVELOPMENT OF A SOFTWARE APPLICATION FOR ANALYZING THE SPATIAL DISTRIBUTION OF EYE MOVEMENTS

Eye-trackers manufacturers always provide software applications for playing back and analyzing the eye-movement data that are recorded by the system. Most of these applications provide several interesting features, sometimes much more than those required by the investigators using them. Indeed, the great deal of functionalities makes these applications resource-consuming and way too complicated for rapid and easy manipulation of coordinate data. With that in mind, we have developed A Simple Tool for Examining Fixations (ASTEF), whose primary function is to deal with point patterns. ASTEF was coded using C# and runs on Microsoft® Windows machines.

Defining Areas of Interest in ASTEF. Inspection of the scanpath is one of the primary tasks accomplished by ASTEF. This is a common task for many researchers that need to examine the sequence of fixations one by one in order to identify Areas Of Interest (AOI).

ASTEF implements area selection in three different ways:

1. by dragging the diagonal of a rectangle (during this procedure the area size and the mouse pointer coordinates are always visible in the status bar);
2. by moving the four sides of the rectangle separately, dragging the four corresponding cursors by mouse;
3. by clicking on the "Manual Selection" icon and inserting the exact coordinates.

ASTEF also provides the possibility to invert the selection. This may be useful in order to operate on the points outside an AOI (e.g. delete all the points outside the AOI).

All the selected AOIs can be named and saved for further use from the "AOIs" menu.

Fixation Identification Tool. ASTEF also provides a tool for identifying fixations from a raw file of gaze coordinates. In order to obtain fixations, the user is required to set two parameters: Min Fixation (in milliseconds), which is the minimum duration of the fixation, and Radius (in pixels), which is the minimum fixation radius. The latter is nothing more than the projection on the screen of the "threshold" visual angle. Default values are those frequently reported in the literature (Salvucci & Goldberg, 2000; Hornof & Halverson, 2002; Jacob & Karn, 2002; Jainta et al., 2002; Kramer & McCarley, 2003): ½° - 1° of visual angle and 100-200 ms of duration. For a 4:3 - 17" display having a 1024 x 768 resolution, the projection of 1° visual angle, at an approximate distance of 50 cm, is equivalent to a 25px radius.

Noise Filter. Sporadic points falling outside the fixations may be found during the identification process (Alpern, 1962; Ditchburn, 1980; Hornof & Halverson 2002). Sometimes, after a first outsider point, several other points may fall into the identified fixation. Ignoring those points may cause a biased

estimate; for this reason, it has been implemented a noise filter that checks for the timing of those points occurring after the outsider point (see figure 1).

Validity codes. Generally, the quality of a recorded gaze is affected by many factors like wearing glasses and contact lens, as well as by head movements. Most eye-tracking software suites provide "validity" or "confidence" codes for the recorded gazes, which are informative about the quality of a sample.

ASTEF also implements two columns in its data files that refer to the sampling quality. In order to work properly, the codes need to be consistent with those used by the Tobii's ClearView®. That software suite uses a 0-4 validity range, where "0" represents the best tracking quality. However, in processing the data file, ASTEF implements a rigid sample selection taking into consideration only those gazes having the maximum tracking validity ("0" coded). Such strictness is due to the fact that lower validity means lack of information about some features of the gaze (e.g. either the left or right eye coordinates are missing), and it is our opinion that it is much more appropriate to exclude those samples.

Spatial statistics with ASTEF. Some applications already exist for computing the NNI and other spatial statistics indices. CrimeStat (Levine, 2004), for example, is one of such applications committed to the spatial analysis of criminal acts. Also an increasing number of R packages, such as "spatstat" (Baddeley & Turner, 2005), are available. These packages allow the researcher to have full control over the analysis s/he runs. Nevertheless, R is intended for the advanced user and, despite its utility, this may discourage many researchers who are not familiar with it.

Usable spreadsheet-based software also exists. Prior to develop ASTEF, we have executed NNI computation using Paleontological Statistics (PAST: Hammer, Harper, & Ryan, 2001), which is a software application including many functions that are specific to Paleontology and Ecology, including NNI. Although this represented a viable solution, it also prevented the use of a tool that is specifically committed to the analysis of eye movement data. PAST does not provide visualization tools, and performing simple tasks -such as computing mean fixation duration- might be tricky.

ASTEF allows the use of two different areas for computing the Nearest Neighbor Index: Convex Hull and Smallest Rectangle. The first is derived by the Delaunay's algorithm (Delaunay, 1934), which creates a temporary hull from the first 3 points, and then adds other triangles for each outer point. The second is based on an algorithm that creates a bounding box for defining the rectangle having the smallest area comprising all the examined points. For the convex hull, ASTEF also implements the Donnelly's edge effect adjustment method (Donnelly, 1978).

All the analyses functions can be accessed from the "Analyze" menu appearing by right-clicking on the screen selection, as well as from the main menu.

Integration with commercial eye-tracking systems. Even if ASTEF works with any ASCII file properly formatted, in the future it could import files created with any commercial eye-tracking system. Conversion algorithms will be implemented according to users' needs, and the availability of proprietary file-structure information. The current version of ASTEF only imports Combined-Data-File (CMD)

created with Tobii's ClearView®, since this is the system we have used for the studies hereby reported. The import function has been tested with ClearView® v. 2.5.1. The import function can be accessed from the "Tools" menu.
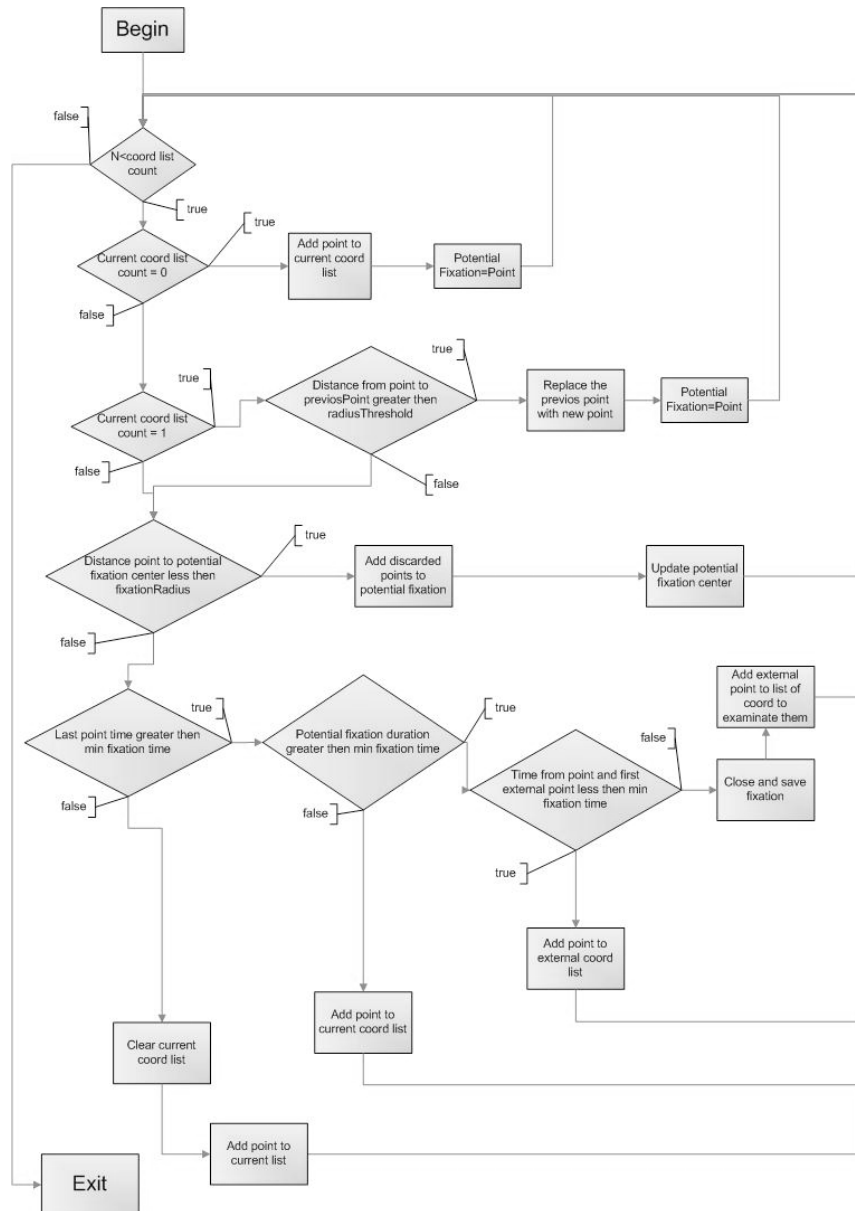
Figure 1 - Noise filter for the fixation identification tool implemented in ASTEF (pseudocode).

PILOT STUDY

The aim of this pilot study was to select three among ten levels of difficulty of a visuo-motor task (the Tetris game). These three levels should clearly generate "high", "intermediate" and "low" workload levels to be implemented as taskload conditions (hereinafter reported as "hard", "medium", "easy") in the successive experiments. This is a necessary step in order to assess the validity of the proposed index.

METHOD

Subjects. Twenty participants (10 females; mean age = 23 years, st. dev. = 2.41) volunteered in this study. All participants were right-handed, with normal hearing and normal or correct to normal vision.

Apparatus. The Tetris game used in this study was coded using C# and the .Net standard libraries (GDI+). The game area consisted of 300 cells deployed on 15 rows. Each block was randomly extracted from a pool of 7 different block types and descended at a constant speed. In order to generate the ten levels of difficulty, the speed was varied from 600 ms per cell (level 1) to 60 ms per cell (level 10).

Procedure. Participants received training prior experimentation and were included in the sample only when they became able to play for 10 minutes without filling completely the game area. Participants sat in dark and sound-attenuated room and were asked to play the game gaining as many points as possible. Order of presentation of the levels of difficulty was randomized across participants. After each block participants compiled the NASA-Task Load indeX (NASA-TLX: Hart & Staveland, 1988) for the subjective assessment of mental workload.

DATA ANALYSIS AND RESULTS

NASA-TLX weighted scores and number of completed lines (an index of performance in the Tetris game) were analyzed by ANOVA designs using the level of difficulty as independent variable. Results showed a main effect of the level of difficulty in both cases ($F_{9,171}=56.56$ p<.0001 and $F_{9,171}=37.01$ p<.0001, respectively). According to Duncan post-hoc testing and the comparison between the two measures, conditions 6, 7 and 8 (showing significant differences between them) were selected.
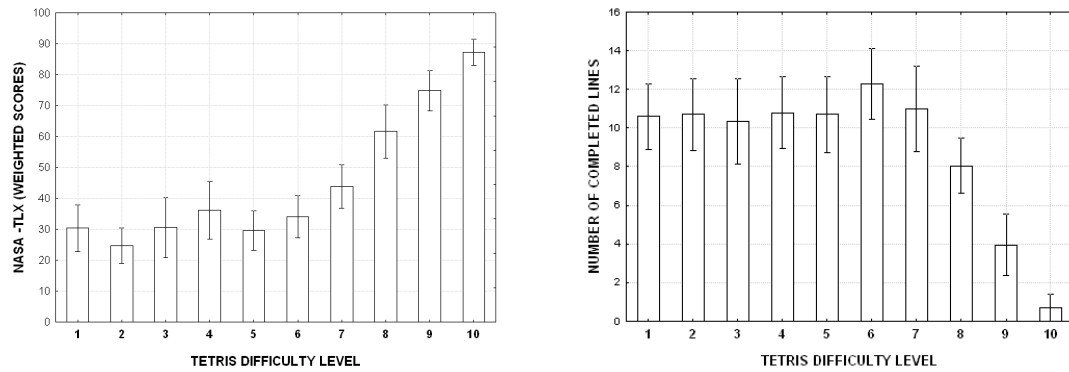
Figure 2 - NASA TLX weighted scores and number of completed lines separately for level of difficulty. Spreads denote .95 confidence intervals.

DISCUSSION

The Tetris game is a common visuo-motor task that has been successfully used for generating mental load in the scientific literature (e.g. Trimmel & Huber, 1998). The game is also well-known, and little participants training is necessary to use it for experimental purposes. One of the primary concerns in this research activity was to define taskload conditions that could clearly generate different amount of mental load. To this aim, ten levels of difficulty were used in the pilot study reported above, and twenty participants were requested to play those levels (randomly presented). Results showed an increment of subjective workload and a performance decrement starting from levels 6 to 10. Levels from 1 to 6 were not significantly different both in terms of gaming performance and subjective workload. Additionally, levels 9 and 10 showed poor performance making them not suitable for the following experiments. Condtions 6, 7, and 8 were instead significantly different and were retained as the "easy", "medium", and "hard" taskload conditions.

EXPERIMENT 1

METHOD

Subjects. Ten participants (5 females; mean age = 23.6 years, st. dev. = 2.01) volunteered in this study. All participants were right-handed, with normal hearing and normal or correct to normal vision.

Apparatus. Three levels of difficulty of the Tetris game (easy, medium, hard), selected according to the results of the pilot study, were used for generating different amounts of mental workload. An odd-ball task was used as secondary task. Three-hundred tones (65 db Spl, 100 ms), were presented through headphones. Seventy-five percent of the tones were 850 Hz (standards) and the remaining 25% were 1100 Hz (targets). These tones were presented randomly intermixed at a variable rate (ISI ranging from 1000 to 1500 ms).

Procedure. After the electrode cap application, participants sat in a sound-attenuated room and were asked to remain relaxed during the recording session. Their task was to play the game gaining as many points as possible, to ignore the standard tones and to count target tones. Order of presentation of the levels of difficulty was randomized across participants. After completing each level of difficulty, participants were requested to rate the amount of mental workload experienced using the NASA-TLX.

EEG Recordings. The EBNeuro Mizar 33 System (for physiological data acquisition and analysis) was used for recording the EEG sampled at 128 Hz for 1 s starting 100 ms prior to each stimulus onset and averaged off-line for target and standard tones separately. Trials judged on a visual inspection as contaminated by artifacts were excluded from the averaging. P300 amplitudes were measured individually for each participant's data as the difference between N2 and P3 (peak-to-peak amplitude).

Ocular activity recordings. The Tobii ET17 eye-tracking system was used for recording ocular activity. This systems allows the researcher to collect ocular data without using invasive and/or uncomfortable head-mounted instruments. Indeed, Tobii uses near infrared diodes to generate reflection patterns on the corneas of the eyes of the user. These reflection patterns, together with other visual information, are collected by a camera. Image processing algorithms identify relevant features, including the eyes and the corneal reflection patterns. Three-dimensional position in space of each eye-ball, and finally the gaze point on the screen are calculated. Sampling rate was approximately 33 Hz.

DATA ANALYSIS AND RESULTS

The NASA-TLX weighted scores were used as dependent variables in a repeated measures ANOVA design (easy vs. medium vs. hard). Results showed a significant difference between the levels of difficulty ($F_{2,18} = 4.83$, $p < .05$). Duncan post-hoc testing showed that the hard condition was significantly different from the other two ($p < .05$).
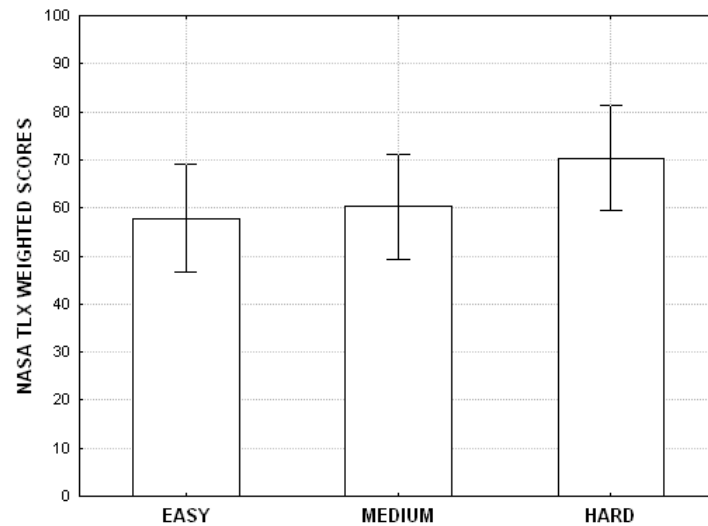

Figure 3 - NASA TLX values (weighted scores) separately for taskload condition. Spreads denote .95 confidence intervals.

Secondary task performance. Counting errors (deviation from the number of target trials as reported by subjects) were used as dependent variables in a repeated measure ANOVA using Taskload (Easy vs. Medium vs. Hard) as repeated factor. Results showed no significant effect of taskload ($p > .05$).

Nearest Neighbor Index. As suggested elsewhere (see Di Nocera et al., in press), the NNI was computed using ASTEF on blocks of 1 minute for each participant. This strategy is necessary because the index evolves over time. NNI fluctuations during time are shown in figure 4. Average NNI values for each subject were used as dependent variables in a repeated measures ANOVA design (easy vs. medium vs. hard). Results showed a main effect of Taskload ($F_{2,18} = 4.22$, $p < .05$). Duncan post-hoc testing showed that the hard condition was significantly different from the other two ($p < .05$).
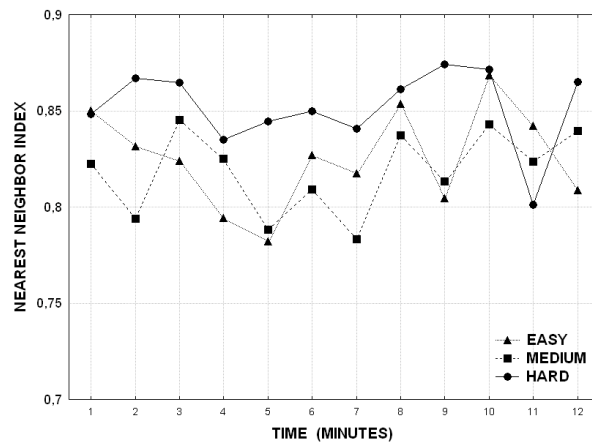
Figure 4 - Variation in time (minutes) of the NNI separately for taskload condition.
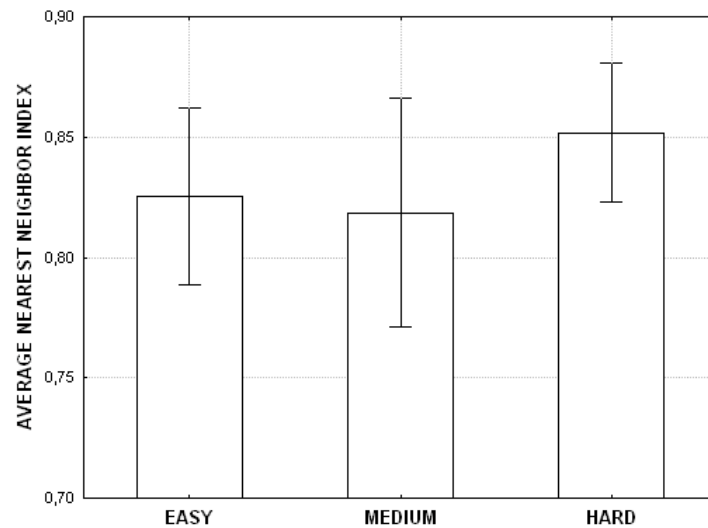
Figure 5 - Average NNI separately for taskload condition. Spreads denote .95 confidence intervals.

Event-Related Brain Potentials. The difference between N2-P3 amplitudes to standard and target stimuli were used as dependent variables in an ANOVA design Taskload (Easy vs. Medium vs. Hard) x Site (Fz vs. Cz vs. Pz). Results showed a main effect of Taskload ($F_{2,18}=4.40$, $p<.05$). Post-hoc Duncan testing showed that only the difference between the "hard" taskload condition and the other two was significant ($p<.05$).
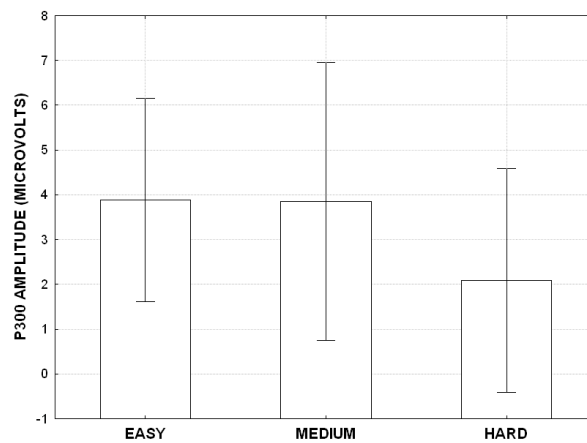
Figure 6 - P300 amplitudes by taskload. Only the difference between the "hard" taskload condition and the other two was significant. Spreads denote .95 confidence intervals.

DISCUSSION

This study represented a first attempt to assess the validity of the dispersion of eye fixations as a measure of mental workload. Two studies (Di Nocera et al., 2006; Di Nocera et al., in press) have reported the usefulness of the NNI, but its validity was not specifically assessed. The strategy adopted here was that of using multiple measures in order to estimate the concurrent validity of the index.

Consistent results were found across the three measures. The most difficult condition was found to generate significantly different values in the NASA-TLX ratings, in the P300 amplitude, and in the NNI values. However, all measures failed to show differences between the easiest and the intermediate taskload conditions. This might be due to two reasons. The first is the great variability affecting the data (mostly in the intermediate condition), which is probably due to the small sample size. Indeed, the three conditions were selected on the basis of the pilot study in which twenty participants have rated their subjective workload, whereas in this case only ten people participated in the experiment. The second reasons might be a lack of perceivable difference between the easy and medium conditions. In fact, in the pilot study ten levels of difficulty were administered, and participants might have been able to experience the entire spectrum of the imposed workload, generating fine-grained assessments. Contrarily, in the experiment reported above participants only experienced three levels of difficulty, making it difficult to generate accurate estimates. This effect is known as "context effect" and has been studies experimentally by Colle & Reid (1998) who demonstrated that subjective estimates of mental workload are biased when participants cannot experience the full range of task difficulty. This explanation is quite convincing for the subjective measure. However, also the ocular strategy and the brain activity showed the same pattern. Can context effect account for those measures too? This is difficult to demonstrate post-hoc. Nevertheless, one could consider the possibility that perceived workload affects the amount of resources actively allocated by the participants during the execution of the task. After all, the operator's perceptions

are always reported to be an important factor in the definition of mental workload as a multidimensional construct.
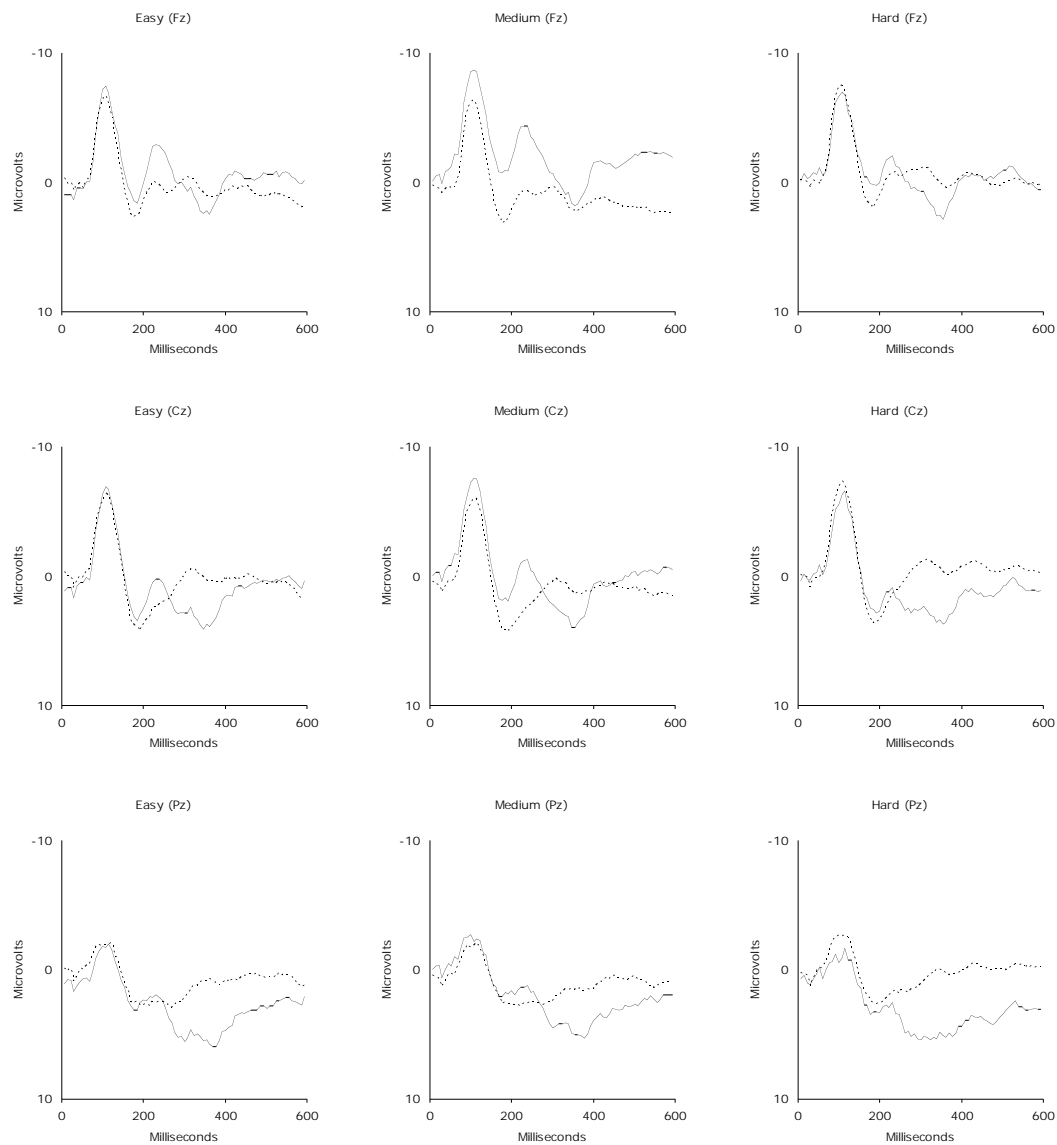
Figure 7 - Grand averages separately for electrode site (Fz, Cz, Pz), taskload condition (Easy. Medium, Hard), and type of stimuli (Standards = dashed; Targets = solid).

RESEARCH ACTIVITY: PHASE 2

Adaptive automation deals with the ability to flexibly adapt to changing situations, realize new intentions, and schedule intended actions, which are central features of human action control. Thus, any change in the level of automation (LOA) can be considered as a change in the task set. Indeed, a LOA is nothing more than a "level of constraint" either for the human or the machine: LOA constrains the space of (possible) action.

The utility of considering LOA as task sets is that these are usually considered as representations utilized to select an action despite the ambiguity of the external context (Mayr & Keele, 2000; Monsell, 1996; see also Rubinstein, Meyer, & Evans, 2001 and Schuch & Koch, 2003 for some accounts related to response selection making use of rules), and this is quite similar to the Norman and Shallice's (1986) perspective on schemata. Their theory is based on the same distinction between automatic and controlled processing which also characterizes other models and theories (see Schiffrin & Schneider, 1977; Schneider & Schiffrin, 1977). This perspective postulates the existence of a mechanism that uses internal representations (or schemata) to coordinate habitual behaviors. Once selected, a schema stays active until it reaches its goal, or it is inhibited by other schemata that are either competing for implementation or located higher in the hierarchy. Besides this type of process, our cognitive system would need also a mechanism to face novelty. Such a "supervisor system" may intervene to shut down the activity of a currently-better schema or to provide a higher level of activation. Hence, different LOA could trigger different schemata, which, in turn, may represent "calls" to specific cognitive processes involved in the remaining tasks that are actively carried out by the individuals. One of the outcomes of this mechanism may be the disengagement of the other functions and processes, which are not easily reacquired when shifting to another LOA. Shifting tasks causes costs that are assumed to reflect configuration processes (Rogers & Monsell, 1995). In the automation domain, shifts are known to affect in several ways human performance. For example, it is well known that the ability to detect automation failures deteriorates under automatic as opposed to manual operating conditions (Parasuraman, Molloy, & Singh, 1993; Parasuraman, Mouloua, & Molloy, 1996), and consequences of such inability may be devastating when error compensation is difficult or even impossible. Therefore, the investigation of the processes underlying LOA shifts, and their consequences on operator's performance seems to be critical for good automation design. This specific aspect has been recently approached by Di Nocera, Lorenz and Parasuraman (2005). However, more studies are needed for gathering a full understanding of these phenomena.

In the following study we have approached this issue investigating automatin shifts and their relation with mental workload.

EXPERIMENT 2

The aim of this second study was to investigate the cost deriving from shifting between levels of automation, employing the same task and measures used in the previous phase. Particularly, we have investigated the effects of the time spent dealing with one specific LOA in adjusting to the next level of automation. Additionally, a side objective of the present experiment was to verify the sensitivity of the NNI to variations in the level of automation. This study also involved a higher number of participants (N = 20), and its results may clarify the role played by sample size in the absence of significant differences between the "easy" and the "medium" difficulty conditions we have reported.

METHOD

Subjects. Twenty participants (8 females; mean age = 21.85 years, st. dev. = 1.04) volunteered in this study. All participants were right-handed, with normal hearing and normal or correct to normal vision.

Apparatus. The previously described three levels of difficulty of the Tetris game were used for generating different amounts of mental workload. Two versions of the game were implemented: automated (see figure 8) and manual. The automated version provided the participants a projection of the falling block for making it easier it. The same odd-ball task used previously was used as secondary task.
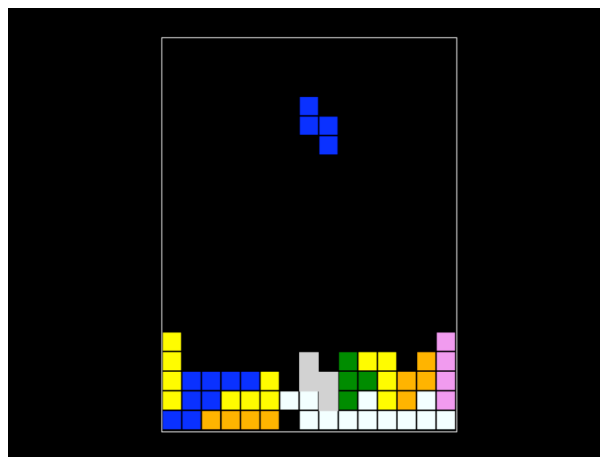
Figure 8 – Automation support was provided by showing a "ghost" block (in grey), a projection of the block that is falling down.

Procedure. After the electrode cap application, participants sat in a dark and sound-attenuated room and were asked to play the game gaining as many points as possible, to ignore the standard tones and to count the target tones. During this task, switches from manual to automatic (and *vice versa*) happened. Two different "LOA permanence" conditions were also implemented: 1-minute (short-term permanence in

LOA) and 3-minute (long-term permanence in LOA). The automation sequence gave rise to three different conditions: forward shift (from manual to automatic), backward shift (from automatic to manual), no shift (a sequence of two identical trials). The following table shows the details of the sequences employed in this study (t = 1 minute).

	t-3	t-2	t-1	t (shift)
Forward (long-term)	Manual control	Manual control	Manual control	Automation support
Forward (short-term)			Manual control	Automation support
Backward (long-term)	Automation support	Automation support	Automation support	Manual control
Backward (short-term)			Automation support	Manual control
Neutral (long-term)	Manual control	Manual control	Manual control	Manual control
Neutral (short-term)			Manual control	Manual control
Neutral (long-term)	Automation support	Automation support	Automation support	Automation support
Neutral (short-term)			Automation support	Automation support

Table 1 - LOA swichting and permanence in LOA.

EEG and Ocular activity recordings. The same instruments and measures employed in the previous phase were used in this case.

**DATA ANALYSIS AND RESULTS**

NASA-TLX weighted scores were used as dependent variables in a repeated measures ANOVA design using Taskload (Easy vs. Medium vs. Hard) as repeated factor. Results showed a main effect of Taskload ($F_{2,38}$ = 18.66, p<.0001). Duncan post-hoc testing showed that the hard condition was significantly different from the other two (p<.01).
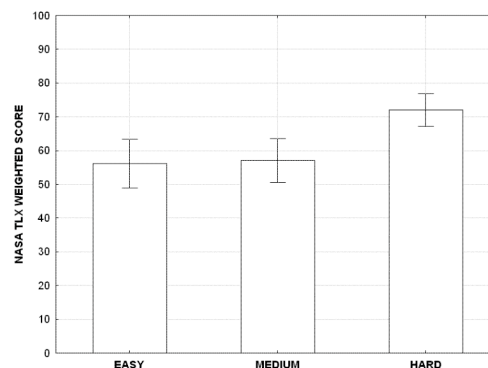

Figure 9 - NASA-TLX scores by Taskload.

Secondary task performance. Counting errors (deviation from the number of target trials as reported by subjects) were used as dependent variables in a repeated measure ANOVA using Taskload (Easy vs. Medium vs. Hard) as repeated factor. Results showed no significant effect of taskload (p>.05).

Event-Related Brain Potentials. The difference between N2-P3 amplitudes to standard and target stimuli were used as dependent variables in an ANOVA design Taskload (Easy vs. Medium vs. Hard) x Site (Fz vs. Cz vs. Pz). Results showed a main effect of Taskload ($F_{2,38}$=3.39, p<.05) and a main effect of electrode site ($F_{2,38}$=6.09, p<.01) due to a larger P300 amplitude in Cz and Pz.
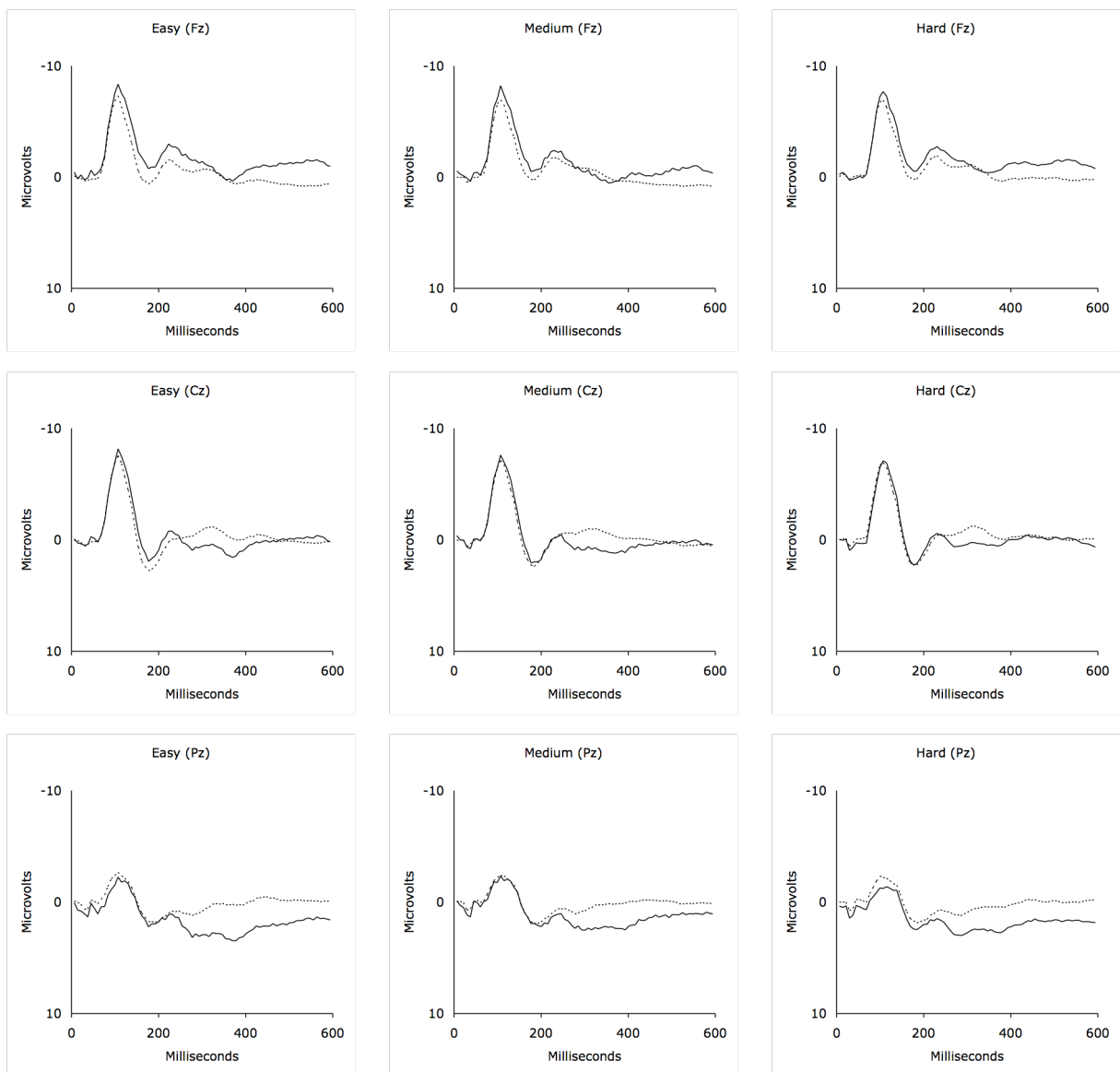

Figure 9 - Grand averages separately for Electrode Site (Fz, Cz, Pz), Taskload (Easy. Medium, Hard), and Stimuli (Standards = dashed; Targets = solid).
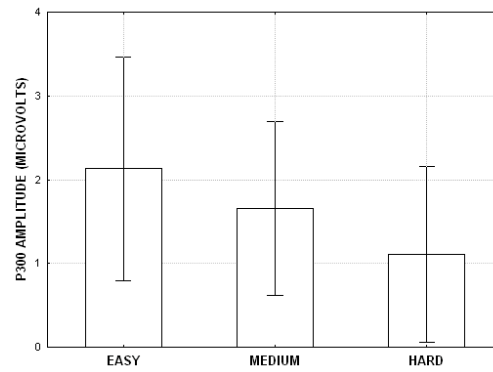
Figure 10 - P300 amplitudes by Taskload.

Performance data. The proportion of completed lines (a row of blocks eliminated during the Tetris game) was used as an index of performance. Differences respect to baseline (the neutral condition) were used as dependent variables in an ANOVA design Taskload (Easy vs. Medium vs. Hard) x Permanence (Long-term vs. Short-term) x Direction (Forward shift vs. Backward shift). Results showed a significant Taskload by Direction interaction ($F_{2,38}$=3.31, p<.05) and a significant Permanence by Direction interaction ($F_{1,19}$=5.46, p<.05).
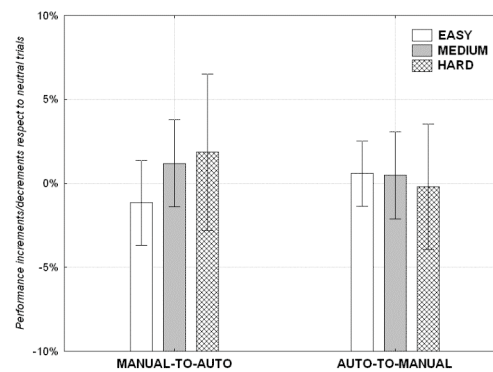

**Figure 11 -** Performance respect to neutral trials by Taskload.
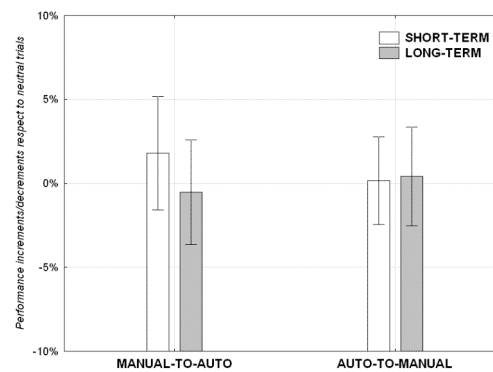
Figure 12 - Performance respect to neutral trials by Permanence in LOA.

Nearest Neighbor Index. The NNI was computed using ASTEF on blocks of 1 minute for each participant. Average NNI values for each subject were used as used as dependent variables in a repeated measures ANOVA design (Easy vs. Medium vs. Hard). Results showed a significant difference between the levels of difficulty ($F_{2,38} = 11.51$, p<.001). Duncan post-hoc testing showed that the hard condition was significantly different from the other two (p<.01).
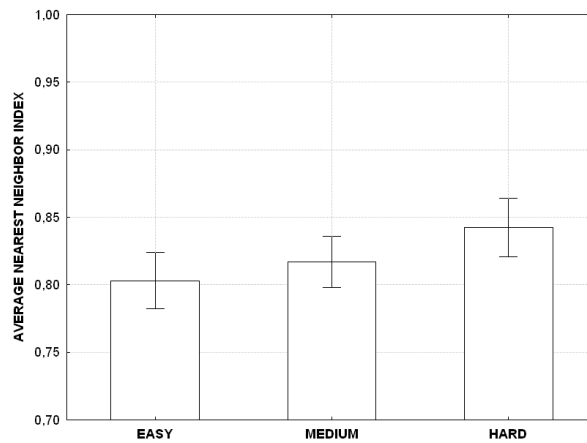
Figure 13 - Average NNI by Taskload.

NNI values recorded in the last two trials (t-1 and t) of each condition were also used as dependent variables in an ANOVA design Taskload (Easy vs. Medium vs. Hard) x Permanence (Long vs. Short) x Direction (Forward vs. Backward vs. Neutral) x Trial (t-1 vs. t). Results showed a significant Taskload by Direction interaction ($F_{4,76}$=4.21, p<.01). Duncan testing showed that backward shifts generated a significant higher mental workload only in the hard condition (figure 14).
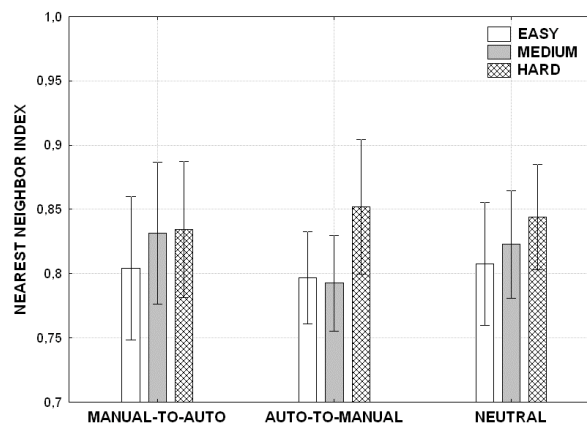
Figure 14 - Average NNI by Taskload and Direction.

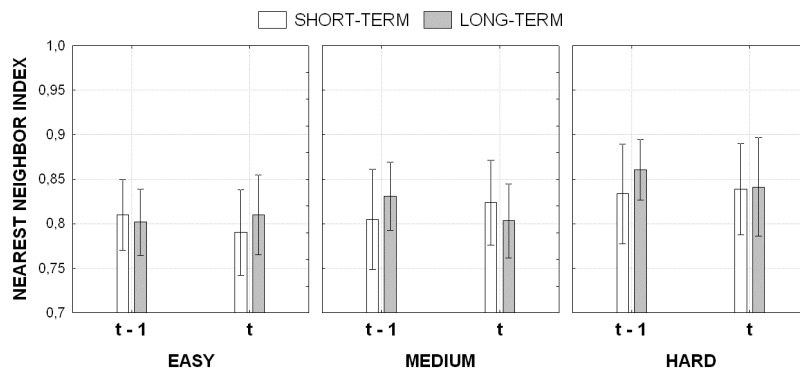A significant Permanence by Direction interaction was also found ($F_{2,38}$=3.59, p<.05). As shown by figure 15, this interaction is due to a reduced NNI value in the short-term permanence condition when the shift is backward.
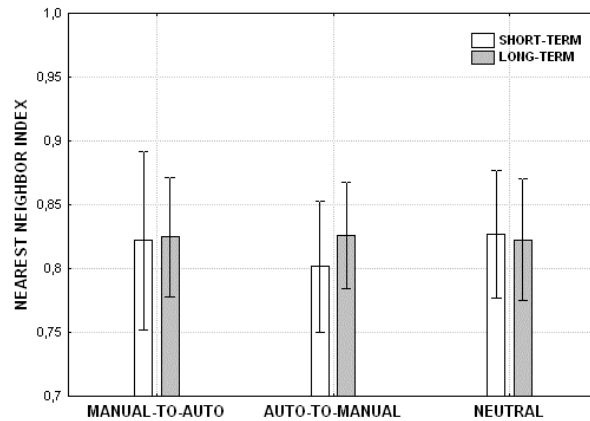
Figure 15 - Average NNI by Taskload and Permanence.

Permanence was also found to interact significatively with the Taskload and Trial factors ($F_{2,38}$=4.67, p<.05). As shown by figure 16, taskload and permanence affect the change in NNI values during the shift differently.

Figure 16 - Average NNI by Taskload, Permanence and Trial.

Furthermore, results showed a significant Direction by Trial interaction ($F_{2,38}$=24.45, p<.0001). Neutral trials showed approximately the same NNI value at t-1 and t, backward trials NNI values at time t were higher than those at time t-1, whereas NNI values in the forward shift were lower at time t than at time t-1. Post-hoc testing showed that only the forward shift was significant (p<.05), while the backward shift only showed a tendency towards statistical significance (p=.13).
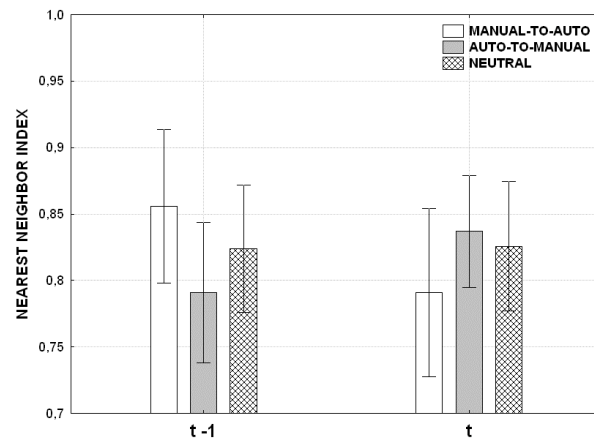
Figure 17 - Average NNI by Direction and Trial.

Direction and Trial were also found to interact with Taskload. However, this interaction only showed a tendency towards statistical significance (p=.07).
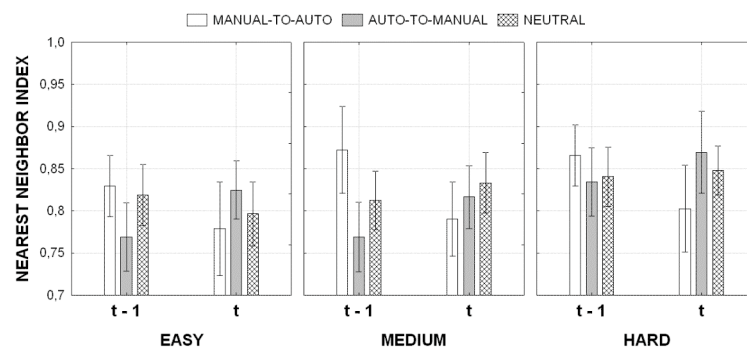
Figure 18 - Average NNI by Taskload, Direction and Trial.

Figure 19 – NNI changes in time for the Easy condition.

Figure 20 – NNI changes in time for the Medium condition.

Figure 20 – NNI changes in time for the Hard condition.

DISCUSSION
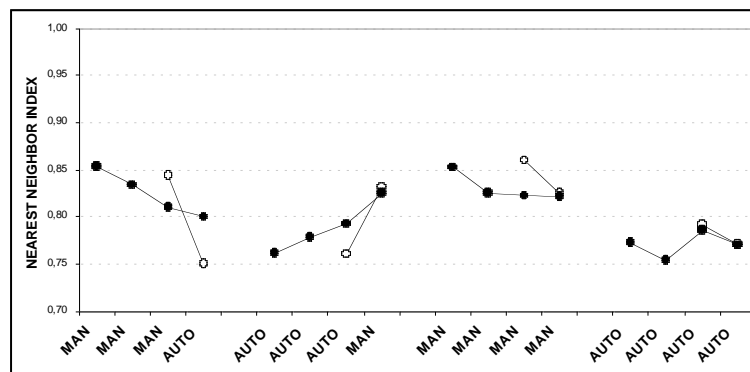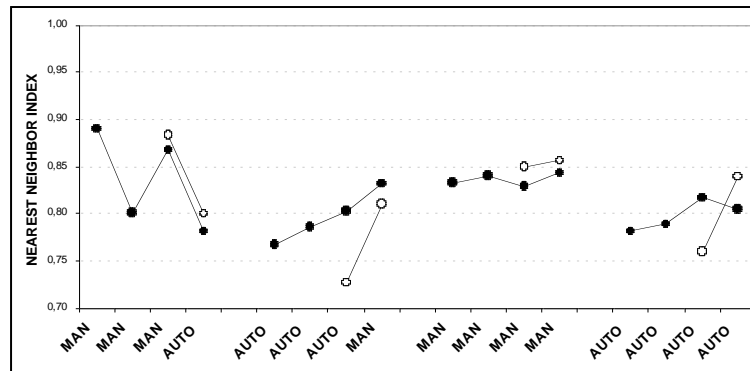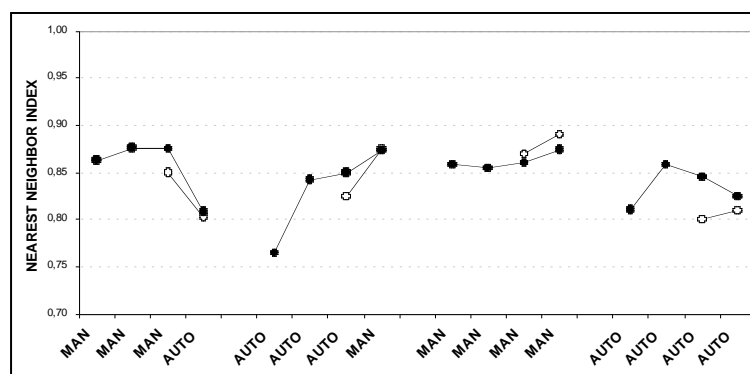
Results of this second study confirmed and extended those of study 1 by showing sensistivity of the NNI to variations in taskload, as well as the absence of differences between the easiest and the intermediate taskload conditions. Moreover, ERPs showed significant differences between the easiest and the hardest taskload conditions, suggesting that the intermediate condition might be the problematic one. This is also supported by the great variability affecting the data in this very condition (also in study 1).

However the main aim of this second study was to investigate the effects on performance and workload of the shifiting between levels of automation: from manual to automatic and from automatic to manual. It was expected to find switching costs in both directions, not only in the backward shift. This prediction was based on the general idea that in both cases there is the engagement / disengagement of cognitive processes (mental rotation, in this case). Results seem to support this view, but the effect of LOA-switching seems to be also modulated by taskload. Indeed, a better performance was associated with a forward shift in the hard taskload condition, whereas the same forward shift was detrimental in the easiest taskload condition.

Another aim of this study was to investigate the effects of the time spent in a LOA (permanence). It was expected that the longer an individual interacted with a task at a particular LOA the most difficult would

have been to switch to another LOA. Results showed that the advantage of the forward shift (which makes it generally easier the task) was eliminated by the long-term permanence in a LOA. Switching direction and permanence also affected workload. NNI showed sensitivity to variations in the type of shift, and differential patterns were found as a function of taskload and type of shift. Particularly, the three taskload conditions generated progressively higher NNI values only in the neutral sequences, whereas the switching conditions generated differential patterns. The NNI was also affected by permanence, and showed differential patterns as a function of taskload. NNI values were lowered by short-term permanence and increased by long-term permanence in the easiest condition, whereas they showed the opposite pattern in the intermediate taskload condition. The effect is not much clear in the hardest taskload condition. Nevertheless, we found a reduction of the NNI values associated with the long-term permanence. This results should be taken carefully, because the variability affecting the data does not allow to clearly isolate these effects.

Moreover, it is worth noting that -albeit the experiment reported here was aimed at studying LOA shifts effects for future development of adaptive automation- a main difference exist between our experimental setup and actual adaptive systems. Indeed, in adaptive systems LOA shifts would happen according to some modification in human physiology and/or behavior, whereas in this study they have been programmed by the experimenters. That should be taken into consideration for interpreting the results. Nevertheless, the outcome of this study may be of interest for understanding what type of response we may expect from operators when LOA shifts are inconsistent or partially unrelated to the operator functional state.

GENERAL DISCUSSION AND CONCLUSIONS

On the real time assessment of mental workload. One of the most important issues for effective implementation of AA is the choice of the index to use for triggering the system when the functional state of the operator significantly deviates from optimal levels. Several indexes have been discussed in the literature (see Di Nocera et al., 2003 for an up-to-date discussion on this topic). Most of them are psychophysiological, representing the physiological response to events mediated by the cognitive system, and it is commonly believed they represent the most valid and reliable method of obtaining real-time information about the state of the operator (Boucsein & Backs, 2000; Scerbo et al., 2001). Behavioral, subjective and physiological measures can, however, be used according to particular needs. Nevertheless, great care should be taken when selecting the index to use. Sensitivity of a parameter may be affected by different factors such as the number of samples used to compute it. One may find, for example, that the selected measure can work well in one task environment, or indeed the laboratory, but not in another. For example, in a comparative study on different techniques for evaluating psychomotor load, Wierwille and Connor (1983) showed that sensitivity of measures might vary widely, strongly affecting the workload assessment.

Among the psychophysiological indices of mental workload, the ocular activity has recently received considerable attention, even if its use can be traced back to Fitts' work (Fitts, Jones, & Milton,1950). This recent interest towards eye movements is primarily due to the advancements in the technology for recording them (namely, eye-trackers), which is becoming increasingly usable and affordable. Moreover, eye-tracking systems that do not need to be head-mounted (infrared based) open new possibilities for eye movements recording in ecological settings.

Compared to other psychophysiological indices (e.g. Event-Related Brain Potentials, Heart Rate Variability) that have been proposed as candidate measures for triggering adaptive systems, eye movements show many benefits: they are insensitive to limbs movements (they can also be adjusted for head movements), no much training is necessary for setting up the equipment (at least the infrared-based system used in the present research activity), and the calibration procedure can be accomplished in a short time. Results of the studies presented here have confirmed that the NNI computed on eye fixations is sensitive to variations in mental workload, thus replicating previous findings and providing additional support to the robustness of this index. Moreover, specific workload-related fixations patterns were found using eye-movement data collected during the execution of a visuo-motor task along with subjective reports and brain activity. As expected, higher NNI values were associated to high workload conditions. The lack of significant results in some of the post-hoc comparisons should not be considered as indicating lack of sensitivity of the measure, because that was presumably due to the taskload condition that have been selected, and the variability affecting the data. Indeed, the easy and medium conditions showed very high variability. Also, both the subjective and ERPs data showed the same pattern, thus suggesting that those two conditions might have been too close in terms of resources request. Likely, falling speed of the blocks is not the best manipulation in order to obtain clearly different taskload conditions. Future studies may take into consideration other aspects such as type of blocks, color combinations, and the like.

Overall, the evidences provided here allow suggesting the implementation of this index as a real-time measure of mental workload, hence as a trigger for automated systems. One additional benefit of the proposed index is that it does not necessarily need extreme precision and high temporal resolution: in fact, the comparison is made between the actual distribution of points and the expected random distribution of the same number of points. The index itself is a rough estimate of grouping and having more points (i.e. more than 100-200), in our experience, does not enrich its meaningfulness.

As found by Di Nocera et al. (2006; in press), also in this case the direction of the NNI pattern was found to diverge from that expected on the basis of the entropy studies run by Harris, Glover and Spady (1986), Hilburn et al. (1997), and Tole et al. (1983). The functional significance of the index can be summarized as follows: under high mental workload conditions, more dispersed patterns may be due to a strategy aimed at optimizing promptness to incoming information. Indeed, as hypothesized by Smith, Valentino and Arruda (2003), endogenous mechanisms that cause organisms to automatically alternate their attention between focusing and casting a wide net may have evolved. This is also compatible with a finding reported by Pelz and Canosa (2001) that individuals might use "look-ahead" fixations serving future tasks. The cyclical pattern shown in figure 4 (and replicating that reported by Di Nocera et al., in press) seems to support this view. At this stage of development of the research it is impossible to address the basic mechanisms involved in the generation of this effect. However, the results provided in this report may indicate a fluctuation of attentional resources. From a logical standpoint, we could think of three possible strategies for resources allocation: 1) on-demand, 2) continuous, and 3) cyclical. The first would be a strategy based on minimizing the resources expenditure when they are not needed, and allocating them only when required. That, of course, strongly reduces the promptness of the individual to react. The second strategy would involve a continuous expenditure of attentional resources in order to put the individual always in condition to react properly. This is clearly impossible, considering that mental resources are limited in nature. The third, instead, considers the possibility of a cyclical allocation of the attentional resources (a parsimonious strategy), so that a certain degree of promptness is always (cyclically) available to the individual. Such a cyclical "rise and fall" would then allow the individual to take advantage of the level of mental resources made available, and to use that level as a starting point for voluntary resources management.

In other words, when task demands are high, it becomes mandatory to monitor everything in the shortest time, without "wasting time" (and fixations) on the same parts of the interface. Fixations, after all, are "pauses over informative regions of interest" (Salvucci & Goldberg, 2000 p. 71). Similar considerations have been made about the course of ocular inspection of pictures: fixations are usually shorter when we start viewing a picture (high workload condition, so to say). This phenomenon has been also reported by Kahneman (1973), who found it puzzling, because that is exactly the phase when we need to gather more information, and fixations are supposed to last longer on an object. However, the type of information we need in the initial phases (or in the most difficult phases) may do the difference. Indeed, "structural" more than "semantic" information may be extracted, and that can be accomplished with few (and short) fixations. Also this account is compatible with recent findings. Irwin and Zelinsky (2002) reported a continuous increase in fixation duration during the inspection time (over a 15-fixation long period), and

Unema et al. (2005) found a shift (that is function of the inspection time) from shorter fixations and saccades with longer amplitudes to longer fixations and shorter saccades. The authors interpret this effect in terms of two different spatial representations underlying the early and late phases in picture viewing. The focus of that work is on the "what" and "where" systems and their relation to the "ventral" and "dorsal" visual pathways (Ungerleider & Mishkin, 1982). An extensive discussion of this topic is outside the aim of the present report. However, for sake of completeness, Unema et al. (2005) report that the transition from short-fixations/large-saccades to long-fixations/short-saccades may suggest that "two qualitatively different competitive processes negotiate whether to keep fixating or to go on to the next salient object" (p. 491).

In conclusion, the application of the Nearest Neighbor Index to eye fixation data provides a domain-independent measure that could be eventually used in operational environments for gathering real-time information on operator load. This is of critical interest in several domains from Air Traffic Control to baggage screening.

On the costs of switching between levels of automation. Adaptive automation is thought to be a key towards optimizing the benefits of automation for system performance. These systems should adapt to operators' overt and covert behavior, but changes in the system behavior could affect operators as well. For example, the direction of the automation shift (toward full manual or full automatic control) as well as the distance between two successive LOA (one, two, or more "jumps" in the hierarchy) could differently affect the performance of an individual (see Di Nocera, Lorenz, & Parasuraman, 2005). Also, as discussed by Parasuraman et al. (2000), automation can differ in type and complexity. Some forms of automation may simply organize information sources or integrate them. Such forms of "information automation" differ from automation of decision-making functions, in which decision options that best match the incoming information are provided to the user. Automation support at any or all of these stages of processing could be engaged and disengaged by the human operator. In doing so, operators (or adaptive systems) could trigger different shifts in the distance between LOA.

This study investigated this issue using a simple visuo-motor task. The hypothesis was that switching from one LOA to another would affect individuals' performance because of the costs associated with the engagement/disengagement process. Indeed, we found that a better performance was associated with a forward shift in the hard taskload condition, whereas the same forward shift was detrimental in the easiest taskload condition (confirming what was reported by Di Nocera et al., 2005). Also, the advantage of the forward shift (which makes it generally easier the task) was eliminated by the long-term permanence in a LOA, and differential patterns were found as a function of taskload and type of shift. It is worth stress again that the three taskload conditions generated progressively higher workload only in the neutral sequences, whereas the switching conditions generated differential patterns.

The commonsense consideration that only shifts toward a lower level of automation should reflect poor performance is unsupported. Forward shifts may affect performance as well, particularly when workload is moderate.

Overall, these findings suggest that, when individuals perform a task, their cognitive systems set to a particular level (represented by the activation of a particular set of behaviors) and no costs are observed until the level (or rule) remains the same. Actually, under some circumstances, no shifts can even lead to a better performance.

Final considerations and future work. The research activity reported here has reached two main goals. First, it has demonstrated the validity of a novel index of mental workload. This measure is totally non-invasive and opens new frontiers for the real-time assessment of mental load in a variety of tasks. Second, it showed that LOA transitions should be taken into consideration when designing adaptive systems, because they produce costs both in terms of increased cognitive load and performance detriment. However, in order to predict these costs, a different approach to the construct of Level of Automation is necessary. Indeed, the traditional approach to the concept of "Level Of Automation" (LOA) is qualitative in nature: it simply describes the trading of system control between humans and computers. Since Sheridan's seminal work, many taxonomies have been proposed, but they are domain- and task-dependent. This makes it difficult to compare results from different studies. Recently, Terenzi, Camilli, & Di Nocera (2006) have introduced a different approach that will eventually allow to define LOAs quantitatively. This approach is founded upon the idea that LOAs may be characterized in terms of the amount of information traded by humans and machines. For example, at the information-acquisition level (see Parasuraman et al. 2000), LOAs can be defined in terms of number of features of an object to be identified. Thus, automation providing reliable information on 1 out of 4 possible features would have LOA=.25, whereas a system providing aid on 2 out of 4 features would have LOA=.50. A first study showed that it is possible to ascertain the mathematical relations that exist between LOAs and human performance. In other terms, it was possible to predict performance benefits and costs associated with a specific proportion. Considering what has been reported so far, this seems to be best method to study the shift between levels of automation, its costs, and the potential remedies.

REFERENCES

Alpern, M. (1962). Type of movement. In H. Davson (Ed.), *The Eye,* 3, 63-151. New York: Academic Press.

Baddeley, A., & Turner, R. (2005). Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12(6), 1-42.

Bellenkes, A.H., Wickens, C.D. & Kramer, A.F. (1997). Visual scanning and pilot expertise: The role of attentional flexibility and mental model development. *Aviation Space and Environmental Medicine*, 68, 569-579.

Boucsein, W., & Backs, R. W. (2000). Engineering psychophysiology as a discipline: Historical and theoretical aspects. In: R. W. Backs, & W. Boucsein (Eds.), *Engineering Psychophysiology. Issues and Applications* (pp. 3-29). Mahwah, N. J.: Lawrence Erlbaum

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

Bunecke, J. L. (1987). Quantifying some information processing requirements of the pilot's instrument crosscheck. *Proceedings of the Human Factors Society Annual Meeting*, 31, 1301-1305.

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

Clark, P.J., Evans F.C. (1954). Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology,* 35, 445-453.

Colle, H.A., & Reid, G.B. (1998). Context effects in subjective mental workload ratings. *Human Factors*, 40(4), 591-600.

Crocoll, W.M., & Coury, B.G. (1990). Status or recommendation: Selecting the type of information for decision aiding. *Proceedings of the Human Factors Society 34th Annual Meeting*, 12524-1528.

Delaunay, B. (1934). Sur la sphère vide. *Izvestia Akademii Nauk SSSR, Otdelenie Matematicheskikh i Estestvennykh Nauk*, 7, 793-800.

Di Nocera, F., Camilli, M., & Terenzi, M. (in press). A random glance to the flight deck: pilot's scanning strategies and the real-time assessment of mental workload. *Journal of Cognitive Engineering and Decision Making*.

Di Nocera, F., Lorenz, B., & Parasuraman, R. (2005). Consequences of shifting from one level of automation to another: main effects and their stability. In D. de Waard, K.A. Brookhuis, R. van Egmond, and Th. Boersema (Eds.), *Human Factors in Design, Safety, and Management* (pp. 363-376). Maastricht, the Netherlands: Shaker Publishing.

Di Nocera, F., Lorenz, B., Tattersall, A.J., & Parasuraman, R. (2003). New Possibilities for Adaptive Automation and Work Design. In G.R.J. Hockey, A.W.K. Gaillard, & O. Burov (Eds.), *Operator Functional State: The Assessment and Prediction of Human Performance Degradation in Complex Tasks* (pp. 363-372). Amsterdam: IOS Press.

Di Nocera, F., Terenzi, M., & Camilli, M. (2006). Another look at scanpath: distance to nearest neighbour as a measure of mental workload. In D. de Waard, K.A. Brookhuis, A. Toffetti (Eds.),

Developments in Human Factors in Transportation, Design, and Evaluation (pp. 295-303). Maastricht, the Netherlands: Shaker Publishing.

Diez, M., Boehm-Davis, D. A., Hansberger, J. T., Pinney, M. E., Hansberger, J. T., & Schoppek, W. (2001). Tracking pilot interactions with flight management systems through eye movements. *Proceedings of the 11th International Symposium on Aviation Psychology, Columbus, OH.*

Ditchburn, R. W. (1980). The function of small saccades. *Vision Research*, 20, 271-272.

Donnelly, K.P. (1978). Simulations to determine the variance and edge effect of total nearest-neighbor distance. In I. Hodder (Ed.), *Simulation Studies in Archaeology* (pp. 91-95). London, UK: Cambridge University Press.

Endsley, M.R. & Kaber, D.B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42(3), 462-492.

Endsley, M.R., & Kiris, E.O. (1995). The out-of-the-loop performance problem and level of control. *Human Factors*, 37, 381-394.

Ephrath, A. R., Tole, J.R., Stephens, A. T., & Young, L. R. (1980). Instrument scan – Is it an indicator of the pilot's workload? *Proceedings of the Human Factors Society Annual Meeting*, 24, 257-258.

Fitts, P. M., Jones, R.E. & Milton, J.L. (1950). Eye movements of aircraft pilots during instrument-landing approaches. *Aeronautical Engineering Review*, 9(2), 24-29.

Gevins, A.,Smith, M.E. (1999). Detecting transient cognitive impairment with EEG pattern recognition methods. *Aviation, Space, and Environmental Medicine*, 70, 1018-1024.

Gevins, A.S., Leong, H., Du, R., Smith, M.E., Le, J., DuRousseau, D., Zhang, J., & Libove, J. (1995). Towards measurement of brain function in operational environments. *Biological Psychology*, 40, 169-186.

Gopher, D., & Donchin, E. (1986). Workload: An examination of the concept. In K.R. Boff, L. Kaufman and J. P. Thomas (Eds.) *Handbook of perception and human performance*. New York: Wiley.

Haining, R. (2003). *Spatial Data Analysis: Theory and Practice*, Cambridge, UK: Cambridge University Press.

Hammer, Ø., Harper, D.A.T., & Ryan, P.D. (2001). PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontologia Electronica*, 4(1). Retrieved April 5, 2006, from http://palaeo-electronica.org/2001_1/past/main.htm

Hancock, P.A., Chignell, M.H., and Lowenthal, A. (1985). An Adaptive Human-Machine System. *Proceedings of the IEEE Conference on System, Man, and Cybernetics*, 15, 627-629.

Hankins, T.C., & Wilson, G.F. (1998). A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight. *Aviation, Space, and Environmental Medicine*, 69, 360–367.

Harris, R.L., Glover, B.L., & Spady, A.A. (1986). Analytic techniques of pilot scanning behavior and their application. *NASA technical paper 2525*.

Hart, S.G., & Staveland, L.E. (1988). Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam, The Netherlands: Elsevier.

Heinonen T., Lahtinen A., Häkkinen V. (1999). Implementation of Three-Dimensional EEG Brain mapping. *Computers and Biomedical Research*, 32(2): 123-131.

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. In M. Mouloua and J. Koonce (Eds.) *Human-Automation Interaction: Research and Practice* (pp. 84-91). Mahwah, NJ, Erlbaum Associates.

Hornof, A. J., & Halverson, T. (2002). Cleaning up systematic error in eye tracking data by using required fixation locations. *Behavior Research Methods, Instruments, and Computers,* 34(4)*,* 592-604.

Irwin, D. E., & Zelinsky, G. J. (2002). Eye movements and scene perception: Memory for things observed. *Perception & Psychophysics*, 64(6), 882-895.

Jacob, R.J.K., & Karn, K.S. (2003). Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises (Section Commentary). In *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, ed. by J. Hyona, R. Radach, and H. Deubel, 573-605.

Jainta, S., Jaschinski, W. (2002). Fixation disparity: Binocular vergence accuracy for a visual display at different positions relative to the eyes. *Human Factors,* 44, 443-450.

Kaber, D.B., Onal, E., & Endsley, M.R. (1998). Level of automation effects on telerobot performance and human operator situation awareness and subjective workload. In M. Scerbo and M. Mouloua (Eds.), *Automation Technology and Human Performance: Current Research and Trends* (pp. 165-170). Mahwah: Lawrence Erlbaum Associates.

Kaber, D.B., Onal, E., & Endsley, M.R. (2000). Design of Automation for Telerobots and the Effect on Performance, Operator Situation Awareness, and Subjective Workload. *Human Factors and Ergonomics in Manufacturing*, 10 (4), 409–430.

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

Kramer, A. F. and Weber, T. (2000). Applications of psychophysiology to human factors. In J.T. Cacioppo, L.G. Tassinary and G.G. Berntson (eds), *Handbook of psychophysiology*. 2nd edn Cambridge, UK: Cambridge University Press.

Kramer, A. F., & McCarley, J. S. (2003). Oculomotor behaviour as a reflection of attention and memory processes: neural mechanisms and applications to human factors. *Theoretical Issues in Ergonomics Science*, 4(1-2), 21–55.

Kramer, A.F. (1991). Physiological metrics of mental workload: a review of recent progress. In D.L. Damos (Ed.), *Multiple-task performance*. (pp. 279-328). London: Taylor & Francis.

Levine, N. (2004). CrimeStat: *A Spatial Statistics Program for the Analysis of Crime Incident Locations (v 3.0)*. Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC. May.

Lorenz, B., Di Nocera, F., & Parasuraman, R. (2004). Examination of the proximity-compatibility principle in the design of displays in support of fault management in automated systems. In D. de

Waard, K. Brookhuis, and C. Weikert (Eds.) *Human Factors in Design* (pp. 1-17). Maastricht, the Netherlands: Shaker Publishing.

Lorenz, B., Di Nocera, F., Röttger, S., & Parasuraman, R. (2002). Automated Fault-Management During Simulated Space Flight. *Aviation, Space, and Environmental Medicine*, 73, 886-897.

Mayr, U., Keele, S. (2000). Changing internal constraints on action: The role of backward inhibition. *Journal of Experimental Psychology: General*, 129, 4-26.

Miller, J. M. (1973). Visual behavior changes of student pilots flying instrument approaches. *Proceedings of the Human Factors Society Annual Meeting*, 17, 208-214.

Monsell, S. (1996). Control of mental processes. In V. Bruce (Ed.), *Unsolved mysteries of the mind* (pp. 93-148). Hove: Erlbaum.

Moray, N., Inagaki, T., Itoh, M., (2000). Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology: Applied*, 6, 44-58.

Morrison, J.G., & Gluckman, J.P. (1994). Definitions and prospective guidelines for the application of adaptive automation. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: current research and trends* (pp. 256-263). Hillsdale, NJ: Erlbaum.

Norman, D.A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In Davidson, R., Schwartz, G., & Shapiro, D. (Eds.), *Consciousness and Self Regulation: Advances in Research and Theory, 4*,1–18. Plenum, New York, NY.

O'Donnell, R.D. & Eggemeier, F.T. (1986). Workload assessment methodology. In K.R. Boff, L. Kaufman & J.P. Thomas (Eds.), *Handbook of perception and human performance*. Volume II, cognitive processes and performance. (pp 42/1-42/49). New York: Wiley.

Parasuraman, R., & Caggiano, D. (2002). Mental workload. In V. S. Ramachandran (Ed.), *Encyclopedia of the Human Brain*. San Diego: Academic Press.

Parasuraman, R., Molloy, R., Singh, I. L. (1993). Performance consequences of automation induced complacency. *The International Journal of Aviation Psychology*. 3, 1-23.

Parasuraman, R., Mouloua, M., Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors*. 38, 665-679.

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

Pelz, J.B., & Canosa, R. (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research*, 41, 3587–3596.

Recarte & Nunes (2000). Effects of verbal and spatial imagery tasks on eye fixations while driving. *Journal of Experimental Psychology: Applied*, 6, 31-43.

Recarte & Nunes (2003). Mental workload and driving: Effects on visual search, discrimination and decision making. *Journal of Experimental Psychology: Applied*, 9, 119-137.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, 124, 207–231.

Rovira, E., McGarry, K., & Parasuraman, R. (2002). Effects of unreliable automation on decision making in command and control. *Proceedings of the Annual Meeting of the Human Factors Society*, 428-432.

Rubinstein, J., Meyer, D., Evans, J. (2001), Executive Control of Cognitive Processes in Task Switching. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 4, 763-797.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the Eye Tracking Research and Applications Symposium.*

Sarter N.B., & Schroeder, B. (2001). Supporting decision making and action selection under time pressure and uncertainty: the case of in-flight icing. *Human Factors*, 43, 573-583.

Scerbo, M.W., (1996). Theoretical perspectives on adaptive automation. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 37-63). Mahwah, NJ: Erlbaum.

Scerbo, M.W., Freeman, F.G., Mikulka, P.J., Parasuraman, R., Di Nocera, F., & Prinzel, L.J. (2001). The efficacy of psychophysiological measures for implementing adaptive technology. *Technical Paper NASA/TP-2001-211018*. Hampton, VA: NASA Langley Research Center.

Schneider, W., and Shiffrin, R.M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. Psychological Review, 84, pp1-66.

Schuch, S., & Koch, I. (2003). The role of response selection for inhibition of task sets in task shifting. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 92-105.

Sheridan, T.B. (2002). *Humans and Automation: System Design and Research Issues*. New York: Wiley.

Shiffrin, R. M., and Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual and learning, automatic attending, and a general theory. Psychological Review, 84, 127-190.

Smith, K.J., Valentino, D.A., & Arruda, J.E. (2003). Rhythmic oscillations in the performance of a sustained attention task. *Journal of Clinical and Experimental Neuropsychology*, 25, 561-570.

Terenzi, M., Camilli, M., Di Nocera, F. (2006). Making it Quantitative: Development of a New Taxonomy for Levels of Automation. *Poster presented at the Human Factors & Ergonomics Society – Europe Chapter Meeting*. Sheffield, UK.

Tole, J.R., Stephens, A.T., Vivaudou, Eprath, A. & Young, L.R. (1983*).* Visual scanning behavior and pilot workload. *NASA Contractor Report 3717*. Hampton, Virginia: NASA Langley Research Center.

Trafton, J.G., Altmann, E.M., Brock, D.P., & Mintz, F.E. (2003). Preparing to resume an interrupted task: effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies*, 58(5), 583-603.

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

Unema, P., 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.

Ungerleider, L.G. & Mishkin, M. (1982). Two cortical visual systems. In D.J. Ingle, M.A. Goodale & R.J.W. Mansfield (Eds). *Analysis of Visual Behavior* (pp.549-586). Cambridge, MA: MIT Press.

Van Orden, K., Makeig, S, Limbert, W., & Jung, T-P. (2001). Eye activity correlates of workload during a visualspatial memory task. *Human Factors*, 43(1),111-21.

Wierwille, W. W., Connor, S. A. (1983). Evaluation of 20 workload measures using a psychomotor task in a moving-base aircraft simulator, *Human Factors*, 25, 1-16.

Wickens, C.D., & Hollands, J.G. (2000). *Engineering Psychology and Human Performance*. Upper Saddle River, NJ: Prentice Hall.

Wilson, G. F., Lambert, J. D., and Russell, C.A. (1999). Performance enhancement with real-time physiologically controlled adaptive aiding. *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting,* 3, 61-64.

Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6), 767-791.