Using theta and alpha band power to assess cognitive workload in multitasking environments

Puma [1]	Matton [1 2]	Paubel [1]	Raufaste [1]	El-Yagoubi [1]
Sébastien	Nadine	Pierre-V.	Éric	Radouane

[1] University of Toulouse & CNRS (UMR5263). 5 allées Antonio Machado, 31058 Toulouse Cedex 9.
[2] ENAC École Nationale d'Aviation Civile. 7 Avenue Édouard Belin, 31000 Toulouse.

Abstract

Cognitive workload is of central importance in the fields of human factors and ergonomics. A reliable measurement of cognitive workload could allow for improvements in human machine interface designs and increase safety in several domains. At present, numerous studies have used electroencephalography (EEG) to assess cognitive workload, reporting the rise in cognitive workload to be associated with increases in theta band power and decreases in alpha band power. However, results have been inconsistent with some failing to reach the required level of significance. We hypothesized that the lack of consistency could be related to individual differences in task performance and/or to the small sample sizes in most EEG studies. In the present study we used EEG to assess the increase in cognitive workload occurring in a multitasking environment while taking into account differences in performance. Twenty participants completed a task commonly used in airline pilot recruitment, which included an increasing number of concurrent sub-tasks to be processed from one to four. Subjective ratings, performances scores, pupil size and EEG signals were recorded. Results showed that increases in EEG alpha and theta band power reflected increases in the involvement of cognitive resources for the completion of one to three subtasks in a multitasking environment. These values reached a ceiling when performances dropped. Consistent differences in levels of alpha and theta band power were associated to levels of task performance: highest performance was related to lowest band power.

Keywords

Cognitive workload; Electroencephalography; Multitasking, spectral power.

Highlights

Theta band spectral power increased with increase of cognitive load and reached a plateau after overload.

Theta and alpha bands spectral power allowed to distinguish participants based on their performances during task completion, better performers generating lower levels of spectral power.

Alpha band spectral power allowed differences to be distinguished according to the cognitive resource involvement during the task.

Acknowledgement

The authors wish to thank Robert Ruiz for his valuable comments on signal processing during the experiment and Antonio Rafael Hidalgo Munoz and Noémie Te-rietmolen for their comments and contributions on all aspects of this document.

Corresponding author

Correspondence should be addressed to Sébastien Puma: sebastien.puma@univ-tlse2.fr

Introduction

Cognitive workload is considered as an important factor in human performance, affecting human error, system safety, productivity and operator satisfaction (Xie & Salvendy, 2000). It can be defined as "the proportion operator information processing capacity or resources that is actually required to meet system demands" (Eggemeier, Wilson, Kramer & Damos, 1991; Cain, 2007; see also Moray, 1979; Vidulich & Tsang, 2012 for reviews), the amount of cognitive resources being limited (*e.g.*, Broadbent, 1958). These cognitive resources mainly refer to attentional resources (Patten, Kircher, Östlund, Nilsson, & Svenson, 2006; Hollands & Wickens, 2000; Wickens, 1991; 2008) and to working memory capacity (Brouwer *et al.*, 2012; Berka *et al.*, 2007; Grimes, Tan, Hudson, Shenoy & Rao, 2008) both of them representing the cognitive processes involved in cognitive workload (Sauseng, Klimesch, Schabus, & Doppelmayr, 2005). This concept of cognitive workload has raised many theoretical concerns (Tricot & Chanquoy, 1996), but "perhaps the most basic issue in the study of cognitive workload is the problem of how to actually measure it" (Gevins & Smith, 2003).

When assessing cognitive workload, three different measurements are usually distinguished: behavioral, subjective and physiological (Vidulich & Tsang, 2012; Cegarra & Chevalier, 2008; Kramer, 1990; Cain, 2007). They provide different information and are only rarely correlated (Funke *et al.*, 2013), which lead to the hypothesis that these measurements reflect different aspects of the cognitive workload phenomenon (Matthews, Reinerman-Jones, Barber & Abich, 2015; Cain, 2007).

The present study focused on physiological measurements of cognitive workload, mainly using electroencephalography (EEG), however other techniques such as pupillometry might also provide valuable insight.

Pupil diameter is assumed to reflect general arousal and has also been shown to reflect variations of workload (Beatty & Lucero-Wagoner, 2000 for a review) either during laboratory experiments (Kahneman & Beatty, 1966; Peavler, 1974) or during more ecologically-valid tasks (Just & Carpenter, 1993; Ahlstrom & Friedman-Berg, 2006; Stein, 1992). Pupil size increases with cognitive effort (Kahneman, Tursk, Shapiro, & Crider, 1969;

Iqbal, Adamczyk, Zheng & Bailey, 2005), in response to inhibition which is assumed to consume attentional resources (Laeng, Ørbo, Holmlund, & Miozzo, 2011; Chiew & Braver's, 2013) and thus also with attentional load (Lisi, Bonato & Zorzi, 2015). However, pupil size may vary with luminosity (Beatty & Lucero-Wagoner, 2000 for a review) and with other, non-cognitive, factors, such as physical effort (Richer & Beatty, 1985). Despite these limitations, pupil size analysis remains a good indicator of cognitive load variations in laboratory experiments (Beatty & Lucero-Wagoner, 2000).

EEG is another extensively used to assess cognitive workload (Ke *et al.*, 2014). When using EEG, it seems necessary to rely on studies that focus on the cognitive determinants of the cognitive workload, mainly attention and working memory (*e.g.*, Wickens, Kramer, Vanasse & Donchin, 1983). By explaining the implications of different resources, studies on cognitive determinants may help to reconcile divergent results on cognitive workload measurements. While studies of cognitive processes used laboratory settings, cognitive workload studies often used more ecologically-valid tasks. This difference might be the source of inconsistencies between the two domains. Nonetheless, explaining attention and working memory processes might help to understand variations in the measurements of cognitive workload. In the next sections, we will firstly present empirical findings of laboratory studies on EEG markers of attention and working memory. Secondly, we will present empirical findings of ecologically-valid studies on EEG cognitive workload assessments.

EEG markers of attention and working memory

Attention and working memory share parts of the same cerebral regions. However, it remains unclear whether they share the same networks and are different functions emerging from these networks or whether they rely on distinct ones (LaBar, Gitelman, Parrish, & Mesulam, 1999). Working memory is supported by prefrontal cortex and parietal areas (Sauseng *et al.*, 2005), the left parietal lobe supporting the phonological loop (Ravizza, Behrmann, & Fiez, 2005) and the right parietal lobe supporting the visuo-spatial sketchpad (see d'Esposito *et al.*, 1998 for a review, but see also LaBar *et al.*, 1999). Activation of frontal and right parietal cerebral regions, reflected by a synchronization in the theta band (4-8 Hz) and a desynchronization in the alpha band (8-12 Hz), is sensitive to working memory load (see Schacter, 1977; Basar, Basar-Eroglu, Karakas & Schürmann, 2001; Kahana, Seelig, & Madsen, 2001; Klimesch, 1999 for review). Fronto-parietal theta power has been linked to working memory capacity in numerous studies (Sauseng, Griesmayr, Freunberger & Klimesch, 2010; Klimesch, 1996), with a higher level of theta band spectral power elicited reflecting lower working memory capacity (Klimesch, Vogt & Doppelmayr, 1999; Klimesch, 1999). These differences might be due to different amounts of cognitive resources available as well as to differences in strategies used to complete the task or perhaps an interaction between the two hypotheses (Gulbinaite, Johnson, de Jong, Morey & van Rijn, 2014).

In a similar manner, the solicitation of attentional resources has been linked mainly to a desynchronization of the alpha band (Klimesch, 1996; Klimesch, Doppelmayr, Russeger, Pachinger & Schwaiger, 1998) and theta band synchronization (Gevins & Smith, 2000). Both processes share the same cerebral regions and vary in the same way for numerous tasks, but alpha band synchronizations were also found during tasks soliciting frequent task switching

(Pope, Bogart & Bartolome, 1995). Other studies also found alpha band power to increase with task demand; (Borghini, Astolfi, Vecchiato, Mattia & Babiloni, 2014; Kamzanova, Kustubayeva & Matthews, 2014; Zhao, Zhao, Liu & Zheng, 2012). Recently, it was proposed that both alpha band synchronization and desynchronization might be responsible for two different working memory maintenance mechanisms (Capilla, Schoffelen, Paterson, Thut & Gross, 2014). As a result, alpha band synchronization would support interfering item inhibition (Rihs, Michel & Thut, 2007) while alpha band desynchronization would support relevant item maintenance (Fukuda, Mance & Vogel, 2015).

EEG related workload assessment

Despite differences between laboratory settings and ecologically-valid experiments, results obtained in both fields are mostly consistent. Indeed, EEG has often been used to assess changes in mental workload and is probably the "most studied mental workload indicator" (Ke *et al.*, 2014; Gevins *et al.*, 1998). Considering EEG methods, an increase in workload is said to be associated with theta synchronization and with an alpha desynchronization, mainly at frontal and parietal sites, (Smith, McEvoy, & Gevins, 1999; Antonenko, 2007).

In eliciting cognitive workload, two approaches are usually employed. The first consists in increasing the difficulty of the task, with the assumption that the more processing steps the task requires in a time unit, the higher the cognitive workload (Johannsen, 1979). The second way is to use multitasking paradigms, since the number of concurrent tasks to be processed is one of the major determinants of cognitive workload (Schvaneveldt, 1969; Yeh & Wickens, 1988; Rogers & Monsell, 1995).

Alpha spectral power variations

Alpha spectral band power has been shown to decrease with increased task difficulty (Sterman & Mann, 1995; Klimesch, 1999; Ota, Toyoshima, & Yamauchi, 1996), as well as with increased memory load (Fairclough & Venables, 2006; Ryu & Myung, 2005; Sterman & Mann, 1995; Fairclough, Venables, & Tattersall, 2005; Fournier, Wilson & Swain, 1999; Gevins *et al.*, 1998; Smith, Gevins, Brown, Karnik, & Du, 2001). In the same way, alpha band power decreases with the increase in experienced time pressure (Slobounov, Fukada, Simon, Rearick, & Ray, 2000). This decrease in alpha brain waves is mainly located in the occipital and parietal brain locations and may be modulated by high inter-individual variations (Klimesch, 1999; Kramer, 1990). It is usually attributed to modulation due to task related attention demand, but the mere onset of the task may sometimes be sufficient to cause the suppression of alpha waves (Valentino, Arruda, & Gold, 1993).

Theta spectral power variations

On the other hand, theta spectral power is thought to increase along with numerous other factors, such as time pressure (Slobounov *et al.*, 2000) cognitive resource demand (see Vidulich & Tsang, 2012 for a review) and the number of concurrent tasks to be processed (Yamada, 1998; Fairclough & Venables, 2006; Fairclough, Venables, & Tattersall, 2005). This increase is mainly observed in fronto-central regions, though these locations may be modulated by age (McEvoy, Pellouchoud, Smith, & Gevins, 2001). However, using increasingly difficult tasks to elicit consistent patterns of increasing theta spectral power has been proved inefficient in numerous studies (Käthner, Wriessnegger, Müller-Putz, Kübler, & Halder, 2014; Fournier, Wilson and Swain, 1999; Baldwin and Penaranda, 2012; Funke *et al.*, 2013) or revealed inconsistent patterns (Brookings, Wilson & Swain, 1996; Pigeau, Hoffman, Purcell & Moffitt, 1988). For example, Gevins and his colleagues (1995) reviewed three of the experiments of their team using tasks of increasing difficulty (Gevins & Schaffer, 1980; Gevins *et al.*, 1979; Gevins, Zeitlin, Doyle, Schaffer, & Callaway, 1979). None of these revealed a significant increase of theta band power in relation to the difficulty of the task. Increasing the number of concurrent tasks to be performed simultaneously, also led to either no pattern or an inconsistent one in different studies (Holm, Lukander, Korpela, Sallinen, & Müller, 2009; Fournier, Wilson & Swain, 1999). Moreover, according to a review by Kramer (1990), theta band power should decrease with an increasing cognitive workload, a result already reported in Sirevaag, Kramer, DeJong and Mecklinger (1988) and in Natani and Gomer (1981).

Hypothesis on results differences

The lack of consistent variations in EEG theta rhythms and an incoherent pattern might arise from two possible methodological issues. Either these studies used paradigms where the low workload condition demanded too many cognitive resources to allow for significant variations with other workload conditions (see Kramer, 1990 for a discussion of this point) or the inter-individual differences overshadowed the variations elicited by the task manipulations. The first explanation was suggested by Kramer (1990) who compared EEG patterns with regard to theta rhythm in three studies (Sirevaag *et al.*, 1988; Natani & Gomer, 1981; Pigeau *et al.*, 1988). He remarked that the differences in theta rhythm were due to differences in the difficulty in the initial task.

The second concerns a statistical issue occurring when too few subjects with high differences perform the same tasks. Differences between participants will elicit high variations while differences due to the task are considered as negligible using a statistical test of hypothesis. None of these studies have tried to form clusters of participants based on their performances and to compare the theta spectral measurements of these clusters. Such cluster analysis might reduce the influence of inter-individual differences since different participants might generate different levels of theta spectral activity according to their cognitive resources (Beatty & O'Hanlon, 1979; Beatty, 1977; Valentino, Arruda & Gold, 1993).

The present study

In this experiment, we took into account both concerns. Firstly, we induced cognitive workload increase by increasing the number of subtasks to perform simultaneously, starting with a very easy monitoring subtask and ending with a complex multitasking environment. Thus, we maximized the likelihood of variations in the level of cognitive workload. Moreover, we used also subjective and pupil size measurements in order to check for these variations. Secondly, we collected data on a large group of participants (n = 20) in order to group participants with similar performances and thus avoid to confound inter-individual differences and differences among workload conditions.

More precisely, we used a multitasking protocol employed in a French airline pilot recruitment. It proposes an increasing number of tasks to be processed, rising from one to four concurrent multitasks. Since the number of concurrent sub-tasks has proved to be an efficient way to induce increasing levels of cognitive workload, we expected this protocol to generate increasing cognitive workload starting with a low level. We used behavioral, subjective and pupil size measurements as overall cognitive workload indicators to compare with EEG variations. We computed a cluster analysis based on performance and calculated EEG spectral power for each homogeneous group.

We anticipated that theta band power level would increase with a rising number of concurrently processed sub-tasks, this level being higher for the lower performance group. As our multitasking setting imposed a larger need to inhibit concurrent sub-tasks while processing than to maintain relevant items in working memory, we also expected alpha band power to increase, with differences between the performance-based clusters. However, given the nature of the task used we did not expect to be able to make any hypothesis on the nature of the differences of processing observed, whether related to cognitive resource capacity or to strategy use.

Method

Participants

Twenty participants took part in this study (14 women and 6 men), aged from 19 to 38 (mean age = 27.25 *sd* = 3.88). Most of them (16) were right handed. The mean number of college years after graduation was 4.1 (*sd* = 1.57). They all volunteered for this experiment and were not paid. They all gave an informed verbal consent before their participation.

Material

Pupil size was recorded using an EyeLink 1000 remote eye-tracker (SR Research Ltd., Mississauga, Ontario, Canada) connected to a Pentium IV 3.0GHz computer. This eye-tracker has a spatial accuracy greater than 0.5, and a 0.01 spatial resolution. The sampling rate was set to 1000Hz. The camera was placed at a distance of 20cm from the screen and the eye-camera distance was 60cm. A chin and forehead rest was used to maintain these distances and to avoid head movements. We used a display screen DELL 19” with a refresh rate of 75 Hz and a resolution of 1024 x 768 pixels. All eye-tracking data was extracted using the SR Research default algorithm. Simulation room lighting was maintained constant. The luminance was constant across all phases of the study (*i.e.*, 212 lux as measured a posteriori with an analogical luxmeter Extech 401025).

EEG was continuously recorded using a 32-channel BioSemi ActiveTwo system (BioSemi, Amsterdam, Netherlands) connected to an Intel i5 3.0 GHz computer. Electrodes were mounted on an elastic head cap and located at standard positions on the left and right hemispheres at frontal, central, parietal, occipital and temporal locations (10/20 international system): Fp1, AF3, F7, F3, FC1, FC5, T7, C3, CP1, CP5, P7, P3, Pz, PO3, O1, Oz, O2, PO4, P4, P8, CP6, CP2, C4, T8, FC6, FC2, F4, F8, AF4, Fp2, Fz, Cz. Due to the frontal localization of the electrodes Fp1 and Fp2, their data was excluded as it was contaminated by occasional contacts with the forehead rest, producing exceedingly high electrical values. Only the data from the 30 remaining electrodes was used. The impedance of the electrodes was always lower than 3 kΩ, and EEG data was recorded with a 512 Hz sampling frequency and filtered during the offline analysis with a Butteworth band pass of 0.01-40 Hz. In addition, the Horizontal Electro-OculoGram (HEOG) was recorded from a bipolar installation with electrodes placed 1 cm to the left and right of the external canthi; the Vertical (VEOG) was recorded from a bipolar montage with electrodes placed beneath and above the right eye, to detect blinks and vertical eye movements. The EEG was re-referenced offline from the algebraic average of two electrodes positioned on the left and right mastoids. EEG data was then processed using the BrainVision Analyzer (v2.4) software and corrected for eye movements by using an independent component analysis (ICA) based correction with the addition of VEOG and HEOG channels.

Subjective workload measurements were recorded after each phase of the task, using a computerized version of Nasa-Tlx. Participants rated each of the six dimensions by moving a cursor on a scale ranging from 0 to 100. Behavioral performances were recorded during the entire experiment based on instantaneous performance on each subtask.

Procedure

The multitasking situation adopted was the Priority Management Task, currently used at the ENAC (Ecole Nationale de l'Aviation Civile, the French civil aviation university) for airline pilot student selection. It includes four phases of progressive difficulty during which participants had to manage one to four tasks simultaneously. During the first phase, called the gauge monitoring task, participants had to monitor four gauges by using the left joystick. Periodically (each 15s), one or more of the gauges needles drifted from their location. Before the first phase, participants were allocated one minute to train with the same task. During the second phase, participants had to deal simultaneously with the first sub-task (gauge monitoring) and the second sub-task, called the tracking task. In this second sub-task, participants had to maintain a white cross in a white circle by using the right joystick. Periodically (each 15s), the white circle moved rapidly from its previous location. This second phase was also preceded by one minute of training with the two sub-tasks. During the third phase, participants had to simultaneously manage the first two sub-tasks in addition to a third, the letter detection by using the keys F1 to F9. In this sub-task, a series of three target letters were presented and participants had to detect them in a block of nine letters by pressing the corresponding keys on the keyboard. Periodically (each 15s), the series of nine letters was renewed. This third phase was preceded by one minute of training with the three sub-tasks. During the fourth phase, represented in Figure 1, participants had to manage the first three sub-tasks and a fourth, a mental arithmetic sub-task. In this sub-task, participants had to answer to a series of easy mental arithmetic problems (for example: "what is 11% of 500?") by using the number keypad and then validating their answer. A new problem was presented each 15s. This fourth phase was preceded with training on the four sub-tasks for one minute. Importantly, the subtask events were synchronized, forcing the participant to make choices.

During each phase, instantaneous performance on each sub-task and global performance were displayed to participants. At the end of each of the four phases, participants rated the six dimensions of the Nasa-Tlx questionnaire (Hart & Staveland, 1988) using a computer version and after the last phase they completed both the dimension and the pair comparisons. The entire experimental protocol lasted one hour.

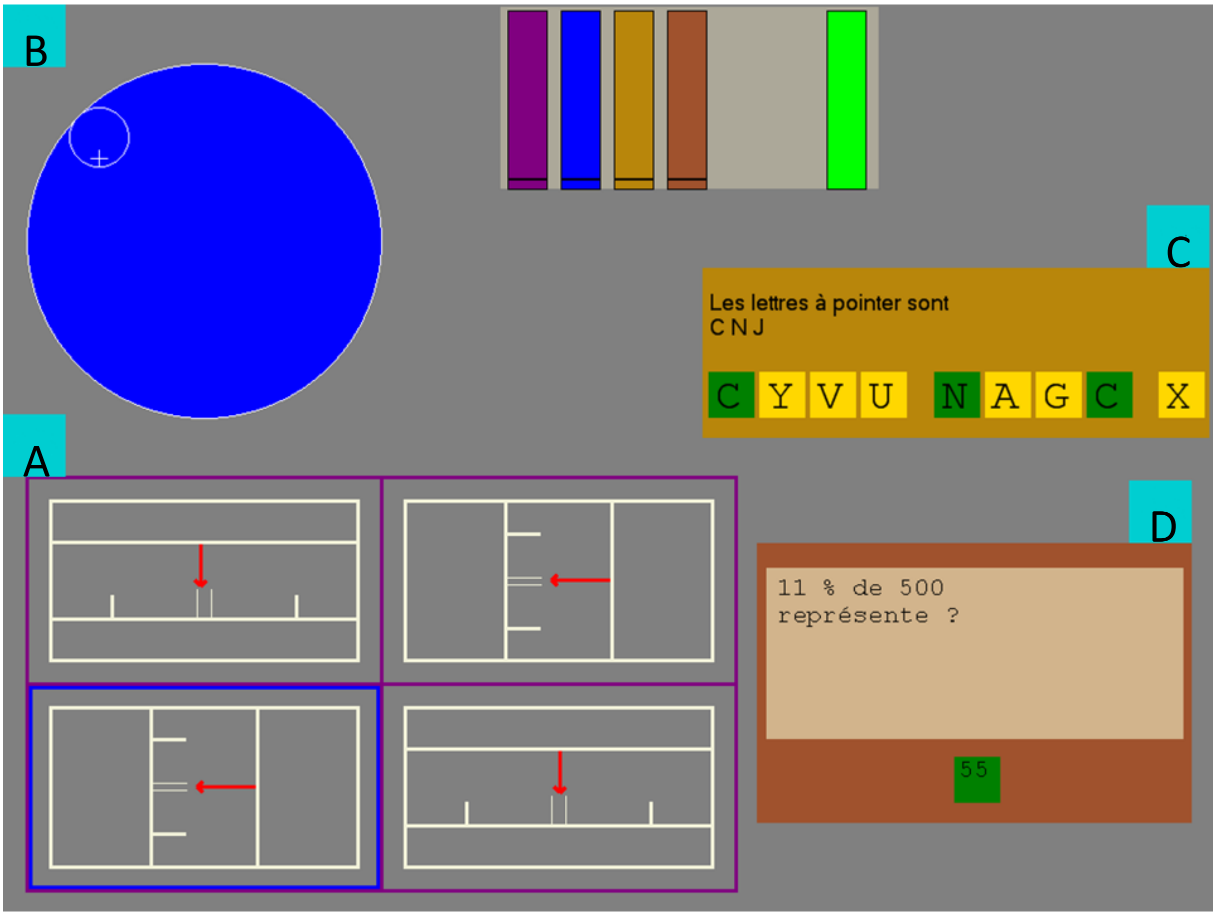

Figure 1: Experimental protocol. During Phase 1, participants had to manage task "A", the gauge monitoring task. During Phase 2, they had to simultaneously perform task "A" and "B", the tracking task. In Phase 3, the task "C", letter detection was added to the first two tasks and during Phase 4, the participants had to solve small calculation problems at the same time as the three precedent sub-tasks. During the whole task, instantaneous performance was displayed to the participants (top of the screen).

Data processing

EEG signal processing: The signal was segmented at the end of each phase, resulting in four segments of different length according to the phase. Then the signal was divided into 10s length segments with an overlap of 5s. Then a Fast Fourier Transform (FFT) using a hamming window was applied to each segment to avoid differences due to segment size on band power calculation. The average power associated with each frequency band according to the two rhythms Theta (4-8Hz) and Alpha (8-12Hz) during each of the four experimental phases were then calculated. These values were then log transformed to achieve normal distributions.

Dependent variables and statistical analysis

Raw Nasa-Tlx scores for each dimension were averaged to compute a global score. Performance data were obtained each 100ms for each subtask during each phase. A global performance measurement was computed by averaging sub-task performance. To prevent the participants giving up with one sub-task, global performance dropped to zero when one of the four sub-task performances fell below a threshold of 10%. A global score was then computed by averaging the global performance across the phase. This global performance score ranged from 0 to 150. Pupil size measurements were transformed into a z-score taking training phases into account.

Results

The results of the different measures are summarized in Table 1 and detailed below.

Nasa-Tlx: The global Nasa-Tlx score increased significantly with the number of tasks to process, $F_{(3,57)} = 74.56$, $p<.001$, $\eta^2_{partial} = .797$, see Figure 2 left panel. Further pair comparisons using a LSD test showed a significant increase between phases (all *ps*<.001).

Pupil size analysis: The z-score measurement of pupil size increased significantly with the number of tasks to process, $F_{(3,\ 57)} = 21.57$, $p<.001$, $\eta^2_{partial} = .532$. Post-hoc analysis revealed a significant increase between the three first phases (all *ps*<.02) but only a marginally significant difference between phases 3 and 4 ($p = .10$). Since performances feedback represented less than 5% of the fixations at any phase, this AOI was not further considered.

		Phase 1	Phase 2	Phase 3	Phase 4
Nasa-Tlx	Overall	3,28 (1,57)	5,17 (1,75	5,93 (1,68)	7,22 (1,68)
	Cluster High	3,68 (1,41)	5,05 (1,84)	5,58 (1,63)	7,02 (1,33)
	Medium	2,37 (1,10)	5,01 (1,83)	5,98 (1,72)	7,12 (2,18)
	Low	4,88 (2,66)	6,41 (0,78)	7,37 (0,39)	8,57 (0,31)
Performance	Overall	126,07 (27,50)	118,57 (40,47)	112,64 (32,77)	92,676 (26,55)
	High	137,36 (7,85)	135,94 (6,99)	127,77 (7,16)	111,87 (4,97)
	Medium	128,27 (8,72)	125,80 (7,68)	116,68 (8,49)	84,638 (10,15)
	Low	60,81 (58,80)	2,74 (1,79)	20,85 (13,21)	28,845 (10,03)
DP	Overall	-0,79 (0,62)	-0,45 (0,33)	-0,03 (0,31)	0,18 (0,18)
	High	-1,02 (0,60)	-0,44 (0,30)	0,050 (0,19)	0,20 (0,15)
	Medium	-0,68 (0,42)	-0,38 (0,37)	-0,16 (0,42)	0,20 (0,16)

	Low	-0,14 (1,20)	-0,74 (0,24)	0,026 (0,29)	-0,01 (0,34)
EEG Theta	Overall	4,88 (0,82)	5,01 (0,79)	5,34 (0,91)	5,33 (0,91)
	High	4,55 (0,77)	4,72 (0,78)	5,10 (0,99)	5,11 (1,01)
	Medium	5,11 (0,72)	5,24 (0,66)	5,54 (0,74)	5,48 (0,74)
	Low	5,58 (0,65)	5,59 (0,66)	5,76 (0,74)	5,80 (0,73)
EEG Alpha	Overall	4,17 (0,91)	4,31 (0,89)	4,46 (0,89)	4,44 (0,91)
	High	3,80 (0,80)	3,97 (0,83)	5,10 (0,87)	5,11 (0,89)
	Medium	4,47 (0,81)	4,58 (0,73)	4,72 (0,74)	4,68 (0,78)
	Low	4,78 (1,09)	4,92 (1,08)	4,85 (1,05)	4,92 (1,03)

Table 1 : Means (and Standard Deviation) of the measures of the study, across all participants and by cluster based on performances.

Performances: Performance scores varied significantly with the number of tasks to process, $F_{(3,57)} = 22.24$, $p<.001$, $\eta^2_{partial} = .539$. When comparing phases, performances did not significantly decrease between phase 1 and phase 2 ($M_1 = 126.07$ and $M_2 = 118.57$ respectively, $p = .161$) but phase 1 was significantly higher than phase 3 ($M_3 = 112.645$, $p = .003$) and higher than phase 4 ($M_4 = 92.676$, $p<.001$). Phase 2 was not significantly higher than phase 3 though marginally different ($p = .063$), but was significantly higher than phase 4 ($p<.001$). Performance decreased significantly between phase 3 and phase 4 ($p<.001$). The lack of significant difference between performance in phases 1 and 2 might be accounted for by the lack of sensitivity of performance based workload measurements already mentioned: Performance measurements might be maintained at a specified acceptable level while the cognitive workload rises to maintain this level (Wickens, 2002; 2008).

Participants could be gathered into three different groups, particularly during the fourth phase. Thus we used performance at this most demanding phase (phase 4) to split participants into different clusters. A hierarchical cluster analysis using the Ward method revealed three clusters of participants based on their performances (see Figure 2, top panel). The first cluster gathered higher performers, with raw scores on phase 4 ranging from 105.6 to 118.9. A second cluster gathered medium performers, with performances on phase 4 ranging from 67 to 97.3. The third cluster comprised two participants with very low performances (ranging from 21.7 to 35.9). When splitting performance data by cluster, a repeated measure ANOVA following the phase performed on each cluster revealed significant differences for cluster 1, “higher performers” ($F_{(3,27)} = 42.11$, $p<.001$, $\eta^2_{partial} = .82$) and for cluster 2, “medium performers” ($F_{(3,21)} = 36.97$, $p<.001$, $\eta^2_{partial} = .84$). The third cluster, comprising only two participants, showed no significant performance differences according to the phase ($F_{(3,3)} = 1.30$, $p = .42$) with their performances remaining at a low level for each phase. For higher performers, pair comparisons revealed no decrease between phase 1 and phase 2 ($p = .29$) but a significant decrease between phase 2 and phase 3 ($p = .009$) and also between phase 3 and phase 4 (p = .001). The same analysis for the medium performers revealed a similar pattern: No differences between phase 1 and phase 2 (p = .57), nor between phase 2 and phase 3 though there was a trend (p = .051) and a significant decrease between phase 3 and phase 4 (p<.001).

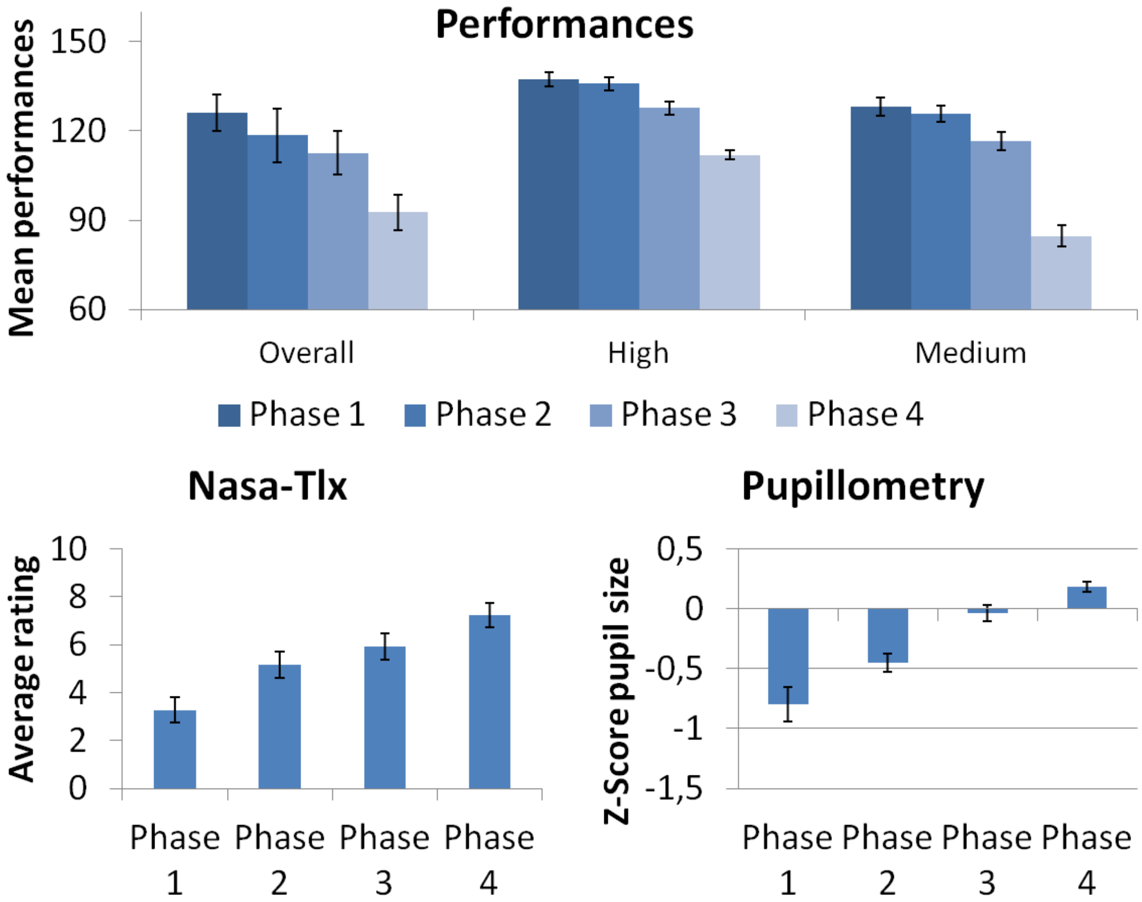

Figure 2: Means and standard errors (1 SE) for subjective ratings and performance data. Two main clusters emerged from the performance data, the two remaining participants being gathered into a third cluster. Subjective ratings increased between each phase while performance decreased significantly only between phase 3 and phase 4. Pupil size increased between all phases significantly except between phase 3 and phase 4, where the increase was only marginal.

Electroencephalography: For each rhythm we performed a two-way ANOVA with Electrodes $_{(28)}$ and Phases $_{(4)}$ as within factors. Since most of the previous studies focused on frontal electrodes when analyzing theta spectral feature data and on parietal electrodes regarding the alpha rhythm, we reproduced our analysis on five electrodes centered on these areas. Considering the low sample size of the low performers cluster (n = 2), this cluster was removed from all ANOVAs. However, low performers' data was analyzed descriptively as we were interested in comparing their levels of frequency band spectral powers to those of medium and higher performers. We focused our analysis (1) on global level differences between medium and higher performer clusters, (2) on band frequency differences across phases for each cluster, when the number of concurrent tasks increased. We then used planned comparisons to compare clusters and phases.

Theta rhythm: A repeated measures Anova with the Cluster as a between subject factor and the 28 electrodes and the four phases as intra-participants variables revealed a main effect of the cluster, for the theta rhythm $F_{(2,162)} = 5.20$, $p = .017$, $\eta^2_{partial} = .060$. Analyses of the theta rhythm are illustrated in Figure 3. Considering cluster comparisons, the medium performers cluster exhibited a significantly higher level of theta spectral power than the higher performers cluster at phase 1, $F_{(1,16)} = 7.23$, $p = .02$, at phase 2, $F_{(1,16)} = 6.93$, $p = .02$, and at phase 3, $F_{(1,16)} = 4.80$, $p = .04$, but not at phase 4, $F_{(1,16)} = 2.91$, $p = .11$. Thus, except for the highest workload level, higher performers had lower theta spectral power than medium performers. Moreover, in descriptive terms, theta spectral power for low performers was higher than for that of the medium performers cluster at each phase.

Considering phase comparisons, when focusing on the higher performers cluster, theta spectral power increased between phase 1 and phase 2, $F_{(1,16)} = 13.36$, $p = .002$, between phase 2 and phase 3, $F_{(1,16)} = 21.52$, $p<.001$, but not between phase 3 and phase 4, $F_{(1,16)} = 0.08$, $p = .77$. When considering the medium performers cluster, the same pattern was observed with an increase between phase 1 and phase 2, $F_{(1,16)} = 6.12$, $p = .02$, between phase 2 and phase 3, $F_{(1,16)} = 10.69$, $p = .005$, but not between phase 3 and phase 4, $F_{(1,16)} = 0.96$, $p = .34$. To summarize, for higher as for medium performers, theta spectral power increased from one to three concurrent tasks and then reached a plateau. Adding a fourth task resulted in no increase. In the meantime, both low performers exhibited a smaller increase between phases 1 and 4 than the mean of medium or high performers.

Based on the literature, five electrodes were selected as particularly relevant for the analysis of the theta band frequency on frontal localization. A repeated measures Anova, including the two main regions of interest (frontal and parietal) and the cluster as a between subject factor revealed a significant effect of the phase ($F_{(3,51)} = 18.29$, $p < .001$, $\eta^2_{partial} = .518$) the region ($F_{(2,51)} = 8.17$, $p = .011$, $\eta^2_{partial} = .325$ and of the cluster ($F_{(2,17)} = 4.09$, $p = .036$, $\eta^2_{partial} = .325$). The theta rhythm appeared to be higher for the frontal region than the parietal region ($p = .011$) though the interaction between region and cluster was not significant ($F_{(2,17)} = 0.13$, $p = .883$, $\eta^2_{partial} = .014$) nor was the interaction between phase and cluster ($F_{(3,51)} = 1.42$, $p = .227$, $\eta^2_{partial} = .143$).

Performing the same analysis as previously, based on cluster with these five electrodes centered only in the frontal area (Fz, F3, F4, F7 and F8) did not allow to discriminate between phase 1 and phase 2 neither for the higher performers cluster, $F_{(1,16)} = 2.67$, $p = .121$, nor for the medium performers cluster, $F_{(1,16)} < 0.01$, $p = .954$.

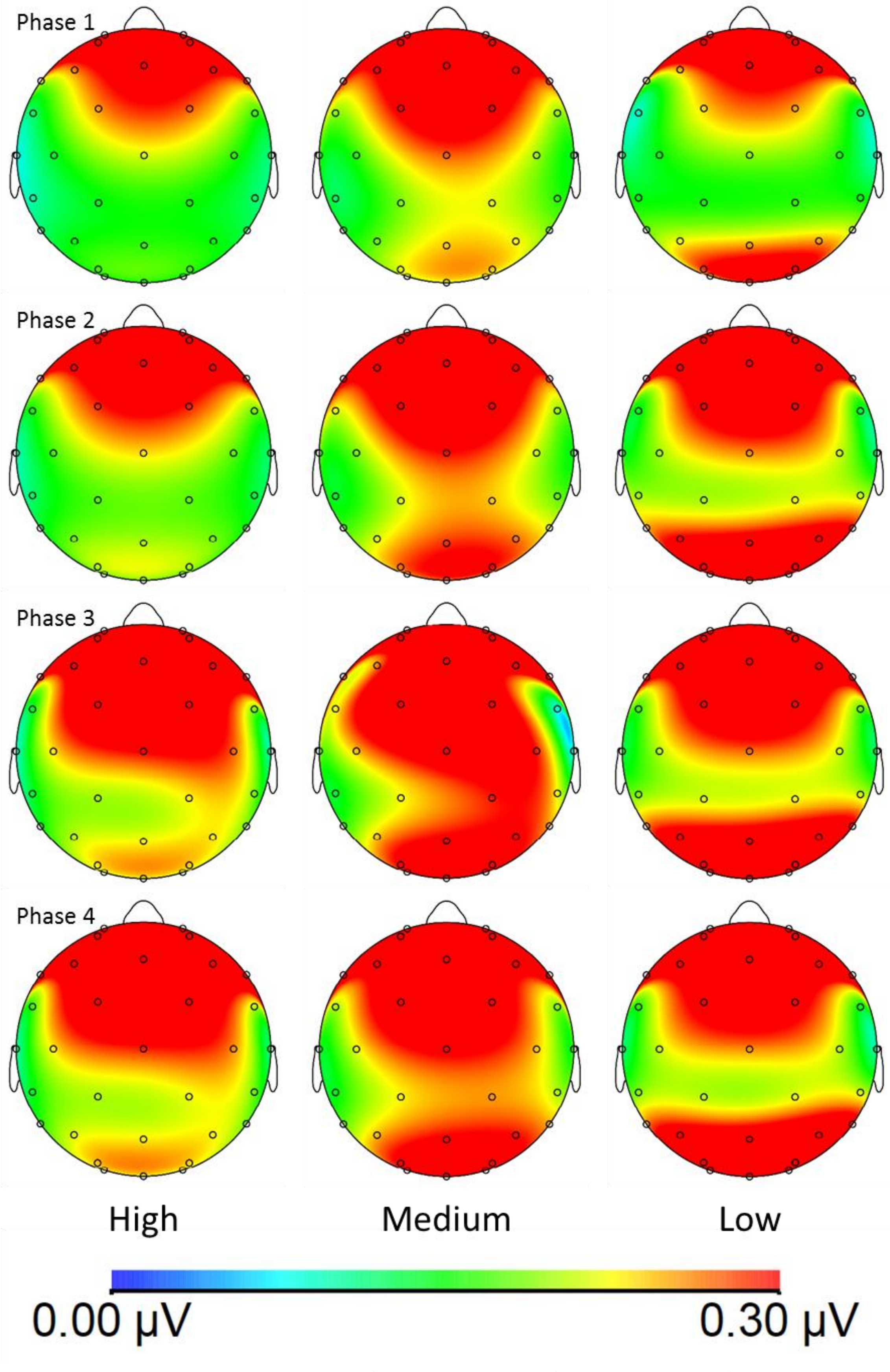

Figure 3: EEG Theta spectral power of the entire scalp for phase 1 to 4 from the top to the bottom respectively. From the left to the right, higher performers, medium performers and lower performers.

Alpha rhythm: A repeated measures Anova with the Cluster as a between subject factor and the 28 electrodes and the four phases as intra-participants variables revealed no significant main effect of the cluster for the alpha rhythm $F_{(2.162)} = 2.21$, $p = .140$, $\eta^2_{partial} = .027$. Following Howell (2007, pp. 366-367) commenting Wilcox (1987), further analysis were nevertheless performed similarly as with theta rhythm. Analyses of the alpha rhythm are illustrated in Figure 4. Considering cluster comparisons, the medium performer cluster exhibited higher alpha spectral power than the higher performer cluster during the first three phases though these differences were only marginally significant (phase 1: $F_{(1,16)} = 3.90$, $p = .07$; phase 2: $F_{(1,16)} = 3.70$, $p = .07$; phase 3: $F_{(1,16)} = 3.66$, $p = .07$). There was no significant difference at phase 4, $F_{(1,16)} = 2.92$, $p = .11$. Moreover, from a descriptive viewpoint, low performers exhibited higher levels of alpha spectral power than the medium and higher performers clusters during each phase. Thus, alpha spectral power tended to be higher for lower than for higher performers.

Considering phase comparisons, higher performers' alpha spectral power increased significantly between phase 1 and phase 2 $F_{(1,16)} = 10.20$, $p = .006$, between phase 2 and phase 3, $F_{(1,16)} = 14.73$, $p = .001$ but not between phase 3 and phase 4, $F_{(1,16)} = 0.20$, $p = .66$. The medium performers cluster presented the same pattern, with a marginally significant difference between phase 1 and phase 2 $F_{(1,16)} = 3.97$, $p = .06$, a significant increase between phase 2 and phase 3 $F_{(1,16)} = 5.89$, $p = .03$ and no difference between phase 3 and phase 4 $F_{(1,16)} = 0.54$, $p = .47$. Therefore, as for theta spectral power, alpha spectral power tended to increase from one to three concurrent tasks and to reach a plateau after the addition of the fourth task.

Based on the literature, five electrodes were gathered to examine parietal variations in the alpha band frequency. A repeated measures Anova, including the two main regions (frontal and parietal) and the cluster as a between subject factor revealed a significant difference following the phase ($F_{(3,51)} = 10.38$, $p < .001$, $\eta^2_{partial} = .379$) but not following the region ($F_{(1,17)} = 0.59$, $p = .455$, $\eta^2_{partial} = .033$) nor the cluster ($F_{(2,71)} = 1.86$, $p = .185$, $\eta^2_{partial} = .180$) though the interaction between region and cluster appeared marginally significant ($F_{(2,17)} = 3.16$, $p = .068$, $\eta^2_{partial} = .271$. The interaction between phase and cluster showed no significant effect ($F_{(6,51)} = 1.25$, $p = .398$, $\eta^2_{partial} = .128$).

Considering only the five electrodes (Pz, P3, P4, P7 and P8) in the parietal area showed an increase in alpha spectral power between phase 1 and phase 2 for both groups ($F_{(1,16)} = 9.84$, $p = .006$ and $F_{(1,16)} = 6.11$, $p = .02$ for higher and medium performers clusters respectively). Alpha spectral power also increased between phase 2 and phase 3 for both higher and medium performers, $F_{(1,16)} = 6.71$, $p = .02$ and $F_{(1,16)} = 6.59$, $p = .02$ respectively but not between phase 3 and phase 4, $F_{(1,16)} < 0.01$, $p = .97$ and $F_{(1,16)} = 0.99$, $p = .33$ respectively. Thus, the same pattern was observed as for the entire scalp.

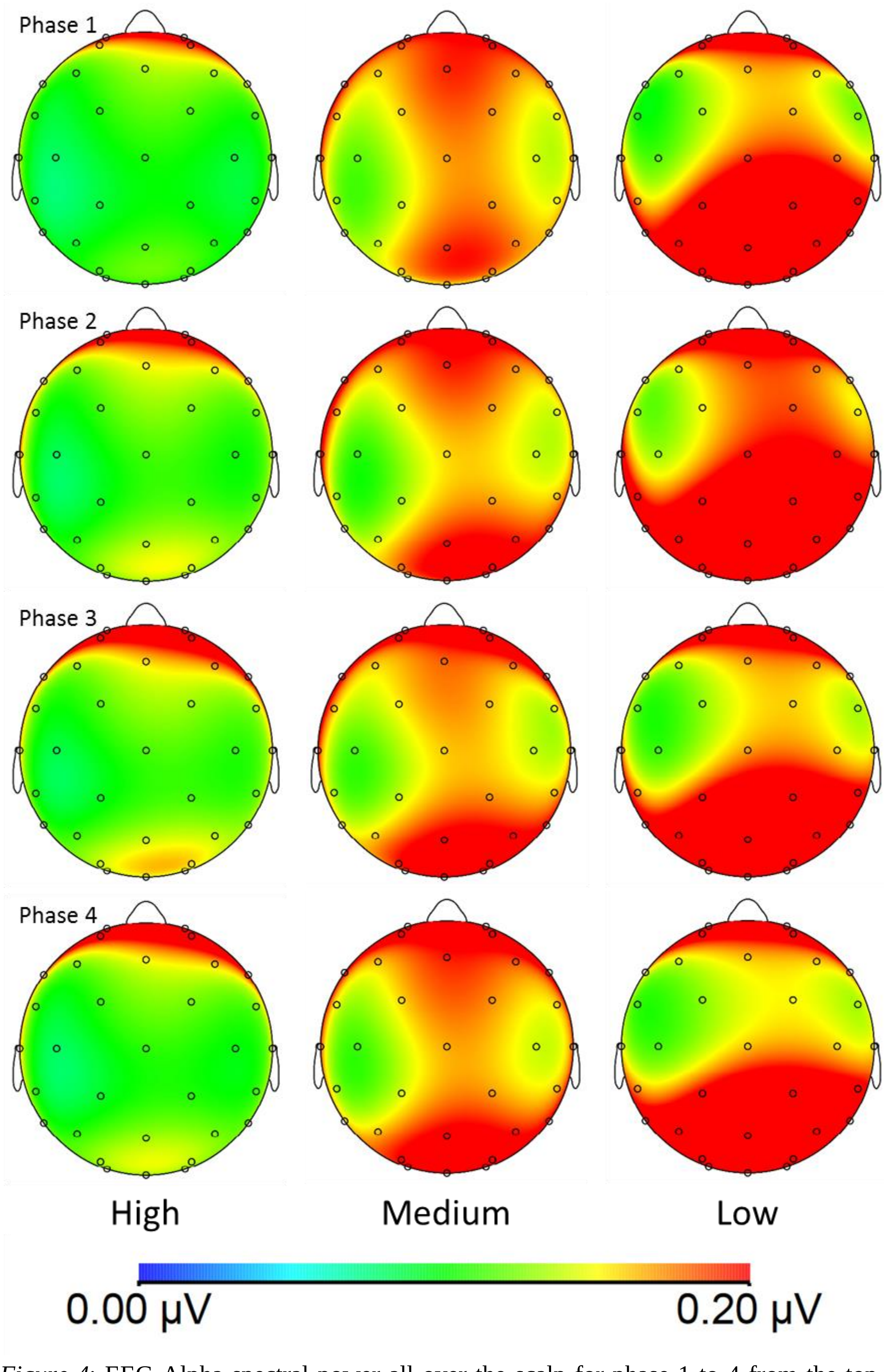

Figure 4: EEG Alpha spectral power all over the scalp for phase 1 to 4 from the top to the bottom respectively. From the left to the right, higher performers, medium performers and lower performers.

Discussion

Many authors consider that an increase in cognitive workload is associated with an increase in theta band frequency and a decrease in alpha band frequency (Vidulich & Tsang, 2012, for a review) during ecologically-valid tasks. Nevertheless, a deeper look at the empirical findings revealed that these differences were not systematically observed (Kramer, 1990; Sirevaag *et al.*, 1988; Natani & Gomer, 1981; Fournier, Wilson & Swain, 1999; Holm *et al.*, 2009; Funke *et al.*, 2013). However, on laboratory tasks more systematic variations on theta and alpha band frequencies characterized working memory and attention solicitation (Klimesch, 1999).

We hypothesized, based on previous studies on cognitive workload, firstly that some workload manipulations might not have been effective or might have reached a plateau. Secondly, there could have been individual differences in absolute spectral band frequencies relative to performance differences or in attention allocation and working memory solicitation strategies used to perform the task. This might have brought noise into the data and masked band frequency differences. During this experiment, (1) we recorded subjective ratings as well as pupil size in order to confirm the workload manipulation and (2) we collected data for a large sample of participants (20) in order to enable cluster analysis based on performances. Our participants had to manage an increasing number of concurrent subtasks in a complex multitasking environment used in pilot selection. Subjective ratings as well as pupil size variations increased with the addition of concurrent tasks from one to four. Thus, converging elements were in favor of an increase in cognitive demands imposed by the task across phases. Furthermore, behavioral results allowed distinguishing three different clusters of participants based on their performances at the fourth task phase.

Electroencephalographic data indicated an increase in the theta band in line with the number of concurrent tasks to perform until reaching a plateau for three and four concurrent tasks. This pattern was observed for medium and higher performers but not for low performers who appeared to reach a plateau from the beginning of the first task. This increase was consistent with previous studies indicating a link between theta spectral power and task demand (*e.g.*, Klimesch, 1999 for review) or with the number of tasks to be performed (Fournier, Wilson & Swain, 1999; Gevins & Smith, 2003; Holm *et al.*, 2009; see Borghini *et al.*, 2014 for a review) and while other studies failed to replicate this increase in theta spectral power according to the task demands, our results showed the expected pattern. Some studies failed to discriminate between one and two concurrent tasks (Holm *et al.*, 2009), while others succeeded in discriminating one versus two or one versus four simultaneous tasks, but failed to discriminate different levels of difficulty for the same task (Matthews *et al.*, 2015; Käthner *et al.*, 2014; Fournier, Wilson & Swain, 1999). However, using only five frontal electrodes did not reveal a difference between one and two concurrent tasks. This might be accounted for by the differences in the nature of the four tasks used in this experiment. This result was also consistent with the claim that working memory uses large scales networks distributed over frontal and parieto-occipital regions (Kahana, Seelig & Madsen, 2001; d'Esposito *et al.*, 1998). Moreover, the results of the lower performers indicated that with a group of participants who have difficulties with the task, one could find no differences in theta spectral power.

The plateau effect might reflect saturation in cognitive resource allocation. Although largely speculative, this interpretation could be supported by the behavioral performance differences, which were significant between the third and fourth phases. If all available cognitive resources were used to perform three concurrent tasks and maintain performance at an acceptable level (Wickens 2002), then adding a fourth would have detrimental effects on performance. The participants would no longer be able to compensate for the increasing cognitive load and maintain performance at the same level, leading to the observed performance diminution (Wickens, 2002; 2008). This interpretation of the observed plateau as reflecting a saturation of cognitive resources allocation is further supported by the significant increase in subjective ratings and decrease in performance. When comparing the clusters based on performance, significant differences appeared, higher theta spectral power being associated with lower performances. The medium performers exhibited higher theta power than the higher performers cluster at least until the third phase. No difference appeared between the two clusters during the fourth phase. From a descriptive viewpoint, the lower performers cluster exhibited higher theta spectral power during the whole experiment. This result was in line with previous findings suggesting that the power of theta rhythm could predict performance with a negative correlation: the higher the theta spectral power, the lower the performance (Beatty & O'Hanlon, 1979; Beatty, 1977, see also Kramer, 1990 and Klimesch, 1999 for reviews). However, our experimental protocol did not allow determining why this pattern was observed. For example, some participants might have better performed the task due to a larger amount of cognitive resources available or due to the use of more efficient strategies. More efficient strategies could have reduced the cost of task performance and therefore of working memory solicitation.

When considering the alpha band, results showed a similar pattern with an increase in alpha spectral power according to the number of concurrent tasks, similar to the effects observed with the theta rhythm. This increase was opposite previous findings, which usually note an alpha decrease with an increasing task demand (Klimesch, 1996; see Borghini *et al.*, 2014 and Vidulich & Tsang, 2012 for reviews). The broadly accepted assumption is that when attention demands increase, alpha spectral power decreases, mainly in the parietal and central areas. In this experiment, as the number of tasks increased, the attention demands also increased but the alpha power rose accordingly.

While other studies found an increased alpha level with an increased task difficulty (Pope, Bogart & Bartolome, 1995; Kamzanova, Kustubayeva & Matthews, 2014), another interpretation might arise from the work of Sauseng and his colleagues (2009). In their study, they found that the alpha spectral feature increased with an expanding number of irrelevant items to inhibit. This would be consistent with previous studies showing a disrupting effect of alpha rhythm on cognitive processing (see Foxe & Snyder, 2011, for a review). Indeed, an increase in occipito-parietal alpha band is assumed to reflect an inhibition of irrelevant processing. While some studies found a decrease in alpha spectral power according to task demand (Klimesch, 1997; Fairclough & Venables, 2006; Klimesch, 1999 for review) or the number of items held in working memory (Fukuda, Mance & Vogel, 2015), others found an increase in alpha band associated with irrelevant task inhibition (Rihs, Michel & Thut, 2007).

Both variations might reflect distinct mechanisms involved in working memory performance, an alpha band desynchronization underlying memory maintenance processes and an alpha band synchronization reflecting the inhibition of irrelevant items (Capilla *et al.*, 2014). Thus increases and decreases of alpha spectral power reflect distinct strategies and task demands which are often confounded when considering cognitive workload measurement. In this experiment, the increase in cognitive workload may be due to an increase in inhibition of irrelevant information and therefore might be associated with an increase in alpha spectral power. Finally, our results showed a consistent pattern of mean level of spectral power related to task performance: the more participants were proficient on the task, the lower level of spectral power, both for theta and alpha bands. Such results are consistent with previous findings on working memory solicitation and performances-based groups (Klimesch, Vogt & Doppelmayr, 1999; Klimesch, 1999). This highlights the importance of considering homogeneous groups of participants regarding task performance when assessing the impact of variables (such as cognitive load levels) on EEG spectral power bands.

In conclusion, our results indicated that EEG allows discriminating cognitive resource involvement with a widespread increase of theta rhythm and a more localized increase of alpha rhythm. Both rhythms increased with the number of tasks to perform concurrently until reaching a plateau with three and four simultaneous tasks. These differences are better explained using a laboratory theoretical framework of executive functions, emphasizing the need to bridge the gap between laboratory and ecological experiments. Our study showed differences in relation to the experienced cognitive workload. It also provided evidence that by controlling inter-individual differences and the absolute level of induced cognitive workload, efficient measurements can be derived.

References

Ahlstrom, U. & Friedman-Berg, F.J. (2006). Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics*, *36*, 623–636.

Antonenko, P. D. (2007). The effect of leads on cognitive load and learning in a conceptually rich hypertext environment. Pro-Quest.

Baldwin, C. L., & Penaranda, B. (2012). Adaptive training using an artificial neural network and EEG metrics for within-and cross-task workload classification. *NeuroImage*, 59, 48–56.

Basar, E., Basar-Eroglu, C., Karakas, S., & Schürmann, M. (2001). Gamma, alpha, delta, and theta oscillations govern cognitive processes. *International Journal of Psychophysiology*, 39, 241–248.

Beatty, J. (1977). Learned regulation of alpha and theta frequency activity in the human electroencephalogram. In Schwartz G. & Beatty J. (Eds.). *Biofeedback: Systems, processes, and research*. New York: Academic Press.

Beatty, J. & O'Hanlon, J. (1979). Operant control of posterior theta rhythms and vigilance performance: Repeated treatments and transfer of training. In Birbaumer N. & Kimmel K. (Eds.). *Biofeedback and self-regulation*. Hillsdale, NJ: Erlbaum.

Beatty, J. & Lucero-Wagoner, B. (2000). The pupillary system. In Cacioppo, J.T., Tassinary, L.G. & Berntson, G. (Eds). *Handbook of psycho-physiology* (pp. 142–162). Cambridge, MA: Cambridge University Press.

Berka, C., Levendowski D.J., Lumicao, M.N., Yau, A., Davis, G., Zivkovic, V.T., Olmstead, R.E., Tremoulet, P.D. & Craven, P.L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, Space and Environmental Medicine*, 78, B231-B244.

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 and Biobehavioral Reviews*, 44, 58-75.

Broadbent, D. (1958). *Communication and perception*. Pergamon, New York.

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.

Brouwer, A.M., Hogervorst, M.A., van Erp, J.B.F., Heffelaar, T., Zimmerman, P.H. & Oostenveld, R. (2012). Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of Neural Engineering*, 9, 045008.

Cain, B. (2007). *A review of the mental workload literature*. Defense Research and Development Toronto (Canada).

Capilla, A., Schoffelen, J.M., Paterson, G., Thut, G. & Gross, J. (2014). Dissociated alpha-band modulations in the dorsal and ventral visual pathways in visuospatial attention and perception. Cerebral Cortex, 24, 550-561.

Cegarra, J., & Chevalier, A. (2008). The use of Tholos software for combining measures of mental workload: Toward theoretical and methodological improvements. *Behavior Research Methods*, 40, 988–1000.

Chiew, K.S., & Braver, T.S. (2013). Temporal dynamics of motivation cognitive control interactions revealed by high-resolution pupillometry. *Frontiers in Psychology*, 4, 15.

Eggemeier, F.T., Wilson, G.F., Kramer, A.F. & Damos, D.L. (1991). Workload assessment. In *multi-task environments. Multiple task performance*. D.L. Damos. London, GB, Taylor & Francis, Ltd.: 207-216.

d'Esposito, M., Aguirre, G., Zarahn, E., Ballard, D., Shin, R., & Lease, J. (1998). Functional MRI studies of spatial and nonspatial working memory. *Cognitive Brain Research*, 7, 1–13.

Fairclough, S. H., & Venables, L. (2006). Prediction of subjective states from psychophysiology: A multivariate approach. *Biological psychology*, 71, 100–110.

Fairclough, S. H., Venables, L., & Tattersall, A. (2005). The influence of task demand and learning on the psychophysiological response. *International Journal of Psychophysiology*, 56, 171–184.

Fournier, L. R., Wilson, G. F., & Swain, C. R. (1999). Electrophysiological, behavioral, and subjective indexes of workload when performing multiple tasks: manipulations of task difficulty and training. *International Journal of Psychophysiology*, 31(2), 129–145.

Foxe, J.J. & Snyder, A.C. (2011). The Role of Alpha-Band Brain Oscillations as a Sensory Suppression Mechanism during Selective Attention. *Frontiers in Psychology*, 2, 1-13.

Fukuda, K., Mance, I. & Vogel, E.K. (2015). A Power Modulation and Event-Related Slow Wave Provide Dissociable Correlates of Visual Working Memory. Journal of Neuroscience, 35, 14009-14016.

Funke, G., Knott, B., Mancuso, V. F., Strang, A., Estepp, J., Menke, L., Miller, B. (2013). Evaluation of subjective and EEG based measures of mental workload. In HCI international 2013- posters' extended abstracts (pp. 412–416). Springer.

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

Gevins, A.S. & Smith, M.E. (2000). Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. *Cerebral Cortex,* 829-839.

Gevins, A.S., Smith, M. E., Leong, H., McEvoy, L., Whitfield, S., Du, R., & Rush, G. (1998). Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 40, 79–91.

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

Gevins, A. S., & Schaffer, R.E. (1980). A critical review of electroencephalographic (EEG) correlates of higher cortical functions. *Critical reviews in bioengineering*, 4, 113-164.

Gevins, A. S., Zeitlin, G. M., Doyle, J.C., Schaffer, R.E., & Callaway, E. (1979). EEG patterns during “cognitive” tasks. II. Analysis of controlled tasks. *Electroencephalography and Clinical Neurophysiology*, 47, 704–710.

Gevins, A. S., Zeitlin, G.M., Yingling, C.D., Doyle, J.C., Dedon, M.F., Schaffer, R.E., Yeager, C.L. (1979). EEG patterns during cognitive tasks. I. Methodology and analysis of complex behaviors. *Electroencephalography and clinical neurophysiology*, 47, 693–703.

Grimes, D., Tan, D.S., Hudson, S.E., Shenoy, P. & Rao, R.P. (2008) Feasibility and pragmatics of classifying working memory load with an electroencephalograph. Proceedings of the 26th Annual SIGCHI Conference on Human Factors in Computing Systems, ACM (Florence, Italy), 835-844.

Gulbinaite, R., Johnson, A., de Jong, R., Morey, C.C. & van Rijn, H. (2014). Dissociable mechanisms underlying individual differences in visual working memory capacity. *NeuroImage*, 99, 197-206.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Advances in psychology*, 52, 139–183.

Hess, E.H. & Polt, J.M. (1964). Pupil Size in Relation to Mental Activity during Simple Problem-Solving. *Science*, 143, 1190-1192.

Hollands, J. G., & Wickens, C. D. (2000). *Engineering psychology and human performance*. Prentice Hall.

Holm, A., Lukander, K., Korpela, J., Sallinen, M., & Müller, K. M. (2009). Estimating brain load from the EEG. *The Scientific World Journal*, 9, 639–651.

Howell, D.C. (2007). *Statistical methods for psychology. 7th edition*. Boston, MA: Cengage Learning.

Iqbal, S.T., Adamczyk, P.D., Zheng, S.X., & Bailey, B.P. (2005). Towards an index of opportunity: Understanding changes in mental workload during task execution. In Veer G.v.d. & Gale C. (Eds.). CHI 2005 Proceedings (311–320). New York, NY: ACM Press

Johannsen, G. (1979). Workload and workload measurement. In *Mental workload* (pp. 3–11). Springer.

Just, M.A. & Carpenter, P.A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology*, 47, 310-339.

Kahana, M. J., Seelig, D., & Madsen, J. R. (2001). Theta returns. *Current opinion in neurobiology*, 11, 739–744.

Kahneman, D., Tursk, B., Shapiro, D. & Crider, A. (1969). Pupillary, heart rate and skin resistance changes during a mental task. *Journal of Experimental Psychology*, 79, 164–167.

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

Kamzanova, A.T., Kustubayeva, A.M. & Matthews, G. (2014). Use of EEG workload indices for diagnostic monitoring of vigilance decrement. *Human Factors*, 56, 1136-1149.

Käthner, I., Wriessnegger, S. C., Müller-Putz, G. R., Kübler, A., & Halder, S. (2014). Effects of mental workload and fatigue on the p300, alpha and theta band power during operation of an ERP (p300) brain–computer interface. *Biological psychology*, 102, 118–129.

Ke, Y., Qi, H., He, F., Liu, S., Zhao, X., Zhou, P., Ming, D. (2014). An EEG-based mental workload estimator trained on working memory task can work well under simulated multi-attribute task. *Frontiers in human neuroscience*, 8.

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain research reviews*, 29, 169–195.

Klimesch, W. (1997). EEG-alpha rhythms and memory processes. *International Journal of Psychophysiology*, 26, 319-340.

Klimesch, W. (1996). Memory processes, brain oscillations and EEG synchronization. *International Journal of Psychophysiology*, 24, 61-100.

Klimesch, W., Vogt, F. & Doppelmayr, M. (1999). Inter-individual Differences in Alpha and Theta Power Reflect Memory Performance. *Intelligence*, 27, 347-362.

Klimesch, W., Doppelmayr, M., Russeger, H., Pachinger, T. & Schwaiger, J. (1998). Induced alpha band power changes in the human EEG and attention. *Neuroscience Letters*, 244, 73-76.

Kramer, A. F. (1990). Physiological metrics of mental workload: A review of recent progress (Tech. Rep.). DTIC Document.

LaBar, K. S., Gitelman, D. R., Parrish, T. B., & Mesulam, M.-M. (1999). Neuroanatomic overlap of working memory and spatial attention networks: a functional MRI comparison within subjects. *NeuroImage*, 10, 695–704.

Laeng, B., Ørbo, M., Holmlund, T. & Miozzo, M. (2011). Pupillary Stroop Effects. *Cognitive Processing*, 12, 13-21.

Lisi, M., Bonato, M. & Zorzi, M. (2015). Pupil dilation reveals top–down attentional load during spatial monitoring. *Biological Psychology*, 112, 39-45.

Matthews, G., Reinerman-Jones, L.E., Barber, D.J. & Abich, J. (2015). The Psychometrics of Mental Workload: Multiple Measures Are Sensitive but Divergent. *Human Factors*, 57, 125-143.

McEvoy, L.K., Pellouchoud, E., Smith, M.E. & Gevins, A. (2001). Neurophysiological signals of working memory in normal aging. *Cognitive Brain Research*, 11, 363–376.

Moray, N. (1979). *Mental workload: Its theory and measurement* (Vol. 8). Plenum Publishing Corporation.

Natani, K., & Gomer, F. E. (1981). Electrocortical activity and operator workload: A comparison of changes in the electroencephalogram and in event-related potentials (Tech. Rep.). DTIC Document.

Ota, T., Toyoshima, R., & Yamauchi, T. (1996). Measurements by biphasic changes of the alpha band amplitude as indicators of arousal level. *International journal of psychophysiology*, 24, 25–37.

Patten, C. J., Kircher, A., Östlund, J., Nilsson, L., & Svenson, O. (2006). Driver experience and cognitive workload in different traffic environments. *Accident Analysis & Prevention*, 38, 887–894.

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

Pigeau, R., Hoffman, R., Purcell, S. & Moffitt, A. (1988). The Effect of Endogenous Alpha on Hemispheric Asymmetries and the Relationship of Frontal Theta to Sustained Attention. In Jessen K. (Ed.), *Electric and magnetic activity of the central nervous system: Research and clinical applications in aerospace medicine.* France: NATO Advisory Group for Aerospace Research and Development.

Pope, A.T., Bogart, E.H. & Bartolome, B.S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology*, 40, 187-195.

Ravizza, S. M., Behrmann, M., & Fiez, J. A. (2005). Right parietal contributions to verbal working memory: spatial or executive? *Neuropsychologia*, 43, 2057–2067.

Richer, F. & Beatty, J. (1985). Pupillary Dilations in Movement Preparation and Execution. *Psychophysiology*, 22, 204-207.

Rihs, T.A., Michel, C.M. & Thut, G. (2007). Mechanisms of selective inhibition in visual spatial attention are indexed by alpha-band EEG synchronization. *European Journal of Neuroscience*, 25, 603-610.

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

Ryu, K., & Myung, R. (2005). Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. *International Journal of Industrial Ergonomics*, 35, 991–1009.

Sauseng, P., Klimesch, W., Schabus, M., & Doppelmayr, M. (2005). Fronto-parietal EEG coherence in theta and upper alpha reflect central executive functions of working memory. *International Journal of Psychophysiology*, 57, 97–103.

Sauseng, P., Klimesch, W., Heise, K., Gruber, W., Holz, E., Karim, A., Glennon, M., Gerloff, C., Birbaumer, N., & Hummel, F. (2009). Brain Oscillatory Substrates of Visual Short-Term Memory Capacity. *Current Biology*, *19*, 1846-1852.

Sauseng, P., Griesmayr, B., Freunberger, R. & Klimesch, W. (2010). Control mechanisms in working memory: A possible function of EEG theta oscillations. *Neuroscience and Biobehavioral Reviews*, 34, 1015-1022.

Schacter, D.L. (1977). EEG theta waves and psychological phenomena: A review and analysis. *Biological Psychology*, 5, 47–82.

Schvaneveldt, R. W. (1969). Effects of complexity in simultaneous reaction time tasks. *Journal of Experimental Psychology*, 81, 289.

Sirevaag, E., Kramer, A., DeJong, R. &Mecklinger, A. (1988). A psychophysiological analysis of multi-task processing demands. In *Psychophysiology* (Vol. 25, pp. 482–482).

Slobounov, S., Fukada, K., Simon, R., Rearick, M., & Ray, W. (2000). Neurophysiological and behavioral indices of time pressure effects on visuomotor task performance. *Cognitive Brain Research*, 9, 287–298.

Smith, M. E., Gevins, A., Brown, H., Karnik, A., & Du, R. (2001). Monitoring task loading with multivariate EEG measures during complex forms of human-computer interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 43, 366–380.

Smith, M. E., McEvoy, L. K., & Gevins, A. (1999). Neurophysiological indices of strategy development and skill acquisition. *Cognitive Brain Research*, 7, 389–404.

Stein, E. (1992). *Air traffic control visual scanning* (No. DOT/FAA/CT-TN92/16). Atlantic City International Airport: FAA Technical Center.

Sterman, M., & Mann, C. (1995). Concepts and applications of EEG analysis in aviation performance evaluation. *Biological psychology*, 40, 115–130.

Tricot, A., & Chanquoy, L. (1996). La charge cognitive, " vertu dormitive " ou concept opérationnel ? Introduction. *Psychologie Française, 41*, 313-318.

Valentino, D. A., Arruda, J., & Gold, S. (1993). Comparison of qEEG and response accuracy in good vs. poorer performers during a vigilance task. *International Journal of Psychophysiology*, 15, 123–133.

Vidulich, M. A., & Tsang, P. S. (2012). Mental workload and situation awareness. In *Handbook of human factors and ergonomics* (pp. 243–273). John Wiley & Sons, Inc. Retrieved from http://dx.doi.org/10.1002/9781118131350.ch8

Wickens, C.D. (2008). Multiple Resources and Mental Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society,* 50, 449-455.

Wickens, C.D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Science*, 3, 159-177.

Wickens, C.D. (1991). Processing Resources and Attention. In Damos D.L. (Ed). Multiple Task Performance (3-34). London, Washington DC: Taylor & Francis.

Wickens, C.D., Kramer, A.F., Vanasse, L. & Donchin, E. (1983). Performance of concurrent tasks: A psychophysiological analysis of the reciprocity of information-processing resources. *Science*, 221, 1080-1082.

Wilcox, R. R. (1987). *New statistical procedures for the social sciences*. Hillsdale, NJ: Erlbaum.

Xie, B., & Salvendy, G. (2000). Prediction of mental workload in single and multiple tasks environments. *International Journal of Cognitive Ergonomics*, 4, 213–242.

Yamada, F. (1998). Frontal midline theta rhythm and eye blinking activity during a VDT task and a video game: useful tools for psychophysiology in ergonomics. *Ergonomics*, 41, 678–688.

Yeh, Y.-Y., & Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. Human Factors: The Journal of the Human Factors and Ergonomics Society, 30(1), 111–120

Zhao, C., Zhao, M., Liu, J. & Zheng, C. (2012). Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator. *Accident Analysis and Prevention,* 45, 83-90.