Cognitive State Assessment: 
Examination of EEG-based Measures on a Stroop Task

Lee W. Sciarini¹, Jefferson D. Grubb² and Philip G. Fatolitis³  
¹Naval Postgraduate School  
²Naval Air Systems Command  
³Naval School of Aviation Safety

This effort investigated the ability of a neurophysiological measure to detect changes in workload during a task which is sensitive to cognitive function. A growing collection of research suggests that physiological measures such as EEG can be used to inform the adaptation of systems. However, it has been proposed that such measures often provide a gross interpretation of cognitive workload during complex tasks and are not sensitive to differences in specific cognitive function. To understand the utility of neurophysiological measures for human-machine interaction, we must know if these measures are sensitive to tasks which are sensitive to changes in cognitive function. To begin to answer this question, we investigated the sensitivity of Advanced Brain Monitoring’s EEG-based measures to changes in workload experienced during a Stroop task. Results indicated that ABM’s workload measure can detect changes associated with the attentional demands and cognitive processes linked to the ability to inhibit word naming during tasks involving semantic interference. This indicates that changes in workload associated with the ability to inhibit competing cognitive processes can be identified using neurophysiological workload measures.

INTRODUCTION

From initial use of stone tools over two million years ago, humankind has pursued the development and application of technology (e.g., machines, automation, robots and other technically complex tools) to assist in execution of tasks which were found to be difficult, repetitive, high-risk, dangerous, or even inconvenient. As the eons progressed, technology has evolved in complexity to the point where systems have outpace our ability to take full advantage of their capabilities and could potentially seem too complex to even employ by intended users. All too often, while utilizing technologies, operators are required to assess the current state of a system and operating environment while quickly determining an appropriate course of action. As a result, we create technologies (automation) to assist us in the use of these complex systems. While the utilization of technology can increase the likelihood of successful operations, performance can be adversely impacted by fluctuations in and mismanagement of operator workload. When an operator is paired with technology to accomplish a task, the resulting human-system predominantly relies on a one-way state assessment relationship in which the human is ultimately responsible for maintaining homeostasis. This relationship does have the potential to become more symbiotic. Through the integration of relatively noninvasive neurophysiological measures, it should be possible to create a system that can detect and utilize the cognitive state of its operator and literally result in a truly “human-in-the-loop” system (Sciarini & Nicholson, 2009).

As astutely noted by Reeves, Stanney, Axelsson, Young and Schmorrow (2007) there are several impediments to the application of neurophysiological technologies including (1) the need for valid, reliable, and generalizable cognitive state gauges based on basic neurophysiological sensors; (2) real-time cognitive state classification based on basic cognitive psychology science and applied neurocognitive engineering; and (3) proof of effectiveness which demonstrates generalizable application of mitigations (i.e., the ability to control how/when mitigations are applied). It is important to note that there are exceptional examples of physiological measures being used to achieve system adaptation that predate the 2007 objectives proposed by Reeves et al. (see: Pope, Bogart & Bartolome, 1995; Prinzel et al., 2003; Wilson & Russell, 2003; 2004; 2006; Russell, 2005) and to date, a majority of investigations have examined cognitive state via minimally invasive neurophysiological sensing technologies from a nomothetic perspective. However, it has been suggested that to achieve the goal set forth by Reeves et al. an idiographic approach must be considered and that the identification of neurophysiological measures’ capacity to detect decrements (either alone or in combination with others) in specific cognitive function (e.g., visual, auditory, spatial, etc.) must be empirically demonstrated. Once reliability has been established, candidate measure(s) could then be integrated into a generalizable cognitive state assessment capability and examined in applied domains (Sciarini, 2009).

Central to a discussion of measuring an individual’s cognitive ability, the role of the attention must be considered. Three general categories of attention theories can be found in the literature, including: “cause” theories, in which attention is suggested to modulate information processing (e.g., via a spotlight that functions as a serial scanning mechanism or via limited resource pools); “effect” theories, in which attention is suggested to be a by-product of information processing among multiple systems (e.g., stimulus representations compete for neuronal activation); and hybrids that combine cause and effect theories (Fernandez-Duque & Johnson, 2002). In general, attention is suggested to be a selective process via which stimulus representations are transferred between sensory memory and working memory (WM) and then contributes to the processing of information once in working memory. Attention improves human performance on a wide range of tasks, minimizes distractions, and facilitates access to
awareness (i.e., focused attention). In the best case, attention helps to filter out irrelevant multimodal stimuli. In the worst case, critical information is lost due to overload of incoming information, stimulus competition, or distractions.

Attention

In applied settings, human attention may be defined as the allocation of cognitive resources for task completion. Attentional modulation in real-world settings can be, in part, attributable to changes in workload demands or multitasking (Wickens, 2008). The ability to focus or control attention is an important executive function. Driven by a specific goal, the allocation of attention toward the processing of task-related information and responses is controlled internally. This is evident when competing signals are encountered and conflict arises between the processing of task-relevant information and distractors. Effective attentional control is dependent upon self-monitoring and whether or not goals are met (Luks et al., 2010). A specific subset of this issue involves the effect that interference can have on attention-related performance. For example, in conditions where different attributes of a single stimulus are represented by different sources of information, conflict can occur with a resulting deficit in attentional performance (Wickens & Hollands, 2000), as observed in the Stroop effect (Stroop, 1935).

The Stroop effect is elicited in experiments by manipulating the text of the name of a color, for example, “brown.” The stimulus is manipulated by presenting the text in the same color or in a different color than brown so that there is either congruence or incongruence between text and the color in which it is presented. When there is congruence, reaction time in appropriate identification of the stimulus is shorter. When there is incongruence, reaction time is longer. It has been postulated that longer reaction times for incongruent tasks are caused by a disruption in attention allocation due to competition of potential responses. This is thought to be the result of one response having task relevance while the other has bottom-up saliency which strongly influences the allocation of attention (MacDonald et al., 2000).

According to Wickens and Hollands (2000), the Stroop effect shows that multiple features of a single object are likely to be processed in parallel and performance is facilitated if parallel processing is required, but performance is disrupted if a feature is incongruent. They elaborate that response conflict and redundancy gain are constructs that can help to explain attentional processes in such conditions. Response conflict is a failure of focused attention where there is competition between different response possibilities (Eriksen & Eriksen, 1974). Redundancy gain is identified by an optimization of performance as a function of different stimuli that provide corroborating information. Both response conflict and redundancy gain can be observed where operators are required to use a display, such as can be found in an aircraft cockpit or in a command and control center, and the effect that these constructs can have on performance can be influenced by things such as display clutter (Wickens and Hollands, 2000) and executive control. Executive control involves information management functions and is characterized by such things as the application of rules by which a stimulus is processed, making speed-accuracy tradeoff determinations and managing working memory functions. When incongruent stimuli are presented, executive control is disrupted and can result in response conflict-related performance decrements. The converse is true of redundancy gain with respect to improvements in performance.

It is widely acknowledged that overtasking of cognitive processes (high workload) is related to performance decrements. Further, considering the interrelationships among specific kinds of cognitive activity (such as attention) and performance, it follows that indicators of the effort exerted during tasks requiring executive function (such as attentional control) can be observed using physiological measures and this knowledge can be exploited as a means to improve human-machine interaction.

Electroencephalography derived measures

In 2007, Berka et al. validated the use of EEG for measuring task engagement and mental workload by using baseline data as a model for creating a task engagement index with four levels: high engagement, low engagement and relaxed wakefulness and sleep onset. The first three classifications were created using absolute and relative power spectra variables from a database of over 100 individuals who were either sleep-deprived or fully rested. The fourth classification, sleep onset was derived from the baseline conditions through use of regression equations based on the absolute and relative power of the participant’s own baseline conditions (Berka et al., 2007). In order to examine mental workload, the researchers developed a mental workload index. These measures were derived similarly to the task engagement-index described above and are divided into two classifications: low workload and high workload. Their investigation which used the task engagement and mental workload measures revealed that participants’ EEG-workload index increased on tasks with increasing difficulty and working memory load. Similarly, EEG-engagement was shown to be related to the processes required for completing vigilance tasks (Berka et al., 2007).

This study explored if ABM’s workload (ABM-W) and engagement (ABM-E) measures are sensitive to changes in cognitive effort resultant of the allocation of attentional control while completing the Stroop task.

METHOD

A total of 20 adults (10 Male and 10 Female) with a mean age of 21.6 (18-43) participated in this study. Participants were screened for neurological disorders, spoke English as their primary language, had normal color vision and were all right handed. Participants received $20.00 compensation.

Materials

Biographical data form. All participants completed a biographical data form consisting of information relating to age, gender, handedness, level of education, primary language,
secondary language, and level of comfort working with computers.

*Electroencephalogram.* Advanced Brain Monitoring’s X-10, 9-channel EEG system was used for this investigation. The X-10 records from sensor sites positioned in accordance with the 10-20 system shown in Figure 1.

![Figure 1. EEG Sensor locations (Advanced Brain Monitoring, 2013)](image)

*Software.* ABM’s B-Alert software suite was used to ensure the EEG was recording within prescribed impedance tolerances, provide three baseline cognitive assessment tasks, acquire raw data, perform decontamination, and to provide ABM’s cognitive state and workload metrics.

The Multi-Media Presenter software suite (Barber, Lackey, Reinerman-Jones, 2012a) was used to randomly present Stroop Task images for each condition (no conflict, congruent and incongruent) and record participant responses, reaction times, and computer clock times of the onset of stimulus and the participant reaction.

The PhysioSync Multi-Sensor software suite (Barber, Lackey, Reinerman-Jones, 2012b) was used to synchronize data collected from ABM’s software with the stimulus and response data recorded by the Multi-Media Presenter software.

Subjective workload measure. A uni-dimensional subjective workload rating scale adapted from the Bedford Workload Scale (Roscoe, 1984; 1987; 1990) was used to record subjective workload from participants after each Stroop Task was completed.

**PROCEDURE**

Upon arrival to the lab, participants were asked to read and sign an informed consent form and then completed a biographical data form and randomly assigned to one of two counter balanced (word naming group first or color naming group first) presentation orders. Regardless of presentation order, each task group consisted of six trials (Figure 2): no conflict (NC) word, congruent (CG) word, incongruent (IN) word, NC color, CG color, and IN color. In the word and color NC conditions, the words blue, red, green, brown, and grey were respectively printed in black ink or presented as a solid rectangle of that color. Similarly, in the CG conditions, those colors were printed in their respective ink color or presented as a rectangle of that color. In the IN conditions, those colors were printed in each of the other colors ink (e.g. the word “blue” was printed in green, red, brown and grey ink) which created 20 unique images for use in each IN condition.

![Figure 2. Example of Stroop presentations. Top row: NC word, CG word and IN word; bottom row: NC color, CG color and IN color](image)

Participants were situated in a comfortable chair at the experiment station, outfitted with the ABM X-10 EEG system, and then completed a battery of baseline tasks consisting of a three choice vigilance task, a visual stimulus-response task and an eyes closed, auditory stimulus-response task.

Once baseline procedures were completed, participants completed the six Stroop tasks in their assigned presentation order. Each trial was preceded by a short training event to ensure participants understood each task. In each task group, 20 test images were randomly presented for 2000ms with an interstimulus interval of 2000ms. Participants completed the subjective workload assessment at the conclusion of each task.

**RESULTS**

*Performance*

To confirm that the Stroop effect was being observed, a repeated measures analysis of variance (ANOVA) was conducted on the reaction times in the NC, CG and IN for each presentation (word and color) condition. A main effect was observed for both the word presentation \(F(2,18) = 16.75, p < .0005, \eta^2 = .65\) and the color presentation \(F(2,18) = 17.857, p < .0005, \eta^2 = .67\). Paired sample t-tests were used to make post hoc comparisons between conditions for each presentation type to determine differences in reaction time. Table 1 shows results of the word condition with NC \((M = .85, SD = .14), CG (M = .75, SD = .19)\) and IN \((M = 1.01, SD = .21)\) and Table 2 displays for the color condition NC \((M = .81, SD = .12), CG (M = .80, SD = .12)\) and IN \((M = 1.01, SD = .20)\).

Accuracy in naming was consistent across all conditions with no significant differences detected.

![Table 1. Paired Samples t-test Reaction Time – Word Condition](image)

<table>
<thead>
<tr>
<th>Word Condition</th>
<th>(t)</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Conflict – Congruent</td>
<td>2.151</td>
<td>0.045</td>
</tr>
<tr>
<td>No Conflict – Incongruent</td>
<td>-3.560</td>
<td>0.002</td>
</tr>
<tr>
<td>Congruent – Incongruent</td>
<td>-5.915</td>
<td>0.000</td>
</tr>
</tbody>
</table>

![Table 2. Paired Samples t-test Reaction Time – Color Condition](image)

<table>
<thead>
<tr>
<th>Color Condition</th>
<th>(t)</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Conflict – Congruent</td>
<td>0.597</td>
<td>0.557</td>
</tr>
<tr>
<td>No Conflict – Incongruent</td>
<td>-4.753</td>
<td>0.000</td>
</tr>
<tr>
<td>Congruent – Incongruent</td>
<td>-5.840</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Subjective Workload

Repeated measures ANOVA was used to determine if there was a difference in perceived workload experienced by participants for each presentation condition. A significant effect was shown for both the word \(F(2,18) = 14.17, p < .0005\), partial \(\eta^2 = .61\) and the color presentation \(F(2,18) = 16.42, p < .0005\), partial \(\eta^2 = .65\).

As with performance, paired sample t-tests were used to examine differences in perceived workload for each presentation type. Results for the word condition with NC \((M = 2.55, SD = .95)\), CG \((M = 2.20, SD = .89)\) and IN \((M = 3.95, SD = 1.40)\) are shown in Table 3 and the results of the color condition for NC \((M = 2.30, SD = 1.26)\), CG \((M = 2.10, SD = 1.12)\) and IN \((M = 3.95, SD = 1.64)\) are reported in Tables 4.

Table 3. Paired Samples t-test Perceived Workload – Word Condition

<table>
<thead>
<tr>
<th></th>
<th>t</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Conflict – Congruent</td>
<td>1.926</td>
<td>0.069</td>
</tr>
<tr>
<td>No Conflict – Incongruent</td>
<td>-4.172</td>
<td>0.001</td>
</tr>
<tr>
<td>Congruent – Incongruent</td>
<td>-5.411</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table 4. Paired Samples t-test Perceived Workload – Color Condition

<table>
<thead>
<tr>
<th></th>
<th>t</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Conflict – Congruent</td>
<td>0.890</td>
<td>0.385</td>
</tr>
<tr>
<td>No Conflict – Incongruent</td>
<td>-5.320</td>
<td>0.000</td>
</tr>
<tr>
<td>Congruent – Incongruent</td>
<td>-5.663</td>
<td>0.000</td>
</tr>
</tbody>
</table>

EEG

A significant main effect was shown for AMB-W in the word and color presentation conditions \(F(2,18) = 8.031, p = .003\), partial \(\eta^2 = .47\) and \(F(2,18) = 6.939, p = .006\), partial \(\eta^2 = .30\) respectively. Significant effects were not observed for the ABM-E measure for either the word or color presentation conditions.

In order to determine the differences in AMB-W between presentation conditions, paired sample t-tests were conducted. Significant differences were found between the IN and both the NC and CG groups for each presentation condition. For the word condition, NC \((M = .55, SD = .08)\), CG \((M = .56, SD = .08)\) and IN \((M = .62, SD = .08)\) are shown in Table 5 and the results shown in Table 6 is for the color condition for NC \((M = .57, SD = .09)\), CG \((M = .55, SD = .10)\) and IN \((M = .62, SD = .09)\).

Table 5. Paired Samples t-test AMB Workload – Word Condition

<table>
<thead>
<tr>
<th></th>
<th>t</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Conflict – Congruent</td>
<td>2.030</td>
<td>0.057</td>
</tr>
<tr>
<td>No Conflict – Incongruent</td>
<td>-2.643</td>
<td>0.016</td>
</tr>
<tr>
<td>Congruent – Incongruent</td>
<td>-3.653</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Table 6. Paired Samples t-test AMB Workload – Color Condition

<table>
<thead>
<tr>
<th></th>
<th>t</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Conflict – Congruent</td>
<td>-3.300</td>
<td>0.778</td>
</tr>
<tr>
<td>No Conflict – Incongruent</td>
<td>-3.954</td>
<td>0.001</td>
</tr>
<tr>
<td>Congruent – Incongruent</td>
<td>-3.351</td>
<td>0.003</td>
</tr>
</tbody>
</table>

DISCUSSION

This study explored the sensitivity ability of ABM’s workload and engagement measures to changes in cognitive effort which resulted from the allocation of attentional control while completing the Stroop task. Performance and subjective workload results confirmed that the current incarnation of the Stroop task resulted in reaction response times and perceived workload differences as expected. It is important to note that the lack of difference between NC and CG in the color condition was predictable as identical stimuli were used for each presentation. The lack of difference in perceived workload between the NC and CG in the word condition is likely attributable to a lack of discriminant diagnostic sensitivity of the modified Bedford Scale on Stroop performance. In the color conditions, this result holds no meaning as they were identical presentations.

It was expected that the ABM measures would be sensitive to semantic interference (naming the font color of a word when the word is itself is another color name) and semantic facilitation (when the word matches the font color). Results presented here suggest that the EEG measure of ABM-W is only sensitive to changes associated with the attentional demands and cognitive processes linked to the ability to inhibit the naming of the word during tasks involving semantic interference (NC or CG vs. IN).

As discussed in the opening section of this paper, an examination of the utility of neurophysiological measures at an idiosyncratic level is needed in order to obtain the goals set forth by Reeves, Stanney, Axelsson, Young and Schmorrow (2007). The results presented here suggest that the overtasking of cognitive processes related to executive function can in fact be partially measured through neurophysiological measures. While an important achievement, the capability of detecting changes via brain based measures is just one step towards reliable neuroadaptive systems (Grubb & Chon, 2011). These results suggest that ABM-W can be used as an unobtrusive measure which is sensitive to changes in participants’ workload when confronted with a response conflict task.

Considering that the overtasking of cognitive processes has been demonstrated to result in performance decrements and that there exists interrelationships between cognitive activity and performance, it follows that the ABM-W measure shows promise as a surrogate measure for identifying the inhibitive activity required for the allocation of attention when adjudicating competing processes and that this measure can be exploited as a means to improve human-machine interaction through applications such as adaptive automation.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Daniel Barber at UCF’s ACTIVE Lab for his support with the PhysioSync software and the service members and civilians at the Naval Air Warfare Center-Training Systems Division. This effort was funded by Naval Air Systems Command Basic and Applied Research Program. The views expressed in this paper are solely those of the authors and not necessarily those of the
United States Navy, Marine Corps or the Department of Defense.

REFERENCES


