



A novel approach to the physiological measurement of mental workload

Matthew W. Miller^{a,1}, Jeremy C. Rietschel^{a,1}, Craig G. McDonald^b, Bradley D. Hatfield^{a,*}

^a Neuroscience and Cognitive Science Program, Department of Kinesiology, University of Maryland, College Park, MD, USA

^b Department of Psychology, George Mason University, Fairfax, VA, USA

ARTICLE INFO

Article history:

Received 19 October 2010

Received in revised form 14 January 2011

Accepted 4 February 2011

Available online 20 February 2011

Keywords:

Mental Workload

Attention

Event-Related Potential

Performance

ABSTRACT

While performing a visuo-motor task under incrementally-varied levels of difficulty, individuals were probed with a variety of novel, task-irrelevant, auditory stimuli. To determine the effect of task load on cerebral-cortical processing of these stimuli, event-related potentials were recorded while participants performed the task. We found that N1, P2, P3 and late positive potential (LPP) component amplitudes were inversely related to task-difficulty. This suggests that a variant of the oddball paradigm – in which the stimulus stream comprises novel sounds – is capable of providing a reliable index of mental workload.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

The efficient allocation of neural resources is crucial while individuals perform mentally demanding tasks. Such a need is based on the inverse relationship between mental workload and attentional reserve (Wickens et al., 1983), which when depleted can be expected to limit cognitive processing for any additional demands, resulting in performance decrement. Thus, the manner by which neural resources are allocated during the performance of a task is a question of fundamental interest. Such an understanding is dependent upon the attainment of an objective measure of mental workload.

Beginning with the seminal work by Wickens et al. (1977), a number of studies have employed the event-related potential (ERP) technique to assess mental workload. These early efforts (e.g., Isreal et al., 1980a; Isreal et al., 1980b; Kramer et al., 1987; Sirevaag et al., 1989; Wickens et al., 1983) employed dual-task paradigms in which ERPs to a secondary task were measured while participants performed a primary task of interest (e.g., attending to auditory stimuli while solving arithmetic problems as the primary task). Many of these studies revealed an inverse relationship between amplitude of the parietal-P3 component and difficulty of the primary task. However, such dual-task paradigms are not optimal for assessing the mental workload required by a primary task given that the addition of a secondary task may inherently change the primary task, thus compromising the assessment of the demand it imposes (Kramer et al., 1985).

In order to avoid the limitation of dual-task paradigms, ERPs to task-irrelevant stimuli should be measured while participants focus exclusively on a given task (Papanicolaou and Johnstone, 1984). Several studies have employed such an irrelevant-probe technique and reported decreases in N1 and/or P3 component amplitudes with changes in task difficulty (Bauer et al., 1987; Kramer et al., 1995; Sirevaag et al., 1993; Ullsperger et al., 2001; Wilson and McCloskey, 1988). However, some of these studies used visual probes, which may not have been detected by participants (e.g., Bauer et al., 1987; Wilson and McCloskey, 1988), while others did not observe graded difficulty-dependent changes with respect to task load (e.g., Kramer et al., 1995) or were limited in the number of participants and recording sites analyzed (Sirevaag et al., 1993). As such, further research employing the task-irrelevant probe technique was warranted.

Building upon the previous work, recent studies have also employed task-irrelevant probes to measure mental workload (e.g., Allison and Polich, 2008; Ullsperger et al., 2001). Ullsperger et al. (2001) challenged participants in 4 task conditions: an oddball task, a visuo-motor task, an arithmetic task, and a simultaneous performance of the visuo-motor and arithmetic tasks. Throughout each condition participants were presented with common and rare pure tones as well as rare novel sounds. The authors reported greater parietal-P3 amplitude to rare tones and greater novelty-P3 amplitude to novel sounds during the oddball task relative to the other 3 conditions. Further, the novelty-P3 tended to decrease slightly as a function of task complexity (although this trend failed to reach statistical significance), whereas this trend was not observed for the parietal-P3. This finding supports novel sounds as being advantageous over pure tones in gauging mental workload. The lack of a significant graded difficulty-dependent effect may be due to the fact that task difficulty was manipulated by having participants perform one of two

* Corresponding author at: 2351 SPH Building, College Park, MD 20742–2611, USA.
Tel.: +1 301 405 2485.

E-mail address: bhatfield@umd.edu (B.D. Hatfield).

¹ These authors contributed equally to the work.

primary tasks in two separate blocks, and in a third block having them simultaneously perform both tasks. It seems reasonable to suggest that incrementally varying the difficulty of a single primary task might be more efficacious in demonstrating significant monotonic differences in relation to task difficulty.

More recently, Allison and Polich (2008) published a study using the task-irrelevant probe technique (specifically, a modified oddball using pure tones) in which the difficulty of a single task was incrementally varied. In this study, participants viewed a video game and also played the game at easy, moderate and hard levels of difficulty. Although the authors reported a significant reduction in the amplitude of a number of ERP components between play (regardless of difficulty) and view conditions, they found little evidence of reductions with respect to changes in game difficulty. Specifically, the authors reported larger P2, N2, and P3 component amplitudes during the view condition relative to all the playing conditions as well as a decrease in N2 amplitude during the hard condition relative to the medium condition. It is possible that the lack of significant graded difficulty-dependent differences in component amplitude might be attributable to the relatively low salience of pure tones, resulting in modest attentional capture (Friedman et al., 2001).

Therefore, in the present study we combined the strengths of the approaches taken by Allison and Polich (2008) and Ullsperger et al. (2001). Specifically, we used a single task that was incrementally varied with respect to difficulty while probing participants with novel sounds. We reasoned that novel stimuli would be more successful in indexing mental workload as they have been shown to be more effective in capturing attention than simple tones. The superior efficacy of the novel stimuli is due to their resemblance to real-world compulsory attention capture by novel or unexpected events (Friedman et al., 2001). In the current study participants played Tetris®, a video game that requires executive control (planning), visual-spatial processing, and motor execution under two levels of difficulty, and in a third condition viewed the game, but did not engage in play. Throughout each condition participants were intermittently presented with a set of novel, task irrelevant, auditory stimuli (e.g., a woman coughing, the sound of breaking glass). One-second epochs of the EEG time-locked to each of the auditory stimuli were extracted and the resultant epochs were averaged within each condition. Based on the prior literature, we predicted that the amplitude of ERP components elicited by the probes would be inversely related to task difficulty.

2. Materials and methods

2.1. Participants

Twenty-three, right-handed young adults (11 women) were recruited from a large Mid-Atlantic metropolitan area. Because the data from three of the participants were discarded due to poor electrophysiological recordings, the final sample was comprised of 20 individuals (10 women, with an age range of 20–33, $M = 24.4$, $SD = 4.1$ years). Tetris® playing experience ranged from never having previously played to having played more than 50 h. All participants provided informed written consent.

2.2. Procedures

Participants were seated in front of a 15" monitor and played Tetris® while the song "Korobeiniki" ("Music A" in the standard Tetris® game) was played (72–76 dB SPL) from a speaker next to the monitor. During the view condition participants fixated on a paused game while the music continued to play. During the easy and difficult conditions, game play began at levels 1 and 8, respectively. After completing a level (i.e., completing 10 lines), the participants immediately advanced to the next level of the game. For each

successive level, the game became more difficult due to an increased rate of speed with which the game pieces fell. Although the game allows the player to manually increase the speed of the pieces, speed was held constant within a level as the participants were instructed not to manually advance the pieces. The change in speed was thought to impact upon mental workload as the participants had to more quickly decide where to place the current piece, execute the placement, and update their planning for the successive pieces. In the event that a participant could not complete a level, the experimenter restarted the game at the level at which the participant began. This occurred rapidly (under 3 s) so as to minimize the probability of a sound being presented during this interruption. A paired sample *t*-test revealed that there were more restarts during the hard condition ($M = 2.3$) than the easy condition ($M = 0.15$) ($t(19) = 10.302$, $p < 0.001$), suggesting a successful manipulation of difficulty. Despite individual differences in game playing experience, all participants ranged between levels 1 and 3 during the easy condition and between levels 8 and 11 in the hard condition.

In each experimental condition, participants were probed with a set of 30 familiar auditory stimuli randomly selected from a larger collection obtained from the New York State Psychiatric Institute (Fabiani et al., 1996). The stimuli were presented in random order (87–96 dB SPL; interstimulus interval = 6–30 s.) from two speakers positioned 70 cm behind the participants.

2.3. Data collection and signal processing

Scalp EEG was collected using tin electrodes housed within a stretchable lycra cap, (Electro-Cap International, Inc.). Data were acquired from 30 sites referenced to linked earlobes and a common ground (FPz). Electrode placement was adapted from the 10–20 international system (Jasper, 1958). Additional electrodes were placed above and below the right eye over the orbicularis oculi muscle and the outer canthi of both eyes to record eyeblinks. Impedances were kept below 10 kΩ for electrodes of interest (Fz, Cz, and Pz) throughout the experiment. All channels were amplified 500 times using Neuroscan Synamps 1, linked to Neuroscan acquisition/edit software (version 4.3). Online bandpass filters were set at .01–100 Hz with a sampling rate of 1000 Hz. Offline, data were processed with the Neuroscan software. After setting a zero-phase shift low-pass filter at 20 Hz (24 dB octave/slope), an ocular artifact algorithm was applied (Semlitsch et al., 1986). ERPs were obtained by extracting the epoch of 100 ms prior to stimulus onset through 900 ms post-stimulus, then baseline corrected with reference to the pre-stimulus interval. Next, each of the 30 trials was visually inspected and any trials containing marked artifact were excluded from subsequent analysis. The remaining trials were then averaged. Each ERP was based on at least 20 trials. The mean amplitude for each component was calculated using the approach suggested by Handy (2005), which recommends the use of narrow time windows centered around the peaks of the components in the grand average waveform. Accordingly, the time windows used were as follows: N1 = 140–160 ms; P2 = 225–255 ms; P3 = 290–320 ms; LPP = 570–610 ms for each of the three midline electrodes of interest (Fz, Cz, and Pz).

2.4. Statistical design

Mean amplitude of each component was subjected to separate 3×3 (Condition $\times$ Electrode) repeated measures ANOVAs. Significant interaction effects were followed by one-way ANOVAs applied to each electrode. Finally, all simple mean effects were determined using paired *t*-tests. Conventional degrees of freedom are reported throughout the results, and the Greenhouse–Geisser correction is provided when sphericity was violated. The *p*-values reported are based upon the corrected degrees of freedom. Additionally, Cohen's *d* effect sizes are also provided when appropriate.

Correlational analyses were also conducted between Tetris experience and all component amplitudes as well as Tetris performance and all component amplitudes.

3. Results

Fig. 1A illustrates the grand average ERPs recorded from the midline electrodes Fz, Cz, and Pz for each condition. The N1, P2, P3, and LPP (late positive potential) components are evident. The statistical analysis revealed a Condition $\times$ Electrode interaction for the N1 component ($F(4,76) = 4.072$; $p = 0.013$, $\epsilon = 0.685$), the P3 component ($F(4,76) = 5.371$; $p = 0.004$; $\epsilon = 0.650$) and the LPP component ($F(4,76) = 4.891$; $p = 0.001$). Additionally, there was a main effect for Condition for the P2 component ($F(2,38) = 6.026$; $p = 0.010$, $\epsilon = 0.757$).

3.1. N1 component

Post hoc analyses revealed that for the N1 at the Cz electrode, the mean amplitude was significantly larger in both the view and easy conditions than in the hard condition (Fig. 1B) (view > hard, $p = 0.003$, $d = 0.717$; easy > hard, $p = 0.045$, $d = 0.473$). The N1 component is believed to reflect compulsory, early sensory processing, exhibit a frontocentral scalp distribution, and be sensitive to attention (Hillyard et al., 1973; Parasuraman and Beatty, 1980).

3.2. P2 component

Post hoc analyses revealed that for the P2 regardless of electrode, the mean amplitude was significantly larger in both the view and easy

conditions than in the hard condition (Fig. 1B) (view > hard, $p < 0.01$, $d = 0.740$; easy > hard, $p < 0.05$, $d = 0.697$). Like the N1, the P2 component is believed to reflect compulsory, early sensory processing and exhibit a frontocentral scalp distribution, and be sensitive to attention (Peters et al., 2005; Picton and Hillyard, 1974).

3.3. P3 component

As predicted, mean amplitudes of the P3 at the Pz electrode differed across all three experimental conditions in a graded difficulty-dependent manner (Fig. 1B) (view > easy, $p = 0.046$, $d = 0.402$; view > hard, $p = 0.003$, $d = 0.906$; easy > hard, $p = 0.012$, $d = 0.674$). The P3 is believed to represent non-obligatory, cognitive evaluation of stimuli and generate a parietal maximal distribution (Parasuraman and Beatty, 1980; Ruchkin et al., 1988). Importantly, the P3 at Pz reflects cognitive processes independent of variations in motor processes (Makeig et al., 2004).

3.4. LPP component

Like the P3, mean amplitudes of the LPP at the Pz electrode differed across all three experimental conditions in a graded difficulty-dependent manner (Fig. 1B) (view > easy, $p = 0.003$, $d = 0.652$; view > hard, $p < 0.001$, $d = 1.717$; easy > hard, $p = 0.004$, $d = 1.037$). Similar to the P3, the LPP is believed to represent non-obligatory, cognitive evaluation of stimuli and generate a parietal maximal distribution (Ruchkin et al., 1988). However, it has been reported much less frequently in the literature than the P3. See Table 1 for all means and post hoc analyses.

The correlational analysis revealed that neither Tetris® experience nor task performance was significantly correlated with any of the ERP component amplitudes.

4. Discussion

For more than three decades, researchers have been using the ERP technique to measure mental workload. The assessment protocols were greatly improved with the use of the task-irrelevant probe technique. Two contemporary studies have further advanced the measurement of mental workload by incrementally-varying task difficulty or using intermittently presented novel sounds (Allison and Polich, 2008 and Ullsperger et al., 2001, respectively). The present study combined elements of these two innovations by incrementally-varying task difficulty while probing participants with novel, task-irrelevant auditory stimuli.

The results demonstrate the utility of this ERP paradigm in indexing mental workload. Unlike many of the prior efforts to index workload, the present approach provided compelling evidence of an inverse relationship between incremental changes in task load and

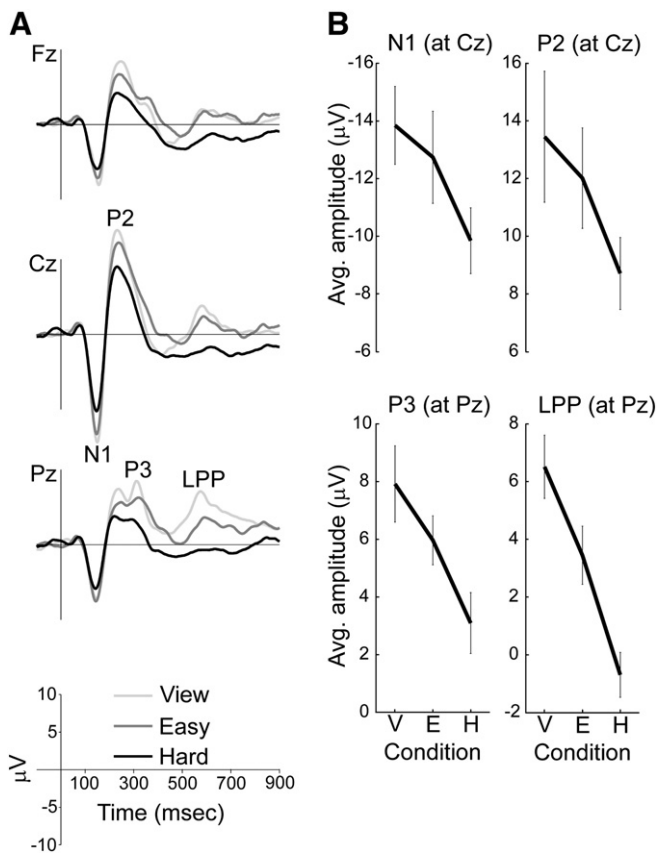


Fig. 1. A) Grand-average ERPs recorded from the Fz, Cz and Pz electrode locations time-locked to the auditory probes. Data from the three experimental conditions are superimposed. B) Average amplitude of the N1, P2, P3 and LPP components as a function of experimental condition (View, Easy, and Hard).

Table 1
Means and summary of statistical results.

Component	Electrode	Condition		
		View	Easy	Hard
N1	Fz	-7.425	-6.933	-5.815
	Cz	-13.844	-12.738	-9.846 ^{a, b}
	Pz	-6.871	-7.090	-5.454
P2	Main Effect	9.632	8.030	5.419 ^{a, b}
P3	Fz	4.034	4.040	2.224 ^{a, b}
	Cz	6.220	7.102	3.940 ^b
	Pz	7.918	5.964 ^a	3.094 ^{a, b}
LPP	Fz	1.893	1.578	-1.756 ^{a, b}
	Cz	3.481	2.181	-2.026 ^{a, b}
	Pz	6.513	3.445 ^a	-0.699 ^{a, b}

^a Significantly different from view.

^b Statistically different from easy.

ERP component amplitude. Due to their psychometric similarity, the Easy vs. Hard comparison was this study's critical contrast of interest. All of the evaluated ERP components differed significantly between the Easy and Hard conditions. Additionally, the P3 and LPP differed among all three conditions suggesting that these components may be the most sensitive (in terms statistical significance and effect sizes) to changes in task difficulty. While modulation of P3 amplitude as a function of task difficulty has been previously reported (Kramer et al., 1987, 1995; Isreal et al., 1980a; Isreal et al., 1980b; Wickens et al., 1977; Sirevaag et al., 1989, 1993), the LPP appears to have provided the most robust index of task load. However, this component is not well-characterized in the literature. Thus, the present findings indicate that future investigation of the LPP is warranted.

The P3 and LPP components are more sensitive to changes in mental workload than the N1 and P2 components. However, it is possible that early cognitive processing, as indexed by N1 and P2, influences these latter components (P3 and LPP). Specifically, given that mental workload is inversely related to attentional reserve and N1 and P2 have been shown to be modulated by attention (Hillyard et al., 1973; Picton and Hillyard, 1974), it is likely that the reduction in N1 and P2 represent a reduction in the allocation of attention to the probe stimuli. This putative reduction in attentional resource allocation can be expected to diminish the stimulus information available to the higher order processes indexed by P3 and LPP, leading to a further reduction in amplitude of these components as well.

The detection of graded difficulty-dependent reductions in the ERP in the present study likely resulted from taking advantage of select innovations developed by Allison and Polich (2008) and Ullsperger et al. (2001). Specifically, the use of novel, environmental sounds, as opposed to pure tones, appears to have elicited increased electrocortical activity to the sounds. Additionally, incrementally varying the difficulty on a single task, as opposed to switching tasks or concurrently performing multiple tasks, likely facilitated the detection of monotonic differences in relation to task load. Future work should examine more gradations of task difficulty to determine if reductions in electrocortical activity are concomitant with increases in mental workload when performance remains relatively stable.

This measure of mental workload has broad implications and is also easy to implement in that it requires a small number of trials (30) to generate an informative index. Notably, such a measure could be employed to assess the demands of various cognitive tasks (e.g., reading, operating a machine) which could then be applied to various learning environments and human–machine interfaces. In summary, the present effort provided a unique contribution to the assessment of mental workload using the ERP technique. By building upon previous innovations, the current study was able to capture graded difficulty-dependent reductions in the ERP.

Acknowledgments

We would like to thank Drs. Raja Parasuraman and Scott Kerick for their helpful comments on an earlier draft of this manuscript. We

would also like to thank the anonymous reviewer for general comments and insightful interpretation of the ERP results.

References

- Allison, B.Z., Polich, J., 2008. Workload assessment of computer gaming using a single-stimulus event-related potential paradigm. *Biol. Psychol.* 77, 277–283.
- Bauer, L., Goldstein, R., Stern, J., 1987. Effects of information processing demands on physiological response patterns. *Hum. Factors* 29, 213–234.
- Fabiani, M., Kazmerski, V.A., Cycowicz, Y.M., Friedman, D., 1996. Naming norms for brief environmental sounds: effects of age and dementia. *Psychophysiology* 33, 462–475.
- Friedman, D., Cycowicz, Y.M., Gaeta, H., 2001. The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neurosci. Biobehav. Rev.* 25, 355–373.
- Handy, T.C., 2005. Basic principles of ERP quantification. In: Handy, T.C. (Ed.), *Event-related Potentials: A Methods Handbook*. MIT Press, Cambridge, pp. 33–56.
- Hillyard, S.A., Hink, R.F., Schwent, V.L., Picton, T.W., 1973. Electrical signs of selective attention in the human brain. *Science* 182, 177–180.
- Isreal, J., Chesney, G., Wickens, C., Donchin, E., 1980a. P300 and tracking difficulty: evidence for multiple resources in dual-task performance. *Psychophysiology* 17, 259–273.
- Isreal, J., Wickens, C., Chesney, G., Donchin, E., 1980b. The event-related potential as an index of display monitoring workload. *Hum. Factors* 22, 211–224.
- Jasper, H.H., 1958. The ten-twenty electrode system of the international system federation. *Electroencephalogr. Clin. Neurophysiol.* 10, 371–375.
- Kramer, A.F., Wickens, C.D., Donchin, E., 1985. Processing of stimulus properties: evidence for dual task integrality. *J. Exp. Psychol. Hum.* 11, 393–408.
- Kramer, A.F., Sirevaag, E.J., Braune, R., 1987. A psychophysiological assessment of operator workload during simulated flight missions. *Hum. Factors* 29, 145–160.
- Kramer, A.F., Trejo, L.J., Humphrey, D., 1995. Assessment of mental workload with task-irrelevant auditory probes. *Biol. Psychol.* 40, 83–100.
- Makeig, S., Delorme, A., Westerfield, M., Jung, T.P., Townsend, J., Courchesne, E., et al., 2004. Electroencephalographic brain dynamics following manually responded visual targets. *PLoS Biol.* 2, 747–762.
- Papanicolaou, A., Johnstone, J., 1984. Probe evoked potentials: theory, method and applications. *Int. J. Neurosci.* 24, 107–131.
- Parasuraman, R., Beatty, J., 1980. Brain events underlying detection of weak sensory signals. *Science* 210, 80–83.
- Peters, J., Suchan, B., Zhang, Y., Daum, I., 2005. Visuo-verbal interactions in working memory: evidence from event-related potentials. *Cogn. Brain Res.* 25, 406–415.
- Picton, T.W., Hillyard, S.A., 1974. Human auditory evoked potentials II: effects of attention. *Electroencephalogr. Clin. Neurophysiol.* 36, 191–199.
- Ruchkin, D.S., Johnson Jr., R., Mahaffey, D., Sutton, S., 1988. Toward a functional categorization of slow waves. *Psychophysiology* 25, 339–353.
- Semlitsch, H.V., Anderer, P., Schuster, P., Presslich, O., 1986. A solution for reliable and valid reduction of ocular artifacts, applied to P300 ERP. *Psychophysiology* 23, 695–703.
- Sirevaag, E.J., Kramer, A.F., Coles, M.G.H., Donchin, E., 1989. Resource reciprocity: an event-related brain potential analysis. *Acta Psychol.* 70, 77–97.
- Sirevaag, E., Kramer, A., Wickens, C., Reisweber, M., Strayer, D., Grenell, J., 1993. Assessment of pilot performance and mental workload in rotary wing aircraft. *Ergonomics* 36, 1121–1140.
- Ullsperger, P., Freude, G., Erdmann, U., 2001. Auditory probe sensitivity to mental workload changes — an event-related potential study. *Int. J. Psychophysiol.* 40, 201–209.
- Wickens, C.D., Isreal, J.B., Donchin, E., 1977. The event-related cortical potential as an index of task workload. *Proceedings of the Human Factors Society. Human Factors Society, Santa Monica, CA.*
- Wickens, C., Kramer, A., Vanasse, L., Donchin, E., 1983. Performance of concurrent tasks: a psychophysiological analysis of the reciprocity of information-processing resources. *Science* 221, 1080–1082.
- Wilson, G., McCloskey, K., 1988. Using probe evoked potentials to determine information processing demands. *Proceedings of the Human Factors Society. Human Factors Society, Santa Monica, CA.*