Ergonomics
Vol. 55, No. 5, May 2012, 526–537

Mental workload during brain–computer interface training

Elizabeth A. Felton[a], Justin C. Williams[b], Gregg C. Vanderheiden[b] and Robert G. Radwin[b]*

[a]*Department of Neurology, School of Medicine, Johns Hopkins University, Baltimore, MD, USA;* [b]*Department of Biomedical Engineering, University of Wisconsin-Madison, Madison, WI, USA*

(*Received 30 May 2011; final version received 26 January 2012*)

It is not well understood how people perceive the difficulty of performing brain–computer interface (BCI) tasks, which specific aspects of mental workload contribute the most, and whether there is a difference in perceived workload between participants who are able-bodied and disabled. This study evaluated mental workload using the NASA Task Load Index (TLX), a multi-dimensional rating procedure with six subscales: Mental Demands, Physical Demands, Temporal Demands, Performance, Effort, and Frustration. Able-bodied and motor disabled participants completed the survey after performing EEG-based BCI Fitts' law target acquisition and phrase spelling tasks. The NASA-TLX scores were similar for able-bodied and disabled participants. For example, overall workload scores (range 0–100) for 1D horizontal tasks were 48.5 ($SD = 17.7$) and 46.6 (SD 10.3), respectively. The TLX can be used to inform the design of BCIs that will have greater usability by evaluating subjective workload between BCI tasks, participant groups, and control modalities.

Practitioner Summary: Mental workload of brain–computer interfaces (BCI) can be evaluated with the NASA Task Load Index (TLX). The TLX is an effective tool for comparing subjective workload between BCI tasks, participant groups (able-bodied and disabled), and control modalities. The data can inform the design of BCIs that will have greater usability.

Keywords: brain–computer interface (BCI); mental workload; NASA Task Load Index, NASA-TLX; Fitts' law; electroencephalogram (EEG)

1. Introduction

A brain–computer interface (BCI) is a communication system that does not depend on the brain's normal input/output pathways of peripheral nerves and muscles (Wolpaw *et al.* 2000). It is anticipated that the more direct connection between a computer and neural commands should give BCIs a would-be advantage over other augmentative communication methods. This technology offers the promise to provide functionality and independence to individuals with severe motor disabilities by enabling them to directly control assistive devices such as computers, wheelchairs or prosthetic limbs.

Previous BCI studies have demonstrated that humans can move a computer cursor to a target using signals from surface electrodes (i.e. electroencephalogram or EEG) (Wolpaw *et al.* 2000, Wolpaw and McFarland 2004, Kubler *et al.* 2005), subdural electrodes (i.e. electrocorticogram or ECoG) (Leuthardt *et al.* 2004, Felton *et al.* 2007b), or electrodes implanted within the brain (i.e. local field potentials and action potentials) (Kennedy *et al.* 2000, Hochberg *et al.* 2006). Although the methods vary, cursor control is often accomplished by training the participant to use motor imagery to modulate neural signals (Wolpaw *et al.* 2002, Leuthardt *et al.* 2004, Wolpaw and McFarland 2004, Felton *et al.* 2007b).

Although BCI research is rapidly advancing, there are still limitations to overcome before BCIs become a widely available assistive technology. It is not well understood how people perceive the difficulty of performing BCI tasks, which specific aspects of workload contribute the most, and whether there is a difference in perceived workload between users who are able-bodied and disabled. Equally important is an understanding of how to best design the interface to reduce mental workload and prevent fatigue, both of which are likely to be limiting factors of the current technology (Curran and Stokes 2003, Riccio *et al.* 2011).

This study evaluated mental workload using the NASA Task Load Index (NASA-TLX) for EEG-based BCI target acquisition and phrase spelling tasks executed by participants who are able-bodied and motor-disabled. The target acquisition task experiment involved training participants to use motor imagery to control the movement of a computer cursor in one dimension (1D) or two dimensions (2D). Fitts' law, which is a predictor of human

*Corresponding author. Email: radwin@engr.wisc.edu

ISSN 0014-0139 print/ISSN 1366-5847 online

http://dx.doi.org/10.1080/00140139.2012.662526
http://www.tandfonline.com

movement time, was used to determine the information transfer rate, which was based on target acquisition time and target difficulty (Felton *et al.* 2009). The phrase-spelling task experiment involved training participants to use similar brain-controlled cursor movements to interface with a text-entry system to write short phrases (Felton *et al.* 2007a). For comparison, able-bodied participants also performed the tasks using a manual control device (joystick or mouse).

1.1. Mental workload

Mental workload should be considered when designing tasks or procedures that require user attention. Research on mental workload started in the 1960s when investigators evaluated human–machine systems such as ground transportation, air traffic control, and process control (Parasuraman and Hancock 2001). Many studies have evaluated strategies to reduce mental effort when performing computer tasks (Wickens and Hollands 2000), which tend to require high visual and cognitive demands, leading to an increase in mental workload (Hjortskov *et al.* 2004).

A systematic evaluation of mental workload during BCI training and use has not yet been performed. BCI research is typically conducted in the well-controlled environment of an isolated, distraction-free laboratory where participants are instructed to sit still and concentrate on the task. In addition, the types of tasks tend to be emotionally neutral, which clearly will not always be the case for people who might use BCI as their main mode of communication. In actual use, BCI users will have a higher cognitive load due to increased attentional demands from distractions, emotions, and interactions with other people (Curran and Stokes 2003). Kennedy *et al.* reported that one participant with an implanted neurotrophic electrode had reduced BCI performance when medicated, ill, or fatigued (in general or from repeat BCI use) (Kennedy *et al.* 2000). When the researchers noticed a decline in performance and asked about the participant's sense of effort, the participant reported that it was maximal (Kennedy *et al.* 2000). This follows the theory that high workload can lead to stress and errors (Wickens and Hollands 2000).

1.2. NASA-Task Load Index

The NASA-TLX was developed by the Human Performance Group at the NASA Ames Research Center (Hart and Staveland 1988). It is a multi-dimensional rating procedure with six subscales: Mental Demands, Physical Demands, Temporal Demands, Performance, Effort, and Frustration. The first three subscales relate to demands on the participant and the last three relate to the interaction of the participant with the task. Numerical ratings for each subscale are obtained by having the participant mark a line divided into 20 equal intervals, which is converted to a rating on a 0–100 scale. The line is anchored on each side with bipolar descriptors (for example, low and high). A raw workload score is calculated by adding the rating given to each of the six subscales and dividing by six (Byers *et al.* 1989). The more commonly used overall workload score is calculated based on a weighted average rating of the subscales. Weights are determined by the participant's evaluation of how much each factor contributed to the task workload. After completing the initial ratings, 15 pair-wise comparisons of the 6 subscales are presented individually, and the participant chooses the subscale contributing most to the task workload. The weight for each subscale is the total number of times it was selected divided by 15. An overall workload score is calculated by multiplying each rating by the weight given to that subscale, and can range from 0–100 (Hart and Staveland 1988). As a multidimensional scale, the NASA-TLX is useful for obtaining more detailed and diagnostic data than a unidimensional scale (Hill *et al.* 1992). The NASA-TLX has been used in many investigations of human–computer interaction as a method of assessing mental workload (Smith *et al.* 2001, Vitense *et al.* 2003).

2. Methods

2.1. Participants

Twelve participants with disabilities (see Table 1 for demographic information) and 32 able-bodied participants were enrolled, however, the requirement for inclusion in this data analysis was completion of at least six one-hour BCI training sessions. Therefore, 12 able-bodied participants (7 men and 5 women, aged 19–34) and 7 participants with disabilities (4 men and 3 women, aged 46–60) were included in the analysis. The majority of participants excluded from the analysis did not complete enough sessions. The only exceptions were two participants with disabilities for whom there were difficulties in administering the TLX, so accuracy of responses could be not be assured. The participants were naïve to BCI tasks upon initial enrolment in the study, with the exception of one able-bodied participant who participated in ECoG-based BCI experiments 1.5 years prior to

Table 1. Demographics of participants with disabilities.

#	Description of disability	Gender	Age	Duration of disability	Speech affected (Y/N)*	Mobility Affected (Y/N)	> 5 BCI sessions (Y/N)	Cursor Control (Y/N)	NASA-TLX (Y/N)
1	Spinal muscular atrophy (type 2)	F	27	Birth	Y	Y	Y	N	N
2	Spinal cord injury (level T-10)	M	51	8 years	N	Y	N	N	N
3	Locked-in syndrome	M	59	20 years	Y	Y	Y	N	N
4	Amyotrophic lateral sclerosis	M	50	18 years	Y	Y	Y	Y	Y
5	Amyotrophic lateral sclerosis	F	46	19 years	Y	Y	Y	Y	Y
6	Spinal cord injury (level C4-5)	M	50	9 years	N	Y	Y	N	Y
7	Post-polio	F	65	51 years	N	Y	N	N	N
8	Post-polio	M	60	53 years	N	Y	Y	Y	Y
9	Muscular dystrophy (MD) and post-polio	M	54	MD – Birth, Polio – 51 years	Y	Y	Y	Y	Y
10	Cerebral palsy	M	47	Birth	Y	Y	N	N	N
11	Spinal muscular atrophy (type 2/3)	F	49	Birth	N	Y	Y	Y	Y
12	Spinal muscular atrophy (type 3)	F	47	Birth	N	Y	Y	N	Y

Note: *The degree of speech impairment varied from difficult to understand to no speech at all.

commencing participation in the EEG experiments presented here. The tasks performed in the prior experiments were different than those used in the current study, and her BCI task performance did not differ significantly from the participants who were naïve to BCI tasks. All participated with informed consent and the protocol was approved by the University of Wisconsin-Madison Health Sciences Institutional Review Board.

2.2. *Experimental hardware and software*

Participants wore a 16-channel electrode cap (Electro-Cap International Inc., Eaton, OH), which recorded and transmitted their EEG signals to a computer via a g.USBamp amplifier (Guger Technologies, Graz, Austria) or Pentusa amplifier and signal processing system (Tucker-Davis Technologies, Alachua, FL). The pure tin electrodes were 9 mm in diameter and the cap configuration followed the standard 10–20 system for electrode placement (see (Sharbrough *et al.* 1991) for details).

BCI2000, a general-purpose brain–computer interface software package created by investigators at the Wadsworth Center in Albany, New York, was used for these experiments (Schalk *et al.* 2004). BCI2000 uses a standardised data format for offline analysis and is based on a model that can be applied to any BCI system. The code for BCI2000 is open-source, so the programme was modified with in-house software routines for the tasks used in this experiment (Schalk *et al.* 2004).

2.3. *Experimental procedure: screening and parameter selection*

A screening procedure was performed during the first experimental session to determine the EEG spectral frequency components of specific electrodes that participants could self-modulate using motor imagery. Participants imagined different types of movements (e.g. clenching their hands, tapping their feet, sticking out their tongue) in response to visual cues presented on a computer screen. These data were processed in the frequency domain using autoregressive spectral analysis to find frequency bands of specific electrodes that changed in power between the imagery (active) and blank screen (rest) time segments. If a high correlation ($r^2 > 0.3$) was observed between the power change and an active response time segment, it was concluded that the participant could use that imagery to self-modulate the specific signal component. Based on the analysis, 3–5 Hz wide frequency bands (e.g. 12–15 Hz) of specific electrodes were assigned a preferred direction for cursor movement (e.g. left or right). A specific electrode, frequency range, and weighting factor were used to control each direction of cursor movement (up, down, left, right). The selected electrodes usually corresponded to locations over sensorimotor cortex, such as C3 and C4 (based on 10–20 electrode system convention). During each trial, the EEG power spectral content for each electrode was measured continuously, and the signal magnitude was entered as the independent variable in a linear calibration equation that controlled cursor motion in real time. Calibration occurred at the beginning of each session, so the weighting factors could vary slightly between sessions. Details on the calibration and weighting algorithms can be found in previous work describing the BCI2000 platform (Schalk *et al.* 2004).

Once the parameters selected from the screening tasks were established, participants were trained to use motor imagery to acquire onscreen targets with a computer cursor controlled by their neural signals. Participants were instructed to use the same imagery as during the screening tasks and adjust as necessary to what seemed to work best for them. By using imagery to control the cursor movement, they were effectively self-modulating the EEG signals.

2.4. Experimental procedure: target acquisition tasks

At the start of each trial, a rectangular target appeared on the screen (Figure 1) and then 2 s later, a circular 'cursor' appeared at the centre of the screen. The participant was instructed to use specific types of imagery to control the cursor movement in each direction. For example, cursor movement in the horizontal direction was typically accomplished by instructing participants to imagine clenching their left or right hand to make the cursor move left or right, respectively. The task was to move the cursor to the target as quickly and accurately as possible and hold it within the target for 500 ms. When the centre of the cursor moved into the target, the target changed colour to indicate the cursor was in the target and the participant needed to hold it there for 500 ms. After 500 ms, the target flashed and disappeared to indicate a successful acquisition. If the cursor did not come in contact with the target after 15 s, the trial ended. However, if the cursor came in contact with the target, but did not dwell 500 ms, the timer kept going until the target was acquired, or 15 s passed without contacting the target. There was a 2 s rest period between trials and participants completed as many trials as possible in sets lasting 3 min. The targets appeared in a random order over the duration of the set. A typical training session would include 10 of the 3 min sets, and each set had 10–20 trials depending on how fast targets were acquired.

The Fitts' index of difficulty (ID) describes the relative difficulty of a ballistic movement in a target acquisition task emphasising speed over accuracy, and is based on the distance (A) from the starting point to the target, and the width (W) of the target (Equation 1) such that movement time (MT) is a function of ID (Equation 2). The information transfer rate (bits/sec) is calculated by taking the reciprocal of the slope, *b*, in the regression equation.

$$ID = \log_2 \frac{2A}{W} \tag{1}$$

$$MT = a + b\left[\log_2\left(\frac{2A}{W}\right)\right] = a + b(ID) \tag{2}$$

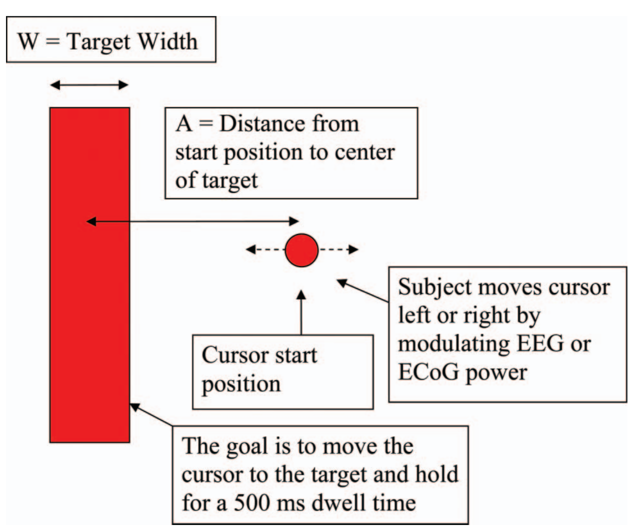

Figure 1. Target acquisition task based on Fitts' law. The task is to move the circular cursor to the rectangular target. One target is shown, but targets of varying indices of difficulty are presented one at a time on each side of the screen during the task. Index of difficulty is varied by changing the width (W) of the target and distance from the start position of the cursor to the centre of the target (A).

BCI task degrees of freedom increased from unidirectional (1D) to orthogonal directional (OD) to two-dimensional (2D). The 1D task constrained presented targets along a single coordinate axis and the cursor could only move along that axis (either left–right horizontal OR up–down vertical). The OD task targets could be presented in any of the four directions and the cursor could move along either of the two orthogonal coordinates (left–right horizontal and up–down vertical). The 2D tasks involved free movement in any two dimensional direction (e.g. up–right diagonal). Participants typically learned horizontal 1D movement first and then they learned vertical 1D movement. Even as participants advanced to more difficult tasks as described below, there was always time at the beginning of each session to practice movements learned previously.

There were seven indices of difficulty (ID) for the 1D tasks: 0.58, 0.74, 1.14, 1.22, 2.0, 2.59, and 3.70 bits. This meant there were 28 possible target positions (seven each on the top, bottom, left, and right sides of the screen), although the horizontal and vertical dimensions were evaluated separately because the cursor movement for each dimension required different types of imagery. For example, in a 1D horizontal task set, targets would appear only on the left or right side of the screen and the cursor movement was constrained to the horizontal (left–right) direction. After at least five sessions of learning how to control both the horizontal and vertical dimensions of cursor movement, these were combined in the orthogonal directional (OD) task where the six of the same seven indices of difficulty (0.58–2.59 bits) were evaluated, for a total of 24 possible targets positions that appeared randomly in a test set. In order to perform the Fitts' law analysis, which is reported separately (Felton *et al.* 2009), the task difficulty is controlled and a minimal level of task accuracy is expected. For this study BCI cursor control is definite as $\geq$80% task accuracy.

Finally, two participants who were successful with the OD tasks ($\geq$90% accuracy) advanced to a true 2D task where the targets could be located diagonally from the cursor start position. Participants could acquire targets by simultaneously or sequentially performing the imagery attributed to each direction of cursor movement. For example, if a target was located to the bottom-left of the cursor, and the participant used foot movement imagery for downward cursor movement and left hand clenching imagery for leftward cursor movement, they could either think about the foot and hand movement together or sequentially. The 2D task only had two levels of ID: 1.59 and 1.92 bits.

2.5. *Experimental procedure: Dasher-BCI*

In addition to the basic BCI target acquisition tasks, eight able-bodied participants who completed at least six 1-hour BCI sessions with 1D task accuracy $\geq$90% for three consecutive sessions were trained to use similar brain-controlled cursor movements to interface with the Dasher text-entry system. Participants with disabilities were not trained to do the Dasher-BCI because none of them achieved the consistent high level of task accuracy required. Dasher is a user interface for computer text entry intended for situations when a keyboard cannot be used. The programme uses a language model to predict the probabilities of characters that come next, making more likely text easier to write. For detailed information about Dasher and details of the Dasher-BCI experiment see (Wills and MacKay 2006, Felton *et al.* 2007a).

Participants wrote six predetermined short phrases ('good morning,' 'turn tv on,' 'clean glasses,' 'too cold,' 'lights off,' and 'head hurts') in Dasher using brain-controlled cursor movements. The Dasher-BCI tasks used the direction (horizontal or vertical) of cursor movement most accurately controlled by the participant. For example, if a participant was able to best control horizontal cursor movement using left and right hand motor imagery, they used the same motor imagery and screen orientation for Dasher-BCI tasks. In Dasher, the letters could be set up to appear at the right side or bottom of the screen, requiring 1D vertical or horizontal cursor movement, respectively. Participants moved the cursor in one direction to select letters and if an error was made, they moved the cursor in the opposite direction to remove unwanted letters.

2.6. *Experimental procedure: manual cursor control*

Manual control tasks using a joystick or mouse were also carried out with able-bodied participants. All participants reported some prior experience using both a joystick and mouse, and none reported difficulty with their operation. The BCI target acquisition tasks were performed with a hand-held joystick (Logitech Attack 3, Logitech Inc.) and the Dasher phrase spelling tasks were completed with a standard hand-held mouse. This procedure took place during a session at the end of BCI training, so a direct comparison between manual and brain-controlled cursor movement could be made. Nine able-bodied participants did the joystick task and seven did the Dasher-mouse task. The purpose of this part of the study was to produce a baseline level of performance using a conventional computer input device to compare against the BCI performance.

2.7. Mental workload analysis

Participants were administered the computerised NASA-TLX at the end of BCI training, after completing at least six one-hour sessions. Therefore, it was administered after participants had adequate practice with the tasks. Separate evaluations were completed for brain-controlled target acquisition tasks, joystick-controlled target acquisition tasks, brain-controlled phrase writing in Dasher, and mouse-controlled phrase writing in Dasher. The mean raw workload, overall workload, and weighted subscale scores were calculated for each task. The raw and overall workload values indicated the tasks requiring the most mental workload, while the weighted subscale scores identified the workload factors contributing most to each task. The highest possible score for the raw and overall workload is 100, while the highest score for the weighted subscales is 33.3. The independent variables were task (target acquisition or Dasher) and input modality (brain or manual input). The dependent variables were overall workload and each weighted subscale score.

3. Results

3.1. NASA-TLX scores

NASA-TLX mean raw, overall workload, and weighted subscale scores are shown in Tables 2 and 3 for participants who are disabled (D) and able-bodied (AB). Table 2 shows the data for participants who achieved BCI cursor control ($\geq$80% task accuracy for 1D tasks) and Table 3 shows the data for participants who did not. The six NASA-TLX subscales are: Mental Demands, Physical Demands, Temporal Demands, Performance, Effort, and Frustration. Note that these are subjective ratings; so for example, the perceived performance does not necessarily reflect the participant's actual performance. The raw score is the average of the six subscale ratings, while the overall workload score is the total after weighting each subscale rating.

For participants who achieved cursor control, the tasks with the highest and lowest overall workload scores were Dasher-BCI (81.1, $SD = 14.3$) and 1D joystick (31.3. $SD = 13.5$), respectively. However, when evaluating the weighted subscale ratings, the Dasher-BCI task was highest on only two of the six subscales: effort (19.9, $SD = 9.1$), and frustration (15.8, $SD = 10$).

A one-way repeated measures analysis of variance (ANOVA) indicated no significant effects for overall workload score between brain-controlled target acquisition tasks performed by able-bodied participants who were able to achieve cursor control for the basic 1D horizontal, 1D vertical, and orthogonal directional tasks, with mean scores of 48.5, 69.1, and 60.8, respectively. There also were no significant differences ($p < 0.05$) for any of the subscales between the three tasks.

3.2. Comparison between participant groups for 1D BCI tasks

A between groups ANOVA was performed to determine the effect of participant group (able-bodied or disabled) on the 1D brain-controlled task conditions (horizontal and vertical). There were no statistically significant differences ($p > 0.05$) between raw TLX, overall workload, or subscale scores between groups for participants with cursor control.

3.3. Comparison between cursor control groups for 1D BCI tasks

A between groups ANOVA was performed to determine the effect of ability to perform the BCI tasks (cursor control and no cursor control) on the 1D brain-controlled task conditions (horizontal and vertical). For able-bodied participants, there were no statistically significant differences ($p > 0.05$) between raw TLX, overall workload, or subscale scores between those who were and were not able to perform the 1D BCI tasks. For participants with disabilities, there was a statistically significant difference ($p < 0.05$) for the overall workload and raw TLX scores for the 1D horizontal task. Participants with disabilities who had cursor control had an overall workload score of 46.6 (SD 10.3) compared to 81 (SD 0) for those who did not. All other differences between scores for the 1D horizontal and vertical tasks were not statistically significant. Note, the low participant numbers in each group must be considered when interpreting these results.

3.4. Dasher-BCI tasks

Dasher BCI tasks were only performed by the able-bodied participants. A one-way repeated measures ANOVA indicated statistically significant effects ($p < 0.05$) for overall workload score between both 1D tasks (horizontal

Table 2. NASA-TLX mean (*SD*) score for each task – Participants with cursor control.

Task	Brain Control (Y/N)	Participant Group (AB/D)	*N*	Raw TLX (0–100)	Overall Workload (0-100)	Mental Demand (0–33.3)	Physical Demand (0–33.3)	Temporal Demand (0–33.3)	Performance (0–33.3)	Effort (0–33.3)	Frustration (0–33.3)
1D Horizontal	Y	D	5	39.8 (8.1)	46.6 (10.3)	13.2 (8.9)	0.7 (1.5)	9.3 (7)	5.2 (6.1)	14.6 (6.4)	3.6 (3.3)
1D Vertical	Y	D	4	57.9 (12.7)	67.4 (14.9)	13.5 (10.8)	1.3 (1.6)	4.7 (3.2)	17.7 (8.7)	18.1 (6.5)	12.2 (8.2)
1D Horizontal	Y	AB	8	38.1 (14.5)	48.5 (17.7)	16.7 (8)	0.3 (0.7)	6.2 (8.6)	6 (6.9)	13.5 (6.2)	5.9 (6.6)
1D Vertical	Y	AB	7	58 (13.8)	69.1 (18.6)	21.0 (10.1)	0 (0)	12.1 (7.4)	9 (6.6)	14.6 (8.1)	12.3 (7.9)
OD	Y	AB	4	50.8 (7.6)	60.8 (7.9)	17.6 (2.2)	0 (0)	11.7 (5.1)	8.3 (7.8)	17.5 (9.6)	5.7 (3.5)
2D	Y	AB	2	53.8 (12.4)	64 (12.7)	17.0 (0.5)	0 (0)	16.3 (3.3)	8 (4.2)	16.3 (14.6)	6.3 (5.2)
Joystick	N	AB	9	28.1 (11.6)	31.3 (13.5)	3.2 (3.2)	8.1 (8.5)	4.6 (7.1)	5.5 (7.2)	9.2 (9.6)	0.7 (1.8)
Dasher-BCI	Y	AB	8	69.2 (13.9)	81.1 (14.3)	20.8 (7.9)	0.7 (1.9)	13.7 (12.3)	10.1 (5.6)	19.9 (9.1)	15.8 (10)
Dasher-Mouse	N	AB	7	32.6 (14.7)	38.9 (18.2)	7.0 (7.1)	3.9 (8.6)	13.7 (11.1)	6.1 (5.7)	6.4 (4.3)	1.7 (2.3)

Note: N = number of participants completing the evaluation for each task; AB – able bodied participants; D – participants with disabilities;; OD – orthogonal directional; Raw TLX = Raw Task Load Index score, composed only of the scale ratings, range of possible scores is 0 – 100; Overall Workload = Combined weighted scores from each subscale, range of possible scores is 0 – 100.

Table 3. NASA-TLX mean (*SD*) score for each task – Participants without cursor control.

Task	Brain Control (Y/N)	Participant Group (AB/D)	*N*	Raw TLX	Overall Workload	Mental Demand	Physical Demand	Temporal Demand	Performance	Effort	Frustration
1D Horizontal	Y	D	2	70 (1.2)	81 (0)	27.2 (6.4)	0 (0)	15.2 (3.1)	11.2 (8.2)	17.5 (2.1)	10 (7.1)
1D Vertical	Y	D	2	70.4 (1.8)	83.3 (4.7)	30 (0)	4.3 (6.1)	1.8 (2.6)	12.7 (4.7)	20.8 (2.6)	13.7 (6.1)
1D Horizontal	Y	AB	4	51.9 (9.4)	66 (13.9)	23.1 (6.1)	0.5 (1)	7.5 (2.9)	8 (4.7)	19.2 (9.8)	7.7 (12.7)
1D Vertical	Y	AB	2	47.1 (6.5)	60.3 (7.5)	26.7 (0)	0 (0)	12.5 (0.7)	4 (1.9)	15.3 (4.7)	1.8 (0.2)

Note: See Table 2 for key.

and vertical) and Dasher-BCI tasks. The mean overall workload scores for 1D horizontal, 1D vertical, and Dasher-BCI were 48.5 ($SD = 17.7$), 69.1 ($SD = 18.6$), and 81.1 ($SD = 14.3$), respectively. The only statistically significant subscale score ($p < 0.05$) was between 1D horizontal and Dasher-BCI tasks for frustration, with a mean subscale score of 5.9 ($SD = 6.6$) and 15.8 ($SD = 10$), respectively.

3.5. Manual control tasks

Comparisons were also made between the brain and manual controlled tasks performed by able-bodied participants. The manual control tasks were identical to the brain-controlled tasks, but were carried out with either a joystick or mouse. Scores for manual devices were used to determine baseline mental workload for each task using a familiar device and can be used to determine sources of workload that differ between modalities.

In general, the most important contributors to the workload for the basic BCI target acquisition tasks (1D horizontal and vertical) were mental demand and effort, while physical demand and effort were most important for the equivalent task carried out with a joystick. There were statistically significant effects ($p < 0.05$) for overall workload between 1D horizontal BCI and joystick tasks, with a mean score of 48.5 ($SD = 18$) and 31.3 ($SD = 13.5$), respectively. There were also statistically significant effects between 1D BCI and joystick controlled tasks ($p < 0.05$) for the mental demand and effort subscales.

There were statistically significant effects ($p < 0.01$) between Dasher-BCI and Dasher-mouse tasks, with a mean overall score of 69.2 ($SD = 13.9$) and 32.6 ($SD = 14.7$), respectively. There were also statistically significant effects between Dasher-BCI and Dasher-mouse tasks ($p < 0.05$) for the mental demand, effort, and frustration subscales.

4. Discussion

This study investigated mental effort for brain-controlled target acquisition and phrase-spelling tasks. This information was studied in order to understand the practicality of using BCIs for assistive technology and alternative control tasks, and to make comparisons between participant groups for different computer input modalities. Although BCI technology is targeted for individuals with disabilities, the majority of BCI research participants are able-bodied (AB). It was anticipated that persons with disabilities who may benefit from BCI would perceive these tasks as less mentally demanding.

Evaluating BCI technology in AB participants provides a good starting point for understanding BCI-related mental workload. However, many differences can exist between AB and populations with disabilities, so the results may not translate directly. For example, users with disabilities may report less mental workload because the brain-controlled tasks are easier for them to execute than manual tasks. Or, even if the brain-controlled tasks are more difficult than tasks carried out using an alternative computer input method, the difference may be smaller than that experienced by AB participants who can easily perform computer tasks with a joystick or mouse.

This study had several limitations. The total number of study participants, particularly those with disabilities, was limited due to multiple factors including the number of training sessions needed for this study. Further, the participants with disabilities varied in their type and duration of disability (see Table 1), making this a relatively heterogeneous group. Recruitment of participants with disabilities is difficult and although the numbers in this study were small, they are comparable or larger than other BCI studies of this type. The mean age of the two participant groups was quite different, so there may be other factors, such as general comfort with technology, which influenced the results of the younger able-bodied group. Also, the NASA-TLX was administered after several training sessions, once the participants were familiar with the task. However, it would be interesting to administer the tool at regular intervals throughout training to evaluate how perceptions change over time. Finally, the manual cursor control conditions were not counterbalanced in this experiment – they were done at the end of BCI testing.

Despite the limitations, this study has shown that the NASA-TLX scale can be an effective tool for evaluating mental workload for a range of BCI tasks and participant groups. NASA-TLX scores can be used to inform the design of BCI applications because it provides data on both overall and individual contributors to workload. The scores can be used to directly compare different designs to find those requiring the least amount of mental workload.

The unidirectional (1D) horizontal and vertical tasks were similar in that they required the use of motor imagery to move the cursor in one dimension (i.e. left–right or up–down) only. However, different types of imagery were used for each movement direction and they may not be equally easy to execute. The orthogonal directional (OD) tasks put the two separate dimensions of cursor movement together, so participants were expected to use four different types of imagery depending on the target location. The cursor did not always move directly to the target, so unless immediately acquired, more than one type of imagery may have been necessary for a particular target.

Despite this, there were no significant differences ($p < 0.05$) in the NASA-TLX scores between the 1D horizontal, 1D vertical, and OD brain-controlled tasks. The finding of no differences between the 1D tasks supports results reported separately showing no significant differences between Fitts' law regression lines for the four movement directions (up, down, left, right) (Felton *et al.* 2009).

The Dasher-BCI task was a different implementation of the same brain-control used for the 1D target acquisition tasks. The direction of cursor movement (left–right or up–down) best controlled by the participant was used to continuously steer the cursor towards letters to complete phrase-writing tasks. Although the same brain-control is used for both tasks, there are key differences between the tasks. The standard BCI target acquisition tasks required participants to move the cursor to a target after which a two-second break was given before the next target appeared. In addition, whether they were able to acquire a particular target had no influence on the next target; they were essentially able to 'start-over' with each next target. By contrast, the Dasher-BCI tasks required continuous brain-control of the cursor without any breaks. In addition, a certain sequence of letters had to be selected, so if the incorrect letter was selected, the participant had to make a correction. Significant differences were found between the standard BCI target acquisition and Dasher-BCI phrase-spelling tasks. These findings are important because in order for BCI technology to be pragmatic, the programme will likely involve more than simple target acquisition, even if meaning were to be attributed to the targets (e.g. yes/no, letters, etc.). Programmes such as Dasher have been successful for other types of alternative input modalities, such as the eye tracker and breath mouse (Ward *et al.* 2002, Ward and MacKay 2002, Shorrock *et al.* 2005, Wills and MacKay 2006). Information about the highest sources of workload for that task can be used to make strategic changes to the programme to enhance user performance and reduce mental workload. Because the same type of cursor control was used for the 1D target acquisition tasks, that task can be used as a baseline for comparison. While the overall workload was significantly higher for Dasher-BCI and all subscale scores were higher, the only subscale score that achieved statistical significance was frustration, which was higher for Dasher-BCI.

The NASA-TLX scores for mental workload were similar for able-bodied and participants with disabilities. Mental demand and effort were the subscales contributing most to overall workload in both participant groups. The similarity between groups suggests that participants with disabilities do not have a greater sense of subjective workload when completing BCI tasks. This was despite differences in their cursor movement times as determined by the Fitts' law analysis which is reported separately (Felton *et al.* 2009). For example, participants with disabilities had average 1D movement times ranging from 3.20 s ($SD = 0.84$ s) to 7.01 s ($SD = 1.14$ s) at the lowest and highest indices of difficulty, 0.58 and 2.59, respectively. Whereas, able-bodied participants had average times ranging from 1.78 s ($SD = 0.10$ s) to 5.62 s ($SD = 0.4$ s). Note that although the participants with disabilities had longer movement times, they still met criteria for control, so were able to perform the task. The increased movement time required for participants with disabilities can be accounted for in the task design.

This similarity in mental workload scores will be useful because different types of tasks can be screened in able-bodied participants, who are readily available for BCI studies. Once a set of tasks that requires low workload are selected, these can be evaluated in participants with disabilities. Information from the NASA-TLX can be used to inform the design of BCI applications because it provides data on both overall and individual contributors to workload. The scores can be used to directly compare different designs to find those requiring the least amount of mental workload. For computer-based BCI tasks, the ultimate goal is to reduce workload to that from using a standard mouse or joystick.

An interesting observation was the significantly higher overall and raw workload scores for BCI tasks performed by participants with disabilities who did not have cursor control compared to those who did. It is reasonable to assume that participants who have difficulty performing the BCI task may experience a higher subjective workload. However, that the scores were higher in the disabled versus able-bodied participants who did not have cursor control is also worth noting. Although the participant numbers in each group were low, these differences should be explored further in future studies that include both participants who are able-bodied and disabled.

In addition to making comparisons between different groups of participants there is a need for comparison of workload and performance between modalities. In this study, a manual joystick was compared against EEG based BCI tasks. However, future studies will need to make mental workload comparisons between manual tasks and other brain-controlled modalities, such as those from electrocorticogram (ECoG) and single unit brain recordings. Although those methods are more invasive, they have the potential of an improved signal-to-noise ratio making brain control easier. The NASA-TLX can be a useful tool for comparing across these modalities.

NASA-TLX is just one method that can be used for mental workload analysis. Future studies should consider other objective measures of mental workload. One example of such a measure is heart rate variability (HRV), which describes the variations of instantaneous HR and RR intervals (distance between subsequent R waves on an ECG recording) (Camm *et al.* 1996). HRV has been shown to be sensitive to acute stress, with mental load lowering HRV (Hjortskov *et al.* 2004). Objective measures of workload for BCI tasks may also prove useful for studies that use non-human primates as BCI test participants.

The application of methods from the human factors, ergonomics, and neuroergonomics to BCI research will help take the technology out of the lab and into the real world where it can benefit people with motor disabilities.

Acknowledgements

This work was supported in part by: National Institute on Disability and Rehabilitation Research (NIDRR); US Department of Education (H133E030012); Clinical Neuroengineering Training Program (NIH T90 DK070079); Training and Education to Advance Multidisciplinary Clinical Research Program (NIH K12 HD049077); and Medical Scientist Training Program. The authors wish to thank J. Adam Wilson, PhD for programming assistance; Gerwin Schalk, PhD of the Wadsworth Center for technical and software assistance; Joel Warm, PhD of the University of Cincinnati for first suggesting the use of NASA-TLX for this research; and the study participants for their time, effort and helpful feedback.

References

Byers, J.C., Bittner, A.C., and Hill, N.J., 1989. Traditional and raw task load index (TLX) correlations: are paired comparisons necessary? *In*: A. Mital, ed. *Advances in industrial ergonomics and safety*. Philadelphia, PA: Taylor & Francis, 481–486.

Camm, A.J., Malik, M., Bigger, J.T., Breithardt, G., Cerutti, S., Cohen, R.J., Coumel, P., Fallen, E.L., Kennedy, H.L., Kleiger, R.E., Lombardi, F., Malliani, A., Moss, A.J., Rottman, J.N., Schmidt, G., Schwartz, P.J., and Singer, D., 1996. Heart rate variability – standards of measurement, physiological interpretation, and clinical use. *Circulation*, 93 (5), 1043–1065.

Curran, E.A. and Stokes, M.J., 2003. Learning to control brain activity: a review of the production and control of EEG components for driving brain–computer interface (BCI) systems. *Brain & Cognition*, 51 (3), 326–336.

Felton, E.A., Lewis, N.L., Wills, S.A., Radwin, R.G., and Williams, J.C., 2007a. Neural signal based control of the Dasher writing system. *In*: *3rd International IEEE/EMBS Conference on Neural Engineering*, 2–5 May, Kohala Coast, HI. Piscataway, NJ: IEEE, 366–370.

Felton, E.A., Radwin, R.G., Wilson, J.A., and Williams, J.C., 2009. Evaluation of a modified Fitts law brain–computer interface target acquisition task in able and motor disabled individuals. *Journal of Neural Engineering*, 6 (5), 056002.

Felton, E.A., Wilson, J.A., Williams, J.C., and Garell, P.C., 2007b. Electrocorticographically controlled brain-computer interfaces using motor and sensory imagery in patients with temporary subdural electrode implants: report of four cases. *Journal of Neurosurgery*, 106 (3), 495–500.

Hart, S.G. and Staveland, L.E., 1988. Development of NASA-TLX (task load index): results of empirical and theoretical research. *In*: P.A. Hancock and N. Meshkati, eds. *Human mental workload*. Amsterdam: North Holland, 139–184.

Hill, S.G., Iavecchia, H.P., Byers, J.C., Bittner, A.C., Zaklad, A.L., and Christ, R.E., 1992. Comparison of four subjective workload rating-scales. *Human Factors*, 34 (4), 429–439.

Hjortskov, N., Rissen, D., Blangsted, A.K., Fallentin, N., Lundberg, U., and Sogaard, K., 2004. The effect of mental stress on heart rate variability and blood pressure during computer work. *European Journal of Applied Physiology*, 92 (1–2), 84–89.

Hochberg, L.R., Serruya, M.D., Friehs, G.M., Mukand, J.A., Saleh, M., Caplan, A.H., Branner, A., Chen, D., Penn, R.D., and Donoghue, J.P., 2006. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, 442 (7099), 164–171.

Kennedy, P.R., Bakay, R.A., Moore, M.M., Adams, K., and Goldwaithe, J., 2000. Direct control of a computer from the human central nervous system. *IEEE Transactions on Rehabilitation Engineering*, 8 (2), 198–202.

Kubler, A., Nijboer, F., Mellinger, J., Vaughan, T.M., Pawelzik, H., Schalk, G., McFarland, D.J., Birbaumer, N., and Wolpaw, J.R., 2005. Patients with ALS can use sensorimotor rhythms to operate a brain–computer interface. *Neurology*, 64 (10), 1775–1777.

Leuthardt, E.C., Schalk, G., Wolpaw, J.R., Ojemann, J.G., and Moran, D.W., 2004. A brain–computer interface using electrocorticographic signals in humans. *Journal of Neural Engineering*, 1 (2), 63–71.

Parasuraman, R. & Hancock, P.A., 2001. Adaptive control of mental workload. *In*: P.A. Hancock and P.A. Desmond, eds. *Stress, workload, and fatigue*. Mahwah, NJ: Lawrence Erlbaum Associates, 305–320.

Riccio, A., Leotta, F., Bianchi, L., Aloise, F., Zickler, C., Hoogerwerf, E.J., Kubler, A., Mattia, D., and Cincotti, F., 2011. Workload measurement in a communication application operated through a p300-based brain-computer interface. *Journal of Neural Engineering*, 8 (2), 025028.

Schalk, G., McFarland, D.J., Hinterberger, T., Birbaumer, N., and Wolpaw, J.R., 2004. Bci2000: a general-purpose brain–computer interface (BCI) system. *IEEE Transactions on Biomedical Engineering*, 51 (6), 1034–1043.

Sharbrough, F., Chatrian, G., Lesser, R., Luders, H., Nuwer, M., and Picton, T., 1991. American electroencephalographic society guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, 8 (2), 200–202.

Shorrock, T.H., Mackay, D.J.C., and Ball, C.J., 2005. Efficient communication by breathing. *In*: J. Winkler, M. Niranjan and N. Lawrence, eds. *Deterministic and statistical methods in machine learning*. Berlin: Springer, 88–97.

Smith, M.E., Gevins, A., Brown, H., Karnik, A., and Du, R., 2001. Monitoring task loading with multivariate EEG measures during complex forms of human–computer interaction. *Human Factors*, 43 (3), 366–380.

Vitense, H.S., Jacko, J.A., and Emery, V.K., 2003. Multimodal feedback: an assessment of performance and mental workload. *Ergonomics*, 46 (1–3), 68–87.

Ward, D.J., Blackwell, A.F., and Mackay, D.J.C., 2002. Dasher: a gesture-driven data entry interface for mobile computing. *Human–Computer Interaction*, 17 (2–3), 199–228.

Ward, D.J. and Mackay, D.J.C., 2002. Artificial intelligence: fast hands-free writing by gaze direction. *Nature*, 418 (6900), 838–838.

Wickens, C.D. and Hollands, J.G., 2000. *Engineering psychology and human performance*. 3rd ed. Upper Saddle River, NJ: Prentice Hall.

Wills, S.A. and Mackay, D.J.C., 2006. Dasher – an efficient writing system for brain–computer interfaces? *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14 (2), 244–246.

Wolpaw, J.R., Birbaumer, N., Heetderks, W.J., Mcfarland, D.J., Peckham, P.H., Schalk, G., Donchin, E., Quatrano, L.A., Robinson, C.J., and Vaughan, T.M., 2000. Brain–computer interface technology: a review of the first international meeting. *IEEE Transactions on Rehabilitation Engineering*, 8 (2), 164–173.

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

Wolpaw, J.R. and Mcfarland, D.J., 2004. Control of a two-dimensional movement signal by a noninvasive brain–computer interface in humans. *Proceedings of National Academy of Sciences USA*, 101 (51), 17849–17854.