An evaluation of mental workload with frontal EEG

So, Winnie K. Y.; Wong, Savio W. H.; Mak, Joseph N.; Chan, Rosa H. M.

Published in:
PLoS ONE

Published: 01/04/2017

Document Version:
Final Published version, also known as Publisher’s PDF, Publisher’s Final version or Version of Record

License:
CC BY

Publication record in CityU Scholars:
Go to record

Published version (DOI):
10.1371/journal.pone.0174949

Publication details:

Citing this paper
Please note that where the full-text provided on CityU Scholars is the Post-print version (also known as Accepted Author Manuscript, Peer-reviewed or Author Final version), it may differ from the Final Published version. When citing, ensure that you check and use the publisher's definitive version for pagination and other details.

General rights
Copyright for the publications made accessible via the CityU Scholars portal is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights. Users may not further distribute the material or use it for any profit-making activity or commercial gain.

Publisher permission
Permission for previously published items are in accordance with publisher’s copyright policies sourced from the SHERPA RoMEO database. Links to full text versions (either Published or Post-print) are only available if corresponding publishers allow open access.

Take down policy
Contact lbscholars@cityu.edu.hk if you believe that this document breaches copyright and provide us with details. We will remove access to the work immediately and investigate your claim.
An evaluation of mental workload with frontal EEG

Winnie K. Y. So¹, Savio W. H. Wong², Joseph N. Mak³, Rosa H. M. Chan¹ *

¹ Department of Electronic Engineering, City University of Hong Kong, Hong Kong, Hong Kong, ² Centre for Brain and Education and Department of Special Education and Counselling, The Education University of Hong Kong, Hong Kong, Hong Kong, ³ NeuroSky Hong Kong Research Laboratory, Hong Kong, Hong Kong

¹ rosachan@cityu.edu.hk

Abstract

Using a wireless single channel EEG device, we investigated the feasibility of using short-term frontal EEG as a means to evaluate the dynamic changes of mental workload. Frontal EEG signals were recorded from twenty healthy subjects performing four cognitive and motor tasks, including arithmetic operation, finger tapping, mental rotation and lexical decision task. Our findings revealed that theta activity is the common EEG feature that increases with difficulty across four tasks. Meanwhile, with a short-time analysis window, the level of mental workload could be classified from EEG features with 65%–75% accuracy across subjects using a SVM model. These findings suggest that frontal EEG could be used for evaluating the dynamic changes of mental workload.

Introduction

The construct of mental workload can be understood as the level of cognitive engagement which has a direct impact on the effectiveness and quality of a learning process [1]. While an optimal level of mental workload facilitates efficient learning, mental overload could negatively affect task performance and result in more errors [2]. An overloaded individual may even exhibit psychological symptoms, such as frustration, stress and depression [3]. Yet, there lacks a real-time measure of mental workload which can help an individual identify the optimal level of mental workload and hence enhance one’s learning performance.

Conventionally, the level of mental workload is assessed through the verbal or written feedback of an individual. However, the reliability of such self-reported measurements depends on the metacognition skills of the individual [4]. In an educational setting, a continuous assessment of student’s cognitive engagement can be used to determine the pace of teaching and enhance the effectiveness of the learning process. Nevertheless, it is a challenging task for a teacher to evaluate the cognitive engagement of 30–40 students in a typical classroom setting. Although a teacher can evaluate the learning performance of students based on their coursework and examination (i.e., offline assessment), the immediate need of students during the learning process may not be addressed due to the lack of a real-time assessment of mental workload.
Department of Special Education and Counselling at the Education University of Hong Kong. Any opinions, findings, conclusions or recommendations expressed in this material/event (or by members of the project team) do not reflect the views of the Government of the Hong Kong Special Administrative Region, the Innovation and Technology Commission or the Panel of Assessors for the Innovation and Technology Support Programme of the Innovation and Technology Fund. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The funder provided support in the form of salaries for author JNM, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of this author are articulated in the ‘author contributions’ section.

Competing interests: NeuroSky is the industry co-applicant with the project PI, Dr. Rosa Chan, in two research grants from the Innovation and Technology Fund (ITF) provided by the Innovation and Technology Commission of Hong Kong. NeuroSky manufactures the mobile EEG device used in the present study. The first author, Ms. Winnie So has been supported by the ITF grants during the project. One of the co-authors, Dr. Joseph Mak, was a full-time employee of NeuroSky and had a role in the study conceptualization, data collection, and preparation of manuscript. Dr. Savio Wong and Dr. Rosa Chan did not receive any financial benefit from Neurosky or other company with potential conflicts of interest. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

To close the feedback loop in the teaching and learning system, researchers have looked into the use of cutting edge technologies for real-time evaluation of learning performance. For instance, wearable and ambient sensors were used to collect the external environmental information, such as location, surrounding temperature, and people in contact, and provide contextual data in supporting reflective learning of employees in a workplace setting [5]. In a study of behavioral engagement, Liu et al. reported that the writing performance of participants was benefited from the feedback of a learning analytic system which determines the level of engagement based on the intermediate states of document development and how the document is modified [6]. However, many of these technologies are task-specific or bounded by the task nature and characteristics.

Recent research has looked into the use of physiological responses to quantify individual mental workload. From animal experiment such as using invasive electrode [7], to human experiment using non-invasive device. Kapoor et al. used several body sensors, included eye tracking, mouse sensitivity, skin conductance and chair pressure, to estimate the mental workload of an individual with an accuracy of 80% [8]. Studies have also used EEG technologies to determine mental workload based on brain activities. Hogervorst et al. used traditional multi-channel EEG setup to examine the mental workload of 2-minutes period and offered a high classification accuracy (>80%) [9]. So et al. correlate the EEG signal to muscle EMG single to investigate motor performance [10]. Nevertheless, the setup of traditional EEG with wet electrodes requires at least 30–60 minutes. Such setting bounds the usage of conventional EEG in a controlled environment, like research laboratory.

Recently, a range of mobile EEG systems, which only have a few electrodes channels and transmit the recorded neural signal to a computer wirelessly, have been developed to measure brain activities outside the laboratory setting [11–13]. For example, Wong et al. examined the frontal EEG spectra associated with motor acquisition task using a single channel wireless EEG system [14]. Researchers have also used the dry sensor EEG system to develop a neuro-feedback training program for children with Attention Deficit Hyperactivity disorder [12]. Furthermore, mobile EEG has been used in developing Brain-Computer Interface (BCI) for entertainment [15, 16]. Nevertheless, the potential of quantifying mental workload with a mobile EEG system has yet to be explored.

In this study, we aim to examine the feasibility in developing a bio-marker of mental workload based on the frontal activities measured by a mobile single channel EEG system. Previous studies have demonstrated that EEG signals, in particular alpha and theta activities, has a close relationship with cognitive performance and mental effort [17]. In a memory study, Raghayachan et al. reported that event-related Theta activity increases with memory load and decreases sharply at the end of the task [18]. However, these findings were obtained from conventional multi-channel EEG with wet electrodes. To investigate how frontal EEG signals collected from a dry sensor may vary with changes of mental workload, four cognitive and motor tasks (i.e. arithmetic operation, lexical decision, mental rotation and finger tapping task) with different level of difficulties were used to elicit a dynamic change of mental workload in this study. Based on the findings of traditional EEG studies, we hypothesized that the level of mental workload can be distinguished based on the alpha and theta activities which are collected at the frontal cortex with a single channel dry sensor EEG system. More specifically, we hypothesized that, relative to tasks with low level of difficulty, higher event-related theta activities would be observed in tasks with high level of difficulty. We have also identified the key EEG spectral feature associated with mental workload and explored the feasibility to classify different levels of mental workload from EEG features using Support Vector Machine (SVM).
Method

Experimental design

This study consisted of four different cognitive and motor tasks, namely arithmetic operation, finger tapping, mental rotation and lexical decision task which were programmed with Matlab Psychtoolbox [19]. Participants were asked to complete three difficulty levels in the order: low, medium, high for each task. In each difficulty level, there are 25 trials grouped into 5 sessions (i.e. 5 trials per session). In total, each participant had to complete 75 trials of each task. The maximum duration for each trial was 2.5s. After each session, participants gave a subjective mental workload rating on a continuous SMEQ 0–150 range questionnaire [20] see Fig 1. The order of tasks was counterbalanced across participants. The whole experiment lasted for one hour.

Arithmetic (Simple calculation task). In this task, subjects were told to determine the correctness of arithmetic equations showing on the computer screen and responded with button pressing—Left arrow for correct and Right arrow for wrong. Three different difficulty levels were applied in this task—low: single-digit addition, medium: double-digit addition or subtraction with carry set and high: mixed arithmetic operations. Example equations for each difficulty level are shown in Table 1.

Finger tapping (Visual-motor coordination task). In this task, subjects were asked to follow the pattern presented on computer screen and perform specific finger tapping pattern on a keyboard. Their wrist and arm maintained stationary, with their fingers other than the thumbs ready on FGHJ buttons and ASDFJKL buttons for single-hand task and two-hand task respectively. Three difficulty levels were established in the task—low: single hand single finger, medium: single hand multiple fingers and high: two hands multiple fingers. For medium and high level trials, subjects were instructed to press all the keys at the same time instead of one-by-one. Fig 2 showed the finger positions and examples of trials.

Mental rotation (Visual-spatial task). In the mental rotation task [21], subjects were asked to compare a pair of figures presented on computer screen and determine if they were

---

Table 1. Example questions of three difficulty levels in the arithmetic task.

<table>
<thead>
<tr>
<th>Low</th>
<th>Answer</th>
<th>Medium</th>
<th>Answer</th>
<th>High</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>$3 + 4 = 6$</td>
<td>F</td>
<td>$13 + 18 = 51$</td>
<td>F</td>
<td>$7 \times 3 - 12 = 19$</td>
<td>F</td>
</tr>
<tr>
<td>$4 + 2 = 6$</td>
<td>T</td>
<td>$61 - 12 = 49$</td>
<td>T</td>
<td>$63/3 - 32 = -19$</td>
<td>T</td>
</tr>
<tr>
<td>$2 + 7 = 4$</td>
<td>F</td>
<td>$52 - 67 = -8$</td>
<td>F</td>
<td>$78 + 11 \times 3 = 71$</td>
<td>F</td>
</tr>
<tr>
<td>$3 + 6 = 9$</td>
<td>T</td>
<td>$-11 + 28 = 17$</td>
<td>T</td>
<td>$(21 - 13) \times 4 = 32$</td>
<td>T</td>
</tr>
<tr>
<td>$5 + 2 = 8$</td>
<td>F</td>
<td>$-13 - 56 = -60$</td>
<td>F</td>
<td>$6 \times 6 - 79 = -23$</td>
<td>F</td>
</tr>
</tbody>
</table>

https://doi.org/10.1371/journal.pone.0174949.t001
the same object. The pair of figures could be the same object with different rotation, mirror images or different objects. Easy level contained 6 squares in 2D plane, whereas medium and high level were in 3D space with 6 and 9 cubes respectively. Examples trials were shown in Fig 3.

Lexical decision (Linguistic task). In the lexical decision task [21], subjects were asked to identify whether the stimulus presented was a real English word or a pseudo-word. Different difficulty levels were applied and created by varying the word usage frequency [22], word length and part of speech (Low: concrete noun, Medium/High: noun, adjective, verb). The pseudo-words were generated using the Wuggy software [23]. Examples of each difficulty level are given in Table 2.

A summary of all experimental tasks are in Table 3 and listing out the variations of task content and difficulty level.

Data collection

Twenty healthy participants (age: 22± 0.71; male/female: 6/14) were recruited in this study. All participants are university students, have normal or corrected-to-normal vision and have no history of neurological or psychological disorder. Participants gave written informed consent to participate by signing the consent form. The experimental procedures were reviewed and approved by the ethics committee at the City University of Hong Kong.

The experiment was conducted in a recording studio room. Participants were told to sit still and relax throughout the experimental. At the beginning of the experiment, 20 seconds of eyes-opened resting state EEG were collected as baseline. Then, participants completed a practice session and became familiarized with the tasks. Data collection began when the participants reached 75% accuracy in 10 consecutive trials or finished 30 trials for each difficulty level. Frontal brain signal was collected at Fp1 channel using a single-channel wireless EEG
device (Neurosky MindWave Mobile headset) at 512Hz sampling rate. The ground and
referencing was at the lobule. Behavioral data including response time and accuracy were
recorded in Matlab. Both the experimental tasks and the EEG recording were controlled with a
tablet computer (Surface Pro 3) connected to an external keyboard (Dell L30U).

Data processing

Normalization of subjective rating. To reduce between-subject effect, subjective rating
was scaled within subject across tasks into [0, 1] range using Eq 1.

\[ X' = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \]  

(1)

The normalized between-subjects ratings were then grouped by subjective ratings at 0–33%
(low), 34–65% (medium) and 66–100% (high) quantile shown as Fig 4.

Pre-processing of EEG signal. Data was first detrended, then bandpass filtered at 0.5 to
45Hz using a FIR filter with a 5th order butterworth window. ICA-based method is commonly
used for eye blink and movement artifact removal. Yet ICA requires multi-channel EEG and
demand relatively heavier computation. In this study, we employed a wavelet-based filter to

Table 2. Lexical decision task: Real word and pseudoword used in the three levels of difficulty.

<table>
<thead>
<tr>
<th>Task</th>
<th>Variation</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real Word</td>
<td>Pseudo Word</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Butter</td>
<td>narrator</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Bread</td>
<td>determined</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hair</td>
<td>dependency</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pepper</td>
<td>forecast</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hind</td>
<td>carriage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hnaqi</td>
<td>moderate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eiloul</td>
<td>rehearsal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>phnqa</td>
<td>inhabit</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Butter narrator steadiness</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bread determined anomalous</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hair dependency supposition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pepper forecast expatriate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hind carriage carround erudition erumition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hnaqi moderate moterate negligible nefflible</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eiloul rehearsal reheandal dehydratde dehyllate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>phnqa inhabit inhabit adjudication aggjudicaiton</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

https://doi.org/10.1371/journal.pone.0174949.t002

Table 3. Summary of all the experimental tasks.

<table>
<thead>
<tr>
<th>Task</th>
<th>Content</th>
<th>Variation</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arithmetic</td>
<td>Determine whether the arithmetic equation is correct or not</td>
<td>Digit</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Arithmetic</td>
<td>+</td>
<td>+ -</td>
<td>+ -</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Type</td>
<td>Simple</td>
<td>Simple</td>
<td>Operation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Carry Set</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Finger Tapping</td>
<td>Follow the spotlight to press the button on the keyboard</td>
<td>Hand</td>
<td>Single</td>
<td>Single</td>
<td>Both</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fingers</td>
<td>Single</td>
<td>Multi</td>
<td>multi</td>
</tr>
<tr>
<td>Mental Rotation</td>
<td>Judge whether the pair of figures is the same object or not</td>
<td>Dimension</td>
<td>2D</td>
<td>3D</td>
<td>3D</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Number of cube</td>
<td>6</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>Lexical Decision</td>
<td>Determine whether the presented word is a real English word or not</td>
<td>Word Frequency Rank</td>
<td>1–5000</td>
<td>5000–10000</td>
<td>&gt;10000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Number of syllable</td>
<td>1–2</td>
<td>3–4</td>
<td>&gt;3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>word length</td>
<td>4–6</td>
<td>8–10</td>
<td>&gt;9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>word type</td>
<td>non word</td>
<td>pseudo</td>
<td>pesudo</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Part of Speech</td>
<td>concrete n</td>
<td>n, v, adj</td>
<td>n, v, adj</td>
</tr>
</tbody>
</table>

Remark: n-noun v-verb adj-adjustive

https://doi.org/10.1371/journal.pone.0174949.t003
the real-time signal channel EEG data to remove eye blink and movement related artifacts [24]. Signal was then segmented into 2.5s epoch according to each trial starting time.

**Time-frequency analysis.** To compute the Time-Frequency matrix (TF Matrix), each trial was segmented with 437.5ms signal window (lower quantile response time of the finger tapping task (low level), 224 data points), with window slide every 31.25ms (16 data points). Welch’s power spectra were calculated over 50% overlapped 2s hamming windows. A trial was rejected if the response time was smaller than 437.5s. Instant Relative power (IRP) is defined as a function of the normalized instantaneous frequency and normalized baseline power in the following form.

$$\text{IRP}_i = \ln \left( \frac{R_{P_i}}{R_{P_b}} \right)$$

where $R_{P_i}(f_1, f_2)$ = $\frac{P_i(f_1, f_2)}{P_i(0.5, 45)}$  

where $(f_1, f_2)$ is the frequency range, $i$ is the instant time interval and $b$ is the baseline period.

---

**Fig 4. Subjective rating result.** The normalized rating of each task. Blue, red, and green indicates low, Medium, and high difficulty levels respectively. The horizontal lines represent the 0.33 and 0.66 quantile respectively.

https://doi.org/10.1371/journal.pone.0174949.g004
We computed the EEG frequency band power using two sets of EEG band distributions: traditional EEG frequency band definitions (theta: 4–8Hz, alpha I: 8–11Hz, alpha II: 11–14Hz, beta I: 14–25Hz, beta II: 25–35Hz, gamma I: 35–40Hz, gamma II: 40–44Hz) and individualized frequency band distribution from individual alpha frequency (IAF) [17, 25]. Concerning about individual differences, theta and alpha ranges were defined from the baseline IAF with the following formula [26]. FBIW theta, alpha I and alpha II were defined as (IAF – 4) to (IAF – 2), (IAF – 2) to (IAF) and (IAF) to (IAF+2) respectively. IBIW theta, alpha I and alpha II were defined as (IAF × 0.6) to (IAF × 0.8), (IAF × 0.8) to (IAF) and (IAF) to (IAF × 1.2). We focused on the time course between trial start and key pressing. The average power spectra across the time series were also computed.

**Visualization of task similarity.** Linear and Kernel Discriminate Analysis (LDA, KDA) [27] were implemented to investigate the similarity among the tasks separately. LDA is a supervise dimensionality reduction method which preserves the class discriminatory information. It tried to find the good linear subspace to project the input data and maximize the separation among classes. KDA extend KDA to nonlinear by transforming the space. The kernel operators used is Gaussian kernel.

Both within subject and across subject analyses had been performed. The input features were the EEG power bands and the label information was the task type. We visualized the first two component of disseminate analysis on the x-y plane. The axes are dimensionless after LDA/KDA transformation, which implies no physical meaning.

**T-Statistics analysis on time-frequency analysis.** Here we investigated the oscillatory activities during the task at different difficulty levels. TF matrix from the time frequency analysis was rescaled with respect to completion rate. Each pixel on TF matrix underwent the minimum t statistic for comparison.

After computing the TF analysis, we calculated the t-value [28] for each time-frequency point using 1-sample t-test within each difficulty level for each subject. Next, pairwise comparison between the t-maps of low and high level was performed for the study of largest task differentiability. At last, we worked out the conjunction by finding the minimum t-value between t-maps from pairs of tasks to investigate the common features.

**Classification of difficulty level.** Support vector machine (SVM) [29] was used to classify EEG data at different task difficulty level. Time averaged power of each frequency band was computed from each 2.5s trial, while both subjective and objective task difficulty level were used as class labels. We used a simple 2-class classification first to distinguish the lowest and highest task difficulty levels. Radial basis function kernel (RBF) was used, and repetitions of 10-fold cross validation was.

**Results**

**Behavioral result**

Repeated measure ANOVA (Table 4) and pairwise analysis (Table 5 and Fig 5) were performed on EEG features and the behavioral performance measures, namely reaction time, missing rate, accuracy and subjective rating within each task. All the behavioral measures exhibited a significant main effect of Task and Difficulty Level (p < 0.001). Significant interaction effect between task and difficulty level was observed in the subjective rating (p < 0.01), response time, missed rate and accuracy (p < 0.001). Post hoc multiple comparisons analysis was conducted to compare each pair of levels. The behavioral data also showed that when the difficulty level increased, the response time and missing rate increased and the accuracy decreased across four tasks. The Spearman correlation between objective level and the subjective rating also showed a significant positive correlation for all four tasks (p < 0.05; Table 6). These results
Table 4. Repeated measure ANOVA of EEG feature and behavioral responses. The within-subjects factors are task (4 levels) and difficulty (3 levels), ***p < 0.001, **p < 0.05 and *p < 0.01.

<table>
<thead>
<tr>
<th>Task</th>
<th>Level</th>
<th>Task* Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F(3, 17)</td>
<td>p</td>
</tr>
<tr>
<td>Subjective Rating</td>
<td>10.069</td>
<td>***</td>
</tr>
<tr>
<td>Normalized Rating</td>
<td>19.162</td>
<td>***</td>
</tr>
<tr>
<td>Response Time</td>
<td>313.797</td>
<td>***</td>
</tr>
<tr>
<td>Missed</td>
<td>22.754</td>
<td>***</td>
</tr>
<tr>
<td>Accuracy</td>
<td>118.634</td>
<td>***</td>
</tr>
<tr>
<td>4–8Hz (θ)</td>
<td>8.331</td>
<td>**</td>
</tr>
<tr>
<td>8–11Hz (α1)</td>
<td>4.061</td>
<td>***</td>
</tr>
<tr>
<td>11–14Hz (α2)</td>
<td>5.046</td>
<td>**</td>
</tr>
<tr>
<td>14–25Hz (β1)</td>
<td>1.687</td>
<td>*</td>
</tr>
<tr>
<td>25–35Hz (β2)</td>
<td>2.441</td>
<td>*</td>
</tr>
<tr>
<td>35–40Hz (γ1)</td>
<td>1.58</td>
<td>0.782</td>
</tr>
<tr>
<td>40–44Hz (γ2)</td>
<td>2.42</td>
<td>*</td>
</tr>
</tbody>
</table>

An evaluation of mental workload with frontal EEG implied that the task design successfully created different levels of difficulty for each task. Meanwhile, the difficulty levels between the tasks were not necessarily the same. Because a consistent significant difference in the behavioral measures is observed only between the comparison of low and high difficulty trials (Table 5), therefore the analysis of the EEG data focused on the comparison of the low and high difficulty trials.

TF analysis: Dynamic change in trials

Fig 6 shows the time-frequency analysis of each task. The baseline has been taken into account using Eq 2. The entire trial from the start point to 2.5s is shown and the vertical line indicates the minimum time response within all the trials. Comparing the low and high difficulty levels,

Table 5. Statistics of behavioral responses. RT: Response Time(s), Missed: Number of Missed trial/Total Number of Trial, Acc: Accuracy Rate (excluded missed trial), Rating: SMEQ subjective rating scaled to 0–1 range, m: mean value, **p < 0.001 and *p < 0.05.

<table>
<thead>
<tr>
<th>Task</th>
<th>Low Mean</th>
<th>SD</th>
<th>Medium Mean</th>
<th>SD</th>
<th>High Mean</th>
<th>SD</th>
<th>p</th>
<th>m&lt;sub&gt;low&lt;/sub&gt; - m&lt;sub&gt;medium&lt;/sub&gt;</th>
<th>m&lt;sub&gt;low&lt;/sub&gt; - m&lt;sub&gt;high&lt;/sub&gt;</th>
<th>m&lt;sub&gt;medium&lt;/sub&gt; - m&lt;sub&gt;high&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cal</td>
<td>RT</td>
<td>0.96</td>
<td>0.11</td>
<td>1.80</td>
<td>0.21</td>
<td>1.97</td>
<td>0.16</td>
<td>&lt;0.001</td>
<td>0.05</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>Rating</td>
<td>21.44</td>
<td>0.11</td>
<td>41.70</td>
<td>0.15</td>
<td>58.00</td>
<td>0.19</td>
<td>&lt;0.001</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Missed</td>
<td>0.15</td>
<td>0.02</td>
<td>2.90</td>
<td>0.11</td>
<td>7.35</td>
<td>0.13</td>
<td>&lt;0.001</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Acc</td>
<td>0.95</td>
<td>0.03</td>
<td>0.58</td>
<td>0.03</td>
<td>0.47</td>
<td>0.03</td>
<td>&lt;0.001</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Fin</td>
<td>RT</td>
<td>0.58</td>
<td>0.06</td>
<td>0.79</td>
<td>0.13</td>
<td>1.34</td>
<td>0.18</td>
<td>&lt;0.001</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Rating</td>
<td>19.61</td>
<td>0.10</td>
<td>32.02</td>
<td>0.15</td>
<td>48.45</td>
<td>0.18</td>
<td>&lt;0.001</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Missed</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.20</td>
<td>0.02</td>
<td>&lt;0.05</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Acc</td>
<td>0.99</td>
<td>0.02</td>
<td>0.92</td>
<td>0.02</td>
<td>0.72</td>
<td>0.02</td>
<td>&lt;0.001</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Rot</td>
<td>RT</td>
<td>1.66</td>
<td>0.19</td>
<td>1.86</td>
<td>0.17</td>
<td>1.94</td>
<td>0.11</td>
<td>&lt;0.001</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Rating</td>
<td>41.16</td>
<td>0.20</td>
<td>53.63</td>
<td>0.18</td>
<td>67.33</td>
<td>0.17</td>
<td>&lt;0.001</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Missed</td>
<td>1.90</td>
<td>0.09</td>
<td>4.00</td>
<td>0.13</td>
<td>5.40</td>
<td>0.14</td>
<td>&lt;0.001</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Acc</td>
<td>0.72</td>
<td>0.03</td>
<td>0.58</td>
<td>0.03</td>
<td>0.44</td>
<td>0.03</td>
<td>&lt;0.001</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Lin</td>
<td>RT</td>
<td>0.70</td>
<td>0.08</td>
<td>1.14</td>
<td>0.21</td>
<td>1.40</td>
<td>0.26</td>
<td>&lt;0.001</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Rating</td>
<td>20.08</td>
<td>0.11</td>
<td>29.46</td>
<td>0.20</td>
<td>41.76</td>
<td>0.23</td>
<td>&lt;0.001</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Missed</td>
<td>0.00</td>
<td>0.00</td>
<td>0.30</td>
<td>0.03</td>
<td>0.70</td>
<td>0.04</td>
<td>&lt;0.05</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Acc</td>
<td>0.98</td>
<td>0.02</td>
<td>0.83</td>
<td>0.02</td>
<td>0.69</td>
<td>0.02</td>
<td>&lt;0.001</td>
<td>0.15</td>
<td>0.30</td>
</tr>
</tbody>
</table>

https://doi.org/10.1371/journal.pone.0174949.t005

https://doi.org/10.1371/journal.pone.0174949.t006
Fig 5. Power of different EEG frequencies in four cognitive tasks. 1, 2 and 3 are low, medium and high difficulty level respectively. Blue line indicated the medium level. Red line means it is significantly different from medium level and gray line means not significant.

https://doi.org/10.1371/journal.pone.0174949.g005
it can be seen there was a relative increase in the frequency components around 20–35Hz when the subject was executing the task. Following completion, frequency components in the 20–30Hz range were suppressed, while the theta range (4–8Hz) increased. We also observed that the response time for the low level is shorter than that for the high level, and we are actually comparing the period during task and after task. This observation represents the change between engagement during task and relaxation after activity.

Medium and high difficulty levels of arithmetic operation and mental rotation require longer response time, we observed the power bursted across the time course in theta and alpha I range. As for the low difficulty level, the shorter the time response, the larger suppression in beta and gamma observed.

KDA analysis: Identification of task similarity

In Figs 7 and 8, the results of KDA transformation with the task class label are presented. Each point presents a trial (Blue: Arithmetic operation, Red: Finger tapping, Green: Mental Rotation, Black: Lexical decision). Different kernel and parameter values were tested. Fig 7A which shows the four tasks overlapping with each other, is transformed by linear discriminate analysis. In Fig 7B, a Gaussian Kernel is used with different parameter values were explored and finally, parameter = 10 showed better results as several clusters are clearly observed. This implies the non-linearity of EEG dynamics, as such that non-linear analysis provides a better approach to characterize the properties.

The KDA transformation for each subject are presented in Fig 8. The distance between pairs of clusters implies how different brain signal changes are among tasks. Shorter distances mean that they were more similar. Generally, arithmetic operation and mental rotation clusters overlapped in most of the subjects, except subjects 3 and 13. On the other hand, finger tapping is more separable from other tasks. From this transformation, we observed the similarity level
of the brain signal characteristic among the four tasks. This result demonstrates a possibility to group the cognitive tasks together if we want to build a more general model in the future.

Changes in band power: Spectral characteristic

Fig 9 presents the frequency power change between high and low level of difficulty. The overall changes in arithmetic operation and finger tapping are more significant. An increase in theta

---

Fig 7. LDA and KDA results. A: LDA without and kernel. B: KDA with all the subjects trials, Gaussian kernel with parameter = 10.

https://doi.org/10.1371/journal.pone.0174949.g007

Fig 8. KDA results for each subject. C: KDA within each subject using Gaussian Kernel.

https://doi.org/10.1371/journal.pone.0174949.g008
power is a common feature among all the tasks. This is consistent with previous findings that theta is related to workload demand [17, 30].

Previous KDA transformation suggested that arithmetic operation and mental rotation are more similar. The frequency power analysis supported this in that they shared the same trend in theta, IAF alpha I and both beta I and II frequency ranges.

Working memory consists of a phonological loop and a visual spatial sketchpad [31]. Among the four tasks, finger tapping and mental rotation rely heavily on the visuospatial loop whereas lexical decision and arithmetic operation involve mainly the phonological loop. Yet, previous research [32] indicated that visual spatial skill also involved in arithmetic operation causing it be more similar with mental rotation. Moreover, both of them require a problem solving process instead of pure long-term memory retrieval or motor coordination.

Fig 9 has showed different patterns of frequency power change which matched with other research groups’ studies. For example, increase of Beta II in lexical decision task referred to the orthographic and semantic difference in the choice of vocabulary [33]. Meanwhile, the gamma increase related to the vocabulary recollection in high difficulty level rather than only familiarity [34]. As for the finger tapping task, the increase in IAF theta, IAF alpha I and II matched our previous study about the motor skill acquisition by using a mirror drawing experiment [35]. These band powers showed that they positively correlated with perceived difficulty level of the task.

T-statistic of TF analysis

Time-frequency T-map analysis allowed us to observe the dynamic change in frequency range across the time course. Fig 10 shows the time frequency difference between low and high level after the trial re-sampling. Here, we focus on the change in the common feature, theta activities. Although all four tasks showed an increase in theta, the time of occurrence could differ. Arithmetic operation and mental rotation tasks had theta power increase in over the whole trial whereas it only occurred in the middle of the finger tapping and lexical decision task.

Fig 11 is the conjunction between two pairs of tasks and showing the minimum T value. Color in red and blue mean the common synchronization and desynchronization in two tasks respectively. The blank color means both task were in opposite trend. The rightmost graph is the conjunction of all the tasks and we discovered that theta, beta and the IAF analysis have same trend in synchronization or desynchronization in some time points.

Fig 9. Ratio of absolute power change. Frequency power change between high and low level of difficulty in each task. P-value <0.05.

https://doi.org/10.1371/journal.pone.0174949.g009
SVM classification: Real-time analysis model

As a practical application, ability to predict the workload using the frequency power feature by classification technique is necessary. Preliminary result in Fig 5 shows that EEG features from the medium level of difficulty always overlap with either low or high level. To simplify the classification model, we have studied the cases with the lowest and highest difficulties. Table 7 presents the test sample accuracy in 10-fold cross validation during 2-class SVM classification. The samples were taken from across all the subjects and the accuracy reached greater than 70% in Arithmetic operation, finger tapping and lexical decision. Mental rotation had a relative lower accuracy of around 64%, because of the relatively small difficulty gap between the levels as indicated in Fig 4. This test has demonstrated the potential real application of short term prediction of the mental workload using a single channel EEG device.

Discussion

This study aims to develop an EEG-based mental workload-detection application by building a generalized model for four different cognitive and motor tasks. Our findings showed that
the frontal theta activity is a common feature across these tasks. This result is consistent with previous studies that theta activities increase with the level of mental effort [17, 18]. Meanwhile, the correlation of mental workload level and other frequency bands is task-dependent [3, 36].

With a 2.5s analysis window size, the accuracy of mental workload classification could reach 65%–75%, which is slightly higher than other EEG studies with around 60% accuracy [36, 37]. It might be related to three factors: intra subject variance, task duration and the selection of EEG channel.

First, previous studies have reported that individually adjusted frequency bands are useful for the analysis of event-related potential [26]. The individual differences in alpha peak were evaluated by computing the IAF-defined power value from baseline. Second, unlike previous study which took long measurement and had subject to give an overall rating afterwards, the current experiments was composed of very brief trials and required participants to feedback on the subjective mental workload level immediately after each session. The short-time analysis window indicated the moment when the subjects were engaging in the task. Indeed, as Dai et al. suggested that using all channel in the analysis might cause large variance and result to a poorer classification of the task [38]. Active EEG channels should be selected whereas unrelated channels should be discarded in order to improve the accuracy. Our findings illustrated that EEG signals collected from a single-channel dry sensor at Fp1 provide sufficient information in generating a reasonable bio-marker of mental workload.

This study takes the first step in bringing laboratory research technique to real life application with the latest mobile EEG technology. Our findings indicated that event-related frontal EEG theta frequency band power is a common feature of mental workload across different cognitive and motor tasks. The advantage of this model is the ability to detect short term

| Objective | Cal | 75.40% | Fin | 76.00% | Rot | 60.40% | Lin | 74.42% |
| Subjective | Cal | 73.91% | Fin | 73.31% | Rot | 64.91% | Lin | 73.03% |

https://doi.org/10.1371/journal.pone.0174949.t007
mental workload in real time. Due to the limited sample size of the current study, counterbal-
ance was performed on the order of the four different tasks but the difficulty level. Although
the current design resembles to the real-world practice that we usually complete the relatively
easy task before heading to the difficult one, we could not rule out the possible practice effect
across the three difficulty levels within each condition. The practice effect, if any, should
improve the performance of the later trials within each condition (i.e. the high difficulty trials)
and hence reducing the behavioral difference between the low and high difficulty trials. However,
even under a potential practice effect, robust differences in the behavioral measures were
still observed between the low and high difficulty trials. On the other hand, as the order of the
four conditions is counterbalanced across subjects, the low, medium and high difficulty trials
of the four conditions were interleaved within the experiment which reduced the chance of
having the effect of fatigue and boredom biasing the data of a particular difficulty level. Future
studies with a larger sample size may investigate how EEG activities may be affected if the
order of difficulty level is changed.

With the goal of bridging the gap between fundamental neuroscience research and real-
world application (e.g. evaluating the real-time mental workload of students in a classroom set-
ting), our study has provided a proof of concept in using single channel frontal EEG for short
term mental workload detection. The EEG-based workload detection provides alternative
approach to evaluate the study progress of students by monitoring the physiological response.
The application can also be extended to self-study outside the classroom. When students are
spelling vocabularies or solving an arithmetic problem, the short-term workload detection can
provide feedback for self-evaluation. This study also shed new light on the possibility in devel-
oping a biomarker for quantifying mental workload and providing a real-time feedback on the
dynamic change of mental workload.

Supporting information
S1 Dataset. EEG and behavioral data of 20 subjects.
(ZIP)

Author Contributions
Conceptualization: SHWJ NRM RHMC.
Data curation: WKYS SHWJ NRM RHMC.
Formal analysis: WKYS.
Funding acquisition: SHWJ NRM RHMC.
Investigation: WKYS.
Methodology: WKYS.
Project administration: RHMC.
Resources: SHWJ NRM RHMC.
Software: WKYS.
Supervision: SHWJ NRM RHMC.
Validation: RHMC.
Visualization: WKYS.
Writing – original draft: WKYS.
Writing – review & editing: SWHW JNM RHMC.

References


