Functional Near-Infrared Spectroscopy Adaptive Cognitive Training System (FACTS) for Cognitive Underload and Overload Prevention: A Feasibility Study

WEI CHUN UNG\textsuperscript{1}, (Graduate Student Member, IEEE), FABRICE MERIAUDEAU\textsuperscript{1}, (Member, IEEE), MASASHI KIGUCHI\textsuperscript{2}, AND TONG BOON TANG\textsuperscript{1}, (Senior Member, IEEE)

\textsuperscript{1}Centre for Intelligent Signal and Imaging Research (CISIR), Department of Electrical and Electronic Engineering, Universiti Teknologi PETRONAS, Perak 32610, Malaysia
\textsuperscript{2}Center for Exploratory Research, Research and Development Group, Hitachi, Ltd., Tokyo 185-8601, Japan

Corresponding author: Tong Boon Tang (tongboon.tang@utp.edu.my)

This work was supported in part by Yayasan UTP (cost centre: 0153AA-E99), the Ministry of Higher Education, Malaysia, under grant FRGS/1/2018/TK04/UTP/02/10, and the Higher Institution Centre of Excellence (HiCoE) Scheme awarded to the Centre for Intelligent Signal and Imaging Research (CISIR).

ABSTRACT Mismatch between mental workload and working memory capacity can cause mental underload or overload. Adopting the Yerkes–Dodson law as the framework, functional near-infrared spectroscopy adaptive cognitive training system (FACTS) has been developed, whereby the mental workload shall never exceed an individual’s capacity, to prevent those unintended conditions. It works by monitoring mental workload in real time and performing dynamic difficulty adjustment accordingly. The feasibility study involved thirty-seven healthy participants undergoing mental arithmetic task with and without FACTS. Without FACTS, the participants not only showed higher NASA Task Load Index scores but also poorer task performance and a significant drop in DLPFC activation towards the end of the task, signifying more severe mental overload. Conversely, they continued to exhibit manifestation of productive learning with FACTS despite showing early signs of mental overload. The study results demonstrated that it is feasible to implement the concept of FACTS. The actual gains from cognitive training will be investigated in future longitudinal study.

INDEX TERMS Functional near-infrared spectroscopy, dynamic difficulty adjustment, cognitive training, mental workload, cognitive underload, cognitive overload.

I. INTRODUCTION

Working memory is defined as the set of cognitive processes in charge of temporary storage and manipulation of information to execute more complex tasks [1]. Dubbed as the engine of learning, working memory has enormous implications in learning. Nevertheless, the working memory capacity, which refers to the limited amount of information an individual can retain at a given time, plays an influential role in learning. Therefore, it is of vital importance to keep the workload to a suitable level so that a larger proportion of the working memory capacity can be devoted to effective cognitive engagement and learning.

Several studies have attempted to improve working memory through cognitive training. One of them has observed that trained individuals could remember more digits [2] while another study has reported enhancement in cognitive functions like speed, reasoning and memory [3]. Cognitive training in the form of arithmetic training has been proven to be successful in boosting mathematical abilities not only in healthy young adults [4] but also in dementia patients [5]. In addition, mental arithmetic has been proposed as an effective form of cognitive training to train working memory [6]. However, varying training difficulty can result in different
amount of mental workload, potentially impeding performance and learning [7]. Mental workload unmatched by working memory capacity can lead to unintended mental underload or overload, whereby the learners may lose focus, learn less, and become less productive and more error-prone [8], [9]. Ideally, the amount of mental workload should fall just beyond the capacity limits of the learners [10] so that they are challenged to invest mental effort in the learning process. Nevertheless, overloading the capacity can give rise to detrimental mental overload [11], in which the gains of cognitive training may be reduced or even eliminated [12].

To avoid mismatch between mental workload and working memory capacities, personalization is the key as it allows learners to train at their own pace [13]. Nonetheless, defining the right amount of mental workload remains a challenge. One potential solution is the implementation of dynamic difficulty adjustment (DDA), a technique for adapting the task difficulty in the event of any undesirable state arising from unmatched mental workload. DDA can facilitate learning and boost productivity by tailoring the pace of learning accordingly to the capacity of each individual. Since there has been a widespread application of DDA to maximize engagement in gaming [14], [15], similar approach can possibly be adopted in cognitive training or rehabilitation environment but studies on the subject remain relatively scarce. Therefore, the current paper aims to explore how such design concept can be used in cognitive training.

In comparison to cardiac activity, eye blink, performance and subjective-based measures, neurophysiological metrics offer more promise as input to a DDA system because they: 1) are more sensitive to task difficulty [12]; 2) respond predictably to mental workload [16]; and 3) can provide continuous and unobtrusive monitoring without interrupting the ongoing work [17]. A previous study has observed greater engagement and skill development when neurological biomarkers of anxiety instead of performance were used as the input [15]. Neurophysiological metrics are measurable through the use of neuroimaging techniques such as electroencephalography (EEG) or functional near infrared spectroscopy (fNIRS).

EEG, an electrophysiological approach to records electrical activity in the brain, not only costs less but it also possess outstanding temporal resolution and is able to estimate workload [18]. Nonetheless, EEG has limited spatial resolution [19] and only a small proportion of the raw EEG signals originates from the deeper layers of the brain [20]. EEG is also more prone to artifacts, especially jaw-related motion artifacts [21]. On the other hand, fNIRS is a relatively new neuroimaging modality that monitors brain activity non-invasively through hemodynamic responses [22]. fNIRS possesses a spatial resolution of $\approx 10$ mm as well as reasonable temporal resolution of $\approx 0.1$ seconds [23]. It also offers portability, ease of setup, lenient subject constraints, customizable experimental environment and higher tolerance of motion artifacts [24]. Owing to those advantages, there has been an increasing deployment of fNIRS in recent years [25]–[32] and it is evident that fNIRS is the most suitable neuroimaging modality for DDA applications. Since EEG and fNIRS complement each other in neuroimaging perspective as well as temporal and spatial resolution, they can viably form a superior bimodal DDA. However, the current paper focuses only on fNIRS so hybrid EEG–fNIRS-based DDA remains an interesting avenue for future research.

So far, no previous work has been reported so far on the application of neurophysiological DDA in longitudinal cognitive intervention like training and rehabilitation. The closest ones are EEG studies which have found success in improving cognition in healthy older adults [33] and children with attention deficit hyperactivity disorder [34], but their methodology leaned more towards neurofeedback rather than cognitive training with adaptive difficulty. The participants in both studies learned to regulate their brain activity to be more attentive through feedback from neural correlates of attention sampled during cognitive training. Thus, it will be interesting to see what benefits neurophysiological DDA can bring to cognitive intervention.

The Yerkes–Dodson law dictates that there exists a direct relationship between performance and arousal [35]. Performance can be boosted by increasing the arousal but only up to a certain point; beyond the point where arousal becomes excessive, performance declines. Converging evidence from previous fNIRS studies suggests that the same law can also be applied to hemodynamics [25], [26]. It has been revealed that the bilateral DLPFC in working memory is associated with task demands [36] and that its activity exhibits a similar inverted U-shaped relationship (as suggested by the Yerkes–Dodson law) with increasing demands of a supervisory control task [37]. Given its implications in working memory demands coping and easy accessibility by fNIRS, the DLPFC makes a good feedback source for real-time mental workload monitoring.

An fNIRS–DDA system to aid learners to avoid mental underload and overload, hereinafter referred to as fNIRS adaptive cognitive training system (FACTS), has been developed and was first introduced in a proceedings paper [38]. In contrast to another fNIRS-based DDA system that acts on user state predicted based on a pre-established model [39], FACTS adopts a more relative approach so that not only no model or prior calibration is required but the system itself can also be easily incorporated into any task. FACTS employs fNIRS-based neurophysiological DDA technology and uses the Yerkes–Dodson law as the framework to personalize cognitive training in terms of difficulty i.e., by making the difficulty adaptive. Through a counterbalanced repeated measures design in which the participants performed the same task with and without FACTS, the study examined the feasibility of FACTS using mental arithmetic as an example of cognitive training environment. The primary hypotheses were that the mental workload would be higher as assessed using the NASA Task Load Index (NASA-TLX) and the participants would suffer from more severe mental overload in the absence of FACTS. In a simplified manner, the 4-page proceedings
paper introduces FACTS, reports on the unfinished study (only 25 participants at the time of publication), and draws conclusion solely based on hemodynamics [38]. Meanwhile, the current paper not only describes FACTS comprehensively but also employs a different post hoc analysis, including a side-by-side inspection of behavioral and hemodynamic outcomes, on the now-completed data collected from 37 participants.

II. MATERIALS AND METHODS

A. PARTICIPANTS
The participants were recruited through purposive sampling with specific inclusion criteria: 1) Right-handed; 2) No history of psychiatric or neurological disorder; 3) Have normal or corrected-to-normal vision; 4) Able to communicate in English. A total of 38 healthy right-handed local university students matched for age and gender participated in the feasibility study. The medical research ethics committee of Universiti Kuala Lumpur Royal College of Medicine Perak has consented to the study protocol used in the study (reference number: UniKLR-CMP/MREC/2018/019) and the study was performed in compliance with the Declaration of Helsinki. Informed consent along with demographic information were obtained and the participants were remunerated for their participation. In the end, the data collected from one participant, who did not complete his second session due to a technical glitch, was excluded. The final group of participants was reduced to 19 males and 18 females, with a mean ± standard deviation age of 23.41 ± 3.32 years. All but one underwent the second session at least a week apart from the first (mean ± standard deviation duration between sessions, 8.00 ± 2.60 days). The participants reported similar sleep duration the night before their first and second session, with a mean ± standard deviation sleep of 6.65 ± 0.22 and 6.96 ± 0.21 hours respectively.

B. TASK
Mental arithmetic was used as the cognitive training task. After resting for 15 s, participants were asked to solve as many on-screen four-choice mathematical problems mentally as possible, at their own pace and within their capability, during the following 15-s task period. A 4-button Serial Response Box (Psychology Software Tools, Inc.) was also utilized to collect the participants’ responses. Fig. 1b shows the paradigm for a single task trial. The mathematical problems only involved addition or subtraction or a mix of both.

There were six levels of difficulty, each with a distinctive set of operands: 1) Two single digits; 2) Three single digits; 3) Two double digits; 4) Two single and one double digits; 5) One single and two double digits; and 6) Three double digits. The single-digit operands ranged from 1 to 9 while the three-digit operands lied between 10 and 99. All the operands and answers were non-negative. Besides that, there were no repeating number in all the mathematical problems.

To avoid frustration at carelessness or failing to reach a solution, no answer feedback was presented.

First, a 10-minute pre-task assessment was carried out not only to determine the participants’ optimal level of difficulty but also to sample their baseline performance. The participants sat through mental arithmetic task of all six difficulty levels (three trials per level) in a randomized sequence. In accordance to Yerkes–Dodson Law in the context of hemodynamics, peak brain activation is achieved when the task level at hand is appropriate [37]. An activation consists of an increase in oxygenated hemoglobin (oxy-Hb) and an almost antiphase decrease in deoxygenated hemoglobin (deoxygen-Hb) [40]. In the task, each trial yielded an activation value and the final 18 activation values were averaged down to only one average value per level. For each participant, the level eliciting the largest average value was eventually picked as the participant’s optimal level.

To assess the feasibility of FACTS, the use of FACTS (adaptive condition) had to be compared to a control condition (nonadaptive condition). In environmental settings similar to those in the pre-task assessment, participants underwent two extended sessions of the mental arithmetic task under nonadaptive and adaptive conditions in a counterbalanced design within at least two weeks. To intentionally induce mental overload, both sessions contained 60 trials and lasted approximately 30 minutes. In order to prevent bias, the participants were blinded to the experimental condition (nonadaptive or adaptive) and the difficulty level of the mathematical problems. Moreover, they did not acknowledge not only the details of the pre-task assessment and both experimental conditions but also how FACTS works. After all, the verbal instruction given to them could be summarized as follow: solve every on-screen mathematical problem.

Throughout nonadaptive condition, the task difficulty was fixed at the participant-specific optimal level predetermined from the pre-task assessment. Meanwhile, the incorporation of FACTS meant that the difficulty level in adaptive condition would be constantly adjusted in accordance to real-time mental workload monitoring. In adaptive condition, the first and following three trials were set at level 1 and 2 respectively. Beyond the sixth trial, FACTS’ DDA (refer to section II-C2) applied itself to adjust the difficulty level once every three trials until the end of the session.

C. fNIRS ADAPTIVE COGNITIVE TRAINING SYSTEM (FACTS)
Developed entirely on Windows using MATLAB (Mathworks, inc.), FACTS adopts a closed-loop control system by which fNIRS metrics of mental workload are taken into consideration in the DDA process (refer to Fig. 1a). It monitors mental workload by comparing the hemodynamics between task trials and adjusting the task difficulty accordingly. OT-R40 (Hitachi Medical Corporation, Japan) is used to provide non-invasive means to measure brain activity through fNIRS. Since the device itself is incapable of real-time fNIRS data processing and algorithm execution,
FIGURE 1. (a) FACTS’ closed-loop control system block diagram. (b) The paradigm for a single trial, where \( t_{Pn} \) is the time taken to respond to \( n \)th mathematical problem. The green and red shaded regions represent the period in which oxy-Hb rest(\( n \)) and oxy-Hb task(\( n \)) are computed. (c) Schematic diagram of FACTS. fNIRS data are transferred and processed in an external computer before being ultimately translated into feedback to adjust the task difficulty. (d) The experimental setup and FACTS in use. The participants were instructed to avoid movement, keep their left hand on the arm rest and their right hand on the Serial Response Box throughout each session. (e) The 3×11 probe set used in FACTS is sufficient to cover the entire prefrontal cortex. Sources 23 and 28 are consistently placed at the positions T4 and T3 respectively in the international 10-20 system [41]. Real-time mental workload monitoring focuses on the DLPFC, shaded in green. The model has consented to the publication of his photographs.

a computer is included not only to perform the data computation but also to run the computerized mental arithmetic task. FACTS requires both the OT-R40 and computer to stay connected and communicate with one another all the time via two different physical connections — LAN and RS-232C. The former is for real-time transmission of fNIRS data from the OT-R40 to the computer while the latter is for the computer to send predefined commands to operate the unmanned OT-R40. A schematic overview of FACTS is depicted in Fig. 1c while Fig. 1d is a picture of the experimental setup and FACTS in use.

1) fNIRS DATA ACQUISITION AND PREPROCESSING
Multichannel OT-R40 measures brain activity in the units of millimolar-millimeter (mM·mm) at a sampling rate of 10 Hz. It operates based on continuous wave fNIRS concept and utilizes laser diodes (sources) irradiating near-infrared light of 690 and 830 nm directly at the scalp through optical fibers. While being partly absorbed and scattered, the near-infrared light follows an ellipsoid path to be picked up by avalanche photodiodes (detectors) located 30 mm away from respective sources. The hemodynamic responses are usually represented by the relative concentration changes in oxy-Hb and deoxy-Hb converted from the changes in attenuated near-infrared light intensity via the Beer-Lambert law [42]. A pair of source and detector forms a measurement channel. Although FACTS only targets the DLPFC, the standard 3×11 probe set with 52 channels (17 sources and 16 detectors) is used. With a fitting head cap housing the probes, the fNIRS setup is easy, fast and convenient. In spite of the head anatomical variability across individuals, sources 23 and 28 are consistently placed at the positions T4 and T3 respectively in the international 10-20 system [41], covering the entire prefrontal cortex (see Fig. 1e).
Adopting the most commonly used cutoff frequencies [43], a Hamming window fifth-order band-pass finite impulse response filter with cutoff frequencies of 0.01 and 0.5 Hz is applied on every newly acquired fNIRS datum. Cutoff frequencies in this range are recommended to eliminate instrumental noise (3–5 Hz) and physiological noise caused by heart beat (1–1.5 Hz) while avoiding accidental removal of task-elicted hemodynamic responses which usually lie within the block frequency [44]. In accordance to Fig. 1b, the block frequency in this case is 1/30 Hz, overlapping with respiration (0.2–0.4 Hz). Furthermore, the effect of respiration on fNIRS signals is much smaller than the other noise sources [45]. Thus, it is not worth the risk trying to eliminate fNIRS noise due to respiration as the process can unintentionally remove task-related components. Next, the data are moving-averaged (≤ 5 s) for further denoising. Baseline and motion correction are excluded as the former normally require a minimum number of observations while a perfectly accurate model or prior distribution of noise is a necessity for the latter, both of which are unrealistic in real-time applications like FACTS whereby the uncertainty is high as the task difficulty is always changing. Relative to deoxy-Hb, oxy-Hb not only possesses higher signal-to-noise ratio [46] but is also more highly correlated with functional magnetic resonance imaging blood oxygen level-dependent signals [47] and more sensitive towards task-evoked changes [48]. As a consequence, oxy-Hb is chosen as the primary source for DDA in FACTS. The average of the oxy-Hb signals sampled from eight channels of interest (channel 8, 18, 19, 29 for the left DLPFC and 3, 13, 14, 24 for the right DLPFC [49]) is used for further analysis. A previous study has successfully employed a similar approach to evaluate different levels of workload on unmanned aerial vehicle pilots [27].

2) FEATURE EXTRACTION AND DYNAMIC DIFFICULTY ADJUSTMENT (DDA)
The 5–7-s hemodynamic delay [48] has to be taken in account. For each trial (taking Fig. 1b as an example), the average oxy-Hb value between 8 and 15 s, and between 23 and 30 s (end of task period) are hereinafter referred to as \( \text{oxy-Hb}_{\text{rest}(n)} \) and \( \text{oxy-Hb}_{\text{task}(n)} \) respectively, where \( n \) is the number of trials. By subtracting the former from the latter, each trial yields an activation value, \( \text{oxy-Hb}_{\text{act}(n)} \). Since task repetition has been demonstrated to be effective in uncorrelated trend noise reduction [50], the activation values are averaged every three trials, as follow:

\[
\text{oxy-Hb}_{\text{act}(n-2,n)} = \frac{1}{3} \sum_{n=2}^{N} \text{oxy-Hb}_{\text{act}(n)}
\]

Representing the hemodynamic changes progressing from previous three trials to current three trials, \( \text{oxy-Hb}_{\text{act}(n-2,n)} \) (current) is constantly compared to \( \text{oxy-Hb}_{\text{act}(n-5,n-3)} \) (previous). However, the first comparison between the average activation values can only take place right after the sixth trial \( (n = 6) \). From there onward, similar comparison is made once every three trials \( (3n) \) to decide the difficulty level of the next three trials. The process is repeated until the total number of trials \( (N) \) is reached. Such quantitative pairwise comparison between average activation values can help to minimize the interference of baseline drift and scalp-hemodynamics which is often assumed to be globally uniform [51].

The arousal factor in the Yerkes–Dodson law, in this case, is replaced by mental workload in the form of task difficulty. An increase in task workload is often accompanied by an increase in oxy-Hb level when the mental workload at hand does not exceed the working memory capacity limits [27], [28]. In this case, the workload lies within the lower arousal region as per the Yerkes–Dodson law and there are ample cognitive resources to deal with it. A larger \( \text{oxy-Hb}_{\text{act}(n-2,n)} \) can be an indication of such situation. In order to prevent mental underload and to push the mental workload past the capacity limits to promote learning [10], the upcoming three trials will be made harder by one level. If the participants are already at the highest level, the difficulty level will remain unchanged. Contrariwise, once the task workload exceeds the working memory capacity limits, hemodynamics starts to follow a downward trend. Excessive difficulty shifts the mental workload to the higher arousal region and may cause anxiety, stress, indecision, lower concentration and so on, all of which can reduce brain activity. The existence of a sudden decrease during hemodynamic monitoring in the form of a smaller \( \text{oxy-Hb}_{\text{act}(n-2,n)} \) can be interpreted as a manifestation of excessive mental workload. To prevent disruption in both learning and motivation [11], the difficulty of the next three trials will be reduced by a level or remain at the minimum. The DDA can be represented via the flowchart and pseudocode depicted in Fig. 2a and 2b.

D. OUTCOME ANALYSIS
1) BEHAVIORAL
a: PERCEIVED WORKLOAD
One of the outcome measures was defined by the difference in the NASA-TLX score between nonadaptive and adaptive conditions. The NASA-TLX, a multidimensional tool to measure subjective workload [52], was administered at the end of each session. It assesses six distinct dimensions: 1) mental demand — amount of cognitive activity required; 2) physical demand — intensity of physical activity needed; 3) temporal demand — level of time pressure due to the task pace; 4) effort — quantity of hard work invested to perform the task; 5) performance — degree of success in task completion; and 6) frustration — complacency felt during the task. These six subscale altogether contribute to the overall workload score.

b: TASK PERFORMANCE
The number of problems solved, response time, and accuracy were metrics that could reliably reflect task performance. Nevertheless, the difficulty level in adaptive
condition which was designed to adjust every three trials had a straightforward influence on these metrics with every round of adjustment. Therefore, it was inappropriate to evaluate the outcome measures by directly comparing both experimental conditions on the basis of these task performance metrics. Benchmarking their performance i.e., calculating changes in these metrics relative to the baseline performance obtained in the pre-task assessment constituted a more valid approach to assess the outcome measures. Each experimental session consisted of 60 trials, with each trial yielding a trio of task performance metrics. For each participant, the changes in task performance relative to baseline were calculated by subtracting the baseline performance values from these metrics. To observe the time-on-task effects, these metrics were split evenly and averaged according to three time intervals: $T_1$ (0–10 minutes), $T_2$ (10–20 minutes) and $T_3$ (20–30 minutes). This left only a trio of relative task performance metrics per time interval per participant. The pattern of change across time intervals in these metrics, potentially indicating mental underload or overload, was considered another outcome measure.

2) HEMODYNAMICS

Similarly, each participant yielded 60 DLPFC activation values in each of the experimental sessions. fNIRS measurement is always relative to baseline or a starting value. Since baseline correction could not be done during acquisition and since fNIRS data in both experimental sessions were recorded separately, it was inappropriate to directly compare the hemodynamics across the participants and experimental conditions. Instead, min-max normalization was first applied to correct the individual and sessional variability in baseline/starting values [23] using the formula

$$z_i = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)}; \quad i = [1, 37]$$

where $i = \text{participant number}$, $x_i = \text{i}^{\text{th}} \text{participant’s original activation values}$, and $z_i = \text{i}^{\text{th}} \text{participant’s normalized activation values}$. In the normalization process, both the numerator and denominator were in the same units (mM·mm), leaving the resulting normalized data in arbitrary units (a.u.). The normalized DLPFC activation was analyzed in a similar manner as the relative task performance metrics. The normalized activation values were split evenly into three time intervals and averaged. Any drop in the activation value over time on task could signify mental underload or overload. Via a side-by-side monitoring of the DLPFC activation and relative task performance metrics, the feasibility of FACTS in preventing mental underload and overload could be evaluated more comprehensively.

E. STATISTICS

The differences in the NASA-TLX overall and subscale scores between the two experimental conditions were determined using the Wilcoxon signed-rank tests [53] with Bonferroni–Holm correction [54]. Similar tests were repeated not only to test for intra-condition differences across three time intervals but also to check for inter-condition differences in difficulty level. A two-way repeated measures analysis of variance (ANOVA) with experimental condition (non-adaptive and adaptive) and time interval ($T_1$, $T_2$ and $T_3$) as within-subjects factors was performed to examine the

**FIGURE 2. Representation of FACTS’ DDA algorithm in the form of: (a) flowchart and (b) pseudocode.**
interaction effect between both factors on the relative number of problems solved, response time, accuracy and normalized DLPFC activation. This was followed by simple-effects analysis of the factor(s) showing a significant effect. On the behavioral level, all three relative task performance metrics were also statistically assessed. The relative number of problems solved, response time, and accuracy in each condition was analyzed using multiple paired $t$-tests with Bonferroni–Holm correction only to investigate the time-on-task effects i.e., the degree of detriment in these metrics over increasing time on task. In both experimental conditions, differences across time intervals in the normalized DLPFC activation were also evaluated using the same tests to see how the hemodynamics and time on task were associated. Finally, three paired $t$-tests with Bonferroni–Holm correction was used to determine if the normalized DLPFC activation differed between both experimental conditions in each of the time intervals.

### III. RESULTS

#### A. BEHAVIORAL

1) **PERCEIVED WORKLOAD**

Table 1 records the NASA-TLX overall and subscale scores. In adaptive condition, the overall and subscale scores were consistently lower except in the temporal demand subscale. The differences were not significant though.

2) **TASK DIFFICULTY**

The difficulty level the participants went through in both experimental conditions is also tabulated in Table 1. In adaptive condition, the initial difficulty level in $T_1$ was significantly lower than that in nonadaptive condition ($p = 0.001$) due to the contrasting starting point. Adaptive condition always started from level 1 but nonadaptive condition was not. Nevertheless, in $T_2$ and $T_3$, the participants gradually
advanced to significantly higher levels as compared to \(T_1\) (both \(p < 0.001\)). There was no significant difference between both experimental conditions in terms of difficulty level in \(T_2\) and \(T_3\).

3) TASK PERFORMANCE

The relative number of problems solved, response time, and accuracy are also reported in Table 1. The two-way repeated measures ANOVA revealed no interaction effect of experimental condition and time interval on all these task performance metrics. Both the relative number of problems solved and response time exhibited an inverted U-shaped curve in both experimental conditions; they increased from \(T_1\) to \(T_2\) but dropped from \(T_2\) to \(T_3\), albeit insignificantly. The relative task accuracy in nonadaptive condition also followed the same pattern. Conversely, the participants’ relative task accuracy improved consistently along the course of the mental arithmetic task in adaptive condition. Nonetheless, no significant inter-condition difference was found in these metrics.

B. HEMODYNAMICS

1) PRE-TASK ASSESSMENT

The difficulty level that evoked the largest activation value among all six levels in the pre-task assessment was considered the participant’s optimal level. The activation values for all six difficulty levels and 37 participants are tabulated in Table 2, with the participant-specific optimal difficulty level shaded in gray. Throughout nonadaptive mental arithmetic task, the participants were only tested with mathematical problems of their respective optimal difficulty level. According to the table, roughly 13.51, 16.22, 18.92, 8.11, 18.92 and 24.32% of the participants underwent level 1, 2, 3, 4, 5 and 6 respectively.

2) NORMALIZED DLPFC ACTIVATION

Grand block averages of the preprocessed oxy-Hb and deoxy-Hb signals (before normalization) across participants and difficulty levels for both nonadaptive and adaptive conditions are illustrated in Fig. 3. In both graphs, there appeared to be a negative relationship between oxy- and deoxy-Hb signals. The observation was consistent to the typical cerebral hemodynamic response to neural activation [40] as a result of neurovascular coupling [55].

According to the two-way repeated measures ANOVA, there was no interaction effect of experimental condition and time interval on the normalized DLPFC activation. However, time interval was found to have a significant effect on the normalized DLPFC activation not only in nonadaptive condition \([F(1,477,53.159) = 11.884, p < 0.001]\) but also in adaptive condition \([F(2,72) = 15.086, p < 0.001]\). In both experimental conditions, the participants’ normalized DLPFC activation exhibited a consistent downward trend over the course of the task, from \(T_1\) to \(T_3\). At the halfway mark \((T_2)\) of nonadaptive condition, the participants’ activation decreased significantly \((p = 0.019)\) as compared to the starting point \((T_1)\). In the last time interval \((T_3)\), their activation continued to diminish to a significantly lower amplitude than that in \(T_1\) and \(T_2\) \((p < 0.001\) and 0.003 respectively). Similarly, the participants’
DLPFC activation in adaptive condition differed significantly between $T_1$ and $T_2$ ($p < 0.001$). Although their activation further dropped in $T_3$, the decrement was smaller in contrast to that in nonadaptive condition. As a result, the amplitude in $T_3$ was only significantly smaller than that in $T_1$ ($p < 0.001$) but not in $T_2$. No significant difference between experimental conditions was found.

**IV. DISCUSSION**

**A. PERCEIVED WORKLOAD**

The workload was lower in adaptive condition, as evidenced by the lower NASA-TLX overall score. The above observation was expected when comparing between personalized and non-personalized task. In adaptive condition, the participants also reported lower scores in all the subscales except the temporal demand subscale. Similar scores in the temporal demand were expected as the participants undertook the task at their own pace in both experimental conditions; they were not pressured to solve as many mathematical problems as quickly as possible. In addition, it came no surprise that the participants registered lower score in the mental demand in adaptive condition as compared to nonadaptive condition. The latter was deliberately made more cognitively taxing, in attempt to induce more severe mental overload. However, despite not being statistically significant ($p = 0.140$), the difference might be of clinical significance. Furthermore, the higher frustration subscale scores registered in nonadaptive condition could possibly reflect how irritated, stressed and annoyed the participants were. As a further matter, increase in mental demand has been known to increase effort, as indicated by the effort subscale. All of these observed in nonadaptive condition could be a manifestation of mental overload associated with problem-solving tasks like mental arithmetic. Still, there was no significant difference between both experimental conditions in the overall and subscale scores. Perhaps increasing the sample size would uncover some as it would more reliably represent the population mean.

**B. TASK PERFORMANCE**

All the participants started the cognitive task the same way i.e., three level 1 trials followed by three level 2 trials in adaptive condition. Afterwards, only the task difficulty depended on their own hemodynamics. Most participants eventually progressed to significantly higher difficulty levels that were comparable to those in nonadaptive condition. The pattern of change in the relative task performance metrics (number of problems solved, response time, and accuracy) was analyzed to monitor the participants’ performance along the course of the 30-minute task.

In both experimental conditions, the participants’ number of problems solved and response time increased then decreased. The inverted U-shaped trend observed conforms to the empirical theory proposed by the Yerkes–Dodson law [35]. With arousal, performance increases until a certain point where the arousal becomes immoderate then it begins to decline. Productive learning might be the cause of the enhanced task performance while sustained cognitive engagement for prolonged periods could eventually lead to mental overload, which could be the culprit behind the drop in task performance. Mental overload not only worsens the latency of response inhibition and selection [56] but also causes one to lose focus and commit error [8], [9]. Interestingly, the participants’ accuracy in nonadaptive condition also followed an inverted U-shaped pattern while in adaptive condition, their accuracy consistently improved over time. The only logical explanation behind the different patterns in task performance observed in nonadaptive and adaptive conditions is that the participants might have been suffering from more severe mental overload in the former while they might have been just starting to experience mental overload in the latter. The fewer problems solved, slower response time and higher accuracy towards the end of adaptive condition could be an implication that the participants simultaneously exhibited indications of productive learning and early signs of mental overload.

**C. HEMODYNAMICS**

The DLPFC activation in nonadaptive and adaptive conditions followed a decreasing trend over time, from $T_1$ to $T_3$. Under both experimental conditions, the brain activation dropped significantly as the participants progressed from $T_1$ to $T_2$. The only observable difference between nonadaptive and adaptive conditions was that the activation dropped significantly from $T_2$ to $T_3$ in the former but not in the latter. Contrastingly, lower oxy-Hb levels for improved task performance and lower oxy-Hb levels for poorer performance could be interpreted completely differently. Looking at the DLPFC activation, task performance and perceived workload altogether, two possible explanations have been proposed for the drop in activation: 1) productive learning; 2) mental overload.

1) **PRODUCTIVE LEARNING**

The decrease in the DLPFC activation, coupled with better performance, could be attributed to productive learning following repeated exposure to the same task [57]. The observation was in agreement to the neural efficiency hypothesis [58], [59] which posits that decreased activation yielding better performance indeed reflects enhanced efficiency of the underlying neural circuits. The DLPFC is more active when novice individuals engage in a particular task. As they learn and become experts, they can attain similar or even higher levels of task performance albeit consuming less neural resources. It may very well be the case given the simultaneous reduced brain activation and improved performance from $T_1$ to $T_2$. In both experimental conditions, the drop in brain activation in $T_2$ could be due to productive learning. The participants were able to perform the task with better accuracy using less neural resources, a sign of more efficient processing after repeated exposure to the task. Consistently, a previous study training participants on Tetris, a tile-matching
puzzle video game, also found that performance and brain activation are inversely proportional i.e., the better the performance gets, the less active the brain is [60].

2) MENTAL OVERLOAD
The participants’ simultaneous drop in DLPFC activation and all three task performance metrics towards the last interval (T3) observed only under nonadaptive condition, taken together with higher NASA-TLX scores, could be a sign of more severe mental overload. Solving mathematical problems of fixed complexity repetitively could function as a frustrating agent. Therefore, prolonged periods of solving similar mathematical problems in nonadaptive condition was more cognitively taxing and aversive as reflected in the higher NASA-TLX mental demand and frustration subscale scores. Sustained cognitive engagement over long periods of time might manifest so-called time-on-task effects whereby human performance decreases and frustration arises because of mental overload [61], [62]. Several studies have also reported similar inverted U-shaped relationship between mental workload and DLPFC hemodynamics [25], [26], [37]. The observation is aligned with the hypothesis which fore-saw more severe mental overload in nonadaptive condition. Nonetheless, the participants continued to show trends associated with productive learning i.e., reduced DLPFC activation coupled with improved task accuracy despite the early signs of mental over load (drop in number of problems solved and response time). Taken together, these findings could mean that the participants were more immersed in the task and more successful in avoiding mental overload with the incorporation of FACTS.

D. GENERALIZABILITY
Accounting for the 5–7-s hemodynamic delay [48], fNIRS studies normally employ rest and task period which are not too short (≥10s) [43] so there will not be any timing issue to incorporate FACTS. Else, the period used for averaging can also be adjusted accordingly in FACTS. Just that FACTS is working memory-oriented, it focuses on the DLPFC due to its implications in working memory demands coping. Thus, it might be used potentially in conjunction with any other working memory task. The Yerkes-Dodson law that constitutes the working principle of FACTS applies not only to mental arithmetic task but also to almost all tasks. N-back task, one of the most widely used working memory task, is no exception. A previous study has observed that with increasing level of N-back task, hemodynamics increases to a point then it decreases [63]. Therefore, it is safe to assume that FACTS is generalizable for other working memory tasks but it requires more studies adopting FACTS in task of different nature in order to say the same for other tasks.

E. LIMITATIONS
A minor limitation lies in the mathematical problems used in the mental arithmetic task. There might be instances in which mathematical problem of higher difficulty level was as difficult as or easier than that of lower difficulty level. For example, a level 6 problem (20 + 30 + 40) would seem like and could be solved like a level 2 problem (2 + 3 + 4). However, the probability of occurrence was extremely small (≤0.395%). New cognitive intervention must undergo rigorous tests before they can be offered to people. This testing process needs to prove that it works as advertised and that it has the edge over existing intervention. Such rigorous testing includes, but not limited to, a side-by-side comparison of adaptive difficulty against static difficulty (either too easy or too hard or both) or random difficulty. It will be worthwhile to assess the performance and robustness of FACTS with respect to those scenarios. It is also postulated that further task prolongation can yield more obvious and significant results to ponder over the efficacy of FACTS. On top of that, testing FACTS on different age or cognitively impaired or left-handed groups can provide even more comprehensive validation as current results were solely based on healthy young right-handed university students. One last limitation is that this feasibility study only looked into the cross-sectional effects that FACTS brings within a 30-minute session. What is more interesting is the long-term impact of FACTS in longitudinal cognitive training and rehabilitation. Comparison between FACTS and other neurophysiological DDA systems in cognitive training is another area worthwhile for further investigation.

V. CONCLUSION
FACTS works by monitoring mental workload in real time and adapting the task difficulty when detrimental state such as mental underload or overload is detected. Its novelty can aid learners to avoid those unintended conditions. Given the higher perceived workload and severe signs of mental overload observed only in nonadaptive but not in adaptive condition, the feasibility of FACTS was proven. The results conclude that the implementation of FACTS is feasible. A neurophysiological system to personalize cognitive training or rehabilitation using DDA technology can potentially maximize the gains by boosting learners’ engagement for the long haul. Since FACTS is highly flexible that it can be bundled with existing longitudinal cognitive training or rehabilitation programs effortlessly, future studies will need to investigate if incorporating FACTS is more efficacious. As of now, FACTS will be employed with the same mental arithmetic task in an upcoming longitudinal study in attempt to decelerate the cognitive deterioration in Alzheimer’s disease. It is believed that FACTS can be further improved by including extracerebral artifacts correction such as the real-time scalp signal separating algorithm [30] so it can offer more accurate assessment of mental workload.

REFERENCES
[1] A. R. A. Conway, M. J. Kane, and R. W. Engle, “Working memory capacity and its relation to general intelligence,” Trends Cognit. Sci., vol. 7, no. 12, pp. 547–552, Dec. 2003.
[2] K. A. Ericsson, W. G. Chase, and S. Faloon, “Acquisition of a memory skill,” Science, vol. 208, no. 4448, pp. 1181–1182, Jun. 1980.
[3] K. Ball, D. B. Berch, K. F. Helmers, J. B. Jobe, M. D. Leveck, M. Marsiske, J. N. Morris, G. W. Rebok, D. M. Smith, S. L. Tennstedt, F. W. Unverzagt, and, S. L. Willis, “Effects of cognitive training interventions with older adults: A randomized controlled trial,” J. Am. Med. Assoc., vol. 288, no. 18, pp. 2271–2281, 2002.

[4] H. Takahashi, Y. Takeda, N. Iwata, H. Hashizume, A. Sekiguchi, A. Fukushima, and R. Kawashima, “Working memory training using mental calculation impacts regional gray matter of the frontal and parietal regions,” PLoS ONE, vol. 6, no. 8, Aug. 2011. Art. no. e23175.

[5] R. Kawashima, K. Orita, R. Yamazaki, N. Tajima, H. Yoshida, M. Taira, K. Iwata, T. Sasaki, K. Maeyama, N. Usui, and K. Sugimoto, “Reading aloud and arithmetic calculation improve frontal function of people with dementia,” J. Gerontol. A, Biol. Sci. Med. Sci., vol. 60, no. 3, pp. 380–384, Mar. 2005.

[6] D. DeStefano and J. LeFevre, “The role of working memory in mental arithmetic,” Eur. J. Cognit. Psychol., vol. 16, no. 3, pp. 353–386, May 2004.

[7] D. Giakoumis, A. Vogiannou, I. Kosunen, K. Moustakas, D. Tzovaras, and M. S. Young and N. A. Stanton, “It’s all relative: Defining mental workload,” in Proc. B-Interface negative mood induced during HCI,” in Of Augmented Cognition, vol. 6, no. 8, Aug. 2011, Art. no. e23175.

[8] K. Nathan and J. L. Contreras-Vidal, “Negligible motion artifacts in functional near-infrared spectroscopy (fNIRS) neuroimaging in exercise–cognition science: A systematic, methodology-focused review,” J. Clin. Med., vol. 7, no. 12, p. 466, Nov. 2018.

[9] F. Zheng, S. Fantini, and R. J. K. Jacob, “Using fNIRS brain sensing in realistic HCI settings: Experiments and guidelines,” in Proc. 22nd Ann. ACM Symp. User Interface Softw. Technol. (USIT), 2009, pp. 157–166.
**[43]** P. Pinti, F. Scholkmann, A. Hamilton, P. Burgess, and I. Tachtsidis, “Current status and issues regarding pre-processing of fNIRS neuroimaging data: An investigation of diverse signal filtering methods within a general linear model framework,” *Frontiers Hum. Neurosci.*, vol. 12, p. 505, Jan. 2019.

**[44]** S. Brigadoi, L. Ceccherini, S. Cutini, F. Scarpa, P. Scatturin, J. Selh, L. Gagnon, D. A. Boas, and R. J. Cooper, “Motion artifacts in functional near-infrared spectroscopy: A comparison of motion correction techniques applied to real cognitive data,” *NeuroImage*, vol. 85, pp. 181–191, Jan. 2014.

**[45]** Y. Tong, K. P. Lindsey, and B. D. Frederick, “Partitioning of physiological noise signals in the brain with concurrent near-infrared spectroscopy and fMRI,” *J. Cerebral Blood Flow Metabolism*, vol. 31, no. 12, pp. 2352–2362, Dec. 2011.

**[46]** F. Homae, H. Watanabe, T. Nakano, and G. Taga, “Prosodic processing in the developing brain,” *Neurosci. Res.*, vol. 59, no. 1, pp. 29–39, Sep. 2007.

**[47]** X. Cui, S. Bray, D. M. Bryant, G. H. Glover, and A. L. Reiss, “A quantitative comparison of fNIRS and fMRI across multiple cognitive tasks,” *NeuroImage*, vol. 54, no. 4, pp. 2808–2821, Feb. 2011.

**[48]** S. Moghimi, A. Kushki, S. Power, A. M. Guerguerian, and T. Chau, “Hemodynamic noise signals in the brain with concurrent near-infrared spectroscopy and functional near-infrared imaging,” *J. Cerebral Blood Flow Metabolism*, vol. 15, no. 4, pp. 511–525, Apr. 2005.

**[49]** S. Holm, “A simple sequentially rejective multiple test procedure,” *Scand. J. Stat.*, vol. 9, no. 1, pp. 65–70, 1979.

**[50]** G. Gratton, M. R. Goodman-Wood, and M. Fabiani, “Comparison of neuronal and hemodynamic measures of the brain response to visual stimulation: An optical imaging study,” *Hum. Brain Mapping*, vol. 13, no. 1, pp. 13–25, 2001.

**[51]** Z. Guo, R. Chen, X. Liu, G. Zhao, Y. Zheng, M. Gong, and J. Zhang, “The impairing effects of mental fatigue on response inhibition: An ERP study,” *PLoS ONE*, vol. 13, no. 6, Jun. 2018, Art. no. e0198206.

**[52]** S. G. Hart and L. E. Staveland, “Development of nasa-tlx (task load index): Results of empirical and theoretical research,” in *Advances in Psychology*, vol. 52, Amsterdam, The Netherlands: Elsevier, 1988, pp. 139–183.

**[53]** F. Wilcoxon, “Individual comparisons by ranking methods,” in *Breakthroughs in Statistics*, New York, NY, USA: Springer, 1992, pp. 196–202.

**[54]** S. Holm, “A simple sequentially rejective multiple test procedure,” *Scand. J. Stat.*, vol. 6, no. 2, pp. 45–70, 1979.

**[55]** G. Gratton, M. R. Goodman-Wood, and M. Fabiani, “Comparison of neuronal and hemodynamic measures of the brain response to visual stimulation: An optical imaging study,” *Hum. Brain Mapping*, vol. 13, no. 1, pp. 13–25, 2001.

**[56]** Z. Guo, R. Chen, X. Liu, G. Zhao, Y. Zheng, M. Gong, and J. Zhang, “The impairing effects of mental fatigue on response inhibition: An ERP study,” *PLoS ONE*, vol. 13, no. 6, Jun. 2018, Art. no. e0198206.

**[57]** C. Bartels, M. Wegzyn, A. Wiedl, V. Ackermann, and H. Ehrenreich, “Practice effects in healthy adults: A longitudinal study on frequent repetitive cognitive testing,” *BMC Neurosci.*, vol. 11, no. 1, p. 118, Dec. 2010.

**[58]** A. Curtin, H. Ayaz, Y. Tang, J. Sun, J. Wang, and S. Tong, “Enhancing neural efficiency of cognitive processing speed via training and neurostimulation: An fNIRS and TMS study,” *NeuroImage*, vol. 198, pp. 73–82, Sep. 2019.

**[59]** B. Dunst, M. Benedek, E. Jauk, S. Bergner, K. Koschutnig, M. Sommer, A. Ischebeck, B. Spinath, M. Arendasy, M. Bühner, H. Feinsteinhalter, and A. C. Neubauer, “Neural efficiency as a function of task demands,” *Intelligence*, vol. 42, pp. 22–30, Jan. 2014.

**[60]** R. J. Haier, B. V. Siegel, A. MacLachlan, E. Soderling, S. Lottenberg, and B. Dunst, M. Benedek, E. Jauk, S. Bergner, K. Koschutnig, M. Sommer, A. Ischebeck, B. Spinath, M. Arendasy, M. Bühner, H. Feinsteinhalter, and A. C. Neubauer, “Neural efficiency as a function of task demands,” *Intelligence*, vol. 42, pp. 22–30, Jan. 2014.

**[61]** R. I. Hau, B. V. Siegel, A. MacLachlan, E. Soderling, S. Lottenberg, and M. S. Buchsbaum, “Regional glucose metabolic changes after learning a complex visuospatial/motor task: A positron emission tomographic study,” *Brain Res.*, vol. 570, nos. 1–2, pp. 134–143, Jan. 1992.

**[62]** T. Möckel, C. Beste, and E. Wascher, “The effects of time on task in response selection—An ERP study of mental fatigue,” *Sci. Rep.*, vol. 5, Sep. 2015, Art. no. 10113.

**[63]** J. Lim, W.-C. Wu, J. Wang, J. A. Detre, D. F. Dinges, and H. Rao, “Imaging brain fatigue from sustained mental workload: An ASL perfusion study of the time-on-task effect,” *NeuroImage*, vol. 49, no. 4, pp. 3426–3435, Feb. 2010.

**[64]** S. H. Fairclough, C. Burns, and U. Kreplin, “Enrich activity in the prefrontal cortex and motivational intensity: Impact of working memory load, financial reward, and correlation-based signal improvement,” *Neurophotonics*, vol. 5, no. 3, 2018, Art. no. 035001.

**[65]** W. C. Ung et al.: FACTS for Cognitive Underload and Overload Prevention: A Feasibility Study

WEI CHUN UNG (Graduate Student Member, IEEE) was born in Penang, Malaysia, in 1992. He received both the B.Eng. (Hons.) and M.Sc. degrees in electrical and electronic engineering from Universiti Teknologi PETRONAS, Seri Iskandar, Malaysia, in 2016 and 2018, respectively, where he is currently pursuing the Ph.D. degree. From 2014 to 2015, he was a research intern with the Research and Development Group, Center for Exploratory Research, Hitachi Ltd., Saitama, Japan. His research interests include neuroimaging modalities and applications, neurofeedback, brain-computer interfaces, and technology-based cognitive training and rehabilitation. He was a recipient of the IEEE Signal Processing Society Malaysia Research Excellence Award—Best Master Thesis Award 2018 and Best Journal Paper Award 2017.

FABRICE MERIAUDEAU (Member, IEEE) was born in Villeurbanne, France, in 1971. He received the master’s degree in physics, the Engineering degree (first) in material sciences, and the Ph.D. degree in image processing from Dijon University, France, in 1994 and 1997, respectively. He held a postdoctoral position with the Oak Ridge National Laboratory for one year. From 2011 to 2016, he was the Director of Le2i (UMR CNRS), which has more than 200 staff members, where he is currently the Professeur des Universites. He has coordinated an Erasmus Mundus Master in the field of computer vision and robotics, from 2006 to 2010. He was the Vice President of the International Affairs for the University of Burgundy from 2010 to 2012. He has authored or coauthored more than 150 international publications. He holds three patents. Since 2016, he has been a Professor with the Universiti Teknologi Petronas. His research interests include image processing for non-conventional imaging systems (UIR, IR, and polarization) and, more recently, the medical biomedical imaging.

MASASHI KIGUCHI has studied various mental problems, including working memory, emotions, depression, blood flow, mental stress, and has been taking the lead in the development of new techniques for observing brain activities to open new research fields and in basic studies for putting them to practical use.

TONG BOON TANG (Senior Member, IEEE) was born in Johor, Malaysia, in 1976. He received the B.Eng. (Hons.) and Ph.D. degrees from The University of Edinburgh. He is currently an Associate Professor and the Director of the Institute of Health and Analytics, Universiti Teknologi PETRONAS. His research interests are in biomedical instrumentation, from device and measurement to data fusion. He serves as the Secretary of the Higher Centre of Excellence (HiCENG) Council, a Board Member of the INCF, and the Chair of the IEEE Circuits and Systems Society Malaysia Chapter. He received the Lab on Chip Award, in 2006, and the IET Nanobiotechnology Premium Award, in 2008.