Optimizing Mental Workload Estimation by Detecting Baseline State Using Vector Phase Analysis Approach

Lam Ghai Lim, *Graduate Student Member, IEEE*, Wei Chun Ung, *Graduate Student Member, IEEE*, Yee Ling Chan, *Member, IEEE*, Cheng-Kai Lu, *Senior Member, IEEE*, Tsukasa Funane, Masashi Kiguchi, and Tong Boon Tang, *Senior Member, IEEE*

***Abstract*— Improper baseline return from the previous task-evoked hemodynamic response (HR) can contribute to a large variation in the subsequent HR, affecting the estimation of mental workload in brain-computer interface systems. In this study, we proposed a method using vector phase analysis to detect the baseline state as being optimal or suboptimal. We hypothesize that selecting neuronal-related HR as observed in the optimal-baseline blocks can lead to an improvement in estimating mental workload. Oxygenated and deoxygenated hemoglobin concentration changes were integrated as parts of the vector phase. The proposed method was applied to a block-design functional near-infrared spectroscopy dataset (total blocks = 1384), measured on 24 subjects performing multiple difficulty levels of mental arithmetic task. Significant differences in hemodynamic signal change were observed between the optimal- and suboptimal-baseline blocks detected using the proposed method. This supports the effectiveness of the proposed method in detecting baseline state for better estimation of mental workload. The results further highlight the need of customized recovery duration. In short, the proposed method offers a practical approach to detect task-evoked signals, without the need of extra probes.**

***Index Terms*— Baseline state, brain-computer interface, functional near-infrared spectroscopy, hemodynamic response, mental workload, vector phase analysis, working memory.**

Manuscript received December 20, 2020; revised January 28, 2021; accepted February 22, 2021. Date of publication February 24, 2021; date of current version March 8, 2021. This work was supported in part by the Ministry of Higher Education Malaysia through the Higher Institutional Centre of Excellence (HICoE) grant awarded to the Centre for Intelligent Signal and Imaging Research (CISIR), National Grant, under Grant FRGS/1/2018/TK04/UTP/02/10, and in part by the Yayasan UTP under Grant 0153A-E99. *(Corresponding author: Tong Boon Tang.)*

Lam Ghai Lim, Wei Chun Ung, Yee Ling Chan, and Tong Boon Tang are with the Centre for Intelligent Signal and Imaging Research (CISIR), Institute of Health and Analytics, Universiti Teknologi PETRONAS, Seri Iskandar 32610, Malaysia (e-mail: limlamghai@gmail.com; ungweichun@gmail.com; elingx10@hotmail.com; tongboon.tang@utp.edu.my).

Cheng-Kai Lu is with the Department of Electrical and Electronic Engineering, Universiti Teknologi PETRONAS, Seri Iskandar 32610, Malaysia (e-mail: chengkai.lu@utp.edu.my).

Tsukasa Funane and Masashi Kiguchi are with the Center for Exploratory Research, Research and Development Group, Hitachi Ltd., Tokyo 185-8601, Japan (e-mail: tsukasa.funane.sb@hitachi.com; masashi.kiguchi.py@hitachi.com).

This article has supplementary downloadable material available at https://doi.org/10.1109/TNSRE.2021.3062117, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2021.3062117

I. Introduction

NEUROVASCULAR coupling takes place when there is an increase in brain activity, leading to changes in cerebral oxygenation and hemodynamics [1]. A number of brain imaging modalities, such as functional magnetic resonance imaging (fMRI) and functional near-infrared spectroscopy (fNIRS), have been used to quantify the hemodynamic response (HR). Due to the slow nature of HR, block design has been extensively employed in hemodynamic-based experimental paradigm [2]. The block design typically consists of multiple repetitions of functional task/stimulation period alternating with a control period for the recovery of task-evoked HR to the baseline [2]. The amount of time required for the recovery has been observed to prolong with an increase in the task duration [1]. Hence, it is recommended that the control period should have at least the same duration as, if not longer than, the task period [3]. Besides, younger and older people have different shapes and timings of the HR [4], [5], which can cause more variations in the recovery time.

The most widely used control condition is *rest* [6], where the participants are asked to stay relax and fixate their eyes on a target in a computer display. However, unrelated thoughts or mind wandering, the so-called default mode network, is likely to be active during such resting state [7], [8]. These task-independent ongoing or spontaneous brain activities in either short or long rest period could be substantially stronger or even higher than that of the task period [6], [9]. Such cognitive processes have resulted different effects on the task-evoked HR [10]–[12]. In the worst-case scenario, improper baseline return can change the time-shape of task-evoked HR, resulting in a false identification of the observed activation [13]. This may also lead to the cancellation effects when performing the block averaging [14].

In recent years, fNIRS-based brain-computer interface (BCI) has shown a rapid development in the field of neuroscience such as for mental workload assessment [15]–[17]. Using the averaging technique over the repetitions of each task difficulty level, a common finding has been reported that the magnitude of brain activation increases along with the increase in difficulty level [15], [18]. Although this technique has demonstrated a significant

finding at the inter-subject level, it is of vital importance to ensure the effectiveness of using fNIRS modality for single-trial estimation of mental workload. Considering the baseline-related factors, the estimation of mental workload can be greatly affected when the previous task-evoked HR does not return to a proper level of baseline. Hence, there is a need to develop a baseline detection method.

In this study, we propose a method using vector phase analysis, aiming to detect the baseline state of the task-evoked HR (before the onset of new functional stimulation) as being optimal or suboptimal. We exploit the facts that (i) fNIRS can quantify both oxygenated (HbO) and deoxygenated (HbR) hemoglobin concentration changes [19], and (ii) the occurrence of brain activation is detectable using vector phase analysis [20]. In addition, the peak of task-evoked HR is taken into consideration as the recovery duration scales proportionally with it, and a novel vector phase diagram is specifically designed to detect (with a set of design criteria) if the task-evoked HR had returned to baseline. The proposed method was assessed with a block-design fNIRS dataset containing hemodynamic measurements over the prefrontal cortex (PFC) of 24 human subjects [21]. At both inter- and intra-subject analyses, we hypothesize that the estimation of mental workload improves with an optimal baseline detected with our novel vector phase method. Due to the lack of a clear reference of an individual's mental workload, the behavioral performance, i.e. response time [22], was used to evaluate the estimation of mental workload with fNIRS activation.

II. Methodology

A. Proposed Method

Fig. 1(a) shows a standard block design employed in hemodynamic-based experiment, with n denotes the current block. A pre-task period ($n = 0$) is preceded and followed by a repetition of block consists of task and rest periods ($n \geq 1$). The notations t_g^n and t_h^n are used to denote the timeline, representing the end of task and rest periods for n^{th} block, respectively. The proposed method is formulated using the time-shape of conventional HR, where HbO/HbR increases/decreases after the task onset and returns to baseline during the rest period [19]. The baseline state is identified prior to task onset, with the aid of vector phase diagram. At each time point t, the time-series HbO and HbR signals are transformed to two-dimensional vector components $x(t)$ and $y(t)$, respectively, with respect to their origins. Hence, the following equations are derived:

$$x(t) = \text{HbO}(t) - \text{HbO}_{\text{base}}^{n-1} \tag{1}$$

$$y(t) = \text{HbR}(t) - \text{HbR}_{\text{base}}^{n-1} \tag{2}$$

where $\text{HbO}(t)$ and $\text{HbR}(t)$ are the time-series HbO and HbR signals, respectively. $\text{HbO}_{\text{base}}^{n-1}$ and $\text{HbR}_{\text{base}}^{n-1}$ are the estimated HbO and HbR baselines from $(n-1)^{\text{th}}$ block, respectively. These estimated baselines, acting as the origin, are used for the execution of n^{th} block. Both $x(t)$ and $y(t)$ are computed every 0.5 s, as such time interval is neither too long nor too short for capturing the changes in HR. The expression of the magnitude

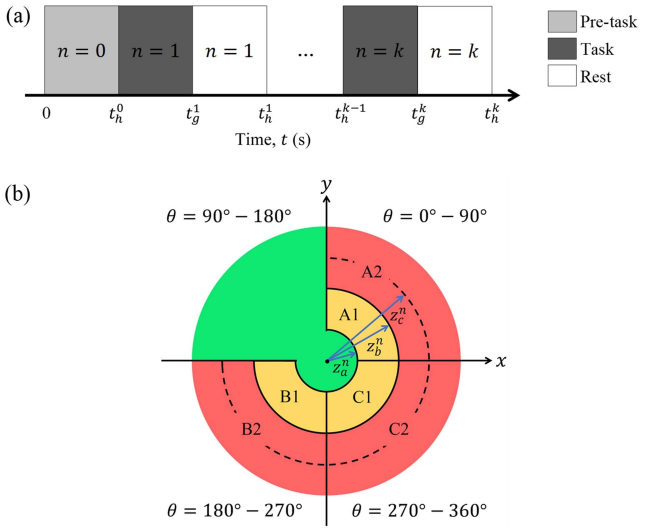

Fig. 1. (a) A typical block design used in hemodynamic-based experiment. (b) The design of a vector phase diagram with multiple threshold circles.

$z(t)$ and the phase $\theta(t)$ of $[x(t), y(t)]$ are then obtained as:

$$z(t) = \sqrt{(x(t))^2 + (y(t))^2} \tag{3}$$

$$\theta(t) = \tan^{-1}\frac{y(t)}{x(t)} \tag{4}$$

Fig. 1(b) displays the design of a vector phase diagram with multiple threshold circles for the estimation of baseline state in n^{th} block prior to task onset of $(n+1)^{\text{th}}$ block. Three major zones are defined in different shaded colors (green, yellow, and red) to describe the baseline state of task-evoked HR. The minor zones are labeled as A1, A2, B1, B2, C1, and C2 in accordance to the threshold circles and phases for a more detailed explanation of the yellow and red zones. Basically, green zone indicates that the task-evoked HR has returned to an optimal baseline. The yellow zone (A1, B1, and C1) implies that the task-evoked HR is somewhere in between peak activation and optimal baseline. The red zone (A2, B2, and C2) signifies that the task-evoked HR is rising towards the peak activation. Three threshold circles are designed to separate the zones, denoting as z_a^n, z_b^n, and z_c^n, from the most inner to the outer layer, respectively. z_c^n denotes the peak activation obtained from $z(t)$, computing from the time interval of the identified $[\text{HbO}_{\text{base}}^{n-1}, \text{HbR}_{\text{base}}^{n-1}]$ to t_g^n. Note that, the vector phase diagram considers the range beyond z_c^n, as this may happen when the HR in rest period is higher than the task period [6]. The relationships among the threshold circles are expressed as:

$$z_a^n = 0.3 z_c^n \tag{5}$$

$$z_b^n = 0.7 z_c^n \tag{6}$$

where the coefficients 0.3 and 0.7 are chosen empirically.

$\text{HbO}_{\text{base}}^n$ and $\text{HbR}_{\text{base}}^n$ are estimated before the task onset of $(n+1)^{\text{th}}$ block. As HbO/HbR gradually decreases/increases to baseline after the end of task [19], we identify the time points (t_{HbO}^n and t_{HbR}^n) whereby $\text{HbO}(t_{\text{HbO}}^n)$ and $\text{HbR}(t_{\text{HbR}}^n)$ are at minimum and maximum level, respectively. To counter the delay between HbO and HbR signals in returning to baseline [13], we compute the midpoint between

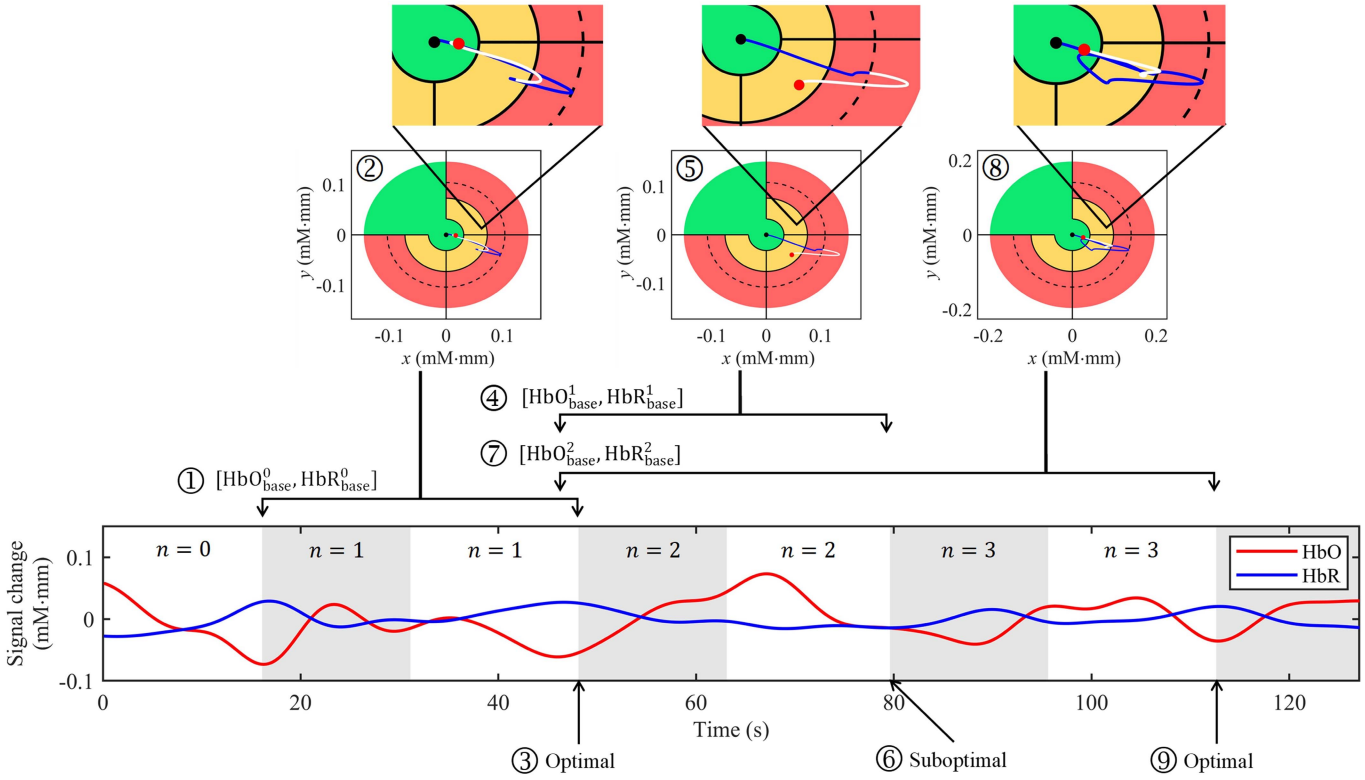

Fig. 2. An illustration of the proposed method in detecting baseline state using a sample of time-series fNIRS signals with a total of three complete sets of block ($n = 1$ to 3), where each block consists of a task (shaded in gray) and a rest period. The customized vector phase diagrams and the enlarged views of the projected trajectory $[x(t), y(t)]$ are shown for every block. Steps ①–⑨ are the execution of the proposed method. [HbO$^0_{base}$, HbR$^0_{base}$] is identified before 1st block (step ①). This is followed by the construction of vector phase diagram for 1st block (step ②). The blue-color projected trajectory starts from t^0_{mid} (the time where [HbO$^0_{base}$, HbR$^0_{base}$] is identified) to the end of task period and followed by the white-color projected trajectory till the end of rest period. The black- and red-color points indicate the starting and ending of the projected trajectory, respectively. Based on its ending location (green zone), an optimal baseline is detected for 2nd block (step ③). Steps ①–③ are repeated for every new block.

t^n_{HbO} and t^n_{HbR} as:

$$t^n_{mid} = \frac{t^n_{HbO} + t^n_{HbR}}{2} \tag{7}$$

Using t^n_{mid}, HbO$^n_{base}$ and HbR$^n_{base}$ are obtained as HbO(t^n_{mid}) and HbR(t^n_{mid}), respectively.

Fig. 2 illustrates the steps of applying the proposed method on fNIRS signals. It begins with the pre-task period ($n = 0$) by identifying HbO$^0_{base}$ and HbR$^0_{base}$ for the execution of next block ($n = 1$). For the subsequent n^{th} block, $x(t)$, $y(t)$, and $z(t)$ are computed using (1), (2), and (3), respectively, from t^{n-1}_{mid} to t^n_g. The design parameters of vector phase diagram are then obtained and computed using (5) and (6). After that, $x(t)$, $y(t)$, $z(t)$, and $\theta(t)$ are computed using (1), (2), (3), and (4), respectively, over the time interval of rest period. The zone of $[x(t), y(t)]$ is simultaneously identified using the customized vector phase diagram.

As the green zone indicates the most closer return of HR to its initial starting values, a set of design criteria is applied to determine the time interval for the computation of HbO$^n_{base}$ and HbR$^n_{base}$. It is selected based on the time interval of the first satisfying condition in the following sequence: (i) The green zone is detected over the rest period and (ii) the green zone is detected from t^{n-1}_{mid} to t^n_g. When none of the above-mentioned conditions (i) and (ii) are satisfied, HbO$^n_{base}$ and HbR$^n_{base}$ are remained the same from the previous $(n-1)^{th}$ block.

TABLE I
THE DEFINITION OF BASELINE STATE IN RELATION TO THE ZONE IDENTIFIED AT THE END OF REST PERIOD

The zone or zone transition	Baseline state
Green	Optimal
Green to yellow	Optimal
Red to yellow	Suboptimal
Red	Suboptimal

Table I summarizes the definition of baseline state in relation to the zone identified at the end of rest period, i.e. the zone at $[x(t^n_h), y(t^n_h)]$. The $(n+1)^{th}$ block is detected to have an optimal baseline when $[x(t^n_h), y(t^n_h)]$ is located in the green zone. A suboptimal baseline is detected when $[x(t^n_h), y(t^n_h)]$ lies in the red zone. Due to the uncertainty state of the yellow zone, a different approach, i.e. considering the zone transition, is used to detect the baseline state of the block that ends in this zone. When $[x(t), y(t)]$ transits from the green to yellow zone, this indicates that the HR begins to activate and we assume that there is still a high potential of obtaining the time-shape of conventional HR. As such, an optimal baseline is detected in this case. Oppositely, the transition from the red to yellow zone implies that the HR starts to return to baseline and it is therefore not suitable for the next task to begin. As such, a suboptimal baseline is detected in this case.

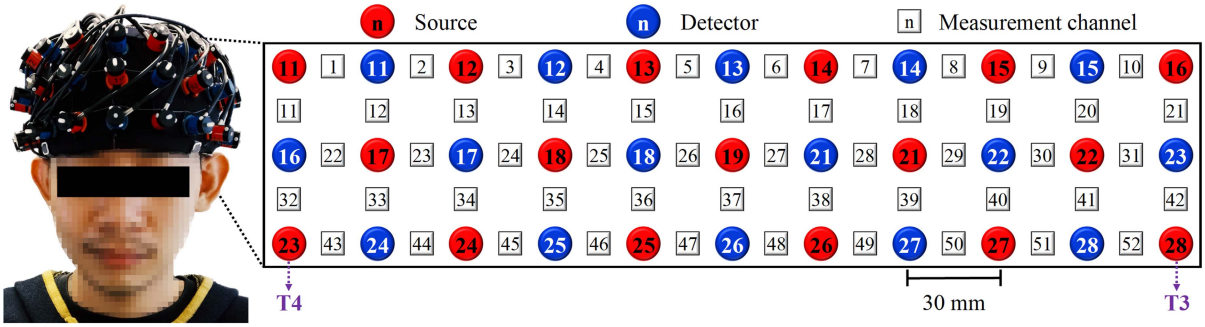

Fig. 3. A 3 × 11 probe layout with a standard source-detector distance of 30 mm was used to cover the entire PFC in [21]. In accordance to the international 10-20 system [24], sources 23 and 28 were constantly placed at the positions T4 and T3, respectively. The authors [21] have permitted the use of the figure.

B. fNIRS Dataset

The proposed method was applied on the lab-based fNIRS dataset published in [21]. Briefly, 10-Hz sampling rate fNIRS data were collected from a total of 38 right-handed healthy university students performing mental arithmetic task using a closed-loop brain training system, specifically to optimize the mental workload at WM-related brain region. A 3 × 11 probe layout with a standard source-detector distance of 30 mm was used for the fNIRS recording over the PFC (see Fig. 3). The experimental paradigm began with a 25-s pre-task and followed by a total of 60 block repetitions consists of 15-s task and 15-s rest periods. The whole protocol lasted approximately 30 min. During the task period, subjects were asked to mentally calculate and solve as many four-choice mathematical questions as possible. Addition or subtraction or a mixture of both were used for the mathematical questions, comprising of six difficulty levels from the easiest two single digits to the hardiest three double digits. As the detection of baseline state is only applicable after the task period of 1st block, the evaluation of task-evoked HR began from the 2nd block onwards. Hence, a maximum of 59 blocks per subject were available for assessment.

There are a few reasons of selecting the dataset to assess our proposed method. Firstly, due to the high variations at the intra- and inter-subject levels, we expected that there would be a mixture of baseline state in the dataset. As such, it offers a sizable amount of blocks with optimal or sub-optimal baseline to assess the effectiveness of the proposed method. Secondly, the experimental paradigm fulfilled a typical block design, where the duration of the rest period is recommended to have at least equal to or more than that of the task period [3]. Thirdly, the utilization of multiple task difficulty levels provides a means to evaluate the estimation of mental workload under the influence of baseline state. In this case, it is vital to select the blocks with optimal behavioral performances. As the time taken in responding to a task is associated with the level of mental workload [22], the blocks without a single attempted question were excluded. In addition, the blocks with an average response time of more than 1.5 interquartile range were further excluded to remove the non-optimal workload states, such as disengagement and overload [23].

C. Channel of Interest

MATLAB (Mathworks Inc., New York, USA) was primarily used for the data analysis. Prior to analysis, the channels with poor optode-to-scalp coupling were removed [25]. The light intensity data were transformed to the optical density data. The wavelet-based filtering with Daubechies 5 setup was chosen as a channel-by-channel approach for motion artifacts correction [26]. For every channel, the optical density data under the wavelengths of 695 and 830 nm were converted to the time-series HbO and HbR signals using the modified Beer-Lambert law [27]. These signals were obtained in the unit of millimolar millimeter (mM·mm), a product of the hemoglobin concentration change and optical path length. A third order Butterworth bandpass filter with cutoff frequencies of 0.01 to 0.09 Hz was then applied to remove the slow drifts (<0.01 Hz), Mayer wave (∼0.1 Hz), respiration (∼0.3 Hz), heartbeat (∼1 Hz), and high frequency noise (3–5 Hz) [28].

The proposed method was developed based on the time-shape of the conventional HR model. As such, a channel of interest (COI), i.e. the most activated channel, was selected for each subject to avoid misinterpretation of fNIRS signals arising from noisy or non-activated channels [29]. The time-series fNIRS signals were analyzed using general linear model (GLM)-based statistical parametric mapping [30]. A widely used two-gamma hemodynamic response function (HRF) with the standard parameter values were adopted to represent the conventional HR model [31]. Canonical HRF (cHRF) was generated by convolving the HRF with the boxcar function that corresponds to the time-piece of the experimental paradigm. The design matrices included a constant term, cHRF, and the temporal and dispersion derivatives of the cHRF [13]. The regression coefficient (β) and statistical t-value (t_{stat}) for the time-series HbO and HbR signals of every channel were estimated for the cHRF model using MATLAB's *fitlm* function. The association of the HbO/HbR signal with the conventional HR model was confirmed with a significant positive/negative t_{stat} ($p < 0.05$) [30]. The COI was selected from the channel with the largest absolute average of t_{stat} obtained from both HbO and HbR signals. To identify the Brodmann area (BA), the probe layout was mapped onto the head model applied in [25] as the reference position

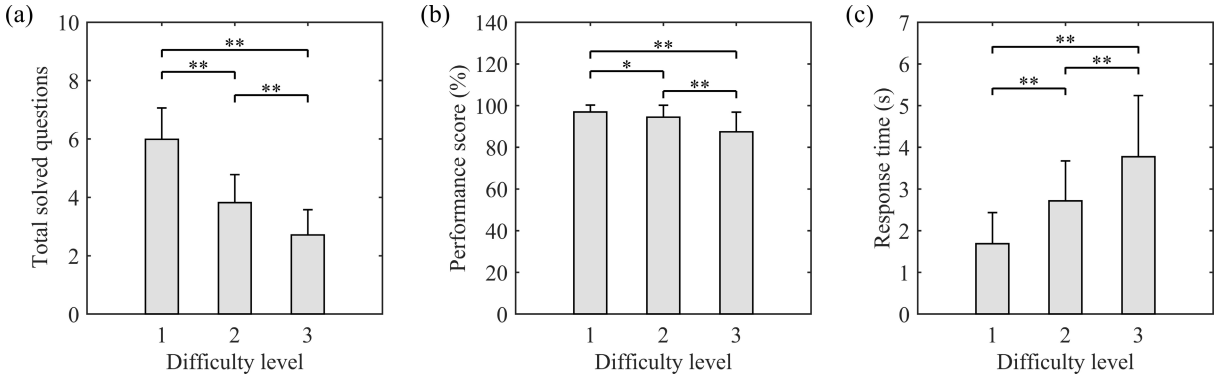

Fig. 4. (a) Total solved questions, (b) performance score, and (c) response time at each difficulty level. Error bars indicate SD. Statistical significance of Bonferroni *post hoc* test is shown: *$p < 0.01$ and **$p < 0.001$.

was similar. Prior to the execution of the proposed method, the identified COI was downsampled by averaging the latest five data points for every 0.5 s.

D. Evaluation

1) Behavioral Performance: Based on the findings of the closed-loop brain training [21], every subject managed to achieve and solve a minimum of third difficulty level for the mental arithmetic task. Due to the difference in attempted difficulty levels across subjects, it was not possible to one-by-one compare the difficulty level. Hence, the difficulty level was simplified to three levels, with levels 1 and 2 maintained, and levels 3 and above were considered as level 3. The total solved questions, response time, and performance score were averaged at each difficulty level for every subject. The main effect of difficulty level on each behavioral outcome was examined using one-way repeated measures analysis of variance (ANOVA). Geisser-Greenhouse correction was used to deal with the violation of sphericity and Bonferroni *post hoc* test was employed to identify the locus of the main effect.

2) Task-Evoked fNIRS Signals and Activation: Using the zone identified at the end of rest period, the total blocks were computed for each zone. The original dataset was referred as raw-baseline block (RBB). Accordingly, the blocks detected with our proposed method were labeled as an optimal-baseline block (OBB) or suboptimal-baseline block (SBB). The average value computed over a time interval of 0–0.5 s before the task onset was used to perform baseline correction for each block. The block average of HbO and HbR signals were conducted in accordance to the detected baseline state at both inter- and intra-subject levels. The task-evoked HbO/HbR activation was computed from the average of HbO/HbR signal amplitudes over the entire duration of the task period for every block [32], [33].

3) Mental Workload Estimation: At the inter-subject level, HbO/HbR activation was further averaged within each difficulty level and the detected baseline state. One-way repeated measure ANOVA with Geisser-Greenhouse correction (if necessary) and Bonferroni *post hoc* test were performed on HbO/HbR activation to access the effect of difficulty level. As the response time demonstrates a strong association with the difficulty level [22], Pearson's correlation coefficient r was obtained to individually examine its association with HbO/HbR activation in RBB, OBB, and SBB. In order to perform statistical comparison, Fisher's Z-transformation [34] was applied to convert the sampling distribution of the Pearson's correlation coefficient r to the normally distributed variable Z_r. The correlation coefficient Z_r across subjects was then statistically evaluated using Bonferroni-corrected paired t-test.

III. Results

A subject's data was excluded due to a technical glitch during the experiment. Six subjects' fNIRS data were severely corrupted with the motion artifacts over multiple time intervals, causing the failure of COI identification. Besides, the poor synchronization of either HbO or HbR or both signals with the task timeline led to another failure of COI identification in 7 subjects. Overall, GLM analysis identified COI in the remaining 24 subjects (identification number: S01–S24), located either over the right (17 subjects) or left (7 subjects) PFC. The mean and standard deviation (SD) for HbO signal were $\beta = 0.032 \pm 0.024$ and $t_{\text{stat}} = 34.9 \pm 24.4$ at $p < 0.001$, whereas for HbR signal were $\beta = -0.009 \pm 0.007$ and $t_{\text{stat}} = -29.2 \pm 22.4$ at $p < 0.001$. A total of 1416 blocks (24 subjects × 59 blocks/subject) were available. However, 32 blocks (2.3%) that left unattempted or detected as non-optimal workload states were further excluded. Thus, the remaining 1384 blocks (97.7%) were used for the assessment.

A. Behavioral Data

Fig. 4 shows the total solved questions, performance score, and response time at each difficulty level across 24 subjects. One-way repeated measures ANOVA with Geisser-Greenhouse correction returned a significant main effect of difficulty level on the total solved questions [$F(1.5,35.3) = 461.4$, $p < 0.001$], performance score [$F(1.4,31.1) = 27.9$, $p < 0.001$], and response time [$F(1.1,25.3) = 85.6$, $p < 0.001$]. For the total solved questions, Bonferroni *post hoc* test showed that the level 1 was significantly greater than all other levels (both $p < 0.001$) and level 3 was significantly smaller than all other levels (both $p < 0.001$). Similarly, for the performance score, level 1 was significantly higher than all other levels (both $p < 0.01$) and level 3 was significantly lower than

TABLE II
THE DISTRIBUTION OF BLOCKS IN ACCORDANCE TO THE DETECTED BASELINE STATE

Zone		Block (mean ± SD)	RBB		OBB		SBB	
			Quantity	Percentage (%)	Quantity	Percentage (%)	Quantity	Percentage (%)
Green		17.3 ± 6.6	416	30.06	416	30.06	0	0
Yellow	A1	4.4 ± 2.6	106	7.66	58	4.19	48	3.47
	B1	1.3 ± 1.6	32	2.31	16	1.15	16	1.16
	C1	21.1 ± 7.9	506	36.56	195	14.09	311	22.47
Red	A2	2.6 ± 3.1	62	4.48	0	0	62	4.48
	B2	0.3 ± 0.4	6	0.43	0	0	6	0.43
	C2	10.7 ± 4.9	256	18.50	0	0	256	18.50
Total			1384	100.00	685	49.49	699	50.51

all other levels (both $p < 0.001$). For the response time, opposite trends were observed, where level 1 was significantly faster than all other levels (both $p < 0.001$) and level 3 was significantly slower than all other levels (both $p < 0.001$).

B. Baseline Detection

Table II summarizes the number of blocks in RBB, OBB, and SBB for each zone. The largest number of blocks was identified in C1 zone (506 blocks; 36.5%), followed by the green zone (416 blocks; 30.1%), and C2 zone (256 blocks; 18.5%), equivalent to an overall of 1178 blocks (85.1%). A minimal number of blocks were identified in A1, B1, A2, and B2 zones, contributing to the remaining 206 blocks (14.9%).

A total of 416 blocks in the green zone (30.1%) were detected with an optimal baseline. Meanwhile, 269 (19.4%) and 375 (27.1%) blocks in the yellow zone were detected with an optimal and suboptimal baseline, respectively. The remaining 324 blocks in the red zone (23.4%) were detected with a suboptimal baseline. For those blocks detected in the yellow and red zones, we further quantified the number of blocks where $[x(t), y(t)]$ has ever reached the green zone in any time point over the rest period. Such a trend was found in 293 (45.5% of 644) and 113 (34.9% of 324) blocks for the yellow and red zones, respectively. The total blocks in OBB and SBB across subjects were ranged from 18 to 42 and 17 to 40 (see Table III), respectively.

C. Block-Averaged Time-Series fNIRS Signals

Fig. 5 displays the block-averaged time-series fNIRS signals in RBB, OBB, and SBB at each difficulty level over 24 subjects. In general, HbO signal showed greater changes compared to HbR. As observed in RBB, a large variability was present in the time-shape of task-evoked HR. It was worth noting that OBB was more closely resembled to the time-shape of conventional HR and showed larger changes in both HbO and HbR signals. In contrast, such time-shape was severely affected in SBB.

Although the time-shape of conventional HR was present over the task period in OBB, not all task-evoked fNIRS signals showed a proper return to baseline over the rest period. This was notably observed in subjects S03, S04, S08, S09, S12, S13, S15, S18, S19, and S21 (see Supplementary Fig. 1(b) as a sample), where the average fNIRS signals were considerably

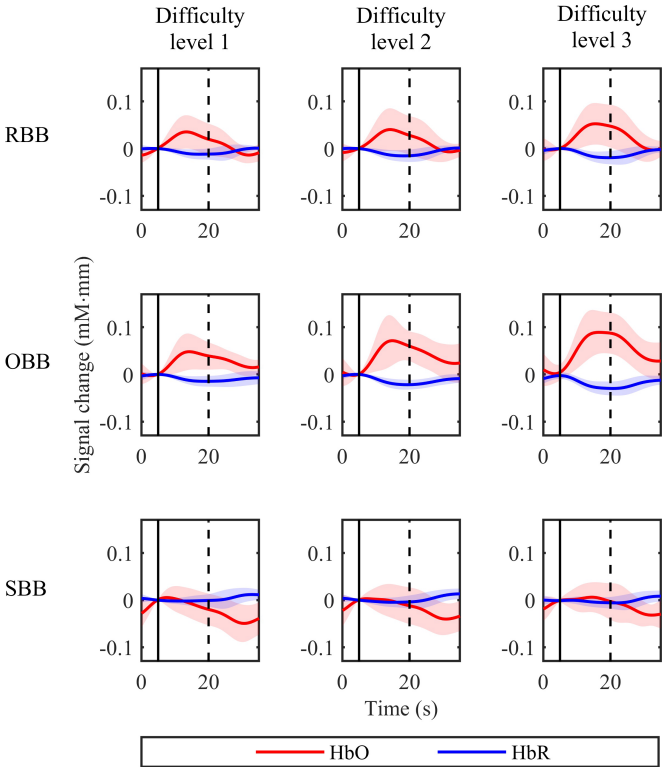

Fig. 5. Block-averaged time-series fNIRS signals in RBB, OBB, and SBB at each difficulty level over 24 subjects. The transparent shaded areas indicate SD at each time point. The task starts and finishes at time 5 and 20 s, as indicated by vertical solid and broken lines, respectively.

far from reaching the initial baseline. Hence, the chances of being detected as an optimal baseline are likely to decrease for the subsequent block. As a result, the total blocks in OBB tended to be smaller than that in SBB. On the other hand, the average fNIRS signals in OBB returned to about a similar level of its initial baseline in subjects SO1, S02, S05, S06, S07, S10, S11, S14, S16, S17, S20, S22, S23, and S24 (see Supplementary Fig. 1(a) as a sample). This explained the higher tendency of obtaining larger numbers of total blocks in OBB as compared to that in SBB.

D. Inter-Subject fNIRS Activation

Fig. 6 shows the HbO and HbR activations in RBB, OBB, and SBB at each difficulty level across 24 subjects. One-way repeated measures ANOVA returned a significant main effect of difficulty level on HbO activation in RBB [$F(2,46) = 4.3$,

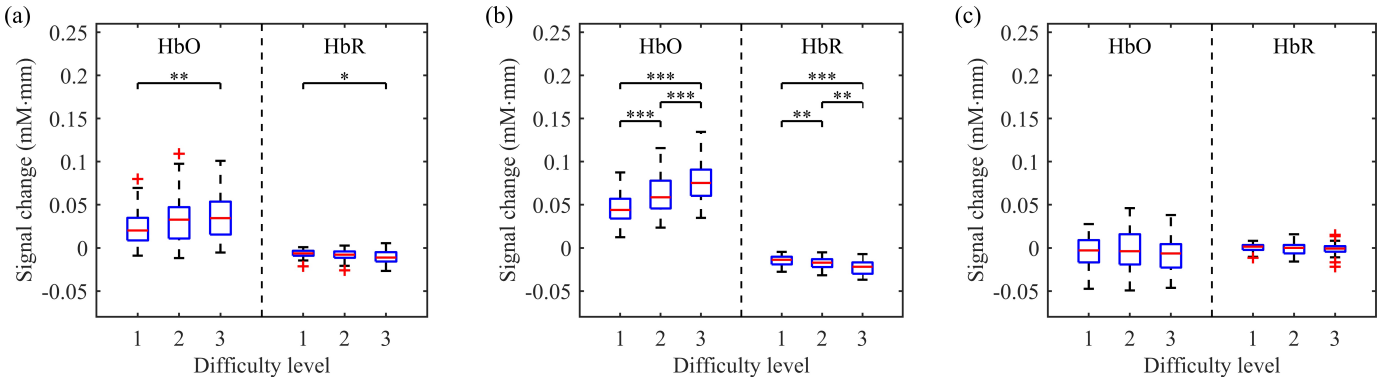

Fig. 6. HbO and HbR activations in (a) RBB, (b) OBB, and (c) SBB at each difficulty level across 24 subjects. Statistical significance of Bonferroni *post hoc* test is shown: *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

TABLE III
PEARSON'S CORRELATION COEFFICIENT r OF FNIRS ACTIVATION WITH THE RESPONSE TIME

Subject	COI	BA	Correlation r (HbO and response time)			Correlation r (HbR and response time)		
			RBB	OBB	SBB	RBB	OBB	SBB
S01	25	46	0.44*** (59)	0.54*** (42)	−0.03 (17)	−0.30* (59)	−0.39* (42)	0.07 (17)
S02	13	46	0.33* (58)	0.52** (34)	−0.30 (24)	−0.15 (58)	−0.44** (34)	0.10 (24)
S03	13	46	0.20 (59)	0.48* (22)	−0.14 (37)	−0.20 (59)	−0.51* (22)	0.01 (37)
S04	28	46	0.01 (58)	0.32 (22)	−0.10 (36)	0.01 (58)	−0.37 (22)	0.16 (36)
S05	24	45	0.27* (54)	0.45* (27)	0.07 (27)	−0.39** (54)	−0.64*** (27)	−0.11 (27)
S06	3	9	0.44*** (58)	0.48** (37)	0.27 (21)	−0.36** (58)	−0.38* (37)	−0.27 (21)
S07	13	46	0.08 (58)	0.41* (29)	−0.02 (29)	0.03 (58)	−0.41* (29)	0.20 (29)
S08	39	46	0.05 (58)	0.41 (21)	0.04 (37)	−0.16 (58)	−0.37 (21)	−0.29 (37)
S09	19	46	0.30* (59)	0.46* (22)	−0.09 (37)	−0.36** (59)	−0.39* (22)	−0.11 (37)
S10	24	45	0.46*** (59)	0.63*** (42)	0.31 (17)	−0.21 (59)	−0.46** (42)	0.25 (17)
S11	13	46	−0.07 (59)	0.16 (24)	−0.03 (35)	−0.11 (59)	−0.24 (24)	−0.16 (35)
S12	25	46	−0.08 (59)	0.58** (24)	−0.25 (35)	0.04 (59)	−0.62** (24)	0.32 (35)
S13	13	46	−0.04 (58)	0.40 (18)	−0.08 (40)	0.05 (58)	−0.46 (18)	0.14 (40)
S14	14	46	0.07 (56)	0.21 (39)	−0.30 (17)	−0.13 (56)	−0.18 (39)	0.06 (17)
S15	24	45	0.18 (58)	0.55** (23)	0.31 (35)	−0.06 (58)	−0.34 (23)	−0.11 (35)
S16	28	46	0.06 (55)	0.49* (25)	−0.12 (30)	−0.07 (55)	−0.57** (25)	0.11 (30)
S17	25	46	0.24 (58)	0.55** (26)	0.03 (32)	−0.23 (58)	−0.41* (26)	−0.09 (32)
S18	24	46	0.29* (59)	0.53** (27)	0.15 (32)	−0.27* (59)	−0.44* (27)	−0.12 (32)
S19	8	9	0.37** (55)	0.47* (20)	0.14 (35)	−0.16 (55)	−0.23 (20)	0.08 (35)
S20	39	46	0.50*** (59)	0.60*** (29)	0.19 (30)	−0.36** (59)	−0.53** (29)	−0.01 (30)
S21	13	46	0.44*** (59)	0.46* (23)	0.24 (36)	−0.38** (59)	−0.51* (23)	−0.11 (36)
S22	19	46	0.29* (58)	0.46** (40)	−0.11 (18)	−0.24 (58)	−0.33* (40)	0.10 (18)
S23	24	45	0.24 (52)	0.63*** (34)	0.19 (18)	−0.23 (52)	−0.48** (34)	−0.33 (18)
S24	34	45	0.42** (59)	0.48** (35)	0.15 (24)	−0.41** (59)	−0.47** (35)	−0.15 (24)

The number in () indicates the total blocks. Statistical significance is shown: *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

$p = 0.02$] and OBB [$F(2,46) = 72.8$, $p < 0.001$], but not in SBB [$F(2,46) = 1.0$, $p = 0.371$]. Similarly, the main effect of difficulty level on HbR activation was significant in RBB [$F(2,46) = 4.2$, $p = 0.021$] and OBB [$F(1.5,35.3) = 22.9$, $p < 0.001$], but not in SBB [$F(2,46) = 0.3$, $p = 0.709$].

In RBB, Bonferroni *post hoc* test showed that the magnitudes of both HbO and HbR activations at the level 3 were significantly greater than that of level 1 (both $p < 0.05$). In OBB, Bonferroni *post hoc* test revealed that the magnitudes of both HbO and HbR activations at the level 2 were significantly larger than that of level 1 (both $p < 0.01$). Not only that, the magnitudes of both HbO and HbR activations at the level 3 were significantly larger than that of levels 1 and 2 (all $p < 0.01$).

E. Intra-Subject Correlation Between fNIRS Activation and Response Time

Table III presents the correlations between fNIRS activation and response time at the intra-subject level. The COI and dominant BA number are listed as well. In RBB, 12 subjects showed significantly weak positive correlation of HbO activation with the response time ($r = 0.27$ to 0.50, $p < 0.05$), whereas 8 subjects demonstrated significantly weak negative correlation of HbR activation with the response time ($r = -0.27$ to -0.41, $p < 0.05$). Extracting OBB using the proposed method demonstrated an improvement in the correlation strength, where 19 subjects showed significantly weak to moderate positive correlation of HbO activation with the response time ($r = 0.41$ to 0.63, $p < 0.05$), and 17 subjects showed significantly weak to moderate negative correlation of HbR activation with the response time ($r = -0.33$ to -0.64, $p < 0.05$).

On the other hand, there was no significant correlation in SBB for both HbO and HbR activations with the response time (all subjects: $p > 0.05$). Overall, the correlation coefficient Z_r between response time and HbO/HbR activation across 24 subjects are shown in Supplementary Fig. 2. The magnitude of positive/negative correlation between response time and

HbO/HbR activation in OBB was significantly larger than that in RBB and SBB (all $p < 0.001$).

IV. Discussion

In this study, we proposed a method using vector phase analysis to detect the baseline state of the brain activation, with the ultimate aim of optimizing the estimation of mental workload. Despite the difference in the number of detected OBB and SBB for every subject, our findings reached similar conclusions: (i) the large variations observed in the task-evoked HR were contributed by the difference in baseline state and (ii) the blocks detected with an optimal baseline demonstrated a significant improvement in the estimation of mental workload.

A. Behavioral Data

As the difficulty level of the mental arithmetic task increases, subjects required more time to mentally compute the arithmetic questions. Therefore, they solved lesser questions within the task period at higher difficulty level. This tends to increase the mistake, leading to a drop in the performance score. Such behavioral outcomes were expected and in agreement with the previous studies utilizing multiple WM loads [15], [16], [35]. As reflected in the very minor number of excluded blocks, subjects showed that they were able to focus and perform the task despite a lengthy brain training session. This further justified the validity of using these datasets for the assessment of our proposed method.

B. Baseline Detection

It has been a well-known and widely-accepted finding on the time-shape of conventional HR in the activation region, where HbO/HbR increases/decreases after functional stimulation and gradually returns to baseline during the control period [19]. Note that, the COI of every subject was either located on dorsolateral (BAs 9 and 46) or ventrolateral (BA 45) PFC. This was expected as both regions are highly responsible for the WM processes related to arithmetic functions [36], [37]. Ideally, from the task onset to the end of rest period, the trajectory of fNIRS signals is supposed to follow such path: green $\rightarrow$ C1 $\rightarrow$ C2 $\rightarrow$ C1 $\rightarrow$ green. With the proper estimation of $\mathrm{HbO}_{\mathrm{base}}^{n-1}$ and $\mathrm{HbR}_{\mathrm{base}}^{n-1}$ for the execution of n^{th} block, the proposed method returned significantly high percentage of blocks identified in green, C1, and C2 zones (85.1%). At present, the thresholds of 0.3 and 0.7 were chosen empirically. These thresholds can be further optimized to improve the overall performance of the proposed method. Conversely, a minor number of blocks were identified in A1, A2, B1, and B2 zones (14.9%). This may occur when there is a large lagging between HbO and HbR signals in following the time-shape of conventional HR.

One may argue that 15-s rest period may not be sufficiently long enough to ensure the complete return of HR to baseline. Here, we showed that 45.5% and 23.4% of the blocks in the yellow and red zones had been in the green zone, respectively. This implies that even when the HR reaches an optimal baseline before the end of rest period, it may not sustain at that particular level. The systemic fluctuation (both HbO and HbR synchronously change in the same direction) and spontaneous brain activity (HbO and HbR change in the opposite direction) are likely to coexist during the rest period [1], [19]. The proposed method took into account the influences from both cases, whereby the thresholds and zones of the vector phase diagram were properly designed and optimized for the detection of baseline state. In this case, it is very likely that the detection of suboptimal baseline was resulted from the spontaneous or ongoing brain activity [6]–[8].

There are many other factors that have been suggested to account for the large variability in the time-shape of task-evoked HR, such as the experimental design [38], task complexity [38], and the systemic physiological changes [19], [39]. Our findings demonstrated that the baseline state is also one of the factors contributing to the large variability in the time-shape of task-evoked HR, agreeing with the previous studies [6], [9], [13]. The time-shape of conventional HR was certainly affected when the HR did not return to an optimal baseline, which can increase the false negatives in identifying the brain activation [13]. Overall, a different number (18–42) of blocks detected with an optimal baseline was obtained across subjects, implying that some subjects were able to rest at most of the time while some had serious difficulty to rest their minds. A part of the blocks showed that the task-evoked HR sustained throughout the rest period and did not completely return to baseline. This situation was observed in other studies as well [13], [40], suggesting that more time is needed for the task-evoked HR to return to baseline.

C. Mental Workload Estimation

A number of cognitive tasks with multiple difficulty levels have been applied in fNIRS and fMRI studies for the assessment of mental workload, reaching a general conclusion that the brain activation in WM-related regions increases along with the increase in difficulty level [15], [16], [18], [32]. After detecting the blocks with an optimal baseline, such trends were better observed in both HbO and HbR activations. On top of that, we took a step further to examine the relationship between fNIRS activation and response time at the intra-subject level. It is worth noting that, despite the variations in the number of detected blocks with an optimal baseline, a majority of them demonstrated an improvement in the correlation strength for both HbO and HbR activations with the response time. These findings demonstrate the reliability of fNIRS modality in single-trial estimation of mental workload, providing that the hemodynamic baseline is at optimal state prior to functional stimulation.

Judging from our findings, the presence of systemic physiological change, such as from the scalp hemodynamics, may not critically affect the findings as it is reported to have a much more uniform change [41], [42]. In addition, both HbO and HbR signals were considered for the identification of COI to avoid misinterpretation of brain activity [29]. Hence, the activation difference among the task periods should mainly reflect the changes in neuronal activity associated with the WM loads. At present, there are many features that have been extracted from fNIRS signals for task-related and mental workload classifications using machine learning approaches,

e.g. slope, mean, peak, variance, skewness, and kurtosis [16], [43]–[45]. To the best of our knowledge, none of the studies have considered the variability arising from the baseline state. As such, it is highly anticipated that the classification performance can be further enhanced when incorporating with our proposed method.

In [3], a different set of time intervals has been selected for the baseline correction (commonly ranging from 0–2 s to 0–10 s before the task onset) and task-evoked activation (entire task or peak activation period). The selection of these time periods was shown to greatly affect the findings and the conclusion drawn [3], [46]. This issue may worsen when there is a large HR variability observed in the task and rest periods. In this study, the entire task period was selected as every change in the HR during the task period is reflecting the change evoked by the functional stimulation. Note that, we have additionally analyzed the activation by choosing several time periods for the baseline correction (data not shown), up to 0–5 s from the current 0–0.5 s. We observed that the correlation strength reduced with the increase in the time period used for baseline correction, implying the large HR variability in the rest period can affect the computation of task-evoked activation to a certain degree. Yet, the overall improvement was still observed in OBB.

D. Limitations and Future Perspective

A total of 7 subjects did not have a COI, where the measured fNIRS signals were poorly resembled the time-shape of conventional HR. Based on the visual observation in the WM-related regions, a majority of the rest periods showed a stronger hemodynamic activity than that of the task periods. This may imply a complete failure of using the rest as a control. Applying the proposed method to these data led to high detection number of SBB, which makes it not suitable for the assessment. Besides, the issues with the selection of time periods for the baseline correction and task-evoked activation need to be further investigated.

Although our findings have demonstrated an overall improvement in the estimation of mental workload, rejecting a large number of blocks detected with a suboptimal baseline may not be a feasible approach when the block repetitions are limited. Taking a future perspective, we see a need to implement this method in a real-time fNIRS-based BCI as a means to estimate the optimal duration for the rest period, i.e. ensuring the baseline state is at optimal before initiating a new functional stimulation. Doing so, we believe that the rejection rate can be greatly minimized or even ensure that all blocks are accepted for analysis. This method is also suitable for other cognitive tasks used in block-design fNIRS studies.

V. Conclusion

Taking the consideration of brain activation evoked during the functional stimulation, we proposed a method using vector phase analysis to detect its baseline state as being optimal or suboptimal. A vector phase diagram was designed and customized for each functional stimulation. In addition, a set of design criteria was applied to identify the location of baseline, which acts as the vector origin. The proposed method was applied to a continuous block-design fNIRS dataset with 60 block repetitions. Despite the unequal numbers in detecting baseline state, our findings emphasize the importance of ensuring an optimal baseline before the onset of functional stimulation. The blocks detected with an optimal baseline demonstrated a better resemblance with the time-shape of conventional HR, leading to an improvement in the estimation of mental workload.

In short, the findings supported our hypothesis that the estimation of mental workload improves when selecting the neuronal-related HR associated with an optimal baseline. The existing technique used for removing systemic activity, such as the implementation of shorter source-detector separation [19], requires an extra cost and effort to fit the additional probes. The proposed method offers a solution to this issue as it is practically useful without any additional hardware and also because of the importance of improving the significance level in statistical analysis. Considering the variability in HR among different individuals is inevitable, this method is anticipated to enhance and improve the data analysis in fNIRS-based BCI.

References

[1] K. Uludağ and P. Blinder, "Linking brain vascular physiology to hemodynamic response in ultra-high field MRI," *NeuroImage*, vol. 168, pp. 279–295, Mar. 2018.
[2] Z. Y. Shan *et al.*, "Modeling of the hemodynamic responses in block design fMRI studies," *J. Cerebral Blood Flow Metabolism*, vol. 34, no. 2, pp. 316–324, Feb. 2014.
[3] F. Herold, P. Wiegel, F. Scholkmann, and N. Müller, "Applications of 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.
[4] T. Arichi *et al.*, "Development of BOLD signal hemodynamic responses in the human brain," *NeuroImage*, vol. 63, no. 2, pp. 663–673, Nov. 2012.
[5] K. L. West *et al.*, "BOLD hemodynamic response function changes significantly with healthy aging," *NeuroImage*, vol. 188, pp. 198–207, Mar. 2019.
[6] C. E. Stark and L. R. Squire, "When zero is not zero: The problem of ambiguous baseline conditions in fMRI," *Proc. Nat. Acad. Sci. USA*, vol. 98, no. 22, pp. 12760–12766, 2001.
[7] J. Smallwood and J. W. Schooler, "The restless mind," *Psychol. Bull.*, vol. 132, no. 6, p. 946, 2006.
[8] M. F. Mason, M. I. Norton, J. D. Van Horn, D. M. Wegner, S. T. Grafton, and C. N. Macrae, "Wandering minds: The default network and stimulus-independent thought," *Science*, vol. 315, no. 5810, pp. 393–395, Jan. 2007.
[9] D. A. Gusnard and M. E. Raichle, "Searching for a baseline: Functional imaging and the resting human brain," *Nature Rev. Neurosci.*, vol. 2, no. 10, pp. 685–694, Oct. 2001.
[10] A. Arieli, A. Sterkin, A. Grinvald, and A. Aertsen, "Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses," *Science*, vol. 273, no. 5283, pp. 1868–1871, Sep. 1996.
[11] S. Makeig, "Dynamic brain sources of visual evoked responses," *Science*, vol. 295, no. 5555, pp. 690–694, Jan. 2002.
[12] X. Liu, X.-H. Zhu, and W. Chen, "Baseline BOLD correlation predicts individuals' stimulus-evoked BOLD responses," *NeuroImage*, vol. 54, no. 3, pp. 2278–2286, Feb. 2011.
[13] M. Uga, I. Dan, T. Sano, H. Dan, and E. Watanabe, "Optimizing the general linear model for functional near-infrared spectroscopy: An adaptive hemodynamic response function approach," *Neurophotonics*, vol. 1, no. 1, Aug. 2014, Art. no. 015004.
[14] J. A. Meltzer, M. Negishi, and R. T. Constable, "Biphasic hemodynamic responses influence deactivation and may mask activation in block-design fMRI paradigms," *Hum. Brain Mapping*, vol. 29, no. 4, pp. 385–399, 2008.

[15] H. Ayaz, P. A. Shewokis, S. Bunce, K. Izzetoglu, B. Willems, and B. Onaral, "Optical brain monitoring for operator training and mental workload assessment," *NeuroImage*, vol. 59, no. 1, pp. 36–47, Jan. 2012.

[16] C. Herff, D. Heger, O. Fortmann, J. Hennrich, F. Putze, and T. Schultz, "Mental workload during n-back task—Quantified in the prefrontal cortex using fNIRS," *Frontiers Hum. Neurosci.*, vol. 7, p. 935, Jan. 2014.

[17] H. Aghajani, M. Garbey, and A. Omurtag, "Measuring mental workload with EEG+fNIRS," *Frontiers Hum. Neurosci.*, vol. 11, p. 359, Jul. 2017.

[18] A. M. Owen, K. M. Mcmillan, A. R. Laird, and E. Bullmore, "N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies," *Hum. Brain Mapping*, vol. 25, no. 1, pp. 46–59, 2005.

[19] F. Scholkmann *et al.*, "A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology," *NeuroImage*, vol. 85, pp. 6–27, Jan. 2014.

[20] A. Zafar and K.-S. Hong, "Neuronal activation detection using vector phase analysis with dual threshold circles: A functional near-infrared spectroscopy study," *Int. J. Neural Syst.*, vol. 28, no. 10, Dec. 2018, Art. no. 1850031.

[21] W. C. Ung, F. Meriaudeau, M. Kiguchi, and T. B. Tang, "Functional near-infrared spectroscopy adaptive cognitive training system (facts) for cognitive underload and overload prevention: A feasibility study," *IEEE Access*, vol. 8, pp. 172939–172950, 2020.

[22] P. Kyllonen and J. Zu, "Use of response time for measuring cognitive ability," *J. Intell.*, vol. 4, no. 4, p. 14, Nov. 2016.

[23] K. Yoshida, D. Sawamura, Y. Inagaki, K. Ogawa, K. Ikoma, and S. Sakai, "Brain activity during the flow experience: A functional near-infrared spectroscopy study," *Neurosci. Lett.*, vol. 573, pp. 30–34, Jun. 2014.

[24] G. H. Klem, H. O. Lüeders, H. H. Jasper, and C. Elger, "The ten-twenty electrode system of the international federation," *Electroencephalogr. Clin. Neurophysiol.*, vol. 52, no. 3, pp. 3–6, 1999.

[25] H. Sato *et al.*, "A NIRS–fMRI investigation of prefrontal cortex activity during a working memory task," *NeuroImage*, vol. 83, pp. 158–173, Dec. 2013.

[26] S. Brigadoi *et al.*, "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.

[27] D. T. Delpy, M. Cope, P. van der Zee, S. Arridge, S. Wray, and J. Wyatt, "Estimation of optical pathlength through tissue from direct time of flight measurement," *Phys. Med. Biol.*, vol. 33, no. 12, p. 1433, 1988.

[28] 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.

[29] G. A. Z. Morais *et al.*, "Non-neuronal evoked and spontaneous hemodynamic changes in the anterior temporal region of the human head may lead to misinterpretations of functional near-infrared spectroscopy signals," *Proc. SPIE*, vol. 5, no. 1, 2017, Art. no. 011002.

[30] J. Ye, S. Tak, K. Jang, J. Jung, and J. Jang, "NIRS-SPM: Statistical parametric mapping for near-infrared spectroscopy," *NeuroImage*, vol. 44, no. 2, pp. 428–447, Jan. 2009.

[31] K. J. Friston, P. Fletcher, O. Josephs, A. Holmes, M. D. Rugg, and R. Turner, "Event-related fMRI: Characterizing differential responses," *NeuroImage*, vol. 7, no. 1, pp. 30–40, Jan. 1998.

[32] M. Causse, Z. Chua, V. Peysakhovich, N. D. Campo, and N. Matton, "Mental workload and neural efficiency quantified in the prefrontal cortex using fNIRS," *Sci. Rep.*, vol. 7, no. 1, p. 5222, Dec. 2017.

[33] S. H. Fairclough, C. Burns, and U. Kreplin, "FNIRS 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.

[34] R. A. Fisher, "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population," *Biometrika*, vol. 10, no. 4, pp. 507–521, May 1915.

[35] F. A. Fishburn, M. E. Norr, A. V. Medvedev, and C. J. Vaidya, "Sensitivity of fNIRS to cognitive state and load," *Frontiers Hum. Neurosci.*, vol. 8, p. 76, Feb. 2014.

[36] V. Menon, S. M. Rivera, C. D. White, G. H. Glover, and A. L. Reiss, "Dissociating prefrontal and parietal cortex activation during arithmetic processing," *NeuroImage*, vol. 12, no. 4, pp. 357–365, Oct. 2000.

[37] S. M. Rivera, A. L. Reiss, M. A. Eckert, and V. Menon, "Developmental changes in mental arithmetic: Evidence for increased functional specialization in the left inferior parietal cortex," *Cerebral Cortex*, vol. 15, no. 11, pp. 1779–1790, Nov. 2005.

[38] C. Issard and J. Gervain, "Variability of the hemodynamic response in infants: Influence of experimental design and stimulus complexity," *Develop. Cognit. Neurosci.*, vol. 33, pp. 182–193, Oct. 2018.

[39] I. Tachtsidis and F. Scholkmann, "False positives and false negatives in functional near-infrared spectroscopy: Issues, challenges, and the way forward," *Neurophotonics*, vol. 3, no. 3, Mar. 2016, Art. no. 031405.

[40] J. Shin, A. von Lühmann, D.-W. Kim, J. Mehnert, H.-J. Hwang, and K.-R. Müller, "Simultaneous acquisition of EEG and NIRS during cognitive tasks for an open access dataset," *Sci. Data*, vol. 5, no. 1, Dec. 2018, Art. no. 180003.

[41] X. Zhang, J. A. Noah, and J. Hirsch, "Separation of the global and local components in functional near-infrared spectroscopy signals using principal component spatial filtering," *Neurophotonics*, vol. 3, no. 1, Feb. 2016, Art. no. 015004.

[42] T. Sato *et al.*, "Reduction of global interference of scalp-hemodynamics in functional near-infrared spectroscopy using short distance probes," *NeuroImage*, vol. 141, pp. 120–132, Nov. 2016.

[43] N. Naseer, F. M. Noori, N. K. Qureshi, and K.-S. Hong, "Determining optimal feature-combination for LDA classification of functional near-infrared spectroscopy signals in brain-computer interface application," *Frontiers Hum. Neurosci.*, vol. 10, p. 237, May 2016.

[44] K.-S. Hong, M. J. Khan, and M. J. Hong, "Feature extraction and classification methods for hybrid fNIRS-EEG brain-computer interfaces," *Frontiers Hum. Neurosci.*, vol. 12, p. 246, Jun. 2018.

[45] L. G. Lim *et al.*, "A unified analytical framework with multiple fNIRS features for mental workload assessment in the prefrontal cortex," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 11, pp. 2367–2376, Nov. 2020.

[46] M. D. Pfeifer, F. Scholkmann, and R. Labruyère, "Signal processing in functional near-infrared spectroscopy (fNIRS): Methodological differences lead to different statistical results," *Frontiers Hum. Neurosci.*, vol. 11, p. 641, Jan. 2018.