Neurophotonics

Neurophotonics.SPIEDigitalLibrary.org

Usability and performance-informed selection of personalized mental tasks for an online near-infrared spectroscopy brain-computer interface

Sabine Weyand
Larissa Schudlo
Kaori Takehara-Nishiuchi
Tom Chau

SPIE.

Usability and performance-informed selection of personalized mental tasks for an online near-infrared spectroscopy brain-computer interface

Sabine Weyand,[a,b] Larissa Schudlo,[a,b] Kaori Takehara-Nishiuchi,[c] and Tom Chau[a,b,*]
[a]Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital, Toronto, Ontario M4G 1R8, Canada
[b]University of Toronto, Institute of Biomaterials and Biomedical Engineering, Ontario M5S 3G9, Canada
[c]University of Toronto, Department of Psychology, Ontario M5S 3G3, Canada

Abstract. Brain–computer interfaces (BCIs) allow individuals to use only cognitive activities to interact with their environment. The widespread use of BCIs is limited, due in part to their lack of user-friendliness. The main goal of this work was to develop a more user-centered BCI and determine if: (1) individuals can acquire control of an online near-infrared spectroscopy BCI via usability and performance-informed selection of mental tasks without compromising classification accuracy and (2) the combination of usability and performance-informed selection of mental tasks yields subjective ease-of-use ratings that exceed those attainable with prescribed mental tasks. Twenty able-bodied participants were recruited. Half of the participants served as a control group, using the state-of-the-art prescribed mental strategies. The other half of the participants comprised the study group, choosing their own personalized mental strategies out of eleven possible tasks. It was concluded that users were, in fact, able to acquire control of the more user-centered BCI without a significant change in accuracy compared to the prescribed task BCI. Furthermore, the personalized BCI yielded higher subjective ease-of-use ratings than the prescribed BCI. Average online accuracies of $77 \pm 12.9\%$ and $73 \pm 12.9\%$ were achieved by the personalized and prescribed mental task groups, respectively. [DOI: 10.1117/1.NPh.2.2.025001]

Keywords: near-infrared spectroscopy; brain-computer interface; personalized mental tasks; user selected mental tasks; user-centered; ease-of-use.

Paper 15002R received Jan. 13, 2015; accepted for publication Apr. 10, 2015; published online May 12, 2015.

1 Introduction

1.1 *Brain–Computer Interfaces*

Brain–computer interfaces (BCIs) allow individuals to interact with their environment using only cognitive activities.[1–4] BCIs exploit a user's brain signals for external device control without requiring intentional muscle activations or peripheral nervous system responses.[3,5] The potential applications of BCIs are vast. BCIs can be used by able-bodied individuals for gaming, entertainment, and to accelerate learning.[1,6] BCIs can also be used by individuals with severe motor impairments as a means of communication, as a way of controlling a wheel chair for mobility, or as a method for controlling devices in their environment.[4–7] Individuals with amyotrophic lateral sclerosis, spinal cord injuries, brain stem stroke, complete paralysis, or muscular dystrophy among other debilitating conditions stand to benefit from BCIs.[6,8] Indeed, BCIs have the potential to significantly increase the quality of life for patients with severe motor impairments.[9]

A BCI consists of an input, a translation algorithm, and an output. The input to a BCI is the physiological signal that is being harnessed. The input can be further broken down into the access modality, which refers to how the signal is collected, and the access pathway, which refers to how a change in that signal is evoked. After the signal is collected, the translation algorithm processes the signal to remove noise, extracts features, extracts the most discriminant features, and trains a classifier to predict the class to which a case belongs. Finally, the output of the BCI involves the control of an external device.[5] For a given BCI, any of the above factors can be modulated in order to improve the BCI. Most papers focus on improving the translation methods, while only a few papers focus on improving the input. The focus of this research is on improving the input access pathway.

1.2 *Near-Infrared Spectroscopy*

Near-infrared spectroscopy (NIRS) is a noninvasive, safe, portable, and low-cost optical neural imaging technique that measures hemodynamic brain activity.[2,3,6,10] Despite temporal limitations, NIRS offers several advantages over other noninvasive BCI access modalities, including, for example, gel and paste-free donning and flexibility of measurement environments. For further discussion of the relative merits of NIRS as an access modality, refer to Refs. 3, 6, 8, and 10. NIRS works by measuring the changes in the absorption of near-infrared light that travels through the skin, periosteum, skull, meninges, and the cerebral cortex of the brain. The amount of light that is absorbed varies with the amount of oxygen in the blood. Through a mechanism known as neurovascular coupling,

*Address all correspondence to: Tom Chau, E-mail: tom.chau@utoronto.ca

2329-423X/2015/$25.00 © 2015 SPIE

areas of the brain that are active have an increase in oxygenated hemoglobin (HbO), an increase in total hemoglobin (tHb), and a decrease in deoxygenated hemoglobin (Hb).[11–13] However, other coupling trends have also been reported.[14–20] NIRS provides an indirect measure of cognitive activity by ascertaining the changes in the concentration of HbO and Hb in the brain.[2,3]

1.3 *Prescribed Mental Tasks*

NIRS is a promising access modality; however, to date, little research has been done on the access pathways accompanying this access modality.[21] Currently, to the best of our knowledge, most NIRS-BCI studies have used prescribed mental activation tasks. The tasks used to control the BCI are chosen by researchers based on previous studies showing differentiability in the activation or deactivation caused by a specific set of tasks. By discriminating between the changes in the NIRS signal resulting from the user performing each task, one is able to control the binary BCI. Several different mental tasks have been used in past NIRS-BCI studies, including mental math,[2,9,14,22–29] mental singing,[9,23,24] word generation,[22,26,28] memory,[6,26,28] mental counting,[30] mental rotation,[22] concentration,[31] motor imagery,[3,8,10,32,33] and rest.[2,6,9,22,24,25,27,29,30]

1.4 *Motivation for User-Selected Personalized Mental Tasks*

An alternative to using prescribed mental tasks is to use personalized mental tasks, where each user has a set of tasks selected specifically for him or her. To date, the exploration of personalized mental tasks in NIRS-BCIs is limited. However, personalized tasks have been explored in fMRI[34] and EEG[21,35,36] BCI studies. It appears that in all fMRI and EEG BCI personalized task studies, the actual tasks were selected by the researcher based solely on performance with the aim of improving the BCI accuracy. Overall, these studies conclude that there is significant intersubject variability in brain activation elicited by the same mental tasks and cognitive processes; and as a result, the tasks that are most effective for controlling a BCI vary among users.[21,34–36]

Although improvements in accuracy are important, improving the BCIs usability has also been concluded in literature to be vitally important. In a review of the first international meeting devoted to BCI research and development, Wolpaw et al. described that the widespread application of BCI-based communication systems will depend on cost, ease of training, ease-of-use, and user satisfaction.[5] Additionally, Bos et al. found that ease-of-use was the second most important factor after accuracy in BCI acceptance.[37,38] Furthermore, a study by Holz et al. found that ease-of-use was one of the most important aspects of the BCI for four severely motor-restricted end-users.[39] From these studies, it can be concluded that BCI usability is very important and even if classification accuracies are very high, if users dislike performing their assigned tasks they are not likely to use the BCI.[5,37] Improving the ease-of-use of a BCI could result in increased user satisfaction and user-friendliness, which could lead to increased adoption and decreased BCI abandonment.[5,37,38]

Currently, when using prescribed tasks or personalized tasks chosen solely based on accuracy, users often find the assigned tasks not suitable, unenjoyable, or difficult to perform.[37,40,41] For example, math tasks may not be suitable for individuals who have difficulty with, minimal knowledge of, or a dislike for arithmetic.[40] Personalized mental tasks that are chosen by the user based on both performance and usability could result in the development of a more user-centered BCI that is easier to use and more enjoyable.

We previously conducted an offline study to compare four mental task frameworks: two user-selected personalized mental task frameworks, a researcher-selected personalized mental task framework, and a prescribed mental task framework.[42] We showed that user-selected personalized tasks have the potential to yield higher perceived ease-of-use ratings.[42] However, further studies are needed to verify the value of personalized tasks by comparing personalized and prescribed task selection schemes, particularly in an online paradigm. To the best of our knowledge, no other NIRS-BCI study has explored the use of personalized tasks and no other BCI studies have allowed users to choose personalized tasks based on both performance and ease-of-use.

1.5 *Objectives*

The aim of this research was to develop a more user-centered personalized mental task access pathway for an NIRS-BCI that allows individuals to choose tasks based on their performance and subjective ease-of-use ratings. This study pursues two research questions: (1) Can individuals acquire control of an online NIRS-BCI via usability and performance-informed selection of mental tasks without compromising classification accuracy? (2) Can the combination of usability and performance-informed selection of mental tasks for NIRS-BCI control yield subjective ease-of-use ratings that exceed those attainable with prescribed mental tasks?

2 Methods

2.1 *Participants*

Twenty able-bodied subjects (eight male) between the ages of 16 and 40 were recruited from the staff and students at Holland Bloorview Kids Rehabilitation Hospital (Toronto, Canada). All participants were right-handed according to the Edinburgh handedness test.[43] Participants had normal or corrected-to-normal vision and had no known motor impairments, degenerative disorders, cardiovascular disorders, trauma-induced brain injuries, drug or alcohol-related conditions, psychiatric conditions, respiratory disorders or metabolic disorders. Participants were asked not to smoke or drink alcoholic or caffeinated beverages 3 h prior to each data collection session. The study was conducted with informed consent and with ethics approval from the Holland Bloorview Kids Rehabilitation Hospital and the University of Toronto.

Half of the participants were randomly allocated to the study group and chose their own personalized mental tasks, and the other half to the control group, which were assigned prescribed mental tasks. Since gender,[44–47] handedness,[46,47] and age[48–50] have been shown to affect NIRS measurements, the two groups were matched based on these three criteria. One participant (male) from the personalized mental task group was excluded from all results analysis since he was not able to follow the experimental protocol. It is noted that the participants in the study group went on to partake in 10 more sessions. Data from those sessions were not used in the present analysis; however, it was used to explore self-regulation as an alternative NIRS-BCI access pathway in Ref. 51.

2.2 *Experimental Setup*

The NIRS data were collected using a multichannel frequency-domain NIRS system with a sampling rate of 31.25 Hz (Imagent Functional Brain Imaging System from ISS Inc., Champaign, IL).[52] The NIRS system was used to measure the blood oxygen content from the prefrontal cortex (PFC). The PFC is involved in higher brain functions, including logical thinking, planning, and emotion.[26,53]

Five laser diodes, each emitting 690 and 830 nm light, and three photomultiplier tube detectors attached to a headband were used. The headband was made out of a rubber polymer (3M 9900 series), which is comfortable on the skin and easily conforms to the shape of any head. Black fabric was sewn on the outside of the headband to create tight, opaque pockets within which the light sources and detectors were fit. These pockets secured the sources and detectors, minimizing their motion while maximizing their contact with the head. The headband was centered on the participant's forehead with reference to the nose and was placed directly above the eyebrows, as shown in Fig. 1(a).

The sources and detectors were arranged in a trapezoidal shape. Each source and adjacent detector was separated by a distance of 3 cm. This distance corresponds to a penetration depth of ~2.5 cm, which has been shown to reach the outer layer of the cerebral cortex.[14,54,55] Several other NIRS-BCI studies have also used a source—detector separation distance of 3 cm over the PFC.[9,14,23–25,27] The source–detector configuration allowed for the interrogation of nine discrete locations. A schematic diagram of the configuration and points of interrogation is shown in Fig. 1(b).

Neurofeedback was provided during all sessions in the form of: (1) a trapezoid topographic image showing the real-time changes in blood oxygenation levels over the PFC and (2) a ball that rose and fell with the average change over the entire interrogation area. The feedback was updated every 125 ms and was calculated using cubic interpolation of the HbO values between the points of interrogation. HbO was chosen for the feedback since it has been cited to be more indicative of activity than Hb and tHb.[3,10] Participants were informed that the red color on the feedback represented an increase in hemodynamic activity, whereas the blue color represented a decrease in hemodynamic activity. The activation feedback is shown in Fig. 2.

2.3 *Personalized Task Measures*

In this study, 11 possible mental tasks were considered based on their deployment in previous BCI studies or their documented ability to induce PFC activity in functional imaging studies. Each of the 11 tasks is described in Table 1. To facilitate a user-centered approach to task selection that allows one to strike a personal balance between usability and performance, two task measures were invoked, namely, a total ease-of-use score and a weighted slope (WS) score.

2.3.1 *Total ease-of-use score*

Subsequent to performing an iteration of each task, users rated their perceived ease-of-use on a five-point Likert-type scale, ranging from "very easy" to "very difficult," as per recommendations for measuring post-task usability.[65,66] A total ease-of-use

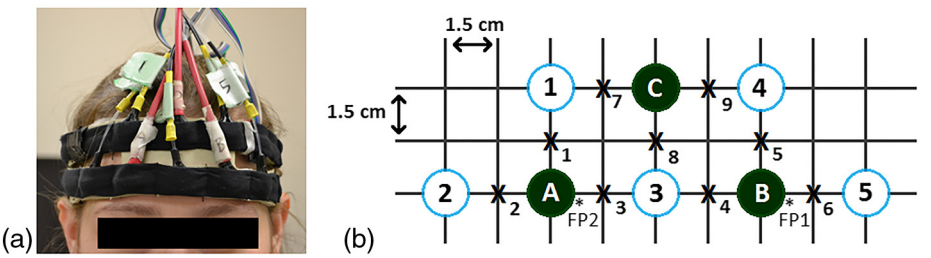

Fig. 1 (a) NIRS headband placed over the forehead. (b) Experimental source and detector configuration. The solid circles represent detectors; the open circles represent light source pairs; the x's represent points of interrogation; and the starred areas represent the FP1 and FP2 positions of the international 10 – 20 EEG system.

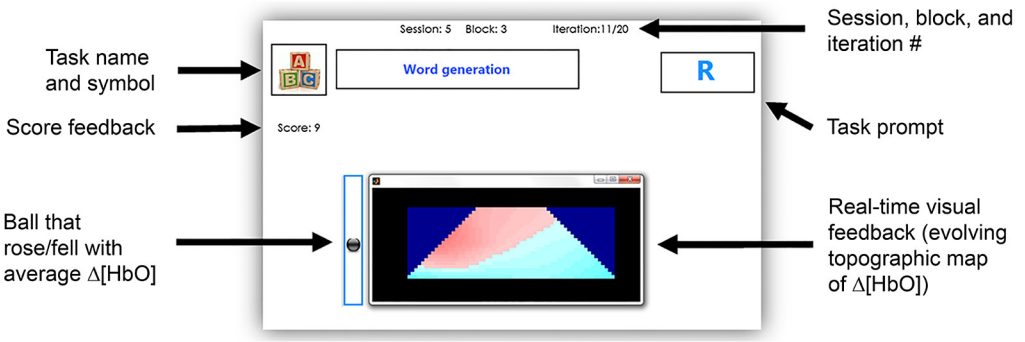

Fig. 2 User interface for session 5, blocks 2 and 3 (online classification with score feedback).

Table 1 Eleven mental tasks used in sessions 1 to 3.

Task	Description
Mental math (Math)	Participants were prompted with a math problem that appeared in the top right corner of the screen, and they were asked to repeatedly subtract a two digit number from a three digit number. For example, given 986-12; the participant would mentally evaluate 986-12=974; 974-12=962; 962-12=950; and so on. Numbers were randomly generated. This task has been used in several previous NIRS-BCI studies.[2,9,14,22–29,56,57]
Mental singing (Music)	Participants were asked to sing a song in their head. They were informed they could choose to sing any song they wanted, but they were asked to pick a song that they liked. This task has been used in previous NIRS-BCI studies.[9,22–24]
Word generation (Words)	Participants were asked to think of as many words as possible that start with a specific letter. For example, if the letter "D" appeared on the screen, the user may think of the words: dog, draft, door, deli, and so on. Letters (excluding x and z) were randomly generated. This task has been used in previous NIRS-BCI studies.[22,26,28,58]
Tangram puzzle (Rotation)	Participants were prompted with a tangram puzzle in the top right corner of the screen, and were asked to imagine rotating the pieces to make a final picture. This task was chosen because it has been shown to alter PFC activity, measured using NIRS in a non-BCI study.[59] A similar task was also used in a previous NIRS-BCI study.[22,59]
Relaxing with counting (Counting)	Participants were asked to slowly count in their heads while relaxing. A similar task has been used in a previous NIRS-BCI study.[30,57]
Happy thoughts (Happy)	Participants were asked to think about the details of a past event in their life that made them very happy. This task has been used in a previous NIRS-BCI study.[60] This task also uses episodic memory, which has been shown to alter activity in the PFC.[19]
Word color (Stroop)	Participants were prompted with a series of color names. The words were also colored, but the color of the words did not always match the written word. For example, the word blue may have been written in red ink. The participants were asked to say the real color of the word in their head. This task is commonly referred to as the stroop task. This task has been used in a previous NIRS non-BCI study.[61,62]
Visualizing the future (Future)	Participants were asked to imagine their life in five years, specifically focusing on their future day-to-day activities. This task was chosen for its potential to alter activity in the PFC. The PFC is part of the default mode network and has been shown to be activated when envisioning the future and during self-relevant mentalization.[20]
Relaxing with focus on the feedback (Focus)	Participants were asked to relax and focus on the feedback. A similar task has been used in a previous NIRS-BCI study.[31]
Motor imagery (Motor)	Participants were asked to imagine moving their arms or legs. Motor imagery has been investigated in previous NIRS-BCI studies, but strictly over the motor cortex.[3,8,33] This task was included since it has been shown that the PFC is also affected by motor imagery.[63,64]
Rest (Rest)	Participants were asked to relax and let their minds rest. This task has been used in previous NIRS-BCI studies.[2,6,9,22,24,25,27,29,30,56]

score for each task was determined as the average ease-of-use rating across all iterations of the task.

2.3.2 *WS score*

Task performance was captured by a task-specific WS score that represents the ability of a subject to consistently increase or decrease their hemodynamic activity by performing a task. Specifically, the WS score, WS_i, for the i'th task, was defined as

$$\mathrm{WS}_i = \frac{\frac{1}{N}\sum_{j=1}^{N} m_{ij}}{\sqrt{\frac{1}{N}\sum_{j=1}^{N}\left[m_{ij} - \left(\frac{1}{N}\sum_{k=1}^{N} m_{ik}\right)\right]^2}}, \quad (1)$$

where m_{ij} and m_{ik} are the slopes of the least square line of best-fit to the hemodynamic activity over time for the j'th or k'th iteration of the i'th task, and N is the number of times the task was performed. For each iteration j or k, of each task, i, an average response is computed from the hemodynamic response ($\Delta[\mathrm{HbO}]$) from each of the nine channels (i.e., nine points of interrogation). A slope value, m_{ij} or m_{ik}, is extracted from the best-fit line to each channel-averaged response. The WS score is then computed as the mean of all slopes for each iteration of a given task divided by the corresponding standard deviation, providing a measure of the tendency for a task to consistently increase or decrease hemodynamic activity.

2.4 *Data Collection Sessions*

All participants took part in five data collection sessions on five separate days, spanning a period of one week. In each session, participants were seated in front of a computer in a dimly lit room. The general protocol was the same for all sessions.

Each session started with a short warm-up period which allowed the user to become familiar with the interface. Following the warm-up, each participant took part in three data collection blocks. During each data collection block, the participant performed either 22 task intervals (sessions 1, 2, and 3) or 20 task intervals (sessions 4 and 5). Each task interval involved a task being performed for 17 s, followed by a 20 s rest. The duration of the task and rest activities was chosen on the basis of preliminary data and past work.[24,25] A schematic illustration of the overall study, session, and block structure is shown in Fig. 3.

2.4.1 *Sessions 1 to 3*

Participants performed each of the 11 tasks twice per block. The tasks were presented in a random order. Immediately after performing each task and before the 20 s rest, the user was asked to rate the task in terms of its ease-of-use and desirability for BCI control. The 20 s rest period allowed cortical hemodynamic changes from the previous task and the ease-of-use scoring to dissipate. The goal of these three sessions was to determine the magnitude of change in blood oxygenation when the participants performed each task and to determine the level of subjective enjoyment of each task. By the end of the third session, participants had performed each task 18 times (3 sessions × 3 blocks × 2 iterations), and thus $N = 18$, in Eq. (1) above.

2.4.2 *Session 4*

The control group was assigned the mental math and rest tasks irrespective of their performance in the first three sessions. These tasks represent the current state-of-the-art in NIRS-BCIs.[23,25–27] On the other hand, the study group was instructed to choose their own pair of personalized tasks from among the 11 possibilities. To inform their choice, participants were provided with their total ease-of-use rating for each task from sessions 1 to 3. In addition, for each task, participants were presented with line graphs depicting HbO concentration changes averaged over the nine points of interrogation. Graphs for the top three tasks for increasing hemodynamic activity (highest WS scores) and the best three tasks for decreasing the hemodynamic activity (lowest WS scores) were shown. Participants were asked to choose a preferred task that tended to increase their hemodynamic response, their "increasing task," and another that tended to decrease their hemodynamic response, their "decreasing task." Subsequently, participants were prompted to perform their two tasks: mental math and rest for the comparison group, and personalized tasks for the study group.

2.4.3 *Session 5*

In the first block, participants were prompted to perform their two tasks as in session 4. During the second and third blocks of the fifth session, participants continued to perform their two tasks when prompted by the interface (Fig. 2) but received the corresponding label (i.e., increase or decrease task) estimated by the computer. A score counter was updated, displaying the number of times that the computer correctly labeled (classified) the task performed by the user. On average, classification took 1 to 2 s; however, this will vary depending on the speed of the computer. An example of the user interface from session 5, block 3, is shown in Fig. 2.

2.5 *Signal Treatment*

2.5.1 *Filtering*

NIRS data are affected by various sources of physiological noise. In particular, the NIRS signal is contaminated with the Mayer wave at a frequency of 0.1 Hz, the respiration cycle at a frequency of 0.2 to 0.4 Hz, and the cardiac cycle at a frequency of 0.5 to 2 Hz.[23,67] To mitigate the influences of these noise sources, the NIRS signal was passed through a digital low-pass filter in real-time using a third-order Chebyshev infinite impulse response cascade filter with a pass-band from 0 to 0.1 Hz, a transition band from 0.1 to 0.5 Hz, a stop-band from 0.5 Hz onwards, and a pass band ripple of 0.1 dB.

2.5.2 *Calculating hemoglobin concentrations*

After filtering the data, the changes in concentrations of HbO, Hb, and tHb, i.e, $\Delta[\mathrm{HbO}]$, $\Delta[\mathrm{Hb}]$, and $\Delta[\mathrm{tHb}]$, were calculated using the modified Beer–Lambert's law.[3,5,68,69]

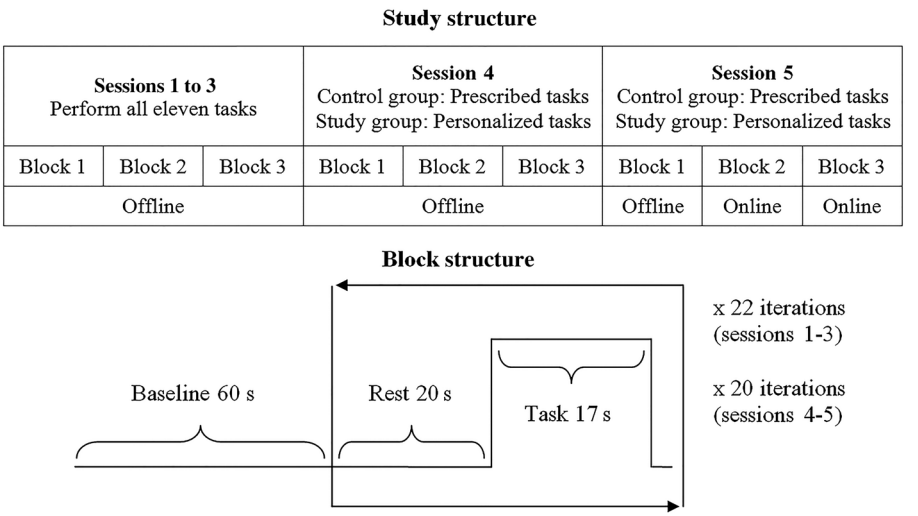

Fig. 3 Study, session, and block structure.

$$\Delta[\mathrm{HbO}] = \frac{\varepsilon_{\mathrm{Hb}}^{\lambda_2}\left(\frac{\log\left(\frac{I_B^{\lambda_1}}{I_A^{\lambda_1}}\right)}{\mathrm{DPF}^{\lambda_1}}\right) - \varepsilon_{\mathrm{Hb}}^{\lambda_1}\left(\frac{\log\left(\frac{I_B^{\lambda_2}}{I_A^{\lambda_2}}\right)}{\mathrm{DPF}^{\lambda_2}}\right)}{r\left(\varepsilon_{\mathrm{Hb}}^{\lambda_2}\varepsilon_{\mathrm{HbO}}^{\lambda_1} - \varepsilon_{\mathrm{Hb}}^{\lambda_1}\varepsilon_{\mathrm{HbO}}^{\lambda_2}\right)}, \quad (2)$$

$$\Delta[\mathrm{Hb}] = \frac{\varepsilon_{\mathrm{HbO}}^{\lambda_2}\left(\frac{\log\left(\frac{I_B^{\lambda_1}}{I_A^{\lambda_1}}\right)}{\mathrm{DPF}^{\lambda_1}}\right) - \varepsilon_{\mathrm{HbO}}^{\lambda_1}\left(\frac{\log\left(\frac{I_B^{\lambda_2}}{I_A^{\lambda_2}}\right)}{\mathrm{DPF}^{\lambda_2}}\right)}{r\left(\varepsilon_{\mathrm{Hb}}^{\lambda_1}\varepsilon_{\mathrm{HbO}}^{\lambda_2} - \varepsilon_{\mathrm{Hb}}^{\lambda_2}\varepsilon_{\mathrm{HbO}}^{\lambda_1}\right)}, \quad (3)$$

$$\Delta[\mathrm{tHb}] = \Delta[\mathrm{Hb}] + \Delta[\mathrm{HbO}], \quad (4)$$

where I_B^{λ} is the mean light intensity measured at the baseline at wavelength λ, I_A^{λ} is the light intensity measured at any given time at wavelength λ, $\varepsilon_{\mathrm{Hb}}^{\lambda}$ and $\varepsilon_{\mathrm{HbO}}^{\lambda}$ are the specific extinction coefficient of deoxygenated and oxygenated hemoglobin, respectively, at wavelength λ, DPF^{λ} is the differential path factor at wavelength λ, and r is the geometric distance between the emitter and detector. For a derivation of these equations, see Refs. 69 and 60. In this study,

$r = 3$ cm, $\varepsilon_{690\ \mathrm{nm,Hb}} = 2.1382\ \mathrm{mM^{-1}\,cm^{-1}}$,[70]
$\varepsilon_{830\ \mathrm{nm,Hb}} = 0.7804\ \mathrm{mM^{-1}\,cm^{-1}}$,[70]
$\varepsilon_{690\ \mathrm{nm,HbO}} = 0.3123\ \mathrm{mM^{-1}\,cm^{-1}}$,[70]
$\varepsilon_{830\ \mathrm{nm,HbO}} = 1.0507\ \mathrm{mM^{-1}\,cm^{-1}}$,[70]
$\mathrm{DPF}_{690\ \mathrm{nm}} = 6.51$ and $\mathrm{DPF}_{830\ \mathrm{nm}} = 5.86$.[71]

2.5.3 *Feature extraction and feature selection*

Features were extracted over four time windows (0 to 5 s, 0 to 10 s, 0 to 15 s, and 0 to 17 s). Features included the temporal changes in the three chromophores (Hb, HbO, and tHb) at each of the nine points of interrogation (4 timewindows × 3 chromophores × 9 points of interrogation = 108 features) and the spatiotemporal changes of the zero to fourth order discrete orthogonal Chebyshev image moments (4 time windows × 3 chromophores × 15 image moments = 180 features).[25] A total of 288 features were thus extracted from the data.

The temporal feature extraction involved normalizing each measured response by subtracting the mean and dividing by the standard deviation, and then determining the least square line of best-fit slope of the concentration change of each of the three chromophores over each of the four time windows. For example, the first feature was the least square line of best-fit slope of [Hb] at the first point of interrogation over the first 5 s that the task was performed. Temporal features have been previously deployed with NIRS-BCIs.[25,27,29,72,73]

To derive the spatial features, topographic images for Δ[HbO], Δ[Hb], and Δ[tHb] were generated by spatial interpolation of the data at the nine points of interrogation. Specifically, cubic interpolation at equally spaced intervals between the points of interrogation was performed, as in Ref. 25, to create a trapezoidal image, 21 pixels in height and with parallel sides 21 and 61 pixels in length. To account for intertrial variability, the pixel values were normalized to fall between 0 and 1, as in Ref. 25. To summarize the spatial changes, zero to fourth order discrete orthogonal Chebyshev polynomial image moments were extracted from each image at every instance in time. Image moments are a weighted average of the image pixel intensities and take the general form:

$$M_{mn} = \sum_{x=0}^{N_x-1}\sum_{y=0}^{N_y-1} P_m(x)P_n(y)f(x,y), \quad (5)$$

where, $f(x,y)$ is the intensity distribution of the N_x by N_y image, x and y are the pixel coordinates, m and n are the degrees (orders) of the Chebyshev polynomials, $m+n$ is the moment order, and $P_m(x)$ and $P_n(y)$ are the two-dimensional orthogonal Chebyshev polynomials,[74] calculated using Eqs. 9 and 12 and Table 2 from Ref. 74. A total of 15 image moments were extracted from each image at each time point, one for each possible permutation of moment orders of 0, 1, 2, 3, and 4, as shown in Table 2. Finally, the simple least square line of best-fit slope of each image moment signal was calculated over each time window. For example, the first feature was the least square line of best-fit slope of the change of the zeroth order moment ($m = n = 0$) of [Hb] over the first 5 s that the task was performed.

The sequential forward floating search (SFFS) algorithm was used to select a subset of eight features from the total feature set used for classifier training.[25,75] In general, SFFS uses a criterion function to assess the discriminative capabilities of each candidate feature set. Starting with an empty feature set, the algorithm sequentially adds features with the largest criterion value. At each iteration, the method also removes a previously added feature that is presently the least significant with respect to the criterion function.[75] We used the Fisher criterion to assess the discriminatory capabilities of each feature set, as in Refs. 25

Table 2 Degree m and n of each of the 15 image moments.

Image moment number	Moment order ($m+n$)	m	n
1	0	0	0
2	1	0	1
3	1	1	0
4	2	0	2
5	2	2	0
6	2	1	1
7	3	0	3
8	3	3	0
9	3	1	2
10	3	2	1
11	4	0	4
12	4	4	0
13	4	1	3
14	4	3	1
15	4	2	2

and 29. For this study, the target number of eight features was chosen on the basis of preliminary data and past work.[29,72]

2.5.4 *Pattern classification*

An ensemble of three classifiers was used to differentiate between task-induced changes in the hemodynamic response, as in Ref. 25. In particular, for each participant, three linear discriminant analysis classifiers (LDAs) were trained. In general, LDAs seek to separate classes by projecting the training samples onto a line that maximizes class separabilty.[76] The first classifier was trained using eight features selected from the 108 temporal features; the second classifier was trained using eight features selected from the 180 spatiotemporal features; and the third classifier was trained using eight features selected from all 288 features (temporal and spatiotemporal). Feature extraction and selection are described in Sec. 2.5.3. The trained classifier was used to label testing data into one of the two classes. Each classifier predicted the class of the test data, and the overall classification was determined using a majority vote,[77] a decision scheme in which the final label is taken to be the one predicted by the majority of classifiers. For example, if the decisions of the three classifiers are respectively, class 1, class 2, and class 1, the majority vote would yield class 1 as the predicted label. The data used for training and testing the classifier are described in Sec. 2.6.1. For more information on classification and LDA, please refer to Ref. 76.

2.6 *Data Analyses*

2.6.1 *Determining accuracies*

Classification of NIRS signals can either be performed offline, following the completion of data collection, or online, in real-time, as the data are being collected. In general, the aim of offline classification is to provide an estimate of how a classifier, trained on the data collected, would perform on similar future data. Offline classification also provides the ability to make adjustments to the analysis methods, such as extracting and selecting different features. By contrast, online classification involves training a classifier using previously collected data, and then predicting the class of new data in real-time as the task is being performed. Online classification enables real-time control and can provide the user with immediate feedback.

Offline accuracies were calculated retrospectively after all the offline data had been collected. Specifically, all data collected in session 4 and the first block of session 5 were pooled together, and accuracies were determined using 30 iterations of five-fold cross-validation. Cross-validation is a well-established method for statistically estimating classifier performance, namely, how well the classifier will generalize when presented with previously unseen data.[78] Specifically, cross-validation involved randomly partitioning the data into five equally sized portions (folds). Next, each fold was used as testing data, while the other four folds were used as training data. The training data were used for feature selection and classifier training, and the testing data were used to estimate classification accuracy. This process was repeated until all folds had been used for testing and five classification accuracies had been obtained. Five-fold cross-validation was then repeated 29 more times, with new, randomly partitioned folds. Finally, the 150 accuracies were averaged to provide an overall mean offline accuracy for each participant (30 iterations $\times$ 5 folds $=$ 150 accuracies).

Online accuracies were calculated in real time. The classifier was trained using all the offline data (session 4 and the first block of session 5), and each new task was classified immediately after being performed. Specifically, in the final two blocks of session 5, the data were classified using an online classifier trained on the offline data.[27]

2.6.2 *Comparison of ease-of-use*

The ease-of-use ratings for each task were summed across all 18 instances where the task was performed. These sums were then used to rank the tasks based on ease-of-use for that participant, where a rank of 1 represented the hardest task to perform and a rank of 11 represented the easiest task to perform. Finally, the ordinal ease-of-use rankings of the two groups were compared using a two-tailed Mann–Whitney U-test. For all statistical tests, normality of the data was confirmed using the Shapiro–Wilk normality test.

2.6.3 *Comparison of accuracies*

The offline and online accuracies achieved over sessions 4 and 5 by the personalized mental task group were compared to the prescribed mental task group using a two-tailed Student's t-test for two independent means ($\alpha = 0.05$). Additionally, the personalized mental task groups offline classification accuracies between the participant's personalized tasks and the state-of-the-art prescribed mental strategies (mental math and rest) at the end of session 3 were compared using a two-tailed Student's t-test for two dependent means ($\alpha = 0.05$). The offline classifications were performed using 10 iterations of five-fold cross-validation and using two extracted features from the data collected in sessions 1 to 3, which consisted of 18 data points per task.

2.6.4 *Evaluation of WS score*

In order to verify the suitability of the WS score, a Pearson correlation between the WS score and the online accuracy achieved by the participants was investigated. The total WS score for each participant was determined by the absolute difference between the WS scores for the increasing and decreasing tasks after the third session, as shown in Eq. (6)

$$\mathrm{Total}\mathrm{WS}_{ID} = |\mathrm{WS}_I - \mathrm{WS}_D|, \tag{6}$$

where I is the increasing task in a given pair and D is the decreasing task in a given pair.

2.6.5 *Evaluation of feedback*

At the end of the fifth session, participants were asked to rate on a seven-point Likert-type scale how helpful they found the continuous activation feedback to be, with 1 denoting not helpful and 7 meaning most helpful. The helpfulness of the feedback was compared between the two groups using a two-tailed Student's t-test for two independent means ($\alpha = 0.05$).

2.6.6 *Analysis of time windows of selected features*

A frequency count of the features selected from the different time windows (0 to 5 s, 0 to 10 s, 0 to 15 s, and 0 to 17 s) was conducted with all the data collected during the offline sessions. This count was completed for the selected eight features, including all 19 participants, for each of the three classifiers (temporal, spatiotemporal, and temporal combined with

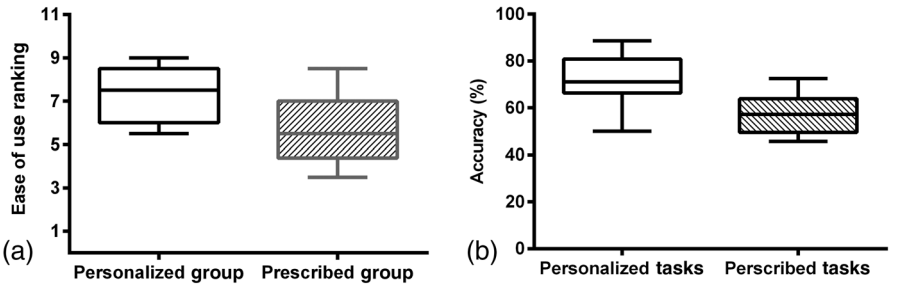

Fig. 4 (a) Ease-of-use rankings for personalized and prescribed task groups. (b) Classification accuracy of personalized tasks (chosen after session 3) and prescribed tasks (mental math and rest) for the personalized mental task group in sessions 1 to 3.

spatiotemporal). For feature selection methods, please refer to Sec. 2.5.3. In total, 456 features were considered.

3 Results

3.1 *Ease-of-Use: Personalized Versus Prescribed Tasks*

The perceived ease-of-use of the personalized task group and prescribed task group are shown in Fig. 4(a). A Mann–Whitney U-test revealed that the overall tasks' ease-of-use of the BCI was significantly higher for the personalized mental task group compared to the prescribed mental task group ($z = 2.16$, $p = 0.0308$).

A high variability was observed in the perceived ease-of-use ratings for tasks that users found to be easiest among both personalized and prescribed task groups. Each task was rated very high (5/5) by some users and very low (1 or 2/5) by other users. The intersubject variability in the tasks' ease-of-use supports the notion that different individuals find different mental tasks easy to use.

3.2 *Offline and Online Classification Accuracies*

The offline and online classification accuracies achieved in sessions 4 and 5 by the personalized mental task group and the prescribed mental task group are shown in Tables 3 and 4. Both groups achieved average online classification accuracies greater than 70%, which has been cited as the accuracy required for an effective BCI.[79] However, only the personalized mental task group achieved an average offline accuracy >70%. On average, the personalized mental tasks group achieved an offline accuracy of $75\% \pm 10.8\%$ and an online accuracy of $77\% \pm 12.9\%$, whereas the prescribed mental task group achieved an offline accuracy of $68\% \pm 12.9\%$ and an online accuracy of $73\% \pm 12.9\%$. Statistically, the classification accuracies achieved by the two groups were not significantly different, as evaluated by a two-tailed t-test for two independent means (offline accuracies: $t = 1.29$, $p = 0.213$; online accuracies: $t = 0.554$, $p = 0.587$).

For the personalized mental task group, offline analysis was conducted to compare the classification accuracies between the participant's personalized tasks and the state-of-the-art prescribed mental strategies (mental math and rest) at the end of session 3. The average offline classification accuracy for the personalized tasks was $71.8 \pm 11.5\%$ versus $57.7 \pm 8.8\%$ for prescribed tasks. A Student's t-test for two dependent means revealed that the personalized task accuracies were significantly higher than the prescribed task accuracies ($t = -2.90$, $p = 0.0198$). These results are shown in Fig. 4(b).

Table 3 Accuracies achieved by prescribed task group.

ID	Offline (%)	Online (%)
101	62.2	52.5
102	76.8	82.5
103	68.2	67.5
104	92.8	95.0
105	57.4	65.0
106	71.6	77.5
107	79.3	82.5
108	47.7	60.0
109	63.2	82.5
110	59.5	67.5
Average	67.9	73.3

Table 4 Accuracies achieved by personalized task group.

ID	Offline (%)	Online (%)
1	81.7	95.0
2	69.2	57.5
3	71.6	77.5
4	76.6	60.0
5	51.4	72.5
7	75.3	72.5
8	90.0	92.5
9	82.5	76.3
10	76.4	85.0
Average	75.0	76.5

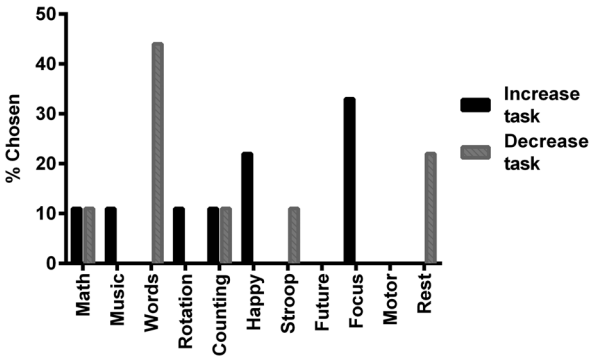

Fig. 5 Personalized tasks chosen by users to increase and decrease their hemodynamic activity.

3.3 *Variability in Personalized Tasks*

The tasks chosen by the participants of the personalized mental task group as their increase and decrease tasks are shown in Fig. 5.

Note that a variety of tasks were chosen for both increase and decrease tasks. Nine of the eleven tasks were chosen at least once, and the only tasks that were not chosen at all were motor imagery and visualizing the future. Of the 11 tasks, relaxing-with-focus was chosen most often as the increasing task, and word generation was most often chosen as the decreasing task. This variability in task choice supports the notion that different individuals prefer different mental tasks. Interestingly, mental math was chosen as an increase task by one participant and a decrease task by another participant. The same phenomenon was observed with the relaxing-with-counting task. This observation suggests a high intersubject variability in the hemodynamic response produced by each task.

A Hinton plot of the WS scores at the end of session 3 for the personalized mental task group is shown in Fig. 6. Intersubject variability in WS scores is evident; each of the 11 tasks resulted in positive WS scores in some users and negative scores in others.

As previously mentioned, two tasks were not chosen by any of the participants, motor imagery and visualizing the future. Yet, as seen in Fig. 6, they were among the top three increase and decrease tasks for all participants, except P8. Moreover, motor imagery was the top increase task for P5 and the top decrease task for P10 and visualizing the future was the top increase task for P7. This raises the question as to why these tasks were never chosen by any of the participants. Upon further analysis of the questionnaires and written comments, many of the participants did not enjoy performing these tasks. They found performing motor imagery to be cumbersome and visualizing the future to be very abstract and difficult to perform consistently.

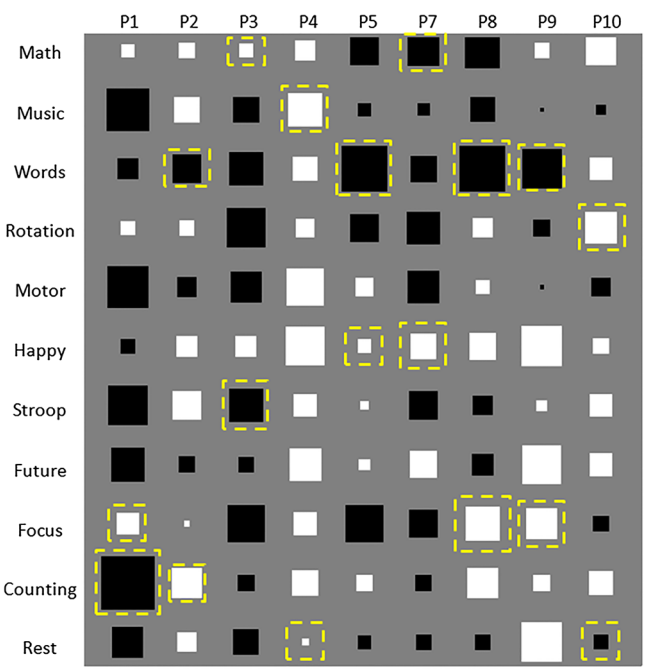

Fig. 6 Hinton diagram of weighted slope (WS) scores at the end of session 3. Positive and negative values are represented by white and black squares, respectively, and the size of each square is proportional to the magnitude of each WS score. Chosen tasks are indicated by a dashed box surrounding the corresponding black or white square. The largest square represents a magnitude of 1.63.

3.4 *Selection of Personalized Tasks Using the WS Score*

Figure 7 is a scatter plot of each participant's online classification accuracy along with their corresponding total WS score.

The WS scores have a strong positive Pearson's correlation with the online accuracy ($\rho = 0.61$, $p < 0.01$).[80,81] This suggests that there is potential in using the WS score as a measure of task suitability for controlling an NIRS-BCI.

Additionally, we computed the correlation between the offline accuracy achieved in sessions 1 to 3 to the total WS score at the end of session 3 over all participants and for all 55 pairwise combinations of tasks. A moderate positive Pearson's correlation was found between the WS scores and the offline accuracies ($\rho = 0.4$, $p < 0.001$).[80,81] This finding reinforces the potential in using the WS score as a measure of task suitability for controlling an NIRS-BCI.

3.5 *Helpfulness of Feedback*

On average, users found the feedback moderately helpful with a rating of 5.2 ± 1.2 on a seven-point Likert scale. The personalized mental task group had an average helpfulness rating of 5.8 ± 1.0, whereas the prescribed mental task group had an

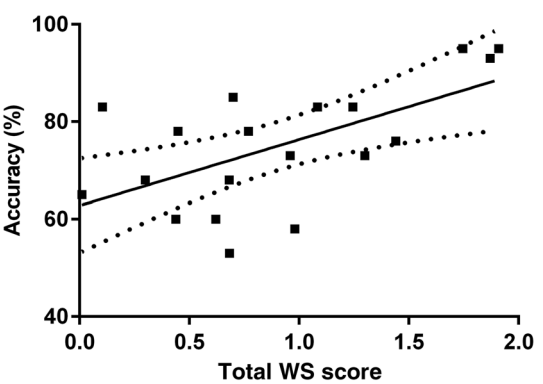

Fig. 7 Plot of online accuracy in session 5 versus WS score. A regression line was fit to the plot with a slope of 13.54 ± 4.3. The 95% confidence intervals are plotted as dotted lines.

average helpfulness rating of 4.7 ± 1.1. Based on a Student's t-test for two independent means, the personalized mental task group had a significantly higher helpfulness ratings than the prescribed mental task group ($t = 2.27$, $p = 0.036$). Moreover, all participants in the personalized mental task group, other than P7, found the feedback to be helpful (rating > 4). By contrast, only five participants in the prescribed mental task group found the feedback to be helpful (P102, P104, P105, P108, and P109). Overall, this indicates that participants in the personalized mental task group found the feedback significantly more helpful than participants in the prescribed mental task group.

3.6 *Time Window Feature Selection Analysis*

All four time windows (0 to 5 s, 0 to 10 s, 0 to 15 s, and 0 to 17 s) were frequently chosen during feature selection for each participant. Overall, the 0 to 17 s time window was chosen most often, followed by, in descending frequency of selection, the 0 to 5, 0 to 15 and 0 to 10 s time windows, as shown in Fig. 8. This result is in line with previous findings by Power et al.[24] The largest and smallest time windows were chosen most frequently likely because they capture both gradual and early changes in the hemodynamic signal. Additionally, the overall distribution of selected time windows was similar for both the personalized and prescribed mental task groups.

4 Discussion

4.1 *Ease-of-Use*

Relating to our second research question, the task usability ratings for the personalized task NIRS-BCI were found to be significantly higher than those of the prescribed task NIRS-BCI. This finding is nontrivial because participants selected tasks based on both ease-of-use and WS scores of each task. Each participant had the opportunity to choose their task by evaluating their own personal ease-of-use/effectiveness tradeoff. Incidentally, our previous offline single-group study also identified a significantly greater perceived ease-of-use for user-selected personalized mental tasks compared to prescribed tasks.[42]

The significance of developing an easier to use BCI has been well established in literature. As described in Sec. 1.4, ease-of-use is an identified enabler in the development of BCIs. Ease-of-use is recognized as one of the key attributes to the widespread application of BCI-based communication,[5] one of the most important factors in BCI acceptance,[37,38] and one of the most important aspects of the BCI for four severely motor-restricted end-users.[39] Furthermore, ease-of-use has been linked to satisfaction, which has been shown to positively impact adoption and BCI abandonment.[5,37,38] Overall, an easier to use BCI is vitally important.

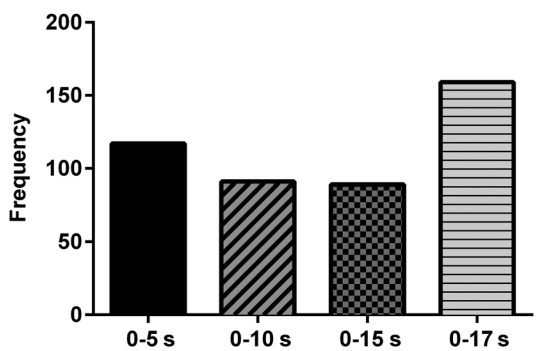

Fig. 8 Frequency of occurrence of each time window (0 to 5 s, 0 to 10 s, 0 to 15 s, and 0 to 17 s) among the selected features.

4.2 *Online and Offline Classification*

In support of our first research question, it was determined that individuals can acquire control of an online NIRS-BCI via usability and performance-informed selection of mental tasks while maintaining classification accuracies statistically comparable to those of the prescribed task group. This finding corroborates that of our previous offline single-group study, where no significant difference was observed between the accuracies of user-selected personalized mental tasks and prescribed tasks.[42]

This study adds to the expanding literature of online NIRS-BCI research. Online classification is a critical step toward real-world BCI applications and presents various challenges not applicable to offline classification, including hardware and software adaptations to allow for immediate classification, and to address classifier generalization issues.[29] The online accuracies achieved in this study are on par with those reached by Schudlo et al.,[29] and Coyle et al.,[3] and exceed the accuracies of other online NIRS-BCI studies, such as those by Chan et al.[82] and Stangl et al.[83] Our training paradigm was similar to that of previous online NIRS-BCIs (i.e., used in Ref. 29) but with fewer samples for classifier training and a shorter task performance interval of 17 s compared to 20 s used by Schudlo et al.[29] This shorter response interval can improve the communication rate and decrease the mental demand placed on BCI users.

It should be noted that it is possible that a small nonsignificant increase in the accuracy of personalized mental tasks was actually also present. The power of the online test was calculated to be only 9.8%, and the associated Cohen's d effect size was only 0.3 (a small effect). A sample size of 166 participants per group would be necessary to increase the power of this analysis to 80%. Since the effect size appears to be small and a very large number of participants would be required to detect a significant difference, the authors conclude that conducting further analysis using this design is not justified.

However, it is possible that future studies using other tasks or a different length of testing may result in significant differences. For example, with fewer tasks, there may be a larger effect size or smaller standard deviation of the personalized mental task group. It is also possible that if longitudinal data were taken for both groups, a greater difference in accuracy may emerge. The ease-of-use of selected tasks may be amplified during extended use, and this could have an effect on the BCI accuracy over time.

Additionally, it should also be noted that users chose their personalized tasks based on both subjective evaluation of performance and usability of the task. Had task choice been exclusively based on performance, a change in accuracy may have been more apparent.[42] However, our findings collectively suggest that perceived ease-of-use may trump accuracy for some users and may facilitate BCI control. For example, the benefits of personalization in initial acquisition and learning have been demonstrated in other areas of research. In education, personalization has increased learning, motivation, and depth of engagement.[84] In an air traffic control training study, researchers found that personalized adaptive task selection based on both efficiency and preference led to more efficient training than nonpersonalized task selection.[85]

In line with previous literature, users in the present study achieved significantly higher offline accuracies in some tasks than other tasks.[21,34–36] To the best of our knowledge, no other BCI research study has compared online or offline classification accuracies between a personalized and prescribed mental task group. However, one study by Dobrea et al. conducted a within group comparison of personalized mental tasks and prescribed mental tasks. In this EEG BCI study, Dobrea et al. explicitly compared the offline accuracy of the chosen personalized tasks to a set of prescribed state-of-the-art tasks. Dobrea et al. found that the best combination of four tasks from a choice of 12 tasks (chosen based on classification accuracies) achieved a greater accuracy than the state-of-the-art quartet of mental tasks for all four participants.[21] In line with this result, our study also concluded that the personalized task group achieved significantly higher accuracies using their personalized tasks than the state-of-the-art prescribed tasks (mental math and rest), based on the session 3, offline, within-subject classification results.

4.3 *Variability in Hemodynamic Changes*

Overall, all tasks elicited increases in hemodynamic activity in some participants and decreases in others (Fig. 6). The anterior PFC is known to be involved in various executive functions, including: working memory, decision making, predicting future events, multitasking, maintaining attention, and emotional control.[19,53,60] Additionally, the medial anterior PFC is part of the default mode network (DMN), which is associated with deactivations below resting baseline levels during various goal-directed cognitive tasks and is also activated during autobiographical memory and envisioning the future.[19,20]

The task that resulted in the most consistent increase in hemodynamic activity across participants appears to be happy thoughts. Happy thoughts elicited an increase in activity in all but one participant (P1). This could be due to the fact that happy thoughts involve emotional control, which is believed to be a function of the PFC, and it also involves autobiographical memory, which is known to activate the DMN.[19,20]

Interestingly, the task that appears to result in the most consistent decrease in hemodynamic activity across participants was word generation; it was also the most commonly chosen decrease task. Word generation resulted in a strong decrease in hemodynamic activity in all participants other than P4 and P10. Word generation has often been associated with activations in the left PFC;[58,86] however, other trends have also been observed.[17] The observed decrease in activation with respect to the baseline may be a consequence of measuring mainly over the medial PFC, since the main language areas are predominantly situated on the left side of the brain.[87] Furthermore, the decrease in hemodynamic activity may be attributable to a deactivation in the DMN or resource sharing with the adjacent verbal areas.[88]

The prescribed tasks of mental math and rest were associated with both activations and deactivations among participants. Rest was usually accompanied by decreased hemodynamic activity (all participants other than P2, P4, and P10), while math usually resulted in increased hemodynamic activity (all participants other than P5, P7, and P8). Similar trends have been observed in literature.[14,16,27] The increase in hemodynamic activity when performing mental math could be attributed to the engagement of working memory,[73] while the decrease in hemodynamic activity associated with the rest task could be related to mental relaxation. On the other hand, the math task-induced decrease and rest task-associated increase may be related to role of the medial PFC in the DMN.[19,20]

Intersubject differences in cortical hemodynamic responses may, in part, be related to interindividual differences in cognitive processing and brain anatomy. Researchers have shown that there is a large intersubject variation in the size, shape, and position of various regions of the brain.[89,90] Thus, it may not be surprising that functional activation of the PFC (the region of focus in our study) varied among participants. EEG BCI researchers have drawn similar conclusions about the diversity of thought patterns between individuals.[21,35]

The large intersubject variability that appears to be present in most tasks confirms the need for personalized mental strategies. Our results corroborate research showing that the most effective task for controlling a BCI will vary among users.[21,22,24,26,34,91]

4.4 *Suitability of Personalized Task Selection Method*

Personalized tasks were chosen on the basis of both performance and ease-of-use. Incidentally, research in human–computer interactions has identified these considerations to be the two most important factors for BCI acceptance.[37]

The WS score was proposed as a measure to aid users in choosing their own personalized mental tasks. By providing a method to evaluate each task's effectiveness, irrespective of task pairings, the WS score simplified the selection of personalized mental tasks. Moreover, when using the WS score to select personalized tasks, the user is only concerned with one value per task (the task's effectiveness); by contrast, when using classification accuracies, the user is overwhelmed with all 55 pairwise classification accuracies. The positive correlation between the WS score and accuracy (Fig. 7) supports the use of the WS score as a measure of task effectiveness.

4.5 *Helpfulness of Feedback*

On average, the personalized and prescribed task groups found the continuous activation feedback somewhat helpful. This is in line with the findings of a previous NIRS-BCI study by Schudlo et al. where a similar form of feedback was deployed and found to be moderately helpful (3.13 ± 1.25 on a five-point Likert scale) by users.[29]

The personalized mental task group found the feedback to be significantly more useful than did the prescribed mental task group. This could be due to the fact that users in the former group chose their tasks based, in part, on the feedback. Continuous rather than intermittent (e.g., score feedback) feedback may better support long-term use of the BCI. Specifically, continuous feedback may promote adaptation of mental strategies and could potentially increase the accuracy and usability of the BCI over time.[92,93]

4.6 *Significance*

To date, personalized mental tasks have been explored in fMRI[34] and EEG[21,35,36] BCIs. To the best of our knowledge, these studies have only explored researcher-selected tasks based solely on performance, with the aim of improving BCI accuracy. By contrast, the present study investigated user-selected personalized tasks with the aim of improving ease-of-use. To the best of our knowledge, to date, the exploration of personalized mental tasks in NIRS-BCIs is limited to one offline, single-group study

that illustrated the potential of user-selected tasks in increasing ease-of-use.[42] The present study extends that earlier work by evaluating user-selected personalized mental tasks online in a two-group design.

4.7 *Limitations and Future Work*

The study was conducted exclusively with able-bodied participants. The findings reported herein likely do not reflect the performance of individuals with severe motor impairments. Further, it would be challenging to perform the personalized mental task protocol with individuals who have reached the total locked-in stage.[92] Nonetheless, with minor adjustments, we anticipate that the proposed protocol could be applied to clients with incomplete locked-in syndrome who retain reliable visual gaze and a yes/no response. For example, the ease-of-use ratings would need to be administered via a binary selection, scanning paradigm. There are several potential reasons why an individual with locked-in syndrome could stand to benefit from a BCI. First, even if eye gaze has been maintained, muscle fatigue could limit effective communication. Second, conditions such as amyotrophic lateral sclerosis are progressive; therefore, when clients transition from a locked-in to a total locked-in state, eye gaze may no longer be a viable access pathway. Literature has suggested that gaining control of the BCI prior to reaching total locked-in syndrome may increase the rate of success.[4,94–96] Further research and testing on the target population is necessary before conclusions about the effectiveness of personalized mental tasks in a communication BCI can be drawn.

Second, this study was conducted under ideal environmental conditions (quiet and dimly-lit room) that may not be indicative of most real-world settings. Further research should be conducted to assess the effect of environmental conditions on the system's performance.

Finally, when using NIRS as an access modality for a BCI, there is the potential for systemic contributions to the signal.[27,29,97] Since near-infrared light travels through the scalp and skull before reaching the brain, the recorded signal may contain systemic artefacts. Some researchers have proposed using simultaneous shallow measurements to remove the systemic portion of the deep NIRS signal.[82,98] However, the effect of such filtering on classification accuracies has yet to be fully quantified.[82] Other studies by Hoshi et al. and Villinger et al. reported minimal task-related changes in the systemic blood flow.[15,99] Furthermore, for the purpose of BCI design, it can be argued that as long as the system is able to differentiate between mental states, the exact origin and composition of the signal may be a moot point.

5 Conclusion

This study explored the possibility of allowing participants to choose their own personalized mental tasks, based on both performance and usability, to control an online NIRS-BCI. Our findings suggest that individuals can acquire control of an online personalized NIRS-BCI with classification accuracies comparable to those of an NIRS-BCI with prescribed, state-of-the-art tasks. The personalized mental task NIRS-BCI was significantly easier to use than its prescribed mental task counterpart. Users appeared to be able to effectively choose personalized mental tasks using the WS score as the measure of performance and post-task ease-of-use ratings as the measure of usability. Overall, the personalized mental task NIRS-BCI provided a more user-centered and easier-to-use online BCI without compromising accuracy. Personalized mental tasks may support the development of more user-friendly BCIs.

Acknowledgments

We would like to thank all of the members of the PRISM Lab, specifically Ka Lun Tam for his technical help with this project, and Amanda Fleury for her help with the headband. We would also like to thank Holland Bloorview Research Institute, NSERC, NSERC Create: Care, and University of Toronto Institute for Biomaterials and Biomedical Engineering for their support.

References

1. E. V. C. Friedrich, R. Scherer, and C. Neuper, "The effect of distinct mental strategies on classification performance for brain-computer interfaces," *Int. J. Psychophysiol.* **84**(1), 86–94 (2012).
2. K. Ang, J. Yu, and C. Guan, "Extracting and selecting discriminative features from high density NIRS-based BCI for numerical cognition," in *2012 Int. Joint Conf. on Neural Networks*, Vol. 10, pp. 1–6 (2012).
3. S. M. Coyle, T. E. Ward, and C. M. Markham, "Brain-computer interface using a simplified functional near-infrared spectroscopy system," *J. Neural Eng.* **4**(3), 219–226 (2007).
4. N. Birbaumer, "Breaking the silence: brain–computer interfaces (BCI) for communication and motor control," *Psychophysiology* **43**(6), 517–532 (2006).
5. J. Wolpaw et al., "Brain-computer interface technology: a review of the first international meeting," *IEEE Trans. Rehabil. Eng.* **8**(2), 164–173 (2000).
6. H. Ayaz et al., "Detecting cognitive activity related hemodynamic signal for brain computer interface using functional near infrared spectroscopy," in *3rd Int. IEEE/EMBS Conf. on Neural Engineering*, pp. 342–345, IEEE (2007).
7. S. Blain, A. Mihailidis, and T. Chau, "Assessing the potential of electrodermal activity as an alternative access pathway," *Med. Eng. Phys.* **30**(4), 498–505 (2008).
8. R. Sitaram et al., "Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain-computer interface," *Neuroimage* **34**(4), 1416–1427 (2007).
9. M. Naito et al., "A communication means for totally locked-in ALS patients based on changes in cerebral blood volume measured with near-infrared light," *IEICE - Trans. Inf. Syst.* **E90-D**(7), 1028–1037 (2007).
10. S. Coyle et al., "On the suitability of near-infrared (NIR) systems for next-generation brain–computer interfaces," *Physiol. Meas.* **25**(4), 815–822 (2004).
11. M. Wolf et al., "Different time evolution of oxyhemoglobin and deoxyhemoglobin concentration changes in the visual and motor cortices during functional stimulation: a near-infrared spectroscopy study," *Neuroimage* **16**(3), 704–712 (2002).
12. A. Villringer and B. Chance, "Non-invasive optical spectroscopy and imaging of human brain function," *Trends Neurosci.* **20**(10), 435–442 (1997).
13. M. Izzetoglu et al., "Functional brain imaging using near-infrared technology," *IEEE Eng. Med. Biol. Mag.* **26**(4), 38–46 (2007).
14. G. Bauernfeind et al., "Development, set-up and first results for a one-channel near-infrared spectroscopy system," *Biomed. Tech.* **53**(1), 36–43 (2008).
15. A. Villringer et al., "Near infrared spectroscopy (NIRS): a new tool to study hemodynamic changes during activation of brain function in human adults," *Neurosci. Lett.* **154**(1–2), 101–104 (1993).
16. G. Pfurtscheller et al., "Focal frontal (de)oxyhemoglobin responses during simple arithmetic," *Int. J. Psychophysiol.* **76**(3), 186–192 (2010).
17. V. Quaresima et al., "Bilateral prefrontal cortex oxygenation responses to a verbal fluency task: a multichannel time-resolved near-infrared topography study," *J. Biomed. Opt.* **10**(1), 011012 (2005).
18. Y. Hoshi et al., "Non-synchronous behavior of neuronal activity, oxidative metabolism and blood supply during mental tasks in man," *Neurosci. Lett.* **172**, 129–133 (1994).

19. H. Koshino et al., "Anterior medial prefrontal cortex exhibits activation during task preparation but deactivation during task execution," *PLoS One* **6**(8), e22909 (2011).
20. R. L. Buckner, J. R. Andrews-Hanna, and D. L. Schacter, "The brain's default network: anatomy, function, and relevance to disease," *Ann. N. Y. Acad. Sci.* **1124**, 1–38 (2008).
21. M.-C. Dobrea and D. M. Dobrea, "The selection of proper discriminative cognitive tasks–a necessary prerequisite in high-quality BCI applications," in *2nd Int. Symp. on Applied Sciences in Biomedical and Communication Technologies*, pp. 1–6 (2009).
22. C. Herff et al., "Classification of mental tasks in the prefrontal cortex using fNIRS," in *Conf. Proc. Annual Int. Conf. IEEE Engineering Medical Biology Society*, pp. 2160–2163 (2013).
23. S. D. Power, T. H. Falk, and T. Chau, "Classification of prefrontal activity due to mental arithmetic and music imagery using hidden Markov models and frequency domain near-infrared spectroscopy," *J. Neural Eng.* **7**(2), 026002 (2010).
24. S. Power, A. Kushki, and T. Chau, "Towards a system-paced near-infrared spectroscopy brain-computer interface: differentiating prefrontal activity due to mental arithmetic and mental singing from the no-control state," *J. Neural Eng.* **8**(6), 066004 (2011).
25. L. C. Schudlo, S. D. Power, and T. Chau, "Dynamic topographical pattern classification of multichannel prefrontal NIRS signals," *J. Neural Eng.* **10**(4), 046018 (2013).
26. H. Ogata, T. Mukai, and T. Yagi, "A study on the frontal cortex in cognitive tasks using near-infrared spectroscopy," in *Proc. 29th Annual Int. Conf. of the IEEE EMBS*, pp. 4731–4734 (2007).
27. S. D. Power, A. Kushki, and T. Chau, "Intersession consistency of single-trial classification of the prefrontal response to mental arithmetic and the no-control state by NIRS," *PLoS One* **7**(7), e37791 (2012).
28. K. Utsugi et al., "Development of an optical brain-machine interface," in *Proc. 29th Annual Int. Conf. of the IEEE EMBS*, pp. 5338–5341 (2007).
29. L. C. Schudlo and T. Chau, "Dynamic topographical pattern classification of multichannel prefrontal NIRS signals: II. Online differentiation of mental arithmetic and rest," *J. Neural Eng.* **11**(1), 016003 (2014).
30. N. Naseer and K.-S. Hong, "Functional near-infrared spectroscopy based discrimination of mental counting and no-control state for development of a brain-computer interface," in *35th Annual Int. Conf. Proc. IEEE Engineering in Medicine and Biology Society*, pp. 1780–1783 (2013).
31. K. Izzetoglu et al., "The evolution of field deployable fNIR spectroscopy from bench to clinical settings," *J. Innov. Opt. Health Sci.* **04**(03), 239–250 (2011).
32. S. Kanoh et al., "A NIRS-based brain-computer interface system during motor imagery: system development and online feedback training," in *Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, pp. 594–597 (2009).
33. N. Naseer and K.-S. Hong, "Classification of functional near-infrared spectroscopy signals corresponding to the right- and left-wrist motor imagery for development of a brain-computer interface," *Neurosci. Lett.* **553**, 84–89 (2013).
34. B. Sorger et al., "Another kind of 'BOLD Response': answering multiple-choice questions via online decoded single-trial brain signals," *Prog. Brain Res.* **177**, 275–292 (2009).
35. R. Palaniappan, "Utilizing gamma band to improve mental task based brain-computer interface design," *IEEE Trans. Neural Syst. Rehabil. Eng.* **14**(3), 299–303 (2006).
36. R. Chai et al., "Mental non-motor imagery tasks classifications of brain computer interface for wheelchair commands using genetic algorithm-based neural network," in *Int. Joint Conf. on Neural Networks*, pp. 1–7 (2012).
37. D. Bos, M. Poel, and A. Nijholt, "A study in user-centered design and evaluation of mental tasks for BCI," *Lect. Notes Comput. Sci.* **6524**, 122–134 (2011).
38. D. Tan and A. Nijholt, *Brain-Computer Interfaces: Applying our Minds to Human-Computer Interaction*, pp. 149–178, Springer, London Dordrecht, Heidelberg (2010).
39. E. M. Holz et al., "Brain-computer interface controlled gaming: evaluation of usability by severely motor restricted end-users," *Artif. Intell. Med.* **59**(2), 111–120 (2013).
40. E. Curran et al., "Cognitive tasks for driving a brain-computer interfacing system: a pilot study," *IEEE Trans. Neural Syst. Rehabil. Eng.* **12**(1), 48–54 (2004).
41. E. V. C. Friedrich, R. Scherer, and C. Neuper, "User-appropriate and robust control strategies to enhance brain–computer interface performance and usability," in *Brain-Computer Interface Research SpringerBriefs in Electrical and Computer Engineering*, C. Guger, B. Z. Allison, and G. Edlinger, Eds., pp. 15–22, Springer Berlin Heidelberg, Berlin, Heidelberg (2013).
42. S. Weyand, K. Takehara-Nishiuchi, and T. Chau, "Exploring methodological frameworks for a mental task-based near-infrared spectroscopy brain-computer interface," Manuscript submitted for publication, (2015).
43. R. C. Oldfield, "The assessment and analysis of handedness: the Edinburgh inventory," *Neuropsychologia* **9**(1), 97–113 (1971).
44. K. Marumo et al., "Gender difference in right lateral prefrontal hemodynamic response while viewing fearful faces: a multi-channel near-infrared spectroscopy study," *Neurosci. Res.* **63**(2), 89–94 (2009).
45. H. Yang et al., "Gender difference in hemodynamic responses of prefrontal area to emotional stress by near-infrared spectroscopy," *Behav. Brain Res.* **178**(1), 172–176 (2007).
46. F. Okada et al., "Gender- and handedness-related differences of forebrain oxygenation and hemodynamics," *Brain Res.* **601**(1–2), 337–342 (1993).
47. M. Tamura, Y. Hoshi, and F. Okada, "Localized near-infrared spectroscopy and functional optical imaging of brain activity," *Philos. Trans. R. Soc. B Biol. Sci.* **352**(1354), 737–742 (1997).
48. M. L. Schroeter et al., "Age dependency of the hemodynamic response as measured by functional near-infrared spectroscopy," *Neuroimage* **19**(3), 555–564 (2003).
49. I. Kwee and T. Nakada, "Dorsolateral prefrontal lobe activation declines significantly with age–functional NIRS study," *J. Neurol.* **250**(5), 525–529 (2003).
50. A. Duncan et al., "Measurement of cranial optical path length as a function of age using phase resolved near infrared spectroscopy," *Pediatr. Res.* **39**(5), 889–894 (1996).
51. S. Weyand, K. Takehara-nishiuchi, and T. Chau, "Weaning off mental tasks to achieve voluntary self-regulatory control of a near-infrared spectroscopy brain-computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.* In Press (2015).
52. ISS Inc., "ISS, imagent: functional brain imaging system using infrared photons," ISS, 2012, http://iss.com/biomedical/instruments/imagent.html (September 2012).
53. W. Gao et al., "The unique properties of the prefrontal cortex and mental illness," in *When Things Go Wrong - Diseases and Disorders of the Human Brain*, pp. 3–26, InTech, Croatia (1990).
54. F. B. Haeussinger et al., "Simulation of near-infrared light absorption considering individual head and prefrontal cortex anatomy: implications for optical neuroimaging," *PLoS One* **6**(10), e26377 (2011).
55. E. Okada et al., "Theoretical and experimental investigation of near-infrared light propagation in a model of the adult head," *Appl. Opt.* **36**(1), 21–31 (1997).
56. N. Naseer, M. J. Hong, and K.-S. Hong, "Online binary decision decoding using functional near-infrared spectroscopy for the development of brain-computer interface," *Exp. Brain Res.* **232**(2), 555–564 (2014).
57. M. J. Khan, M. J. Hong, and K.-S. Hong, "Decoding of four movement directions using hybrid NIRS-EEG brain-computer interface," *Front. Hum. Neurosci.* **8**, 244 (2014).
58. A. Faress and T. Chau, "Towards a multimodal brain-computer interface: combining fNIRS and fTCD measurements to enable higher classification accuracy," *Neuroimage* **77**, 186–194 (2013).
59. H. Ayaz et al., "Tangram solved? Prefrontal cortex activation analysis during geometric problem solving," in *34th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, pp. 4724–4727 (2012).
60. K. Tai and T. Chau, "Single-trial classification of NIRS signals during emotional induction tasks: towards a corporeal machine interface," *J. Neuroeng. Rehabil.* **6**(39), 1–14 (2009).
61. M. L. Schroeter et al., "Near-infrared spectroscopy can detect brain activity during a color-word matching Stroop task in an event-related design," *Hum. Brain Mapp.* **17**(1), 61–71 (2002).
62. A.-C. Ehlis et al., "Multi-channel near-infrared spectroscopy detects specific inferior-frontal activation during incongruent Stroop trials," *Biol. Psychol.* **69**(3), 315–331 (2005).
63. T. F. D. Kanthack, M. Bigliassi, and L. Altimari, "Equal prefrontal cortex activation between males and females in a motor tasks and different

visual imagery perspectives: a functional near- infrared spectroscopy (fNIRS) study," *Mot. Rio Claro* **19**(3), 627–632 (2013).
64. D. R. Leff et al., "Assessment of the cerebral cortex during motor task behaviours in adults: a systematic review of functional near infrared spectroscopy (fNIRS) studies," *Neuroimage* **54**(4), 2922–2936 (2011).
65. D. P. Tedesco and T. S. Tullis, "A comparison of methods for eliciting post-task subjective ratings in usability testing," in *Usability Professionals Association Annual Conf.*, pp. 1–9 (2006).
66. J. Sauro and J. S. Dumas, "Comparison of three one-question, post-task usability questionnaires," in *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, pp. 1599–1608 (2009).
67. H. Ayaz et al., "Assessment of cognitive neural correlates for a functional near infrared-based brain computer interface system," *Lect. Notes Comput. Sci.* **5638**, 699–708 (2009).
68. N. Birbaumer and L. G. Cohen, "Brain-computer interfaces: communication and restoration of movement in paralysis," *J. Physiol.* **579**(3), 621–636 (2007).
69. D. T. Delpy et al., "Estimation of optical pathlength through tissue from direct time of flight measurement," *Phys. Med. Biol.* **33**(12), 1433–1442 (1988).
70. M. Cope, *The Application of Near Infrared Spectroscopy to Non Invasive Monitoring of Cerebral Oxygenation in the Newborn Infant*, pp. 2–305, University of London, London (1991).
71. A. Duncan et al., "Optical pathlength measurements on adult head, calf and forearm and the head of the newborn infant using phase resolved optical spectroscopy," *Phys. Med. Biol.* **40**(2), 295–304 (1995).
72. S. D. Power, A. Kushki, and T. Chau, "Automatic single-trial discrimination of mental arithmetic, mental singing and the no-control state from prefrontal activity: toward a three-state NIRS-BCI," *BMC Res. Notes* **5**(141), 1–10 (2012).
73. S. D. Power and T. Chau, "Automatic single-trial classification of prefrontal hemodynamic activity in an individual with Duchenne muscular dystrophy," *Dev. Neurorehabil.* **16**(1), 67–72 (2013).
74. H. Zhu et al., "General form for obtaining discrete orthogonal moments," *IET Image Process.* **4**, 335 (2010).
75. P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit. Lett.* **15**(11), 1119–1125 (1994).
76. C. Bishop, *Pattern Recognition and Machine Learning*, pp. 1–743, Springer, New York (2006).
77. R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits Syst. Mag.* **6**, 21–45 (2006).
78. P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-validation," in *Encyclopedia of Database Systems*, **71**, pp. 532–538, Springer, New York (2009).
79. A. Kübler et al., "Brain-computer communication: Self-regulation of slow cortical potentials for verbal communication," *Arch. Phys. Med. Rehabil.* **82**(11), 1533–1539 (2001).
80. R. Taylor, "Interpretation of the correlation coefficient: a basic review," *J. Diagn. Med. Sonogr.* **6**(1), 35–39 (1990).
81. J. F. Hemphill, "Interpreting the magnitudes of correlation coefficients," *Am. Psychol.* **58**(1), 78–80 (2003).
82. J. Chan, S. Power, and T. Chau, "Investigating the need for modelling temporal dependencies in a brain-computer interface with real-time feedback based on near infrared spectra," *J. Near Infrared Spectrosc.* **20**(1), 107–116 (2012).
83. M. Stangl et al., "A haemodynamic brain-computer interface based on real-time classification of near infrared spectroscopy signals during motor imagery and mental arithmetic," *J. Near Infrared Spectrosc.* **21**(3), 157–171 (2013).
84. D. Cordova and M. Lepper, "Intrinsic motivation and the process of learning: beneficial effects of contextulization, personalization, and choice," *J. Educ. Pyschol.* **88**(4), 715–730 (1996).
85. R. J. C. M. Salden, F. Paas, and J. J. G. van Merriënboer, "Personalised adaptive task selection in air traffic control: effects on training efficiency and transfer," *Learn. Instr.* **16**(4), 350–362 (2006).
86. M. J. Herrmann, A.-C. Ehlis, and A. J. Fallgatter, "Frontal activation during a verbal-fluency task as measured by near-infrared spectroscopy," *Brain Res. Bull.* **61**(1), 51–56 (2003).
87. R. Schlösser et al., "Functional magnetic resonance imaging of human brain activity in a verbal fluency task," *J. Neurol. Neurosurg. Psychiatry* **64**(4), 492–498 (1998).
88. I. Shapira-Lichter et al., "Portraying the unique contribution of the default mode network to internally driven mnemonic processes," *Proc. Natl. Acad. Sci. U. S. A.* **110**(13), 4950–4955 (2013).
89. J. Nie, L. Guo, and T. Liu, "A computational model of cerebral cortex folding," *Med. Image Comput. Comput. Interv.* **12**(2), 458–465 (2009).
90. J. Xiong et al., "Intersubject variability in cortical activations during a complex language task," *Neuroimage* **12**(3), 326–339 (2000).
91. H. Nai-Jen and R. Palaniappan, "Classification of mental tasks using fixed and adaptive autoregressive models of EEG signals," in *Proc. 26th Annual Int. Conf. of the IEEE EMBS*, Vol. 1, pp. 507–510 (2004).
92. N. Birbaumer et al., "Neurofeedback and brain-computer interface clinical applications," *Int. Rev. Neurobiol.* **86**, 107–117 (2009).
93. E. V. C. Friedrich, C. Neuper, and R. Scherer, "Whatever works: a systematic user-centered training protocol to optimize brain-computer interfacing individually," *PLoS One* **8**(9), e76214 (2013).
94. N. Weiskopf et al., "Real-time functional magnetic resonance imaging: methods and applications," *Magn. Reson. Imaging* **25**(6), 989–1003 (2007).
95. L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors* **12**(2), 1211–1279 (2012).
96. A. Kübler and N. Birbaumer, "Brain-computer Interfaces and communication in paralysis: extinction of goal directed thinking in completely paralysed patients?," *Heal. Psychol.* **119**(11), 2658–2666 (2008).
97. I. Tachtsidis et al., "Measurement of frontal lobe functional activation and related systemic effects: a near-infrared spectroscopy investigation," *Adv. Exp. Med. Biol.* **614**, 397–403 (2008).
98. T. Funane et al., "Quantitative evaluation of deep and shallow tissue layers' contribution to fNIRS signal using multi-distance optodes and independent component analysis," *Neuroimage* **85**, 150–165 (2014).
99. Y. Hoshi et al., "Recognition of human emotions from cerebral blood flow changes in the frontal region: a study with event-related near-infrared spectroscopy," *J. Neuroimaging* **21**(2), 94–101 (2011).

Sabine Weyand is a PhD candidate in the Institute of Biomaterials and Biomedical Engineering at the University of Toronto, Canada. She received her BASc from Queen's University, Canada, and her MASc from the University of Ottawa, Canada. Her current research is in the area of active brain–computer interface development using near-infrared spectroscopy.

Larissa Schudlo is a PhD candidate in the Institute of Biomaterials and Biomedical Engineering at the University of Toronto, Canada. She received her BEng from McMaster University, Canada, and her MASc from the University of Toronto. Her current research is in the area of active brain–computer interface development using near-infrared spectroscopy.

Kaori Takehara-Nishiuchi is an assistant professor of psychology at the University of Toronto. She received her PhD in the Department of Pharmaceutical Sciences at the University of Tokyo and did a postdoctoral training at the University of Arizona. By applying *in vivo* single-neuron recording techniques to rodent models, her current research is directed at neurophysiological mechanisms of long-term memory and its disruption in Alzheimer's disease.

Tom Chau is vice president, researcher, and director at Bloorview Research Institute at Holland Bloorview Kids Rehabilitation Hospital. He is also a full professor in the Institute of Biomaterials & Biomedical Engineering at the University of Toronto. He received a PhD from the University of Waterloo in systems design engineering. His research interests revolve around access innovations for children and youth with severe disabilities, kinetics of handwriting and drawing, and automatic classification of swallowing safety.