Brain-Computer Interfaces

ISSN: 2326-263X (Print) 2326-2621 (Online) Journal homepage: https://www.tandfonline.com/loi/tbci20

# Development of a ternary hybrid fNIRS-EEG brain–computer interface based on imagined speech

Alborz Rezazadeh Sereshkeh, Rozhin Yousefi, Andrew T Wong, Frank Rudzicz & Tom Chau

**To cite this article:** Alborz Rezazadeh Sereshkeh, Rozhin Yousefi, Andrew T Wong, Frank Rudzicz & Tom Chau (2019) Development of a ternary hybrid fNIRS-EEG brain–computer interface based on imagined speech, Brain-Computer Interfaces, 6:4, 128-140, DOI: 10.1080/2326263X.2019.1698928

**To link to this article:** https://doi.org/10.1080/2326263X.2019.1698928

BRAIN-COMPUTER INTERFACES
2019, VOL. 6, NO. 4, 128–140
https://doi.org/10.1080/2326263X.2019.1698928

Taylor & Francis
Taylor & Francis Group

ARTICLE



# Development of a ternary hybrid fNIRS-EEG brain–computer interface based on imagined speech

Alborz Rezazadeh Sereshkeh [a,b], Rozhin Yousefi [a,b], Andrew T Wong [a,b], Frank Rudzicz [c,d] and Tom Chau [a,b]

[a]Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Canada; [b]Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital, Toronto, Canada; [c]Department of Computer Science, University of Toronto, Toronto, Canada; [d]Vector Institute, University of Toronto, Toronto, Canada

**ABSTRACT**

There is increasing interest in developing intuitive brain-computer interfaces (BCIs) to differentiate intuitive mental tasks such as imagined speech. Both electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS) have been used for this purpose. However, the classification accuracy and number of commands in such BCIs have been limited. The use of multi-modal BCIs to address these issues has been proposed for some common BCI tasks, but not for imagined speech. Here, we propose a multi-class hybrid fNIRS-EEG BCI based on imagined speech. Eleven participants performed multiple iterations of three tasks: mentally repeating 'yes' or 'no' for 15 s or an equivalent duration of unconstrained rest. We achieved an average ternary classification accuracy of 70.45 ± 19.19% which is significantly better than that attained with each modality alone ($p < 0.05$). Our findings suggest that concurrent measurements of EEG and fNIRS can improve classification accuracy of BCIs based on imagined speech.

**ARTICLE HISTORY**
Received 21 March 2019
Accepted 24 November 2019

**KEYWORDS**
Brain-computer interface; imagined speech; hybrid BCI; fNIRS; EEG; regularized linear discriminant analysis (RLDA); discrete wavelet transform (DWT)

## 1. Introduction

The primary goal of most brain–computer interface (BCI) research is to provide a communication pathway for individuals with severe motor impairments who have very limited or no voluntary movement [1]. A BCI can be a suitable means of communication for these individuals, as they are only required to perform a mental task or attend to a stimulus, without the need for muscle activity. However, a majority of common BCI activation tasks, such as mental arithmetic or word generation, have little or no correlation with typical communication [2]. For example, a user may be required to perform mental arithmetic to answer basic yes or no questions, to move his/her wheelchair around, or to control a switch [3,4]. This non-intuitiveness makes the assistive device difficult to use, limiting its potential to meaningfully improve quality of life [5].

Another common BCI activation task is motor imagery, which involves the imagined movement of a specific part of the body. This mental task can be considered intuitive for certain applications, such as navigation or robotic control. However, it can be difficult or impossible for individuals with congenital or long-term motor impairments [6,7].

A mental task which has gained attention recently as an intuitive BCI task is 'imagined speech'. In this task, the BCI user is instructed to covertly say or repeat a phrase without moving the articulators [8]. Although 'imagined speech' overcomes the aforementioned shortcomings of other BCI mental tasks, it can be difficult to detect and classify using only noninvasive brain recording modalities, such as electroencephalography (EEG) or functional near-infrared spectroscopy (fNIRS) [8]. An extensive review of previous BCIs based on 'imagined speech' is provided by Schultz et al. [9].

EEG classification of imagined speech has been investigated in various levels of language such as vowels [5], syllables [10], and complete words [2]; as well as in different languages, such as Chinese [11] and Spanish [12]. In most of those studies, the classification accuracy between different imagined speech tasks exceeded the chance level. However, the reported average accuracies were considerably below 70% (the suggested minimum threshold for practical BCI use [13], even in binary classification problems [2]). Brain regions that provided more discriminatory electrical information in imagined speech tasks were also explored in those studies. In imagined speech BCI based on 'yes' and 'no', task-specific changes in EEG beta and gamma power in

**CONTACT** Alborz Rezazadeh Sereshkeh a.rezazadeh@utoronto.ca; alborzrs@gmail.com Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Canada

language-related brain areas (such as the Broca's area and the superior temporal gyrus) tended to provide discriminatory information [2,5].

Due to the slow nature of the hemodynamic response, early applications of fNIRS in imagined speech focused on distinguishing different patterns of hemodynamic responses during trials of full sentences (e.g., inner recitation of hexameter versus prose [14]) as opposed to decoding small units of language such as nouns [9]. Later, several studies investigated fNIRS-BCIs for answering yes or no questions by simply thinking 'yes' or 'no' and reported binary classification accuracies in the range of 70% to 76% [15–17]. Recently, Sereshkeh et al. introduced a ternary fNIRS-BCI for the classification of covert rehearsal of the words 'yes' or 'no' and an equivalent duration of unconstrained rest and reported a ternary classification accuracy of 64% across 12 participants [18]. They also showed that for most participants, fNIRS channels in the left temporal and temporoparietal cortex provided the most discriminative information.

As mentioned, most imagined speech BCI studies based on noninvasive measurements reported low average accuracies, especially for more than two classes. To realize a reliable 'imagined speech' BCI, classification accuracy must be improved further. One solution for improving the performance of a BCI without changing the activation task is to combine two or more brain recording modalities [19–22]. Specifically, previous work has utilized EEG in conjunction with fNIRS for BCI use. These modalities are mutually complementary: EEG has a high temporal resolution but low spatial resolution, while fNIRS has a low temporal resolution but superior spatial resolution [21] and [22] provided extensive reviews of multimodal BCIs.

In this paper, we present a hybrid fNIRS-EEG BCI for ternary classification of imagined speech (mentally rehearsing the phrases 'yes' and 'no' to answer yes versus no questions and an idle state). To the best of our knowledge, this is the first report of a combination of fNIRS and EEG to classify imagined speech. Furthermore, a novel technique is proposed for the fusion of the two classifiers trained using the data from each modality alone.

## 2. Materials and methods

### 2.1. Participants

Eleven typically developed, right-handed participants (six males) between the ages of 23 and 33 (mean age: 28.3 ± 3.0 years) participated in this study. Participants were fluent in English, had a normal or corrected-to-normal vision, and had no health issues that could adversely affect the measurements or the ability to follow the experimental protocol. These issues included cardiovascular, psychiatric, respiratory, neurological, degenerative, metabolic or alcohol-related conditions. This study was approved by the research ethics boards of the University of Toronto and Holland Bloorview Kids Rehabilitation Hospital. Written consent was obtained from all participants prior to participating in the study.

### 2.2. Instrumentation

EEG measurements were taken from 32 locations spanning all cortical regions, with a higher density of electrodes in the temporal speech-related regions. EEG recording was done using dry EEG electrodes (an Acticap Xpress Twist) connected to a BrainAmp DC amplifier (Brain Products GmbH, Germany) with a sampling rate of 1 kHz. Reference and ground (GND) electrodes were placed on the left and right earlobes, respectively (A1 and A2). Fp1 and Fp2 electrodes were only used to detect and remove EOG artifacts. See Figure 1 for the location of the electrodes.

fNIRS data were collected using a near-infrared spectrometer (ETG-4000 Optical Topography System, Hitachi Medical Co., Japan) from the temporal, frontal, and parietal cortices. Each emitter consisted of two laser diodes that concurrently emitted light at wavelengths of 695 nm and 830 nm. The reflected light was captured using the detectors with a sampling frequency of 10 Hz. Two 3 × 5 rectangular grids were used to fit 16 emitters and 14 photodetectors, spaced 3 cm apart. Optical signals were acquired only from source-detector pairs separated by 3 cm (henceforth referred to as 'channels'). As a result, fNIRS signals were collected from a total of 44 locations distributed equally and symmetrically between the two hemispheres (see Figure 1).

The EEG electrode holders, near-infrared (NIR) emitters and NIR photodetectors were integrated into a custom-made cap (EasyCap, Germany).

### 2.3. Experimental protocol

Each participant attended two sessions (~1 h each) on two separate days. During each trial, participants were asked to perform one of the three mental tasks: unconstrained rest, 'yes' trials, and 'no' trials. In the 'yes' and 'no' trials, participants answered 'yes' and 'no' questions by thinking 'yes' or 'no' while mentally repeating the phrase 'yes' or 'no' in response to stimuli. The first session consisted of an offline block of 36 trials, followed by two online blocks of 24 trials each. The second session consisted of 4 online blocks of 24 trials each. Each

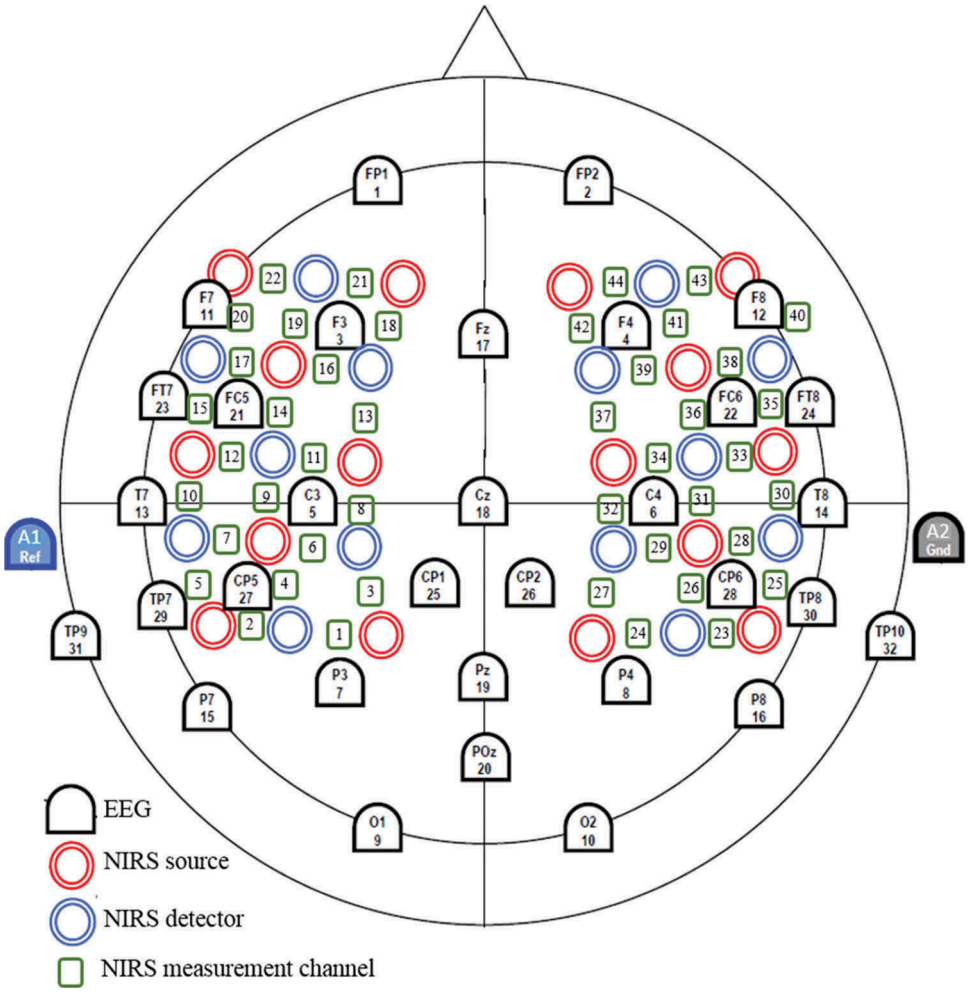


**Figure 1.** The placement of EEG electrodes and fNIRS sources and detectors. EEG positions are marked using the nomenclature of the international 10–20 system along with the corresponding channel number.

block contained an equal number of each task presented in pseudorandom order, with each trial lasting for 15 s.

During the third task, 'unconstrained rest', participants were only asked to refrain from performing the other two imagined speech tasks. The primary motivation behind choosing 'unconstrained rest' as the third class is to facilitate potential asynchronous implementation. Specifically, for users who can achieve a reliable 3-class accuracy in our suggested paradigm, imagined speech might eventually be utilized in a 2-class asynchronous BCI. Such an asynchronous BCI can be activated by mentally repeating the phrases 'yes' or 'no' and potentially be used as a binary switch (i.e., with two activation modes) for an assistive device [18].

In the online trials, real-time feedback was provided after the completion of the mental task. The feedback was calculated by a classifier trained on fNIRS data. The online fNIRS classification results were previously reported [18]. Note that EEG data were collected from only 11 of the 12 participants of that study and presented here.

In this paper, the performance of the hybrid system is evaluated and compared to the performance of each modality alone. The timing diagram of the experiment is presented in Figure 2.

### 2.4. Training and test set

The entirety of the first session plus the first block of the second session was used as the training dataset for offline classification, resulting in 108 trials (36 per class). This training set was used for the selection of the feature extraction and classification methods, as well as their required hyper-parameters. The method and/or parameter(s) which yielded the highest cross-validation (CV) accuracy (100 runs of 10-fold CV) on the training set was selected.

The remaining 72 trials (24 per class) i.e., last 3 blocks from session 2 were used as the test set. Prior to each test block, the classifier was retrained with the accumulated

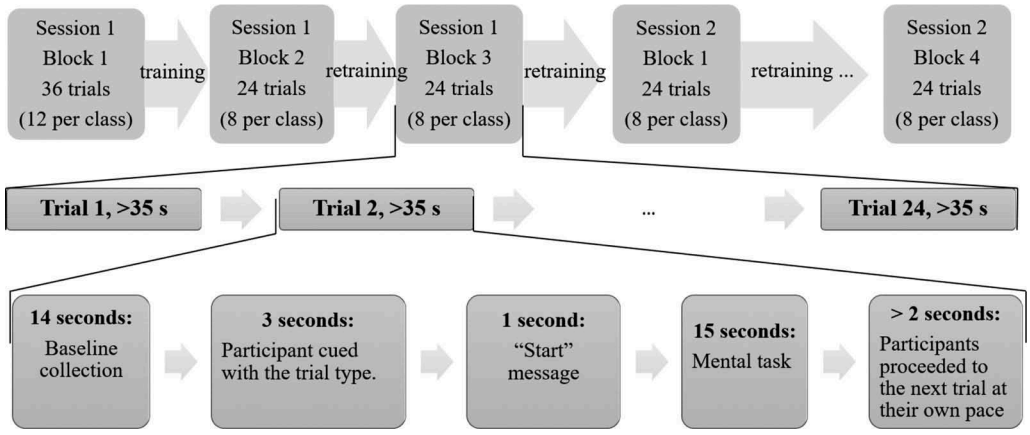


**Figure 2.** The timing diagram of the experiment.

data from all previous blocks, following a pseudo-online paradigm. For example, the second test block would utilize a classifier trained on 132 trials, comprising the training set (108 trials) plus data from the first test block (24 trials).

There were two reasons for choosing the last three blocks as the test dataset. First, we wanted to be consistent with the presentation of results in an earlier paper on the same study [18]. Secondly, since each participant attended two sessions on separate days, we designed the first session to be the training session, and the first block of the second session to be the same-day calibration block. As a result, the rest of the blocks (i.e., last three blocks) were used as the test dataset.

## 2.5. EEG data analysis

### 2.5.1. Signal preprocessing

EEG data were first filtered using a 0.5–40 Hz bandpass FIR filter. The low-pass cutoff frequency of 40 Hz was suggested by the hardware manufacturer (Brain Products GmbH, Germany) as the maximum reliable frequency of the dry electrodes.

In order to remove electrooculography (EOG) artifacts, the ADJUST algorithm [23] was deployed. Independent components due to eye blinks, and horizontal and vertical eye movements were removed. The remaining components were used to reconstruct the EEG data. The reconstructed artifact-free signals for each of the 30 electrodes were subjected to further analysis (Fp1 and Fp2 data were solely used for EOG artifact detection and were not considered thereafter). Note that for the ADJUST algorithm, the independent components were calculated using only the training set data.

### 2.5.2. Feature extraction

Discrete wavelet transform (DWT) coefficients were extracted from each trial [24]. Other common types of features for EEG classification, such as autoregressive components, common spatial patterns, and spectral power estimates were also tested on the training set, but DWT features using the Symlet-10 (sym10) wavelet yielded the highest 10-fold cross validation (CV) accuracy on the training set and hence was selected for the test set. DWT features have previously proven discriminatory for EEG signals accompanying imagined speech [25].

DWT has been frequently deployed EEG analysis given its ability to localize information in both frequency and time domains [26–28]. Six levels of decomposition yielded the lowest 10-fold CV classification error in the training set. The root-mean-square (RMS) of the outputs from each DTW decomposition level were used as features for classification. These six levels represent the following frequency ranges: 40–31.3 Hz, 31.3–15.6 Hz, 15.6–7.8 Hz, 7.8–3.9 Hz, 3.9–2.0 Hz, and 2.0–1.0 Hz. A total of 180 DWT features (30 channels × 6 frequency ranges) were generated from each trial.

## 2.6. fNIRS data analysis

The signal processing, baseline removal and feature extraction steps for the fNIRS data are the same as the steps described by [18]. In short, the optical intensities were converted to oxygenated hemoglobin concertation changes, [HbO], using the modified Beer–Lambert Law [29]. The [HbO] data were filtered using a Chebyshev type II low-pass filter with a passband cutoff frequency of 0.1 Hz and a stopband cutoff frequency of 0.5 Hz. A trial-specific mean baseline was removed using a 1.5 s period just prior to stimulus presentation. The mean value of [HbO] for each of the 44 channels, over the

entire length of a trial, constituted the input features. Other common types of fNIRS features, such as variance, slope, kurtosis and skewness of changes in deoxygenated and oxygenated hemoglobin concentrations were examined, but the mean of [HbO] yielded the lowest 10-fold CV classification error in the training set.

## 2.7. Classification

### 2.7.1. Regularized linear discriminant analysis (RLDA)

Linear discriminant analysis has been extensively used in BCI studies. While the curse of dimensionality and overfitting are very common problems in BCI classification [30], a large number of these studies did not regularize their LDA models [30–32]. Strother et al. [33] compared three linear discriminant models (LDA, support vector machines and logistic regression) and concluded that comprehensive optimization of the regularization parameter(s) may be much more important than the choice of the model. In this paper, we used the regularized linear discriminant analysis (RLDA) algorithm for classification [34]. The regularization parameter was optimized separately for each participant, each modality and each test block.

### 2.7.2. Choosing the regularization parameter

The test set consisted of three blocks. For each block, the classifier was re-trained with data from all previous trials. In other words, for the first test set block, the classifier was trained using only the training set data. For the second test set block, the classifier was trained using the training set data and the first test set block. And finally, for the third test set block, the classifier was trained using the training set data and the first two test set blocks.

The regularization parameter was also optimized for each of the aforementioned classifiers using 100 runs of 10-fold CV on the available training data (as explained in the previous paragraph). In other words, the average CV accuracy was calculated for $\gamma = 0.05, 0.1, 0.15, \ldots, 1$ and the $\gamma$ which provided the highest CV accuracy was used to train the classifier. A separate classifier for EEG and fNIRS was trained and the value of $\gamma$ was optimized for each of these two classifiers separately (see Figure 3). Specifically,

$$\gamma^*_{EEG} = \underset{\gamma}{argmax}\left(A_{EEG\gamma=0.05},\ A_{EEG\gamma=0.1},\ \ldots,A_{EEG\gamma=1}\right) \tag{1}$$

and,

$$\gamma^*_{fNIRS=} = \underset{\gamma}{argmax}\left(A_{fNIRS_{\gamma=0.05}},\ A_{fNIRS_{\gamma=0.1}},\ \ldots,A_{fNIRS_{\gamma=1}}\right) \tag{2}$$

In Equations (1) and (2), $A_{EEG}$ and $A_{fNIRS}$ are the average classification accuracies over 100 runs of 10-fold CV on all previous trials (the entire training set as well as previous test blocks, if any) using EEG and fNIRS measurements, respectively. In total, six classifiers were trained (three test blocks over two modalities). In this study, 10-fold CV was used over leave-one-out CV (used by [18]) since it provides better generalizability and less variance [35].

We note that the proposed method of optimizing the regularization parameter can be computationally expensive, which is not ideal for cases when the process needs to be done multiple times during an online data collection session. Future work may include exploring automatic optimization methods, such as the automatic estimation of the optimal shrinkage parameter proposed by [36].

### 2.7.3. Fusion of EEG and fNIRS classifiers

After optimizing the regularization parameters and training two classifiers, one using EEG data and one using fNIRS data, these two classifiers were combined using a probabilistic model to make predictions on the test set. The classifier predicted the class, $C$, of a single trial according to:

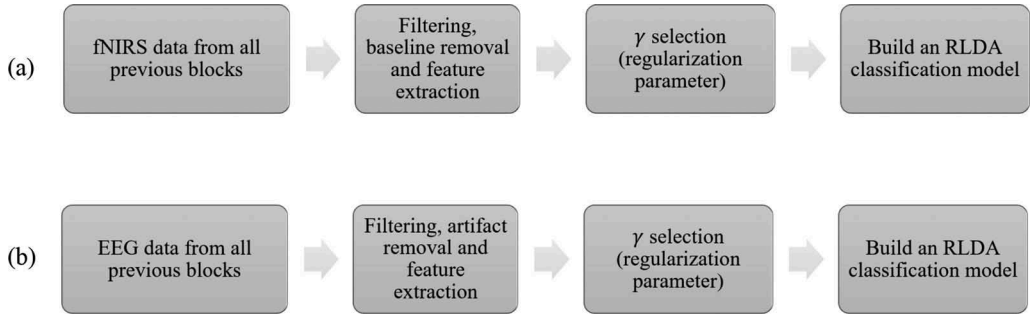


**Figure 3.** Flowchart illustrating the pathway for building (a) the fNIRS classifier and (b) the EEG classifier (RLDA = regularized linear discriminant analysis).

$$C(X_{EEG}, X_{fNIRS}) = \underset{c=''yes'',''no'',''rest''}{argmax} P(C = c|X_{EEG}, X_{fNIRS}) \quad (3)$$

where $X_{EEG}$ and $X_{fNIRS}$ are feature vectors of that particular trial in the test set. The probabilities that the trial belonged to each of three classes were obtained using the following three equations:

$$P(C = ''no''|X_{EEG}, X_{fNIRS}) = max(A^*_{EEG} \times P(C = ''no''|X_{EEG}) \quad A^*_{fNIRS} \times P(C = ''no''|X_{fNIRS})) \quad (4)$$

$$P(C = ''yes''|X_{EEG}, X_{fNIRS}) = max(A^*_{EEG} \times P(C = ''yes''|X_{EEG}) \quad A^*_{fNIRS} \times P(C = ''yes''|X_{fNIRS})) \quad (5)$$

$$P(C = ''rest''|X_{EEG},X_{fNIRS}) = max(A^*_{EEG} \times P(C = ''rest''|X_{EEG}), \quad A^*_{fNIRS} \times P(C = ''rest''|X_{fNIRS})) \quad (6)$$

where $A^*_{EEG}$ and $A^*_{fNIRS}$ are the average CV classification accuracies obtained with the optimized regularization parameters, $\gamma^*_{EEG}$ and $\gamma^*_{fNIRS}$ for each modality.

In other words, the class labels for a given test trial were taken as that predicted either by the EEG classifier or the fNIRS classifier, depending on the confidence of each classifier's prediction and the classifier's prior probability.

## 3. Results

### 3.1. Ternary classification accuracies

Table 1 provides the ternary classification accuracy across the three test blocks for all participants using EEG only, fNIRS only, and the proposed hybrid system. The statistical significance level of each accuracy is calculated based on the permutation test approach [37], with 10,000 permutations. Accuracies exceeding the significance levels of 0.05, 0.01 and 0.001 are marked with one, two and three asterisks, respectively.

By using EEG data only, an average classification accuracy of $63.8 \pm 20.9\%$ (over the entire test set) was reached across participants with six participants surpassing the chance level ($p < 0.001$, using the permutation test approach). With fNIRS data only, an average classification accuracy of $63.6 \pm 21.1\%$ was obtained across participants with seven participants exceeding the same limit.

After the fusion of fNIRS and EEG classifiers using the proposed method, the average classification accuracy across participants improved to $70.5 \pm 19.2\%$ (>6.5% improvement compared to EEG and fNIRS alone) with nine participants surpassing the chance level (two and three more participants compared to fNIRS and EEG, respectively). Figure 4 illustrates the average classification accuracy over the entire test set across participants using each modality on its own and in combination.

In order to statistically compare the hybrid accuracies with the single-modality ones, we did two comparisons: hybrid versus EEG and hybrid versus fNIRS. For each of these comparisons, we used a two factor, within-subject, repeated measures ANOVA. The two factors were the measurement device (EEG or hybrid for the first comparison, and EEG or fNIRS for the second comparison) and the test block number (one, two or three). The calculated p-values for the first factor (i.e., the measurement device) were corrected using the Holm–Bonferroni method to account for multiple (two) comparisons. As a result, the hybrid BCI provided a significantly higher accuracy compared to both unimodal systems ($p < 0.05$).

**Table 1.** Ternary classification accuracies (%) of participants across three test blocks.

| | Test Block 1 | | | Test Block 2 | | | Test Block 3 | | | All Test Blocks | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | EEG | fNIRS | Hybrid | EEG | fNIRS | Hybrid | EEG | fNIRS | Hybrid | EEG | fNIRS | Hybrid |
| P1 | 100*** | 83.3*** | 100*** | 95.8*** | 83.3*** | 100*** | 95.8*** | 95.8*** | 95.8*** | 97.2*** | 87.5*** | 98.6*** |
| P2 | 33.3 | 79.2*** | 62.5** | 62.5** | 79.2*** | 79.2*** | 58.3* | 79.2*** | 79.2*** | 51.4** | 79.2*** | 73.6*** |
| P3 | 79.2*** | 66.7*** | 79.2*** | 91.7*** | 70.8*** | 91.7*** | 83.3*** | 83.3*** | 83.3*** | 84.7*** | 73.6*** | 84.7*** |
| P4 | 41.7 | 29.2 | 33.3 | 50.0 | 50.0 | 54.2* | 58.3* | 12.5 | 54.2* | 50.0** | 30.6 | 47.2** |
| P5 | 75.0*** | 95.8*** | 95.8*** | 87.5*** | 100*** | 100*** | 95.8*** | 100*** | 100*** | 86.1*** | 98.6*** | 98.6*** |
| P6 | 25.0 | 37.5 | 41.7 | 54.2* | 45.8 | 62.5** | 50.0 | 66.7*** | 66.7*** | 43.1 | 50.0** | 56.9*** |
| P7 | 54.2* | 58.3* | 62.5** | 41.7 | 50.0 | 41.7 | 41.7 | 83.3*** | 66.7*** | 45.8* | 63.9*** | 56.9*** |
| P8 | 79.2*** | 54.2* | 70.8*** | 41.7 | 50.0 | 45.8 | 91.7*** | 70.8*** | 100*** | 70.8*** | 58.3*** | 72.2*** |
| P9 | 66.7*** | 70.8*** | 66.7*** | 91.7*** | 62.5** | 87.5*** | 79.2*** | 87.5*** | 87.5*** | 79.2*** | 73.6*** | 80.6*** |
| P10 | 33.3 | 45.8 | 45.8 | 33.0 | 45.8 | 41.7 | 33.0 | 25.0 | 37.5 | 33.3 | 38.9 | 41.7 |
| P11 | 37.5 | 29.2 | 45.8 | 66.7*** | 54.2* | 70.8*** | 75*** | 54.2* | 75.0*** | 59.7*** | 45.8* | 63.9*** |
| AVG | 56.82 | 59.09 | 64.02 | 65.2 | 62.88 | 70.45 | 69.32 | 68.94 | 76.89 | 63.76 | 63.64 | 70.45 |
| SD | 24.50 | 22.35 | 21.67 | 23.1 | 18.11 | 22.70 | 22.2 | 28.10 | 19.66 | 20.93 | 21.14 | 19.19 |

Accuracies with statistical significance levels of 0.05, 0.01 and 0.001 are marked with *, **, and ***, respectively. These significant levels were calculated using the permutation test approach.

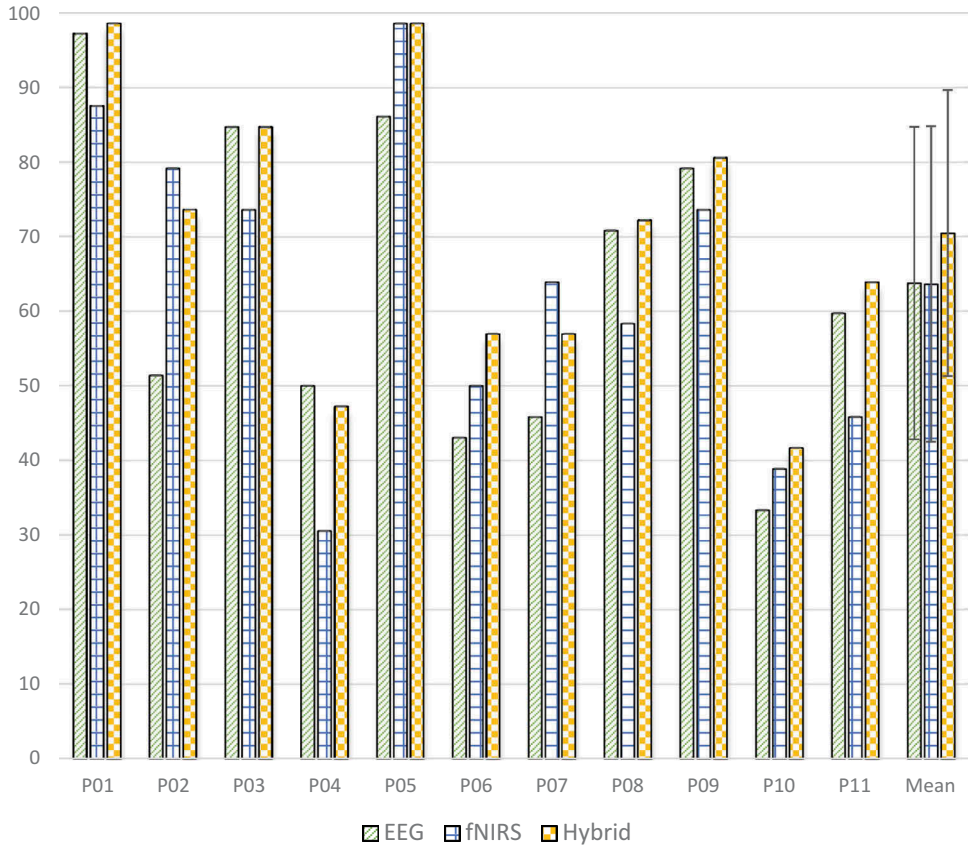


**Figure 4.** The average classification accuracy (yes versus no versus rest) over the entire test set across participants using each modality separately and together in the hybrid system.

If we consider the average accuracies across the three test blocks separately, the hybrid BCI yielded a ternary classification accuracy of 69.3%, 68.9%, and 76.9% in the first, second and third test blocks, respectively. Recall that the classifier used for each test block was trained on data from all previous blocks. Figure 5 depicts the changes in the classification accuracy across the three test blocks using each single modality and the hybrid system.

Table 2 provides the confusion matrix for the hybrid classifier on the entire test set. The classifier was most successful in correctly detecting Rest and No trials. This observation is in-line with the findings reported in [2]

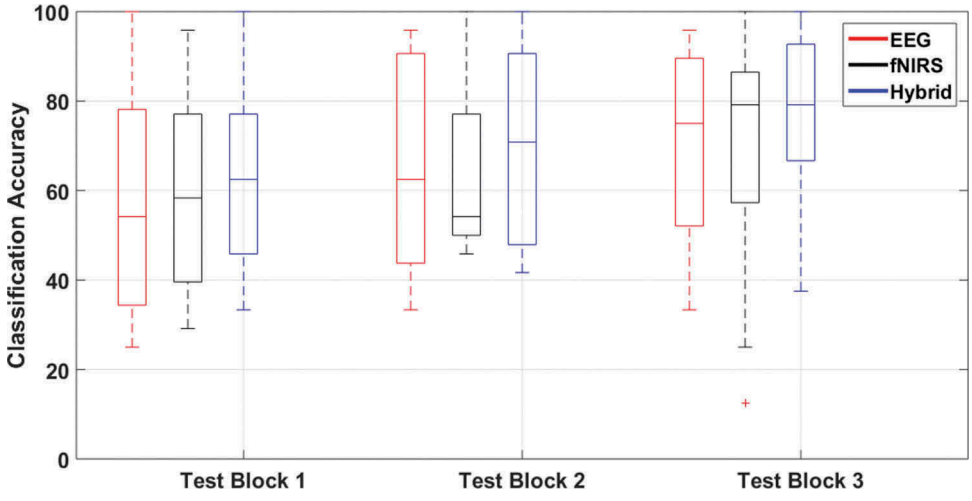


**Figure 5.** The classification accuracy (yes versus no versus rest) across the three test blocks using each modality separately and together in the hybrid system.

**Table 2.** The confusion matrix for the hybrid classifier on the entire test set.

| | Detected as Rest | Detected as Yes | Detected as No |
|---|---|---|---|
| Rest | 76.1% | 12.9% | 11.0% |
| Yes | 18.2% | 62.5% | 19.3% |
| No | 12.5% | 14.8% | 72.7% |

and [25]. We also note that for imagined speech trials ('yes' and 'no') detected incorrectly, the predicted labels were distributed approximately equally between the other two classes.

### 3.2. The contribution of each modality in the hybrid system

Figure 6 illustrates the contribution of each modality (i.e., the percentage of trials for which the decision was made by that modality). Four participants (P1, P3, P5, and P9) had almost all trials classified using a single modality. For all other participants, a mix of EEG and fNIRS data was used, with the overall mean being 64.7% of trials labeled by the EEG-BCI. The EEG modality may have had a bigger contribution in this experimental setup as the EEG electrodes covered more brain regions (central, parietal, and occipital regions) compared to fNIRS channels (see Figure 1)

### 3.3. The role of different EEG and fNIRS channels in providing discriminative information

To highlight brain regions that exhibit task-specific EEG patterns of activity, we used the average value of the Fisher criterion across participants for each frequency range and each electrode. As a reminder, RLDA ranks each feature's discriminative capability based on the Fisher criterion, with the highest score being most discriminant. Six wavelet features were extracted from each channel, representing six pseudo-frequency levels. For each of these frequency levels, only one feature per channel was generated. Figure 7 depicts the topographic maps of the Fisher criterion (averaged across all test blocks and participants) for each frequency range. To highlight the variation across participants, the standard deviation of the Fisher criterion values across participants are shown in Figure 8 using the same topographic map.

Evidently, more EEG channels provided discriminative information in higher frequency ranges (gamma and beta) than in lower frequency ranges. This finding is consistent with the previous classification of imagined speech using EEG [2]. However, the channels which provided the highest Fisher criterion value also have a large variance across participants. This inconsistency between participants could be attributed to subject-specific performance of imagined speech tasks, as well as inter-individual variations in the size and shape of different brain regions. Determining precisely which Brodmann regions provided the highest activation would require fMRI and structural data for each individual [38].

Similar analysis for the fNIRS measurements in this study are provided by [18], which showed that the fNIRS channels in the left temporal and temporoparietal regions provided the highest Fisher criterion value.

### 3.4. Reducing the duration of trials

During the experiment, participants were asked to perform 15 s of each task. To increase the information transfer rate, BCI performance for shorter task durations

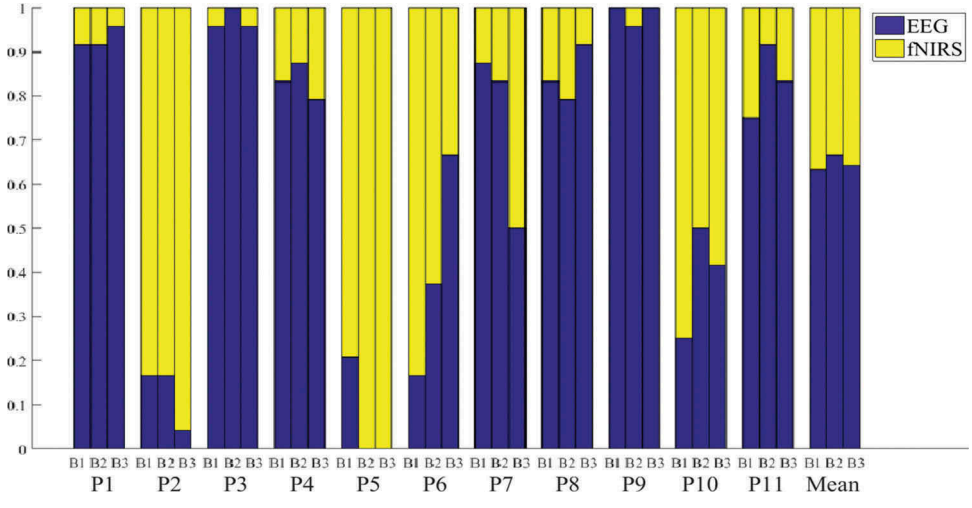


**Figure 6.** The contribution of each modality (i.e. the percentage of trials for which the decision was made by each modality) for different participants and in different blocks (B1, B2, B3 denote blocks 1, 2 and 3).

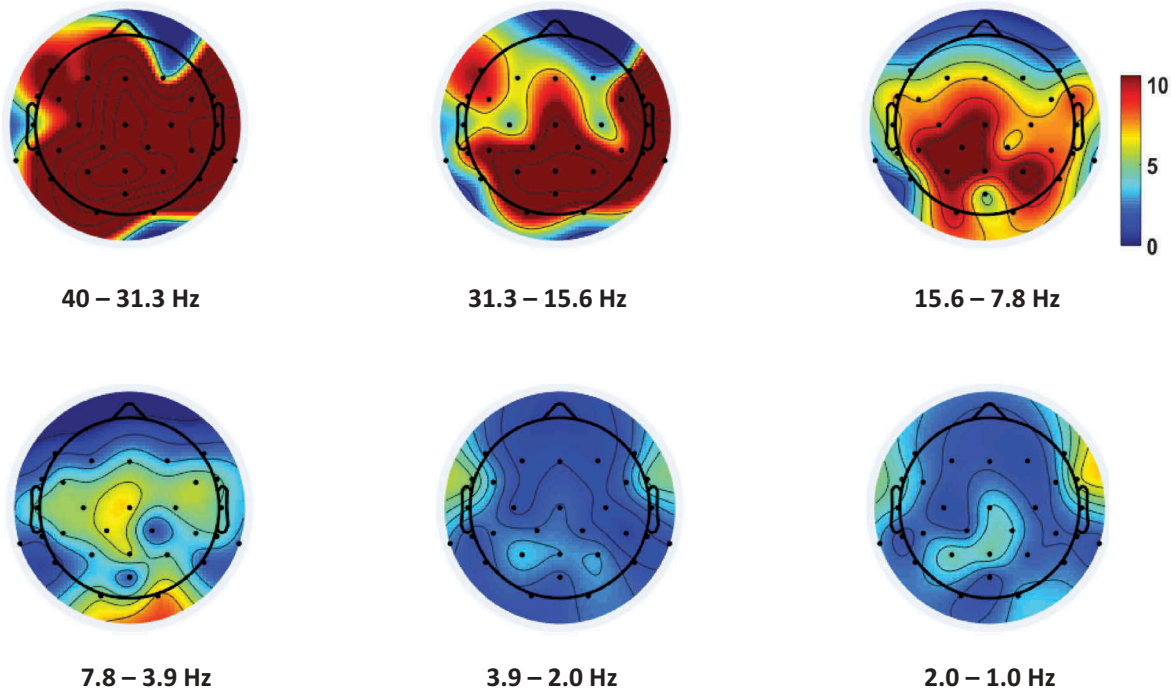


**Figure 7.** Topographic maps of the mean of the Fisher criterion (across all test blocks and participants) for each frequency range (yes versus no versus rest).

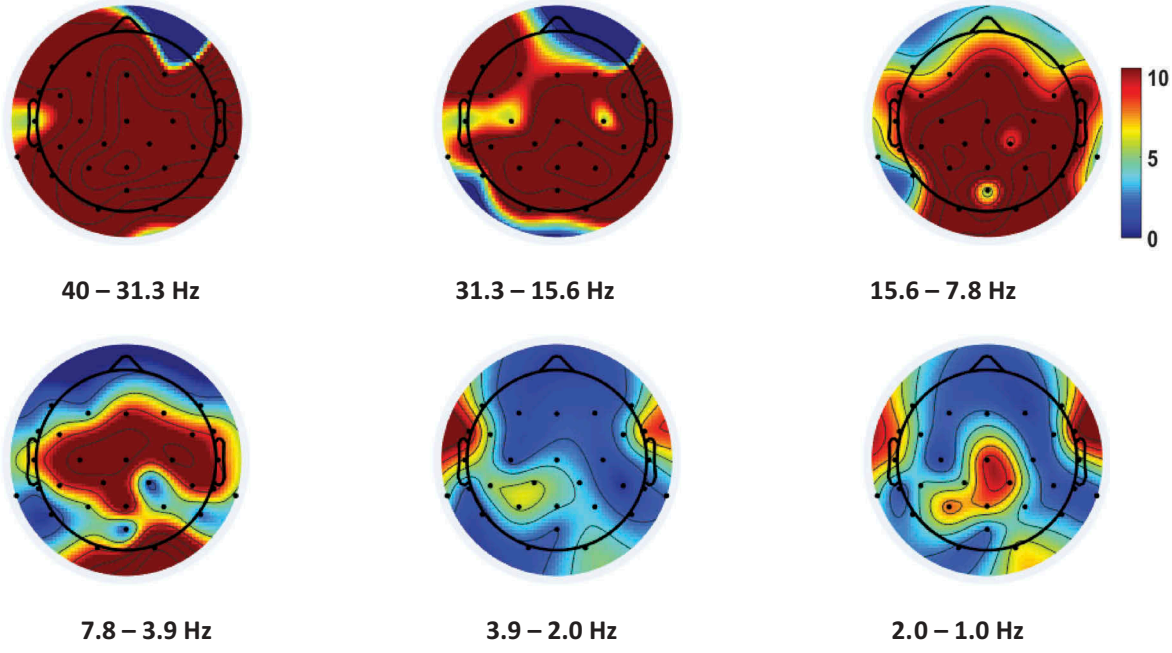


**Figure 8.** Topographic maps of the standard deviation of the Fisher criterion across participants for each frequency range (yes versus no versus rest).

were explored. Figure 9 illustrates the yes versus no versus rest classification accuracies (averaged across the three test blocks), had the duration of each trial been reduced. These hypothetical accuracies were estimated for eight different trial durations, from 8 s to 15 s (with 8 s suggested as the minimum fNIRS recording duration for observing a change in the hemodynamic response in a speech-related trial [39]). The average fNIRS accuracies increased from 59.2% to 63.6% as the duration of trials was increased from 8 to 15 s. For EEG, the average

accuracy changed from 59.2% at 8 s to 63.8 at 15 s, with some small fluctuations in between.

For the hybrid BCI, the mean accuracy across participants increased from 65.5%, at 8 s, to 70.5%, at 15 s. In general, there was a trade-off between the duration of each trial and the hybrid BCI accuracy. For most participants, the information transfer rate can be nearly doubled without much loss in accuracy. However, two participants (P6 and P7) surpassed the chance level ($p < 0.001$) only when the duration was longer than 13 s. This suggests that there may be user-specific enhancements to optimize ITR while maintaining comparable accuracy.

## 4. Discussion

### 4.1. Comparison with previous hybrid fNIRS-EEG BCIs

We proposed a 3-class hybrid fNIRS-EEG BCI based on imagined speech. An average ternary classification

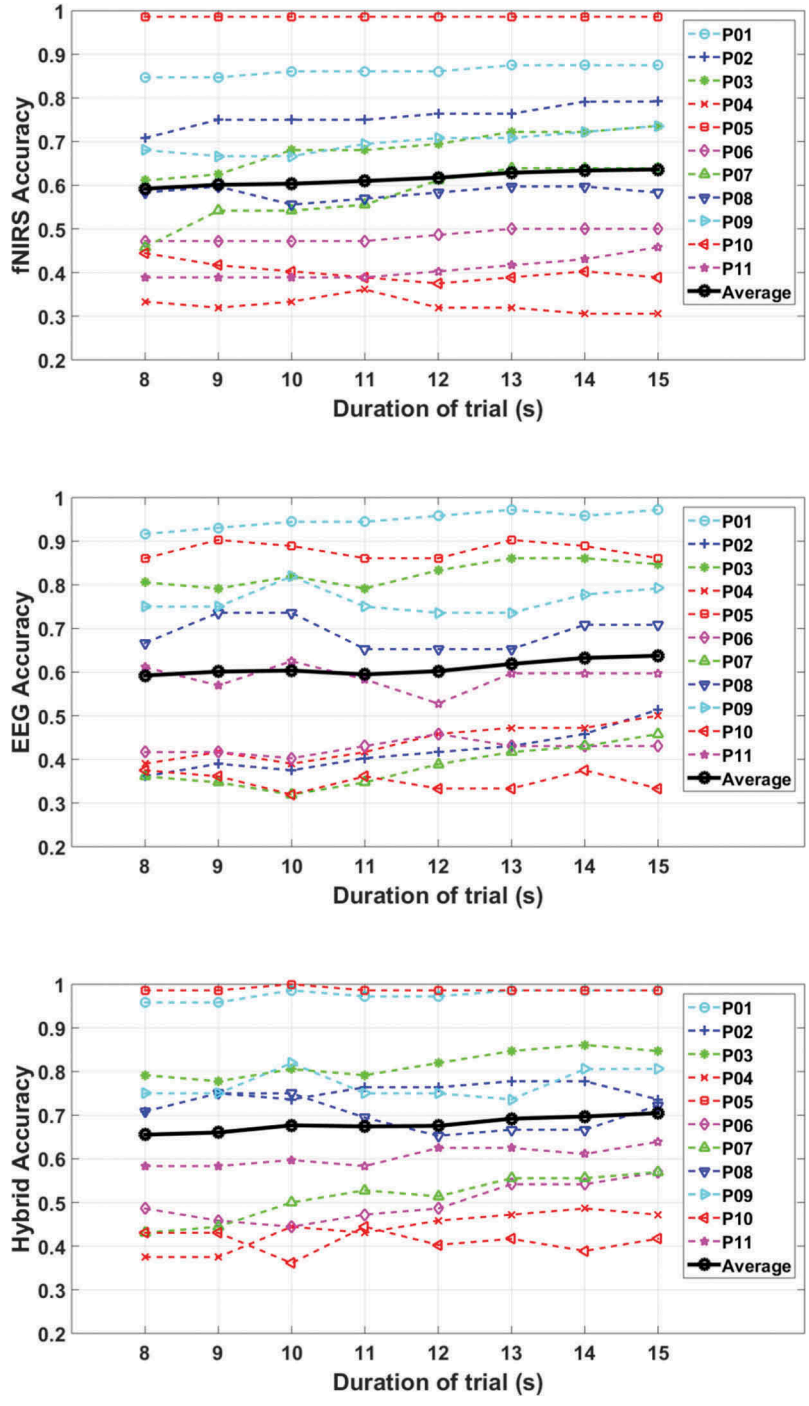


**Figure 9.** The ternary classification accuracies (averaged across the three last blocks) for different trial durations.

accuracy of 70.5 ± 19.2% was reached across all participants, with 9 out of 11 participants surpassing the upper limits of 99.9% confidence limits of chance. When averaged across participants, the hybrid BCI significantly outperformed both EEG and fNIRS BCIs with more than 6.5% enhancement in classification accuracy.

Most previous hybrid fNIRS-EEG BCIs based on active tasks focused on binary classification, either between two mental tasks [40–42] or a mental task versus the idle state [43] and mostly reported accuracy improvements of ~5% compared to each modality alone [19]. Shin et al. [44] developed one of the first hybrid 3-class BCIs (motor imagery, mental arithmetic, and idle state) and reported an average hybrid classification accuracy of ~82% with ~6% and ~18% improvements compared to EEG and fNIRS (covering only the prefrontal cortex), respectively. However, the main reason for this dramatic improvement compared to fNIRS alone may be due to the fact that motor imagery and mental math (i.e. working memory) are expected to elicit activation in different regions (sensorimotor vs. prefrontal) while the fNIRS channels were only covering the prefrontal cortex.

### 4.2. Comparison of the information transfer rate (ITR) with previous BCIs with analogous activation tasks

In this section, we compare the ITR of the proposed BCI to that of four studies using analogous activation tasks [2,16,17,25]. In all these studies, participants were asked to answer 'yes versus no' questions, presented visually or auditorily. However, the exact instruction they were given was slightly different. Sereshkeh et al. instructed participants to 'mentally rehearse the phrases "yes" and "no" without any muscle and tongue movements' after reading the question ([2,25]). Hwang et al. presented the questions auditorily and asked participants to internally answer 'yes' or 'no' [16]. And finally, Chaudhary et al. presented the questions auditorily and instructed participants to think 'ja' or 'nein' (German for 'yes' and 'no') and to avoid imagining the answer aurally or visually [17]. In our study, the questions were presented visually and the exact instruction which was given to participants was 'to think "yes" or "no" while repeating the answer mentally without any muscle or tongue movements'.

For calculating the ITR, the following equation by [45] was used:

$$ITR = \frac{60}{\tau}.\left[\log_2 N + P.\log_2 P + (1-P).\log_2 \frac{1-P}{N-1}\right] \quad (7)$$

where $\tau$ is the trial duration, $N$ is the number of classes and $P$ is the classification accuracy.

The calculated ITRs of the binary 'yes' versus 'no' BCIs presented by [16,17], and [2] were 1.08 ± 1.36, 1.06 ± 0.25, 0.94 ± 0.12 bits/min, respectively. For the ternary yes versus no versus idle state BCI presented by [18], the ITR was 0.94 ± 0.83 bits/min. In this study, we achieved the ITR of 2.03 ± 1.84 bits/min.

### 4.3. fNIRS-EEG fusion model

We proposed a novel technique to combine the data from two recording modalities. Additionally, we applied two previously suggested fNIRS-EEG fusion techniques on our dataset for comparison: (I) normalizing and merging the feature vectors from the two modalities [46] and (II) training a classification model for each classifier and use a metaclassifier for combining these two classifiers [40]. However, we found that the proposed technique provided the best performance.

The first technique is a straightforward solution, i.e., combining data from two different modalities and leverages the discriminative information from both modalities in decision-making. However, merging two feature vectors increases the input dimensionality and risk of overfitting. Hence, the number of trials needs to be sufficiently large for this technique to be optimal. Furthermore, the normalization parameters for each modality needs to be optimized for each participant, which increases the number of hyperparameters.

The second solution, which uses a metaclassifier, resembles our proposed technique. However, feeding the output scores of two classifiers into a metaclassifier does not necessarily take the reliability of each classifier into account. For instance, if a classifier is overfitted on the training data, the output scores (which serve as the input to the metaclassifier) will be high, while in reality, the classifier will fail to perform well on the test set. Using the cross-validation accuracy as a prior probability to adjust the output scores of each classifier can address this problem. As seen in Figure 6, for most participants, one recording modality tended to dominate the decision-making, although the dominant modality varied across participants. In other words, the proposed technique appears to be ideal when the preferred BCI modality is participant-specific or may change from trial to trial.

### 4.4. Limitations and future directions

For future studies, this BCI can be developed with the option of making the decision sooner than the entire trial period, if the classifier's confidence surpasses a certain

threshold. Our findings showed that some participants reached their highest performance in less than 10 s, while others required a longer duration to surpass chance. The option of selecting the class type earlier than the end of a trial can improve the ITR for some participants.

Future research can also investigate the effect of increasing the number of sessions, allowing for more trials for training the classifier. In this study, the average classification accuracy increased in each test block (see Figure 5), which is possibly due to the increased volume of training data. With more trials, the maximum achievable classifier performance could be determined.

To progress toward a subject-independent hybrid BCI, future work may investigate the merit of a general model based on data from all participants, and using transfer learning [47] to fine tune the model using a small dataset from each new participant. This may be more feasible due to the innate subject-dependent nature of active BCIs.

Finally, prior to any clinical translation of these findings, the results should be replicated on the main target population of BCI research, i.e., individuals who present as locked-in.

## 5. Conclusion

This study investigated a hybrid 3-class fNIRS-EEG-BCI based on imagined speech. Eleven participants, across two sessions, performed multiple iterations of three different mental tasks: thinking 'yes' or 'no' while mentally repeating the word for 15 s, or an equivalent duration of unconstrained rest. BCI classification was performed using EEG data and fNIRS data alone, as well as a hybrid fNIRS-EEG classification model. The hybrid model provided an average ternary classification accuracy of 70.5 ± 19.2% (>6.5% improvement compared to EEG and fNIRS alone) across participants with nine out of eleven participants surpassing chance (compared to seven for EEG or fNIRS alone). Our findings suggest that concurrent measurements of EEG and fNIRS can improve both classification accuracy and the information transfer rate of BCIs based on imagined speech. We also calculated and provided the performance of the BCI for shorter durations of mental tasks and showed that for most participants, the ITR of the hybrid system can be almost doubled with a negligible drop in the accuracy. To the best of our knowledge, this is the first report of a multimodal and multiclass BCI based on imagined speech and moves toward a more reliable intuitive BCI.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## ORCID

Alborz Rezazadeh Sereshkeh http://orcid.org/0000-0003-2680-646X
Rozhin Yousefi http://orcid.org/0000-0003-2402-5467
Andrew T Wong http://orcid.org/0000-0002-1881-5561
Frank Rudzicz http://orcid.org/0000-0002-1139-3423
Tom Chau http://orcid.org/0000-0002-7486-0316

## References

[1] Nicolas-Alonso LF, Gomez-Gil J. Brain computer interfaces, a review. Sensors. 2012;12(2):1211–1279.
[2] Sereshkeh AR, Trott R, Bricout A, et al. Online EEG classification of covert speech for brain–computer interfacing. Int J Neural Syst. 2017;27(08):1750033.
[3] Yousefi R, Rezazadeh Sereshkeh A, Chau T. Online detection of error-related potentials in multi-class cognitive task-based BCIs. Brain-Comput Interfaces. 2019;6:1–12.
[4] Yousefi R, Sereshkeh AR, Chau T. Development of a robust asynchronous brain-switch using ErrP-based error correction. J Neural Eng. 2019;16:066042.
[5] DaSalla CS, Kambara H, Sato M, et al. Single-trial classification of vowel speech imagery using common spatial patterns. Neural Networks. 2009;22(9):1334–1339.
[6] Cramer SC, Lastra L, Lacourse MG, et al. Brain motor system function after chronic, complete spinal cord injury. Brain. 2005;128(12):2941–2950.
[7] Conson M, Sacco S, Sarà M, et al. Selective motor imagery defect in patients with locked-in syndrome. Neuropsychologia. 2008;46(11):2622–2628.
[8] Schultz T, Wand M, Hueber T, et al. Biosignal-based Spoken Communication : A Survey. IEEE/ACM Trans Audio Speech Lang Process2015;14(8):1–15.
[9] Schultz T, Wand M, Hueber T, et al., Biosignal-based spoken communication: a survey. IEEE/ACM Trans Audio Speech Lang Process. 2017; 25:2257–2271
[10] Brigham K, Kumar BVKV, "Subject identification from Electroencephalogram (EEG) signals during imagined speech," *IEEE 4th International conference of biometrics theory, Appl Syst BTAS 2010*, Washington, DC, 2010.
[11] Wang L, Zhang X, Zhong X, et al. Analysis and classification of speech imagery EEG for BCI. Biomed Signal Process Control. 2013;8(6):901–908.
[12] Torres-García AA, Reyes-García CA, Villaseñor-Pineda L, et al. Implementing a fuzzy inference system in a multi-objective EEG channel selection model for imagined speech classification. Expert Syst Appl. 2016;59:1–12.
[13] Kübler A, Mushahwar VK, Hochberg LR, et al. BCI Meeting 2005 - Workshop on clinical issues and applications. IEEE Trans Neural Syst Rehabil Eng. 2006;14(2):131–134.
[14] Scholkmann F, Wolf M, Wolf U. The effect of inner speech on arterial CO2 and cerebral hemodynamics and oxygenation: A functional NIRS study Oxygen Transport to Tissue XXXV. In: S Van Huffel et al editors. Advances in experimental medicine and biology; vol 789; 2013. p. 81–7.

[15] Gallegos-Ayala G, Furdea A, Takano K, et al. Brain communication in a completely locked-in patient using bedside near-infrared spectroscopy. Neurology. 2014;82:1930–1932.
[16] Hwang H-J, Choi H, Kim J-Y, et al. Toward more intuitive brain–computer interfacing: classification of binary covert intentions using functional near-infrared spectroscopy. J Biomed Opt. 2016;21(9):091303.
[17] Chaudhary U, Xia B, Silvoni S, et al. Brain–computer interface–based communication in the completely locked-in state. PLoS Biol. 2017;15(1):1–26.
[18] Sereshkeh AR, Yousefi R, Wong AT, et al. Online classification of imagined speech using functional near-infrared spectroscopy signals. J Neural Eng. 2018;16(1):016005.
[19] Ahn S, Jun SC. Multi-modal integration of EEG-fNIRS for brain-computer interfaces – current limitations and future directions. Front Hum Neurosci. 2017;11(October):1–6.
[20] Chiarelli AM, Zappasodi F, Di Pompeo F, et al. Simultaneous functional near-infrared spectroscopy and electroencephalography for monitoring of human brain activity and oxygenation: a review. Neurophotonics. 2017;4(04):1.
[21] Choi I, Rhiu I, Lee Y, et al. A systematic review of hybrid brain-computer interfaces: taxonomy and usability perspectives. PloS one, 2017;12(4),e0176674..
[22] Hong KS, Khan MJ. Hybrid brain-computer interface techniques for improved classification accuracy and increased number of commands: A review. Front Neurorobot. 2017;11(Jul):35.
[23] Mognon A, Jovicich J, Bruzzone L, et al. ADJUST: an automatic EEG artifact detector based on the joint use of spatial and temporal features. Psychophysiology. 2011;48(2):229–240.
[24] Subasi A. EEG signal classification using wavelet feature extraction and a mixture of expert model. Expert Syst Appl. 2007;32(4):1084–1093.
[25] Sereshkeh AR, Trott R, Bricout A, et al. Eeg classification of covert speech using regularized neural networks. IEEE/ACM Trans Audio Speech Lang Process. 2017;25 (12):2292–2300.
[26] Adeli H, Zhou Z, Dadmehr N. Analysis of EEG records in an epileptic patient using wavelet transform. J Neurosci Methods. 2003;123(1):69–87.
[27] Ocak H. Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy. Expert Syst Appl. 2009;36(2 PART 1):2027–2036.
[28] Subasi A. Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients. Expert Syst Appl. 2005;28(4):701–711.
[29] Yamashita Y, Maki A, Koizumi H. Wavelength dependence of the precision of noninvasive optical measurement of oxy-, deoxy-, and total-hemoglobin concentration. Med Phys. 2001;28:1108–1114.
[30] Lotte F, Congedo M, Lécuyer A, et al. A review of classification algorithms for EEG-based brain-computer interfaces. J Neural Eng. 2007;4(2): R1–R13.
[31] Bharne PP. Classification techniques in brain computer interface : a review. An International Journal of Engineering & Technology (AIJET). 2015;2(1):1–7. http://www.aijet.in/ eISSN: 2394-627X
[32] Naseer N, Hong K-S. fNIRS-based brain-computer interfaces: a review. Front Hum Neurosci. 2015;9 (January):1–15.
[33] Strother SC, Rasmussen PM, Churchill NW, et al. Stability and Reproducibility in fMRI Analysis. Pract Appl Sparse Model. 2012 (2009).
[34] Guo Y, Hastie T, Tibshirani R. Regularized linear discriminant analysis and its application in microarrays. Biostatistics, 2007;8(1):86–100.
[35] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Appears in the International Joint Conference on Articial Intelligence (IJCAI), Montreal, Quebec, Canada, 1995.
[36] Ledoit O, Wolf M. Honey, I shrunk the sample covariance matrix. J Portf Manag. 2004;30(4):110–119.
[37] Combrisson E, Jerbi K. Exceeding chance level by chance: the caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. J Neurosci Methods. 2015;250:126–136.
[38] Koessler L, Maillard L, Benhadid A, et al. Automated cortical projection of EEG sensors: anatomical correlation via the international 10-10 system. Neuroimage. 2009;46(1):64–72.
[39] Walsh B, Tian F, Tourville JA, et al. Hemodynamics of speech production: an fNIRS investigation of children who stutter. Sci Rep. 2017;7(1):1–13.
[40] Fazli S, Mehnert J, Steinbrink J, et al. Enhanced performance by a hybrid NIRS-EEG brain computer interface. Neuroimage. 2012;59(1):519–529.
[41] Blokland Y, Spyrou L, Thijssen D, et al. Combined EEG-fNIRS decoding of motor attempt and imagery for brain switch control: an offline study in patients with tetraplegia. IEEE Trans Neural Syst Rehabil Eng. 2014;22(2):222–229.
[42] Morioka H, Kanemura A, Morimoto S, et al. Decoding spatial attention by using cortical currents estimated from electroencephalography with near-infrared spectroscopy prior information. Neuroimage. 2014;90:128–139.
[43] Shin J, Müller KR, Schmitz CH, et al. Evaluation of a compact hybrid brain-computer interface system. Biomed Res Int. 20172017;2017:1–11.
[44] Shin J, Kwon J, Im C-H. A ternary hybrid EEG-NIRS brain-computer interface for the classification of brain activation patterns during mental arithmetic, motor imagery, and idle state. Front Neuroinform. 2018;12 (February):1–9.
[45] Dornhege G, Del Millán JR, Hinterberger T, et al. Toward brain-computer interfacing. Neural Inf Process Ser. MIT press; 2007 (December 2014):507.
[46] Yin X, Xu B, Jiang C, et al. A hybrid BCI based on EEG and fNIRS signals improves the performance of decoding motor imagery of both force and speed of hand clenching. J Neural Eng. 2015;12(3):36004.
[47] Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng. 2010;22(10):1345–1359.