Visual and haptic feedback in detecting motor imagery within a wearable brain-computer interface

Visual and haptic feedback in detecting motor imagery within a wearable brain-computer interface

Pasquale Arpaia[a,b,c], Damien Coyle[e], Francesco Donnarumma[a,f], Antonio Esposito[a,d], Angela Natalizio[a,g], Marco Parvis[g]

[a] *Augmented Reality for Health Monitoring Laboratory (ARHeMLab).*
[b] *Department of Electrical Engineering and Information Technology (DIETI), Universita' degli Studi di Napoli Federico II, Naples, Italy.*
[c] *Centro Interdipartimentale di Ricerca in Management Sanitario e Innovazione in Sanità (CIRMIS), Università degli Studi di Napoli Federico II, Naples, Italy.*
[d] *Centro Servizi Metrologici e Tecnologici Avanzati (CeSMA), Università degli Studi di Napoli Federico II, Naples, Italy.*
[e] *Intelligent Systems Research Centre, University of Ulster, Derry, Northern Ireland.*
[f] *Institute of Cognitive Sciences and Technologies, National Research Council (ISTC-CNR), Rome, Italy.*
[g] *Department of Electronics and Telecommunications (DET), Polytechnic of Turin, Turin, Italy.*

Abstract

This paper presents a wearable brain-computer interface relying on neurofeedback in extended reality for the enhancement of motor imagery training. Visual and vibrotactile feedback modalities were evaluated when presented either singularly or simultaneously. Only three acquisition channels and state-of-the-art vibrotactile chest-based feedback were employed. Experimental validation was carried out with eight subjects participating in two or three sessions on different days, with 360 trials per subject per session. Neurofeedback led to statistically significant improvement in performance over the two/three sessions, thus demonstrating for the first time functionality of a motor imagery-based instrument even by using an utmost wearable electroencephalograph and a commercial gaming vibrotactile suit. In the best cases, classification accuracy exceeded 80 % with more than 20 % improvement with respect to the initial performance. No feedback modality was generally preferable across the cohort study, but it is concluded that the best feedback modality may be subject-dependent.

Keywords: brain-computer interface, motor imagery, electroencephalography,

extended reality, haptic, neurofeedback.

1. Introduction

Motor imagery consists of imagining a movement without executing it. Interestingly, the neuronal activities during both the execution and the imagination of a movement are compatible. Both of them induce "event-related desynchronization" and "event-related synchronization" of the μ and β rhythms [1–3]. Because of this, motor imagery is widely exploited in building brain-computer interfaces (BCI) [4] as it offers an alternative way to communicate motor intentions without involving peripheral nerves or muscles [5]. Motor imagery-based BCIs are powerful tools both for people with [6–9] and without motor disabilities [10]. Application examples range from controlling a wheelchair [7] or a robotic arm [8] to navigating a virtual environment [10] or assessing awareness in disorder of consciousness [11] or implementing a speller [9]. Such BCIs typically rely on electroencephalography (EEG) to measure brain activity due to its non-invasiveness, low cost, and wearability [12, 13]. However, in contrast with other common BCI paradigms [14, 15], the user must be trained to properly control a BCI based on motor imagery. In this framework, neurofeedback helps the user to self-learn to modulate sensorimotor rhythms intentionally.

Fig. 1 represents a closed-loop metrological chain where neurofeedback is used to support the modulation of EEG rhythms [16, 17]. As a consequence, this aims to enhance the performance in BCI control applications [18]. According to the literature, unimodal feedback such as visual, auditory, and haptic, are compatible in terms of performance [19–21]. Among them, haptic (somatosensory) feedback could improve the sense of agency in motor imagery BCI's [21]. Moreover, it has the potential to enhance cortical activation and system performance as well as increase the pertinence of provided feedback [21–23].

Multimodal feedback is also sought to enhance user engagement [24, 25]. The most commonly investigated multimodal feedback combines visual and haptic somatosensory feedback modalities. Recent studies showed virtual hands ap-

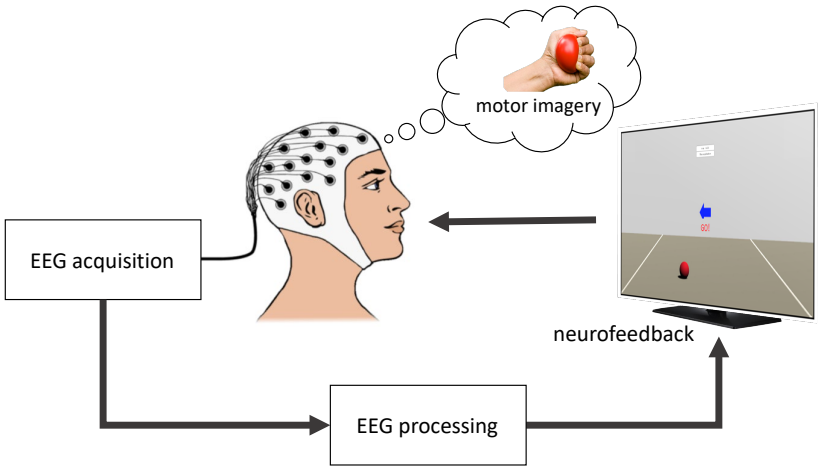


Figure 1: Representation of visual neurofeedback in a brain-computer interface based on motor imagery. EEG: electroencephalography.

pearing on a screen while an electrical [26] or vibrotactile [27] stimulation was delivered to the user's hand. In both cases, results suggest that multimodal feedback was beneficial for motor imagery detection, even in comparison with a unimodal feedback modality. Meanwhile, the authors in [28] demonstrated that better detection accuracy[1] was associated with visual or multimodal feedback if compared to the vibrotactile feedback alone. In that case, haptic feedback was provided by two vibrating motors placed on the wrists. Finally, three tactile actuators were used in [30] to stimulate the shoulder blade. However, no significant differences were found in terms of detection accuracy between visual or visual-haptic guide.

There is thus evidence that neurofeedback enhances the detection of motor imagery. However, there is no consensus about what is the best feedback modality or what is the best way to provide it/them. Notably, the number of

[1]The concept of classification accuracy is taken into account, namely the ratio between the number of correctly classified tasks and the total number of tasks. This should not be confused with measurement accuracy, namely "the closeness of agreement between a measured quantity value and a true quantity value of a measurand" [29].

EEG acquisition channels in the above-mentioned studies ranged from 11 to 64. However, previous research also suggests that three is a minimum number of channels to properly measure sensorimotor rhythms [31], so that a lower number of channels could be used to achieve user comfort, wearability, portability, and ease of use [32, 33].

On these premises, the present work investigates different neurofeedback modalities within the implementation of a wearable BCI based on motor imagery. In acquiring brain signals, only three differential channels were chosen a-priori to achieve utmost wearability and portability. Meanwhile, visual and haptic feedbacks were presented through a custom virtual reality scenario. For the first time, a wearable haptic suit was exploited as an actuator for chest vibrotactile feedback. The remainder of the paper is organized as follows. Section 2 presents the proposed system with specific regards to hardware, software, and signal processing. Then, Section 3 discusses the experimental procedures adopted to validate the instrument prototype, while Section 4 reports the results of an experimental campaign carried out according to those methods.

2. Proposal

The proposed closed-loop wearable BCI is presented in this section. Two feedback modalities were adopted either singularly or simultaneously, namely visual and vibrotactile modalities. The former consisted of a rolling virtual ball, while the latter was a vibration delivered on the chest. As this study aims to maximize user engagement and comfort, EEG signals were acquired through a recently commercialized wireless cap, FlexEEG [34], while the chest-based feedback was delivered with a suit designed for immersive experiences.

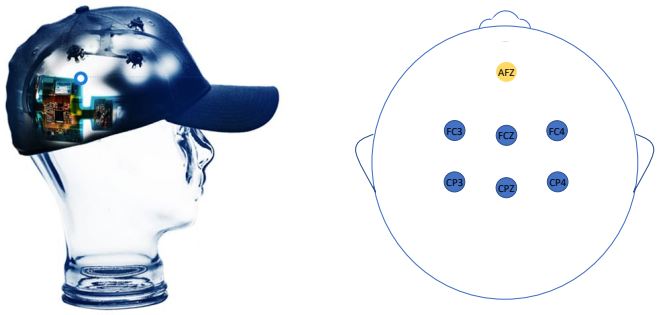
The block diagram of the BCI system is shown in Fig. 2. The EEG signals, acquired from the user's scalp, are sent via Bluetooth to a custom Simulink model embedding online signal processing. The EEG processing output is sent to a purposely designed Unity application, and it is employed for modulating the sensory feedback. Thus, the loop is closed by delivering the neurofeedback

Figure 2: Block diagram of the wearable brain-computer interface. The information exchanged between blocks (black) and the exploited communication protocols (blue) are highlighted. EEG: electroencephalography, UDP: User Datagram Protocol, NF: neurofeedback.

to the user. In addition, the Unity application also dictates the timing for the motor imagery tasks (synchronous cue-based paradigm [35]). Details about the system implementation are discussed in the following subsections.

2.1. Wearable hardware

The hardware of the wearable BCI system involves two main devices: a commercial electroencephalograph, and a haptic suit for vibrotactile feedback. Visual feedback was instead delivered through the screen of a personal computer, though, in a real application, the visual feedback may be naturally provided as an effect of the control task.



(a) (b)

Figure 3: Wearable and portable EEG acquisition system: (a) EEG cap with electrodes, (b) FlexMI channels configuration with three pairs and a reference electrode.

EEG acquisition was carried out with the FlexEEG headset by NeuroCON-

(a) (b)

Figure 4: Wearable and portable haptic suit with a double vibration motors matrix: (a) suit, (b) front view of matrix with motors.

CISE Ltd[2] shown in Fig. 3. The headset was used with the FlexMI channels montage, which is specifically designed to record the sensorimotor area of the brain. Notably, it consists of three differential channels placed at FC3-CP3, FCz-CPz, and FC4-CP4, while the ground electrode is at AFz (Fig. 3(b), see [36] for the international 10/20 EEG standard locations). Conductive gel was used to ensure low contact impedance and high stability at the scalp interface. The EEG signals were filtered and amplified by the electronic board. Then, these signals were digitized with 16-bit resolution by sampling at 250 Sa/s and then down-sampling to 125 Sa/s. The data were finally transmitted via Bluetooth 2.0.

The hardware for the haptic feedback consists of the vibrotactile suit from bHaptics Inc [3] shown in Fig. 4. This is wearable and portable, and it is primarily commercialized for gaming. It provides a double 5×4 matrix with motors installed on the front and back of the torso. Vibration can be modulated in terms of intensity per each single motor, and patterns can be created to give a specific haptic sensation to the user. In the current application, vibration

[2]https://www.neuroconcise.co.uk/technology/

[3]https://www.bhaptics.com/tactsuit/tactsuit-x40

patterns consist of activating a column of five motors vibrating at the same time. The column shifts to the left or to the right side of the torso starting from the front centre. They are controlled through Bluetooth via the Unity application according to online classification of EEG signals.

2.2. Software application

A virtual scenario was developed with the Unity games engine to deliver the feedbacks and dictate the timing of the experiment and trials. The visual feedback consisted of a virtual ball with the capability to roll in one dimension to the left or the right side of a screen. Gravity was applied to the ball to keep it tied to the virtual floor (Fig. 5). Its movement was controlled in accordance with online classification of EEG signals where the class and score (i.e., class probability) associated with the EEG signals determine the force applied to the ball in terms of direction and intensity, respectively. Meanwhile, the haptic pattern (columns of vibrating motors) could be modulated in terms of position and intensity, again according to class and score.

Through the graphical interface, the experimenter could select the feedback to deliver: visual, vibrotactile, or both. When the visual feedback was not wanted, the virtual ball disappeared. Instead, if the vibrotactile feedback was unwanted, the suit was shut down. In any case, the task indication was always provided with an arrow appearing on the screen. Details on the timing will be provided in Section 3.

2.3. Signal processing

In order to translate the brain activity into control commands, the acquired EEG data were processed both online and offline with an algorithm based on "filter-bank common spatial pattern" (FBCSP) [37, 38]. This approach was successfully replicated and tested on benchmark datasets [37–40], and some studies even showed its efficacy in analysing differential channel data [41]. FBCSP involves the following steps:

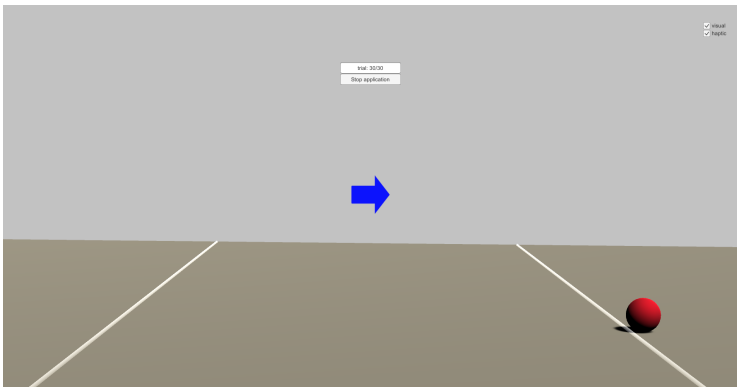


Figure 5: Visual feedback consisting of a ball rolling according with motor imagery.

- the EEG signals are filtered with an array of 17 bandpass Type II Chebyshev filters (tenth order, attenuation 50 dB) from 4 Hz to 40 Hz (4-8, 6-10, 8-12, ..., 36-40 Hz);
- the data from the three channels are spatially filtered using the Common Spatial Pattern (CSP) algorithm, which transforms the raw EEG signals to maximize the variance of one class while minimizing the variance of the other;
- the most informative features are selected by means of the Mutual Information-based Best Individual Features (MIBIF) algorithm;
- the signal features were classified using a Bayesian approach, namely the Naive Bayesian Parzen Window (NBPW).

The steps of the FBCSP are shown in Fig. 6. Further details about the implementation of each processing block are reported in the literature [37, 38]. It is worth mentioning that the adopted classifier assigns a probability to the two possible classes (left and right), and hence the most probable class is assigned to the processed EEG data. By exploiting the class and its probability as a score, the feedback could be modulated in terms of direction and intensity, respectively.

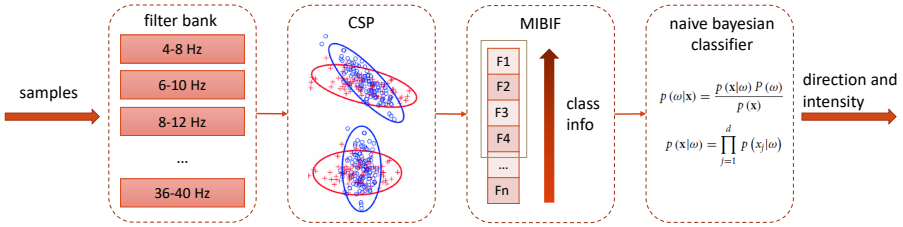


Figure 6: EEG signals processing in the proposed wearable brain-computer interface. CSP: Common Spatial Pattern, MIBIF: Mutual Information-based Best Individual Features.

Note that the algorithm must be trained before being used for online classification. For this purpose, a first set of data was acquired subject-by-subject without any pre-processing being applied (more details in Sec. 3.1). The model for EEG processing was thus identified with these initial data for each subject. Then, in a second step, the identified model was used for online classification of unlabelled EEG data. In the online version, the EEG data stream was processed with a sliding window of fixed duration in order to provide a continuous feedback. The width of the sliding window was fixed at 2.00 s while the shift between consecutive windows was fixed at 0.25 s. These choices were made both according to empirical evidence from preliminary measures and according to literature [42]. The same FBCSP approach was adopted in offline analyses as well.

3. Experimental validation

The experimental protocol is described in the following section and the methods for offline analyses of the results are discussed. Some details are also given about the participant in this preliminary validation.

3.1. Experimental protocol

The experiments were carried out in two or three sessions on different days, each lasting about two hours. Participants were asked to imagine the movement of the left or right hand. A single experimental session is depicted in Fig. 7.

In each session, they first performed motor imagery without any feedback (noF block), which was required for identifying the online classification model. Then, they received online feedback (NF block). The presentation of the three feedback modalities were randomized for each subject to avoid biases associated with the sequence of presentation. A questionnaire (see Tab. 1) was administered during each session to monitor changes in the participants' mental and physical state between blocks. This was adapted from [43] to include neurofeedback-related aspects.

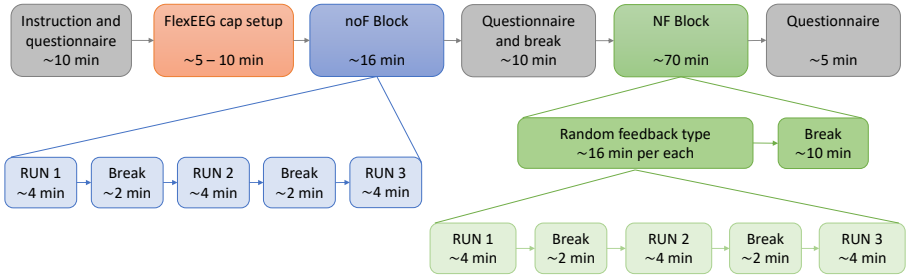


Figure 7: Structure of a single session of the experimental campaign. noF: no feedback, NF: neurofeedback.

The noF block consisted of three runs with 30 trials each (15 per class) with about two-minute breaks in between. The order of the cue-based motor imagery was again randomized to avoid any bias. The timing of a single trial was recalled from the standard paradigms of BCI competitions [44]. In particular, it consists of an initial relax, a cue at $t = 2\,\mathrm{s}$ indicating the task to carry on, motor imagery starting at $t = 3\,\mathrm{s}$, and motor imagery ending at $t = 6\,\mathrm{s}$. Final relaxation was then presented, and its duration was randomized between 1 s to 2 s.

After the first block, a 10-minute break was employed to continue the questionnaire and identify the online classification model. In particular, the FBCSP was used in a 5-fold cross validation with 10 repetitions for selecting the best time window in terms of optimal classification accuracy. The time-varying classification accuracy and the associated standard deviation were calculated with a 2.00 s wide sliding window to span the 0.00 s to 7.00 s range with a 0.25 s

Experimental Information at start	
Date	yyyy:mm:dd
Session	#
Starting time	hh:mm
Handedness	1: left / 2: right / 3: both
Age	#
Sex	1: male / 2: female
Do you practice any sport?	0: no / 1: yes / 2: professional
BCI experience	0: no / 1: active / 2: passive / 3: reactive / 4: multiple types
Biofeedback experience	0: no / number: how many times
How long did you sleep?	number: hours
Did you drink coffee within the past 24 h?	0: no / number: hours before
Did you drink alcohol within the past 24 h?	0: no / number: hours before
Did you smoke within the past 24 h?	0: no / number: hours before
How do you feel?	Anxious 1 2 3 4 5 Relaxed
	Bored 1 2 3 4 5 Excited
	(Physical state) Tired 1 2 3 4 5 Very good
	(Mental state) Tired 1 2 3 4 5 Very good
Which motor imagery are you confindent with?	1: grasp / 2: squeeze / 3: kinesthetic / 4: other
After training block	
How do you feel?	(Attention level) Low 1 2 3 4 5 High
	(Physical state) Tired 1 2 3 4 5 Very good
	(Mental state) Tired 1 2 3 4 5 Very good
Have you nodded off/slept a while?	No 1 2 3 4 5 Yes
How easy was motor imagery?	Hard 1 2 3 4 5 Easy
How do you feel?	(Attention level) Low 1 2 3 4 5 High
	(Physical state) Tired 1 2 3 4 5 Very good
	(Mental state) Tired 1 2 3 4 5 Very good
Have you nodded off/slept a while?	No 1 2 3 4 5 Yes
Did you feel to control the feedback?	(Visual) No 1 2 3 4 5 Yes
	(Haptic) No 1 2 3 4 5 Yes
	(Multimodal) No 1 2 3 4 5 Yes
How easy was motor imagery?	Hard 1 2 3 4 5 Easy
After the motor imagery experiment	
Which type of feedback did you prefer?	0: v / 1: h / 2: v-h
How do you feel?	Anxious 1 2 3 4 5 Relaxed
	Bored 1 2 3 4 5 Excited
How was this experiment?	(Duration) Too long 1 2 3 4 5 Good
	(Timing) Too fast 1 2 3 4 5 Good
	(Environment) Poor 1 2 3 4 5 Good
	(System) Uncomfortable 1 2 3 4 5 Comfortable

Table 1: Questionnaire administered to the participants during each experimental session.

shift. For each subject, the best time window was chosen in terms of maximum classification accuracy during the motor imagery task and minimum difference between within class accuracies. The online model was hence trained by using such a window. A further run of the noF block was repeated if the classification results were compatible with randomness.

Once the model was trained, participants performed the NF block, namely they received feedback in response to the motor imagery task. In each of these trials, the online processing began 0.50 s after the cue at $t = 2$ s and it relied on a sliding time window of 2.00 s shifting of 0.25 s until the end of the motor imagery task. For each type of feedback, three runs with 30 trials each and two classes

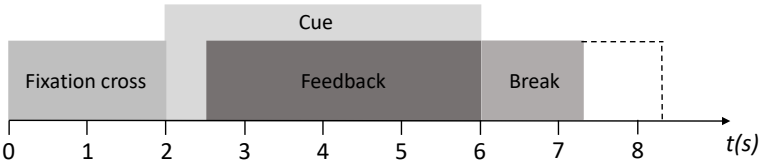


Figure 8: Timing diagram of a single trial in the BCI experiment with neurofeedback.

of imagery were recorded in total. The timing of a trial with feedback is shown in Fig. 8. Note that, unlike the previous block, in this case the imagination starts from the cue. The participants were asked to maintain high concentration throughout the entire motor imagery task, even if the feedback did not respond correctly. Between feedback types, a 10-minute break was given.

Regarding the visual feedback, the goal for the user was to overcome the white line (Fig. 5). Instead, for the haptic feedback, the goal for the user was to activate the haptic feedback maximally on the respective side of the chest. Finally, in the multimodal feedback case, the aforementioned feedbacks were jointly provided. During feedback, the virtual ball or the haptic pattern could move only if the obtained EEG class was equal to the assigned task (positive bias [45]). Otherwise, no feedback was provided, and the virtual ball was drawn towards the centre of the screen while the vibration intensity was stopped.

3.2. Offline data analysis

After the experiments, the 360 available trials were analysed offline per session per subject. Firstly, baseline removal was applied by considering the 100 ms before the cue. Then, the time-varying accuracy was calculated for all subjects, blocks and sessions by means of cross validation [46]. A permutation test was performed for each session, subject and block. The purpose was to validate the results obtained in the time-varying analysis by evaluating how far these were from random classification. Hence, the labels associated with the left and right motor imagery tasks were randomly permuted and the time-varying analysis was repeated. In both cases, the cross-validation was performed.Finally, the

comparison between the results with permuted and non-permuted labels was carried out by using the non-parametric Wilcoxon test.

Next, by relying on the best 2.00 s time window in terms of classification accuracy, the one-way analysis of variance (ANOVA) was performed to compare the accuracies in different conditions. To check for the normality assumption of the distributions, the Jarque-Bera test [47] was performed. Instead, the homoscedasticity was tested by means of the Bartlett's test. When the assumption of homoscedasticity was violated, Welch's correction to ANOVA was applied. When the distributions were not normally distributed, the Kruskal–Wallis non-parametric test was used.

3.3. Subjects

Eight right-handed volunteers (three males, mean age 28 years, and five females, mean age 25 years) participated in the experiments. These were conducted at the Augmented Reality for Health Monitoring Laboratory (ARHeM-Lab, University of Naples Federico II) in Italy. When designing the experiments, the number of subjects was chosen according to the expected effect size [48, 49] due to neurofeedback. Notably, a preliminary study [50] suggested a Cohen's d index greater than 0.8, namely a large effect size [51]. Consequently, a sample size of six was enough to achieve a statistical power around 95 % [52]. Nonetheless, eight subjects were actually considered in accordance with recent studies in this field [53, 54].

All subjects had no brain injury or motor impairment, and did not report any other medical or psychological illness/medication. Moreover, they had normal or corrected to normal vision. All subjects signed an informed consent before taking part to the experiment. By means of the questionnaire, it emerged that half of the participants played sport (S01, S03, S05, S08), though none of the participants practiced them at a professional level. Subjects S03 and S05 had previous experience with multiple BCI paradigms, S08 had previous experience with motor imagery only, while S01 and S07 only had experience with evoked potentials. The remaining three participants had never used a BCI before. The

subjects S03, S05 and S08 had already experienced neurofeedback.

By analysing the experimental sample as a whole, it resulted that the night before the experiment subjects slept about 7 ± 1 h. In considering the subjects' answers to the questionnaire from all sessions, about 35 % of cases reported that they did not drink coffee within the 24 h prior to the experiment. When coffee was consumed, it was drunk 4 ± 2 h earlier. No subjects drank alcohol within the 24 h prior to the experiment and only 20 % of times participants smoked a cigarette approximately an hour before starting (subjects S04 and S08).

Prior to the experiment beginning, subjects were instructed with information about the experimental protocol. The goal of the experiment was first explained, and then they were asked to try different ways of imagining hand movement (kinaesthetic sensation, squeezing a ball, grasping an object, snapping their fingers, imagining themselves or another person performing the movement) to identify the one they were most confident with. Once chosen, they were asked to keep it constant throughout the single session. Finally, they were instructed to avoid muscle and eye movements and eye blinks during the motor imagery task.

4. Experimental results

Experimental data were analysed in accordance with the previous section and the results are reported hereafter. Among the eight volunteers, four subjects participated in three sessions and four subjects participated in two sessions.

4.1. Permutation test

The time-varying accuracies generated with the original and randomly permuted labels are shown in Fig. 9 for the subject S03 (four feedback types, three sessions). In accordance with the previous discussion, they were obtained through a sliding window of width 2.00 s on the -1.00 s to 8.00 s range. Therefore, the accuracy at $t = 0.00$ s corresponds to the -1.00 s to 1.00 s window, the point in $t = 0.25$ s corresponds to the -0.75 s to 1.25 s window, and so on. This allows to span the 0.00 s to 7.00 s range as a whole with a 0.25 s step.

The different sessions are reported on rows, while the different feedback modalities are shown in columns. Per each plot, the time in seconds is reported on the x-axis, while the mean classification accuracy along with its associated standard deviation are reported in percentage on the y-axis. The blue curves correspond to the results obtained with the true labels while the red curves indicate the accuracy corresponding to the permuted labels. The two bounded lines overlap up to the cue at $t = 2\,\mathrm{s}$ as expected during the baseline period. Then, the curves are separated during the motor imagery (event related) period in most cases.

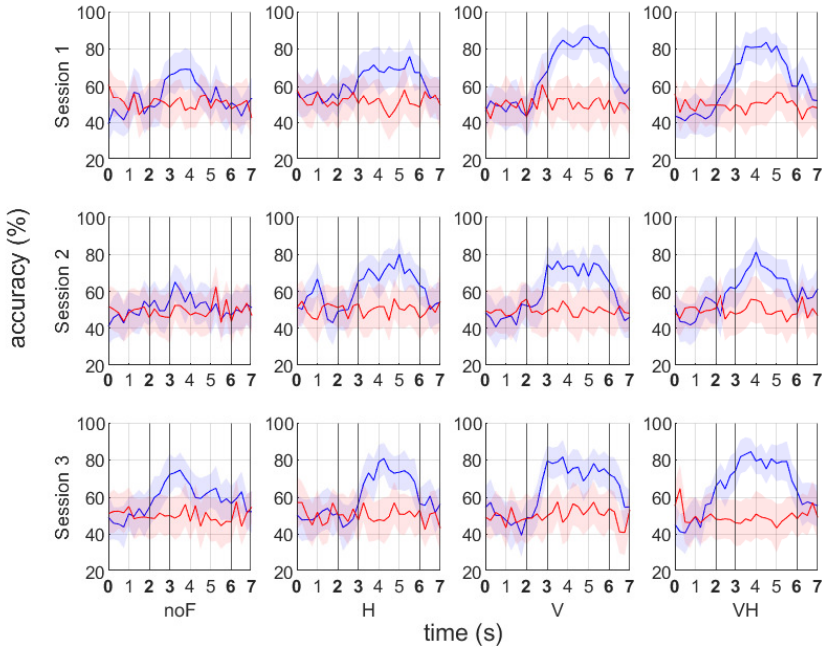


Figure 9: An example of the accuracy and permutation test accuracy for subject S03. Mean classification accuracy and associated standard deviation are calculated in time with cross-validation. The blue line corresponds to actual classification (original labels), while the red line corresponds to random classification (permuted labels). The relevant time instants are reported in bold according to the trial timing diagram of Fig. 8. noF: no feedback, H: haptic, V: visual, M: multimodal

The significance of the difference between the curves was proven with the

Wilcoxon test. The results are reported in Tab. 2 for all subjects. They refer to a test executed by considering the only motor imagery window and a 5 % significance level. A significant difference resulted in the first session for only half of the participants when no feedback was provided. However, the number of subjects associated with significant, non-random, classification rose to five, seven, and five with the haptic, visual, and multimodal feedback, respectively. Regarding the second session, statistically significant results were obtained for six subjects with no feedback, seven subjects with haptic feedback, six subjects with visual feedback, and seven subjects with multimodal feedback. Finally, for the last session, all the four subjects obtained a statistically significant result without receiving feedback, while only three out of four obtained a significant result for each of the feedbacks. These results prove the functionality of the wearable BCI and already suggest the effectiveness of providing neurofeedback during the motor imagery task. Moreover, a training effect across sessions is indicated, since a greater percentage of significant results was obtained in the first block (no feedback provided) during the second and third sessions. This is confirmed by the classification results discussed next.

	p-value											
	Session 1				Session 2				Session 3			
	noF	H	V	M	noF	H	V	M	noF	H	V	M
S01	**0.002**	0.349	**0.011**	**<0.001**	0.999	**0.004**	0.491	**<0.001**				
S02	0.967	**<0.001**	**<0.001**	**<0.001**	**0.001**	**<0.001**	**<0.001**	**<0.001**				
S03	0.076	**<0.001**	**<0.001**	**<0.001**	**0.042**	**<0.001**	**0.003**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**
S04	0.272	**0.002**	0.212	0.212	**0.013**	**0.004**	**0.004**	**0.009**				
S05	0.225	**0.040**	0.119	0.262	**<0.001**	0.586	**<0.001**	0.832	**<0.001**	**<0.001**	0.258	0.185
S06	**0.036**	**<0.001**	**<0.001**	**0.034**	**<0.001**	**<0.001**	**0.001**	**<0.001**	**<0.001**	0.094	**<0.001**	**<0.001**
S07	0.756	**<0.001**	**0.023**	**0.013**	0.791	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.01**	**<0.001**	**<0.001**
S08	0.194	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.001**	**<0.001**				

Table 2: P-values associated with the permutation test. The significant results of the Wilcoxon non-parametric test (5 % significance level) with respect to the permuted results are marked in boldface. noF: no feedback, H: haptic, V: visual, M: multimodal

4.2. Classification results

In analysing offline classification results, the best 2.00 s-wide time window was selected per each subject, session, and feedback modality. In particular, the window associated with the highest mean accuracy during the motor imagery task was selected. The results are reported in Tab. 3 in terms of mean accuracy, by reporting results in bold that were significantly different from the "no feedback" case with a significance level equal to 5 %. The statistical significance was tested by performing the ANOVA, eventually adjusted according to Sec. 3.2. The mean accuracy across the subjects is also reported along with the associated standard deviation.

	accuracy (%)											
	Session 1				Session 2				Session 3			
	noF	H	V	M	noF	H	V	M	noF	H	V	M
S01	70	60	**89**	60	57	59	**65**	64				
S02	57	**83**	**65**	**84**	67	**78**	66	**75**				
S03	70	**75**	**87**	**82**	62	**80**	**77**	**82**	74	**80**	**82**	**85**
S04	64	60	60	60	67	66	62	58				
S05	63	65	64	**68**	65	60	61	64	72	67	67	59
S06	62	61	61	57	74	75	59	69	75	59	72	65
S07	57	**68**	59	**65**	61	**80**	62	**76**	60	**83**	**75**	**70**
S08	64	**90**	**84**	**88**	76	**90**	70	**90**				
Mean	63	70	71	71	66	73	65	72	70	72	74	70
Uncertainty	2	4	5	4	2	4	2	4	4	6	3	6

Table 3: Classification accuracies using a 5-folds cross validation with 10 repetitions. The significant results from the ANOVA (5 % significance level) with respect to the no feedback block are marked in boldface. noF: no feedback, H: haptic, V: visual, M: multimodal. Recall that, in Session 3, some values are missing because those subject were only involved in two experimental sessions.

The results confirm that, thanks to the neurofeedback, the system performance improves with respect to the absence of feedback. The accuracy improvement indicates that this is especially true for the first session. Notably, also the mean accuracy among the subjects increases across sessions in the "no feedback" case, thus suggesting the training effect that was already indicated

by the permutation tests. In detail, the quantitative increase is from $(63 \pm 2)\,\%$ to $(70 \pm 4)\,\%$. However, when considering the feedback modalities, the mean accuracies do not significantly change between sessions. Finally, by considering the subjects as a whole, no significant differences emerged between the different feedback conditions.

4.3. Discussion

In validating the proposed system, the permutation test and the ANOVA applied to classification results led to compatible results. Both tests confirmed two important aspects: (i) the feedback improves system performance with respect to the absence of feedback, and (ii) a training effect subsists between different sessions. When comparing the classification accuracy associated with the "no feedback" case, it is worth noting that the results of the third session are compatible with those obtained when the feedback is provided. Such accuracy values are above $70\,\%$ in most cases, which is often considered as an empirical threshold for motor imagery-based control in BCI.

However, no mean improvement is demonstrated by the current results. This is justified by the fact that subjects show variegated performance associated with motor imagery detection. Therefore, although improvements can be highlighted in a subject-by-subject analysis, the current results do not prove a statistically significant improvement on average across the cohort. Furthermore, the results do not suggest an overall best feedback modality. This is compatible with literature findings on unimodal feedbacks [19], although the results do not provide sufficient evidence to prefer the multimodal feedback either.

It should be pointed out that the multimodal feedback occasionally appeared less effective than single feedback modalities, even when both single feedbacks improved the detection. This suggests that the multimodal feedback does not simply consist of a combination of single feedbacks, and that delivering multiple feedback simultaneously could be distracting or less engaging for the user and/or may require additional sessions to enable to the subject to gain familiarity with simultaneous feedback modality presentation. In this regard, investigating a

greater experimental sample would be desirable to better understand the differences between the proposed neurofeedback modalities. This would mean both to involve more subjects and to consider more sessions.

Questionnaire results also provide some complementary indications to the classification results. First, in motor imagery, participants preferred to imagine squeeze in 50 % of circumstances, the kinaesthetic sensation associated with touching an object for 20 % of times, grasp in 15 %, and other imagery, such as to snap the fingers or to dribble, in the remaining 15 %. Only two subjects, S05 and S07, changed the type of motor imagery between sessions. Regarding feedback, participants stated that they preferred visual and multimodal over haptic feedback. Finally, the degradation of classification performance occurring in some sessions could be correlated with the worsening in physical and mental state of the participant. As an example, if considering the "no feedback" and the haptic feedback for subject S03, there was an accuracy diminishing of about 10 % from the first to the second session, but the subject declared that he was mentally tired and bored during the second session. More motivating feedback with improved accuracy enabled through a number of additional EEG channels may indeed enhance accuracy.

4.4. Limitations

At the end of the experiments, some issues could be also pointed out. In the future, a balanced sample in terms of dominant hand should be considered because of the relevance that handedness has on motor imagery control [55]. Moreover, a single experimental session was still long enough to tire the participants indeed. Therefore, the usage of transfer learning techniques would be desirable during future developments to simultaneously reduce the time required for model calibration and to improve the classification performance by means of preliminary acquired data. Notably, improving the accuracy would also be desirable to better engage the user during motor imagery. This could be accomplished by means of a better online pipeline of EEG data analysis, e.g., by dealing with non-stationarity and lowering its computational cost [56].

After shortening the single experimental session, a greater number of subjects could be considered to investigate the correlation between the best feedback modality and subject's profile. In a daily-life application, the BCI should continuously measure the brain activity and also detect when the user is willing to act (asynchronous BCI). Moreover, as a further development, imagining both hands, or both feet, or tongue movements will be also considered, thus increasing the number of possible commands.

Finally, dry electrodes should be foreseen to enhance usability of the system, though a major problem with dry electrode may be the need to apply pressure to reduce impedance. Furthermore, this will also require effective artefact removal strategies, whose necessity was limited in the present work due to the usage of conductive gel.

5. Conclusions

In this paper, neurofeedback was applied to engage the user during motor imagery, and aimed to improve the detection of the associated neurophysiological phenomena. Visual and haptic feedback modalities were compared, and multimodal sensory feedback was explored by providing both feedback modalities simultaneously. A closed-loop wearable brain-computer interface based on the detection of motor imagery was designed and implemented by including an innovative vibrotactile chest-based feedback, wearable and portable EEG, and online signal processing.

The system was validated with an experimental campaign involving eight subjects in two or three sessions taken on different days in accordance with a standard synchronous paradigm. Results demonstrated that the feedback improves the classification with respect to the absence of feedback, even with an utmost wearable system. However, a statistically significant mean improvement was not always observed because of the limited subject sample with performance variation. Moreover, no feedback modality was generally preferable. Hence, further experiments are recommended, in which the overall system should be

enhanced by means of a more immersive feedback, better online processing, and artefacts removal strategies. Finally, more experimental sessions with more subjects would be needed especially to highlight the training effect and differences between feedbacks in more depth.

Acknowledgements

This work was carried out as part of the "ICT for Health" project, which was financially supported by the Italian Ministry of Education, University and Research (MIUR), under the initiative 'Departments of Excellence' (Italian Budget Law no. 232/2016), through an excellence grant awarded to the Department of Information Technology and Electrical Engineering of the University of Naples Federico II, Naples, Italy. DC is supported by a UKRI Turing AI Fellowship 2021-2025 funded by the EPSRC (grant number EP/V025724/1). FD is supported by the project "Free energy principle and the brain: neuronal and phylogenetic mechanisms of Bayesian inference" funded by the MIUR PRIN2020 - Grant N. 2020529PCP. The authors also thanks Leah Hudson for her proofreading of the work.

References

[1] G. Pfurtscheller and C. Neuper, "Motor imagery activates primary sensorimotor area in humans," Neuroscience letters, vol. 239, no. 2-3, pp. 65–68, 1997.

[2] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," Proceedings of the IEEE, vol. 89, no. 7, pp. 1123–1134, 2001.

[3] T. Mulder, "Motor imagery and action observation: cognitive tools for rehabilitation," Journal of neural transmission, vol. 114, no. 10, pp. 1265–1278, 2007.

[4] C. Neuper, R. Scherer, S. Wriessnegger, and G. Pfurtscheller, "Motor imagery and action observation: modulation of sensorimotor brain rhythms during mental control of a brain–computer interface," Clinical neurophysiology, vol. 120, no. 2, pp. 239–247, 2009.

[5] M. Xu, X. Xiao, Y. Wang, H. Qi, T.-P. Jung, and D. Ming, "A brain–computer interface based on miniature-event-related potentials induced by very small lateral visual stimuli," IEEE Transactions on Biomedical Engineering, vol. 65, no. 5, pp. 1166–1175, 2018.

[6] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, T. M. Vaughan et al., "Brain-computer interface technology: a review of the first international meeting," IEEE transactions on rehabilitation engineering, vol. 8, no. 2, pp. 164–173, 2000.

[7] R. Ron-Angevin, F. Velasco-Álvarez, Á. Fernández-Rodríguez, A. Díaz-Estrella, M. J. Blanca-Mena, and F. J. Vizcaíno-Martín, "Brain-computer interface application: auditory serial interface to control a two-class motor-imagery-based wheelchair," Journal of neuroengineering and rehabilitation, vol. 14, no. 1, pp. 1–16, 2017.

[8] E. Hortal, D. Planelles, A. Costa, E. Iánez, A. Úbeda, J. M. Azorín, and E. Fernández, "SVM-based brain–machine interface for controlling a robot arm through four mental tasks," Neurocomputing, vol. 151, pp. 116–121, 2015.

[9] L. Cao, B. Xia, O. Maysam, J. Li, H. Xie, and N. Birbaumer, "A synchronous motor imagery based neural physiological paradigm for brain computer interface speller," Frontiers in human neuroscience, vol. 11, p. 274, 2017.

[10] F. Lotte, J. Faller, C. Guger, Y. Renard, G. Pfurtscheller, A. Lécuyer, and R. Leeb, "Combining bci with virtual reality: towards new applications and

improved bci," in Towards Practical Brain-Computer Interfaces. Springer, 2012, pp. 197–220.

[11] D. Coyle, J. Stow, K. McCreadie, J. McElligott, and Á. Carroll, "Sensorimotor modulation assessment and brain-computer interface training in disorders of consciousness," Archives of physical medicine and rehabilitation, vol. 96, no. 3, pp. S62–S70, 2015.

[12] R. Abiri, S. Borhani, E. W. Sellers, Y. Jiang, and X. Zhao, "A comprehensive review of eeg-based brain–computer interface paradigms," Journal of neural engineering, vol. 16, no. 1, p. 011001, 2019.

[13] L. Angrisani, P. Arpaia, A. Esposito, and N. Moccaldi, "A wearable brain–computer interface instrument for augmented reality-based inspection in industry 4.0," IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 4, pp. 1530–1539, 2019.

[14] J. Jin, Z. Wang, R. Xu, C. Liu, X. Wang, and A. Cichocki, "Robust similarity measurement based on a novel time filter for SSVEPs detection," IEEE Transactions on Neural Networks and Learning Systems, 2021.

[15] A. Apicella, P. Arpaia, G. Mastrati, and N. Moccaldi, "EEG-based detection of emotional valence towards a reproducible measurement of emotions," Scientific Reports, vol. 11, no. 1, 2021.

[16] G. Prasad, P. Herman, D. Coyle, S. McDonough, and J. Crosbie, "Applying a brain-computer interface to support motor imagery practice in people with stroke for upper limb recovery: a feasibility study," Journal of neuroengineering and rehabilitation, vol. 7, no. 1, pp. 1–17, 2010.

[17] K. A. McCreadie, D. H. Coyle, and G. Prasad, "Learning to modulate sensorimotor rhythms with stereo auditory feedback for a brain-computer interface," in 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2012, pp. 6711–6714.

[18] S. Koyama, S. M. Chase, A. S. Whitford, M. Velliste, A. B. Schwartz, and R. E. Kass, "Comparison of brain–computer interface decoding algorithms in open-loop and closed-loop control," Journal of computational neuroscience, vol. 29, no. 1-2, pp. 73–87, 2010.

[19] K. A. McCreadie, D. H. Coyle, and G. Prasad, "Is sensorimotor bci performance influenced differently by mono, stereo, or 3-d auditory feedback?" IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 22, no. 3, pp. 431–440, 2014.

[20] M. Lukoyanov, S. Y. Gordleeva, A. Pimashkin, N. Grigor'ev, A. Savosenkov, A. Motailo, V. Kazantsev, and A. Y. Kaplan, "The efficiency of the brain-computer interfaces based on motor imagery with tactile and visual feedback," Human Physiology, vol. 44, no. 3, pp. 280–288, 2018.

[21] M. Fleury, G. Lioi, C. Barillot, and A. Lécuyer, "A survey on the use of haptic feedback for brain-computer interfaces and neurofeedback," Frontiers in Neuroscience, vol. 14, 2020.

[22] F. Missiroli, M. Barsotti, D. Leonardis, M. Gabardi, G. Rosati, and A. Frisoli, "Haptic stimulation for improving training of a motor imagery bci developed for a hand-exoskeleton in rehabilitation," in 2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR). IEEE, 2019, pp. 1127–1132.

[23] C. Jeunet, C. Vi, D. Spelmezan, B. N'Kaoua, F. Lotte, and S. Subramanian, "Continuous tactile feedback for motor-imagery based brain-computer interaction in a multitasking context," in IFIP Conference on Human-Computer Interaction. Springer, 2015, pp. 488–505.

[24] H. Gürkök and A. Nijholt, "Brain–computer interfaces for multimodal interaction: a survey and principles," International Journal of Human-Computer Interaction, vol. 28, no. 5, pp. 292–307, 2012.

[25] T. Sollfrank, A. Ramsay, S. Perdikis, J. Williamson, R. Murray-Smith, R. Leeb, J. Millán, and A. Kübler, "The effect of multimodal and enriched feedback on smr-bci performance," Clinical Neurophysiology, vol. 127, no. 1, pp. 490–498, 2016.

[26] Z. Wang, Y. Zhou, L. Chen, B. Gu, S. Liu, M. Xu, H. Qi, F. He, and D. Ming, "A bci based visual-haptic neurofeedback training improves cortical activations and classification performance during motor imagery," Journal of neural engineering, vol. 16, no. 6, p. 066012, 2019.

[27] L. Pillette, B. N'kaoua, R. Sabau, B. Glize, and F. Lotte, "Multi-session influence of two modalities of feedback and their order of presentation on mi-bci user training," Multimodal Technologies and Interaction, vol. 5, no. 3, p. 12, 2021.

[28] B. Ahkami and F. Ghassemi, "Adding tactile feedback and changing isi to improve bci systems' robustness: An error-related potential study," Brain topography, vol. 34, no. 4, pp. 467–477, 2021.

[29] "Vim: International vocabulary of metrology," https://www.bipm.org/en/committees/jc/jcgm/publications, last access: 2022-06-03.

[30] L. Hehenberger, L. Batistic, A. I. Sburlea, and G. R. Müller-Putz, "Directional decoding from eeg in a center-out motor imagery task with visual and vibrotactile guidance," Frontiers in human neuroscience, p. 548, 2021.

[31] R. Leeb, F. Lee, C. Keinrath, R. Scherer, H. Bischof, and G. Pfurtscheller, "Brain–computer communication: motivation, aim, and impact of exploring a virtual apartment," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 15, no. 4, pp. 473–482, 2007.

[32] M. Engin, T. Dalbastı, M. Güldüren, E. Davaslı, and E. Z. Engin, "A prototype portable system for eeg measurements," Measurement, vol. 40, no. 9-10, pp. 936–942, 2007.

[33] M. Xu, F. He, T.-P. Jung, X. Gu, and D. Ming, "Current challenges for the practical application of electroencephalography-based brain–computer interfaces," Engineering, 2021.

[34] N. Du Bois, R. Beveridge, N. McShane, T. Moore, and D. Coyle, "Signal quality assessment of a wearable electroencephalography(EEG) device built on a flexible printed circuit: FlexEEG," in 2022 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence, and Neural Engineering (MetroXRAINE). IEEE, 2022, p. in press.

[35] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," sensors, vol. 12, no. 2, pp. 1211–1279, 2012.

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

[37] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," Frontiers in neuroscience, vol. 6, p. 39, 2012.

[38] P. Arpaia, F. Donnarumma, A. Esposito, and M. Parvis, "Channel selection for optimal eeg measurement in motor imagery-based brain-computer interfaces." International Journal of Neural Systems, pp. 2 150 003–2 150 003, 2020.

[39] C. Zuo, J. Jin, R. Xu, L. Wu, C. Liu, Y. Miao, and X. Wang, "Cluster decomposing and multi-objective optimization based-ensemble learning framework for motor imagery-based brain–computer interfaces," Journal of Neural Engineering, vol. 18, no. 2, p. 026018, 2021.

[40] P. Arpaia, A. Esposito, A. Natalizio, and M. Parvis, "How to successfully classify EEG in motor imagery BCI: a metrological analysis of the state of the art," Journal of Neural Engineering, 2022.

[41] T.-j. Luo, C.-l. Zhou, and F. Chao, "Exploring spatial-frequency-sequential relationships for motor imagery classification with recurrent neural network," BMC bioinformatics, vol. 19, no. 1, pp. 1–18, 2018.

[42] A. D. Bigirimana, N. Siddique, and D. Coyle, "Emotion-inducing imagery versus motor imagery for a brain-computer interface," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 28, no. 4, pp. 850–859, 2020.

[43] H. Cho, M. Ahn, S. Ahn, M. Kwon, and S. C. Jun, "Eeg datasets for motor imagery brain–computer interface," GigaScience, vol. 6, no. 7, p. gix034, 2017.

[44] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "BCI Competition 2008–Graz data set A," Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology, vol. 16, pp. 1–6, 2008.

[45] M. Alimardani, S. Nishio, and H. Ishiguro, "Effect of biased feedback on motor imagery learning in bci-teleoperation system," Frontiers in systems neuroscience, vol. 8, p. 52, 2014.

[46] Scikit Learn, "Cross-validation: evaluating estimator performance," https://scikit-learn.org/stable/modules/cross_validation.html, last access: 2022-06-03.

[47] C. M. Jarque and A. K. Bera, "Efficient tests for normality, homoscedasticity and serial independence of regression residuals," Economics letters, vol. 6, no. 3, pp. 255–259, 1980.

[48] G. M. Sullivan and R. Feinn, "Using effect size—or why the p value is not enough," Journal of graduate medical education, vol. 4, no. 3, pp. 279–282, 2012.

[49] L. Berben, S. M. Sereika, and S. Engberg, "Effect size estimation: methods and examples," International journal of nursing studies, vol. 49, no. 8, pp. 1039–1047, 2012.

[50] P. Arpaia, D. Coyle, F. Donnarumma, A. Esposito, A. Natalizio, and M. Parvis, "Non-immersive versus immersive extended reality for motor imagery neurofeedback within a brain-computer interfaces," in International Conference on Extended Reality. Springer, 2022, pp. 407–419.

[51] University of Colorado, Colorado Springs (UCCS), "Effect size calculators," https://lbecker.uccs.edu/, last access: 2022-11-15.

[52] ClinCalc LLC, "Sample size calculator," https://clincalc.com/stats/samplesize.aspx, last access: 2022-11-15.

[53] C. H. Nguyen, G. K. Karavas, and P. Artemiadis, "Adaptive multi-degree of freedom Brain Computer Interface using online feedback: Towards novel methods and metrics of mutual adaptation between humans and machines for BCI," PloS one, vol. 14, no. 3, p. e0212620, 2019.

[54] B. Abibullaev, J. An, S. H. Lee, and J. I. Moon, "Design and evaluation of action observation and motor imagery based bcis using near-infrared spectroscopy," Measurement, vol. 98, pp. 250–261, 2017.

[55] D. Zapała, E. Zabielska-Mendyk, P. Augustynowicz, A. Cudo, M. Jaśkiewicz, M. Szewczyk, N. Kopiś, and P. Francuz, "The effects of handedness on sensorimotor rhythm desynchronization and motor-imagery BCI control," Scientific reports, vol. 10, no. 1, pp. 1–11, 2020.

[56] J. Jin, R. Xiao, I. Daly, Y. Miao, X. Wang, and A. Cichocki, "Internal feature selection method of CSP based on L1-norm and Dempster–Shafer theory," IEEE transactions on neural networks and learning systems, vol. 32, no. 11, pp. 4814–4825, 2020.