Effects of Head-Mounted Virtual Reality and Haptics in Upper-Limb Brain–Computer Interface Training

Diogo Filipe Simões Batista
diogo.batista@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisbon, Portugal

October 2022

Abstract

Brain–computer interfaces (BCIs) can provide a non-muscular channel for communication and control to patients for assistive or restorative use. Motor-imagery-based BCIs can be augmented with virtual reality (VR) and haptics to provide stroke patients with insufficient motor ability an alternative to conventional therapy. Two questions are addressed in this thesis: (1) What BCI–VR feedback configurations lead to the strongest, most lateralized brain activation in stroke rehabilitation? (2) What conditions and machine-learning algorithms lead to the most robust features and most accurate models? To achieve this, 19 healthy subjects performed motor-imagery training through five conditions with different combinations of abstract vs. realistic feedback through *NeuRow*, head-mounted display vs. monitor, and with or without haptic feedback. The power of alpha and beta rhythms following the motor tasks (event-related desynchronizations [ERDs]) and their hemispheric lateralization (lateralization indices [LIs]) were extracted for analysis. The subjects also answered questionnaires on motor-imagery ability and sense of embodiment. Seven machine-learning algorithms and several hyperparameters were tested for each condition. The results were benchmarked against motor execution. The data suggested that the use of haptic feedback and a virtual environment such as *NeuRow* lead to stronger brain activation, which could become important components in stroke rehabilitation. The support-vector classifier and multilayer perceptrons performed better but are not necessarily more adequate for stroke rehabilitation. The common spatial patterns used to train the models did not correlate significantly with the LIs for the most part, suggesting different features to be used in stroke rehabilitation.

Keywords: Brain–computer interfaces; Upper-limb stroke rehabilitation; Head-mounted virtual reality; Haptic feedback; Machine learning

1. Introduction

Stroke is a leading cause of mortality and disability worldwide [1], and its incidence is predicted to increase throughout the world as the population ages. Victims commonly lose their motor capability, which disrupts their ability to carry out their daily routines. To date, rehabilitation for stroke survivors with severe motor impairments is burdensome, since most current rehabilitation options require some volitional movement to retrain the affected limbs. However, prior research has shown that patients receive increased benefits by combining traditional therapy with emerging technologies like brain–computer interfaces (BCIs) [2] and virtual reality (VR) [3]. In particular, upper-limb rehabilitation of severely affected stroke patients comes with challenges that can be overcome through a technology-based approach.

A BCI can be described as a pattern-recognition system that utilizes the physiological activity from the brain to control external devices (e.g., a prosthesis) [4]. Although various signal-acquisition modalities can monitor brain activity, electroencephalography (EEG) is the most commonly employed, due to its relatively low cost, portability, high temporal resolution, and noninvasiveness [5]. Modulation of EEG in a closed loop can promote plastic changes in the brain, making BCIs an appealing tool for neurorehabilitation [6]. Specifically, motor-imagery-based BCIs (i.e., the subject imagines the movement of their limbs) help promote recovery from brain lesions—particularly in stroke patients [7]—by converting motor imagery into real events, such as exoskeleton [8] or avatar movement [6, 9].

It is, thus, a suitable candidate for the rehabilitation of stroke patients with a motor ability too atrophied to make use of conventional therapy. Nevertheless, the interfaces by themselves may not provide sufficiently engaging feedback to the patient,

which can be augmented with, for example, VR.

Thanks to VR technology, patients are able to interact with engaging virtual environments through a plethora of devices, be it visual, auditory, or haptic. These include screens, head-mounted displays (HMDs), video-capture systems, data gloves, hand controllers, etc. With a vast array of apparatuses that increase the patient's immersion and sense of embodiment, the ability to design engaging motor-related tasks, and superior recovery when combined with conventional therapy [3], virtual rehabilitation seems to be the natural successor to the current rehabilitation paradigm. However, the more severe cases of stroke still prohibit patients from moving and engaging with VR-based rehabilitation. Nevertheless, combining it with BCIs can provide the best of both worlds and fill the bill.

Combining motor-imagery-based BCIs and VR can improve treatment results by directly training the central nervous system [2], providing embodied feedback through avatars, and offering engaging tasks that increase adherence to the treatment. Some, but not that many, studies have tested this synergistic duo with promising results. Nevertheless, given the recency of this field of research, fundamental questions still linger. A couple of these are identified and targeted by this thesis, which hopes to shed some light on them. One such question asks what configurations (i.e., sets of equipment that provide multimodal feedback) lead to the strongest, most lateralized brain activation during motor imagery. A trade-off emerges between the added immersion and equipment cost, which adds to the complexity of the question; however, this thesis strictly compares the brain activation between different configurations.

The use of haptics is scarce and not as prevalent as VR, yet there have been studies assessing its potential in stroke rehabilitation [10]. Together with BCIs and VR, the trio is capable of providing both visual and haptic feedback through a non-muscular channel connecting the patient's motor intention to the avatar's limbs. As such, haptics are taken into consideration in the search for the best configurations.

Another issue is the use of machine-learning algorithms in BCIs, which make the translation between motor imagery and avatar movement possible. Notwithstanding the typical usage of the linear discriminant analysis (LDA) and the support-vector classifier (SVC), there is no standardized combination of algorithms and BCI–VR configurations that lead to the most accurate and robust machine-learning models. Therefore, the second fundamental question appears: what algorithms and configurations lead to the most accurate models?

Given the two questions described in the text above, the object of this thesis is, thus, twofold.

1.1. What BCI–VR configurations lead to the strongest, most lateralized brain activation?

To tackle this question, several subjects performed motor imagery in different combinations of HMD vs. non-HMD and haptic vs. non-haptic configurations, called *conditions*, through either Graz-based abstract feedback [11] or *NeuRow*'s realistic feedback [9]. A sixth condition had the subjects perform motor execution to benchmark the motor-imagery conditions against it. Their EEG signals were analyzed by interpreting a common phenomenon in motor imagery and execution, the event-related desynchronization (ERD) [12]. The spatial distributions of these desynchronizations were also analyzed through the lateralization indices (LIs).

1.2. What BCI–VR configurations and machine-learning algorithms lead to the most accurate models?

Different algorithms and hyperparameters were tested with the EEG signals recorded during the aforementioned experiment. The traditionally used LDA and SVC were included, as well as linear and nonlinear alternatives such as the multilayer perceptrons (MLPs), Gaussian naive Bayes, and the random-forest classifier. The different conditions were likewise compared across the algorithms.

2. Methods

2.1. Participant demographics

The data analysis included 19 subjects in total. The subjects had a mean age of 24.79 years (SD = 3.54 years), with the youngest being 21 years old and the oldest 36. The cohort was 68% male and 32% female. In terms of education, 16% had attended only high school, while 32% had a bachelor's degree, 42% a master's degree, and 11% a doctorate. Subjects S06 up to S18 were also asked to rate their prior experience with BCIs and VR from 1 (nonexistent) to 5 (plenty): the mean scores were 1.36 (SD = 0.67) and 1.73 (SD = 0.65), respectively. Given the nature of the motor-imagery task, some subjects (starting with S06) were asked if they had any rowing experience. At least five subjects had some, while at least four did not.

2.2. Experimental procedure

2.2.1. Conditions

The experiment consisted in having the subjects perform motor imagery of a bimanual rowing task with two individual paddles, one in each hand, under five experimental conditions. Four of these conditions used *NeuRow* [9]—a VR environment made in Unity[1] that renders a set of virtual arms from a first-person perspective—while the other condition used abstract feedback based on the

[1] unity.com

BCI-Graz paradigm [11]. A sixth condition had the subjects perform motor execution, also with the abstract feedback. These last two acted as control conditions. All six conditions and their acronyms are described below:

1. **grazMI:** The standard motor-imagery training, with a fixation cross and directional arrows on a black background guiding the subjects through the experiment.

2. **neurowMIMO:** A motor-imagery training paradigm using *NeuRow*, with a fixation cross and directional arrows overlaid on the VR environment, which was displayed through a monitor.

3. **neurowMIMOHP:** A motor-imagery training paradigm using *NeuRow*, with a fixation cross and directional arrows overlaid on the VR environment, which was displayed through a monitor. Hand controllers also provided haptic feedback through vibrotactile stimulation.

4. **neurowMIMOVR:** A motor-imagery training paradigm using *NeuRow*, with a fixation cross and directional arrows overlaid on the VR environment, which was displayed through a VR headset.

5. **neurowMIMOVRHP:** A motor-imagery training paradigm using *NeuRow*, with a fixation cross and directional arrows overlaid on the VR environment, which was displayed through a VR headset. Hand controllers also provided haptic feedback through vibrotactile stimulation.

6. **grazME:** A fixation cross and directional arrows were displayed on a black background through a monitor and guided the subjects through the experiment by having them tap their fingers on the table accordingly.

The condition order was randomized for each subject to minimize any latent effects originating from preceding conditions that could skew the results. An initial version of the sixth condition, *grazME*, had the subjects perform a circular arm motion similar to the avatar's movement in *NeuRow*. A later version had the subjects then tap their fingers on the desk, as it would be sufficient to induce a similar ERD while avoiding movement artifacts. The results from *grazME* were compared against those obtained with motor imagery in the subsequent data analysis. However, the condition was only implemented after S07's session, so only subjects S08 to S18 performed it.

2.2.2. Setup

A 32-channel EEG cap (actiCAP; Brain Products GmbH, Gilching, Germany) was used to extract the brain's electrical signals with a wireless EEG amplifier (LiveAmp; Brain Products GmbH, Gilching, Germany) with a sampling rate of 500 Hz, which included active electrodes for improving the signal-to-noise ratio. The spatial distribution of the electrodes used the 10–20 system.

Visual feedback was provided through a monitor in all conditions except in *neurowMIMOVR* and *neurowMIMOVRHP*, in which an Oculus Rift CV1 headset (Reality Labs, formerly Facebook, Inc., CA, USA) was used instead. Haptic feedback was provided through vibrotactile stimulation from Oculus Rift hand controllers.

2.2.3. Protocol

Initial questionnaires and hardware setup: The subject would begin by filling in the consent form to participate in the study. Given the informed consent, their blood pressure and resting heart rate were measured with a blood-pressure monitor. Afterward, the subject would answer the 12-item version of the Edinburgh Handedness Inventory questionnaire [13] and the Vividness of Movement Imagery Questionnaire-2 (VMIQ-2) [14]. These questionnaires are a self-assessment of, respectively, one's handedness in different tasks and the ability to perform motor imagery of different tasks in three different ways: by observing the movements through one's eyes (internal visual imagery [IVI]), by observing one's body performing the movements through an external point of view (external visual imagery [EVI]), or by feeling the movements being performed (kinesthetic imagery [KI]).

Interacting with the brain–computer interfaces: The order of the motor-imagery conditions was randomized for each subject to prevent any latent effects of previous conditions from skewing the data. Every condition except *grazME* was approximately eight minutes long, while *grazME* was around five. All of them showed a cross at the beginning of a new trial so the subject could focus on it and minimize eye movement. A left or right arrow would appear five seconds later for one second, in three consecutive trials, telling the subject to begin performing motor imagery for the targeted arm. There were 42 trials in total: 21 left- and 21 right-hand movements. The subject would stop the motor-imagery task when the cross disappeared, ending the trial (see Figure 1). After running all five motor-imagery conditions, the subjects performed a motor-execution task, *grazME*, for five minutes. In its initial version, the subjects would perform a

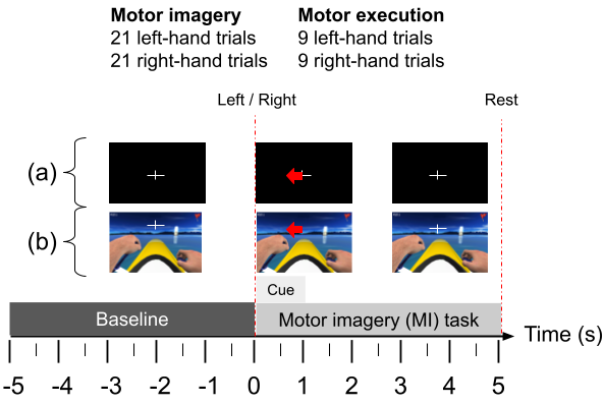


Figure 1: Experimental diagram showing the (a) Graz-based abstract feedback (*grazMI* and *grazME*) and (b) realistic feedback via *NeuRow* (*neurowMIMO*, *neurowMIMOHP*, *neurowMIMOVR*, and *neurowMIMOVRHP*)

circular arm motion similar to the avatar's movement in *NeuRow*, but the later version had the subjects tap their fingers on the desk instead, to prevent artifacts caused by the arm movements. The condition had 18 trials in total, also evenly split. The brain activity recorded during motor imagery could then be benchmarked against the activity recorded during motor execution.

Final questionnaires and hardware removal: After running through all six conditions, the equipment for the signal acquisition was removed from the subject, starting with the auxiliary signals. Lastly, the subjects answered two more questionnaires: one on condition preference and another on the sense of embodiment experienced with *NeuRow* [15].

2.3. Data analysis
2.3.1. Electroencephalography

Data preprocessing: The data was analyzed in MATLAB (R2021b and R2022a; The MathWorks, Inc., Natick, MA, USA) with the EEGLAB toolbox[2] (v2022.0; Swartz Center for Computational Neuroscience, San Diego, CA, USA).

The electrodes were re-referenced using a common average reference (CAR), which subtracts the average electrical signal of all electrodes at all time points. The sampling frequency was downsampled from 500 Hz to 125 Hz to reduce the data size. A band-pass filter from 1 Hz to 40 Hz was applied to the data to include the alpha waves (8–12 Hz) and the beta waves (13–30 Hz). The trials were epoched between five seconds before the arrow cue and five seconds after. Independent component analysis (ICA) was used to decompose the signals into several components and remove those labeled as eye and muscle movements by ICLabel[3] (Swartz Center for Computational Neuroscience, San Diego, CA, USA) with a confidence percentage between 90% and 100%. Trials with artifacts still present after the initial preprocessing were manually removed from the analysis through EEGLAB's interface.

ERD measures: The ERD values were computed as percentages of the baseline through the event-related spectral perturbation (ERSP) values, in decibels, through the formula

$$\text{ERD (\%)} = \left(10^{\text{ERSP}/10} - 1\right) \times 100, \quad (1)$$

which were negative if there was indeed an ERD, null at the baseline, or positive if it was, in fact, an event-related synchronization (ERS). For the sake of simplicity, the percentages are addressed as ERDs unless they are explicitly positive, in which case they are addressed as ERSs instead.

The mean event-related desynchronizations (mERDs) and LIs were computed from the time–frequency ERD maps of each electrode. The alpha waves' mERDs were computed between 8 Hz and 12 Hz, and the beta waves' mERDs were computed between 13 Hz and 30 Hz, from one second after the arrow cue until the end of the trial, as there was usually a delay between the cue and the ERD. The alpha and beta LIs were computed for each condition using the formula

$$\text{LI} = (\text{mERD}_{\text{left, C3}} - \text{mERD}_{\text{left, C4}}) + (\text{mERD}_{\text{right, C4}} - \text{mERD}_{\text{right, C3}}) \,/\, 2, \quad (2)$$

which is positive if the brain activation is mostly contralateral to the arm movement during motor imagery, or negative if it is ipsilateral.

The percentages of the motor-imagery mERD medians relative to *grazME*'s were also analyzed using the formula

$$\text{Relative median mERD (\%)} = \frac{\text{mERD}_{Mdn,\,\text{MI}}}{\text{mERD}_{Mdn,\,\text{ME}}} \times 100, \quad (3)$$

which is the fraction of *grazME*'s mERD median for a given motor-imagery condition.

[2] github.com/sccn/eeglab
[3] github.com/sccn/ICLabel

2.3.2. Machine learning

A Python script was written in Google Colab (adapted from David S. Batista[4]) which measures the algorithms' accuracies, as well as fitting times and precisions and recalls for left- and right-arm motor imagery. The algorithms picked for the script were the LDA, SVC, MLP, *k*-nearest neighbors (*k*-NN), Gaussian naive Bayes, random forest, and Adaptive Boosting (AdaBoost).

Algorithms and hyperparameters: Several hyperparameters were tested to determine the configurations with the highest accuracies by using the *GridSearchCV* function from Scikit-learn[5] so as to have some variety in the configurations while also keeping the script runtime within a couple of hours. The algorithms feature a mix of linear and nonlinear algorithms. For this study, the Google Colab script was adapted in Python (Python 3.9; Python Software Foundation, Wilmington, Delaware, United States) to read the EEG data obtained from the experiments and analyze the model accuracies for each algorithm under different conditions, as well as between algorithms by considering the median accuracies of a given algorithm for all conditions.

Feature extraction: The data preprocessing was different than the one done for the EEG analysis in MATLAB. A band-pass filter between 8 Hz and 30 Hz was used to include just the alpha and beta waves. Afterward, a common spatial pattern (CSP) filter with four components was used, which is the most commonly used feature-extraction tool in BCIs [16]. The data was split into epochs of left- and right-hand trials. For the accuracies, which were offline, 20% of the data was divided into 10 groups of shuffled epochs to b used in all the algorithms.

2.4. Statistical tests

Because the sample size was small—ranging from 10 to 19 subjects, depending on the condition—the Kruskal–Wallis test, which is the nonparametric equivalent of the one-way analysis of variance (ANOVA), was used to determine statistically significant differences between the conditions in the mERD and LI sample groups, and conditions or algorithms in the machine-learning accuracy sample groups, for a significance level of 0.05 ($p < 0.05$). The LI sample groups were also compared to a null lateralization index (LI = 0) to determine if the brain activation was significantly lateralized in any conditions.

[4] davidsbatista.net/blog/2018/02/23/model_optimization
[5] scikit-learn.org

Whenever the null hypothesis was rejected—that the samples from some conditions or algorithms did not come from the same distribution—a post hoc analysis was performed. The analysis consisted of pairwise comparisons using Dunn's test, as it typically follows the Kruskal–Wallis test due to computing the same ranks.

2.4.1. Questionnaire correlations

The VMIQ-2 was used to see whether there were statistically significant ($p < 0.05$) correlations between the answers to its items and the mERDs and LIs obtained for left- and right-hand trials, alpha and beta waves, electrodes C3 and C4, and each condition. The significant correlations were put on a table, where the rows and columns were labeled with numbers to make the table easier to read.

Additionally, the EEG metrics and the embodiment scores were also plotted in scatter plots to better observe any statistically significant correlations between the different conditions.

3. Results

The results were analyzed in an attempt to (1) answer which condition, or BCI–VR configuration, induces the strongest, most lateralized ERD (i.e., brain activation) for stroke rehabilitation and (2) which conditions and machine-learning algorithms lead to the most accurate models in BCIs, also for stroke rehabilitation.

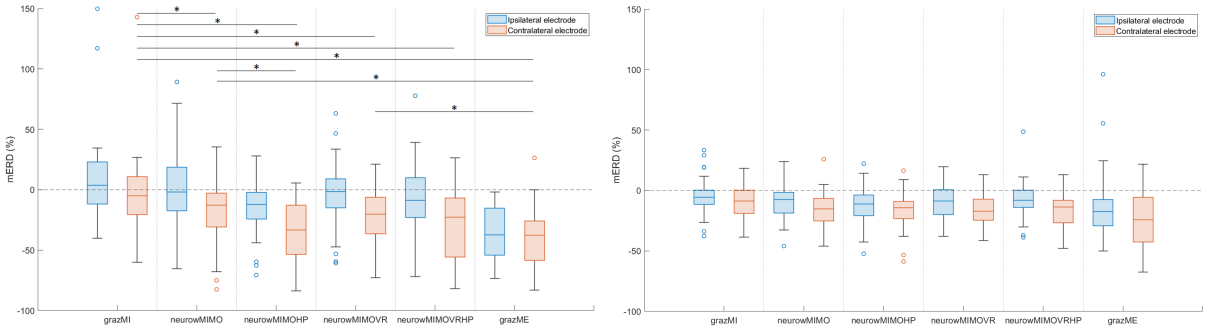
The first results pertain to the EEG metrics, which are the alpha and beta mERDs and LIs. Then, the accuracies of the machine-learning models are shown, and correlations between them and the LIs are mentioned. Finally, questionnaire scores are presented, and correlations between them and EEG metrics are also mentioned.

3.1. Event-related desynchronizations

3.1.1. Power differences

The box plots of the mERDs in electrodes C3 and C4 are shown in Figure 2, which reveal the cortical activation for the different conditions. The mERD medians were all negative except *grazMI*'s ipsilateral alpha mERD (*Mdn* = 3.57%). The condition *grazME* had the strongest mERD medians of all conditions. The contralateral sample groups had overall lower distributions than the ipsilateral ones.

From the ipsilateral alpha mERD medians, *grazME* had the highest value (*Mdn* = −37.45%), followed by *neurowMIMOHP* (*Mdn* = −12.12%), *neurowMIMOVRHP* (*Mdn* = −8.69%), *neurowMIMO* (*Mdn* = −1.79%), *neurowMIMOVR* (*Mdn* = −1.48%), and *grazMI* (*Mdn* = 3.57%). Between the contralateral medians, *grazME* had the highest median (*Mdn* = −37.27%), fol-



(a) Alpha mERDs

(b) Beta mERDs

Figure 2: mERDs for the different conditions, where each sample group contains both left- and right-hand trials for both (a) alpha and (b) beta bands. The asterisks indicate significant differences between the distributions of the contralateral electrodes. ($p < 0.05$)

lowed by *neurowMIMOHP* (*Mdn* = −33.30%), *neurowMIMOVRHP* (*Mdn* = −22.86%), *neurowMIMOVR* (*Mdn* = −20.16%), *neurowMIMO* (*Mdn* = −12.92%), and *grazMI* (*Mdn* = −5.08%) (see Figure 2(a)).

From the ipsilateral beta mERD medians, *grazME* had the highest value (*Mdn* = −17.30%), followed by *neurowMIMOHP* (*Mdn* = −11.08%), *neurowMIMOVR* (*Mdn* = −8.70%), *neurowMIMOVRHP* (*Mdn* = −8.05%), *neurowMIMO* (*Mdn* = −7.50%), and *grazMI* (*Mdn* = −5.47%). Between the contralateral medians, *grazME* had the highest median (*Mdn* = −24.34%), followed by *neurowMIMOVR* (*Mdn* = −17.07%), *neurowMIMO* (*Mdn* = −15.39%), *neurowMIMOHP* (*Mdn* = −14.29%), *neurowMIMOVRHP* (*Mdn* = −13.76%), and *grazMI* (*Mdn* = −8.69%) (see Figure 2(b)).

According to the Kruskal–Wallis test, the alpha mERD was significantly different across all groups (ipsilateral: χ^2 = 33.92, $p < 0.001$; contralateral: χ^2 = 27.54, $p < 0.001$) but there were no significant differences between the beta mERDs (ipsilateral: χ^2 = 10.75, p = 0.057; contralateral: χ^2 = 10.30, p = 0.067). Specifically, according to Dunn's test for the post-hoc pairwise comparisons ($p < 0.05$), *grazMI* was significantly different from all the other conditions for contralateral electrodes (*neurowMIMO*: p = 0.043; *neurowMIMOHP*: $p < 0.001$; *neurowMIMOVR*: p = 0.012; *neurowMIMOVRHP*: $p < 0.001$; *grazME*: $p < 0.001$), while *grazME* was significantly different from *grazMI* ($p < 0.001$) and the *NeuRow* conditions without haptic feedback, *neurowMIMO* (p = 0.009) and *neurowMIMOVR* (p = 0.033). On the other hand, the *NeuRow* conditions with haptic feedback were not significantly different from *grazME* for the contralateral electrodes (*neurowMIMOHP*: p = 0.480; *neurowMIMOVRHP*: p = 0.153). The condition *grazME* also had significantly different ipsilateral mERD distributions from all the other conditions (*neurowMIMOHP*: p = 0.021; the others: $p < 0.001$).

3.1.2. Lateralization indices

The box plots of the LI sample groups are shown in Figure 3, which indicate the lateralization of the ERDs in each condition. The alpha LIs had mostly broader distributions and higher values than the beta LIs. The median alpha LI of *neurowMIMOVR* was the highest (*Mdn* = 15.2), followed by *neurowMIMOHP* (*Mdn* = 14.1), *neurowMIMOVRHP* (*Mdn* = 13.7), and *neurowMIMO* (*Mdn* = 10.8). The conditions *grazMI* and *grazME* had the lowest medians (*Mdn* = 5.3 and *Mdn* = 7.3, respectively).

The condition *grazME* had the highest median beta LI (*Mdn* = 8.6) but also the broadest distribution. The motor-imagery conditions had similar distributions between themselves, with the conditions *neurowMIMOHP* and *neurowMIMOVRHP* having the highest medians (*Mdn* = 5.1 and *Mdn* = 5.4, respectively). However, all conditions included negative indices in their sample groups, which are found up to the lower quartile of the distributions.

According to the Kruskal–Wallis test ($p < 0.05$), none of the sample groups were significantly different (alpha: χ^2 = 6.06, p = 0.300); beta: χ^2 = 3.77, p = 0.582). However, there was a significant difference between the sample groups and a null LI, LI = 0 (alpha: χ^2 = 28.17, $p < 0.001$; beta: χ^2 = 20.03, p = 0.003). Dunn's test for pairwise comparisons found *grazMI* and the *NeuRow* conditions to have significantly different alpha LI sample groups from the null LI (*grazMI*: p = 0.010; *NeuRow* conditions: $p < 0.001$; *grazME*: p = 0.057), while all the beta LI sample groups were significantly different (*grazMI*: p = 0.006; *neurowMIMO* and *neurowMIMOHP*: p = 0.005; *neurowMIMOVR*: p = 0.003; *neurowMIMOVRHP* and *grazME*: $p < 0.001$).

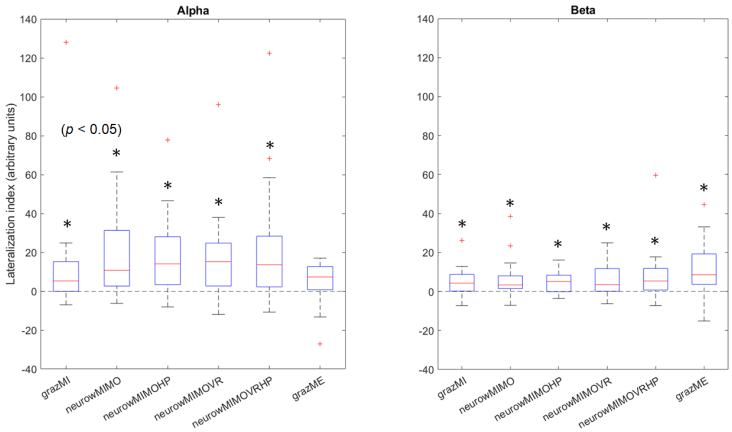


Figure 3: Box plots of the LIs for alpha and beta waves, where the asterisks show the distributions that are significantly different from the null LI, LI = 0 ($p < 0.05$)

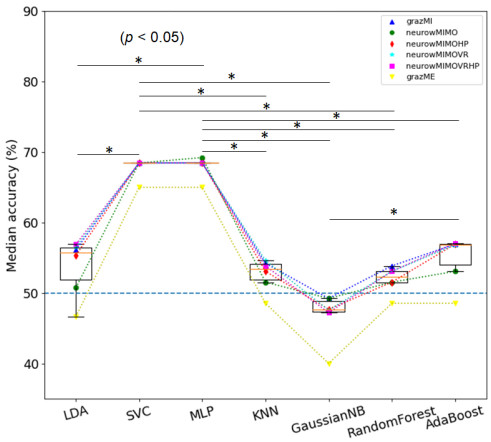


Figure 4: Box plots of the machine-learning models' median accuracies for all conditions, where the asterisk and horizontal lines indicate statistically significant pairwise sample group differences, for $p < 0.05$

3.2. Machine learning

The median accuracies of each model are shown in Figure 4. The algorithms SVC and MLP had the highest median-accuracy medians (*Mdn* = 68.46%), followed by AdaBoost (*Mdn* = 56.88%), LDA (*Mdn* = 55.77%), *k*-NN (*Mdn* = 53.46%), and Gaussian naive Bayes (*Mdn* = 47.63%). The algorithms SVC and MLP had the most compact distributions with a mean accuracy of 67.88% (SD = 1.29%) and 68.01% (SD = 1.38%), respectively. The median accuracies were identical for both algorithms except in *neurowMIMO*, which was 68.46% for SVC and 69.23% for MLP. The Gaussian naive Bayes was the only algorithm with all of its median accuracies below 50%.

The Kruskal–Wallis test showed a significant difference ($p < 0.05$) between the accuracy distributions ($\chi^2 = 32.08$, $p < 0.001$). The pairwise comparisons using Dunn's test further showed that LDA was significantly different from SVC ($p = 0.014$) and MLP ($p = 0.010$); SVC was significantly different from *k*-NN ($p = 0.006$), Gaussian naive Bayes ($p < 0.001$), and random forests ($p = 0.002$); MLP was significantly different from *k*-NN ($p = 0.004$), Gaussian naive Bayes ($p < 0.001$), random forests ($p = 0.001$), and AdaBoost ($p = 0.040$); and Gaussian naive Bayes was significantly different from AdaBoost ($p = 0.015$).

4. Discussion

4.1. What BCI–VR configurations lead to the strongest, most lateralized brain activation?

4.1.1. Strongest event-related desynchronizations

From all the motor-imagery conditions, the ones that use the VR environment (*neurowMIMO*, *neurowMIMOHP*, *neurowMIMOVR*, and *neurowMIMOVRHP*) led to significantly stronger alpha ERDs. That is, the left side of the sensorimotor cortex in right-hand trials and the right side in left-hand trials had stronger alpha desynchronizations after the subject began to perform motor imagery. However, the *NeuRow* conditions that use vibrotactile stimulation as haptic feedback (*neurowMIMOHP* and *neurowMIMOVRHP*) produced the strongest alpha ERDs, which are comparable with motor execution's equivalent (*grazME*). The abstract-feedback-only condition (*grazMI*) performed significantly worse than the *NeuRow* conditions, thus suggesting that *NeuRow* and haptic feedback, together or separately, lead to stronger alpha ERDs. Furthermore, there were no significant differences between *neurowMIMOHP* and *neurowMIMOVRHP*, which implies that, while having haptic feedback, the use of a VR headset does not lead to a significantly stronger desynchronization.

The alpha waves were much more reactive to the different configurations than the beta waves, as there were many more significant differences between the alpha ERDs than the beta counterparts. Nevertheless, motor execution led to a slightly,

though not significantly, stronger beta ERD than motor imagery. The ERD in beta waves has been observed shortly after performing both motor imagery or motor execution [17].

4.1.2. Most lateralized event-related desynchronizations

The LIs were not significantly different between each condition. However, they were significantly different from the null LI (LI = 0), except *grazME*'s alpha LIs. This is ideal in stroke rehabilitation, as the patient's brain activation should be similar to healthy subjects, who have a mostly contralateral alpha activation [18]. Although the motor-imagery conditions led to significantly contralateral brain activation, no significant differences were found between them. Thus, all the conditions may be capable of inducing similar lateralization. Nevertheless, there was a slight increase in the median LIs for the conditions that include a VR headset and haptic feedback, separately or together, which may become significantly different from the other two motor-imagery conditions, *grazMI* and *neurowMIMO*, if a larger sample is gathered.

The *grazME*'s alpha LIs were not significantly different from the null LI, which is in accordance with upper-alpha and lower-beta ERDs being observed to become bilaterally symmetrical shortly before motor execution [17]. Nevertheless, *grazME*'s beta LIs were significantly contralateral, which were computed for both lower and upper beta bands.

4.2. What BCI–VR configurations and machine-learning algorithms lead to the most accurate models?

4.2.1. BCI–VR configurations

In all algorithms, *grazME* performed significantly worse than the motor-imagery conditions, which can be attributed to the fact that CSPs are not usually used for motor execution but prominently implemented in motor-imagery tasks [16]. There were some significant differences between the motor-imagery conditions for some of the algorithms, but these differences varied between them without a clear trend.

As such, the data does not suggest any conditions to perform consistently better for any of the algorithms. Nevertheless, abstract feedback coupled with vibrotactile stimulation has recently been reported to improve classification accuracy compared to abstract feedback only [19]. Moreover, functional electrical stimulation (FES), another type of haptic feedback, has also been used together with a VR headset with an improved classification percentage compared to using just the headset [20].

4.2.2. Machine-learning algorithms

Despite the LDA being commonly used in motor-imagery-based BCIs [21], it performed worse than the SVC and MLPs. The peculiar accuracy distribution of these two algorithms could require further analysis; for example, by looking at their hyperparameters. The use of nonlinear algorithms in BCIs is not commonly seen, as the LDA or SVC with a linear kernel are serviceable for online performance, despite attempts at implementing, for example, convoluted neural networks (CNNs) [22].

Although the accuracies were offline, most of them were close to chance level (i.e., in a binary classifier that detects left- or right-hand movements). The analysis of the models' performances should also consider the fitting times, which is an important factor in online performance, but it was not analyzed in this study. Thus, the machine-learning aspect requires a more thorough investigation that involves online accuracies, fitting times, and possibly other metrics, as well as the analysis of the hyperparameters.

4.3. Limitations

The most pronounced limitation of this study is that the BCIs were not closed loops; that is, the subjects do not control the avatar in *NeuRow*. Rather, the virtual environment assisted the subjects in performing motor imagery, by providing visual aid and vibrotactile stimulation. In stroke rehabilitation, the object of this study, using the more interactive closed-loop BCIs could prove to be more beneficial than displaying the avatar to the patient without any possible input from the patient. While *NeuRow* is capable of closing the loop, the experiment would have taken significantly longer to conduct with all of the conditions, due to the time needed to train the machine-learning models of the BCIs, as well as time constraints. Therefore, the study did not use closed-loop BCIs.

Another limitation of the study was the use of healthy subjects, given that the questions posed in this thesis pertain to stroke patients. Nevertheless, it would have been difficult to gather a meaningful number of stroke patients willing and capable of participating in the study, given the time constraints. Not only that, but the BCIs under study are mainly for stroke patients who have an inadequate motor ability for rehabilitation that requires arm movement, thus making the inclusion criteria even more selective. Therefore, the subjects were all healthy individuals, most of them in their 20s, who had not suffered a stroke before.

The small sample was also a limitation, which also fluctuated between the different conditions. While *neurowMIMO* and *neurowMIMOHP* had 19 runs, or subjects, *grazMI*, *neurowMIMOVR*, and

neurowMIMOVRHP had 18 runs, and *grazME* only 10. As an immediate consequence, the conclusions obtained from the statistical tests were weaker; however, they could signal trends in the data that could be recognized and drawn conclusions from. Another consequence was the lack of demographic comparisons, which would have been too skewed to carry out.

The machine-learning analysis was not close to being exhaustive, due to time constraints, and only a select number of hyperparameters were tested. The analysis of the machine-learning algorithms and conditions that lead to the more robust and accurate models was, therefore, carried out in a broad, but not as deep, search. Not only that, but the accuracies analyzed were offline, as the BCIs did not let the subjects control the avatar in *NeuRow*.

4.4. Future research

Future studies should have a sample criterion that includes stroke patients with poor motor ability instead of healthy participants. They should also test different training features besides standard CSPs, so as to find alternatives that lead to not only accurate classifications in BCIs but also strong contralateral ERDs in the sensorimotor cortex.

Being a new field, the search for the best BCI–VR still requires studies of similar nature (i.e., that test different configurations), albeit with larger samples, the inclusion of stroke patients, and closed-loop BCIs. Haptic feedback, in particular, has seen limited use in research [10], despite the promising results obtained in this study.

5. Conclusions

Based on the acquisition of EEG signals and an analysis of the alpha and beta ERDs and LIs, the use of a virtual environment, *NeuRow*, and haptic feedback—implemented as vibrotactile stimulation in this study—led to significantly stronger contralateral ERDs, which were comparable to motor execution. Furthermore, the VR headset did not lead to such results by itself, being comparable to just using a monitor without haptic feedback. All motor-imagery conditions invoked similarly contralateral desynchronizations.

The SVC and MLPs had the most accurate models by a significant margin but were not influenced by the conditions like the other algorithms such as the LDA and AdaBoost. Moreover, none of the conditions stood out in having more accurate models. Interestingly, however, were the weak correlations of the LIs and the median accuracies across conditions, which, by proxy, relate to the features used to train the models, the CSPs. Given the purpose of the BCI–VR systems, the features should be intimately connected with the brain activation, since the neurofeedback offered to the stroke patients should not only lead to accurate models but also strong contralateral ERDs.

Additionally, the questionnaires, despite arguably not showing interesting results besides the condition preference, were important to bridge the gap between the subjectivity of the subjects' perceived immersion and motor-imagery ability, as well as their preference for certain configurations, and the objectivity of the EEG analysis concerning their brain activation.

There was an attempt in figuring out the configurations and algorithms that led to the strongest brain activation and most accurate models, respectively, but there were several limitations. The absence of stroke patients; the open-loop BCI, as opposed to a closed-loop one; a small sample size; and, by any means, a non-exhaustive machine-learning analysis were some of them. Nevertheless, the promising results pertaining to haptic feedback contributed toward understanding this new field of research better, that of the BCI–VR systems for stroke rehabilitation, which hope to offer a good therapy option to patients who cannot take part in conventional or virtual rehabilitation.

References

[1] GBD 2019 Diseases and Injuries Collaborators, "Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019," *The Lancet*, vol. 396, pp. 1204–1222, October 2020.

[2] K. K. Ang and C. Guan, "Brain–computer interface in stroke rehabilitation," *Journal of Computing Science and Engineering*, vol. 7, pp. 139–146, June 2013.

[3] A. Aminov, J. M. Rogers, S. Middleton, K. Caeyenberghs, and P. H. Wilson, "What do randomized controlled trials say about virtual rehabilitation in stroke? A systematic literature review and meta-analysis of upper-limb and cognitive outcomes," *Journal of NeuroEngineering and Rehabilitation*, vol. 15, March 2018.

[4] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, pp. 767–791, June 2002.

[5] M. L. Martini, E. K. Oermann, N. L. Opie, F. Panov, T. Oxley, and K. Yaeger, "Sensor modalities for brain–computer interface technology: A comprehensive literature review,"

Neurosurgery, vol. 86, pp. 108–117, February 2020.

[6] F. Pichiorri, G. Morone, M. Petti, J. Toppi, I. Pisotta, M. Molinari, S. Paolucci, M. Inghilleri, L. Astolfi, F. Cincotti, and D. Mattia, "Brain–computer interface boosts motor imagery practice during stroke recovery," *Annals of Neurology*, vol. 77, pp. 851–865, 2015.

[7] J. J. Daly and J. R. Wolpaw, "Brain–computer interfaces in neurological rehabilitation," *The Lancet Neurology*, vol. 7, pp. 1032–1043, 2008.

[8] M. Ortiz, E. Iáñez, J. L. Contreras-Vidal, and J. M. Azorín, "Analysis of the EEG rhythms based on the empirical mode decomposition during motor imagery when using a lower-limb exoskeleton: A case study," *Frontiers in Neurorobotics*, vol. 14, pp. 1–13, August 2020.

[9] A. Vourvopoulos, C. Jorge, R. Abreu, P. Figueiredo, J.-C. Fernandes, and S. B. i Badia, "Efficacy and brain imaging correlates of an immersive motor imagery BCI-driven VR system for upper limb motor rehabilitation: A clinical case report," *Frontiers in Human Neuroscience*, vol. 13, July 2019.

[10] 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, June 2020.

[11] G. Pfurtscheller, C. Neuper, G. R. Muller, B. Obermaier, G. Krausz, A. Schlögl, R. Scherer, B. Graimann, C. Keinrath, D. Skliris, M. Wortz, G. Supp, and C. Schrank, "Graz-BCI: state of the art and clinical applications," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, pp. 177–180, 2003.

[12] G. Pfurtscheller and F. H. L. da Silva, in *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, 6th ed.

[13] R. C. Oldfield, "The assessment and analysis of handedness: The Edinburgh inventory," *Neuropsychologia*, vol. 9, pp. 97–113, March 1971.

[14] R. Roberts, N. Callow, L. Hardy, D. Markland, and J. Bringer, "Movement imagery ability: Development and assessment of a revised version of the vividness of movement imagery questionnaire," *Journal of Sport & Exercise Psychology*, vol. 30, pp. 200–221, 2008.

[15] T. C. Peck and M. Gonzalez-Franco, "Avatar embodiment. A standardized questionnaire," *Frontiers in Virtual Reality*, vol. 1, February 2021.

[16] C. S. Nam, A. Nijholt, and F. Lotte, "Introduction: Evolution of brain–computer interfaces," in *Brain–Computer Interfaces Handbook: Technological and Theoretical Advances*. CRC Press, August 2019, p. 3.

[17] G. Pfurtscheller and F. H. L. da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clinical Neurophysiology*, vol. 110, pp. 1842–1857, 1999.

[18] J. E. Downey, K. M. Quick, N. Schwed, J. M. Weiss, G. F. Wittenberg, M. L. Boninger, and J. L. Collinger, "The motor cortex has independent representations for ipsilateral and contralateral arm movements but correlated representations for grasping," *Cerebral Cortex*, vol. 30, pp. 5040–5409, October 2020.

[19] Y. Zhong, L. Yao, J. Wang, and Y. Wang, "Tactile sensation assisted motor imagery training for enhanced BCI performance: A randomized controlled study," *IEEE Transactions on Biomedical Engineering*, pp. 1–9, August 2022.

[20] D. Achanccaray, S.-I. Izumi, and M. Hayashibe, "Visual-electrotactile stimulation feedback to improve immersive brain–computer interface based on hand motor imagery," *Hindawi*, vol. 2021, February 2021.

[21] C. S. Nam, A. Nijholt, and F. Lotte, "A step-by-step tutorial for a motor imagery-based BCI," in *Brain–Computer Interfaces Handbook: Technological and Theoretical Advances*. CRC Press, August 2019, p. 454.

[22] T. Karácsony, J. P. Hansen, H. K. Iversen, and S. Puthusserypady, "Brain computer interface for neuro-rehabilitation with deep learning classification and virtual reality feedback," in *ACM International Conference Proceeding Series*, Reims, France, March 2019.