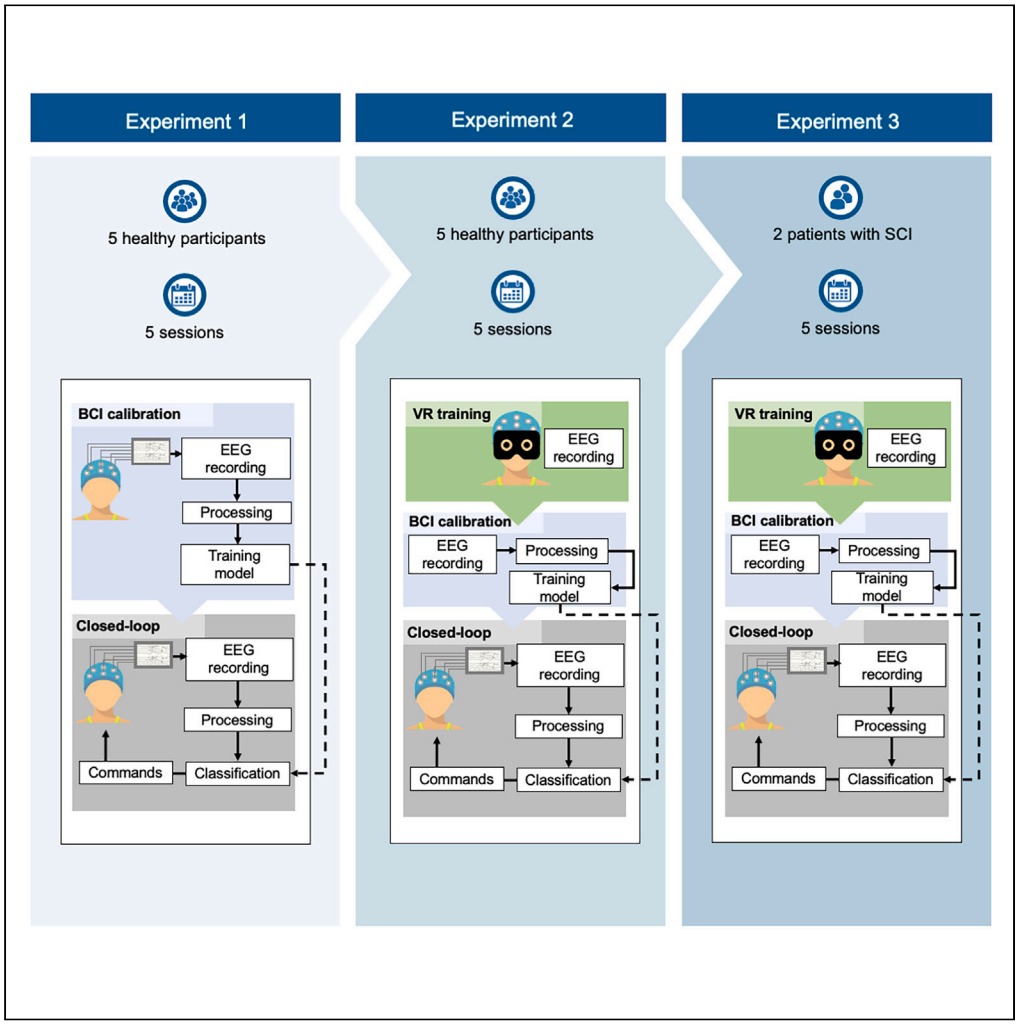
iScience

CellPress
OPEN ACCESS

Article

# Brain-computer interface enhanced by virtual reality training for controlling a lower limb exoskeleton




Laura Ferrero, Vicente Quiles, Mario Ortiz, Eduardo Iáñez, Ángel Gil-Agudo, José M. Azorín

lferrero@umh.es

## Highlights

Virtual reality training system to enhance subjects' ability to perform motor imagery

Speed up brain-computer interfaces (BCIs) training

Closed-loop experiments for controlling a lower-limb exoskeleton by means of BCI

Assessment of the BCI to control an exoskeleton with patients with spinal cord injury

Ferrero et al., iScience 26, 106675
May 19, 2023 
https://doi.org/10.1016/j.isci.2023.106675

iScience

CellPress
OPEN ACCESS

Article

# Brain-computer interface enhanced by virtual reality training for controlling a lower limb exoskeleton

Laura Ferrero,[1,2,4,*] Vicente Quiles,[1,2] Mario Ortiz,[1,2,4] Eduardo Iáñez,[1,2] Ángel Gil-Agudo,[5] and José M. Azorín[1,2,3,4,6]

## SUMMARY

**This study explores the use of a brain-computer interface (BCI) based on motor imagery (MI) for the control of a lower limb exoskeleton to aid in motor recovery after a neural injury. The BCI was evaluated in ten able-bodied subjects and two patients with spinal cord injuries. Five able-bodied subjects underwent a virtual reality (VR) training session to accelerate training with the BCI. Results from this group were compared with a control group of five able-bodied subjects, and it was found that the employment of shorter training by VR did not reduce the effectiveness of the BCI and even improved it in some cases. Patients gave positive feedback about the system and were able to handle experimental sessions without reaching high levels of physical and mental exertion. These results are promising for the inclusion of BCI in rehabilitation programs, and future research should investigate the potential of the MI-based BCI system.**

## INTRODUCTION

Robotic orthoses and exoskeletons have emerged as wearable devices with the potential to enhance physical performance and assist locomotion.[1] Their inclusion in rehabilitation programs has shown to promote the recovery of motor function, particularly among patients who have experienced a stroke or a spinal cord injury (SCI).[2] These debilitating injuries can significantly impact patients' quality of life and autonomy, underscoring the importance of robotic exoskeletons in facilitating their recovery and ability to live independently.

The interfaces used to control these devices typically rely on a combination of mechanical and electrical instruments. For instance, the H3 exoskeleton (Technaid, Spain) is controlled through commands sent via a smartphone application, whereas the Rex exoskeleton (Rex Bionics, New Zealand) is operated via a joystick.[3]

BCIs provide a direct and intuitive mean of communication between subjects and their devices, eliminating the need for external controls.[3,4] These systems record brain activity and translate it into control commands for an output device. Two types of BCI exist based on the control paradigm used: synchronous and asynchronous.

Synchronous BCI rely on external cues to guide the user. Some of them are based on various external stimuli and evoked potentials. Among these, two visually evoked potentials – P300 and steady-state visual evoked potentials (SSVEP)[5,6]- are the most commonly employed. Another type of synchronous BCI uses oscillatory EEG patterns that subjects modulate using specific mental strategies after a cue. An example of such a BCI is motor imagery (MI), where subjects simulate a given action without physically executing it.[7] Brain activity is only analyzed during predefined time intervals, and a cue indicates the beginning and the end of each interval.

In contrast, asynchronous BCI continuously record and analyze brain signals without any temporal constraints. Subjects internally control their brain switches to produce the desired commands for the output device. However, the primary limitation of these BCI relies on its validation. The system either instructs subjects when to switch, or they report themselves changes in their mental procedures.[8]

[1]Brain-Machine Interface System Lab, Miguel Hernández University of Elche, Elche, Spain

[2]Instituto de Investigación en Ingeniería de Elche-I3E, Miguel Hernández University of Elche, Elche, Spain

[3]Valencian Graduate School and Research Network of Artificial Intelligence (valgrAI), Valencia, Spain

[4]The European University of Brain and Technology (NeurotechEU)

[5]Hospital Nacional de Parapléjicos de Toledo, Toledo, Spain

[6]Lead contact

*Correspondence: lferrero@umh.es

https://doi.org/10.1016/j.isci.2023.106675

Recent studies have demonstrated that the practice of motor imagery combined with movement-associated feedback may facilitate motor recovery in patients who have suffered neural injuries.[9] This effect is believed to be mediated by the brain's neuroplasticity, which refers to the nervous system's ability to adapt and recover by reorganizing its neural pathways. The performance of mental motor imagery with motion feedback has shown to promote this property.[10,11]

In the field of BCI, electroencephalography (EEG) is commonly used to monitor brain activity because of its high temporal resolution, non-invasive recording, and portability. EEG primarily measures the activity of pyramidal neurons that are oriented in the same direction with respect to the cortical surface, allowing one to detect their signals without cancellation. However, EEG has some limitations, as the electrical signal from these neurons is attenuated as it travels through the dura, skull, and scalp before reaching the recording electrode.[2,12]

There are more works in the literature that have worked with EEG based on upper limb MI[13–18] than on lower-limb MI. The main limitation of lower-limb MI is that the leg area of the motor cortex is deeply located, around 1–4 cm from the surface. Therefore, EEG cannot accurately record this activity and it can be highly influenced by activity occurring near the surface of the skull.[19] There are some authors that have tried to discriminate leg MI with respect to other types of MI that can be more easily identified by EEG as left-hand, right-hand and tongue MI and results obtained were promising.[20] However, considering a BCI for rehabilitation, it would be interesting to have a non-intention condition in which subjects do not have to perform MI of another task to pause the system.[21] There are previous studies that have assessed this issue having leg MI versus idle state.[9,22–29] During idle state subjects must be relaxed. Accordingly, MI acts as the active condition and idle state as the passive one. Nevertheless, the performance metrics of these approaches are worse than the ones that employ upper-limb MI.

Although there are many studies based on MI BCI, the closed-loop control of an external device remains briefly addressed in the literature, especially in the case of lower-limb MI. Most of the previous studies performed an offline approach in which the subjects first perform certain brain patterns and once they finish the experiment, data are classified and analyzed. This means that no real-time feedback was provided to the subject during the experimentation, working the BCI in an opened-loop control of the external device. Real-time feedback is crucial for subject's rehabilitation.[20,22–24]

Instead of working with MI tasks, some authors have designed experimental approaches for a closed-loop control based on motion intention. Motion intention is a cortical potential produced by the motor cortex just before a movement is performed. In [30] and [31] a patient with spinal cord injury (SCI) walked assisted by an exoskeleton that was triggered by an external operator while their EEG was recorded. Then, motion intention was decoded and used to start/stop the exoskeleton. Moreover, in[32,33] four patients with SCI were instructed to attempt to initiate a movement while they were standing and wearing an exoskeleton with their legs blocked. Then, for the real control, patients were asked to perform the same attempt when they wanted to initiate the gait. To ensure safety during the experiment, the participants were only allowed to have complete and unrestricted control over the exoskeleton during specific time intervals, whereas the rest of the time the exoskeleton was blocked. However, the main benefit of MI in comparison with motion intention is that it requires the subjects to mentally repeat a certain movement and they have to be cognitively involved during the duration of the task which promotes motor recovery.[34]

Barria et al.[9] combined an MI with visual and/or haptic feedback to enhance the performance of the model, whereas Choi et al.[28] included eye blinking in the control loop. Do et al.[29] developed a BCI based purely on MI of the legs for controlling a treadmill suspended robotic gait orthosis and it was evaluated with an able-bodied individual and a patient with SCI. Rodríguez-Ugarte et al.[25] presented an experimental design for controlling the start of the gait of a robotic exoskeleton based on MI of the gait during determined time intervals. They employed transcranial direct current stimulation (tDCS) for improving the practice of MI and it was evaluated with 12 able-bodied subjects. In our previous studies,[26,27] we presented two different BCI based on MI of the gait adding the level of attention of the user in the control paradigm. Therefore, a new command was sent to the exoskeleton only if the attention level to the gait of the user was above a threshold. Whereas the inclusion of this second rule reduced the number of false positives, the sensibility of the model decreased significantly.

**Table 1. Patients' details**

| Subject | P1 | P2 |
|---|---|---|
| Gender | Male | Male |
| Age | 51 | 62 |
| Weigth | 75 kg | 71 kg |
| Injury level | T4 | L3 |
| ASIA injury assessment | C | B |
| Injury type | Incomplete | Incomplete |

One of the primary challenges faced by a BCI is its calibration. During the calibration process, the subject performs a series of mental tasks, such as imagining movements and the BCI records the associated brain activity. These data are then used to train the BCI to recognize the patterns of brain activity associated with specific actions. However, this calibration process can be time-consuming and fatiguing for the subjects, particularly for those with impairments who may require shorter session scenarios.[2,35]

Typically, calibration is done at the beginning of each experimental session because of the non-stationary nature of EEG data. However, researchers have been exploring calibration-free approaches using subject-independent techniques or transfer learning to reduce or eliminate the calibration process altogether.[2,36] However, the effectiveness of these approaches can vary across individuals.[36,37]

Another alternative to reduce the burden of calibration is to improve the quality of the mental tasks performed. In a recent study, participants performed mental imagery tasks while observing an accumulative bar displayed on a screen. The bar increased as the BCI decoded the mental tasks, and subjects were unable to start the experimental session until reaching a certain level of precision.[28]

Another challenge in BCI applications based on MI is the variation in the way subjects perform mental tasks, resulting in high variability in success rates. In this study, we tackled both challenges using a virtual reality (VR) system. Participants were immersed in a VR environment where they practiced MI strategies for gait while receiving visual feedback. Then, they used these mental tasks to calibrate a BCI to control a robotic lower-limb exoskeleton without the VR feedback. As participants had already practiced in VR, less data was necessary for the BCI to achieve a high accuracy ratio, reducing the required training time and physical effort. To test the effectiveness of the VR system, we compared the BCI accuracies between subjects who underwent VR training and a control group that did not. The system was tested in closed-loop control with a lower-limb exoskeleton as presented in our previous research,[27] but this time with 12 subjects, including two patients suffering from SCI.

This study's contributions include:

- The design and proposal of a VR training system to enhance subjects' ability to perform MI tasks and speed up the BCI training process.
- A comparison of BCI accuracies between subjects that used VR for the BCI training and a control group that did not use it.
- Closed-loop experiments for controlling a lower-limb exoskeleton by means of the BCI.
- Assessment of these experiments with not only able-bodied subjects but also patients with SCI, analyzing the usability of the BCI application.

## RESULTS

Ten able-bodied subjects participated in the study. They had no movement impairment and did not report any known disease. Concerning BCI, they did not have any previous experience. Five participants were randomly included in control group and the other five in the VR group. Furthermore, two patients with SCI were recruited from the National Hospital of Paraplegics in Toledo. They had no prior experience with BCI technology. Further patient's details are specified in Table 1.

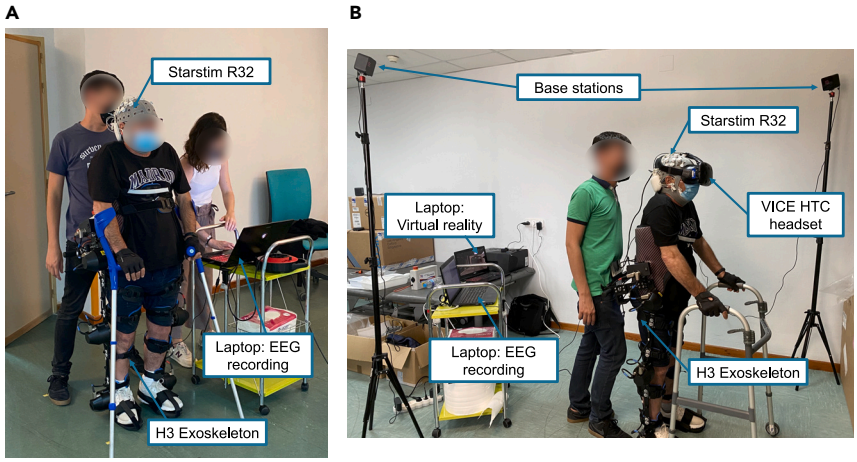



**Figure 1. Experimental setup**
(A) Calibration and closed-loop sessions; and (B)VR immersion.

Participants wore an EEG cap that recorded their brain-activity and wore a lower-limb robotic exoskeleton that assisted the gait. In addition, participants that were included in VR group used a VR environment. The setup of the experiment can be seen in Figure 1.

## Experimental design

Three distinct experiments were conducted, as illustrated in Figure 2. In the first experiment, five participants underwent a conventional calibration phase with the BCI. The calibration was followed by a closed-loop evaluation, where the exoskeleton was controlled by their thoughts resulting in motion feedback. They participated in five sessions scheduled in five different days, each of which consisted of a calibration and a closed-loop phase. Experiment 2 involved the VR system, in which five subjects participated in five experimental sessions too. Each session followed a similar format to the first experiment, but included a practice of mental tasks during VR immersion with visual feedback. The calibration phase was then shortened to half of that in experiment 1 before moving onto the closed-loop phase. Finally, experiment 3 involved two patients with incomplete SCI who underwent the same experimental sessions as in experiment 2, but with a reduced duration of VR immersion. However, in this experiment, the first of the five sessions did not involve closed-loop control phase because patients needed more time to become familiar with the system.

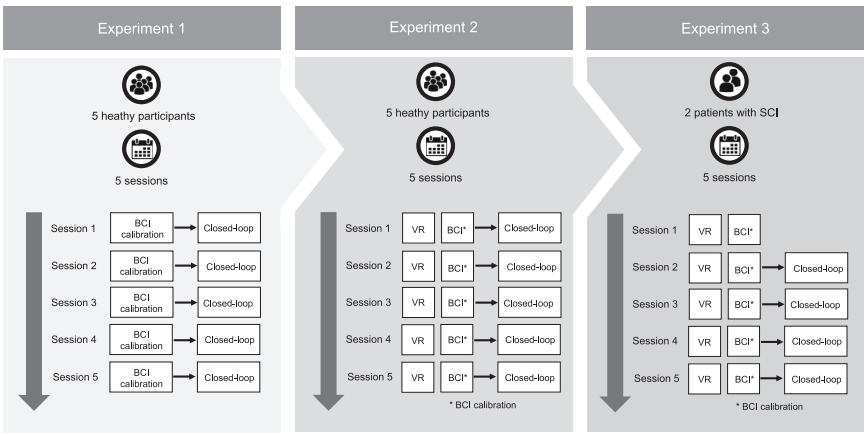



**Figure 2. Schema of the three experiments that were conducted in the present study**
The first two experiments involved able-bodied participants, and were conducted to evaluate the system before its implementation with patients suffering from spinal cord injury (SCI). In the third experiment, subjects only performed the closed-loop phase in sessions 2 to 5.

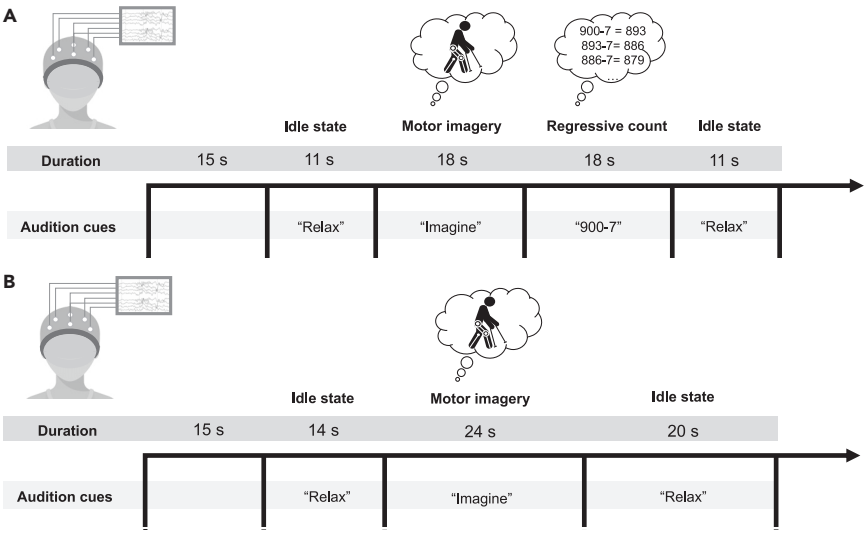



**Figure 3. Sequence of mental practices to perform**

(A) VR immersion and calibration trials (training); and (B) closed-loop trials. At the beginning of each task an audition cue indicated what participants had to do.

### *Experiment 1: Control group*

In Figure 2, we can see that experiment 1 involved five able-bodied participants who took part in five experimental sessions. Each session consisted of a calibration phase, during which the participants performed 22 trials. The sequence of each trial can be seen in Figure 3A. The trial began with a 15-s period in which the participants were not required to perform any mental task. This initial data was utilized to adjust the algorithms that were used to clean the EEG signals. After this, an acoustic cue signaled the start of the first mental task, which involved the participants relaxing in an idle state. Following this, another acoustic cue indicated the start of the MI period, during which the participants were instructed to focus on imagining the movement of their legs as if they were walking (kinesthetic imagination). Then, another cue signaled the transition to the regressive count task, during which the participants had to mentally perform a series of subtractions (e.g., 900-7, 893, 886, 879, etc.). Finally, a cue specified the start of the second idle state period. It is important to note that the data obtained during the regressive count period was not used for this BCI and was discarded. Therefore, only two classes were considered: 'Idle state' and 'MI of the gait'.

As reported by Ferrero et al.,[27] in the experimental setup, participants were assigned to perform two different types of trials: standing still with the exoskeleton (11 trials) and walking with the exoskeleton (11 trials). The exoskeleton was programmed to initiate and terminate gait based on pre-established commands, depending on whether the trial was in static or motion. Notably, the sequence of tasks was identical for all trials (as depicted in Figure 3A), as this was done to facilitate the creation of a dual-state BCI (illustrated in Figure 4), which was then utilized in the closed-loop control phase.

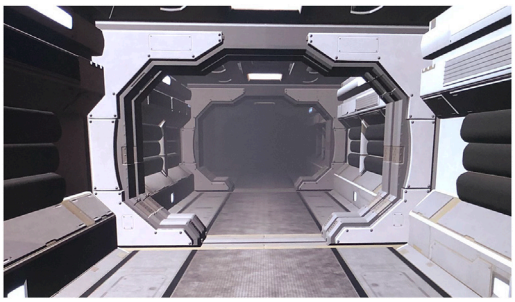


**Figure 4. Operation scheme of the brain-machine interface**

When subjects were standing still with the exoskeleton (static condition), Static model was the one that ruled. It decided to keep the exoskeleton static or to start the gait based on the brain patterns detected from the subject. On the other hand, when subjects were walking assisted by the exoskeleton (moving condition), Motion model oversaw the control. This model could decide to keep on the movement or to stop.

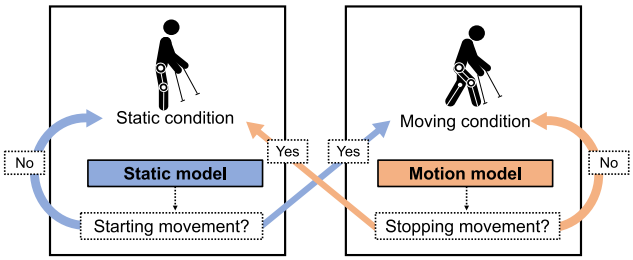



**Figure 5. Virtual reality environment**
It is a corridor inside a spacecraft developed in Unity engine (Unity Technologies, US).

A BCI based on motor imagery usually decodes two mental states to command the exoskeleton: 'Idle state' versus 'MI'. However, to ensure the accuracy of the BCI, the training data used to calibrate the BCI must be obtained under the same conditions as during the closed-loop phase. This means that the idle state and MI mental classes should be collected not only when the exoskeleton is static but also when it is in motion. To achieve this, four different events were collected during training trials: static idle state, idle state in motion, static MI, and MI in motion. These four classes were then used to create two different classifier models: Static ('Static idle state' versus 'Static MI of the gait') and Motion ('Idle state in motion' versus 'MI of the gait in motion').

In the closed-loop control phase, the initiation and cessation of the exoskeleton gait were controlled by the subject's mental practices and the BCI decoding system. During this phase, subjects performed five trials as described in Figure 3Bwhich included the same mental tasks as in the calibration phase but without the regressive count event. The models trained during the calibration phase, namely Static and Motion, were utilized for the control process (see Figure 4). As a dual-state machine, the Static model was used to keep the exoskeleton standing still and to detect the starting of the gait. However, once the exoskeleton initiated the gait, the BCI control model switched to Motion model, which was used to maintain the exoskeleton in motion until the desire to stop was detected. This would prompt a change of the model to Static again. Because the classes used in the model were collected in the same noise conditions, it was ensured that the differences between the classes were solely based on the mental tasks.

### *Experiment 2: VR group*

In experiment 2, the study included five able-bodied participants who were engaged in five experimental sessions. These sessions were similar to experiment 1. However, they also involved a prior training phase which included VR immersion. In this experiment, a shorter calibration phase with only 12 trials (six in static and six in motion) was used. During the VR immersion phase, the participants were placed in a virtual corridor inside a spacecraft in a first-person view, as depicted in Figure 5. They completed ten trials under similar conditions as the calibration phase (see Figure 3A). For five of these trials, the participant's avatar walked through the corridor, and for the other five, it stood still, providing visual feedback. It should be noted that the movement of the avatar was predefined, and did not depend on the participants' mental practices.

The closed-loop phase was identical to experiment 1.

### *Experiment 3: Patients with SCI*

Once the experimental setup was evaluated with able-bodied subjects, it was tested with patients with SCI. Accordingly, they participated in five sessions in different days and each session included three phases: VR, calibration, and closed-loop control of an exoskeleton. The sequence of mental tasks to perform was identical to Figure 3. The only difference with respect to able-bodied participants was that VR sessions were shorter. They performed 6 trials in the VR environment: three with the avatar being static and three with the avatar walking through the corridor. Then, they performed 12 trials for the calibration. And for closed-loop control of an exoskeleton, participants did seven trials. However, only five were in real control and considered for metrics assessment of the BCI, in the other two the BCI was manually forced to work perfectly as if they did the mental practices well. This strategy was employed to prevent them from getting frustrated.

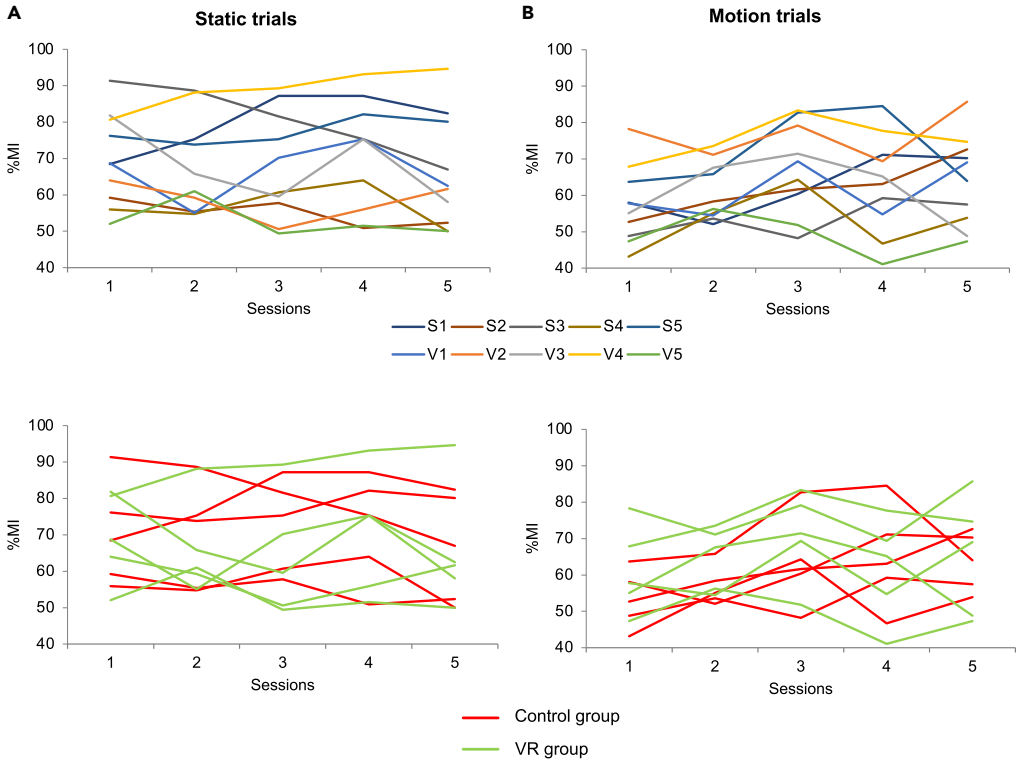



**Figure 6. Results from experiment 1**

(A) Results in static trials; (B) and motion trials. %MI is the accuracy of the model that goes from 0 to 100. S1–S5 are participants included in control group and V1–V5 are participants included in VR group.

During each session, we measured the level of exertion that patients perceived with the scale Borg Rating of Perceived Exertion Scale.[38] This range goes from 6 (“not effort at all”) to 20 (“absolute maximal effort, exhaustion”), and it was used as an indicator to terminate the experimental session if the level perceived was close to the maximum. In addition, after the last experimental session, patients filled two questionnaires: NASA Task Load Index (NASA-TLX)[39] that estimated the perceived mental workload; and Quebec User Evaluation of Satisfaction with Assistive Technology (QUEST 20)[40] that evaluated the satisfaction with the exoskeleton.

## Calibration phase: Control versus VR group

This section compares results obtained in calibration phase between control and VR group. Figure 6 shows the average accuracy (*%MI*) for each session and subject. Static and trials in motion were treated independently because they were produced under different conditions. Figure 6A shows results of cross-validation leave-one-out with static trials from all ten healthy individuals. S1–S5 represent the subjects that were included in the control group from experiment 1 and V1-V5 the subjects that were included in the VR group from experiment 2. Regarding trials in motion, Figure 6B shows their results. The average *%MI* (avg.±std.) of static trials from control group for each experimental session was: 70.2±14.2%, 69.6±14.5%, 72.5±12.9%, 71.9±14.6%, 66.4±15.1%. For trials in movement, performances per session were: 53.3± 8.0%, 57.0±5.4%, 63.5±12.4%, 64.9±14.1%, 63.6±8.0%. The average accuracy of VR group per session with static trials was: 69.5±12.4%, 65.8±13.0%, 63.8±16.5%, 70.2±16.8%, 65.4±17.1%. The average *%MI* of trials in movement from VR per session was: 61.3±12.0%, 64.6±8.7%, 71.0±12.2%, 61.6±14.1%, 65.1±16.7%.

To determine if there were any noteworthy distinctions between the VR group and the control group, a statistical analysis was conducted. However, the variability introduced by the grouping was not the only factor that could affect the results; the experimental sessions could also have an impact. Two separate analyses were conducted for static and trials in motion (further details are included in quantification and statistical analysis in STAR Methods).

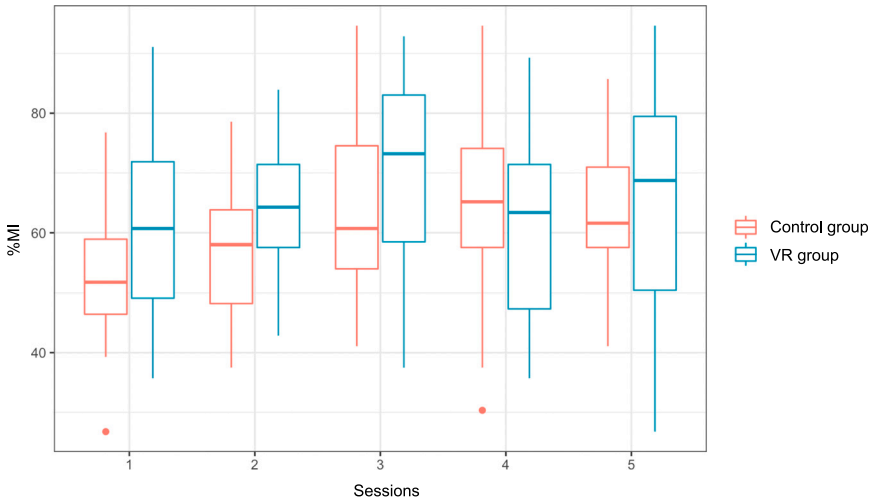



**Figure 7. Boxplot of accuracies obtained in trials in movement for each session and each group**

Subjects who participated in training with a VR environment and subjects who did not (control). %MI is the accuracy of the model that goes from 0 to 100.

- Static trials: No significant differences were found neither among sessions nor among groups.
- Trials in motion: The test showed that significant superior results were seen for VR group. Differences among sessions were significant. In addition, the interaction between group and sessions was significant. It means the evolution of accuracy through sessions varies from VR to control group. Figure 7 shows boxplot of accuracy values for each session and group. It can be observed that VR group outperforms control group in all sessions but session 4.

Figure 8 shows boxplots of accuracies obtained for all trials of each subject. It seemed that the performance of some of them was really different from the rest. Thus, a statistical test was conducted to see if these differences were relevant (further details are included in quantification and statistical analysis in STAR Methods).

- Static trials: The test showed significant differences among subjects.
- Trials in motion: The test showed significant differences among subjects too. Results could be influenced by the fact that subjects had a different training, and it was previously proven that VR group tended to have higher results. In the control group, there were three sub-groups of subjects that showed a significant different behavior: S1 and S2; S3 and S4; and S5. A similar trend was found in the VR group with three sub-groups: V1 and V3; V2 and V4; V5.

Intending to assess the evolution of performance through sessions, metrics were normalized with respect to the first session. This way, it can be studied how the accuracy evolved. New values of accuracy were computed as:

$$\%\widehat{MI}_{i,j} = \frac{\%MI_{i,j} - \%MI_1}{\%MI_1} \quad \text{(Equation 1)}$$

$\%MI_{i,j}$ is the accuracy obtained for a trial $j$ in an experimental session $i$, and $\%MI_1$ is the average accuracy of all trials of the first session. This value ranges from −1 to 1.

Figure 9 shows the average $\%\widehat{MI}$ that goes from −1 to 1. Statistical analysis was repeated with this new metric for static and trials in motion (further details are included in quantification and statistical analysis in STAR Methods).

- Static trials: No differences were found among sessions or the VR and control group.
- Motion trials: Relevant differences were found among sessions and the interaction between sessions and groups was significant. Moreover, the results from control group were substantially better than the ones of the VR group.

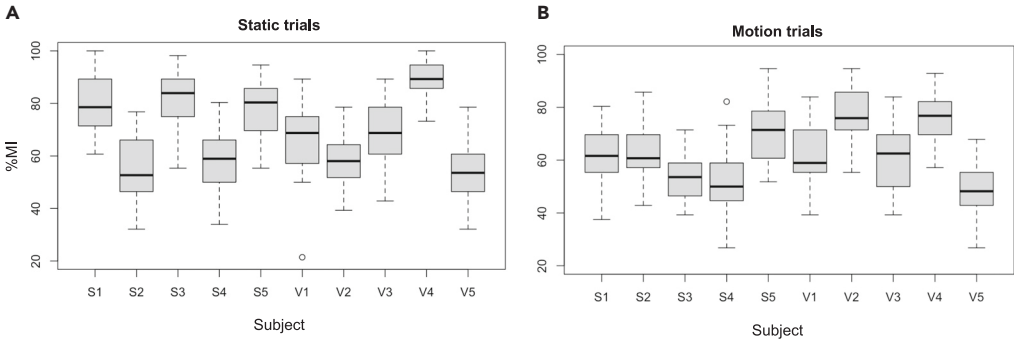



**Figure 8. Boxplot of accuracies obtained**

(A) Accuracies in trials in static, (B) and trials in motion for each subject. Subjects S1-S5 belong to the control group and V1-V5 to the VR group. %MI is the accuracy of the model that goes from 0 to 100.

Figure 10 shows the distribution of $\%\widehat{MI}$ for each subject. Metrics were normalized with respect to the first session. However, after performing a statistical analysis, significant differences were still found in static and motion trials among subjects (further details are included in quantification and statistical analysis in STAR Methods).

To investigate potential differences in neural activity between groups and individuals, power spectral density (PSD) was computed in alpha and beta frequency ranges. These frequency ranges have been associated with motor processes in EEG literature.[41] EEG signals were pre-processed according to

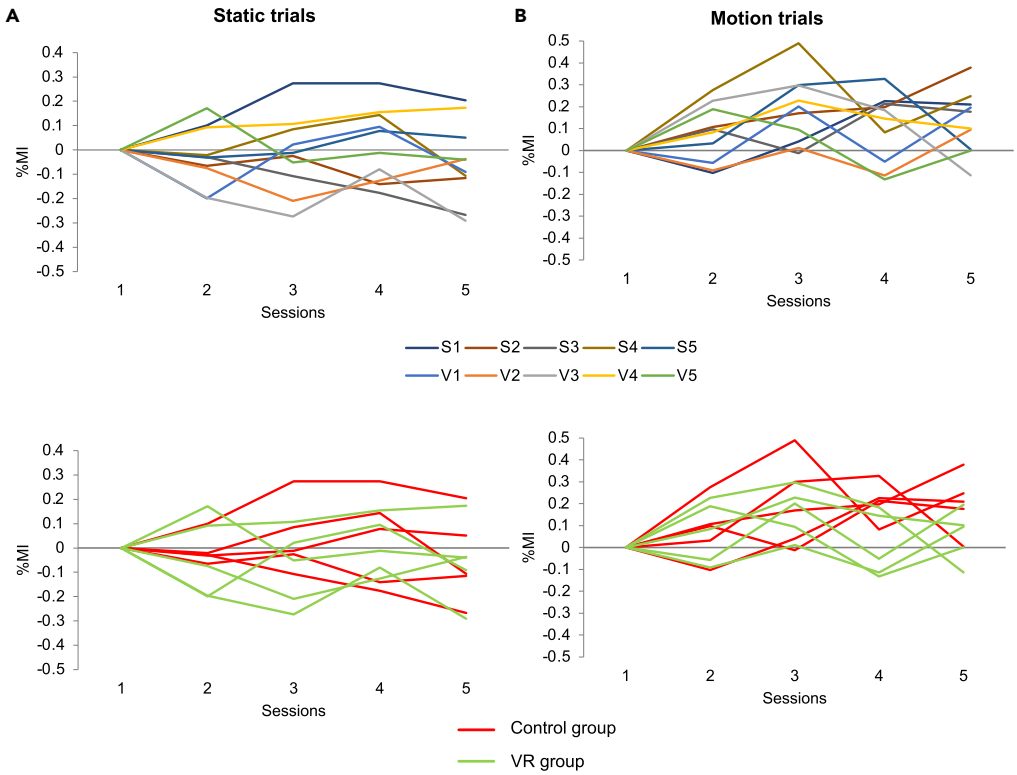



**Figure 9. Results from experiment 1**

(A) Results in trials in static, (B) and in motion trials. %MI is the relative accuracy of the model with respect to the first session. This value goes from −1 to 1. S1–S5 are participants included in control group and V1-V5 are participants included in VR group.

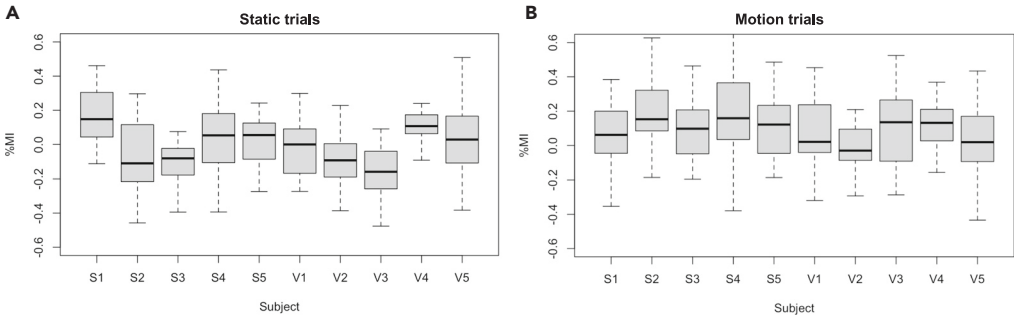



**Figure 10. Boxplot of relative accuracies**

(A) Relative accuracies obtained in static trials; (B) and in motion trials (B) for each subject. Subjects S1–S5 belong to the control group and V1-V5 to the VR group. %MI is the accuracy of the model that goes from 0 to 100.

method details in STAR Methods, and then subjected to independent component analysis (ICA)[42] as a denoising technique. As ICA cannot be used in real-time, it was only used for offline analysis. The ratio of PSD during MI periods and idle state periods was computed using the formula:

$$PSD_{electrode} = \frac{PSD_{electrode}(MI) - PSD_{electrode}(idle\ state)}{PSD_{electrode}(idle\ state} \quad \text{(Equation 2)}$$

This calculation was performed for all electrodes and averaged across all trials and sessions for each subject. The resulting values were then represented as topographical maps of the scalp, providing a spatially-detailed representation of changes in neural activity. The resulting maps are displayed in Figure 11 and 12.

The spectral maps exhibited significant variation among subjects, but a common decrease in power was observed in motor areas, particularly in subjects with high accuracy such as S3, S5, and V4. These observations align with previous research in the field. During the cognitive practice of MI, there appears to be a decrease in relative power of EEG signals in frequencies related to motor function, which is then followed by an increase in power. This phenomenon is commonly referred to as event-related desynchronization/synchronization (ERD/ERS) 42.

No significant differences were observed between the groups, which is advantageous for VR group as they exhibited comparable patterns with less training.

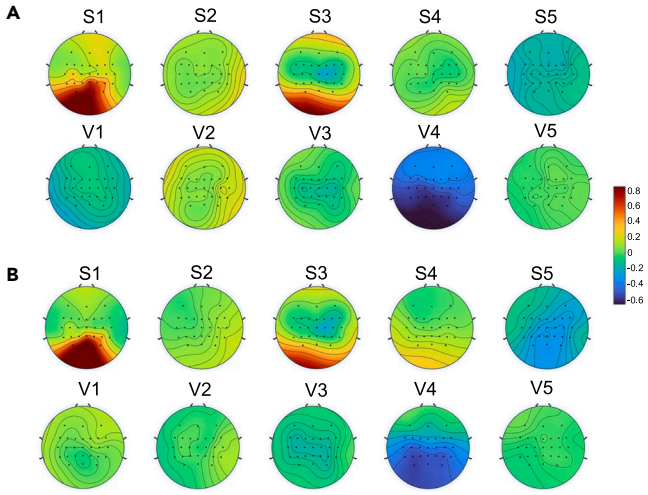



**Figure 11. Spectral maps of neural activity during MI in comparison to idle state**

This figure shows the activity in alpha frequency range (A) and in beta frequency range (B) only in static trials.

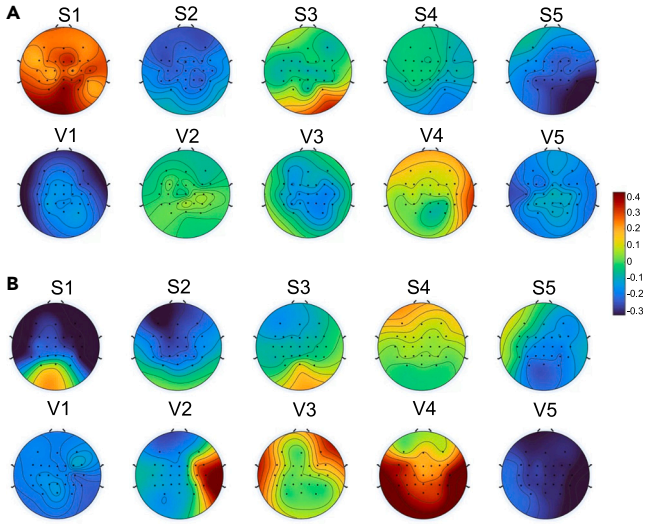



**Figure 12. Spectral maps of neural activity during MI in comparison to idle state**
This figure shows the activity in alpha frequency range (A) and in beta frequency range (B) only in trials in motion.

## Closed-loop control phase: Control versus VR group

This section presents the findings from the closed-loop control phase of experiment 1 and 2, as observed in the five participants in the control group and the five participants in the VR group. Throughout this phase, the exoskeleton was directed by each subject's cognitive techniques. The results from the control group participants are presented in Tables 2, 3, 4, 5, and 6,whereas those from the VR group participants are detailed in Tables 7, 8, 9, 10 and 11.

Table 12 presents a detailed comparison of the average results for both groups. Metrics are further explained in Evaluation metrics in STAR Methods. In both instances, Motion model demonstrated superior performance in comparison to Static model, indicating that the model controlling continuous walking or stopping was more effective. Analyzing the outcome of the experiments conducted with *%MI*,

**Table 2. Results from subject S1 in experiment 1**

| S1 | Session 1 | Session 2 | Session 3 | Session 4 | Session 5 | Avg. |
|---|---|---|---|---|---|---|
| %MI | 58.97 ± 3.48 | 63.28 ± 4.9 | 64.31 ± 5.61 | 64.14 ± 3.14 | 64.83 ± 6.77 | 63.11 ± 63.11 |
| %Commands | 63.59 ± 8.23 | 63.76 ± 7.59 | 64.27 ± 12.91 | 68.03 ± 1.15 | 70.43 ± 9.04 | 66.02 ± 66.02 |
| %Feedback EXO | 65.3 ± 10.16 | 67.52 ± 10.81 | 61.71 ± 34.79 | 56.75 ± 31.86 | 60.85 ± 35.46 | 62.43 ± 62.43 |
| Missing trials | 0 | 0 | 0 | 0 | 0.2 | 0.04 ± 0.04 |
| Static model: TPR | 73.33 ± 25.28 | 80 ± 27.39 | 60 ± 22.36 | 50 ± 35.36 | 30 ± 27.39 | 58.67 ± 58.67 |
| Static model: FPR | 46.67 ± 38.01 | 36.67 ± 34.16 | 30 ± 27.39 | 20 ± 44.72 | 20 ± 44.72 | 30.67 ± 30.67 |
| Static model: Accuracy commands | 56.67 ± 27.89 | 66.67 ± 31.18 | 70 ± 27.39 | 80 ± 44.72 | 60 ± 54.77 | 66.67 ± 66.67 |
| Static model: WD | 0.17 ± 0.55 | 0.35 ± 0.6 | 0.36 ± 0.41 | 0.48 ± 0.83 | 0.28 ± 0.79 | 0.33 ± 0.33 |
| Motion model: TPR | 76.67 ± 32.49 | 63.33 ± 41.5 | 10 ± 22.36 | 30 ± 44.72 | 6.67 ± 14.91 | 37.33 ± 37.33 |
| Motion model: FPR | 80 ± 27.39 | 73.33 ± 25.28 | 80 ± 44.72 | 70 ± 44.72 | 60 ± 54.77 | 72.67 ± 72.67 |
| Motion model: Accuracy commands | 46.67 ± 7.45 | 36.67 ± 21.73 | 20 ± 44.72 | 30 ± 44.72 | 20 ± 44.72 | 30.67 ± 30.67 |
| Motion model: WD | −0.21 ± 0.28 | −0.26 ± 0.53 | −0.64 ± 0.8 | −0.4 ± 0.85 | −0.45 ± 0.79 | −0.39±-0.39 |

%MI refers to the accuracy of the classifier per epochs, %Commands refers to the accuracy of commands per epochs, %Feedback EXO refers to the percentage of epochs with correct feedback from the exoskeleton, TPR is true positive ratio, FPR is false positive ratio and WD is weighted discriminator.

**Table 3. Results from subject S2 in experiment 1**

| S2 | Session 1 | Session 2 | Session 3 | Session 4 | Session 5 | Avg. |
|---|---|---|---|---|---|---|
| %MI | 53.79 ± 7.28 | 46.72 ± 5.43 | 48.45 ± 7.84 | 50.52 ± 13.34 | 62.76 ± 9.8 | 52.45 ± 52.45 |
| %Commands | 60 ± 10.65 | 46.32 ± 7.39 | 50.6 ± 11.47 | 57.78 ± 19.63 | 57.78 ± 12.29 | 54.5 ± 54.5 |
| %Feedback EXO | 30.09 ± 28.38 | 31.79 ± 29.14 | 23.59 ± 32.39 | 44.79 ± 29.35 | 52.82 ± 10.07 | 36.62 ± 36.62 |
| Missing trials | 0 | 0 | 0 | 0 | 0 | 0 ± 0 |
| Static model: TPR | 90 ± 22.36 | 60 ± 22.36 | 80 ± 27.39 | 20 ± 44.72 | 100 ± 0 | 70 ± 70 |
| Static model: FPR | 46.67 ± 36.13 | 61.67 ± 16.24 | 73.33 ± 36.51 | 70 ± 27.39 | 16.67 ± 23.57 | 53.67 ± 53.67 |
| Static model: Accuracy commands | 56.67 ± 25.28 | 38.33 ± 11.18 | 46.67 ± 13.94 | 10 ± 22.36 | 80 ± 27.39 | 46.33 ± 46.33 |
| Static model: WD | 0.23 ± 0.5 | −0.15 ± 0.24 | −0.13 ± 0.29 | −0.56 ± 0.49 | 0.71 ± 0.4 | 0.02 ± 0.02 |
| Motion model: TPR | 66.67 ± 31.18 | 66.67 ± 20.41 | 53.33 ± 44.72 | 33.33 ± 20.41 | 90 ± 22.36 | 62 ± 62 |
| Motion model: FPR | 30 ± 44.72 | 93.33 ± 14.91 | 73.33 ± 25.28 | 30 ± 44.72 | 10 ± 22.36 | 47.33 ± 47.33 |
| Motion model: Accuracy commands | 80 ± 27.39 | 46.67 ± 13.94 | 40 ± 22.36 | 70 ± 44.72 | 90 ± 22.36 | 65.33 ± 65.33 |
| Motion model: WD | 0.45 ± 0.59 | −0.39 ± 0.21 | −0.28 ± 0.25 | 0.25 ± 0.77 | 0.8 ± 0.45 | 0.17 ± 0.17 |

%MI refers to the accuracy of the classifier per epochs, %Commands refers to the accuracy of commands per epochs, %Feedback EXO refers to the percentage of epochs with correct feedback from the exoskeleton, TPR is true positive ratio, FPR is false positive ratio and WD is weighted discriminator.

*%Feedback EXO*, and *%Commands* per epoch, both groups exhibited a similar pattern of behavior. However, in keeping with the trend observed in the calibration trials, the performance of Motion model in the VR group far surpassed that of the control group.

## Patients with SCI

Experiment 3 consisted of evaluating the proposed BCI and training approach with patients with SCI. Moreover, this research evaluated the acceptance and usability perceived from them. QUEST questionnaire evaluated their satisfaction with the lower-limb exoskeleton H3 as an assistive technology. Their responses are shown in Table 13. Patient P1 was in general more satisfied with the exoskeleton than patient P2. Both agreed that it was not very comfortable, and P2 found that the adjustment was not easy. Table 14 shows their responses to NASA-TLX questionnaire that assessed the workload perceived. They felt they had to work hard physically and mentally to reach desired results, but they were satisfied with their

**Table 4. Results from subject S3 in experiment 1**

| S3 | Session 1 | Session 2 | Session 3 | Session 4 | Session 5 | Avg. |
|---|---|---|---|---|---|---|
| %MI | 48.1 ± 6.52 | 62.93 ± 5.81 | 59.83 ± 2.9 | 58.79 ± 8.78 | 0 ± 3.25 | 45.93 ± 45.93 |
| %Commands | 48.03 ± 15.96 | 61.88 ± 8.54 | 58.97 ± 0 | 59.32 ± 7.75 | 0 ± 0 | 45.64 ± 45.64 |
| %Feedback EXO | 52.14 ± 12.78 | 60.34 ± 5.75 | 58.97 ± 0 | 58.12 ± 5.89 | 0 ± 0 | 45.91 ± 45.91 |
| Missing trials | 0 | 0.6 | 1 | 0 | 1 | 0.52 ± 0.52 |
| Static model: TPR | 60 ± 22.36 | 20 ± 27.39 | 0 ± 0 | 80 ± 44.72 | 0 ± 0 | 32 ± 32 |
| Static model: FPR | 80 ± 18.26 | 13.33 ± 18.26 | 0 ± 0 | 30 ± 44.72 | 0 ± 0 | 24.67 ± 24.67 |
| Static model: Accuracy commands | 35 ± 9.13 | 20 ± 27.39 | 0 ± 0 | 70 ± 44.72 | 0 ± 0 | 25 ± 25 |
| Static model: WD | −0.35 ± 0.28 | 0.07 ± 0.09 | 0 ± 0 | 0.44 ± 0.88 | 0 ± 0 | 0.03 ± 0.03 |
| Motion model: TPR | 40 ± 25.28 | 40 ± 54.77 | 0 ± 0 | 10 ± 22.36 | 0 ± 0 | 18 ± 18 |
| Motion model: FPR | 90 ± 22.36 | 40 ± 54.77 | 0 ± 0 | 0 ± 0 | 0 ± 0 | 26 ± 26 |
| Motion model: Accuracy commands | 40 ± 25.28 | 20 ± 27.39 | 0 ± 0 | 20 ± 44.72 | 0 ± 0 | 16 ± 16 |
| Motion model: WD | −0.5 ± 0.42 | −0.12 ± 0.16 | 0 ± 0 | 0.16 ± 0.36 | 0 ± 0 | −0.09±-0.09 |

%MI refers to the accuracy of the classifier per epochs, %Commands refers to the accuracy of commands per epochs, %Feedback EXO refers to the percentage of epochs with correct feedback from the exoskeleton, TPR is true positive ratio, FPR is false positive ratio and WD is weighted discriminator.

**Table 5. Results from subject S4 in experiment 1**

| S4 | Session 1 | Session 2 | Session 3 | Session 4 | Session 5 | Avg. |
|---|---|---|---|---|---|---|
| %MI | 55.69 ± 11.33 | | 48.97 ± 11.59 | 50.34 ± 2.48 | 50.34 ± 10.95 | 51.34 ± 51.34 |
| %Commands | 55.38 ± 7.49 | | 50.43 ± 14.13 | 47.01 ± 3.68 | 55.21 ± 6.83 | 52.01 ± 52.01 |
| %Feedback EXO | 56.41 ± 5.27 | | 47.69 ± 15.26 | 50.09 ± 1.67 | 53.85 ± 6.45 | 52.01 ± 52.01 |
| Missing trials | 0 | | 0 | 0 | 0 | 0 ± 0 |
| Static model: TPR | 50 ± 35.36 | | 90 ± 22.36 | 20 ± 44.72 | 83.33 ± 23.57 | 60.83 ± 60.83 |
| Static model: FPR | 75 ± 27.64 | | 50 ± 11.79 | 90 ± 22.36 | 61.67 ± 11.18 | 69.17 ± 69.17 |
| Static model: Accuracy commands | 31.67 ± 23.86 | | 53.33 ± 7.45 | 10 ± 22.36 | 51.33 ± 15.2 | 36.58 ± 36.58 |
| Static model: WD | −0.36 ± 0.54 | | 0.18 ± 0.1 | −0.76 ± 0.54 | 0.02 ± 0.28 | −0.23±-0.23 |
| Motion model: TPR | 50 ± 37.27 | | 83.33 ± 23.57 | 10 ± 22.36 | 56.67 ± 9.13 | 50 ± 50 |
| Motion model: FPR | 70 ± 44.72 | | 60 ± 22.36 | 30 ± 44.72 | 70 ± 27.39 | 57.5 ± 57.5 |
| Motion model: Accuracy commands | 53.33 ± 38.01 | | 63.33 ± 12.64 | 10 ± 22.36 | 50 ± 11.79 | 44.17 ± 44.17 |
| Motion model: WD | −0.18 ± 0.7 | | 0.11 ± 0.27 | −0.2 ± 0.45 | −0.17 ± 0.33 | −0.11±-0.11 |

%MI refers to the accuracy of the classifier per epochs, %Commands refers to the accuracy of commands per epochs, %Feedback EXO refers to the percentage of epochs with correct feedback from the exoskeleton, TPR is true positive ratio, FPR is false positive ratio and WD is weighted discriminator.

performance. In addition, Borg scale was employed to measure the effort. It was measured after preparing all equipment, at the end of the VR phase, after an initial walk with the exoskeleton without the BCI, so they could get used to it, at the end of the calibration phase and at the end of the phase in which they controlled the exoskeleton with brain tasks. Results from this last scale are shown in Table 15. During each session, the level of exertion increased progressively, but patients never experienced values of exertion close to their own maximum.

The performance of the two patients during the calibration phase (see Figure 13) is within the range of able-bodied individuals (see Figure 6). Because there are significant differences among subjects, it is difficult to say if the performance of patients is in average lower or higher than able-bodied ones. Table 16 and 17show results obtained in closed-loop trials. Contrary to able-bodied individuals, Motion model was less efficient than static one.

**Table 6. Results from subject S5 in experiment 1**

| S5 | Session 1 | Session 2 | Session 3 | Session 4 | Session 5 | Avg. |
|---|---|---|---|---|---|---|
| %MI | 51.38 ± 5.3 | 69.14 ± 2.95 | 67.76 ± 10.66 | 83.28 ± 3.79 | 55.69 ± 10.76 | 65.45 ± 65.45 |
| %Commands | 54.87 ± 4.5 | 70.43 ± 6.09 | 67.86 ± 11.22 | 85.47 ± 5.37 | 65.47 ± 15.03 | 68.82 ± 68.82 |
| %Feedback EXO | 47.35 ± 2.46 | 63.59 ± 7.06 | 59.49 ± 12.41 | 79.83 ± 5.88 | 61.88 ± 13.64 | 62.43 ± 62.43 |
| Missing trials | 0 | 0 | 0 | 0 | 0 | 0 ± 0 |
| Static model: TPR | 90 ± 22.36 | 80 ± 44.72 | 100 ± 0 | 60 ± 54.77 | 30 ± 44.72 | 72 ± 72 |
| Static model: FPR | 60 ± 9.13 | 43.33 ± 9.13 | 56.67 ± 14.91 | 30 ± 27.39 | 66.67 ± 20.41 | 51.33 ± 51.33 |
| Static model: Accuracy commands | 36.67 ± 7.45 | 40 ± 22.36 | 40 ± 9.13 | 50 ± 50 | 13.33 ± 18.26 | 36 ± 36 |
| Static model: WD | −0.02 ± 0.12 | 0.13 ± 0.36 | 0.07 ± 0.2 | 0.24 ± 0.75 | −0.47 ± 0.41 | −0.01±-0.01 |
| Motion model: TPR | 70 ± 18.26 | 70 ± 27.39 | 70 ± 18.26 | 70 ± 27.39 | 53.33 ± 36.13 | 66.67 ± 66.67 |
| Motion model: FPR | 20 ± 44.72 | 0 ± 0 | 0 ± 0 | 0 ± 0 | 20 ± 44.72 | 8 ± 8 |
| Motion model: Accuracy commands | 93.33 ± 14.91 | 100 ± 0 | 100 ± 0 | 100 ± 0 | 66.67 ± 47.14 | 92 ± 92 |
| Motion model: WD | 0.64 ± 0.47 | 0.88 ± 0.11 | 0.88 ± 0.07 | 0.88 ± 0.11 | 0.41 ± 0.69 | 0.74 ± 0.74 |

%MI refers to the accuracy of the classifier per epochs, %Commands refers to the accuracy of commands per epochs, %Feedback EXO refers to the percentage of epochs with correct feedback from the exoskeleton, TPR is true positive ratio, FPR is false positive ratio and WD is weighted discriminator.

**Table 7. Results from subject V1 in experiment 2**

| V1 | Session 1 | Session 2 | Session 3 | Session 4 | Session 5 | Avg. |
|---|---|---|---|---|---|---|
| %MI | 56.9 ± 9.12 | 55.69 ± 2.96 | 56.55 ± 6.26 | 60 ± 10.64 | 58.62 ± 18.79 | 57.55 ± 57.55 |
| %Commands | 58.97 ± 11 | 57.78 ± 2.61 | 64.27 ± 10.68 | 63.08 ± 12.55 | 59.83 ± 26.66 | 60.79 ± 60.79 |
| %Feedback EXO | 58.12 ± 13.73 | 56.92 ± 3.45 | 60.68 ± 6.84 | 59.66 ± 12.18 | 60.17 ± 21.42 | 59.11 ± 59.11 |
| Missing trials | 0 | 0.2 | 0 | 0 | 0 | 0.04 ± 0.04 |
| Static model: TPR | 80 ± 44.72 | 70 ± 44.72 | 90 ± 22.36 | 100 ± 0 | 100 ± 0 | 88 ± 88 |
| Static model: FPR | 60 ± 25.28 | 36.67 ± 21.73 | 78.33 ± 21.73 | 40 ± 25.28 | 50 ± 37.27 | 53 ± 53 |
| Static model: Accuracy commands | 38.33 ± 28.63 | 46.67 ± 27.39 | 34.67 ± 10.63 | 70 ± 18.26 | 63.33 ± 21.73 | 50.6 ± 50.6 |
| Static model: WD | −0.05 ± 0.43 | 0.19 ± 0.12 | −0.22 ± 0.34 | 0.42 ± 0.36 | 0.28 ± 0.48 | 0.12 ± 0.12 |
| Motion model: TPR | 53.33 ± 36.13 | 50 ± 35.36 | 65 ± 25.28 | 73.33 ± 25.28 | 70 ± 29.81 | 62.33 ± 62.33 |
| Motion model: FPR | 60 ± 22.36 | 40 ± 41.83 | 50 ± 35.36 | 40 ± 22.36 | 33.33 ± 31.18 | 44.67 ± 44.67 |
| Motion model: Accuracy commands | 50 ± 28.87 | 50 ± 35.36 | 70 ± 24.01 | 66.67 ± 20.41 | 70 ± 29.81 | 61.33 ± 61.33 |
| Motion model: WD | −0.09 ± 0.52 | 0.1 ± 0.41 | 0.18 ± 0.55 | 0.29 ± 0.41 | 0.37 ± 0.61 | 0.17 ± 0.17 |

%MI refers to the accuracy of the classifier per epochs, %Commands refers to the accuracy of commands per epochs, %Feedback EXO refers to the percentage of epochs with correct feedback from the exoskeleton, TPR is true positive ratio, FPR is false positive ratio and WD is weighted discriminator.

## DISCUSSION

### Calibration phase: Control versus VR group

In this study, the performance of two groups of subjects was compared. The control group underwent a calibration phase consisting of 11 static and 11 movement trials, whereas the VR group underwent an additional previous phase in a VR environment, followed by a calibration phase consisting of six static and six movement trials. It was observed that the VR group experienced less physical exertion because of the reduced number of walking trials with the exoskeleton.

Significant differences were observed among the subjects, with the VR group demonstrating superior performance in trials involving movement. However, no significant differences were observed in static trials. It is noteworthy that even if the performance of the two groups was similar, the VR approach could offer an alternative method of training that requires less physical effort. Furthermore, when the training was divided

**Table 8. Results from subject V2 in experiment 2**

| V2 | Session 1 | Session 2 | Session 3 | Session 4 | Session 5 | Avg. |
|---|---|---|---|---|---|---|
| %MI | 65 ± 7.8 | 60.52 ± 10.1 | 67.41 ± 6.38 | 55.17 ± 7.64 | 72.07 ± 14.11 | 64.03 ± 64.03 |
| %Commands | 71.11 ± 10.73 | 63.25 ± 16.78 | 62.91 ± 9.72 | 54.36 ± 7.75 | 72.14 ± 13.54 | 64.75 ± 64.75 |
| %Feedback EXO | 65.47 ± 11.94 | 62.05 ± 13.76 | 54.19 ± 12.89 | 50.26 ± 8.6 | 68.03 ± 11.91 | 60 ± 60 |
| Missing trials | 0 | 0 | 0 | 0.2 | 0 | 0.04 ± 0.04 |
| Static model: TPR | 70 ± 44.72 | 80 ± 44.72 | 100 ± 0 | 80 ± 44.72 | 80 ± 44.72 | 82 ± 82 |
| Static model: FPR | 60 ± 9.13 | 56.67 ± 25.28 | 40 ± 22.36 | 20 ± 44.72 | 43.33 ± 43.46 | 44 ± 44 |
| Static model: Accuracy commands | 26.67 ± 14.91 | 36.67 ± 21.73 | 53.33 ± 27.39 | 66.67 ± 47.14 | 56.67 ± 43.46 | 48 ± 48 |
| Static model: WD | −0.16 ± 0.19 | −0.03 ± 0.39 | 0.32 ± 0.38 | 0.52 ± 0.67 | 0.23 ± 0.83 | 0.18 ± 0.18 |
| Motion model: TPR | 70 ± 18.26 | 56.67 ± 25.28 | 80 ± 27.39 | 66.67 ± 47.14 | 63.33 ± 41.5 | 67.33 ± 67.33 |
| Motion model: FPR | 20 ± 44.72 | 30 ± 27.39 | 0 ± 0 | 10 ± 22.36 | 0 ± 0 | 12 ± 12 |
| Motion model: Accuracy commands | 93.33 ± 14.91 | 70 ± 27.39 | 100 ± 0 | 70 ± 44.72 | 80 ± 44.72 | 82.67 ± 82.67 |
| Motion model: WD | 0.64 ± 0.47 | 0.35 ± 0.51 | 0.92 ± 0.11 | 0.59 ± 0.57 | 0.73 ± 0.42 | 0.65 ± 0.65 |

%MI refers to the accuracy of the classifier per epochs, %Commands refers to the accuracy of commands per epochs, %Feedback EXO refers to the percentage of epochs with correct feedback from the exoskeleton, TPR is true positive ratio, FPR is false positive ratio and WD is weighted discriminator.

**Table 9. Results from subject V3 in experiment 2**

| V3 | Session 1 | Session 2 | Session 3 | Session 4 | Session 5 | Avg. |
|---|---|---|---|---|---|---|
| %MI | 54.48 ± 9.82 | 54.83 ± 10.53 | 53.97 ± 16.41 | 47.24 ± 5.19 | 47.41 ± 8.47 | 51.59 ± 51.59 |
| %Commands | 58.8 ± 15.5 | 54.87 ± 13.45 | 57.09 ± 20.27 | 49.74 ± 3.16 | 54.36 ± 8.56 | 54.97 ± 54.97 |
| %Feedback EXO | 56.58 ± 15.91 | 55.04 ± 8.94 | 57.09 ± 27.28 | 49.74 ± 2.04 | 50.6 ± 5.25 | 53.81 ± 53.81 |
| Missing trials | 0 | 0 | 0 | 0 | 0 | 0 ± 0 |
| Static model: TPR | 100 ± 0 | 83.33 ± 23.57 | 80 ± 27.39 | 80 ± 44.72 | 73.33 ± 25.28 | 83.33 ± 83.33 |
| Static model: FPR | 55 ± 16.24 | 53.33 ± 38.01 | 63.33 ± 24.72 | 86.67 ± 18.26 | 68.33 ± 23.86 | 65.33 ± 65.33 |
| Static model: Accuracy commands | 54.67 ± 11.69 | 56.67 ± 25.28 | 40 ± 9.13 | 28.33 ± 18.26 | 43 ± 16.09 | 44.53 ± 44.53 |
| Static model: WD | 0.18 ± 0.23 | 0.14 ± 0.55 | −0.07 ± 0.26 | −0.38 ± 0.41 | −0.13 ± 0.4 | −0.05±-0.05 |
| Motion model: TPR | 68.33 ± 20.75 | 66.67 ± 31.18 | 60 ± 25.28 | 43.33 ± 27.89 | 61.67 ± 26.09 | 60 ± 60 |
| Motion model: FPR | 53.33 ± 7.45 | 63.33 ± 41.5 | 60 ± 41.83 | 43.33 ± 25.28 | 80 ± 27.39 | 60 ± 60 |
| Motion model: Accuracy commands | 58.33 ± 11.79 | 53.33 ± 27.39 | 56.67 ± 25.28 | 46.67 ± 27.39 | 51.67 ± 19 | 53.33 ± 53.33 |
| Motion model: WD | 0.09 ± 0.17 | −0.05 ± 0.66 | −0.02 ± 0.55 | 0.02 ± 0.11 | −0.24 ± 0.48 | −0.04±-0.04 |

%MI refers to the accuracy of the classifier per epochs, %Commands refers to the accuracy of commands per epochs, %Feedback EXO refers to the percentage of epochs with correct feedback from the exoskeleton, TPR is true positive ratio, FPR is false positive ratio and WD is weighted discriminator.

into VR and calibration phases, it was perceived as less monotonous, which can be advantageous in keeping the subjects motivated and focused.

The evolution of performance through sessions was only found significant in trials in movement. Accuracy values were significantly different from last session to the first one, but accuracy growth was different among subjects. This growth was significantly different from subjects in VR and controls. The global trend was similar for the first three sessions, but in the fourth one, VR subjects got lower accuracy than in the previous session.

### Closed-loop control phase: Control versus VR group

This study employed a dual-state machine for the BCI, comprising two distinct models. Static model was activated when participants were stationary, whereas Motion model was activated when participants

**Table 10. Results from subject V4 in experiment 2**

| V4 | Session 1 | Session 2 | Session 3 | Session 4 | Session 5 | Avg. |
|---|---|---|---|---|---|---|
| %MI | 51.38 ± 2.77 | 64.14 ± 6.64 | 58.45 ± 8.32 | 56.9 ± 5.21 | 50 ± 2.73 | 56.17 ± 56.17 |
| %Commands | 52.31 ± 3.79 | 63.08 ± 7.73 | 61.37 ± 6.38 | 62.22 ± 8.18 | 52.59 ± 5.94 | 58.19 ± 58.19 |
| %Feedback EXO | 47.52 ± 4.08 | 60.68 ± 6.54 | 54.7 ± 6.81 | 59.15 ± 7.73 | 0.25 ± 0.39 | 54.09 ± 54.09 |
| Missing trials | 0 | 0 | 0 | 0 | 0 | 0 ± 0 |
| Static model: TPR | 70 ± 44.72 | 80 ± 44.72 | 80 ± 44.72 | 100 ± 0 | 3.88 ± 0.79 | 80 ± 80 |
| Static model: FPR | 68.33 ± 20.75 | 36.67 ± 21.73 | 56.67 ± 27.89 | 70 ± 18.26 | 70 ± 44.72 | 62 ± 62 |
| Static model: Accuracy commands | 35 ± 22.36 | 53.33 ± 38.01 | 36.67 ± 21.73 | 36.67 ± 7.45 | 68.33 ± 20.75 | 38 ± 38 |
| Static model: WD | −0.19 ± 0.51 | 0.27 ± 0.55 | −0.03 ± 0.57 | −0.08 ± 0.21 | 35 ± 22.36 | −0.07±-0.07 |
| Motion model: TPR | 73.33 ± 27.89 | 80 ± 27.39 | 70 ± 29.81 | 56.67 ± 14.91 | −0.19 ± 0.51 | 66.67 ± 66.67 |
| Motion model: FPR | 50 ± 35.36 | 23.33 ± 32.49 | 20 ± 27.39 | 30 ± 27.39 | 73.33 ± 27.89 | 30.67 ± 30.67 |
| Motion model: Accuracy commands | 76.67 ± 13.69 | 80 ± 29.81 | 83.33 ± 23.57 | 76.67 ± 22.36 | 50 ± 35.36 | 78 ± 78 |
| Motion model: WD | 0.25 ± 0.39 | 0.57 ± 0.55 | 0.58 ± 0.47 | 0.39 ± 0.42 | 76.67 ± 13.69 | 0.43 ± 0.43 |

%MI refers to the accuracy of the classifier per epochs, %Commands refers to the accuracy of commands per epochs, %Feedback EXO refers to the percentage of epochs with correct feedback from the exoskeleton, TPR is true positive ratio, FPR is false positive ratio and WD is weighted discriminator.

**Table 11. Results from subject V5 in experiment 2**

| V5 | Session 1 | Session 2 | Session 3 | Session 4 | Session 5 | Avg. |
|---|---|---|---|---|---|---|
| %MI | 49.66 ± 9.71 | 48.28 ± 8.34 | 45.52 ± 3.42 | 47.41 ± 5.24 | 54.66 ± 6.99 | 49.11 ± 49.11 |
| %Commands | 49.91 ± 10.39 | 47.18 ± 2.8 | 44.27 ± 11.02 | 48.03 ± 5.78 | 49.57 ± 9.57 | 47.79 ± 47.79 |
| %Feedback EXO | 43.59 ± 9.88 | 47.18 ± 2.8 | 46.67 ± 7.42 | 45.98 ± 6.03 | 54.36 ± 7.68 | 47.56 ± 47.56 |
| Missing trials | 0 | 0 | 0 | 0 | 0 | 0 ± 0 |
| Static model: TPR | 80 ± 27.39 | 100 ± 0 | 63.33 ± 41.5 | 90 ± 22.36 | 100 ± 0 | 86.67 ± 86.67 |
| Static model: FPR | 55 ± 11.18 | 73.33 ± 25.28 | 73.33 ± 25.28 | 56.67 ± 27.89 | 83.33 ± 23.57 | 68.33 ± 68.33 |
| Static model: Accuracy commands | 41.67 ± 11.79 | 46.67 ± 13.94 | 38.33 ± 26.09 | 46.67 ± 13.94 | 50 ± 0 | 44.67 ± 44.67 |
| Static model: WD | 0.02 ± 0.27 | −0.05 ± 0.33 | −0.25 ± 0.43 | 0.07 ± 0.35 | −0.13 ± 0.24 | −0.07±-0.07 |
| Motion model: TPR | 83.33 ± 23.57 | 56.67 ± 27.89 | 50 ± 37.27 | 80 ± 29.81 | 53.33 ± 29.81 | 64.67 ± 64.67 |
| Motion model: FPR | 60 ± 41.83 | 53.33 ± 7.45 | 60 ± 41.83 | 50 ± 35.36 | 63.33 ± 7.45 | 57.33 ± 57.33 |
| Motion model: Accuracy commands | 76.67 ± 13.69 | 55 ± 16.24 | 46.67 ± 36.13 | 70 ± 18.26 | 45 ± 18.26 | 58.67 ± 58.67 |
| Motion model: WD | 0.19 ± 0.46 | 0.02 ± 0.24 | −0.12 ± 0.6 | 0.24 ± 0.47 | −0.15 ± 0.3 | 0.04 ± 0.04 |

%MI refers to the accuracy of the classifier per epochs, %Commands refers to the accuracy of commands per epochs, %Feedback EXO refers to the percentage of epochs with correct feedback from the exoskeleton, TPR is true positive ratio, FPR is false positive ratio and WD is weighted discriminator.

walked with the assistance of an exoskeleton (as illustrated in Figure 4). The performance of the two models varied significantly, with Motion model exhibiting superior performance in both participant groups. This was primarily attributed to the higher FPR of Static model, which made it challenging for participants to remain relaxed when they had control over the exoskeleton, an experience not encountered during the calibration phase.

The *%MI* , *%Commands*, and *%Feedback EXO* indicators evaluated the efficiency of the BCI as a whole, independent of the performance of each model. Both groups displayed similar behavior when considering these indicators. However, focusing on the performance of each model, the weighted discriminator proved to be the most effective metric. Although the performance of Static model was similar across groups, Motion model of the VR group outperformed that of the control group. These findings align with those observed in the calibration phase (refer to section 4.1).

**Table 12. Comparison between VR and control group**

| | Control | VR |
|---|---|---|
| %MI | 55.83 ± 14.68 | 55.69 ± 6.73 |
| %Commands | 57.62 ± 15.18 | 57.3 ± 7.22 |
| %Feedback EXO | 51.87 ± 16.62 | 54.91 ± 6.43 |
| Missing trials | 0.12 ± 0.12 | 0.02 ± 0.02 |
| Static model: TPR | 58.61 ± 31.05 | 84 ± 11.79 |
| Static model: FPR | 44.93 ± 25.41 | 58.53 ± 16.42 |
| Static model: Accuracy commands | 42.35 ± 23.96 | 45.16 ± 11.91 |
| Static model: WD | 0.04 ± 0.35 | 0.02 ± 0.23 |
| Motion model: TPR | 46.67 ± 28.06 | 64.2 ± 10.8 |
| Motion model: FPR | 41.67 ± 33.49 | 40.93 ± 20.87 |
| Motion model: Accuracy commands | 49.86 ± 32.13 | 66.8 ± 14.98 |
| Motion model: WD | 0.07 ± 0.48 | 0.25 ± 0.3 |

%MI refers to the accuracy of the classifier per epochs, %Commands refers to the accuracy of commands per epochs, %Feedback EXO refers to the percentage of epochs with correct feedback from the exoskeleton, TPR is true positive ratio, FPR is false positive ratio and WD is weighted discriminator.

**Table 13. Results from QUEST questionnaire**

| Questions | P1 | P2 |
|---|---|---|
| 1. The **dimensions** (size, height, length, width) of the device? | 4/5 | 3/5 |
| 2. The weight of the device? | 5/5 | 3/5 |
| 3. The **ease in adjusting** (fixing, fastening) the parts of the device? | 4/5 | 2/5 |
| 4. How **safe and secure** the device is? | 4/5 | 5/5 |
| 5. The **durability** (endurance, resistance to wear) of the device? | 4/5 | 4/5 |
| 6. How **easy** is it to use the device? | 4/5 | – |
| 7. How **comfortable** the device is? | 3/5 | 3/5 |
| 8. How **effective** the device is to solve the problem for which you are using it? | 4/5 | 5/5 |
| **Total:** | 32/40 | 25/35 |

1: Not satisfied at all; 2: Not very satisfied; 3: More or less satisfied; 4: Quite satisfied; 5: Very satisfied.

As previously discussed, there were notable variations in performance across participants. Within the control group, S5 exhibited the best results, whereas S2 demonstrated the lowest performance, aligning with their respective performance in the calibration phase. In the VR group, V2 achieved the highest outcomes, whereas V5 showed the lowest. Of interest, during the calibration phase, it was V4 who demonstrated the highest performance, but the values of %*MI* underwent substantial changes from calibration to closed-loop phases. Various factors can influence the system's performance, such as the cognitive state and concentration level of the subject.[43,44]

Within the control group, three subjects did not miss any trials in all sessions, whereas in the VR group, all participants except V2 did not miss any trials in all sessions. These metrics can be interpreted as 88% and 98% of trials where at least one activation command was issued during motor imagery. These results can be compared to the 84.4% accuracy reported by,[32] whose approach was based on motion intention, and the exoskeleton was blocked when subjects were required to relax, giving them control over it only during specific time periods.

In the research presented by,[28] the average accuracy reached 84.4 $\pm$ 5.43% when starting the gait on ten able-bodied subjects. They defined a model that compared gait MI with sit MI so subjects could start the gait with the exoskeleton or could sit. This metric can be compared with our accuracy of commands in Static model. Their approach outperforms our proposed BCI. A possible explanation could be that BCI models perform best when different MI patterns are being compared.[20,44] However, it was crucial in our design to include a non-intention condition.[25] reported an accuracy of 73.4% in seven participants. Because the exoskeleton was only free during periods of MI, these metrics are in accordance with our Motion model that showed an average accuracy of 49.86 $\pm$ 32.13% in control group and 66.8 $\pm$ 14.98% in VR group. In addition, our approach was completely free for the user to transition among tasks.

In our previous studies,[26,27] MI and attention paradigms were combined to control an exoskeleton. In[26], % Commands was 56.77% and in[27] it was 64.08%. Results obtained in this paper are within the same range, but in the VR group less information was employed to train the classifier which supposes a similar performance with a lower protocol time.

**Table 14. Results from NASA-TLX**

| Item | P1 | P2 |
|---|---|---|
| Mental demand | 85/100 | 80/100 |
| Physical demand | 85/100 | 25/100 |
| Temporal demand | 70/100 | 10/100 |
| Performance | 95/100 | 100/100 |
| Effort | 85/100 | 90/100 |
| Frustration level | 90/100 | 0/100 |

**Table 15. Results from borg rating of perceived exertion scale**

| | P1 | | | | | P2 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Sessions | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| Time 1 | 11 | 11 | 10 | 11 | 10 | 8 | 8 | 7 | 6 | 6 |
| Time 2 | 12 | 12 | 11 | 11 | 11 | 10 | 8 | 9 | 7 | 8 |
| Time 3 | 13 | 12 | 11 | 11 | 11 | 13 | 10 | 10 | 11 | 9 |
| Time 4 | 13 | 13 | 12 | 12 | 11 | 15 | 12 | 11 | 11 | 11 |
| Time 5 | – | 13 | 13 | 12 | 12 | – | 13 | 15 | 14 | 15 |

Time 1: after preparing all equipment.
Time 2: at the end of the virtual reality phase.
Time 3: after an initial walk with the exoskeleton without the BCI.
Time 4: at the end of the calibration phase.
Time 5: and at the end of the phase in which they controlled the exoskeleton with brain tasks.

## Patients with SCI

Results from calibration phase of P1 showed a positive trend throughout sessions. This trend was evident in static and trials in motion. P2 did not demonstrate the same learning effect. It is important to note that whereas both patients suffered an incomplete SCI, the injury of P2 affected motor function to a greater extent. P2 may need more practice with the system before reaching an acceptable level of performance.

The ratio of missing trials was 0 for P1 and 0.05 for P1. Therefore, P1 could activate the exoskeleton in all the trials and P2 only missed one trial in last session. Results are in line with.[29] They evaluated their BCI with a patient with SCI and it was only focused on activation of the exoskeleton which corresponds to our Static model. On the other hand,[30,31] defined a closed-loop scenario in which an SCI patient was asked to stop the gait. They obtained 99.31% correct stopping commands. It is difficult to compare this study with our research because in ours, subjects had to stop the exoskeleton in a fixed short period of time instead of having several attempts to do it.[9] evaluated their approach of combining visual/haptic feedback, and MI with five post-stroke patients. They got 50.7 and 68% of successful activation attempts which could be comparable to our TPR in static. However, although patients who suffered an SCI do not have cognitive affectation, patients who suffered a stroke have brain damage, which makes it difficult to do a comparison. TPR in static was 77.71% for patient P1 and 76.67% for P2. These metrics are the number of successful activations of the exoskeleton out of all the attempts. [32,33] used motion intention to start the movement and then, the system kept it. They both were evaluated with four SCI and results from[33] were 79%, 93.10%, 87.10% and 68.81% for each patient. Again, these metrics can be confronted to our TPR in static.[32] reported an average accuracy of 77.6%. However, metrics were computed with different rules as a trial was considered successful if at least an activation command was sent to the exoskeleton. In our results, P1 got a 100% of trials in which at least a command was sent to the exoskeleton (0 missing trials), and P2 got 100% in all sessions, but last one with 80%. The proposed BCI was proven to be at least as effective as previous approaches presented in the literature with more restricting assessing metrics.

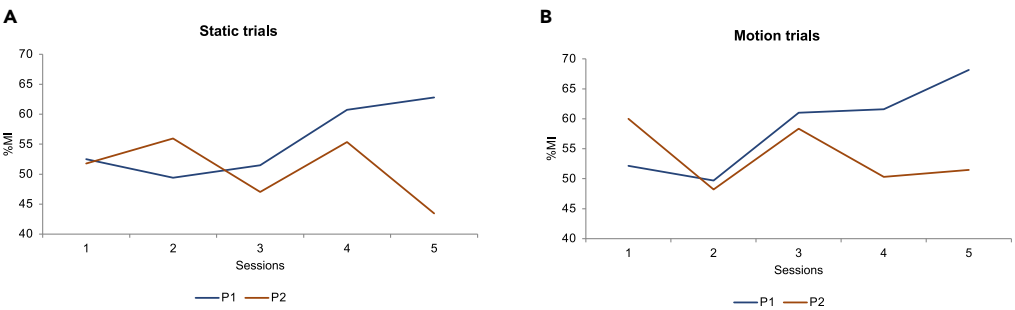



**Figure 13. Results from calibration phase of experiment 3**
Results (A) in static trials, (B)and trials in motion. %MI is the accuracy of the model that goes from 0 to 100.

**Table 16. Results from patient P1 in closed-loop trials of experiment 3**

| P1 | Session 2 | Session 3 | Session 4 | Session 5 | Avg. |
|---|---|---|---|---|---|
| %MI | 50.52 ± 8.35 | 46.55 ± 13.41 | 53.97 ± 3.48 | 50 ± 6.68 | 50.26 ± 3.03 |
| %Commands | 54.19 ± 16.51 | 45.81 ± 19 | 59.15 ± 9.15 | 54.36 ± 15.33 | 53.38 ± 5.54 |
| %Feedback EXO | 57.61 ± 18.98 | 46.67 ± 16.86 | 55.04 ± 8.01 | 54.87 ± 16.24 | 53.55 ± 4.75 |
| Missing trials | 0 | 0 | 0 | 0 | 0 ± 0 |
| Static model: TPR | 73.33 ± 25.28 | 80 ± 27.39 | 87.5 ± 25 | 70 ± 27.39 | 77.71 ± 7.74 |
| Static model: FPR | 70 ± 27.39 | 40 ± 36.51 | 73.33 ± 25.28 | 43.33 ± 9.13 | 56.67 ± 17.43 |
| Static model: Accuracy commands | 44.67 ± 7.67 | 60 ± 36.51 | 30 ± 18.26 | 46.67 ± 7.45 | 45.33 ± 12.28 |
| Static model: WD | −0.14 ± 0.19 | 0.28 ± 0.66 | −0.09 ± 0.25 | 0.13 ± 0.13 | 0.04 ± 0.2 |
| Motion model: TPR | 56.67 ± 43.46 | 73.33 ± 25.28 | 66.67 ± 33.33 | 80 ± 27.39 | 69.17 ± 9.95 |
| Motion model: FPR | 86.67 ± 18.26 | 50 ± 50 | 50 ± 35.36 | 80 ± 27.39 | 66.67 ± 19.44 |
| Motion model: Accuracy commands | 40 ± 25.28 | 63.33 ± 34.16 | 71.67 ± 18.26 | 53.33 ± 7.45 | 57.08 ± 13.63 |
| Motion model: WD | −0.4 ± 0.34 | 0.17 ± 0.77 | 0.2 ± 0.37 | −0.16 ± 0.15 | −0.05 ± 0.29 |

%MI refers to the accuracy of the classifier per epochs, %Commands refers to the accuracy of commands per epochs, %Feedback EXO refers to the percentage of epochs with correct feedback from the exoskeleton, TPR is true positive ratio, FPR is false positive ratio and WD is weighted discriminator.

Furthermore, the BCI implemented in this study provided complete and unrestricted control for patients to initiate, maintain, stop, and remain stationary during gait without any predefined time limitations imposed by the experimental protocol or periods of no control over the exoskeleton. This was achieved by designing a dual-state machine control BCI that relied on two distinct models to initiate and terminate exoskeleton movement, resulting in a more realistic and practical approach to the real-time control of an exoskeleton.

## Limitations of the study

This research has some limitations, especially regarding the dataset. All able-bodied participants shared common features. They were in their 20s, presented right-dominance laterality, and did not have previous experience with a BCI. Nevertheless, there were significant differences among subjects in terms of BCI performance. Thus, it was difficult to draw conclusions regarding performance. For future studies, it would be interesting that each subject tries both experimental conditions with a random order and both spaced in

**Table 17. Results from patient P2 in closed-loop trials of experiment 3**

| P2 | Session 2 | Session 3 | Session 4 | Session 5 | Avg. |
|---|---|---|---|---|---|
| %MI | 56.55 ± 12.4 | 47.76 ± 4.46 | 47.76 ± 10.74 | 42.07 ± 11.36 | 48.53 ± 5.98 |
| %Commands | 55.73 ± 21.21 | 48.21 ± 10.48 | 52.99 ± 5.2 | 42.91 ± 9.62 | 49.96 ± 5.64 |
| %Feedback EXO | 54.53 ± 24.34 | 49.91 ± 11.56 | 49.23 ± 4.7 | 43.42 ± 7.53 | 49.27 ± 4.56 |
| Missing trials | 0 | 0 | 0 | 0.2 | 0.05 ± 0.1 |
| Static model: TPR | 90 ± 22.36 | 70 ± 44.72 | 86.67 ± 18.26 | 60 ± 41.83 | 76.67 ± 14.14 |
| Static model: FPR | 56.67 ± 14.91 | 70 ± 18.26 | 73.33 ± 18.07 | 58.33 ± 16.67 | 64.58 ± 8.32 |
| Static model: Accuracy commands | 43.33 ± 9.13 | 33.33 ± 20.41 | 40 ± 12.25 | 35 ± 25.28 | 37.92 ± 4.59 |
| Static model: WD | 0.05 ± 0.18 | −0.22 ± 0.31 | −0.15 ± 0.27 | −0.13 ± 0.29 | −0.11 ± 0.12 |
| Motion model: TPR | 60 ± 36.51 | 63.33 ± 24.72 | 72 ± 15.92 | 66.67 ± 20.41 | 65.5 ± 5.12 |
| Motion model: FPR | 50 ± 35.36 | 73.33 ± 25.28 | 73.33 ± 25.28 | 75 ± 28.87 | 67.92 ± 11.97 |
| Motion model: Accuracy commands | 60 ± 36.51 | 56.67 ± 9.13 | 60 ± 13.69 | 63.33 ± 24.72 | 60 ± 2.72 |
| Motion model: WD | 0.1 ± 0.71 | −0.14 ± 0.24 | −0.09 ± 0.36 | −0.19 ± 0.34 | −0.08 ± 0.13 |

%MI refers to the accuracy of the classifier per epochs, %Commands refers to the accuracy of commands per epochs, %Feedback EXO refers to the percentage of epochs with correct feedback from the exoskeleton, TPR is true positive ratio, FPR is false positive ratio and WD is weighted discriminator.

time. Half of them would be first included in control group and then in VR, and the opposite for the other half. This way, differences among subjects would be mitigated.

Providing the great dependence on subjects, another limitation was the size of the dataset. Although more subjects participated in this study than for most of BCI applications,[3] the database was not vastly large to represent all possible usage outcomes. Consequently, it was difficult to demonstrate which would be the average efficiency of the system. Future research should consider extending the sample size to have a better representation of the whole population.

On the other hand, for training with the VR environment, subjects wore the exoskeleton, and it was blocked in a standing straight position. It involved less physical exertion than walking with the exoskeleton, but it still required some effort, especially for patients. Looking forward, further attempts could evaluate a training phase with VR devices in which participants remain seated.

### Conclusions

Overall, current research compared two training alternatives for an MI based BCI for controlling a lower-limb exoskeleton. One alternative consisted of conventional calibration training with the BCI and the exoskeleton; and the other consisted of a pre-calibration session with a VR environment combined with a shorter conventional calibration phase. Results demonstrate that the employment of a shorter calibration phase which involves less information for the model did not worsen the effectiveness of the BCI. In fact, the performance of the Motion model improved. Moreover, the BCI was evaluated for controlling a lower-limb exoskeleton in ten healthy subjects and two patients with SCI. Usability and acceptance were evaluated in patients. Results were promising for a lower-limb MI based BCI system. Consequently, future research could continue to explore the inclusion of this type of systems with robotic devices in rehabilitation programs.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Equipment
  - BCI
  - Evaluation metrics
- QUANTIFICATION AND STATISTICAL ANALYSIS

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.106675.

## ACKNOWLEDGMENTS

This research is part of grant RTI2018-096677-B-I00, funded by MCIN/AEI/10.13039/501100011033 and by ERDF A way of making Europe; and of grant PID2021-124111OB-C31, funded by MCIN/AEI/10.13039/501100011033 and by ERDF A way of making Europe. Moreover, it was funded by the Ministry of Science, Innovation and Universities through the Aid for the Training of University Teachers FPU19/03165.

## AUTHOR CONTRIBUTIONS

Conceptualization, M.O., E.I., and J.M.A; Methodology, L.F. and V.Q.; Software, L.F. and V.Q.; Validation, L.F. and V.Q., Formal Analysis, L.F.; Investigation, L.F., V.Q., M.O., E.I., A.G.A., and J.M.A.; Resources, A.G.A., and J.M.A.; Data Curation, L.F.; Writing – Original Draft, L.F.; Writing – Review and Editing, L.F., M.O., E.I., and J.M.A.; Visualization, L.F.; Supervision, M.O., E.I., and J.M.A.; Project Administration, J.M.A.; Funding Acquisition, L.F. and J.M.A.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 25, 2022
Revised: March 6, 2023
Accepted: April 11, 2023
Published: April 15, 2023

## REFERENCES

1. Sirlantzis, K., Larsen, L.B., Kanumuru, L.K., and Oprea, P. (2019). 11 - Robotics. In Handbook of Electronic Assistive Technology, L. Najafi and D. Cowan, eds. (Academic Press), pp. 311–345. https://doi.org/10.1016/B978-0-12-812487-1.00011-9.

2. Singh, A., Hussain, A.A., Lal, S., and Guesgen, H.W. (2021). A comprehensive review on critical issues and possible solutions of motor imagery based electroencephalography brain-computer interface. Sensors *21*, 2173. https://doi.org/10.3390/s21062173.

3. He, Y., Eguren, D., Azorín, J.M., Grossman, R.G., Luu, T.P., and Contreras-Vidal, J.L. (2018). Brain-machine interfaces for controlling lower-limb powered robotic systems. J. Neural. Eng. *15*, 021004. https://doi.org/10.1088/1741-2552/aaa8c0.

4. Tariq, M., Trivailo, P.M., and Simic, M. (2018). EEG-based BCI control schemes for lower-limb assistive-robots. Front. Hum. Neurosci. *12*, 312. https://doi.org/10.3389/fnhum.2018.00312.

5. Kwak, N.-S., Müller, K.R., and Lee, S.-W. (2015). A lower limb exoskeleton control system based on steady state visual evoked potentials. J. Neural. Eng. *12*, 056009. https://doi.org/10.1088/1741-2560/12/5/056009.

6. Kwak, N.S., Müller, K.R., and Lee, S.W. (2017). A convolutional neural network for steady state visual evoked potential classification under ambulatory environment. PLoS One *12*, 1–20. https://doi.org/10.1371/journal.pone.0172578.

7. Jeannerod, M. (1995). Mental imagery in the motor context. Neuropsychologia *33*, 1419–1432. https://doi.org/10.1016/0028-3932(95)00073-C.

8. Leeb, R., Friedman, D., Müller-Putz, G.R., Scherer, R., Slater, M., and Pfurtscheller, G. (2007). Self-paced (asynchronous) BCI control of a wheelchair in virtual environments: a case study with a tetraplegic. Comput. Intell. Neurosci. *2007*, 79642. https://doi.org/10.1155/2007/79642.

9. Barria, P., Pino, A., Tovar, N., Gomez-Vargas, D., Baleta, K., Díaz, C.A.R., Múnera, M., and Cifuentes, C.A. (2021). BCI-based control for ankle exoskeleton T-FLEX: comparison of visual and haptic stimuli with stroke survivors. Sensors *21*, 6431. https://doi.org/10.3390/s21196431.

10. Gharabaghi, A. (2016). What turns assistive into restorative brain-machine interfaces? Front. Neurosci. *10*, 456. https://doi.org/10.3389/fnins.2016.00456.

11. Donati, A.R.C., Shokur, S., Morya, E., Campos, D.S.F., Moioli, R.C., Gitti, C.M., Augusto, P.B., Tripodi, S., Pires, C.G., Pereira, G.A., et al. (2016). Long-term training with a brain-machine interface-based gait protocol induces partial neurological recovery in paraplegic patients. Sci. Rep. *6*, 30383–30416. https://doi.org/10.1038/srep30383.

12. Shih, J.J., Krusienski, D.J., and Wolpaw, J.R. (2012). Brain-computer interfaces in medicine. Mayo Clin. Proc. *87*, 268–279. https://doi.org/10.1016/j.mayocp.2011.12.008.

13. Ramoser, H., Müller-Gerking, J., and Pfurtscheller, G. (2000). Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Trans. Rehabil. Eng. *8*, 441–446. https://doi.org/10.1109/86.895946.

14. Lee, K., Liu, D., Perroud, L., Chavarriaga, R., Millán, J.d.R., and del, R. (2017). A brain-controlled exoskeleton with cascaded event-related desynchronization classifiers. Robot. Autonom. Syst. *90*, 15–23. https://doi.org/10.1016/j.robot.2016.10.005.

15. Liu, D., Chen, W., Pei, Z., and Wang, J. (2017). A brain-controlled lower-limb exoskeleton for human gait training. Rev. Sci. Instrum. *88*, 104302. https://doi.org/10.1063/1.5006461.

16. Gordleeva, S., Lukoyanov, M.V., Mineev, S., Khoruzhko, M.A., Mironov, V., Kaplan, A., and Kazantsev, V. (2017). Exoskeleton control system based on motor-imaginary brain–computer interface. Sovrem. Tehnol. Med. *9*, 31. https://doi.org/10.17691/stm2017.9.3.04.

17. Mousavi, M., and de Sa, V.R. (2021). Motor imagery performance from calibration to online control in EEG-based brain-computer interfaces. In 2021 10th International IEEE/EMBS Conference on Neural Engineering (NER), pp. 491–494. https://doi.org/10.1109/NER49283.2021.9441142.

18. Parashiva, P.K., and Vinod, A.P. (2021). Online Hand Motor Imagery Direction Decoding Using Brain Computer Interface, pp. 17–20.

19. Kline, A., Gaina Ghiroaga, C., Pittman, D., Goodyear, B., and Ronsky, J. (2021). EEG differentiates left and right imagined Lower Limb movement. Gait Posture *84*, 148–154. https://doi.org/10.1016/j.gaitpost.2020.11.014.

20. Ang, K.K., Chin, Z.Y., Zhang, H., and Guan, C. (2008). Filter Bank common spatial pattern (FBCSP) in brain-computer interface. In Proceedings of the International Joint Conference on Neural Networks, pp. 2390–2397. https://doi.org/10.1109/IJCNN.2008.4634130.

21. Rupp, R. (2014). Challenges in clinical applications of brain computer interfaces in individuals with spinal cord injury. Front. Neuroeng. *7*, 38. https://doi.org/10.3389/fneng.2014.00038.

22. Höller, Y., Thomschewski, A., Uhl, A., Bathke, A.C., Nardone, R., Leis, S., Trinka, E., and Höller, P. (2018). HD-EEG based classification of motor-imagery related activity in patients with spinal cord injury. Front. Neurol. *9*, 955. https://doi.org/10.3389/fneur.2018.00955.

23. Formaggio, E., Masiero, S., Bosco, A., Izzi, F., Piccione, F., and Del Felice, A. (2017). Quantitative EEG evaluation during robot-assisted foot movement. IEEE Trans. Neural Syst. Rehabil. Eng. *25*, 1633–1640. https://doi.org/10.1109/TNSRE.2016.2627058.

24. Yu, Z., Li, L., Song, J., and Lv, H. (2018). The study of visual-auditory interactions on lower limb motor imagery. Front. Neurosci. *12*, 509. https://doi.org/10.3389/fnins.2018.00509.

25. Rodríguez-Ugarte, M., Iáñez, E., Ortiz, M., and Azorín, J.M. (2018). Improving real-time lower limb motor imagery detection using tDCS and an exoskeleton. Front. Neurosci. *12*, 757.

26. Ortiz, M., Ferrero, L., Iáñez, E., Azorín, J.M., and Contreras-Vidal, J.L. (2020). Sensory integration in human movement: a new brain-machine interface based on gamma band and attention level for controlling a lower-limb exoskeleton. Front. Bioeng. Biotechnol. *8*, 735. https://doi.org/10.3389/fbioe.2020.00735.

27. Ferrero, L., Quiles, V., Ortiz, M., Iáñez, E., and Azorín, J.M. (2021). A bmi based on motor imagery and attention for commanding a lower-limb robotic exoskeleton: a case study. Appl. Sci. *11*, 4106. https://doi.org/10.3390/app11094106.

28. Choi, J., Kim, K.T., Jeong, J.H., Kim, L., Lee, S.J., and Kim, H. (2020). Developing a motor imagery-based real-time asynchronous hybrid BCI controller for a lower-limb exoskeleton. Sensors *20*, 7309. https://doi.org/10.3390/s20247309.

29. Do, A.H., Wang, P.T., King, C.E., Chun, S.N., and Nenadic, Z. (2013). Brain-computer interface controlled robotic gait orthosis. J. NeuroEng. Rehabil. *10*, 111.

30. Kilicarslan, A., Prasad, S., Grossman, R.G., and Contreras-Vidal, J.L. (2013). High accuracy decoding of user intentions using

EEG to control a lower-body exoskeleton. In Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society, *2013*Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE), pp. 5606–5609. https://doi.org/10.1109/EMBC.2013.6610821.

31. Kilicarslan, A., Grossman, R.G., and Contreras-Vidal, J.L. (2016). A robust adaptive denoising framework for real-time artifact removal in scalp EEG measurements. J. Neural. Eng. *13*, 026013. https://doi.org/10.1088/1741-2560/13/2/026013.

32. López-Larraz, E., Trincado-Alonso, F., Rajasekaran, V., Pérez-Nombela, S., del-Ama, A.J., Aranda, J., Minguez, J., Gil-Agudo, A., and Montesano, L. (2016). Control of an ambulatory exoskeleton with a brain–machine interface for spinal cord injury gait rehabilitation. Front. Neurosci. *10*, 359.

33. Rajasekaran, V., López-Larraz, E., Trincado-Alonso, F., Aranda, J., Montesano, L., Del-Ama, A.J., and Pons, J.L. (2018). Volition-adaptive control for gait training using wearable exoskeleton: preliminary tests with incomplete spinal cord injury individuals. J. NeuroEng. Rehabil. *15*, 1–15. https://doi.org/10.1186/s12984-017-0345-8.

34. Malouin, F., Jackson, P.L., and Richards, C.L. (2013). Towards the integration of mental practice in rehabilitation programs. A critical review. Front. Hum. Neurosci. *7*, 576. https://doi.org/10.3389/fnhum.2013.00576.

35. Zhang, X., She, Q., Chen, Y., Kong, W., and Mei, C. (2021). Sub-band target alignment common spatial pattern in brain-computer interface. Comput. Methods Progr. Biomed. *207*, 106150. https://doi.org/10.1016/j.cmpb.2021.106150.

36. Ferrero, L., Quiles, V., Ortiz, M., Juan, J.V., Iáñez, E., and Azorín, J.M. (2022). Inter-session transfer learning in MI based BCI for controlling a lower-limb exoskeleton. In Bio-inspired Systems and Applications: from Robotics to Ambient Intelligence, J.M. Ferrández Vicente, J.R. Álvarez-Sánchez, F. de la Paz López, and H. Adeli, eds. (Springer International Publishing), pp. 243–252.

37. He, H., and Wu, D. (2020). Transfer learning for brain-computer interfaces: a euclidean space data alignment approach. IEEE Trans. Biomed. Eng. *67*, 399–410. https://doi.org/10.1109/TBME.2019.2913914.

38. Borg, G. (1970). Perceived exertion as an indicator of somatic stress. Scand. J. Rehabil. Med. *2*, 92–98.

39. Hart, S.G., and Staveland, L.E. (1988). Development of NASA-TLX (task Load Index): results of empirical and theoretical research. In Human Mental Workload, P.A. Hancock and P. Meshkati, eds. (North-Holland), pp. 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9.

40. Demers, L., Weiss-Lambrou, R., and Ska, B. (2002). The Quebec user evaluation of satisfaction with assistive Technology (QUEST 20): an overview of recent progress. Technol. Disabil. *14*, 101–105. https://doi.org/10.13072/midss.298.

41. Pfurtscheller, G., Brunner, C., Schlögl, A., and Lopes da Silva, F.H. (2006). Mu rhythm ( de ) synchronization and EEG single-trial classification of different motor imagery tasks. Neuroimage *31*, 153–159. https://doi.org/10.1016/j.neuroimage.2005.12.003.

42. Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods *134*, 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009.

43. Torkamani-Azar, M., Jafarifarmand, A., and Cetin, M. (2020). Prediction of motor imagery performance based on pre-trial spatio-spectral alertness features. In Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, *2020*, pp. 3062–3065. https://doi.org/10.1109/EMBC44109.2020.9175929.

44. Padfield, N., Zabalza, J., Zhao, H., Masero, V., and Ren, J. (2019). EEG-based brain-computer interfaces using motor-imagery: techniques and challenges. Sensors *19*, 1423–1434. https://doi.org/10.3390/s19061423.

45. Rodríguez-Ugarte, M., Iáñez, E., Ortíz, M., and Azorín, J.M. (2017). Personalized offline and pseudo-online BCI models to detect pedaling intent. Front. Neuroinf. *11*, 45. https://doi.org/10.3389/fninf.2017.00045.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Software and algorithms | | |
| MATLAB 2018a | MathWorks, USA | https://es.mathworks.com/products/matlab.html |
| Rstudio 1.4.1106 | Posit, USA | https://posit.co/downloads/ |
| R programming language 1.0.2 | R Foundation, USA | https://www.r-project.org/ |
| NIC V2.0.5c | Neuroelectrics, Spain | https://neuroelectrics.com/resources/software |
| FBCSP | Ang et al.[20] | |

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, José M. Azorín (jm.azorin@umh.es).

#### Materials availability

This study did not generate new unique reagents.

#### Data and code availability

All data reported in this paper will be shared by the lead contact upon request. This paper does not report original code. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

### EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Ten able-bodied subjects participated in the study (mean age, 23.5 $\pm$ 2.0). Five females and five males. They had no movement impairment and did not report any known disease. They were informed about the experiments and signed an informed consent form in accordance with the Helsinki declaration. All procedures were approved by the Responsible Research Office of Miguel Hernández, University of Elche Spain (DIS.JAP.03.18, 22/01/2019). Five participants were randomly included in Control group and the other five in the VR group.

In addition, two patients with spinal cord injury (SCI) were recruited from the National Hospital of Paraplegics in Toledo. They also signed an informed consent form in accordance with the Helsinki Declaration and had no prior experience with BCI technology. Further patient's details are specified in Table 1.

### METHOD DETAILS

#### Equipment

EEG signals were recorded with Starstim R32 (Neuroelectrics, Spain) at 500 Hz. The placement of 27 wet electrodes followed the distribution of the international 10-10 system, with electrode positions including F3, FZ, F4, FC5, FC3, FC1, FCZ, FC2, FC4, FC6, C5, C3, C1, CZ, C2, C4, C6, CP5, CP3, CP1, CPZ, CP2, CP4, CP6, P3, PZ, P4. Ground and reference electrodes were positioned on the right ear lobe. NIC v2.0.5c software (Neuroelectrics, Spain) was used to record EEG in the laptop.

All participants wore H3 exoskeleton (Tecnhnaid, Spain). It is a powered hip-knee-ankle exoskeleton that emulates the human walking. Participants were provided with crutches or a trolley to ensure standing and walking stability. To prevent any risk of falls, a technician also held the exoskeleton from the back. High-level control commands were transmitted from the computer via Bluetooth to initiate and stop the gait at a constant speed.

Virtual reality equipment, including an HTC VIVE headset (HTC, Taiwan) with a resolution of 2160 x 1200 (1080 x 1200 per eye) and a refresh rate of 90 Hz, two base stations to track the headset's location, Steam software (Valve, United States), and a self-developed environment created with Unity engine (Unity Technologies, United States) were used in the study. The equipment setup is illustrated in Figure 1.

## BCI

This section describes the analysis performed to decode mental tasks from EEG data. MATLAB 2018a was used to connect through TCP/IP with NIC v2.0.5c software and get EEG data, analyze it and send control commands to the exoskeleton. While the exoskeleton was controlled in closed-loop, a decision was taken every 0.5 s. This means that every 0.5 s, the BCI could decide to stay static, start walking, keep on walking, or stop the gait. The epochs of analysis were 1 s length with a shifting of 0.5s. Therefore, once 0.5 s of current data were recorded, epochs of 1s had to be pre-processed, processed, classified, and translated into a control instruction before new 0.5 s of data were available.

The analysis of trials in opened-loop control, such as VR and calibration phases, was performed simulating a real-time scenario (pseudo-online analysis). Consequently, although the analysis was performed afterward, it was done with epochs of data of 1 s and 0.5 s of overlapping. Figure S1 shows the overall schema of the BCI.

### *Pre-processing*

The first step of the BCI consisted of increasing the signal to noise ratio. Epochs of 1 s of EEG were filtered with a Notch filter at 50 Hz that mitigated the contribution of the power line, and a high-pass filter at 0.1 Hz for reducing DC offset. The Notch filter was directly applied in the NIC recording software and the high-pass filter was applied with state-variables Butterworth function in MATLAB.

### *Processing*

The following step focused on extracting signal features that could be used to discriminate between different mental patterns. Our BCI used filter-bank common spatial patterns (FBCSP).[20] Firstly, signals were band-pass filtered into different frequency bands: 5–10, 10–15, 15–20, 20–25 Hz. Then, common spatial patterns (CSP) were extracted from each band. CSP estimated a spatial filter that maximized the discriminability between two classes.

$X_1 \sim N \cdot T$ and $X_2 \sim N \cdot T$ are EEG data from two different brain patterns. N is the number of channels and T is the number of samples. In CSP algorithm, their covariance matrices were computed as:

$$C_1 = \frac{X_1 X_1^T}{trace\left(X_1 X_1^T\right)}, C_2 = \frac{X_2 X_2^T}{trace\left(X_2 X_2^T\right)} \quad \text{(Equation 3)}$$

Covariance matrices were calculated independently for each trial and then, averaged. Afterwards, they were combined and factorized:

$$C = \overline{C_1} + \overline{C_2} = U_0 \sum U_0^T \quad \text{(Equation 4)}$$

$\Sigma$ is the diagonal matrix of eigenvalues and $U_0$ is a matrix of eigenvectors. Subsequently, results from previous factorization were used to transform each averaged covariance matrix:

$$P = \Sigma^{\frac{1}{2}} U_0^T \quad \text{(Equation 5)}$$

$$S_1 = P\overline{C_1}P^T, S_2 = P\overline{C_2}P^T \quad \text{(Equation 6)}$$

As a result of Equation 6, $S_1$ and $S_2$ had the same matrix of eigenvectors and the addition of the two matrices of eigenvalues resulted in the identify matrix.

$$S_1 = U\Sigma_1 U^T, S_2 = U\Sigma_2 U^T, \Sigma_1 + \Sigma_2 = I. \quad \text{(Equation 7)}$$

$U$ and $P$ were employed to estimate the spatial transformation matrix, $W$.

$$W = U^T P. \quad \text{(Equation 8)}$$

The original data was transformed with $W$. Resulting $Z$ had the same dimensions as the original data, N·T, but new channels were ordered based on their discriminative power. Top and last channels were more discriminative in terms of variance and middle ones were less discriminative.

$$Z = WX \quad \text{(Equation 9)}$$

For dimensionality reduction, only $m = 3$ first and last discriminant rows of Z were selected as features, $Z_p$. In addition, the variance of each dimension of $Z_p$ was calculated and log-normalized.

$$f_p = log\frac{var(Z_p)}{\sum_{i=1}^{2m} Z_p} \quad \text{(Equation 10)}$$

### *Classification*

As explained in the previous subsection, a vector of features was obtained for each epoch of data. Then, a linear discriminant analysis (LDA) was used as the classifier. It was trained with epochs of class 'Idle state' and 'MI of gait', and it predicted these two classes. Two models were created independently for the static and motion trials.

*Calibration phase.* Trials from calibration phase in the three experiments were evaluated with leave-one-out cross-validation. Moreover, trials in static and trials in motion were assessed independently.

In this study, two approaches were compared, control and VR group (experiment 1 and 2). Therefore, for having a fair comparison, the classifier was trained with the same number of trials. For each subject and session.

- If the subject was in control group:
  - o Leave-one-out cross-validation with last six training trials in static.
  - o Leave-one-out cross-validation with last six training trials in motion.
- If the subject was in VR group:
  - o Leave-one-out cross-validation leave-one-out with all six training trials in static.
  - o Leave-one-out cross-validation with all six training trials in motion.

For both conditions, one trial was selected for evaluation while the remaining five trials were used for training the classifier model. This process was repeated for each of the six trials, so that each trial was used for evaluation once. The accuracy of the BCI system was then calculated as the average of the six cross-validation iterations for both the static and motion conditions.

*Closed-loop control phase.* Two LDA classifiers were trained using the data from the calibration phase for the static and motion conditions, respectively. These classifiers were denoted as Static and Motion. Subsequently, based on the status of the exoskeleton, either stationary or in motion, one of the classifiers was used during each epoch and it predicted either 'Idle state' or 'MI of the gait'. More information about this process is provided in section 2.3 and experiment 1. A depiction of this dual-state machine BCI can be seen in Figure 4.

### *Output commands*

There was an additional step in the BCI for the closed-loop control of an exoskeleton that consisted of translating predictions from the classifier into control instructions. The classifier provided predictions with a value of 1 for epochs labeled as 'MI of the gait' and 0 for epochs labeled as 'Idle state'. To minimize the number of false positives and false negatives, these predictions were smoothed using a moving mean of 4 s. The resulting values ranged from 0 to 1, with values above a'activation threshold' triggering a command to initiate the gait. Additionally, during movement, values below the 'deactivation threshold' would trigger a command to stop the exoskeleton's motion.

To determine the appropriate thresholds for activation and deactivation, calibration data was utilized. Specifically, the results from a leave-one-out cross-validation with six static trials were used to determine the

'activation threshold', while the results from a leave-one-out cross-validation with six motion trials were used to determine the 'deactivation threshold'.

To optimize the 'activation threshold, predictions from six static trials were utilized. These predictions were first smoothed using a moving mean of 4 s. To identify the optimal activation threshold, three plateaus were identified from each static trial: the first idle state, MI, and second idle state periods. A plateau was defined as consecutive points with a slope near 0, ranging between −0.05 and 0.05. Figure S2 demonstrates an example of a trial and the three plateaus that were identified. In cases where more than one plateau was found for each task, the highest one was selected.

Once the plateaus were identified, the intermediate point between the plateau of the first idle state and the plateau of MI was computed and defined as the single trial 'activation threshold' as can be seen in Figure S2. This process was repeated for all six static trials. Outliers were subsequently removed, and the remaining resting values were averaged to identify a reliable activation threshold that could be utilized in the closed-loop control of the exoskeleton.

Similarly, the process was repeated using results from motion trials to define the 'deactivation threshold'. In this case, the threshold was defined as the intermediate point between the plateau of MI and the plateau of the second idle state period. Once again, outliers were removed, and resting values were averaged to obtain a reliable deactivation threshold that could be utilized in the closed-loop control of the exoskeleton.

As the last step, some rules were defined to avoid overfitting and to correlate both thresholds:

- 'Activation threshold' could not be lower than the average prediction value during first idle state period (in static trials).
- 'Deactivation threshold' could not be higher than the average prediction value during MI period (in trials in motion).
- 'Activation threshold' could not be higher than the average prediction value during MI period (in trials in motion).

## Evaluation metrics

### *Calibration phase*

The accuracy was computed as the percentage of epochs with correct predictions (%MI). This value was estimated for each trial without considering the initial task employed for the convergence of the algorithms.

### *Closed-loop control phase*

The following metrics were computed for evaluation.

- **%MI:** percentage of epochs with correct predictions.
- **%Commands:** percentage of epochs with correct commands issued.
- **%Feedback EXO:** percentage of epochs with correct feedback from the exoskeleton.
- **Missing trials:** percentage of trials in which the exoskeleton never activated during MI period.
- Static model
  - **TPR (true positive ratio):** ratio of how many activation commands were issued during MI period out of how many times subjects were static.
  - **FPR (false positive ratio):** ratio of how many activation commands were issued during idle state period of how many times subjects were static.
  - **Accuracy commands:** percentage of correct activation commands issued out of all activation commands.
  - **WD (Weighted discriminator):** this metric was introduced in,[45] it ranges from −1 to 1 and it is calculated as:

$$WD = 0.4\frac{TPR}{100} + 0.6\frac{\text{Accuracy commands}}{100} - \frac{FPR}{100} \quad \text{(Equation 11)}$$

- Motion model
  - o **TPR:** ratio of how many deactivation commands were issued during idle state period out of how many times subjects were moving.
  - o **FPR:** ratio of how many deactivation commands were issued during MI period of how many times subjects were moving.
  - o **Accuracy commands:** percentage of correct deactivation commands issued out of all deactivation commands.
  - o **WD (Weighted discriminator):** it is calculated as detailed above but with metrics corresponding to Motion model.

## QUANTIFICATION AND STATISTICAL ANALYSIS

To determine if there were any noteworthy distinctions between the VR group and the control group in calibration data, a statistical analysis was conducted with Rstudio 1.1.441 and R programming language 4.0.2. The variability introduced by the grouping was not the only factor that could affect the results; the experimental sessions could also have an impact. To account for this, a two-way ANOVA test was performed with accuracy (%MI) as the dependent variable and group and session as independent variables, accuracy ~ group·session. In order to obtain reliable results from the ANOVA test, the dataset must satisfy three assumptions: normality, equality of variances, and independence. Two separate analyses were conducted for static and trials in motion. Saphiro test was employed to test normality with saphiro_test() function and data was organized by group and session variable with group_by(). Equality of variances was evaluated with Barlett test and bartlett.test() function.

- Static trials: Sincedata did not follow a normal distribution, Kruskal-Wallis test was employed instead of ANOVA with kruskal.test() function. No significant differences were found neither among sessions nor among groups (pvalue>0.05).
- Trials in motion: All assumptions were not met, so Kruskal-Wallis test was conducted. It showed that significant superior results were seen for VR group (pvalue<0.05). Differences among sessions were significant (p value<0.05). In addition, the interaction between group and sessions was significant. It means the evolution of accuracy through sessions varies from VR to control group.

Furthermore, to determine the significance of inter-subject differences in accuracy, a separate ANOVA test was performed for each trial type using accuracy as the dependent variable and subjects as the independent variable, accuracy ~ subjects.

- Static trials: Data satisfied all requirements and the results of the ANOVA test showed significant differences among subjects (p value<0.05). The test was performed with aov().
- Trials in motion: Requirements were also met and the results from the ANOVA test showed significant differences among subjects (p value<0.05) too. Results could be influenced by the fact that subjects had a different training, and it was previously proven that VR group tended to have higher results. Therefore, two independent ANOVA tests were conducted for the control and VR subjects. In the control group, there were three sub-groups of subjects that showed a significant different behavior: S1 and S2; S3 and S4; and S5. A similar trend was found in the VR group with three sub-groups: V1 and V3; V2 and V4; V5.

Intending to assess the evolution of performance through sessions, metrics were normalized with respect to the first session. This way, it can be studied how the accuracy evolved. New values of accuracy were computed as:

$$\%\widehat{MI}_{i,j} = \frac{\%MI_{i,j} - \%MI_1}{\%MI_1} \quad \text{(Equation 12)}$$

$\%MI_{i,j}$ is the accuracy obtained for a trial $j$ in an experimental session $i$, and $\%MI_1$ is the average accuracy of all trials of the first session. This value ranges from −1 to 1.

Figure 9 shows the average $\%\widehat{MI}$ that goes from −1 to 1. The two-way ANOVA accuracy∼ group·session was repeated with this new metric for static and trials in motion.

- Static trials: Data did follow a normal distribution, but the variance was not equal for all groups. Thus, Kruskal-Wallis test was conducted and no differences were found among sessions or the VR and control group (p value>0.05).
- Motion trials: As the requisite of equal variances was violated, Kruskal-Wallis test was employed. It found relevant differences among sessions and the interaction between sessions and groups was significant. Moreover, the results from control group were substantially better than the ones of the VR group (p value<0.05).

Lastly, an assessment was made of inter-subject differences in normalized accuracy, $\%\widehat{MI}$. As the distribution of normalized accuracy was non-normal, a Kruskal-Wallis test was utilized. Significant differences were observed among subjects in both static and motion trials (p value<0.05).