Closing the Sensorimotor Loop: Haptic Feedback Facilitates Decoding of Motor Imagery

M. Gomez-Rodriguez[1,2], J. Peters[1], J. Hill[1], B. Schölkopf[1], A. Gharabaghi[3] and M. Grosse-Wentrup[1]

[1] MPI for Biological Cybernetics, Tübingen, Germany

[2] Department of Electrical Engineering, Stanford University, Stanford, CA, USA

[3] Werner Reichardt Centre for Integrative Neuroscience and Department of Neurosurgery, University Hospital, Eberhard Karls University Tübingen, Germany

E-mail: manuelgr@stanford.edu, jan.peters@tuebingen.mpg.de, jeremy.hill@tuebingen.mpg.de, bernhard.schoelkopf@tuebingen.mpg.de, alireza.gharabaghi@uni-tuebingen.de, moritzgw@ieee.org

Abstract. The combination of Brain-Computer Interfaces (BCIs) with robot-assisted physical therapy constitutes a promising approach to neurorehabilitation of patients with severe hemiparetic syndromes caused by cerebrovascular brain damage (e.g., stroke) and other neurological conditions. In such a scenario, a key aspect is how to reestablish the disrupted sensorimotor feedback loop. However, to date it is an open question how artificially closing the sensorimotor feedback loop influences the decoding performance of a BCI. In this article, we answer this issue by studying six healthy subjects and two stroke patients. We present empirical evidence that haptic feedback, provided by a seven degrees-of-freedom robotic arm, facilitates on-line decoding of arm movement intention. The results support the feasibility of future rehabilitative treatments based on the combination of robot-assisted physical therapy with BCIs.

Submitted to: *J. Neural Eng.*

1. Introduction

In the past two decades, research on Brain-Computer Interfaces (BCIs) has evolved from pioneering feasibility studies [1, 2, 3] to a state in which basic communication can be routinely performed after only brief calibration periods with healthy subjects [4, 5] as well as with subjects in early stages of amyotrophic lateral sclerosis (ALS) [6]. Although communication with completely locked-in subjects in late stages of ALS still remains a challenge, this substantial progress has resulted in a growing interest in extending the application domain of BCIs from communication towards control of external actuators and prosthetic devices – even restoration of basic motor functions appears possible by means of a BCI [7, 8]. For example, EEG-based control of an electric wheelchair has been reported in [9], and the feasibility of controlling a mobile robot by means of a non-invasive BCI has been demonstrated in [10].

However, most studies in this field only consider replacing dysfunctional body parts by BCI-controlled artificial actuators. Instead, BCIs may also help to directly facilitate rehabilitation of body parts impaired by neurological conditions such as stroke [11]. While traditional or robot-assisted physical therapy constitutes a key ingredient to rehabilitation after stroke [12, 13], motor imagery has also been shown to have beneficial effects in stroke rehabilitation [7, 8]. Furthermore, successful MEG-based decoding of motor imagery in chronic stroke patients has been demonstrated in [14]. The logical next step is to combine these approaches into an integrated stroke therapy, in which patients exert control over robot-assisted physical therapy through the decoding of movement intentions using a BCI. Such an integrated therapy may have a large impact on stroke rehabilitation, as the synchronization of robot-assisted physical therapy with movement intention is likely to result in increased cortical plasticity due to Hebbian-type learning rules [15, 16, 17].

A key aspect of such a stroke therapy is re-establishing the disrupted sensorimotor feedback loop, *i.e.*, temporarily bypassing the impaired movement execution of stroke

patients through robot-assisted physical therapy controlled by a BCI. Importantly, the effect of artificially closing the sensorimotor feedback loop on BCI-decoding has not yet been studied. It is well known that passive movements [18, 19] as well as active arm movements [20] induce patterns in the electromagnetic field generated by the brain similar to those observed during motor imagery [3]. Moreover, random haptic stimulation has been shown to be beneficial for decoding motor imagery [21]. Nevertheless, it remains an open question how the electromagnetic field generated by the brain changes in response to artificially closing the sensorimotor feedback loop, *i.e.*, by providing sensory feedback on the intended movement without actual movement execution, and whether the resulting feedback processes are beneficial or disadvantageous for decoding movement intentions.

In this article, we study how artificially closing the sensorimotor feedback loop influences BCI decoding of arm movement intention in six healthy subjects and two stroke patients. Specifically, each subject performed motor imagery of arm extension and flexion while being attached to a robot arm. Simultaneously, we performed on-line decoding of movement intention with an EEG-based BCI, and employed the robot arm to move the subject's arm according to the inferred movement intention, *i.e.*, haptic feedback was provided (Section 2). The haptic feedback was provided in a synchronized manner with the subject's intention to move the arm, and not after the motor imagery, as done previously [22]. This is a key point in terms of a future rehabilitation therapy, because it is a strategy that is likely to result in increased cortical plasticity due to Hebbian-type learning rules [15, 16, 17]. By using a block-wise design, we compared subject performance with and without robot-induced movement execution, and we provide evidence that artificially closing the sensorimotor loop substantially increases the decoding accuracy (Section 3). Thereby, our results demonstrate that closing the sensorimotor feedback loop is not only feasible, but even facilitates decoding of movement intention ‡.

2. Materials and Methods

2.1. Human subjects

Six right-handed healthy subjects, four females and two males, and two right-handed stroke patients with a hemiparesis of the left side of their body, a female and a male, took part in this study. Both stroke patients suffered an ischemic stroke in the right hemisphere, in 2003 and 2004. A clinical examination in both stroke patients determined they have intact proprioception. None of the healthy subjects had previous experience with motor imagery. Both stroke patients had been trained in motor imagery of the left hand for a study unrelated to this work. All subjects participated in one session, in which they experienced all the conditions of haptic feedback explained in Section 2.4.

‡ Please note that a preliminary version of this work that does not include the data from the stroke patients has been presented in [23]

Each subject gave informed consent prior to participation and the experiments were in compliance with national legislation and the Code of Ethical Principles for Medical Research Involving Human Subjects of the World Medical Association (Declaration of Helsinki).

2.2. Recording

An EEG electrode cap (by Electro-Cap International, Inc.) in combination with a Quickamp (by Brain Products, GmbH) amplifier were used during our experiments. 35 channels of the EEG electrode cap were fed into the amplifier, with a sampling rate $f_s = 250$ Hz, a built-in common average reference (CAR) montage and a built-in lowpass digital FIR filter with cutoff frequency 67.5 Hz (*i.e.*, $0.27 \times f_s$). The electrode AFz was used as the ground electrode. All electrode impedances were kept below 10 kΩ. The electrode array covered parts of the pre-motor cortex, primary motor cortex and somatosensory cortex as well as several other areas, as shown in Figure 4. EEG signals were acquired from the amplifiers using the general-purpose BCI2000 software [24], and the additional module BCPy2000 [25] was used for on-line signal processing, statistical learning, and transmission of the control signal.

2.3. Experimental paradigm

[Figure 1 about here.]

The subjects' task was to imagine movement of the forearm. Healthy subjects were instructed to imagine the movement with the right forearm while stroke patients were instructed to imagine it with the left forearm (their hemiparesis was on the left side of the body). Every subject participated in one session, divided in eight blocks. A block consisted of 30 consecutive trials separated from each other by 2.5s. In each trial, the subject was instructed to imagine flexion (15 trials per block) or extension (15 trials per block) of the forearm, using the elbow joint as the single degree of freedom (see Figure 1). A trial was divided into three consecutive phases. At the beginning of each phase, cues were delivered visually as text displayed on a computer screen. In addition, each textual cue was spoken out loud as an auditory cue. Both, computer screen and loudspeaker, were located approximately 1.5m in front of the subject. Each trial started with the instruction "Relax". After 3s, subjects were cued on the upcoming type of motor imagery ("Flexion" or "Extension"). After further 3s, subjects were instructed to initiate motor imagery by a "Go"-cue. This phase of motor imagery typically lasted for 5s§. For a better understanding, Figure 2 illustrates the timeline of events during one trial.

Every block was divided in two consecutive sections, a training section and a test section. The training section was sufficiently long to collect 25 seconds of data for each type of movement (*i.e.*, flexion and extension) and for rest. Note that the 25 seconds

§ If the robot arm hit its maximum range, the period ended.

of data for each type of movement and rest were not collected in a consecutive manner but in each period of movement and rest until summing up to a total of 25 seconds. By default, a training section lasted 10 trials (5 trials for each type of movement). Nevertheless, if the robot hit the range limitations at least once during the training section (*i.e.*, one or more of the motor imagery periods were shorter than 5s), the training section lasted more than 10 trials. The test section lasted the remaining number of trials following the training section, *i.e.* 20 trials or less.

In the test section, visual feedback was given in form of an arrow on the computer screen during the movement periods. At the beginning of the movement period of each trial, the arrow was located in the center of the screen and its position was updated every 300ms according to the decoded movement intention (see Section 2.6 for more details regarding the on-line decoding). In order to study the influence of haptic feedback, we had four different conditions (two blocks per condition and session). The robot arm moved the subject's arm during the movement periods of the training section only (Condition II), of the test section only (Condition III), of both sections (Condition I) or none (Condition IV). Please note that during the training section, the robot arm moved according to the instruction provided to the subject while during the test section, the robot arm moved according to the decoded movement intention (see Section 2.4 for more details regarding the conditions and the robot movements).

Trials for both types of movement directions were interleaved randomly in a way that the movement direction could not be inferred a priori in both the training section and the test section‖

[Figure 2 about here.]

2.4. Robotic-based haptic feedback

[Figure 3 about here.]

A WAM robot arm (by Barrett Technology, Inc.) was used to provide haptic feedback during our experiments. This arm was attached to the subject's forearm (see Figure 3). It was programmed to only move the joint that mimics the elbow. Note that we aim to show how haptic feedback influences on-line decoding. A single degree of freedom (DoF) movement suffices for this purpose and was hence chosen for simplicity. The resting position, maximum extension and maximum flexion of the robot's elbow joint were adjusted to the preferences of each subject using the passive, fully compliant mode of the WAM robot arm. We chose an angular velocity for the elbow joint movement of the robot of 0.096 rad/s (5.5 deg/s). Given that the subject's hand was approximately 40 cm from the joint, the subject's hand was moved with a linear velocity of 0.096 rad/s $\times$ 0.4 m = 3.84 cm/s.

‖ The random interleaving of movement directions was violated if the robot arm hit its maximum range as it required the next movement to be in the opposite direction. We cannot not know beforehand if the robot arm will hit its maximum range given the random interleaving of movement directions.

In order to explore the influence of haptic feedback, we introduced four different conditions (I–IV). In Condition I, the robot moved the subject's arm during the training section according to the instruction and during test section according to the decoded intention. In Condition II, the robot moved the subject's arm only during the training section according to the instruction. In Condition III, the robot moved the subject's arm only during the test section according to the decoded intention. In Condition IV, the robot did not move the subject's arm at all (see Table 1). The conditions for each subject were interleaved randomly to have a fair comparison among conditions, avoiding effects caused by training, fatigue or changes in attention over the course of an experiment.

During the movement periods of the training sections in Conditions I and II, the robot moved the arm (i.e., flexion or extension) in a continuous manner according to the instruction given to the subject. During the movement periods of the test sections, we decoded every 300ms whether a movement intent was present, and updated the state of the robot arm (moving/resting) accordingly. The haptic feedback was given during each movement period in a synchronized manner with the subject's intention to move the arm, and not at the end of every trial as done previously [22]. As stated previously, moving the robot throughout each trial, and not at the end of it, is a key point in terms of a future neurorehabilitation therapy, because it is a strategy that is likely to result in increased cortical plasticity due to Hebbian-type learning [15, 16, 17].

Finally, note that if the robot arm hit its maximum range during a movement period, the movement period ended. In this case, the random interleaving of movement directions was violated and the next movement period was forced to be in the opposite direction.

2.5. Signal analysis

First, a centre-surround spatial sharpening filter or *surface Laplacian* [26] was applied to the electrodes that were not located at the borders. Then, the signals of all electrodes were band pass filtered (2–115 Hz). As 20 online features per electrode, we used the normalized average power spectral densities in 2 Hz frequency bins in the frequency range (2–42 Hz), as previously used in motor imagery with EEGs [4]. Welch's method was used to compute an estimate of the power spectral density (PSD). Welch's method divided the time signal into time windows of length 500ms with a 200ms overlap.

The estimate was determined based on incrementally bigger time segments. In each movement or resting period, the first PSD estimate was computed using 500ms and from that moment on every 300 ms, allowing an on-line classification every 300ms (see Section 2.6). Please note that the second estimate based on 800ms fits exactly two 500ms windows with 200ms overlap (second window starts at 300ms), the third estimate based on 1100ms fits three 500ms windows with 200ms overlap (second window starts at 300ms and third window starts at 600ms), and so on. As the movement or resting period advanced, the PSD was estimated from larger segments (*i.e.*, all the signals of

the period up to that point), leading to more reliable and less noisy estimates.

2.6. On-line decoding

On-line classification was carried out to discriminate flexion vs resting, and to discriminate extension vs resting. As explained in Sections 2.3 and 2.4, on-line visual feedback and haptic feedback were provided. For each block, two linear support vector machine (SVM) classifiers [27] were generated on-the-fly after the training section. As stated before, a training section consisted by default of 10 trials. However, if the robot hit the range limitations at least once, the training section lasted more trials. We had multiple time windows per movement period (*i.e.*, (5s-500ms)/300ms = 15) and resting period (*i.e.*, (3s-500ms)/300ms = 8). Each SVM was trained with 75 movement samples and 75 resting samples (note that identical resting samples were used for both classifiers and we had 5 spare resting samples that we did not use). Both classifiers were used only in the test section following the training section within the same block. One classifier distinguished between flexion vs resting and the other one classified between extension vs resting. Due the 20 frequency bins for each of the 35 recorded channels, we had a 700-dimensional feature vector for each data point.

The labels $y_i \in \{-1, 1\}$ indicate resting or movement (both flexion and extension). Hence, during the training section we obtain a set of tuples $\{(\mathbf{x_i}, y_i),\ i = 1 \ldots n\}$. As classifier, a linear SVM classifier $(\mathbf{w}, b)$ was determined by solving the following convex optimization problem for each of the two cases (*i.e.*, flexion vs resting, and extension vs resting) given by

$$\min_{\mathbf{w},\epsilon} \quad \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{n}\epsilon_i, \tag{1}$$

$$\text{subject to} \quad y_i(\mathbf{w}\cdot\mathbf{x_i} - b) \geq 1 - \epsilon_i, \tag{2}$$

$$\epsilon_i \geq 0. \tag{3}$$

Linear SVM classifiers are characterized by providing decision *hyperplanes* that maximize the *margin* between the two classes in the feature space. The slack variables ϵ_i allow for mislabeled examples or outliers. The regularization parameter C was estimated by 10-fold cross-validation within the training set [28]. The outputs of each classifier over the training set were used to fit the parameters of a sigmoid function that mapped the classifier outputs z_i onto probabilities, $p(z_i)$. An output probability $p(z_i) = 1$ indicated 100% confidence when predicting movement (flexion or extension) and $p(z_i) = 0$ indicated 100% confidence when predicting resting.

[Table 1 about here.]

[Figure 4 about here.]

2.7. Conditions comparison

We studied the differences among the different schemes of haptic feedback in three different ways. First, we computed the "negative cross-entropy loss" or NLP loss [29]. Second, we performed a Kruskall-Wallis one way analysis of variance [30] over the NLP losses. Finally, we computed the area under the receiver operating characteristic (ROC) curve (AUC) [31]. Coherent empirical evidence coming in from all of these three measures will strengthen our conclusions on the differences among the different schemes of haptic feedback.

2.7.1. The negative cross-entropy loss [29]. To be able to analyze the confidence of the probabilistic outputs in resting and movement periods together, the NLP loss of each trial was used as a metric of discriminative power. The NLP loss is defined as follows:

$$L = -\frac{1}{n}\left[\sum_{i:y_i=1} \log p(\hat{y}_i = 1|\mathbf{x_i}) + \sum_{i:y_i=0} \log(1 - p(\hat{y}_i = 1|\mathbf{x_i}))\right], \tag{4}$$

where y_i is the real label (flexion or extension depending on the trial and rest), $\hat{y}_i$ is the predicted label, $p(\hat{y}_i = 1|\mathbf{x_i})$ is the probability of movement given by the classifiers for a feature vector $\mathbf{x_i}$ and $p(\hat{y}_i = 0|\mathbf{x_i}) = 1 - p(\hat{y}_i = 1|\mathbf{x_i})$ is the probability of resting given by the classifiers for a feature vector $\mathbf{x_i}$.

This loss penalizes both over-confident and under-confident predictions. Note when the loss becomes larger, the model becomes increasingly certain that the point belongs to the wrong class. The theoretical minimum of this loss is zero and could be achieved if every label of every example (*i.e.*, resting or movement) was predicted correctly with 100% confidence. When one predicts a label incorrectly, the loss is proportional to the confidence of the prediction. For the computation of the NLP, we consider the probabilistic outputs of both binary on-line classifiers together to have a bigger number of samples for the subsequent statistical analysis.

2.7.2. The Kruskal-Wallis one-way analysis of variance [30]. We performed a Kruskal-Wallis one-way analysis of the variance over the NLP loss of each trial and subject, with a Bonferroni adjustment to compensate for multiple comparisons [32]. The Kruskal-Wallis one-way analysis of variance is identical to a traditional one-way analysis of variance with the data replaced by their ranks. Moreover, it is a non-parametric method that does not assume normal populations, unlike the analogous one-way analysis of variance. It is an extension of the Wilcoxon-Mann-Whitney test [33] to three or more groups and it requires independence between sample points.

2.7.3. The ROC curve and the AUC [31]. The Receiver Operating Characteristic (ROC) is a curve of the classifier's *hit* rate against its *false alarm* rate. As one varies the threshold of the classifier, one moves along a curve in this two-dimensional space. A change in the threshold for classifying flexion (extension) vs resting may give more flexion (extension) samples to be classified as flexion (extension), or *hits*, but also more resting samples classified as flexion (extension), or *false alarms*, and vice versa. The area

under the curve (AUC) is a very informative statistic for the evaluation of performance of classification and ranking algorithms. It relies only on the ordering of the points and it is free of any parametric assumptions about the shape of the class distributions. Again, we aggregated flexion vs resting and extension vs resting to obtain more reliable results, given the bigger number of samples.

3. Results

We give important insights on the effects resulting from artificially closing the sensorimotor feedback loop by studying the differences in on-line decoding among conditions (see Table 1 for a definition of the conditions). First, we investigate the informativeness of different features, *i.e.*, the discriminative power of every electrode under the presence and absence of robot movement. Second, we study the differences in classification performance among conditions using the tools described in Section 2.7.

3.1. Spatial and Spectral Features

First, we compare the training periods where the robot arm guided the subject's arm and training periods in which it did not. This comparison is based on the classifier weights per electrode averaged over the frequency bands 8–16 Hz and 18–28 Hz. We focus on these two frequency ranges because the first one contains the μ rhythm (9–14 Hz) and the second range contains the β rhythm (18–24 Hz). Both the μ and β rhythm has been reported as highly discriminative in previous motor imagery experiments [34]. Subsequently, we compute the average classifier weights and power spectra per frequency bin for the electrodes C3, CP3 and C4 in healthy subjects.

3.1.1. Healthy subjects. Figure 5 shows the classifier weights per electrode averaged over the frequency bands 8–16 Hz and 18–28 Hz for the healthy subjects. Note that healthy subjects were instructed to perform motor imagery of the right forearm. In the group level in Figures 5(a)–5(d), we observe empirical evidence of the following facts. First, electrodes over the motor area representing the right arm, *i.e.*, C3, CP3, FC3, FC1, ..., get larger weights (*i.e.*, have a higher discriminative power) when the robot arm guided the subject's arm during the training period in both frequency bands. Second, there is a higher discriminative power of the sensorimotor area during training periods in which the robot arm guided the subject's arm for the frequency band that contains the β rhythm. Third, the spatial distribution of the weights in the classifiers indicates that the classifiers employed neural activity, as the weights in the peripheral locations (*i.e.*, F5, F3, F1, Fz, F2, F4, F6, FC6, C6, CP6, P6, P4, P2, Pz, P1, P3, P5, CP5, C5 and FC5) are low. Hence, it is not likely that electromyographic (EMG) activity coming from movements of the head or the face plays a major role, as such movements are likely to result in strongest weights in more peripheral locations [35].

Figures 7 and 8 show the average classifier weights and power spectra per frequency

bin for the electrodes C3, CP3 and C4. Looking at the figures of the average classifier weights, we observe a shift in discriminative power towards higher frequencies, *i.e.*, from μ rythm desynchronization to β rythm desynchronization, when the robot arm guided the subject's arm during the training period. In terms of the power spectra, we observe a bigger ERD/ERS modulation when the robot was moving the subject's arm. This finding also indicates an increase in BCI-decoding performance achieved by Condition I.

3.1.2. Stroke patients. Figure 6 shows the classifier weights per electrode averaged over the frequency bands 8–16 Hz and 18–28 Hz for the stroke patients. Note that in this case stroke patients were instructed to perform motor imagery of the left forearm. In the group level in Figures 6(a)–6(d), we observe important differences with respect to the group level results for healthy subjects. First, in this case, we observe a higher discriminative power of sensorimotor areas in the μ-frequency range during training periods in which the robot arm guided the subject's arm in comparison to those in which it did not. This indicates that, in agreement with previous studies [36], the stroke patients could experience proprioception. Second, electrodes over the motor area representing the left arm, *i.e.*, C4, CP4, ..., are less informative than electrodes over other areas (*i.e.*, the motor area representing the left leg or even electrodes over the ipsilateral side) for decoding movement intention when the robot arm did not guide the subject's arm during the training session. In terms of a future rehabilitation, this last point is undesirable and, therefore, we decide to re-build the classifiers for the stroke patients using only the spectral features coming from electrodes in the surroundings of the left arm representation in the cortex (specifically, FC2, FC4, FC6, C2, C4, C6, CP2, CP4 and CP6) for the performance analysis. Future studies with more stroke patients are necessary to strengthen the empirical evidence of the discriminative power of the spatial and spectral features.

3.2. Performance

First, we perform a Kruskal-Wallis one-way analysis of variance at significance level $\alpha = 0.05$ over the NLP loss. It rejects the null-hypothesis that the classifier decision values medians were equal for all conditions with a p-value of $3.41 \cdot 10^{-6}$. Subsequently, we perform a multiple t-test comparison procedure also at significance level $\alpha = 0.05$, with a Bonferroni adjustment to compensate for multiple comparisons, over the rank values of every pair of conditions. Figure 10(a) shows the approximate confidence intervals and the average rank per condition of haptic feedback for the multiple comparison procedure. It rejects the null-hypothesis that the classifier decision values means are equal for Condition I in comparison with the remaining conditions. The results suggest an improvement in classification performance when comparing arm movement intention vs resting for the case where the robot provides robot-based haptic reinforcement during the training and the test periods, *i.e.*, Condition I outperforms the rest. Note that during test, the robot arm was programmed to move or stop according to the decoded intention

in an on-line manner during the movement periods.

[Figure 5 about here.]

[Figure 6 about here.]

[Figure 7 about here.]

[Figure 8 about here.]

Average NLP loss values for every condition of haptic feedback as well as for every subject are shown in Figure 10(b). The differences among conditions are consistent with the findings discussed during the analysis of the variance. In Condition I, the classification performance appears to be the highest, for both healthy subjects in a group level and stroke patients in a group level. The numerical values of the total group average (healthy subjects and stroke patients) for the NLP loss are 0.59 for Condition I, 0.65 for Condition II, 0.67 for Condition III and 0.68 for Condition IV. A more traditional measure of performance is the area under the ROC curve (AUC), shown in Figure 9. Note that the AUC values correspond to the on-line binary classification of movement vs resting (aggregation of the AUC values for flexion vs resting and extension vs resting). Again, differences among conditions support our previous findings and Condition I outperforms the rest for both healthy patients in a group level and stroke patients in a group level. The numerical values of the total group average (healthy subjects and stroke patients) for the AUC are 0.76 for Condition I, 0.68 for Condition II, 0.63 for Condition III and 0.64 for Condition IV, and the chance level (binary classification) of the AUC is 0.5.

4. Discussion

In this article, we have demonstrated that artificially closing the sensorimotor feedback loop facilitates decoding of movement intention in both healthy subjects and stroke patients. This result indicates the feasibility of an integrated stroke therapy that combines robot-assisted physical therapy with decoding of movement intention by means of a BCI. Specifically, we have provided evidence that the strength of the sensorimotor rhythm (SMR), as measured by the class probability estimated from the outputs of the SVM, is modulated by the haptic robot-based feedback. Furthermore, our results suggest that this modulation of the SMR is actually beneficial for decoding of arm movement intention. An increased classification accuracy is exhibited by comparing performance with haptic feedback to no haptic feedback (cf. Conditions I and IV in Figure 10).

For healthy subjects, the discriminative power and powerspectra indicate that haptic feedback activates the somatosensory cortex and increases ERD/ERS modulation in the β-frequency range. Both discriminative power and powerspectra are shown in Figures 5 and 7. Interestingly, this observation is in agreement with previous

reports on the effect of haptic stimulation in non-human primates [37]. In addition, our results are also in agreement with previous reports on the effect of passive haptic stimulation [18, 19, 21], despite that in these studies haptic stimulation was performed independently of the actual (or decoded) movement intention. For the stroke patients, it is more difficult to provide strong conclusions. Although the discriminative power (shown in Figure 6) indicates that haptic feedback activates the somatosensory cortex in the μ-frequency range, we also observe how electrodes over the motor area representing the left arm, *i.e.*, C4, CP4, ..., are less informative to decode movement intention than electrodes over other areas (*i.e.*, the motor area representing the left leg or even electrodes over the ipsilateral side) in absence of haptic feedback. Nevertheless, future studies with a larger number of stroke patients are necessary to obtain more insights regarding these points.

The positive effect of closing the sensorimotor loop on the decoding of movement intention allows the speculation that haptic feedback supports subjects in initiating a voluntary modulation of their SMR. Hence, it is a key point that haptic feedback is provided during each trial synchronized to the subject's intention to move the arm, and not at the end of each trial as previously suggested [22]. This strategy is likely to result in increased cortical plasticity due to Hebbian-type learning [15, 16, 17]. However, further investigations into the neuronal correlates of this effect are required.

The positive effect of closing the sensorimotor loop on the decoding of movement intention is also in agreement with previous ECoG studies [38, 39]. These studies indicated a greater performance decoding real (overt) active arm movements vs resting than decoding imagined arm movements vs resting. In this article, we have provided empirical evidence of a greater performance decoding imagined arm movement vs resting when a robot arm moves the subject's arm than decoding imagined arm movement vs resting without robot movement (cf. Conditions I and IV in Figure 10). Nevertheless, we do not aim to address the spatial differences in the neural activity for different movements (in our case, flexion and extension) as in previous ECoG studies [38, 39], given the lack of spatial resolution of EEG when compared to ECoG.

As we did not measure EMG during the experimental sessions, it may be argued that our motor-imagery data could be confounded by actual minimal movement execution. While this is a concern frequently raised in the context of BCIs, we would like to point out that the validity of distinguishing motor imagery from motor execution by EMG is highly disputable. Motor imagery and motor execution share the same neuronal substrate, with both, imagery and execution, resulting in increased muscle tone [40]. Minor movements and motor imagery may thus result in indistinguishable EMG signals. Accordingly, utilizing EMG measurements to control for motor execution suggests a certainty in excluding potential confounders that is not supported by empirical evidence. While this is a major concern when BCIs are used for communication, we wish to point out that this is not the case in our study. We employed motor imagery in healthy subjects to simulate stroke patients, where movement intent is only accompanied by minimal or even no movement execution. As such, we do not consider the potential

presence of minimal movements in our healthy subjects to be a major concern, as such movements are also likely to be present in stroke patients with only residual motor control.

The presented work is substantially different from previous studies that achieved comparable accuracy in full trial binary or four-class classification of motor imagery without the use of haptic feedback [41, 42, 43]. On-line classification of motor imagery based on shorter time-windows than a full trial is a considerably harder problem than full trial classification. Our main take home message is that even if everything else is identical, haptic feedback improves the accuracy.

The presented results indicate the possibility for a future stroke rehabilitation based on robot-assisted physical therapy with BCI-based decoding of movement intention. However, it needs to be pointed out that the support provided by this study hinges on the assumption that the results presented here with healthy subjects and two stroke patients can be transfered to a bigger set of stroke patients with different degrees of hemiparesis and neurological conditions. The beneficial effect reported here is likely to depend on the presence of proprioception. Nevertheless, there is no a-priori reason why stroke patients should not have proprioception. Both stroke patients in this study exhibited proprioception and previous studies have shown that most stroke patients recover proprioception eight weeks post-stroke despite remaining motor disabilities [36].

Besides the relevance of our results for a potential stroke therapy, it is furthermore noteworthy that the positive influence of haptic feedback on decoding accuracy may also be beneficial for other subject groups. For example, subjects in late stages of ALS appear not to be capable of modulating their SMR sufficiently, as indicated by fact that so far no communication with a completely locked-in subject has been established by means of a BCI. While the extent of sensory feedback in late stages of ALS remains unclear, haptic feedback might also support these subjects in initiating volitional modulation of their SMR.

[Figure 9 about here.]

[Figure 10 about here.]

Acknowledgments

Manuel Gomez Rodriguez has been supported in part by a Fundacion Barrie de la Maza Graduate Fellowship. The authors thank Bernd Battes for technical support during the experiments.

[1] L.A. Farwell and E. Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 70:510–523, 1988.

[2] N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kübler, J. Perelmouter, E. Taub, and H. Flor. A spelling device for the paralysed. *Nature*, 398:297–298, 1999.

[3] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzer. EEG-based discrimination between imagination of right and left hand movement. *Electroencephalography and Clinical Neurophysiology*, 103:642–651, 1997.

[4] M. Grosse-Wentrup, C. Liefhold, K. Gramann, and M. Buss. Beamforming in noninvasive brain-computer interfaces. *IEEE Transactions on Biomedical Engineering*, 56(4):1209–1219, 2009.

[5] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller. Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Processing Magazine*, 25(1):41–56, 2008.

[6] A. Kübler, F. Nijboer, J. Mellinger, T.M. Vaughan, H. Pawelzik, G. Schalk, D.J. McFarland, N. Birbaumer, and J.R. Wolpaw. Patients with ALS can use sensorimotor rhythms to operate a brain-computer interface. *Neurology*, 64(10):1775–1777, 2005.

[7] HC Dijkerman, M. Ietswaart, M. Johnston, and RS MacWalter. Does motor imagery training improve hand function in chronic stroke patients? A pilot study. *Clinical Rehabilitation*, 18(5):538, 2004.

[8] S.J. Page, P. Levine, and A. Leonard. Mental practice in chronic stroke: results of a randomized, placebo-controlled trial. *Stroke*, 38(4):1293, 2007.

[9] K. Tanaka, K. Matsunaga, and HO Wang. Electroencephalogram-based control of an electric wheelchair. *IEEE Transactions on Robotics*, 21(4):762–766, 2005.

[10] J.R. Millán, F. Renkens, J. Mouriño, and W. Gerstner. Noninvasive brain-actuated control of a mobile robot by human EEG. *IEEE Transactions on Biomedical Engineering*, 51(6):1026–1033, 2004.

[11] J.J. Daly and J.R. Wolpaw. Brain-computer interfaces in neurological rehabilitation. *The Lancet Neurology*, 7(11):1032–1043, 2008.

[12] R.P.S. van Peppen, G. Kwakkel, S. Wood-Dauphine, H.J.M. Hendriks, Ph.J. van der Wees, and J. Dekker. The impact of physical therapy on functional outcomes after stroke: what's the evidence? *Clinical Rehabilitation*, 18(8):833–862, 2004.

[13] R. Riener, T. Nef, and G. Colombo. Robot-aided neurorehabilitation of the upper extremities. *Medical and Biological Engineering and Computing*, 43(1):2–10, 2005.

[14] E. Buch, C. Weber, L.G. Cohen, C. Braun, M.A. Dimyan, T. Ard, J. Mellinger, A. Caria, S. Soekadar, and A. Fourkas. Think to move: a neuromagnetic brain-computer interface (bci) system for chronic stroke. *Stroke*, 39(3):910, 2008.

[15] W. Wang, J.L. Collinger, M.A. Perez, E.C. Tyler-Kabara, L.G. Cohen, N. Birbaumer, S.W. Brose, A.B. Schwartz, M.L. Boninger, and D.J. Weber. Neural Interface Technology for Rehabilitation: Exploiting and Promoting Neuroplasticity. *Physical Medicine and Rehabilitation Clinics of North America*, 21(1):157–178, 2010.

[16] TH Murphy and D. Corbett. Plasticity during stroke recovery: from synapse to behaviour. *Nature Reviews Neuroscience*, 10(12):861–872, 2009.

[17] L. Kalra. Stroke Rehabilitation 2009: Old Chestnuts and New Insights. *Stroke*, 41(2):e88, 2010.

[18] G.R. Müller, C. Neuper, R. Rupp, C. Keinrath, H.J. Gerner, and G. Pfurtscheller. Event-related beta EEG changes during wrist movements induced by functional electrical stimulation of forearm muscles in man. *Neuroscience Letters*, 340:143–147, 2003.

[19] G.R. Müller-Putz, D. Zimmermann, B. Graimann, K. Nestinger, G. Korisek, and G. Pfurtscheller. Event-related beta EEG changes during passive and attempted foot movements in paraplegic patients. *Brain Research*, 1137:84–91, 2007.

[20] S. Waldert, H. Preissl, E. Demandt, C. Braun, N. Birbaumer, A. Aertsen, and C. Mehring. Hand movement direction decoded from meg and eeg. *The Journal of Neuroscience*, 28(4):1000–1008, 2008.

[21] A. Ramos, S. Halder, and N. Birbaumer. Proprioceptive feedback in bci. In *Proceedings of the 4th International IEEE/EMBS Conference on Neural Engineering*, pages 279–282, 2009.

[22] K.K. Ang, C. Guan, S.G. Chua, B.T. Ang, C. Kuah, C. Wang, K.S. Phua, Z.Y. Chin, and H. Zhang. A clinical study of motor imagery-based brain-computer interface for upper limb robotic rehabilitation. In *Annual International Conference of the IEEE Engineering in Medicine*

and Biology Society (EMBC), pages 5981–5984, 2009.

[23] M. Gomez Rodriguez, J. Peters, J. Hill, B. Schölkopf, A. Gharabaghi, and M. Grosse-Wentrup. Closing the sensorimotor loop: Haptic feedback facilitates decoding of arm movement imagery. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics. Workshop in Shared Control for Brain-Machine Interfaces*, October 2010.

[24] G. Schalk, D.J. McFarland, T. Hinterberger, N. Birbaumer, and J.R. Wolpaw. BCI 2000: A General-Purpose Brain-Computer Interface(BCI) System. *IEEE Transactions on Biomedical Engineering*, 51(6):1034–1043, 2004.

[25] http://bci2000.org/downloads/BCPy2000/.

[26] D.J. McFarland, L.M. McCane, S.V. David, and J.R. Wolpaw. Spatial filter selection for EEG-based communication. *Electroencephalography and Clinical Neurophysiology*, 103(3):386–394, 1997.

[27] B. Schölkopf and A.J. Smola. *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. the MIT Press, 2002.

[28] John C. Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In *Advances in Large Margin Classifiers*, pages 61–74. MIT Press, 1999.

[29] J. Quinonero-Candela, C. Rasmussen, F. Sinz, O. Bousquet, and B. Schölkopf. Evaluating predictive uncertainty challenge. *Machine Learning Challenges Workshop*, pages 1–27, 2006.

[30] W.H. Kruskal and W.A. Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260):583–621, 1952.

[31] N.J. Hill, T. N. Lal, M. Tangermann, T. Hinterberger, G. Widman, C. E. Elger, B. Schlkopf, and N. Birbaumer. Classifying Event-Related Desynchronization in EEG, ECoG and MEG signals. *Toward Brain-Computer Interfacing*, pages 235–260, 2007.

[32] Y. Hochberg and A.C. Tamhane. *Multiple comparison procedures*. Wiley New York, 1987.

[33] H.B. Mann and D.R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(1):50–60, 1947.

[34] J.R. Wolpaw, N. Birbaumer, W.J. Heetderks, D.J. McFarland, P.H. Peckham, G. Schalk, E. Donchin, L.A. Quatrano, C.J. Robinson, and T.M. Vaughan. Brain-computer interface technology: a review of the first international meeting. *IEEE Transactions on Rehabilitation Engineering*, 8(2):164–173, 2000.

[35] I.I. Goncharova, D.J. McFarland, T.M. Vaughan, and J.R. Wolpaw. EMG contamination of EEG: spectral and topographical characteristics. *Clinical Neurophysiology*, 114(9):1580–1593, 2003.

[36] D.L. Smith, A.J. Akhtar, and W.M. Garraway. Proprioception and spatial neglect after stroke. *Age and Ageing*, 12(1):63, 1983.

[37] A.J. Suminski, D.C. Tkach, A.H. Fagg, and N.G. Hatsopoulos. Incorporating Feedback from Multiple Sensory Modalities Enhances Brain-Machine Interface Control. *Journal of Neuroscience*, 30(50):16777, 2010.

[38] P. Shenoy, K.J. Miller, J.G. Ojemann, and R.P.N. Rao. Generalized features for electrocorticographic BCIs. *IEEE Transactions on Biomedical Engineering*, 55(1):273, 2008.

[39] K.J. Miller, G. Schalk, E.E. Fetz, M. den Nijs, J.G. Ojemann, and R.P.N. Rao. Cortical activity during motor execution, motor imagery, and imagery-based online feedback. *Proceedings of the National Academy of Sciences*, 107(9):4430, 2010.

[40] M. Jeannerod. Mental imagery in the motor context. *Neuropsychologia*, 33(11):1419–1432, 1995. The Neuropsychology of Mental Imagery.

[41] D. Huang, P. Lin, D.Y. Fei, X. Chen, and O. Bai. Decoding human motor activity from EEG single trials for a discrete two-dimensional cursor control. *Journal of Neural Engineering*, 6, 2009.

[42] F. Babiloni, F. Cincotti, L. Bianchi, G. Pirri, et al. Recognition of imagined hand movements with low resolution surface Laplacian and linear classifiers. *Medical Engineering & Physics*, 23(5):323–328, 2001.

[43] D.J. McFarland, W.A. Sarnacki, T.M. Vaughan, and J.R. Wolpaw. Brain-computer interface

(BCI) operation: signal and noise during early training sessions. *Clinical Neurophysiology*, 116(1):56–62, 2005.

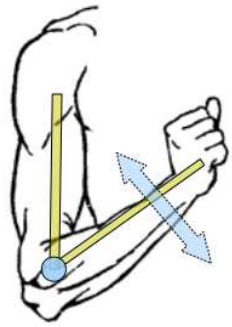


Figure 1. Subject's imagined task. The subject was instructed to imagine flexion or extension of the forearm, using the elbow as the single degree of freedom during the imagined movement. Healthy subjects were instructed to imagine the movement with the right forearm while stroke patients were instructed to imagine it with the left forearm (their hemiparesis was on the left side of the body).

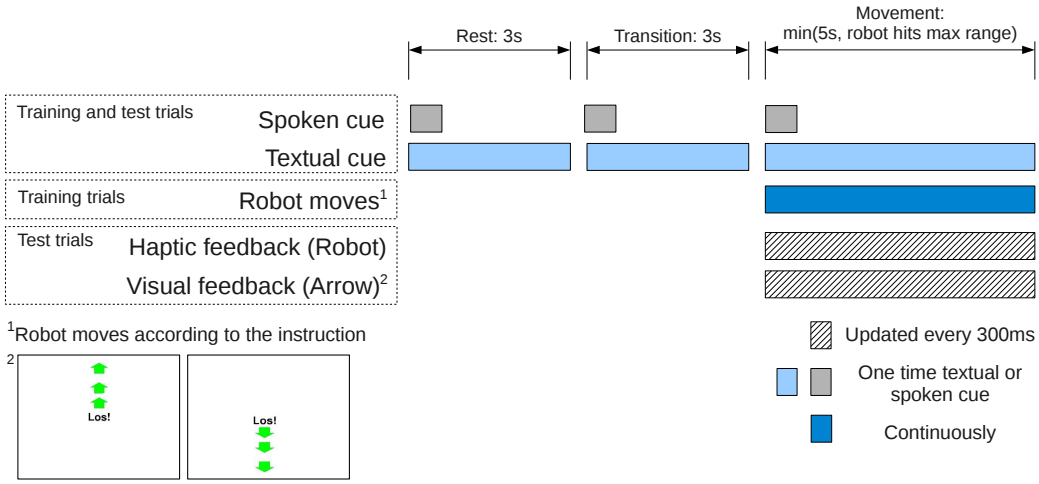


Figure 2. Timeline of events during a trial. Please note that the spoken and textual cues are one-time events, the movement of the robot during movements period of the training sections is continuous, while the haptic and visual feedback are events that are updated every 300ms.

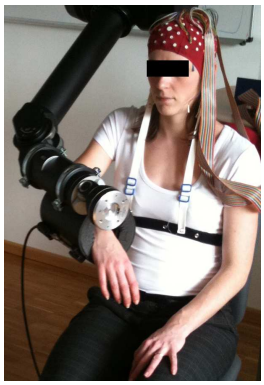


Figure 3. Experimental Setup. A robot arm moves the subject's forearm with the elbow as the single degree of freedom (DoF). The resting position, maximum extension and maximum flexion of the robot arm could be adjusted depending on the subject. The single joint of the robot had an angular velocity of 0.096 rad/s (5.5 deg/s), *i.e.*, every 300ms of movement the joint rotated 0.0288 rad (1.65 deg). Given that the subject's hand was approximately 40 cm from the joint, the hand was being moving with a linear velocity of 0.096 rad/s $\times$ 0.4 m = 3.84 cm/s

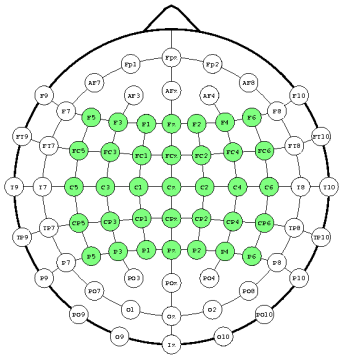
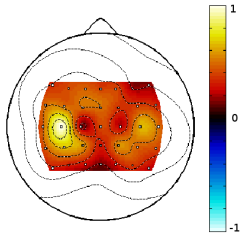
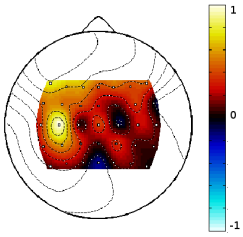


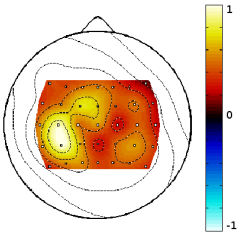
Figure 4. EEG electrode grid configuration used in our experiments. The 35 channels of the EEG electrode cap that were fed into the amplifier are shown in green. The electrodes covered parts of the pre-motor cortex, primary motor cortex and somatosensory cortex as well as several other areas.



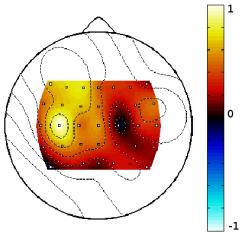
(a) 8–16 Hz, robot (group average)



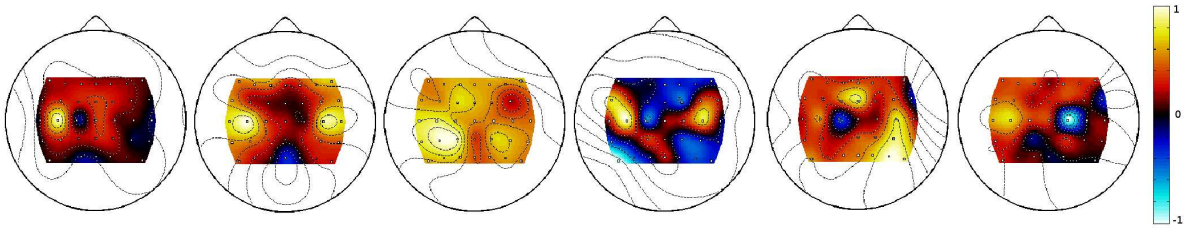
(b) 8–16 Hz, no robot (group average)



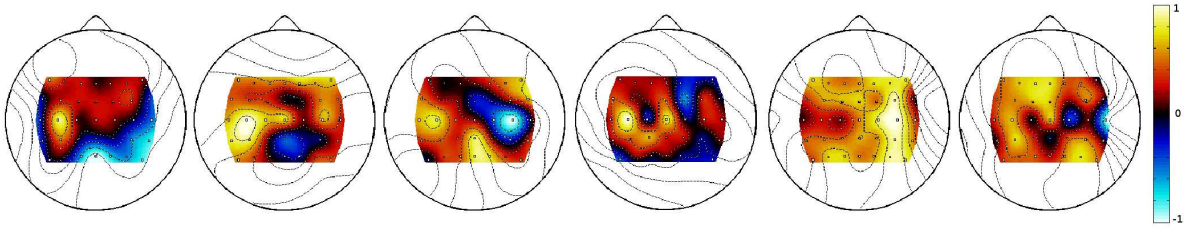
(c) 18–28 Hz, robot (group average)



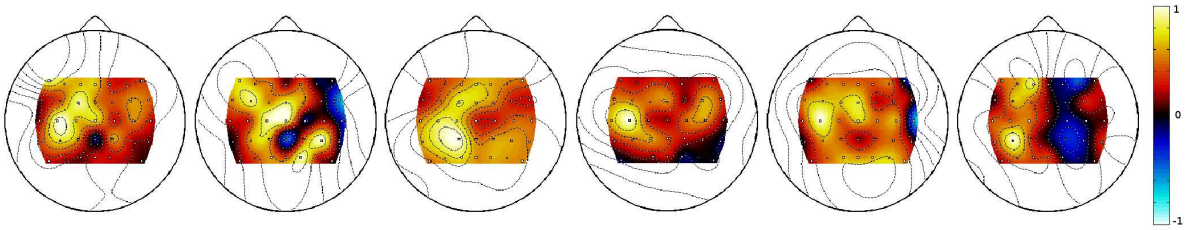
(d) 18–28 Hz, no robot (group average)



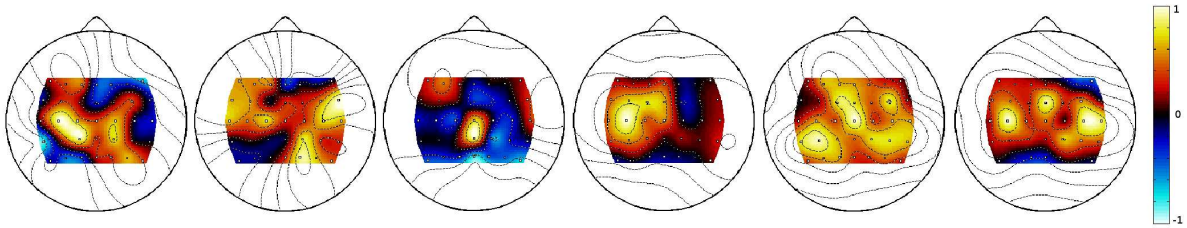
(e) 8–16 Hz, robot (per subject)



(f) 8–16 Hz, no robot (per subject))

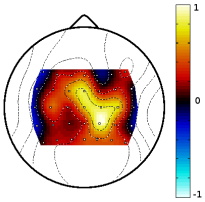


(g) 18–28 Hz, robot (per subject)

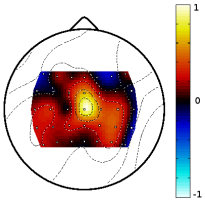


(h) 18–28 Hz, no robot (per subject)

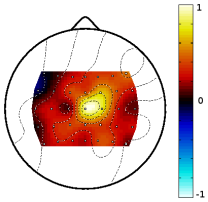
Figure 5. Discriminative power of the electrodes for the healthy subjects: Comparison of the classifier weights for the (a–d) group average and (e–h) per healthy subject for the frequency bands (a,b,e,f) 8–16 Hz and (c,d,g,h) 18–28 Hz, with (a,c,e,g) and without (b,d,f,h) the robot arm guiding the subjects' arm during the training section.



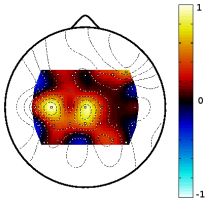
(a) 8–16 Hz, robot (group average)



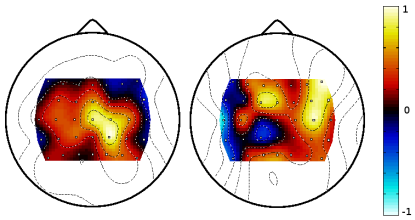
(b) 8–16 Hz, no robot (group average)



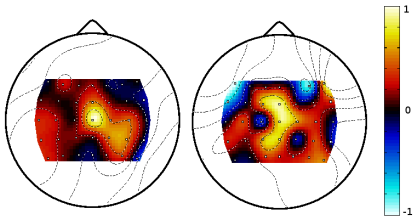
(c) 18–28 Hz, robot (group average)



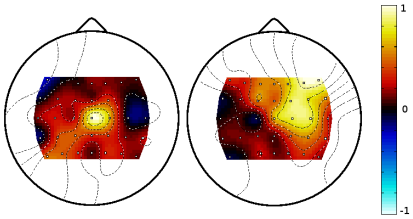
(d) 18–28 Hz, no robot (group average)



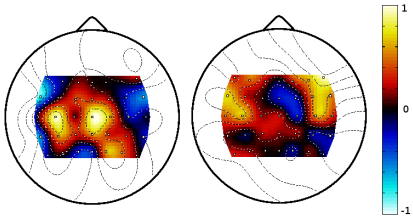
(e) 8–16 Hz, robot (per subject)



(f) 8–16 Hz, no robot (per subject))

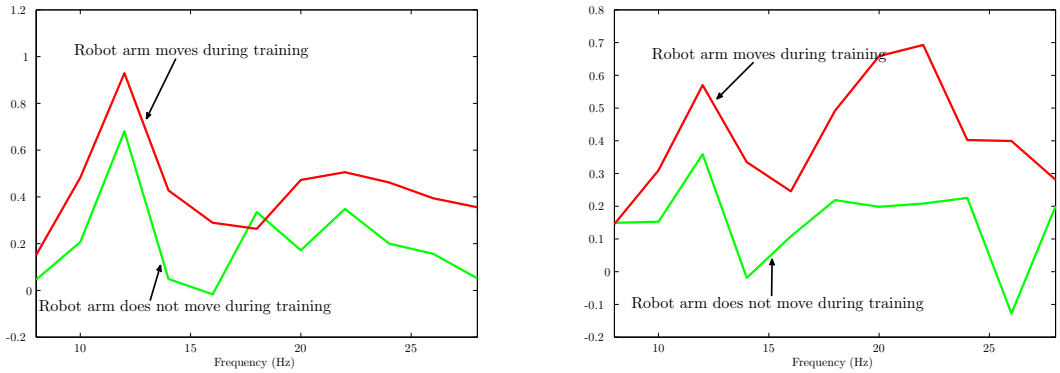


(g) 18–28 Hz, robot (per subject)



(h) 18–28 Hz, no robot (per subject)

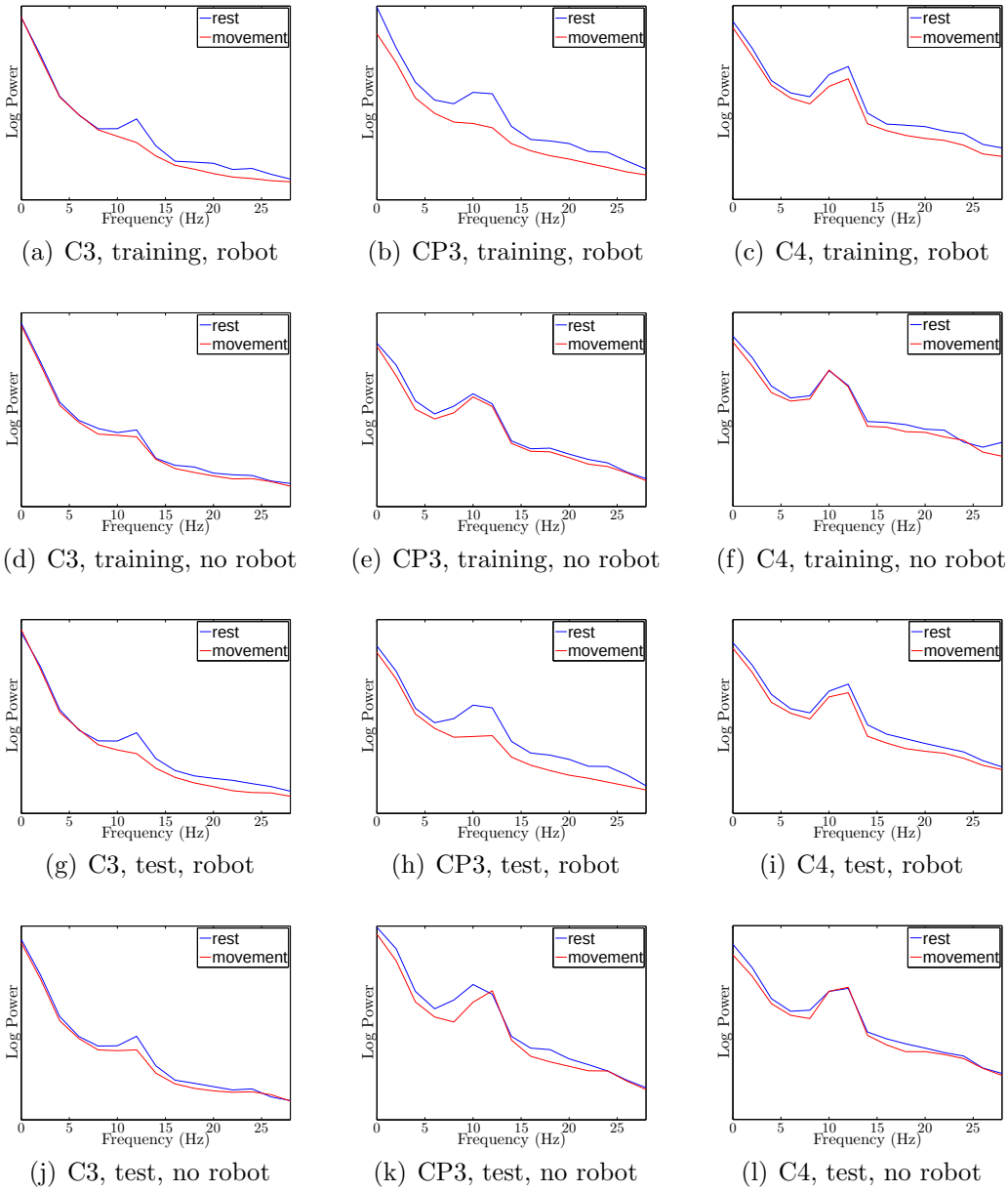
Figure 6. Discriminative power of the electrodes for the stroke patients: Comparison of the classifier weights for the (a–d) group average and (e–h) per stroke patient for the frequency bands (a,b,e,f) 8–16 Hz and (c,d,g,h) 18–28 Hz, with (a,c,e,g) and without (b,d,f,h) the robot arm guiding the subjects' arm during the training section.



(a) Classifier weights for electrode C3

(b) Classifier weights for electrode CP3

Figure 7. Discriminative power of the frequency components of the electrodes C3 and CP3 for healthy subjects: Classifier weights for the electrodes C3 and CP3 in the frequency band 8–30 Hz for the classifiers generated when the robot arm guides the subjects' arm and when the robot arm is not moving during training. Refer to Fig. 4 for electrode grid spatial configuration.



(a) C3, training, robot (b) CP3, training, robot (c) C4, training, robot

(d) C3, training, no robot (e) CP3, training, no robot (f) C4, training, no robot

(g) C3, test, robot (h) CP3, test, robot (i) C4, test, robot

(j) C3, test, no robot (k) CP3, test, no robot (l) C4, test, no robot

Figure 8. Power spectra of the electrodes C3, C4 and CP3 for healthy subjects: Power spectra in movement periods and the rest periods for the electrodes C3, CP3 and C4 in the frequency band 2–30 Hz for the training and test sections with and without the robot arm guiding the subjects' arm. Refer to Fig. 4 for electrode grid spatial configuration.

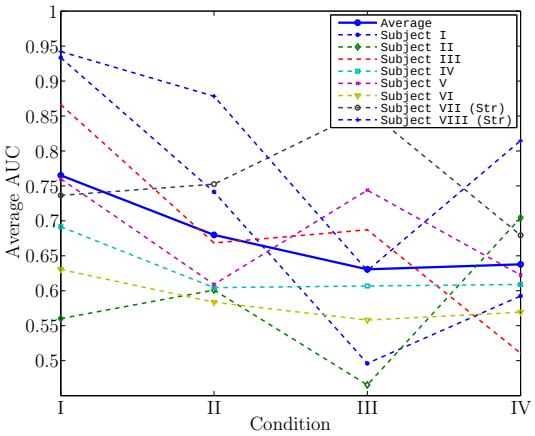
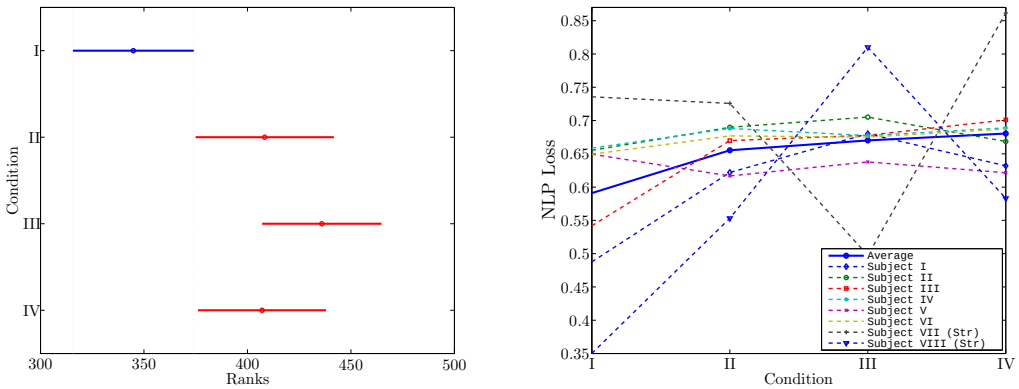


Figure 9. Average AUC values of the on-line classifiers over all trials and subjects for all four conditions, *i.e.*, 30 trials per condition and subject for flexion vs rest and 30 trials per condition and subject for extension vs rest. See Table 1 for a description of the conditions.



(a) ANOVA

(b) Negative Log Probability (NLP) loss

Figure 10. Comparison of discriminative power among Conditions I–IV: (a) Average ranks with confidence intervals given by a multiple comparison procedure computed over the ranks of the NLP loss of each trial and subject using a Bonferroni adjustment to compensate for multiple comparisons and (b) average NLP loss across trials and subjects. A total of 60 trials per condition and subject were recorded. See Table 1 for a description of the conditions.

Condition	**Training**	**Test (Always visual feedback)**
I	Robot moves[1]	Robot moves[2]
II	Robot moves[1]	Robot does not move
III	Robot does not move	Robot moves[2]
IV	Robot does not move	Robot does not move

Table 1. Conditions to explore haptic feedback. Please note that the robot can move [1]according to the instruction in training periods or [2]according to the decoded movement intention in test periods.