# SCIENTIFIC REP🞐RTS

**OPEN**

# Across-subject offline decoding of motor imagery from MEG and EEG

Hanna-Leena Halme[1] & Lauri Parkkonen [1,2]

Long calibration time hinders the feasibility of brain-computer interfaces (BCI). If other subjects' data were used for training the classifier, BCI-based neurofeedback practice could start without the initial calibration. Here, we compare methods for inter-subject decoding of left- vs. right-hand motor imagery (MI) from MEG and EEG. Six methods were tested on data involving MEG and EEG measurements of healthy participants. Inter-subject decoders were trained on subjects showing good within-subject accuracy, and tested on all subjects, including poor performers. Three methods were based on Common Spatial Patterns (CSP), and three others on logistic regression with $l_1$ - or $l_{2,1}$ -norm regularization. The decoding accuracy was evaluated using (1) MI and (2) passive movements (PM) for training, separately for MEG and EEG. With MI training, the best accuracies across subjects (mean 70.6% for MEG, 67.7% for EEG) were obtained using multi-task learning (MTL) with logistic regression and $l_{2,1}$-norm regularization. MEG yielded slightly better average accuracies than EEG. With PM training, none of the inter-subject methods yielded above chance level (58.7%) accuracy. In conclusion, MTL and training with other subject's MI is efficient for inter-subject decoding of MI. Passive movements of other subjects are likely suboptimal for training the MI classifiers.

Brain–computer interfaces (BCI) translate user's brain activity in real time into commands for an external device. Non-invasive BCIs have diverse applications in both basic and clinical neuroscience, including videogames, exercise systems for increasing attention or relaxation, communication devices for disabled people and various tools for neurological rehabilitation. For example, hemiparetic patients can learn to control an orthosis attached to the paretic hand using motor imagery (MI) and concurrent feedback (see a review by Teo and co-workers[1]). This so-called neurofeedback training may enhance recovery of sensorimotor function compared to traditional physiotherapy, but the clinical evidence of BCI therapy still remains vague[2,3].

Electric brain activity can be measured noninvasively with e.g. electroencephalography (EEG) and the corresponding magnetic fields with magnetoencephalography (MEG). MEG has recently gained interest in the context of BCI experiments as well[4–7], mainly because it provides higher signal-to-noise ratio and spatial resolution compared to EEG. MEG can be used to assist the development of eventually EEG-based BCIs, and MEG-based BCIs can be used in basic neuroscientific research.

Unfortunately, BCIs typically require dozens or even hundreds of samples of subject's neurophysiological signals for training the classification algorithms to decode the user's brain states with an accuracy exceeding the chance level. Due to large inter-subject variability of these signals, the user-specific brain responses are typically collected in the beginning of each session for calibrating the BCI and thus the total calibration time can be significant. This process can be exhausting for patients suffering from neurological disorders. Although Foldes and co-workers showed previously that five minutes of MEG data are sufficient for calibrating a subject-specific BCI discriminating MI from rest[7], this time is still away from therapy. In addition, when employing neurofeedback for rehabilitation, the feedback should drive brain activity towards that of a healthy brain and not to reinforce the prevailing pathological state as a patient-specific BCI might do.

In order to overcome these drawbacks, several research groups have made effort to develop inter-subject-generalized BCIs[8–10] which can be trained in advance using data from previously measured subjects. Therefore, a new BCI user can begin the neurofeedback practice immediately without the initial calibration session, which saves both time and effort of the patients.

Successful inter-subject classification requires that globally relevant signal features are extracted from each training subject. Optimal features should accurately discriminate between the brain states of interest and they should generalize across subjects in order to enable inter-subject learning. Feature extraction is a crucial part of

[1]Department of Neuroscience and Biomedical Engineering NBE, Aalto University School of Science, P.O. Box 12200, FI-00076, Aalto, Finland. [2]Aalto Neuroimaging, MEG Core, Aalto University School of Science, Espoo, Finland. Correspondence and requests for materials should be addressed to H.-L.H. (email: hanna-leena.halme@aalto.fi)

a BCI, and selecting optimal features can significantly improve the classification accuracy, as shown e.g. in our previous study[11].

Multi-task learning (MTL[12,13]) aims at selecting a few relevant features from a large feature set such that these features are shared across multiple related tasks; in the case of BCI training, the tasks are usually measurements from different subjects. MTL is beneficial for learning problems in which the number of samples is much lower than the number of features, since it effectively reduces the dimensionality of the feature space. MTL finds a sparse feature set that is correlated across tasks instead of optimizing features for each individual task, which makes the method robust to outliers. In case of inter-subject classification, MTL considers decoding of each subject's data as a separate task. In contrast to data pooling, MTL does not assume that training data from all subjects are equally distributed, which allows variability between subjects.

Regularized logistic regression has been used, among other machine-learning applications, for classifying EEG signals[14,15]. Moreover, logistic regression with $l_{2,1}$-norm regularization is found to be efficient for selecting a small number of relevant features from a high-dimensional feature space[16]. Recently, MTL based on logistic regression and regularization with $l_{2,1}$ norm minimization was successfully used for extracting features and classifying MEG signals in an inter-subject decoding paradigm[17].

Another popular method for feature extraction is spatial filtering with the Common Spatial Patterns (CSP) algorithm, which has been applied to EEG[18,19] as well as to MEG[11]. CSP aims at maximizing the covariance of signals representing one class (e.g. left-hand motor imagery) while simultaneously minimizing it for another class (e.g. right-hand motor imagery). The benefits of CSP are low computational cost and efficiency especially in motor function decoding[20], but the extension to inter-subject classification is challenging, since the spatial patterns of each individual are slightly different, which may result in classifier overfitting.

In order to avoid overfitting and to improve generalization, both the covariance matrices and the objective function of CSP can be regularized towards a more generic filter decomposition by using additional regularization parameters[21]. Regularization of CSP has been found to improve inter-subject classification of EEG data[8,21–23], but to our knowledge has not yet been applied to MEG data. Furthermore, the discriminatory power of CSP can be improved with preceding dimensionality reduction using e.g. Spatio-Spectral Decomposition (SSD[24]), which maximizes the signal power at a frequency of interest while simultaneously minimizing it at the flanking frequencies. SSD is especially beneficial for extracting neural oscillations occurring within a known frequency band, such as sensorimotor rhythms (SMR) consisting of ~10-Hz and ~20-Hz oscillations.

In the current study, we investigate inter-subject decoding of motor imagery (MI) -related MEG and EEG signals using different methods for feature selection and classification. The classification accuracy is evaluated using a dataset from healthy participants. The data include both MEG and EEG collected during passive movements (PM) and MI of the left and right hand. The decoding accuracy is evaluated using 1) MI and 2) PM as training data, separately for MEG and EEG. The rationale for using other subjects' PM for training is based on previous studies showing that PM of the test subject can be used for calibration of an MI-BCI[25–27]. However, there is also evidence that the activity patterns during MI and PM are not similar[28], which likely reduces the accuracy in inter-subject MI classification. Moreover, in previous studies, the decoding was performed between rest and MI, whereas in the current study we are decoding left- vs. right-hand MI. We therefore expect inferior results with PM training compared to MI training.

The evaluated decoding methods are based on either (1) CSP combined with linear discriminant analysis (LDA) classifier, or (2) logistic regression with $l_{2,1}$- or $l_1$-norm regularization. For the CSP-based methods, the following approaches (abbreviation in parentheses) are used for decoding:

a. CSP, classification with LDA (CSP + LDA)
b. CSP, classification with bootstrap aggregating & LDA (CSP + bagging)
c. Regularized CSP, classification with LDA (regCSP)

The functions included in MTJFL[17] and MALSAR[29] Matlab toolboxes were modified and used for the logistic regression –based methods. The following inter-subject decoding approaches, implemented in the MTJFL toolbox, are used for decoding:

a. Pooling with logistic regression and $l_1$-norm regularization (pooling)
b. MTL with logistic regression and $l_1$-norm regularization (L1-MTL)
c. MTL with logistic regression and $l_{2,1}$-norm regularization (L21-MTL)

In addition, we calculate the accuracies for within-subject decoding for comparison with the inter-subject decoding results. In this case, each subject's own PM or MI data are used for training. PM training for MEG is tested in an online BCI paradigm using CSP + LDA. For EEG, similar PM training paradigm is conducted for the offline data. For both modalities, within-subject MI training is calculated offline using 10-fold cross-validation and $l_1$-norm regularized logistic regression.

The main research questions of this study are: (1) do generalized classifiers perform above chance level and how does their performance compare to that of a subject-specific classifier, (2) can generalized classifiers be trained equally well with responses to PM and MI, and (3) whether MEG and EEG data yield different inter-subject decoding results.

## Results

The detailed results for different modalities and training data are presented in the following subsections. All classification results are summarized in Fig. 1.
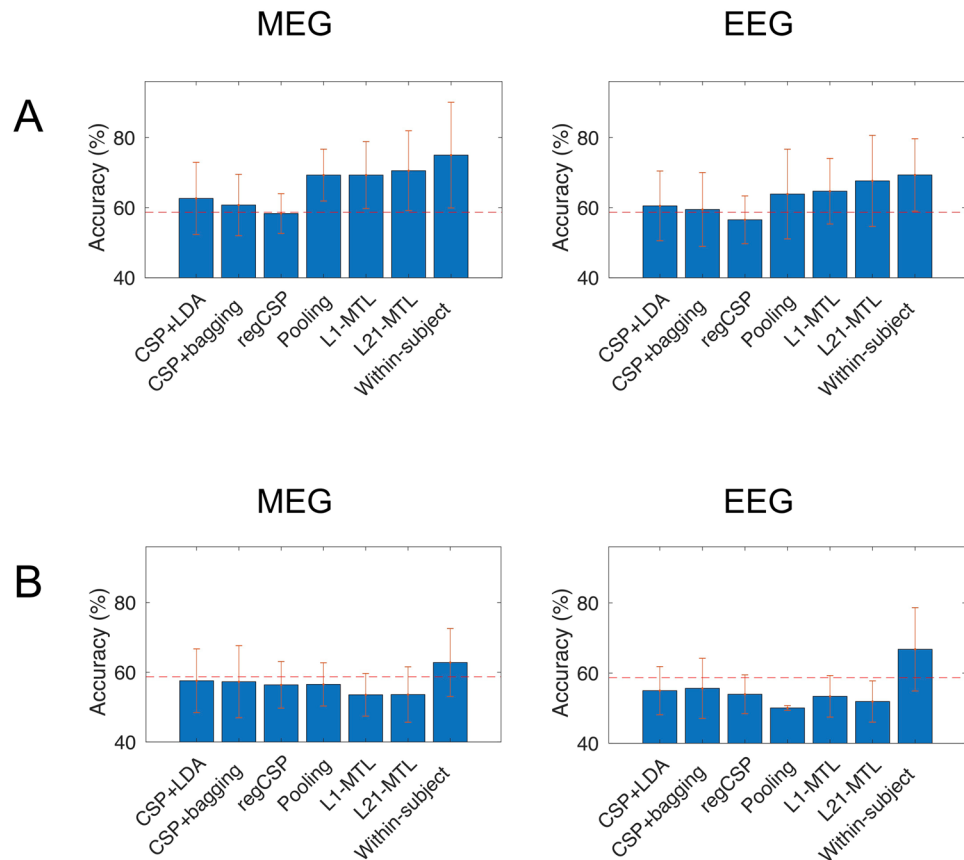
**Figure 1.** Classification results for (**A**) Motor-imagery-trained MEG and EEG, (**B**) Passive-movement-trained MEG and EEG. The chance level is indicated with a dashed line. Error bars represent the standard deviation.

**Within-subject decoding online and offline.** Table 1 represents the results for within-subject MEG decoding (1) in the online BCI paradigm, using subject's own PM for training (see Section 4.4 for a detailed explanation) and (2) in the offline analysis using 10-fold cross-validation with subject's own MI data and $l_1$-norm regularized logistic regression.

The online accuracy was below the sample-size-adjusted chance level of 58.7%[30] for 7 subjects. However, as the online decoding was done with PM training, we assumed these results probably did not reflect the robustness of MI or the ability to perform MI. Therefore, the decoding accuracy was also calculated offline with MI training in order to evaluate the individual MI performance. The mean accuracy for each subject was calculated as the average of online and offline within-subject accuracies. Five subjects with the lowest mean accuracy were discarded from the training data in further analyses, i.e. the poorly performing subjects were used only for testing the decoders. The purpose of discarding bad performers from the training data was to find out whether their results would be improved with good-quality training data from good performers.

**Offline training by motor imagery.** Table 2 shows the results when the classifier was trained by MEG responses to motor imagery; the corresponding results for EEG are shown in Table 3. The best inter-subject mean accuracies of 70.6% for MEG and 67.7% for EEG were obtained with L21-MTL. Both results were significantly (MEG: $p = 0.000$, EEG: $p = 0.007$) above chance level. For MEG, also pooling and L1-MTL yielded good results (66.5% and 67.8%, respectively) that were above chance level (pooling: $p = 0.000$, L1-MTL: $p = 0.000$). For EEG, L1-MTL yielded an average accuracy significantly better than chance ($p = 0.01$). Also, for both modalities within-subject decoding with MI training resulted in average accuracy better than chance ($p = 0.000$ for both MEG and EEG).

The statistical differences between the methods ($p$-values, calculated by paired t-tests and Bonferroni-corrected for multiple comparisons) are presented in Supplementary Table S1 for MEG and in Supplementary Table S2 for EEG.

For MEG, Friedman's test indicated significant effect of method on classification results ($\chi^2 = 34.29$, $p = 0.000$). The post-hoc multiple comparison test revealed that pooling, L1-MTL and L21-MTL were significantly better ($p < 0.05$) than the CSP-based methods. Also, no significant difference was found between within-subject L1 and inter-subject L21-MTL or pooling. For EEG, Friedman's test also indicated a significant effect of method on results ($\chi^2 = 19.84$, $p = 0.003$). In post-hoc multiple comparisons, it was found that L21-MTL was significantly better than CSP + bagging, regCSP and pooling. In this case, no significant difference was detected between within-subject L1 and inter-subject L21-MTL, L1-MTL or pooling.

| Subject | Within-subject L1 | Online decoding | Average |
|---|---|---|---|
| 1 | 86,25 | 71,30 | 78,78 |
| 2 | 88,75 | 77,50 | 83,13 |
| 4 | 71,25 | 75,00 | 73,13 |
| 6 | 85,00 | 67,50 | 76,25 |
| 7 | 82,50 | 62,50 | 72,50 |
| 8 | 58,75 | 85,00 | 71,88 |
| 9 | 86,25 | 68,80 | 77,53 |
| 10 | 63,75 | 63,80 | 63,78 |
| 13 | 95,00 | 57,50 | 76,25 |
| 15 | 95,00 | 53,80 | 74,40 |
| 16 | 77,50 | 51,30 | 64,40 |
| 17 | 86,25 | 56,30 | 71,28 |
| 18 | 91,25 | 55,00 | 73,13 |
| **3** | **48,75** | **47,50** | **48,13** |
| **5** | **60,00** | **52,50** | **56,25** |
| **11** | **66,25** | **60,00** | **63,13** |
| **12** | **53,75** | **66,30** | **60,03** |
| **14** | **53,75** | **58,80** | **56,28** |
| Mean | **75,00** | **62,80** | **68,90** |

**Table 1.** Accuracies for offline 10-fold cross-validation with MI training and online PM-trained decoding for individual subjects. The subjects discarded from the training data for inter-subject classification are written in bold.

| Subject | CSP + LDA | CSP + bagging | regCSP | Pooling | L1-MTL | L21-MTL | Within-subject L1 |
|---|---|---|---|---|---|---|---|
| 1 | 61,25 | 62,50 | 58,75 | 77,50 | 77,50 | 78,75 | 86,25 |
| 2 | 86,25 | 76,25 | 67,50 | 77,50 | 81,25 | 81,25 | 88,75 |
| 4 | 82,50 | 77,50 | 71,25 | 70,00 | 63,75 | 63,75 | 71,25 |
| 6 | 68,75 | 67,50 | 52,50 | 63,75 | 66,25 | 68,75 | 85,00 |
| 7 | 56,25 | 53,75 | 58,75 | 67,50 | 75,00 | 83,75 | 82,50 |
| 8 | 71,25 | 56,25 | 58,75 | 67,50 | 62,50 | 61,25 | 58,75 |
| 9 | 70,00 | 67,50 | 65,00 | 68,75 | 70,00 | 71,25 | 86,25 |
| 10 | 50,00 | 50,00 | 50,00 | 65,00 | 65,00 | 61,25 | 63,75 |
| 13 | 65,00 | 62,50 | 55,00 | 65,00 | 70,00 | 70,00 | 95,00 |
| 15 | 68,75 | 67,50 | 65,00 | 86,25 | 93,75 | 96,25 | 95,00 |
| 16 | 51,25 | 55,00 | 60,00 | 73,75 | 72,50 | 76,25 | 77,50 |
| 17 | 57,50 | 62,50 | 53,75 | 73,75 | 73,75 | 76,25 | 86,25 |
| 18 | 58,75 | 48,75 | 56,25 | 71,25 | 76,25 | 81,25 | 91,25 |
| 3 | 52,50 | 55,00 | 55,00 | 58,75 | 53,75 | 56,25 | 48,75 |
| 5 | 57,50 | 58,75 | 56,25 | 65,00 | 53,75 | 46,25 | 60,00 |
| 11 | 51,25 | 51,25 | 53,75 | 75,00 | 70,00 | 70,00 | 66,25 |
| 12 | 52,50 | 50,00 | 51,25 | 52,50 | 62,50 | 60,00 | 53,75 |
| 14 | 66,25 | 71,25 | 61,25 | 68,75 | 60,00 | 67,50 | 53,75 |
| mean | **62,64** | **60,76** | **58,33** | **69,31** | **69,31** | **70,56** | **75,00** |

**Table 2.** Accuracies of left- vs. right-hand MI decoding for the different methods using MI-trained MEG data.

**Offline training by passive movements.** The classification results for MEG data with training by passive movements are presented in Table 4 and for EEG data in Table 5. For all subjects, the online within-subject MEG decoding yielded an average accuracy of 62.8%, which was not significantly better than chance. For EEG, the offline within-subject decoding resulted in an average accuracy of 66.8%, which was significantly above chance level ($p = 0.008$). For both MEG and EEG data, none of the inter-subject methods yielded accuracies significantly above chance level.

The statistical differences between the methods ($p$-values, calculated by paired t-test, Bonferroni-corrected for multiple comparisons) are presented in Supplementary Table S3 for MEG and in Supplementary Table S4 for EEG.

| Subject | CSP + LDA | CSP + bagging | regCSP | Pooling | L1-MTL | L21-MTL | Within-subject L1 |
|---|---|---|---|---|---|---|---|
| 2 | 51,25 | 52,50 | 60,00 | 68,75 | 66,25 | 72,50 | 65,00 |
| 4 | 52,50 | 50,00 | 55,00 | 57,50 | 62,50 | 61,25 | 72,50 |
| 6 | 47,50 | 42,50 | 50,00 | 37,50 | 55,00 | 45,00 | 62,50 |
| 7 | 58,75 | 57,50 | 50,00 | 82,50 | 73,75 | 93,75 | 88,75 |
| 8 | 78,75 | 67,50 | 61,25 | 57,50 | 58,75 | 62,50 | 73,75 |
| 9 | 71,25 | 80,00 | 75,00 | 68,75 | 66,25 | 71,25 | 65,00 |
| 10 | 70,00 | 68,75 | 52,50 | 67,50 | 71,25 | 73,75 | 56,25 |
| 13 | 55,00 | 58,75 | 56,25 | 77,50 | 65,00 | 76,25 | 80,00 |
| 15 | 56,25 | 62,50 | 53,75 | 81,25 | 72,50 | 86,25 | 85,00 |
| 16 | 52,50 | 53,75 | 53,75 | 66,25 | 71,25 | 72,50 | 67,50 |
| 17 | 51,25 | 50,00 | 53,75 | 57,50 | 53,75 | 65,00 | 73,75 |
| 18 | 66,25 | 58,75 | 60,00 | 80,00 | 86,25 | 86,25 | 83,75 |
| 3 | 67,50 | 60,00 | 53,75 | 42,50 | 46,25 | 50,00 | 56,25 |
| 5 | 58,75 | 56,25 | 56,25 | 62,50 | 65,00 | 57,50 | 70,00 |
| 11 | 60,00 | 57,50 | 50,00 | 70,00 | 67,50 | 61,25 | 51,25 |
| 12 | 50,00 | 50,00 | 50,00 | 46,25 | 51,25 | 51,25 | 65,00 |
| 14 | 81,25 | 85,00 | 70,00 | 62,50 | 67,50 | 63,75 | 62,50 |
| mean | **60,51** | **59,49** | **56,54** | **63,90** | **64,71** | **67,65** | **69,34** |

Table 3. Accuracies of left- vs. right-hand MI decoding for the different methods using MI-trained EEG data.

| Subject | CSP + LDA | CSP + bagging | regCSP | Pooling | L1-MTL | L21-MTL | Online CSP |
|---|---|---|---|---|---|---|---|
| 1 | 61,25 | 68,75 | 57,50 | 63,75 | 57,50 | 65,00 | 71,30 |
| 2 | 81,25 | 72,50 | 63,75 | 46,25 | 58,75 | 56,25 | 77,50 |
| 4 | 78,75 | 87,50 | 50,00 | 53,75 | 50,00 | 51,25 | 75,00 |
| 6 | 58,75 | 51,25 | 61,25 | 55,00 | 56,25 | 53,75 | 67,50 |
| 7 | 52,50 | 50,00 | 56,25 | 58,75 | 56,25 | 57,50 | 62,50 |
| 8 | 57,50 | 50,00 | 76,25 | 63,75 | 57,50 | 63,75 | 85,00 |
| 9 | 62,50 | 66,25 | 62,50 | 53,75 | 53,75 | 52,50 | 68,80 |
| 10 | 60,00 | 50,00 | 61,25 | 56,25 | 57,50 | 61,25 | 63,80 |
| 13 | 52,50 | 53,75 | 56,25 | 47,50 | 48,75 | 46,25 | 57,50 |
| 15 | 43,75 | 48,75 | 50,00 | 66,25 | 60,00 | 65,00 | 53,80 |
| 16 | 51,25 | 50,00 | 51,25 | 47,50 | 46,25 | 40,00 | 51,30 |
| 17 | 56,25 | 53,75 | 57,50 | 63,75 | 63,75 | 60,00 | 56,30 |
| 18 | 50,00 | 50,00 | 56,25 | 63,75 | 56,25 | 57,50 | 55,00 |
| 3 | 51,25 | 50,00 | 48,75 | 51,25 | 52,50 | 46,25 | 47,50 |
| 5 | 58,75 | 56,25 | 50,00 | 58,75 | 53,75 | 53,75 | 52,50 |
| 11 | 51,25 | 57,50 | 53,75 | 61,25 | 47,50 | 41,25 | 60,00 |
| 12 | 52,50 | 65,00 | 50,00 | 50,00 | 36,25 | 40,00 | 66,30 |
| 14 | 56,25 | 50,00 | 52,50 | 56,25 | 51,25 | 53,75 | 58,80 |
| mean | **57,57** | **57,29** | **56,39** | **56,53** | **53,54** | **53,61** | **62,80** |

Table 4. Accuracies of left- vs. right-hand MI decoding for the different methods using PM-trained MEG data.

For MEG, Friedman's test showed a significant difference between methods ($\chi^2 = 16.81$, $p = 0.01$). The post-hoc multiple comparison test showed that within-subject decoding was significantly better than all inter-subject methods. For EEG, the results were quite similar; Friedman's test indicated significant effect of method ($\chi^2 = 34.21$, $p = 0.000$) and the post-hoc multiple comparison test revealed that within-subject decoding was superior to all inter-subject methods. Also for EEG, pooling yielded results inferior to all other inter-subject methods.

## Discussion

In the current study, we investigated methods for inter-subject classification of motor imagery from MEG and EEG signals. We found that multi-task joint feature learning with logistic regression and $l_{2,1}$-norm regularization is an effective method for decoding left- and right-hand MI from both MEG and EEG data. The accuracy of L21-MTL was better than that of the other inter-subject decoding methods both for MEG and EEG data, when other subjects' MI was used for training. In addition, with this method we achieved accuracies significantly

| Subject | CSP + LDA | CSP + bagging | regCSP | Pooling | L1-MTL | L21-MTL | Within-subject CSP |
|---|---|---|---|---|---|---|---|
| 2 | 57,50 | 50,00 | 51,25 | 50,00 | 50,00 | 50,00 | 71,3 |
| 4 | 58,75 | 62,50 | 57,50 | 50,00 | 55,00 | 56,25 | 73,8 |
| 6 | 50,00 | 55,00 | 50,00 | 50,00 | 46,25 | 42,50 | 71,1 |
| 7 | 50,00 | 48,75 | 43,75 | 50,00 | 66,25 | 51,25 | 55,0 |
| 8 | 60,00 | 60,00 | 53,75 | 52,50 | 57,50 | 58,75 | 97,5 |
| 9 | 77,50 | 85,00 | 50,00 | 50,00 | 51,25 | 48,75 | 81,3 |
| 10 | 61,25 | 58,75 | 62,50 | 50,00 | 53,75 | 46,25 | 66,3 |
| 13 | 50,00 | 50,00 | 58,75 | 50,00 | 51,25 | 46,25 | 56,3 |
| 15 | 52,50 | 51,25 | 52,50 | 50,00 | 53,75 | 52,50 | 63,2 |
| 16 | 51,25 | 56,25 | 47,50 | 50,00 | 60,00 | 57,50 | 50,0 |
| 17 | 56,25 | 50,00 | 58,75 | 50,00 | 56,25 | 65,00 | 61,3 |
| 18 | 48,75 | 50,00 | 56,25 | 48,75 | 63,75 | 57,50 | 59,3 |
| 3 | 52,50 | 56,25 | 55,00 | 50,00 | 55,00 | 53,75 | 65,0 |
| 5 | 53,75 | 50,00 | 52,50 | 50,00 | 45,00 | 56,25 | 58,8 |
| 11 | 56,25 | 51,25 | 53,75 | 50,00 | 47,50 | 45,00 | 67,5 |
| 12 | 48,75 | 50,00 | 47,50 | 50,00 | 48,75 | 50,00 | 53,8 |
| 14 | 50,00 | 61,25 | 66,25 | 50,00 | 46,25 | 45,00 | 83,8 |
| mean | **55,25** | **55,66** | **53,97** | **50,07** | **53,38** | **51,91** | **66,76** |

**Table 5.** Accuracies of left- vs. right-hand MI decoding for the different methods using PM-trained EEG data.

above chance level for both modalities. Also L1-MTL resulted in good average accuracy across subjects. When trained by MI, none of the CSP + LDA –based classification methods tested in this study reached an average accuracy equal to or higher than that of L21-MTL, L1-MTL or pooling with either MEG or EEG. Importantly, for both MEG and EEG, there was no statistical difference between within-subject L1 decoding and inter-subject L21-MTL decoding, suggesting that on average the performance of those methods was similar.

Besides the average accuracy across subjects, it is also important to consider how inter-subject decoding affected the performance of individual subjects. Many subjects showed poor performance during the online BCI task, when the decoder was trained using their own PM data. By using the subject's own MI data and cross-validation, the accuracy was improved for most of the poor online performers. This offline analysis implies that the activity patterns of left- and right-hand MI were indeed discriminable in most subjects with low accuracy in the online task. It seems that PM training is not robust enough for MI classification, since there was large inter-subject variability in online decoding.

However, it is also noteworthy that some subjects have poor decoding performance because of insufficient SMR modulation. For example, Subjects 3 and 5 did not show any discernible SMR modulation during MI, and their accuracy remained under chance level even when their own MI was used for training. In case of the mentioned subjects, inter-subject decoding did not remarkably improve the results. On the contrary, for Subjects 11, 12 and 14, L21-MTL yielded better results with MEG (also with EEG for Subjects 11 and 14) compared to within-subject L1. We may thus conclude that for poorly-performing subjects showing at least some SMR modulation, inter-subject learning might improve the performance. On the contrary, for the good performers using other subjects' training data was not beneficial, as most of them got worse results compared to training with their own data. However, inter-subject training still has the advantage of saving time in BCI sessions. It is an interesting question for future studies whether the good performers can adapt their brain activity such that they eventually get equally good accuracy with the generalized decoders.

As expected, training with PM resulted in lower inter-subject accuracy compared to training with MI – in fact, in this case the decoding accuracy was below chance level with all methods and both MEG and EEG. It is likely that this result is due to overfitting. It seems that the relevant features in PM data are different than the features in MI data, and thus it is difficult to classify MI using a classifier trained with other subjects' PM. Some subjects achieved good performance when their own PM data were used for training, but the decoding accuracy decreased substantially when the training data were collected from other subjects. As a conclusion: even if the subject's own PM pattern would be similar enough to the MI pattern to enable decoding of MI, the inter-subject generalized PM features are likely too different from the individual MI features. According to an EEG study by Formaggio and co-workers[31], the activation patterns caused by MI are more lateralized than those caused by PM, which might explain this difference. However, the exact neural mechanisms underlying the difference between PM and MI remains unclear and should be investigated in the future. With regard to these findings, it can be argued that in inter-subject learning the performed mental task should be as similar as possible for the train and test subjects.

In their MEG classification study, Kia and co-workers[17] used a bootstrapping approach for dividing the data into different tasks, i.e. the samples for train and test datasets were randomly selected from a pool of data from different subjects. As samples for every task were collected from the same database, it is inevitable that MEG data samples from one subject are present in multiple tasks as well as in the test data. Therefore, the performance of the classifier may not reflect the true inter-subject generalizability. In the current study, we modified the approach of Kia and co-workers[17] and divided the data in discrete training and test datasets to ensure that no training data were collected from the test subject. It is noteworthy that the modified classifier still yielded

sufficient inter-subject accuracy in MI classification. Kia and co-workers[17] used the L21-MTL method for classifying simulated MEG data and real MEG responses to visual stimuli, but to our knowledge the current study is the first one in which it was used to decode motor imagery.

According to our results, MTL seems to yield slightly but not significantly better results for MEG than EEG. In EEG we used fewer measurement channels and thus less features for classification than in MEG, which might explain the results. However, selecting an equal number of EEG electrodes and MEG sensors would not be neurophysiologically justified, as they would then cover a different area of the scalp. Therefore, we chose to select channels above the motor cortex in both modalities. It is also possible that MEG has a better signal-to-noise ratio than EEG, or that MEG is able to detect signals from sources not clearly visible to EEG. In light of these results, probably the best approach would be to utilize information from both MEG and EEG.

For the offline analyses, we wanted to have optimal data quality. Therefore, we suppressed interference in MEG with the signal-space separation method, aligned the head positions, and manually removed artifacts from EEG. It should be pointed out that this kind of thorough preprocessing is often not possible, or at least not practical, during online analysis. Head position alignment could be implemented in an online MEG data processing pipeline in the near future, so its use in the offline analysis was considered reasonable. In the case of EEG, one should carefully avoid artifacts during online decoding.

The motivation of decoding left-vs-right hand MI is based on the assumption that neural plasticity during stroke recovery might result in decreased functionality if inappropriate activation patterns are enhanced. Many stroke patients with motor disabilities have difficulties in SMR modulation in the affected hemisphere. Instead, the unaffected hemisphere shows increased motor-related activity due to reduced intracortical inhibition. Therefore, discriminating between MI and resting state is not likely a good target for a rehabilitative BCI, since it might only boost the activity of the motor areas on the healthy hemisphere instead of the affected hemisphere. Here we have shown that inter-subject decoding of left-vs-right MI from MEG and EEG yields reasonably good accuracy. The presented L21-MTL method can also be applied to online decoding.

In future studies, we aim at designing an adaptive online BCI paradigm based on the results of the current study. The BCI training could be started with generalized classifiers, and in the following sessions, as the patients' performance improves, their own data could be added to the training set in order to update the classifiers, as has been proposed in recent studies[32,33]. Another valid option is to use only the generalized classifier trained with other subjects. In this case, the test subjects should adapt their neural activity to match the MI representation of the healthy subject population. The latter approach would encourage the patients to change the activity of the damaged motor cortex towards normal function, which could boost rehabilitation. As the key concept in BCI therapy is to enhance activity-dependent brain plasticity[34,35], it is crucial to ensure that the targeted activity corresponds to the normal motor-cortex function. Therefore, calibrating the BCI with healthy subjects' data might actually lead to improved clinical outcome compared to calibration with patient's own data. This issue will also be addressed in future studies.

## Methods

**Participants.**    Eighteen healthy volunteers (5 males, 13 females, age $27.7 \pm 5.0$ y, 3 left handed by self-report) were recruited for this study. None of the subjects reported any neurological illnesses or motor deficits, and all subjects had normal or corrected-to-normal vision. Subjects having any metal objects inside their body or other MEG contraindications were excluded from the study. The study was approved by the Aalto University Research Ethics Committee. The research was carried out in accordance with the guidelines of the Declaration of Helsinki, and the subjects gave written informed consent prior to the measurements. Two of the subjects had participated in our previous experiment[11], and the others had no previous BCI experience.

**Pneumatic hand stimulators.**    The stimulators used in this study consisted of two plastic frames supporting each hand, and of 8 elastic pneumatic artificial muscles (PAM; DMSP-10-100 AM-CM, diameter 10 mm, length of the contracting part 100 mm; Festo AG & Co, Esslingen, Germany) attached to fingers 2–5 of both hands with tape. During the experiment, the subject's hands and arms were resting on the plastic frames and the PAMs were touching the fingertips horizontally from beneath the frame (Fig. 2). The PAMs were connected to the frame via moving hinges, so that the stimulator system could be adjusted to fit the subject's hand.

During the stimulation paradigm, the PAMs moved horizontally according to the internal air pressure (1–4 bar). The pressure to each PAM was switched on and off by a solenoid valve (SY5220-6LOU-01F-Q, SMC Corporation, Tokyo, Japan) controlled by trigger pulses. The valves were placed outside the magnetically shielded room, and 3.5-m long semi-elastic tubes (internal diameter 2.5 mm) conveyed the pressurized air to the PAMs. A computer controlled the valves such that all the four PAMs for either left or right hand contracted sequentially (each for 500 ms) at intervals of 500 ms and flexed the fingers, thus mimicking small finger-tapping movement. After the contraction, each PAM returned to its resting length and extended the corresponding finger, returning it back to the initial positions. This four-finger movement sequence was repeated twice during each epoch.

**MEG and EEG recordings.**    MEG was recorded at the MEG Core of Aalto University School of Science with a 306-channel Elekta Neuromag™ Vectorview system (Elekta Oy, Helsinki, Finland) using a sampling frequency of 1 kHz. Five head-position indicator (HPI) coils were attached to the subject's scalp for head position estimation and alignment with a standard coordinate system. Visual stimuli were delivered on a screen located approximately 50 cm from the subject's eyes via a projector outside the shielded room. EEG was measured simultaneously using a MEG-compatible cap with 60 electrodes and 1-kHz sampling rate. However, only the MEG data were used for real-time classification.

Raw MEG data were continuously written in 300-ms segments to a network-transparent ring buffer[36,37] hosted by the MEG acquisition workstation (6-core Intel Xeon CPU at 2.4 GHz, 64-bit CentOS Linux, version 5.3). The
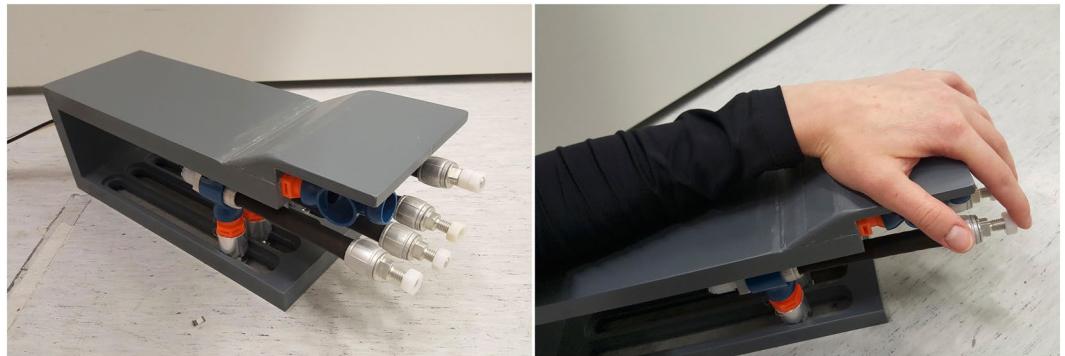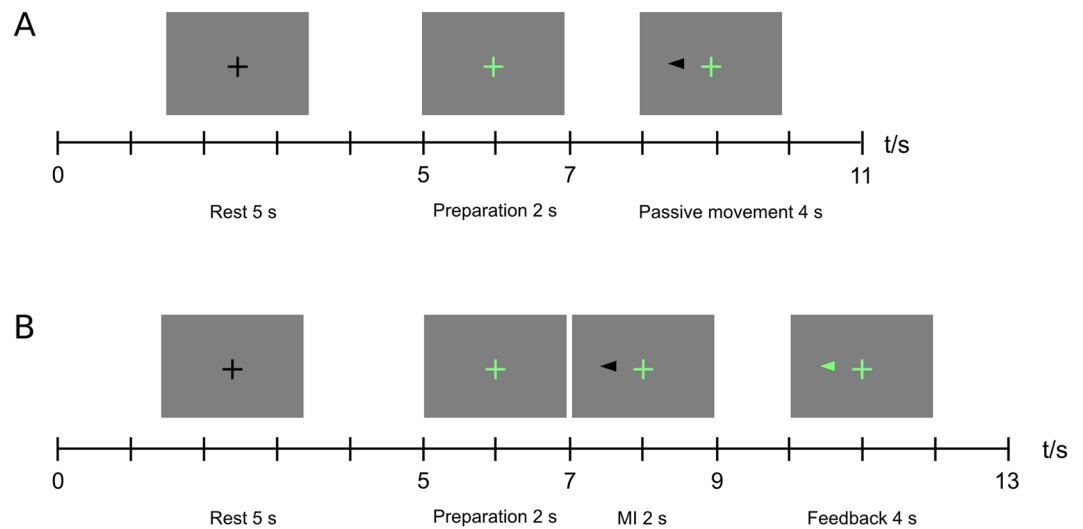
**Figure 2.** Pneumatic passive-movement stimulator.



**Figure 3.** The experimental paradigm in (**A**) passive movement, (**B**) motor imagery tasks.

buffer was read over a local network connection by another computer (Intel Core i7-4771 CPU at 3.5 GHz, 64-bit Ubuntu Linux, version 14.04). This computer processed the MEG data in real time using functions implemented in the MNE-Python software[38], controlled the pneumatic hand stimulators and presented the visual stimuli using PsychoPy version 1.83[39].

**Experimental paradigm.** Each measurement session consisted of the two tasks described below. The whole experimental paradigm is illustrated in Fig. 3. The subjects were allowed to rest eyes closed between the tasks for a few minutes. The duration of each task was approximately 15 min, and the whole experiment lasted for approximately 40 min including breaks.

In the passive-movement task, hand movements were elicited by the pneumatic stimulators according to the predefined sequence. Each sequence onset was accompanied with a visual cue. The subjects were instructed to keep their muscles relaxed and gaze on a fixation cross throughout the task. A total of 80 passive-movement epochs, 40 for each hand in a pseudo-randomized order, were delivered. Each epoch began with a gray background and a black fixation cross in the middle of the screen. At second 5, the fixation cross turned green, indicating preparation for the upcoming movements. At second 7, a black cue arrow appeared on either left or right side of the screen, and the stimulators moved the subject's fingers on the corresponding hand continuously for 4 s. At second 11, a black fixation cross was shown and the PAMs returned to their original positions, leaving the fingers at rest.

After the first task, the subject was instructed to relax for a few minutes without making large head movements. In the meantime, a classifier was trained using the MEG data collected during passive movements. Signals from 64 MEG gradiometers over the parietal areas were included in the online analyses (see Fig. 4). First, the frequency band showing the strongest SMR suppression was estimated by comparing the SMR power at rest, i.e. during the pauses in the first experiment, and during passive movements. Power decrease was calculated in 1-Hz frequency bins in 8–30-Hz range. The center frequency was defined as the frequency with the highest power decrease, and the frequency band of interest as the center frequency ±3 Hz. After that, relevant signal features were extracted with SSD[24] and consecutive Common Spatial Patterns (CSP) filters. The frequency band of interest was used for SSD. Other details for SSD and CSP can be found in our previous study[11]. Finally, a Linear Discriminant Analysis
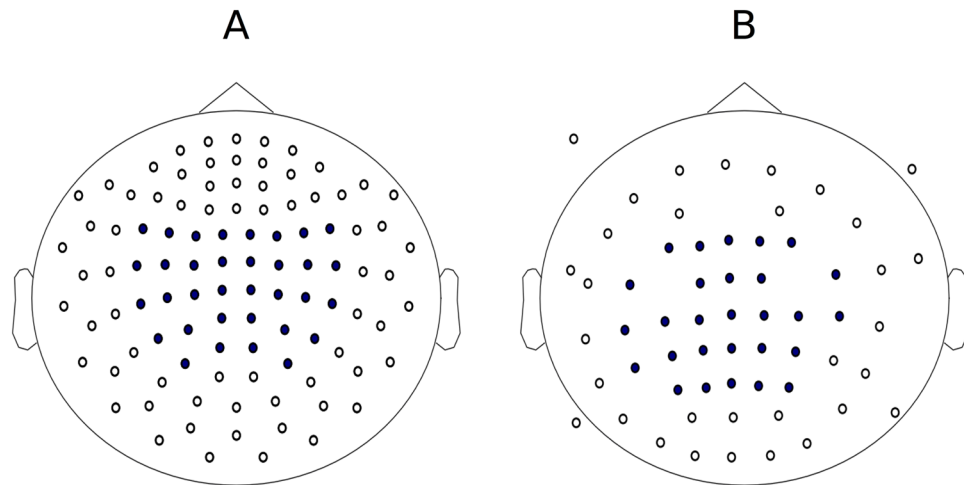
**Figure 4.** Locations of (**A**) MEG sensors and (**B**) EEG electrodes (marked in blue) included in analyses.

(LDA) classifier was fit to the data to discriminate between left- and right-hand-imagery-related suppression of the SMR.

In the motor imagery (MI) task, the subjects were instructed to imagine finger movements similar to the passive movements they received in the earlier task and to pay attention to the visual and tactile feedback following each MI period. The aforementioned SSD and CSP filters and the LDA classifier were applied in real time to each MI epoch to predict left- vs. right-hand MI.

The MI task comprised 80 trials of MI with feedback. Each trial began with a black fixation cross in the middle of the screen. At second 5, the fixation cross turned green, indicating preparation for the upcoming movement. At second 7, a black cue arrow appeared on either side of the screen, indicating the beginning of MI. At second 9, the feedback started; if the MI trial was successful, i.e. the classification probability of the target hand exceeded 50%, the cue arrow turned green and the predefined sequence of PAM stimulator movements started, lasting 4 s. In case of an unsuccessful MI trial, the cue arrow turned red and the PAMs did not move.

**Preprocessing for offline analyses.** Similar preprocessing routines were applied to the data from both tasks. First, the data were processed with the MaxFilter software (version 2.2.15; Elekta Oy, Helsinki, Finland)[40], including transformation of head positions of all subjects into the same orientation, signal space separation for suppressing external magnetic interference and removal of static bad channels. The common-average reference was utilized for all analyses of EEG data. Both MEG and EEG data were bandpass-filtered to 6–45 Hz with a fast Fourier transform (FFT) -based finite impulse response (FIR) filter, with filter length equal to the epoch length. The data were split into 3-s epochs and baseline corrected (epoch $-1.0\ldots2.0$ s, baseline $-1.0\ldots0.0$ s from cue onset). Artifact correction for EEG data was done by performing independent component analysis (ICA; implemented in MNE-Python[38]) and manually removing components including artifacts or high-amplitude noise. 64 MEG sensors and 28 EEG electrodes close to parietal areas were included in further analyses (see Fig. 4).

**Train and test data in offline analyses.** The measurements were split into train and test data in a leave-one-subject-out manner, i.e. the data from all except one subject were used for training the classifiers and the remaining subject's data for testing classification performance. The MEG and EEG signals were evaluated separately. In addition, we evaluated the validity of both passive movements (PM) and MI of other subjects for training the classifier. In case of PM, the training data were PM epochs from all but one subject and the test data were MI epochs from the remaining subject.

**Feature extraction and classification.** *Spatial filtering and LDA.* Before CSP filtering, the data dimensionality was reduced with spatio–spectral decomposition (SSD)[24]. The frequency band of interest was defined as $10\pm3$ Hz in both datasets for all subjects, and 20 SSD components were retained. CSP decomposition was calculated on the SSD-filtered data. Ten CSP components were extracted, averaged over time and log-transformed, and finally used as features for the LDA classifier.

In the CSP + bagging approach, the LDA classifier was first fit to each training subject's data separately, and the individual predictions were aggregated to form the final predictions for the test data. The BaggingClassifier algorithm of the scikit-learn Python package[41] was modified such that instead of dividing the training data into random subsets, each division contained all data from one train subject.

The regularized CSP algorithm was adopted from Regularized CSP (RCSP) Matlab toolbox[21] and implemented in Python. The generic covariance matrices $\mathbf{G}_c$ for each class $c$ were calculated as weighted sums of covariance matrices $\mathbf{C}_c$ from $N$ training subjects[22]:

$$\mathbf{G}_c = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{fn}\mathbf{C}_c^i \tag{1}$$

where $f$ is the Frobenius norm:

$$fn = \sum_{i,j}[\mathbf{C}_{\text{test}}(i, j) - \mathbf{C}_{\text{train}}(i, j)]^2 \tag{2}$$

The objective function of RCSP was formulated as [Lotte & Guan 2010]:

$$J_{P\{1,2\}}(\mathbf{w}) = \frac{\mathbf{w}^{\mathrm{T}}\widetilde{\mathbf{C}}_{\{1,2\}}\mathbf{w}}{\mathbf{w}^{\mathrm{T}}\widetilde{\mathbf{C}}_{\{2,1\}}\mathbf{w}} \tag{3}$$

where

$$\widetilde{\mathbf{C}}_c = (1 - \gamma)\hat{\mathbf{C}}_c - \gamma I \tag{4}$$

and

$$\hat{\mathbf{C}}_c = (1 - \beta)\mathbf{C}_c + \beta\mathbf{G}_c \tag{5}$$

Values for parameters $\beta$ and $\gamma$ were selected by cross-validation such that one subject was used for testing, half of the remaining subjects for parameter optimization and the other half for training. The parameters were optimized discretely on range from 0 to 1 with a step size of 0.1. Leave-one-out cross-validation was used for the parameter optimization dataset to find out the parameters yielding the best classification results, and those parameters were used for the actual training and testing.

*Multi-task joint feature learning.* The MTL optimization problem can be expressed as an empirical risk minimization problem:

$$\mathbf{W} = \operatorname{argmin}\sum_{i=1}^{s}\mathcal{L}(\mathbf{y}^{(i)}, \mathbf{X}^{(i)}\mathbf{w}_{,i}) + \lambda\mathbf{W}_{2,1} \tag{6}$$

where $\mathbf{W}_{2,1}$ is the $l_{2,1}$ norm for matrix $\mathbf{W}$:

$$\|\mathbf{W}_{2,1}\| = \sum_{j=1}^{p}\|\mathbf{W}_{j,:}\|_2 \tag{7}$$

The multi-task logistic regression minimization problem can thus be formulated as:

$$\min_{W,\mathbf{c}}\sum_{i=1}^{t}\sum_{j=1}^{n_i}\log(1 + \exp(-\mathbf{y}_{i,j}(\mathbf{W}_j^T\mathbf{X}_{i,j} + \mathbf{c}_i))) + \rho\|\mathbf{W}\|_{2,1} \tag{8}$$

where $\mathbf{X}_{i,j}$ denotes sample $j$ of the $i$-th task, $\mathbf{y}_{i,j}$ denotes the corresponding label, $\mathbf{W}_i$ and $\mathbf{c}_i$ are the model for task $i$ and the regularization parameter $\rho_1$ controls group sparsity.

Prior to feature selection and classification, the epochs were decimated by 10, i.e. every 10th timepoint was selected, to speed up calculation. The bootstrap data sampling scheme used in the L21-MTL algorithm of MTJFL toolbox was replaced by a leave-one-subject-out scheme, in which no data from the test subject were included in the training set. In the MTL, the data were split such that each task included all data from a single train subject, and the test data included all data from the test subject. In the pooling approach, the data from all except one subject were used for training, and the data from the remaining subject for testing. This train/test split ensured that the classifier actually generalized between the subjects and did not use any of the test subject's data for training. In the within-subject decoding, 10-fold cross-validation was used for dividing each subject's data into train and test sets. In all approaches, the regularization constant lambda was selected by choosing the parameter yielding the best results in L21-MTL classification, and a similar lambda was used also for L1-MTL. The parameter was selected separately for all datasets and conditions.

**Statistical analysis of classification results.** The chance level for left-vs-right MI classification was defined according to the study by Combrisson and Jerbi[30]. The probability for getting a correct classification by chance is 58.7% for 80 samples and a $p$-value of 0.05, and we used this value as a threshold for determining whether the decoding results differ from chance level.

The statistical difference between results of different classification approaches was calculated with Friedman's test, a nonparametric version of two-way analysis of variance. Post-hoc analysis was conducted using Bonferroni correction for multiple comparisons, implemented in MATLAB (R2017a; Mathworks Inc., MA, USA) function *multcompare*.

**Data availability.** The data measured and analysed for the current study are available from the corresponding author on request.

## References

1. Teo, W. P. & Chew, E. Is motor-imagery brain-computer interface feasible in stroke rehabilitation? *PM R* **6**, 723–728 (2014).
2. van Dokkum, L. E. H., Ward, T. & Laffont, I. Brain computer interfaces for neurorehabilitation – its current status as a rehabilitation strategy post-stroke. *Ann. Phys. Rehabil. Med.* **58**, 3–8 (2015).
3. McCrimmon, C. M., Wang, P. T., Nenadic, Z. & Do, A. H. In *Neurorehabilitation Technology* (eds. Reinkensmeyer, D. J. & Dietz, V.) 617–627 (Springer International Publishing), https://doi.org/10.1007/978-3-319-28603-7_30 (2016).
4. Mellinger, J. *et al.* An MEG-based Brain-Computer Interface (BCI). *Neuroimage* **36**, 581–593 (2007).
5. Buch, E. *et al.* Think to Move: a Neuromagnetic Brain-Computer Interface (BCI) System for Chronic Stroke. *Stroke* **39**, 910–917 (2008).
6. Foldes, S. T., Weber, D. J. & Collinger, J. L. MEG-based neurofeedback for hand rehabilitation. *J. Neuroeng. Rehabil.* **12**, 85 (2015).
7. Foldes, S. T., Vinjamuri, R. K., Wang, W., Weber, D. J. & Collinger, J. L. Stability of MEG for real-time neurofeedback. in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* 5778–5781 (IEEE), https://doi.org/10.1109/IEMBS.2011.6091430 (2011).
8. Devlaminck, D., Wyns, B., Grosse-Wentrup, M., Otte, G. & Santens, P. Multisubject learning for common spatial patterns in motor-imagery BCI. *Comput. Intell. Neurosci.* **2011**, 217987 (2011).
9. Olivetti, E., Kia, S. M. & Avesani, P. MEG decoding across subjects. In *Proceedings - 2014 International Workshop on Pattern Recognition in Neuroimaging, PRNI 2014*, https://doi.org/10.1109/PRNI.2014.6858538 (2014).
10. Ray, A. M. *et al.* A subject-independent pattern-based Brain-Computer Interface. *Front. Behav. Neurosci.* **9**, 269 (2015).
11. Halme, H.-L. & Parkkonen, L. Comparing features for classification of MEG responses to motor imagery. *PLoS One* **11**, (2016).
12. Obozinski, G., Taskar, B. & Jordan, M. Multi-task feature selection. *Stat. Dep. UC Berkeley, Tech. Rep* **2** (2006).
13. Argyriou, A., Evgeniou, T. & Pontil, M. Multi-Task Feature Learning. *Adv. Neural Inf. Process. Syst.* 41–48 (2007).
14. Tomioka, R., Aihara, K. & Müller, K.-R. Logistic regression for single trial EEG classification (2007).
15. Tomioka, R. & Müller, K.-R. A regularized discriminative framework for EEG analysis with application to brain–computer interface. *Neuroimage* **49**, 415–432 (2010).
16. Nie, F., Huang, H., Cai, X. & Ding, C. H. In *Advances in Neural Information Processing* Systems 23 (eds Lafferty, J. D., Williams, C. K. I., Shawe-Taylor, J., Zemel, R. S. & Culotta, A.) 1813–1821 (Curran Associates, Inc., 2010).
17. Kia, S. M., Pedregosa, F., Blumenthal, A. & Passerini, A. Group-level spatio-temporal pattern recovery in MEG decoding using multi-task joint feature learning. *J. Neurosci. Methods* **285**, 97–108 (2017).
18. Koles, Z. J., Lazar, M. S. & Zhou, S. Z. Spatial patterns underlying population differences in the background EEG. *Brain Topogr.* **2**, 275–284 (1990).
19. Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M. & Muller, K. Optimizing Spatial filters for Robust EEG Single-Trial Analysis. *IEEE Signal Process. Mag.* **25**, 41–56 (2008).
20. Pfurtscheller, G. & Neuper, C. Motor imagery and direct brain-computer communication. *Proc. IEEE* **89**, 1123–1134 (2001).
21. Lotte, F. & Guan, C. Regularizing Common Spatial Patterns to Improve BCI Designs: Unified Theory and New Algorithms. *IEEE Trans. Biomed. Eng.* **58**, 355–362 (2011).
22. Song, X., Yoon, S. C. & Perera, V. Adaptive Common Spatial Pattern for single-trial EEG classification in multisubject BCI. in *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)* 411–414, https://doi.org/10.1109/NER.2013.6695959 (2013).
23. Cheng, M., Lu, Z. & Wang, H. Regularized common spatial patterns with subject-to-subject transfer of EEG signals. *Cogn. Neurodyn.* **11**, 173–181 (2017).
24. Nikulin, V. V., Nolte, G. & Curio, G. A novel method for reliable and fast extraction of neuronal EEG/MEG oscillations on the basis of spatio-spectral decomposition. *Neuroimage* **55**, 1528–1535 (2011).
25. Arvaneh, M. *et al.* Facilitating motor imagery-based brain–computer interface for stroke patients using passive movement. *Neural Comput. Appl.* 1–14, https://doi.org/10.1007/s00521-016-2234-7 (2016).
26. Kaiser, V., Kreilinger, A., Müller-Putz, G. R. & Neuper, C. First steps toward a motor imagery based stroke BCI: New strategy to set up a classifier. *Front. Neurosci.*, https://doi.org/10.3389/fnins.2011.00086 (2011).
27. Ang, K. K. *et al.* A clinical study of motor imagery BCI performance in stroke by including calibration data from passive movement. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, https://doi.org/10.1109/EMBC.2013.6611069 (2013).
28. Ramos-Murguialday, A. & Birbaumer, N. Brain oscillatory signatures of motor tasks. *J. Neurophysiol.* **113**, 3663–3682 (2015).
29. Zhou, J., Chen, J. & Ye, J. User's Manual MALSAR: Multi-tAsk Learning via StructurAl Regularization (2012).
30. Combrisson, E. & Jerbi, K. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *J. Neurosci. Methods* **250**, 126–136 (2015).
31. Formaggio, E. *et al.* Modulation of event-related desynchronization in robot-assisted hand performance: brain oscillatory changes in active, passive and imagined movements. *J. Neuroeng. Rehabil.* **10**, 24 (2013).
32. Kobler, R. J. & Scherer, R. Restricted Boltzmann Machines in Sensory Motor Rhythm Brain-Computer Interfacing: A study on inter-subject transfer and co-adaptation. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* 469–474, https://doi.org/10.1109/SMC.2016.7844284 (2016).
33. Acqualagna, L., Botrel, L., Vidaurre, C., Kübler, A. & Blankertz, B. Large-Scale Assessment of a Fully Automatic Co-Adaptive Motor Imagery-Based Brain Computer Interface. *PLoS One* **11**, e0148886 (2016).
34. Soekadar, S. R., Birbaumer, N., Slutzky, M. W. & Cohen, L. G. Brain-machine interfaces in neurorehabilitation of stroke. *Neurobiol. Dis.* **83** (2015).
35. Daly, J. J. & Wolpaw, J. R. Brain–computer interfaces in neurological rehabilitation. *Lancet Neurol.* **7**, 1032–1043 (2008).
36. Sudre, G. *et al.* rtMEG: A Real-Time Software Interface for Magnetoencephalography. *Comput. Intell. Neurosci.* **2011**, 1–7 (2011).
37. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive ElectrophysiologicalData. *Comput. Intell. Neurosci.* **2011**, 1–9 (2011).
38. Gramfort, A. *et al.* MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* **7**, 267 (2013).
39. Peirce, J. W. PsychoPy-Psychophysics software in Python. *J. Neurosci. Methods* **162**, 8–13 (2007).
40. Taulu, S., Kajola, M. & Simola, J. Suppression of interference and artifacts by the signal space separation method. *Brain Topogr.* **16**, 269–275 (2004).
41. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

## Acknowledgements

## Author Contributions

H.-L.H. and L.P. designed the studies. Data were collected by H.-L.H. Data analyses were carried out by H.-L.H. The manuscript was written by H.-L.H. and L.P. Figures were prepared by H.-L.H. Both authors reviewed and accepted the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at https://doi.org/10.1038/s41598-018-28295-z.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.