USING MACHINE LEARNING TO IMPROVE MOTOR IMAGERY NEUROFEEDBACK

by

Ross Story

Submitted in partial fulfillment of the
requirements for the degree of
Master of Science

at

Dalhousie University
Halifax, Nova Scotia
July 2015

© Copyright by Ross Story, 2015

Table of Contents

List of Tables

List of Figures

Abstract

Machine Learning (ML) was employed to identify features in MEG recordings of 16 participants performing MI. ML was applied to data obtained using three methods: 1) two sources localized to somatosensory cortex; 2) all 306 MEG sensors; and 3) 80 anatomical sources localized via known coordinates. A linear kernel support vector machine was fit to frequency features extracted by fast fourier transform and accuracy (ability to accurately class trials as rest or imagery) evaluated with leave one batch out cross validation. The two-source model performed poorly with accuracy of 0.7, while the 80-source and sensor-level models performed well with accuracy greater than 0.9. Sensor-level models selected sensors analogous to C4, FC4, and F10 as key sensors. Use of ML to identify features in MI-based neuroimaging data can be applied to make MI interventions clinically feasible options through reducing cost, setup time, and expertise required to employ neurofeedback.

List of Abbreviations Used

ANOVA	Analysis of Variance.
BCI	Brain Computer Interface.
BOLD	Blood Oxygen Level Dependent.
CWT	Continuous Wavelet Transform.
DFL	Dorsal Forelimb.
ECD	Equivalent Current Dipole.
ERD	Event Related Desynchronization.
fMRI	Functional Magnetic Resonance Imaging.
LDA	Linear Discriminant Analysis.
LOBO	Leave One Batch Out.
LSTM	Long Short-Term Memory.
LSVM	Linear Support Vector Machine.
ME	Motor Execution.
MEG	Magnetoencephalography.
MI	Motor Imagery.
ML	Machine Learning.
PCA	Principal Components Analysis.
PMBR	Post-Movement Beta Rebound.
RFE	Recursive Feature Elimination.

RSE	Recursive Sensor Elimination.
RTP	Repetitive Task Practice.
S1	Primary Somatosensory Cortex.
S2	Secondary Somatosensory Cortex.
SQUID	Superconducting Quantum Interference Device.
STFT	Short-Time Fourier Transform.
SVM	Support Vector Machine.

Chapter 1

Introduction

1.1 Executive Summary

There are 50 000 strokes every year in Canada, and over 300 000 Canadians currently live with the effects of stroke [3]. Stroke rehabilitation depends heavily on Repetitive Task Practice (RTP), targeted physical therapy focused on restoring capacity in key quality of life tasks. RTP drives plastic changes in the brain which allow it to relearn and reconnect lost functionality. For patients with a fully or mostly paretic limb, RTP is not an option, and alternative therapies are required. MI, mentally rehearsing an action without executing it, has been shown to activate many of the same brain regions as RTP, though at reduced amplitude [2]. MI intervention could be used to provide therapy to patients who have completely lost motor function in a limb, or as an adjunct to traditional therapy when patients are exhausted or a trained therapist is unavailable.

MI presents its own set of challenges. It's difficult to quantify MI ability, determine how well a patient is performing MI, or even to determine whether the patient is performing MI at all. Neurofeedback, the process of analyzing brain activity in real time and using it to generate feedback for the participant, could be used to overcome all of these problems and make MI into a viable therapeutic tool for stroke rehabilitation. The therapeutic effects of MI have been shown to improve with provision of neurofeedback.

As characterizing the neural activity patterns which underly MI is still an active field of research, it would be helpful to have a data driven method for extracting the signal components which best characterize MI in order to better inform our choice of sources for provision of feedback. Machine learning, the study of building predictive models from complex multidimensional data sets, can provide tools which allow us to take a large quantity of neuroimaging data and extract those features which best discriminate different states. Machine learning presents several advantages over

more traditional statistical analysis. Rather than making a set of assumptions and testing the probability of the gathered data under those assumptions, machine learning assumes only that the labelled classes can be predicted from the gathered data, and experimentally gives a measure of confidence in this hypothesis in the form of the cross validated accuracy. If the model built can successfully predict which distribution new data was drawn from, we can be very confident the model reflects the underlying distributions in our data set well. We will employ a linear kernel support vector machine to classify trials based on whether a participant was performing MI or resting quietly. Using the learned model, we will identify the key features which discriminate MI from Rest in frequency space, and identify potential features to use in the provision of neurofeedback.

Data was previously gathered using MEG from a pool of 18 participants, two of which were rejected due to excessive electromyographic activity during imagery. We sampled this data in three ways: two sources in right and left S1 as identified by median nerve stimulation and equivalent current dipole estimation, all 306 MEG sensors, and finally 80 beamformed sources with anatomical coordinates drawn from the literature. We investigated the locations and frequencies of brain activity in all conditions.

We found that while the two sources localized by median nerve stimulation provided insufficiently reliable data to generalize across participants, the sensor and 80 source level models could accurately predict whether a participant was performing MI in as yet unseen participants, suggesting these models have captured the general pattern of MI shared across most participants. The models found weightings which share many features with the results of other Functional Magnetic Resonance Imaging (fMRI) and MEG studies. Changes in the beta frequency band (15-30 Hz) were weighted in sources including centrolateral and ventrolateral prefrontal cortex, dorsolateral and ventrolateral premotor cortex, S1, and superior parietal cortex. Changes in the mu or alpha bands (8-12 Hz) were weighted in ventral temporal cortex, thalamus, S1, and superior parietal cortex. Feedback effects observed, including dramatic changes in lateralization of MI activity, agree well with previous work. Using source space projection, we localize the differences between these models to specific anatomical regions.

The sensor model was employed to test a data driven system for identifying a reduced set of ten sensors which retained high classification accuracy, potentially facilitating development of a system with a small electrode set for providing clinically feasible neurofeedback in a small form factor and with fast setup time. There are many potential algorithms for identifying the optimal sensors from the data, and further work is required on a larger data set, with the results empirically validated, before a prototype can be developed. Once in the hands of clinicians, this device would allow widespread use of neurofeedback based MI interventions for patients at home, resting in the hospital, or during regular clinical visits.

1.2 Motivation

1.2.1 The Impact of Stroke

Stroke is an interruption of blood flow to, or the hemorrhaging of blood vessels in, the brain, killing affected neural populations and resulting in commensurate loss of function. There are 50 000 strokes every year in Canada, and over 300 000 Canadians currently live with the effects of stroke. Deficits in speech, writing, motor function, and cognitive capacity are common, and in many cases are never fully restored even after years of therapy. Many cannot return to work, and the ongoing costs of stroke in terms of lost quality of life, productivity, and dependence on the health care system for continuing care have been estimated to be as high as $3.6 billion annually by the Public Health Agency of Canada in their 2009 report [3].

1.2.2 The Current State of Post-Stroke Neurorehabilitation

The severity of stroke related impairment has prompted extensive investigation into the time course of and mechanisms underlying stroke recovery. It has long been observed that in the first 2-8 weeks post stroke a period of rapid spontaneous recovery begins, followed by a plateau which typically remains stable in the absence of intervention [4][5]. In rats, increased concentrations of GAP43, a protein produced in large quantities in neuronal growth cones and associated with long-term potentiation, was observed 3-14 days post lesion, while increased concentrations of synaptophysin, a critical synaptic vesicle protein, from 14-60 days post lesion [6]. This suggests that

the spontaneous recovery period after stroke consists of an early reweighting phase where connection strengths are relearned in the absence of the damaged region, and a longer late phase of gross rewiring as new synapses are formed in an attempt to create new pathways which route around the damage. The goal, then, is to both maximize the benefit of this period of heightened plasticity and extend its length.

Experience dependent plasticity in primary motor and premotor cortices have been observed as early as the 1950's in the work of Glees & Cole [7] in monkeys. More recent neurophysiological work by Nudo et al [8] has shown that after lesion of the Dorsal Forelimb (DFL) region of motor cortex in squirrel monkeys, in the absence of task specific activity involving the DFL, the remaining region devoted to the DFL decreased in size. When instead the monkeys were given a course of rehabilitative therapy involving RTP using the DFL, the remaining area of motor cortex dedicated to the DFL increased in size by taking over adjacent regions, and this was strongly correlated with improved task performance [9]. When repeatedly asked to perform a challenging task, dramatic increases in cortical area devoted to involved sensory and motor regions can be observed in healthy squirrel monkeys [10], with preferential recovery of areas related to these tasks post lesion [11]. Even in subcortical stroke experimental work has shown that cortical reorganization is key to functional recovery [12], likely via rerouting of connections between cortex and damaged subcortical regions.

Experience dependent plasticity, then, is a clear candidate for an effective intervention to maximize the restoration of function during the spontaneous recovery period and beyond. Studies examining neuroplastic changes induced in stroke patients by therapy have confirmed many of the results found in primate and rat models [13], unfortunately plasticity can be a double edged sword. Mounting evidence has shown that aspects of spontaneous recovery can be harmful to long term functional recovery [4]. Compensation with the unaffected limb causes increases in limb representation in the contralesional hemisphere, though it can be prevented with RTP [14]. Reciprocal inhibition between motor cortices causes further reductions in cortical representation in the lesioned hemisphere as it can no longer compete. Without therapeutic intervention, this results in further functional deficits in the impaired limb and near total reliance on the unimpaired limb for day to day tasks.

In patients who have stroke induced loss of motor function, exercise therapies have proven themselves highly effective methods of driving plasticity and restoring functionality. Use-dependent therapy, a focus on repeating task specific motions undertaken frequently during everyday life, can preferentially restore functionality which most improves long term quality of life. It serves as a critical component of modern clinical stroke rehabilitation strategies supported by the Canadian Guide on stroke best practices [15]. Constraint induced movement therapy is one example of highly effective use dependent therapy [16], constraining the unimpaired limb to force use of the impaired limb. It improves compliance and focuses on forcing the patient to rely on the damaged regions of cortex.

Converging evidence points to a mechanism by which forced reliance on the lesioned hemisphere may lead to improved patient outcomes post-stroke. Inter-hemispheric inhibition from contralesional M1 to ipsilesional M1 has been shown to be higher post-stroke, likely because of the inability of the lesioned hemisphere to generate reciprocal inhibition of sufficient magnitude [17]. In controls during unilateral motor tasks both primary motor cortices enter into an inhibitory competition, with the dominant hemisphere for the task winning out and leading to facilitation. In patients post-stroke, the lesioned hemisphere often cannot win the inhibitory competition, even when it is the designated hemisphere for the task. This leads to a compensatory reliance on the contralesional hemisphere, and inferior outcomes [18]. A longitudinal fMRI study of 12 patients post-stroke by Askim et al [19] found laterality index, the degree to which Blood Oxygen Level Dependent (BOLD) activity during a unilateral motor task of the contralesional hand was stronger in the ipsilesional hemisphere, was correlated with improvement on the Motor Action Scale and modified Rankin Scale, among many others. They also found increases in contralesional Secondary Somatosensory Cortex (S2) BOLD activity to be correlated with recovery. Dong et al [20] studied 8 patients post-stroke and similarly found laterality index of the BOLD signal towards the ipsilesional hemisphere was correlated with scores on the Wolf Motor Function Test. Lateralization of activity to the ipsilesional cortex is therefore an important goal for any therapeutic intervention post-stroke.

In addition to changes in lateralization, constraint induced movement therapy improves plastic reorganization in and around task related areas. Liepert et al [21]

measured the extent of hand representation in both hemispheres of motor cortex via TMS in 13 stroke patients with upper limb hemiparesis in the chronic stage before they were given a 12 day course of constraint induced movement therapy. Initially the contralesional hemisphere showed a significantly larger hand representation. After therapeutic intervention the extent of hand representation was measured again, and found to be not significantly different across hemispheres. This relative equalization of representation size was correlated with functional recovery as measured by the motor activity log. This is compelling evidence supporting the beneficial effects of use-dependent therapy in driving plastic changes which underly functional recovery.

Unfortunately, as these techniques require prolonged exercise and stroke patients are easily susceptible to fatigue, they can only be performed for a limited period each day. Further, use-dependent therapies are not an option for many patients who are left with a completely paretic limb post stroke. A new therapeutic paradigm is required to restore these patients to the point where they can perform, and potentially as an adjunct to improve, these more traditional therapies. Transcranial magnetic stimulation, stem cell treatments, and neuroregenerative drug therapies are all promising potential avenues for study, but it would be far simpler if a method could be devised to offer the same gains and activate the same networks as proven use-dependent exercise therapies without the requirement to actually move the impaired limb.

1.3 Motor Imagery

MI is the recruitment of motor networks for a planned action in the absence of full execution. The actual motion performed during neurorehabilitative physical therapy are not important for the exercise effects, but for the brain activity it drives. Any methodology which generates the same patterns of brain activity should be similarly effective at driving plastic changes which restore functionality. Neuroimaging and electrophysiological recordings have shown that MI occupies a middle ground between Motor Execution (ME) and planning, activating many of the same networks but without full execution, and thus without the same proprioceptive and kinaesthetic feedback. Several studies have shown that, as MI recruits the same regions of the motor network as actual execution [22][2], neurological insult which prevents action may also prevent effective imagery of that same action [23]. Page et al [24]

administered a course of task specific practice and MI based therapy to 10 stroke patients with loss of upper limb function over three weeks and found improvements on the Fugl-Meyer assessment and the Action Research Arm Test. In addition, increased activation in premotor and primary motor areas bilaterally and superior parietal cortex ipsilateral to the affected hand, and decreased activation in the rest of parietal cortex ipsilateral to the affected hand was found during wrist flexion post treatment as measured by fMRI. These findings together suggest that while stroke patients with motor deficits may find MI related to the affected limbs very challenging, improvement in task specific MI ability has the potential to support plastic changes with commensurate improvements in task specific functional measures.

1.3.1 Motor Imagery as an Adjunct to Exercise Therapy

MI has been shown to reduce the amount of subsequent RTP required to reach a fixed level of proficiency when compared to participants who trained using RTP alone [25]. This improvement in skill acquisition is encouraging for MI's potential as a therapeutic adjunct to physical therapy, allowing patients to continue to improve while unable to participate in further physical therapy either from exhaustion or lack of available hospital resources.

Several studies have had success in using MI therapies alone, showing limited improvements on a variety of functional measures, but widely differing evaluation metrics continue to make comprehensive evaluation challenging [26][27][28]. The wide variety of functional outcome measures chosen and the difficulty or impossibility of translating one score to another, combined with variability in dose and imagery type all make meta-analysis all but impossible in many cases. Further, it is difficult to quantify how skilled participants are at performing MI, or even whether they are following the instructions at all. Without neuroimaging or electrophysiological recording while imagery is being performed, evaluating task performance is completely impossible. As the patterns of activity that lead to optimal recovery are unknown and may vary from participant to participant, determining how close any given participant is to achieving this ideal imagery, or even minimally efficacious imagery, is difficult. The best we can do at the moment is to identify general directions of change in activity patterns which we believe will lead to improved functional outcomes, for example encouraging

lateralization to the ipsilesional hemisphere, thus encouraging dependency upon the damaged region and avoiding compensation with the undamaged hemisphere.

In order for patients to progress along these desirable pathways to improved MI, feedback must be provided. Feedback is a key factor in the learning process, as with no ability to assess and correct error, nor gauge improvement it is both difficult to maintain motivation and quantify progress towards targets [29][30]. Neurofeedback has the potential to dramatically improve the efficacy of MI as a standardized therapeutic tool.

1.4 Neurofeedback

Neurofeedback, the real-time visualization of a participant's measured brain activity, offers both a quantitative evaluation metric and a potentially effective training tool for improving patient's ability to perform MI. By closing the feedback loop between brain activity and imaging (Fig. 1.1) it allows patients to dynamically alter their own brain activity in response to real time neuroimaging visualizations. As a training tool, neurofeedback has been used both for Brain Computer Interface (BCI) [31], the measurement of neural activity as input for a computer system, and altering activity patterns in target brain regions [32].

Research has shown that neurofeedback can improve the ability of MI to modulate brain activity, which can lead to clinically relevant behavioural outcomes and improvements on functional measures. Boe et al [1] provided participants with neurofeedback over three sessions based on cortical oscillations in the beta (15-30 Hz) frequency band from sources localized to bilateral S1 using median nerve stimulation. Neurofeedback consisted of two bars, with a height proportional to the negative log change in beta power from baseline. During an imagined left handed 4-finger 7-button press sequence task, participants were better able to lateralize beta activity to the contralateral hemisphere and demonstrated greater beta desynchronization in primary somatosensory cortex during imagery versus healthy controls who performed MI in the absence of neurofeedback. All participants were right handed, and participants who received neurofeedback were instructed to maximize the size of the bar representing right hemisphere beta ERD and minimize the bar representing left hemisphere beta ERD.

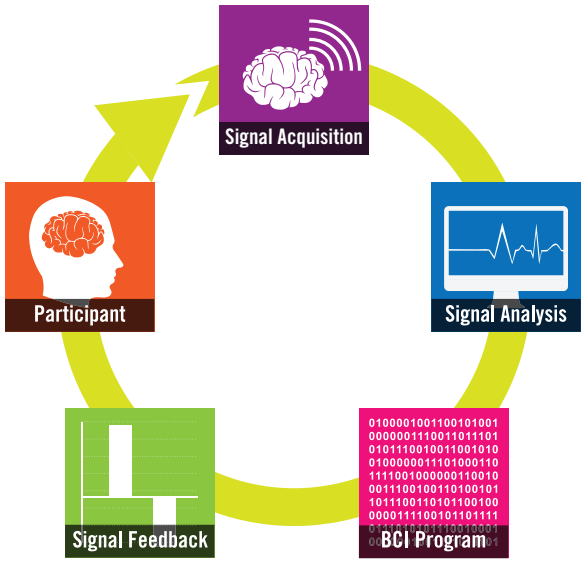

Figure 1.1: Neurofeedback is the process of recording signals generated by neural activity, processing those signals, feeding the signal through a BCI program, generating feedback in some sensory modality, and presenting it to the participant, all in real or near-real time.

Similarly, using fMRI neurofeedback, Berman et al [33] demonstrated patients were able to significantly increase the magnitude of the BOLD signal response in contralateral M1 when training with a two finger tapping task. Unexpectedly, participants were unable to achieve the same effect while performing MI of the same task despite showing BOLD activity in many regions associated with MI. Chiew et al [34] found 7 of their 13 participants who received fMRI based neurofeedback based on degree of lateralization of BOLD signal in M1 were able to significantly increase M1 lateralization over the sham group who received pre-recorded NF from one of the NF group participants. Participants alternated blocks performing left and right handed kinaesthetic MI, with the degree of lateralization required for success in each trial increasing or decreasing as a function of number of successive successful or failed trials, respectively. This difficulty in controlling M1 BOLD activity may simply be a variability in dosage requirements across the population, a limitation of fMRI based neurofeedback, or it may indicate potential difficulty in rehabilitating M1 lesions using MI.

A direct examination of the impact of an EEG based MI neurofeedback system on measures of functional recovery found correlations between altered lateralization

of brain activity in patients given neurofeedback and improvements in subjective and objective functional measures, providing preliminary evidence for the efficacy of neurofeedback assisted MI in a clinical setting [35]. Pichiorri et al [36] compared the effects of MI intervention on functional outcome measures in stroke patients with and without neurofeedback, finding increases on the Fugl-Meyer Assessment, increases in contralateral beta desynchronization, and increases in inter-hemispheric connectivity in the sensorimotor network with provision of feedback. Their feedback system depended on manual feature extraction performed by neurologists examining patient EEG data.

1.4.1 Neurofeedback Presentation

In order to provide neurofeedback, one must first identify the spatiotemporal component of the signal which will form its basis. Correctly isolating activity from task relevant anatomical areas of the brain, and identifying signal components and relevant frequency bands is likely key to providing accurate and efficacious neurofeedback. The sheer volume of information provided by modern neuroimaging techniques can make this process challenging, especially for tasks within as yet unknown networks. Brain activity in several key frequency bands related to ME and MI have been explored in the literature, including mu (8-14 Hz), beta (15-30 Hz), and gamma (30-100 Hz). Beta and gamma have been further subdivided into low beta (15-20 Hz), high beta (20-30 Hz), low gamma (30-60 Hz) and high gamma (70-90 Hz), each based on differential time course and the circumstances under which they are modulated [37].

Modulation of beta oscillations in primary sensorimotor areas has been observed in both active and passive motion, imagery, in anticipation of motor related cues, when observing movement, and during sensory stimulation, to list only the most relevant examples outlined in Kilavik et al's [38] review of sensorimotor beta oscillations. Motor activity consistently generates strong beta Event Related Desynchronization (ERD) followed by a transient Post-Movement Beta Rebound (PMBR) after motor activity ends [39]. The beta ERD persists for as long as movement [40] or gradations in force [41] continue, with no apparent sensitivity to type of movement in degree of beta ERD [40], but movement type and limb dependent changes in peak frequency [42]. Beta ERD can also be generated by other motor-related tasks, most notably

including MI [2][43]. Mu oscillations also show ERD in sensorimotor cortex during motor activity and imagery, but without an analogous post-movement rebound [39]. There are many competing hypotheses for what functional role, if any, beta ERD and the PMBR play in motor function. It has been suggested the beta rhythm represents an idling state for sensorimotor cortex [44], disrupted when activity is required, with the PMBR enforcing the idle state to halt further motion. This is supported by the correlation between beta ERD and the BOLD signal [45]. If beta ERD is a measure of neural population activity, it should correlate with the BOLD measure of cortical energy expenditure. The sensorimotor beta rhythm may also represent a sensory integration process, similar to proposed mechanisms of gamma activity in primary visual cortex [46], taking in somatosensory feedback to evaluate the completed movement and plan the next [47]. Cassim et al [48] found ischaemic nerve block of the arm - restricting blood flow until sensory afferents became non responsive as assessed by the absence of the median sensory nerve action potential - prevented PMBR after a finger movement task, suggesting it is related not to movement termination but termination of movement related sensory feedback. Beta oscillations have been shown to increase corticomuscular coherence, corticocortico coherence, magnitude of sensory evoked potentials, and increases in information flow throughout the motor network as measured by Granger causality, leading some to hypothesize that beta rhythms are part of a system for information transfer between cortical regions [38].

Transient high gamma (70-90 Hz) bursts have also been recorded early in ME [49]. Unlike beta ERD, this transient high gamma ERS has an amplitude proportional to the intensity of movement, is only generated during active movement, and is highly somatotopically localized [50]. This suggests high gamma ERD is very tightly linked to motor commands, and not to sensory feedback.

In order to improve the viability of MI neurofeedback for treating loss of motor function post stroke, we need to identify the optimal set of spatiotemporal features to display to the participant. These are likely the same features which are shared with ME activity in healthy participants, as the desired goal is to guide the patient towards this more functional pattern of activity. Kraeutner et al [2] showed that in the beta band ME and MI share common activation patterns in premotor and supplementary motor areas, differing in degree of beta ERD in M1 and S1. If the

primary differences between MI and ME are simply a matter of degree, then providing feedback which improves MI performance as defined by degree of desynchronization in select frequency bands in specified brain regions should yield many of the same therapeutic effects of RTP.

Neurological insult may change these patterns dramatically, creating both advantages and disadvantages to a more personalized approach focused on learning the features which characterize MI for a single patient. On the one hand, they may not be capable of generating the kind of activity typical of a healthy population; perhaps stroke has caused another neural population to adopt the function of a lesioned region, and the general model doesn't account for this. By identifying the features which characterize this particular participant's MI, we may be able to account for this and provide feedback which will guide this new arrangement to maximally restore functional capability. On the other hand, to learn to provide feedback from poor quality or ineffective imagery could result in leading the patient into perpetuating ineffective patterns like contralesional control. The ability to identify the features which best characterize high quality MI in the general population, and potentially to automatically tailor this feedback system to an individual participant without the need for clinician intervention would therefore be ideal for making neurofeedback a practical and clinically relevant tool.

1.5 Machine Learning

1.5.1 The Principles of Learning

Machine learning is a field of computer science devoted to the study of how to extract predictive models from complex data sets, broadly divided into supervised and unsupervised learning techniques. Data is composed of sets of vectors of features, also called exemplars, each of which may be anything from a voltage sample recorded from a sensor to number of packs a day smoked by a patient. Supervised learning attempts to predict the class of new data after training on data labelled with known correct class information, while unsupervised learning attempts to automatically identify patterns in unlabelled data. As we are interested in the differences between easily labelled data (imagery vs rest, feedback vs control), supervised learning is ideal for

this task. Many techniques in supervised learning rely on the selection of a decision boundary, a hyperplane in the space formed by the feature set such that new data points can be accurately classified based on which side of the boundary they fall upon. The standard selection process employs some modified variant of gradient descent, iteratively shifting the decision boundary in the direction that optimally improves a fitness function like mean squared error or classification error until some local minima is found. Many techniques are employed to avoid settling on poor local minima, but identifying the global minima is impossible in the absence of perfect knowledge of the problem space, and as such the best solution found within some bounded search must be accepted.

Machine learning offers several key advantages over simpler statistical analysis. Learning algorithms cannot simply be satisfied with an effect size and confidence intervals indicating that two groups are different, they must find patterns in the data which reliably allow individual exemplars to be assigned to groups. Further, the ability of the generated model to make predictions allows it to be easily and concretely validated in applications like BCI. The strict demands of prediction and focus on reliably differentiable features generates models which better characterize the essential differences between the phenomena of interest [51].

1.5.2 Evaluation

In supervised learning, the simplest way to evaluate the effectiveness of the model is accuracy, proportion of correctly classified trials, on a withheld data set. The more data withheld, the more robust the evaluation but the less data available for training. This problem can be obviated by the use of cross validation, building multiple models withholding a different segment of the data for testing each time until the model has been evaluated on all of the data without ever having seen it during training [52]. There are a variety of methods of splitting data for cross validation, here we use Leave One Batch Out (LOBO) cross validation, with each batch corresponding to a participant. All trials except those of one participant are used for training, then the model is tested on the trials of the withheld participant and the process is repeated, withholding a different participant each time. The difficulties of generalizing models across participants makes this a more accurate test of whether the model has genuinely

extracted patterns which broadly represent the state of interest - in this case, MI.

1.5.3 The Importance of Representation

Much as arabic numerals allow much faster recognition of patterns in numbers than their roman counterparts, the way in which data is represented when presented to a learning algorithm is key to its success. A linear decision boundary cannot learn to discriminate between one class surrounded by another, but a transformation into polar coordinates can make that non-linear problem easily solved by a linear classifier. When applying machine learning to neuroimaging, the feature set is typically composed of sensors and sources used, and whether signals are presented in the time domain, frequency domain, or both. Preprocessing and feature conditioning are also key steps in establishing a tractable problem for the learning algorithm. The features of interest for classifying a particular pattern of neural activity can vary widely, even across individuals. The challenge of identifying which of these features is best suited to identifying MI will be a central focus of this research.

Considerable work has been done on paring down the large volume of information obtained by whole head analysis into only the most salient features. Santana et al [53] demonstrated superior accuracy when classifying direction of attention from MEG data by using an adaptive sensor and frequency selection system. An adaptive source level selection system may be even more powerful, and capable of dynamically determining which brain regions most reliably alter their patterns of activity for an individual or participant group.

Herman et al [54] found that amongst power spectral density, Continuous Wavelet Transform (CWT), Short-Time Fourier Transform (STFT), and several other representations that power spectral density was the most informative for discriminating right/left MI, and that amongst Linear Discriminant Analysis (LDA), Linear Support Vector Machine (LSVM), and two Gaussian kernel Support Vector Machines (SVMs), that the linear classifiers had the best accuracy and generalizability. Key features in the frequency or size domains during MI included the beta band (typically 15-30 Hz), and the mu band (8-12 Hz).

1.5.4 Recursive Feature Elimination/Addition

For learning algorithms which assign importance to each feature, the algorithm itself can bootstrap feature extraction. By taking the full set of features and iteratively removing the least discriminative features, or taking an empty set of features and iteratively adding the most discriminative, a reduced feature subset can be built. This subset allows building smaller, faster training models which are often more accurate as noisy, misleading, and redundant features are removed. Recursive feature addition has the advantage of being faster running, especially with a small number of target features, as it takes much less time to train models with few features. Recursive Feature Elimination (RFE), while taking longer to compute, can better capture informative combinations of features; if two features are valuable together but individually uninformative, recursive feature addition will likely never select them.

1.6 Linear Discriminant Analysis

LDA seeks to explain a discrete dependent variable, class, by a linear combination of the continuous independent variables, the features. In this respect, it can be considered the complement of the Analysis of Variance (ANOVA), which seeks to explain a continuous dependent variable by a linear combination of discrete independent variables. LDA also shares similarities with Principal Components Analysis (PCA), in that both seek to find a projection onto the axes of greatest variance, but PCA is interested in projecting the data along the axes of greatest absolute variance, while LDA is only concerned with projecting onto the axes along which there is greatest inter-class variance. Like the ANOVA, LDA assumes the features are normally distributed, and that each feature shares the same covariance.

The simplicity and low computational cost of LDA make it useful in applications where learning and classification must be performed very quickly, or where it is desirable that the model be easily interpretable. It has been employed successfully in many BCI and neurofeedback systems [55]. Tangwiriyasakul et al [56] employed LDA and common spatial patterns filtering to classify MI in stroke patients measured with EEG. They found it was possible to train a model based on healthy contralesional activity and predict ipsilesional MI with an accuracy of 0.7 using 3 or 11 EEG sensors

positioned near M1, or up to 0.9 using 45 sensors. Yue et al [57] employed LDA to classify MI in a 15 sensor EEG BCI to control a simulated pendulum balanced on a cart. Five of their six participants were able to successfully use MI of the left or right hand to move the cart and keep the pendulum stable for up to 35 seconds.

1.6.1 Computing Fisher's Linear Discriminant

The degree to which each linear combination of features is a good axis to project the data onto to achieve optimal class separation is evaluated by the following equations:

$$J(v) = \frac{(\mu_1 - \mu_2)^2}{s_1^2 + s_2^2} \tag{1.1}$$

$$s_c = \sum_{i=1}^{n} (c_i - \mu_c)^2 \tag{1.2}$$

where v is the axis, $J(v)$ is Fischer's Discriminant, μ_c is the mean of class c on axis v, s_c is the scatter along v of class c, and c_i is the ith sample of class c along v. Maximizing Fisher's Discriminant thus maximizes the difference between class means, while minimizing within class variance. This helps determine the most useful linear combination of features, and can be used as a simple linear classifier by choosing a linear decision boundary along this most discriminating axis.

1.7 Support Vector Machines

SVM were developed by Cortes & Vapnik [58] as a practical implementation of statistical learning theory capable of tackling difficult supervised learning problems with high dimensionality while maximizing generalizability. Statistical learning theory examines the problem of extracting from an arbitrary labelled data set the function which generated the original classifications. SVMs have achieved strong results in a wide variety of machine learning problems, including many applications in the BCI and neuroimaging domains [55].

1.7.1 Finding the Decision Boundary

The training data is composed of input vectors, which are transformed into some higher dimensional feature space where the data is linearly separable. By choosing

different kernel functions arbitrary decision surfaces in input space can be created, giving the SVM the capacity to find a computationally much more tractable linear decision boundary in feature space to solve a wide variety of nonlinear problems. This mapping can be extremely expensive to compute due to the high dimensionality of the space, but we only require the dot product between our learned weights and our feature vector in this space to produce a prediction. By using the kernel trick, this dot product can be computed for certain kernel functions without solving the full mapping. Once in feature space the parameters of a hyperplane must be chosen such that they maximize the margin between this decision boundary and the training vectors, as seen in Fig. 1.2. This convex optimization problem is typically formulated in either dual or primal form, with dual form being best suited to the case where the number of features exceeds the number of data points. Conveniently, this computationally expensive optimization problem can be vastly simplified by choosing only to optimize the decision boundary with respect to a subset of training vectors which best represent the boundary between the two classes - the eponymous support vectors. This also serves to increase the probability of accurately classifying new exemplars, improving generalizability by reducing the risk of overfitting.

1.7.2 Structural Risk Minimization

The desire to accurately classify all training exemplars can harm generalized accuracy on new data via overfitting - learning a needlessly complex model which attempts to account for noise or outliers. Vapnik and Chervonenkis outlined the VC Dimension to quantify this tradeoff between accuracy on the training set and model complexity [59], and the penalty or soft margin parameter C of the SVM determines the degree to which one or the other is favoured. By properly choosing C during parameter selection, the model can settle on a suitable level of complexity to capture the patterns in the data without sacrificing the ability to smooth out noise to improve generalization.

1.7.3 Regularization

Regularization is a tool for minimizing overfitting by penalizing overconfidence in models. L1 regularization roughly corresponds to a laplacian prior and tolerates individually very high and low weights, while L2 regularization corresponds to a

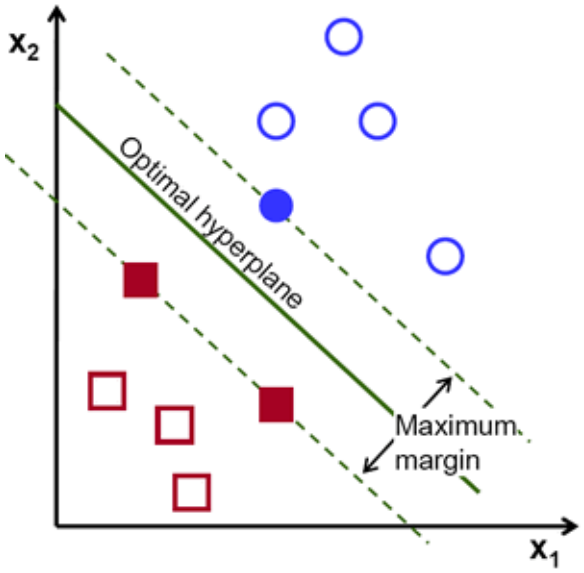

Figure 1.2: Optimizing the margin between the decision boundary and the two classes minimizes the probability of classification error on new data.

gaussian prior and thus is less tolerant of outlying weights. L1 regularization leads to very sparse models, which makes it useful for encoding sparse signals or feature selection, but is generally outperformed by L2 regularization in classification tasks.

1.7.4 Related Work

Yi et al [60] used EEG to classify single and multi-limb MI involving one hand, one foot, both hands, or either hand paired with the contralateral foot. Several variants of common spatial patterns filtering were compared for feature extraction and SVM used for classification, yielding a mean accuracy of 0.7 on this seven class problem.

Tam et al [61] employed RFE using SVM accuracy as the evaluation metric to identify a minimal set of EEG electrodes for driving a MI vs rest based BCI to administer functional electrical stimulation to five stroke patients. They found that between 10 and 20 of their 64 electrodes yielded over 0.9 accuracy, and that repeated calibration sessions contributed little over the first. This suggests inter-session variability is not a large factor in personalizing neurofeedback, dramatically reducing the amount of training and preparation time required to deploy such a system. Despite this accuracy, there were only small functional improvements on the Action Research Arm Test and Fugl-Meyer test.

Adaptive learning systems may be an effective synthesis of generalized and personalized models. Spüler et al [62] explored an adaptive SVM based classification system providing online feedback for MI and mental subtraction. Their system estimated the posterior probability of a given trial being correctly classified, and if it was over a threshold p_t, they added it as a training exemplar of that class and updated the model. They found that when building personalized models on one session and testing it on a second, it outperformed several online and offline learning techniques, yielding mean accuracies of 0.897 offline and 0.846 online. It may be possible to adapt this methodology by providing a generalized model to achieve immediate basic functionality without training and have it adaptively tailor itself to the participant.

In general other studies employing ML in the neurofeedback domain have built subject specific models. Weights and, when used, sensor elimination are fit to an individual and evaluated on that same individual, as in Yi et al [60] and Tam et al [61]. Lal et al [63] explored SVM based RFE for sensor selection in a MI BCI, but only used five participants, having rejected three, and once again performed within-participant testing for their sensor selections and used 10-fold cross validation within each participant to evaluate accuracy. Ideally we would like a subset of sensors which can be used for any participant, and to empirically test its ability to generalize across participants.

Unsupervised learning is typically what is used in zero training BCI systems, learning to classify online as data is gathered, and thus avoiding training scans. Kindermans et al [64] explored an unsupervised training approach which used a randomly initialized classifier which re-examined its previous classifications after each block to improve. Popescu et al [65] explored a similar unsupervised learning based MI BCI with a set of 6 dry electrodes, closely approximating the sort of system a clinic could employ. These systems effectively reduce training times, but are potentially unsuitable to neurofeedback as they very closely fit to individual participants, potentially leading to provision of feedback which perpetuates ineffective patterns of activity. A system capable of fitting to participants who are more skilled at performing MI and generalizing across participants to provide feedback would be preferable from a rehabilitation point of view.

1.8 Magnetoencephalography

MEG is a functional neuroimaging technology which can non-invasively provide superior spatial resolution over EEG due to the reduced signal smearing from the much simpler propagation of magnetic fields through the head as opposed to electrical potentials, and superior temporal resolution to fMRI due to the slow dynamics of the BOLD response. This makes MEG an excellent instrument for measuring spatiotemporal components of cortical activity, especially complex or high frequency oscillatory patterns. Though we eventually plan to transition to EEG for clinical deployment, MEG is a stronger candidate for system development and prototyping.

Our data set consists of MI, ME, and resting blocks. MI and ME are known to induce changes in oscillatory neural activity which are easily detected using MEG. Using source localization techniques we can localize the neuroanatomical generators of these patterns of activity with relatively high specificity, allowing us to discuss our findings in the context of the broad literature of neuroanatomical structure and function.

1.8.1 How MEG Works

MEG measures magnetic fields primarily generated by synchronous firing of cortical neural populations in sulci, and thus oriented parallel to the scalp. These magnetic fields are on the order of fT, and require incredibly sensitive magnetometers to detect. In 1962, Josephson theorized the quantum mechanical properties of a thin insulating layer in a superconducting loop, or Josephson junction, and that they would be suitable for converting voltage to frequency [66]. Anderson & Rowell first observed this effect in 1963 [67], and work began on employing this effect in creating an extremely sensitive magnetometer. This work resulted in Superconducting Quantum Interference Devices (SQUIDs), superconducting wire loops with Josephson junctions, which convert magnetic flux, changes in magnetic field strength, into a voltage oscillation [68]. The high accuracy with which frequency can be measured gives SQUID magnetometers sufficient sensitivity to detect these small cortical magnetic fields.

Two sensor types are created from SQUIDs in our system, magnetometers and planar gradiometers. Magnetometers are traditional SQUIDs, measuring magnetic

flux. Planar gradiometers consist of two adjacent wire loops running in opposite directions. Changes in magnetic flux across the two loops are amplified, while magnetic fields which induce current in both loops equally are cancelled. This attenuates environmental noise while increasing sensitivity to magnetic flux generated by current flowing perpendicular to the axis of the loops. By clustering two planar gradiometers aligned along the lateral and anterior-posterior axes with a magnetometer, a sensitive and specific picture of the local magnetic flux is captured.

1.8.2 Source Estimation in MEG

When localizing sources from sensor arrays there are two main problems of interest: the forward problem, and the inverse problem. The forward problem requires solving for the sensor values which would be measured given a set of sources and their activity. This is a well-defined problem which is easily solved. Its complement, the inverse problem, requires solving for a set of sources and their activity given a set of sensors and their recorded values. This problem has an infinite number of solutions, and thus heuristics and constraints must be used to yield plausible and useful solutions. Sources are typically limited to within the area of the head, and further limited to grey matter regions through boundary element modelling and co-registration with either structural fMRI or a template brain.

In Equivalent Current Dipole (ECD) estimation, sources are assumed to originate from a small number of equivalent current dipoles, point sources of magnetic flux. Given a distance between the sensor and source greater than the distance between a cluster of small sources, this is a good approximation, and works well in practice. By minimizing the error between the proposed set of dipoles and the activity measured, we obtain a set of dipoles which best fit the data.

ECD estimation works best for small numbers of sources. In order to improve its ability to localize activity to many specific anatomical regions and estimate the time-course of the activity at each, a technique developed for RADAR arrays called beamforming is employed. Beamforming can extract time-locked responses even when they are not phase-locked, a significant advantage over evoked field averaging. By using the covariance matrix of the sensor array a set of weights can be assigned to each sensor such that signals generated from areas other than the target source

are strongly attenuated. These weights, once computed, can be applied very quickly, allowing many such weightings to be applied to extract the signal from a large number of sources in real time [69].

1.9 Project Objectives

In this project we aim to improve the quality of neurofeedback by training a machine learning model to identify the components which best discriminate rest from MI, and thus will provide the most accurate feedback. In order to do this, we will analyze previously collected data to answer three main questions:

1.9.1 Objective 1

Question

Can participant state (MI vs rest) be accurately predicted from beta power in primary somatosensory cortex after training?

Rationale

In our group's previous work [1], mean beta power from two sources in right and left somatosensory cortex were used to provide neurofeedback. The degree to which this signal can be used to discriminate MI from rest is not known, and will serve as an important baseline for comparison to further work.

Hypothesis

We predict that beta power in primary somatosensory cortex will be a good general discriminator of MI and rest, and thus we should see good generalized classification accuracy across all participants after training.

1.9.2 Objective 2

Question

Can we identify which MEG sensors are most discriminative between rest and MI?

Rationale

In developing a clinically useful MI feedback tool, a system for identifying the

most discriminative minimal set of sensors for provision of neurofeedback would help reduce the size, complexity, setup time, and cost of the final product. Here we will investigate how effectively we can discriminate MI from rest with sensor level models, and with models based on a small subset of sensors, to evaluate how accuracy changes when a smaller sensor set is used.

Hypothesis

We predict sensors over primary motor cortex, supplementary motor cortex, and posterior parietal cortex will be consistently ranked amongst the highest across all participants.

1.9.3 Objective 3

Question

Using source level modelling of MEG data across the whole brain, can we identify the anatomical regions whose activity best discriminates MI across our participants?

Rationale

In order to better understand the neurophysiological principles underlying MI it is important to localize the relevant signal components to their anatomical sources. Without an understanding of which sources are relevant in which frequency bands, the extensive literature on functional connectivity and task related networks cannot be used to make inferences about the potential effects of feedback based on any features identified.

Hypothesis

We predict sources in primary motor cortex, supplementary motor cortex, and posterior parietal cortex will be consistently ranked amongst the highest across all participants.

1.9.4 Between Group Comparisons

As an extension of both Objective 2 and Objective 3, given that we have group labelled data and previous results of Boe et al [1] to draw comparison with, we will

examine the differences between models built to discriminate imagery on a participant set who received feedback vs those who did not. We expect to see the lateralization effect of feedback observed by Boe et al.

1.9.5 Collective Significance

Once we have the set of most discriminative signal components, the minimal set of sensors required to detect them, and the neurophysiological sources generating them, we have the required information to design and test a prototype system with stroke patients.

Chapter 2

Methods

2.1 Task

The full methodology can be found in Boe et al, 2014 [1], but in brief: 18 right handed non-disabled participants (8 male, mean age 24.7 $\pm$ 3.8 years), each having provided written informed consent, participated in the study. The methodology was approved by the IWK Health Centre Research Ethics Board. Each participant was brought in for three sessions to perform a fine motor task while undergoing an MEG scan in an Elekta Neuromag MEG system, with one group receiving neurofeedback (9 participants) and the other acting as control (9 participants). Two feedback group participants were dropped due to excessive motor activation during imagery blocks as measured by electromyography, leaving a total of 7. The task consisted of a sequence of 7 button presses on a 4 button keypad, to be repeated as many times as possible over a ten second period using the left hand. Data from two participants were rejected due to unacceptably large numbers of artefacts or ME during imagery. Ten second trials of activity would alternate with ten second trials of quiet resting. In each block there were 15 active and 15 rest trials. In half of the blocks the participant was asked to imagine performing the task, rather than actually pressing the buttons (Fig. 2.1). At all times electromyography was recorded from the flexor and extensor muscles of the digits. Imagery trials containing electromyographic activity were rejected by an automated thresholding system which looked for muscle activity that was significantly different from baseline as determined by a paired t-test ($p < 0.05$).

2.2 Data Acquisition

Data was sampled at 1500 Hz, with an immediate bandpass filter of 0.1 - 500 Hz applied. Data was saved for offline analysis, and simultaneously each one second segment was sent for real-time processing.

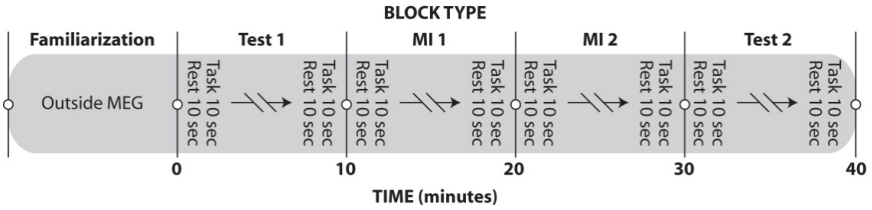

Figure 2.1: The timeline of the task performed by each participant in each session in the study by Boe et al, 2014 [1].

2.2.1 Objective 1: Generating Virtual Depth Electrodes

Two ECD estimates centred on Primary Somatosensory Cortex (S1) in the right and left hemispheres were computed on data gathered from each participant during each session. Median nerve stimulation was used to generate the two initial seeds in S1 for the fitting algorithm. A custom extension of the Fieldtrip package was used to compute the fits. A single sphere model was used for source estimation, with a sphere based on a 200-point head digitization obtained prior to scanning.This dipole fitting yielded two sets of sensor weights which, after being updated with head position information, could be multiplied with the MEG data as it was gathered to generate an estimate of current flow projected into cortex very quickly, facilitating real-time provision of feedback.

2.2.2 Objective 2: Sensor Data

Data from all 204 planar gradiometers and 102 magnetometers was used directly in this analysis. This data set was extracted post-experiment, and was not originally used for provision of feedback.

2.2.3 Objective 3: 80 Virtual Depth Electrodes

80 equivalent current dipole estimates were seeded with relevant anatomical locations drawn from the literature [70][71], relabelling the Frontal Eye Fields as Anterior Premotor Cortex. Elekta Neuromag's beamforming software was used to compute the fits. A boundary element model based on a template brain was used to constrain the fits. This data set was extracted post-experiment, and was not originally used for provision of feedback.

2.3 Preprocessing

For real time feedback and the two-source analysis, data was low-pass filtered at 333 Hz. For the sensor and source level analysis it was down sampled to 250 Hz, low pass filtered at 70 Hz, and had Independent Components Analysis [72] and Temporal Signal Space Separation [73] applied to remove artefacts and attenuate external noise. In all cases the data was epoched around the imagery, motor execution, or rest start event codes, with the segment from 10 seconds before to 10 seconds after each event being used for further processing.

2.4 Feature Extraction

Features for the machine learning algorithm were extracted from sensor or source data as the amplitude of oscillatory activity. For all data sets, the absolute value of the STFT of each trial was taken using 256 sample windows for Objectives 2 & 3, or 1024 sample windows for Objective 1 (1.024 seconds for all) with 128 or 512 sample steps from 1 to 8 seconds (trial) and -9 to -2 seconds (baseline). Both baseline and trial segments had their windows averaged before the averaged trial was divided by the averaged baseline and the log base two taken. The resulting array of log power change from baseline by frequency was used as the feature vector for training. In the case of the two source data, this feature set was reduced by further averaging over four frequency ranges: mu (8-12 Hz), beta (15-30 Hz), low gamma (30-45 Hz), and high gamma (45-70 Hz). These features were used to both match the feature chosen for the real-time feedback and allow for the potential value of other frequency bands. In the sensor and 80-source data sets 72 frequency bins from 0 to 70 Hz were extracted from each sensor or source as features. Participants were instructed to close their eyes during imagery and motor execution blocks, but open them during rest. As engagement of the visual system is known to impact oscillatory activity in primary visual areas, for each sensor and 80-source data set a second set was generated where all sensors more than 30cm posterior of the midline, or sources more then 30 cm posterior of the midline and less than 10 cm above the midline were eliminated as well as all sources more than 70 cm posterior of the midline, to examine the effects of removing features from primary visual areas. These data sets are referred to as

Occipital Excluded, while the full sensor or source array is Occipital Included. Each feature was divided by its standard deviation and had its mean subtracted prior to learning to allow meaningful comparison between learned weights. The Python libraries NumPy, SciPy [74], MNE Python [75][76], Matplotlib [77], and Pandas [78] were used for data processing and feature extraction.

2.5 Training

The LSVM model used L2 regularization, a squared hinge loss function, and solved the dual problem. The hinge loss function is $max(0, 1 - t \cdot y)$, where y is the scalar prediction of the model and t is the true class ($t \in \{-1, 1\}$). C and tolerance were chosen by a stochastic sampling from predefined parameter distributions, generating and testing models for a fixed number of iterations before returning the parameters that yielded the highest cross-validated performance. In practice this led to varying C and tolerance from run to run, but accuracy was largely insensitive to these differences, and it reduced training time over a full grid search.

The feature vectors were arranged into blocks by participant, and a LOBO cross validation system was used to evaluate accuracy (proportion of trials correctly classified), whereby the model trained on all but one participant, and then was evaluated based on its accuracy on the remaining participant. All accuracy and feedback information presented was therefore on a participant the model had never seen before. The stochastic parameter search also used LOBO cross validation, leaving one of the training participants out each time. This effectively resulted in a nested cross-validation, whereby within the training set one participant was held out each time and the parameters which yielded the highest average accuracy over all these held out training participants were selected. These parameters were then used to fit a model on a full training set and finally evaluated on the held out test participant. This was done to ensure the parameters were optimized for ability to generalize across participants. The Python library Scikit Learn [79] was used for training.

When plotting group accuracies, kernel density estimation with bandwidth selected by Silverman's rule was used to approximate the probability density function for accuracy. These plots can be read as smoothed histograms, where the bandwidth is the smoothing factor.

2.5.1 Objective 1: Two Source Models

The dramatically reduced feature set in Objective 1 made the LSVM unnecessarily complex. A simple LDA classifier was used instead, requiring no parameterization. It was still trained and validated as above using LOBO.

2.5.2 Objective 2: Recursive Sensor Elimination

In order to identify the reduced set of sensors which are most informative for discriminating MI, Recursive Sensor Elimination (RSE) was performed, employing the weightings learned using the general training methodology outlined above. This ideally removes spurious and redundant features, keeping only those which are essential to predicting class membership. A sensor's weight was computed as the mean of the absolute value of its ten frequency bins with weights of greatest magnitude. Each pass all of the features from the sensor with the lowest weight were dropped from the feature set, the model was retrained on this reduced feature set, and the process was repeated until only features from the ten most informative sensors remained. A final model was then trained as per the general training methodology outlined above on this reduced feature subset to obtain a final cross validated accuracy. This analysis was performed with occipital sensors excluded.

2.5.3 Objective 3: Source Rankings

For each of two frequency bands, beta and mu, the weight of each feature was normalized to the greatest absolute weight within that band for that cross validation fold, the sensor weighted according to the mean weight of the frequency features in that band, and the mean sensor weight taken over all folds. Only the relative weightings are of interest, and normalization ensured that any change in the total weighting in a given fold would not skew the mean.

Chapter 3

Results

3.1 Objective 1: Can participant state (MI vs rest) be accurately predicted from beta power in primary somatosensory cortex after training?

Participant state could be predicted to a limited extent using only the two sources fitted to the field evoked by median nerve stimulation, but had difficulty generalizing across participants. Simple two source models were built using LDA on average event related change in synchronization in four frequency bands for a total of four features per source: mu (8-12 Hz), beta (15-30 Hz), low gamma (30-45 Hz), and the remaining mid gamma band (45-70 Hz). These models showed extremely variable accuracies (Table 3.1), with a mean accuracy of 0.724±0.131 (Minimum 0.481, maximum 0.870) for the Feedback group and 0.741±0.122 (Minimum 0.520, maximum 1.000) for the Control group, which can largely be attributed to several participants either failing to demonstrate the expected beta ERD or showing beta ERS during imagery trials. As seen in Fig. 3.1, although the localized sources for each participant were on average between S1 and M1 in Talairach space, they varied several cm in both axes. This positional variability likely contributed to the poor generalization ability.

As an example, the data for feedback participant 3 is shown in Fig. 3.2, where it can be seen that the grand mean of the training set follows the expected pattern of beta desynchronization during imagery, but the mean of the testing set deviates to the point that the simple model learned on the training data is no longer valid. Increasing the number of sources involved, or looking at the sensor level, can potentially allow the identification of more reliable regions, or multiple regions which taken together are resistant to unusual patterns in any one area.

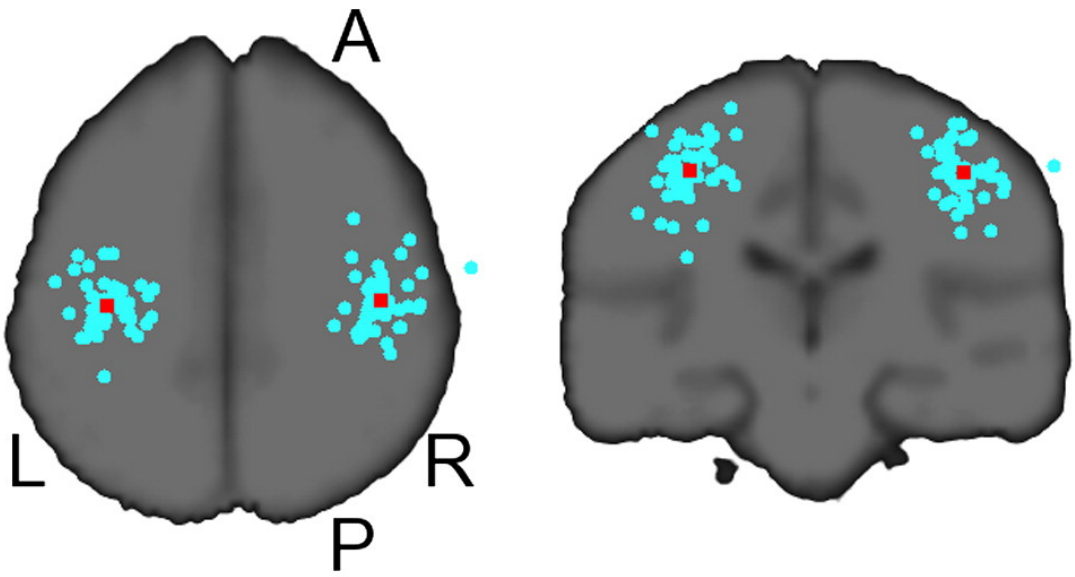

Figure 3.1: The seed locations for neurofeedback for all subjects are shown as cyan circles, with the mean location shown as a red square. [1]

Table 3.1: Classification Accuracies for Two Source Analysis

Task	Imagery	
Group	Control	Feedback
Subj #		
0	–	0.632
1	–	0.737
2	–	0.870
3	–	0.667
4	–	0.820
5	–	0.481
6	–	0.860
7	0.895	–
8	0.769	–
9	0.729	–
10	1.000	–
11	0.520	–
12	0.853	–
13	0.696	–
14	0.542	–
15	0.667	–
Mean	0.741	0.724
CI	0.122	0.131

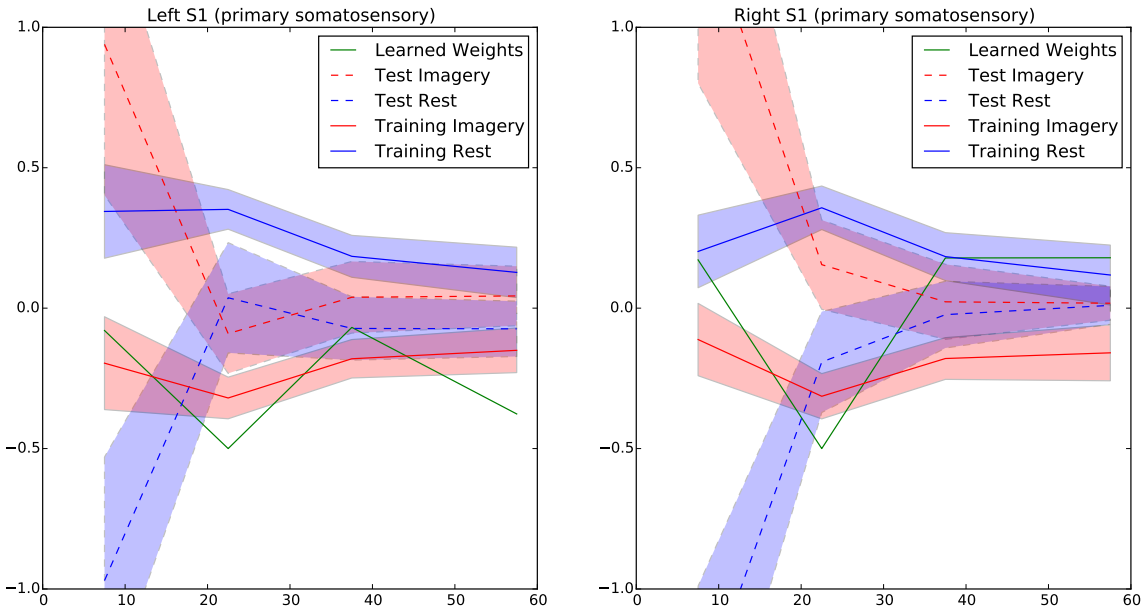

Figure 3.2: Mean feature value vs frequency (Hz) plots for a two source LDA model. The solid red and blue lines represent the class means for the training set, the dotted lines the testing set, and the solid green line the learned weights. Participant 3 was held back to test this model, and the reason for the low accuracy becomes obvious when the right hemisphere beta synchronization in their testing data is contrasted with the desynchronization in the training mean. Several other participants displayed similarly unexpected patterns in alpha and beta, producing similarly low accuracies.

3.2 Objective 2: Can we identify which MEG sensors are most discriminative between rest and MI?

Overall the sensor level models were able to accurately differentiate imagery and rest in most participants with a mean accuracy of 0.929±0.064 (Min. 0.823, Max. 0.980) for the Feedback group and 0.956±0.037 (Min. 0.854, Max. 1.000) for the Control group, as laid out in Table 3.2 and Fig. 3.3. Removing the occipital sensors changed accuracy by an average of 0.0436±0.08505 for Feedback, −0.0257±0.0376 for Control, and -0.0058 ± 0.0060 for Motor. The increases in accuracy were a result of unusual cases where a participant showed misleading alpha activity during imagery. Transfer loss, the difference between cross validated accuracy on the training data and the accuracy on the held out testing participant, is presented in Table 3.3. Transfer loss represents some combination of two main factors: degree of model overfitting, and the degree to which one participant differed from the overall pattern. These are not directly separable, as they are two sides of the same coin. Without outside information it's impossible to know if a participant was difficult to predict because the model overfit and failed to generalize sufficiently well, or because they were mislabelled, i.e., not performing the task correctly. In general, consistent transfer losses indicate overfitting, while outlying transfer losses likely indicate something anomalous about the data. Participant 3's transfer loss of -0.44 is likely an irregularity in the data, as it vastly exceeds the mean transfer loss of -0.10, and will be explored further in the discussion.

Because each fold omits one participant for testing, the choice of which fold to display as representative is not obvious. I have elected to plot the models from the folds where the participant who displayed the greatest transfer loss were held out as the testing set. This biases the plots provided, but as some bias is inevitable removing an outlier seems the least undesirable choice. Further, this makes it possible to contrast this model with the actual data of these outlying participants to better understand why they differ and how the model failed to capture their patterns of activity. Models where participant 3 was held out for testing were chosen for Sensor level Feedback Imagery and Motor plots, and where participant 15 was held out for Control Imagery.

Sensors over primary sensorimotor and premotor areas of cortex are heavily weighted

for beta desynchronization, with the control group favouring the left hemisphere (Fig. 3.4) with maximal left and right hemisphere weights of 0.75 and 0.5 in gradiometers, and 0.8 and 0.2 in magnetometers. Both the feedback and motor groups favour the right hemisphere (Figs. 3.5 and 3.6) with maximal left and right hemisphere weights of 0.6 and 1.2 in gradiometers during MI, 0.3 and 0.9 in magnetometers during MI, 1.6 and 2.0 in gradiometers during ME, and 0.9 and 1.5 in magnetometers during ME, respectively. Sensors over primary visual areas show heavy weighting towards alpha synchronization. Removing occipital areas doesn't eliminate this alpha weighting entirely, as it can still be seen in the most posterior sensors (Figs. 3.7 and 3.8).

3.2.1 Recursive Sensor Elimination

Overall the accuracy of models built on a reduced set of ten sensors remained very high (Table 3.4), demonstrating a reduced sensor set can still be highly effective for discriminating MI on our data set. In the Motor task RSE selected gradiometers over right hemisphere sensorimotor areas, magnetometers over bilateral sensorimotor areas, and gradiometers over left hemisphere prefrontal (Fig. 3.9). Imagery task selections selected left hemisphere sensorimotor in the Control condition, and right hemisphere sensorimotor in the Feedback condition. Right hemisphere sensors over prefrontal areas were reliably selected in both Imagery conditions but not in Motor.

3.2.2 Representative Models vs Mean Weights

In traditional machine learning applications, accurate classification is the primary goal and examining the model is secondary if performed at all. In this work, however, we are primarily concerned with which features the model identified as important, and use accuracy as a validation measure for the confidence we should place in these feature selections. This suggests the average weights over each fold should be evaluated as in a bootstrapped statistical analysis, but this averaged model may not actually be particularly good at accurately classifying the data. Either representative models from noteworthy folds must be selected, or the averaged model validated again. Validating the averaged model requires either a second held out data set, further reducing the training set, or applying the averaged model to the entire data set

Table 3.2: Classification Accuracies for Sensor Level Analysis

Task	Imagery				Motor	
Group	Control		Feedback		All	
Occipital	Included	Excluded	Included	Excluded	Included	Excluded
Subj #						
0	–	–	0.921	0.974	0.983	0.967
1	–	–	0.974	0.974	0.967	0.950
2	–	–	0.978	0.978	0.983	0.967
3	–	–	0.542	0.833	0.950	0.950
4	–	–	0.940	0.940	0.983	0.983
5	–	–	0.865	0.827	0.983	0.983
6	–	–	0.980	0.980	0.967	0.967
7	1.000	1.000	–	–	0.967	0.967
8	0.962	0.962	–	–	0.983	0.983
9	0.938	0.979	–	–	0.983	1.000
10	1.000	1.000	–	–	0.967	0.967
11	0.980	0.960	–	–	0.950	0.950
12	1.000	1.000	–	–	0.983	0.950
13	1.000	0.913	–	–	0.983	0.950
14	0.958	0.938	–	–	1.000	1.000
15	1.000	0.854	–	–	0.950	0.950
Mean	0.982	0.956	0.886	0.929	0.974	0.968
CI	0.018	0.037	0.145	0.064	0.008	0.009

Table 3.3: Transfer Loss for Sensor Level Analysis

Task	Imagery				Motor	
Group	Control		Feedback		All	
Occipital	Included	Excluded	Included	Excluded	Included	Excluded
Subj #						
0	–	–	-0.067	-0.015	-0.014	-0.030
1	–	–	-0.022	-0.022	-0.031	-0.048
2	–	–	0.018	0.010	-0.013	-0.030
3	–	–	-0.444	-0.148	-0.049	-0.049
4	–	–	-0.040	-0.028	-0.016	-0.016
5	–	–	-0.131	-0.169	-0.014	-0.014
6	–	–	-0.008	-0.012	-0.031	-0.031
7	0.008	0.003	–	–	-0.032	-0.032
8	-0.033	-0.036	–	–	-0.016	-0.017
9	-0.062	-0.018	–	–	-0.014	0.002
10	0.003	0.003	–	–	-0.030	-0.030
11	-0.020	-0.040	–	–	-0.049	-0.049
12	0.000	0.011	–	–	-0.016	-0.047
13	0.003	-0.081	–	–	-0.017	-0.049
14	-0.042	-0.062	–	–	0.003	0.003
15	0.000	-0.146	–	–	-0.049	-0.048
Mean	-0.016	-0.041	-0.099	-0.055	-0.024	-0.030
CI	0.019	0.039	0.147	0.067	0.008	0.009

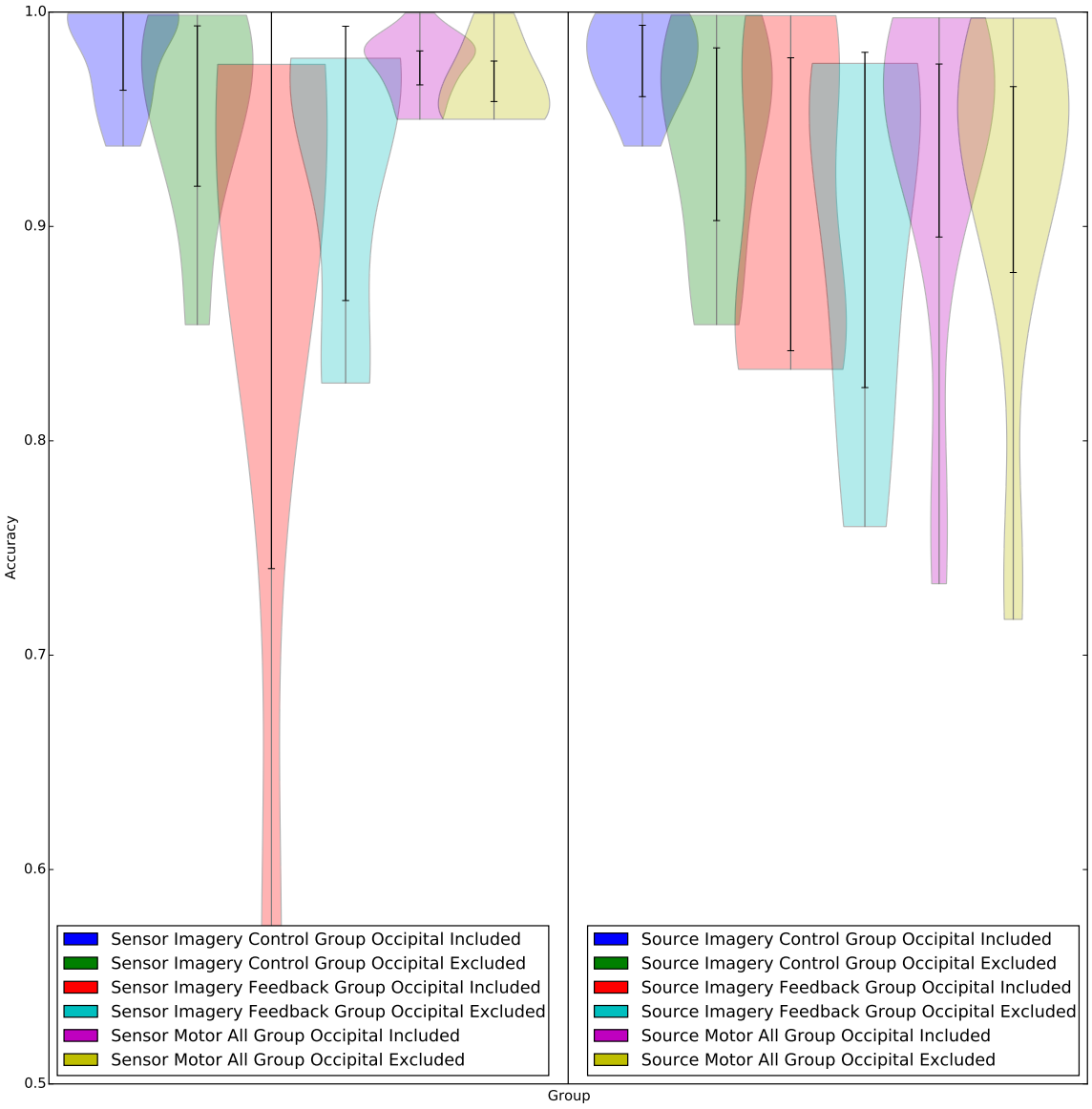

Figure 3.3: The probability density curves for accuracy for each sensor and source level analysis is plotted with error bars representing 95% confidence intervals of the mean. Participant 3 was particularly difficult to classify, as seen in Table 3.2, and caused the extremely large variance in the Sensor Feedback Occipital Included distribution.

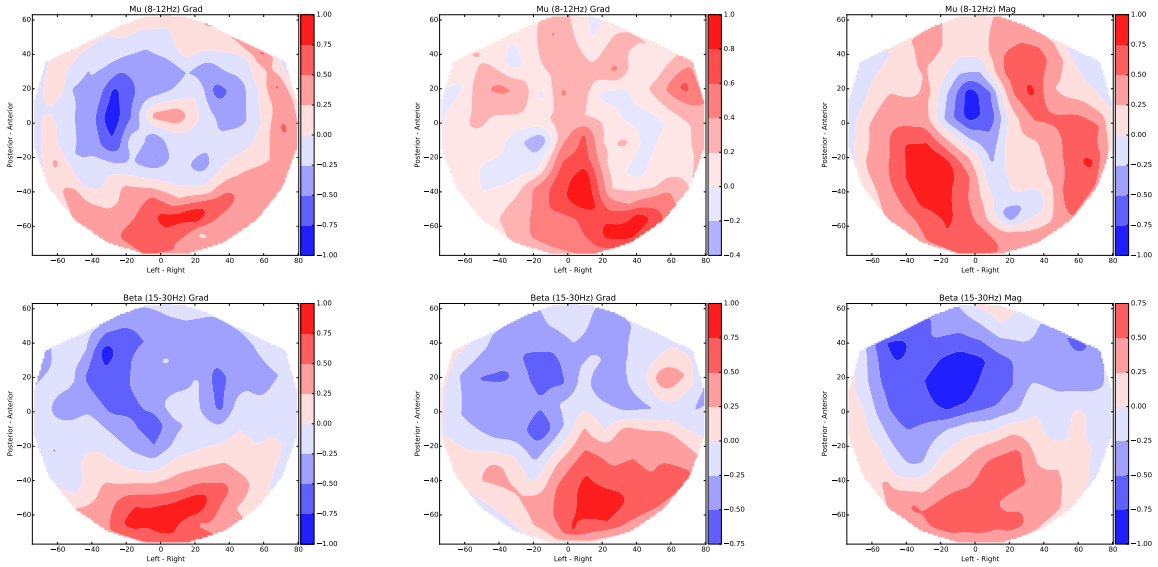

Figure 3.4: Sensor level model of Control group during MI, participant 15 held back for testing. Blue indicates desynchronization was associated with Imagery in the associated frequency band, red synchronization was associated with Imagery. Weights are normalized within each frequency band. Alpha and Beta synchronization in occipital cortex was weighted heavily throughout the control group, to a much higher degree than the feedback group. Mu and Beta desynchronization was weighted in left hemisphere sensors over sensorimotor areas.

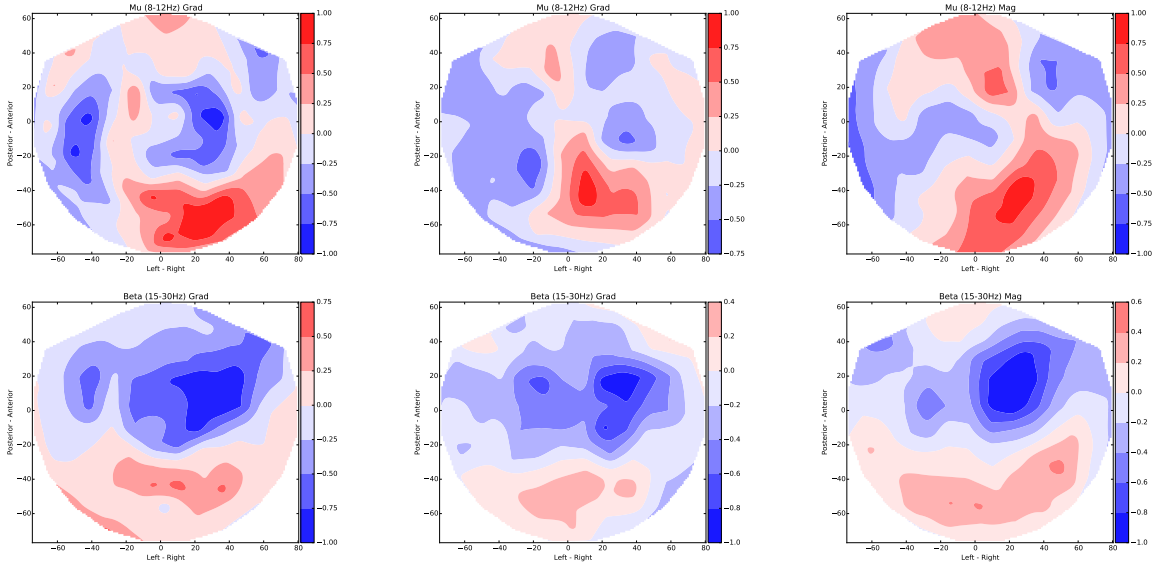

Figure 3.5: Sensor level model of Feedback group during MI, participant 3 held back for testing. Occipital mu synchronization was selected in a larger group of sensors than in ME. Right hemisphere beta desynchronization was again weighted approximately twice as heavily as left. Bilateral mu desynchronization in sensorimotor cortex was also selected in the planar gradiometers.

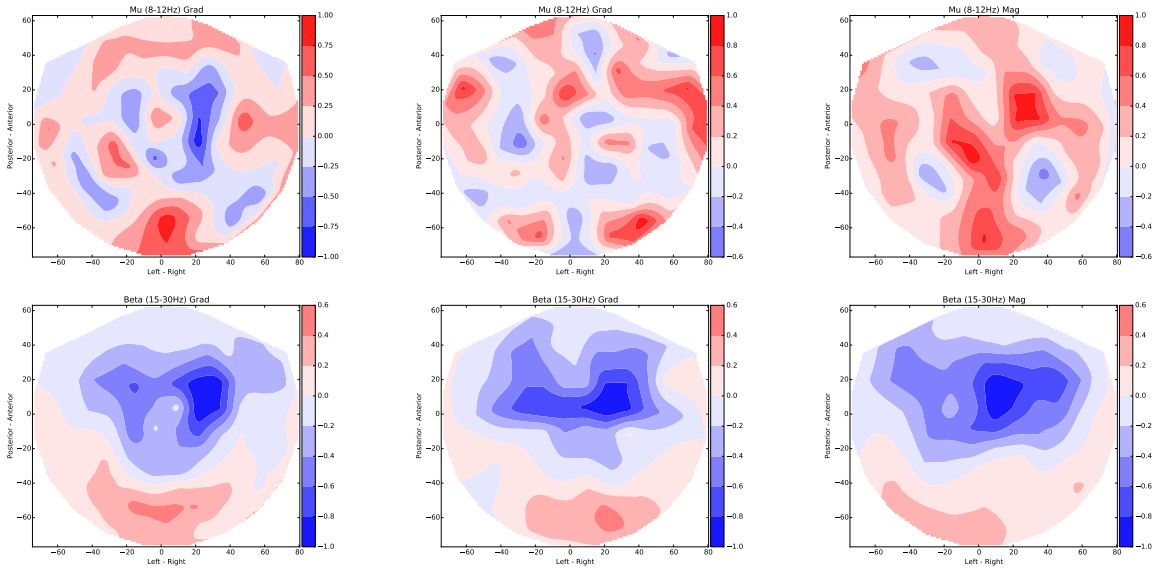

Figure 3.6: Sensor level model of all participants during ME, participant 3 held back for testing. Occipital alpha/mu and beta synchronization, along with right hemisphere sensorimotor beta desynchronization, were preferentially weighted. Right hemisphere sensorimotor beta was weighted approximately twice as heavily as left.

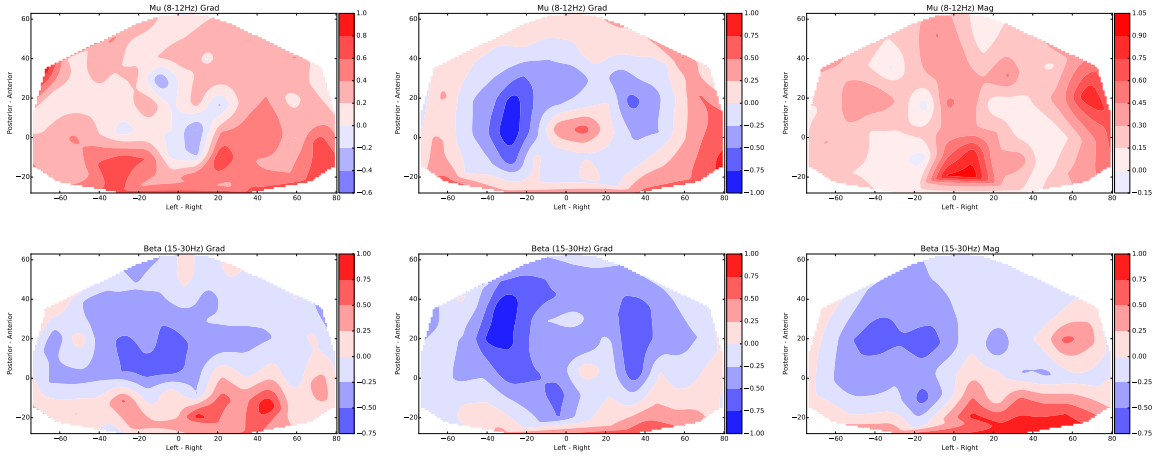

Figure 3.7: Sensor level model of Control group during MI without occipital sensors, participant 15 held back for testing. The same left hemisphere bias in beta and mu desynchronization can be seen. With occipital sensors removed, alpha synchronization weighting can be seen in the remaining most posterior sensors.

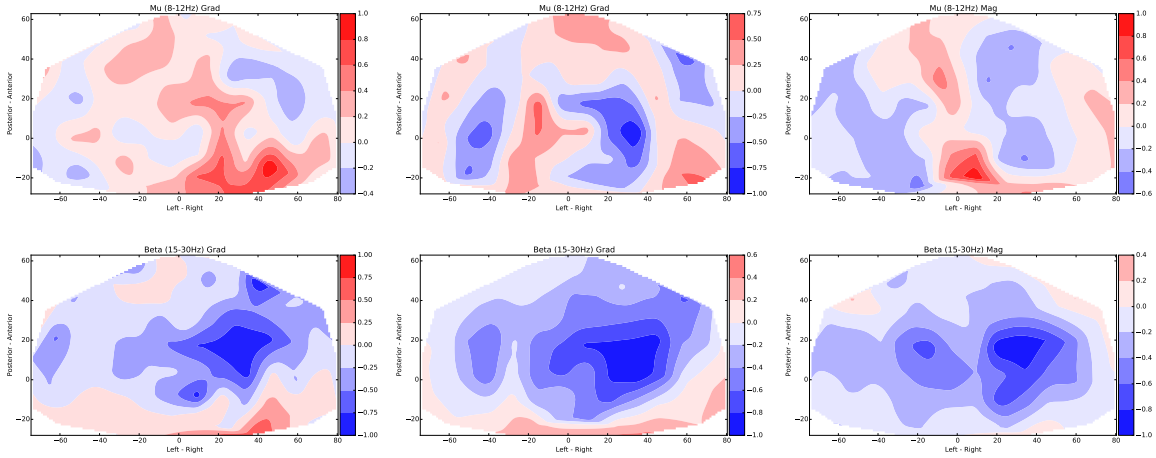

Figure 3.8: Sensor level model of Feedback group during MI without occipital sensors, participant 3 held back for testing. Weighting across beta and alpha/mu are largely unchanged with the removal of occipital sensors. The lower relative weighting of beta synchronization in occipital sensors in the Feedback group suggest that it is less important to differentiating the two classes, and the minimal effect removing these sensors had on the rest of the model seems to support this.

Table 3.4: Classification Accuracies for Recursive Sensor Elimination Analysis

Task	Imagery				Motor	
Group	Control		Feedback		All	
Occipital	Included	Excluded	Included	Excluded	Included	Excluded
Subj #						
0	–	–	0.921	0.921	0.967	0.983
1	–	–	0.974	0.842	0.950	0.950
2	–	–	1.000	0.978	0.950	0.933
3	–	–	0.542	0.625	0.967	0.967
4	–	–	0.900	0.940	0.983	0.983
5	–	–	0.808	0.788	1.000	0.983
6	–	–	0.980	0.980	0.983	0.967
7	1.000	1.000	–	–	0.983	0.983
8	0.904	0.942	–	–	0.967	0.933
9	0.958	0.958	–	–	1.000	1.000
10	1.000	0.964	–	–	0.967	0.983
11	0.980	0.960	–	–	0.967	0.967
12	1.000	1.000	–	–	0.900	0.933
13	0.957	0.804	–	–	0.967	0.917
14	0.958	0.833	–	–	1.000	1.000
15	1.000	0.833	–	–	0.950	0.950
Mean	0.973	0.922	0.875	0.868	0.969	0.965
CI	0.025	0.059	0.149	0.119	0.013	0.014

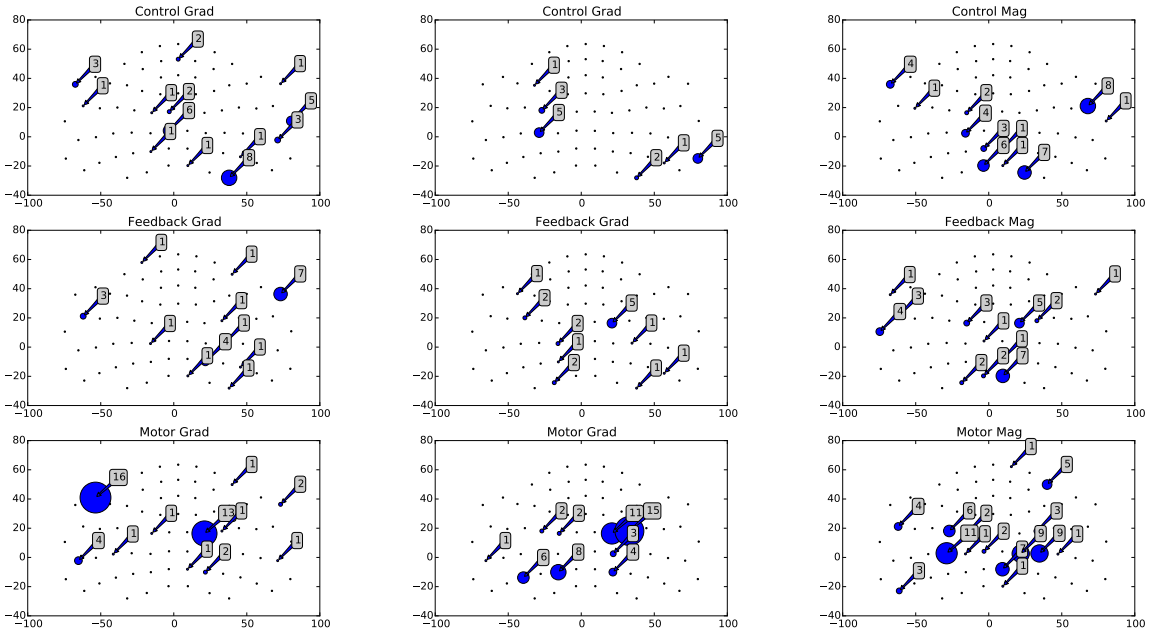

Figure 3.9: Sensor plots with anterior up, posterior down. For each sensor, the number of times it remained in the top ten after recursive sensor elimination in each condition is shown. The Feedback group contained 7 participants, Control 9 participants, and for the motor analysis both groups were pooled for a total of 16 participants. The ME selections reliably clustered around Fc4/C4 in the right hemisphere, F9 on the left hemisphere, and to a lesser extent Fc3/C3 on the left hemisphere. The MI selections were less stable, but this is perhaps not surprising as with each person removed a proportionally larger amount of the data was also removed. On both MI data sets, posterior sources likely sensitive to alpha were selected, further confounding the results. The very reliable response of primary visual alpha leads to high weightings, which make sensors which can detect it very attractive choices to this selection system. Feedback and Control still demonstrated the laterality biases seen in all sensor and the later source models, with Feedback preferring right hemisphere sensors around Fc4/C4 and Control left hemisphere Fc3/C3. Over all sets, magnetometers were the most frequently selected sensors.

and relaxing the requirement that performance be evaluated on as yet unseen data. We presented here representative models, but for completeness the averaged weight figures are presented in the appendix.

3.2.3 Summary of results for Objective 2

RSE generated distinctive patterns clustering around a small subset of sensors, but interestingly these sensors were not necessarily in regions shown to be heavily weighted in all sensor models. For instance, ME models universally chose a gradiometer in left hemisphere prefrontal cortex, which received only small weights in beta and mu in all-sensor models. The accuracy of the models suggests they are still extracting signals capable of discriminating MI effectively, but more analysis is required to determine what exactly makes the bilateral prefrontal sensors selected frequently in all models valuable features.

As more data is gathered, confidence in the discriminative value of the identified clusters will improve. This system is capable of identifying a minimal sensor set that will work for many participants, but even with our small sample size we observed outliers for whom this model did not apply. With a large enough sample, if these outlying participants can be grouped and show reliable patterns, it may be possible to build a voting pool of several models which can collectively discriminate MI in most of the population.

3.3 Objective 3: Using source level modelling of MEG data across the whole brain, can we identify the anatomical regions whose activity best discriminates MI across our participants?

Source level models obtained accuracies lower than sensor level equivalents but within confidence intervals for their means, as seen in Table 3.5 and Fig. 3.3. Transfer loss is shown in Table 3.6. Models with participant 4 held out were chosen to represent Source level Feedback Imagery and Motor, while Participant 15 was again chosen for Control Imagery. Source level models built on ME weighted beta desynchronization in the premotor network including dorsolateral and ventrolateral premotor cortex, superior parietal cortex, and the precuneus heavily, as well as alpha desynchronization in superior parietal cortex and the precuneus (Fig. 3.10). Beta desynchronization

in primary motor areas was of interest but weighted less heavily. Alpha synchronization throughout primary visual cortex was weighted heavily, as participants were instructed to close their eyes during ME to mirror imagery instructions.

Models built on feedback group imagery were similarly focused on beta desynchronization in the premotor and sensorimotor networks, with a similar weighting pattern to motor activity (Fig. 3.11). Increased weight was placed on occipital alpha synchronization in V1, V2, and the Fusiform Gyrus. The importance of these areas is artificially inflated for the task of discriminating MI vs rest by instruction to keep the eyes open during rest. Eliminating occipital cortex sources reduced accuracy by an average of -0.0073 ± 0.0409 for Feedback, -0.0342 ± 0.0320 for Control, and -0.0126 ± 0.0103 for Motor. While beta weighting remained largely unchanged after removal of occipital sources, right hemisphere alpha synchronization showed dramatically increased weighting in the thalamus, parahippocampal cortex, and ventral temporal cortex (Fig. 3.12).

Models built on control group imagery showed a bias towards left hemisphere beta desynchronization in the premotor and sensorimotor networks, in contrast to the right hemisphere focused models built on feedback participants, with an increased emphasis on beta synchronization in primary visual areas (Fig. 3.13). These effects may be attributed to a much stronger or more broad-spectrum synchronization during imagery in primary visual in the control group. When occipital areas are eliminated, the control group displayed a pattern of reweighing similar to the feedback group after occipital was removed, with the notable difference of alpha synchronization being weighted heavily in left hemisphere thalamus, parahippocampal cortex, and ventral temporal cortex (Fig. 3.14). The bias towards left hemisphere beta desynchronization was increased after removing occipital sources.

3.3.1 Source Rankings

The beta source rankings laid out in Fig. 3.15 highlight many of the largest and most promising effects of neurofeedback. Feedback models showed dramatically higher weightings for beta desynchronization in primary motor cortex than Control and even Motor models. Left M1: Control -0.217±0.068, Feedback -0.710±0.144, Motor -0.253±0.050; Right M1: Control -0.068±0.070, Feedback -0.636±0.080, Motor

Table 3.5: Classification Accuracies for Source Level Analysis

Task	Imagery				Motor	
Group	Control		Feedback		All	
Occipital	Included	Excluded	Included	Excluded	Included	Excluded
Subj #						
0	–	–	0.895	0.974	0.783	0.750
1	–	–	1.000	0.947	0.983	0.950
2	–	–	0.978	0.978	0.900	0.883
3	–	–	0.833	0.875	0.933	0.983
4	–	–	0.840	0.760	0.733	0.717
5	–	–	0.846	0.827	0.950	0.917
6	–	–	0.980	0.960	0.967	0.967
7	1.000	1.000	–	–	0.983	0.983
8	0.962	0.942	–	–	0.950	0.917
9	0.938	0.958	–	–	0.967	0.950
10	1.000	0.964	–	–	0.967	0.967
11	0.980	0.980	–	–	0.983	0.983
12	1.000	1.000	–	–	1.000	1.000
13	0.978	0.913	–	–	0.967	0.950
14	0.958	0.875	–	–	0.900	0.883
15	0.979	0.854	–	–	1.000	0.950
Mean	0.977	0.943	0.910	0.903	0.935	0.922
CI	0.017	0.040	0.068	0.078	0.040	0.043

Table 3.6: Transfer Loss for Source Level Analysis

Task	Imagery				Motor	
Group	Control		Feedback		All	
Occipital	Included	Excluded	Included	Excluded	Included	Excluded
Subj #						
0	–	–	-0.101	-0.019	-0.216	-0.247
1	–	–	0.008	-0.045	-0.010	-0.044
2	–	–	-0.006	-0.022	-0.096	-0.114
3	–	–	-0.130	-0.092	-0.050	0.002
4	–	–	-0.144	-0.208	-0.263	-0.280
5	–	–	-0.150	-0.169	-0.048	-0.082
6	–	–	-0.004	-0.024	-0.032	-0.032
7	0.008	0.008	–	–	-0.013	-0.011
8	-0.033	-0.052	–	–	-0.048	-0.080
9	-0.062	-0.039	–	–	-0.031	-0.048
10	0.000	-0.036	–	–	-0.033	-0.033
11	-0.020	-0.020	–	–	-0.017	-0.014
12	0.003	0.006	–	–	0.002	0.003
13	-0.016	-0.075	–	–	-0.027	-0.039
14	-0.039	-0.122	–	–	-0.100	-0.112
15	-0.021	-0.146	–	–	0.017	-0.033
Mean	-0.020	-0.053	-0.075	-0.083	-0.060	-0.073
CI	0.017	0.041	0.066	0.072	0.041	0.044

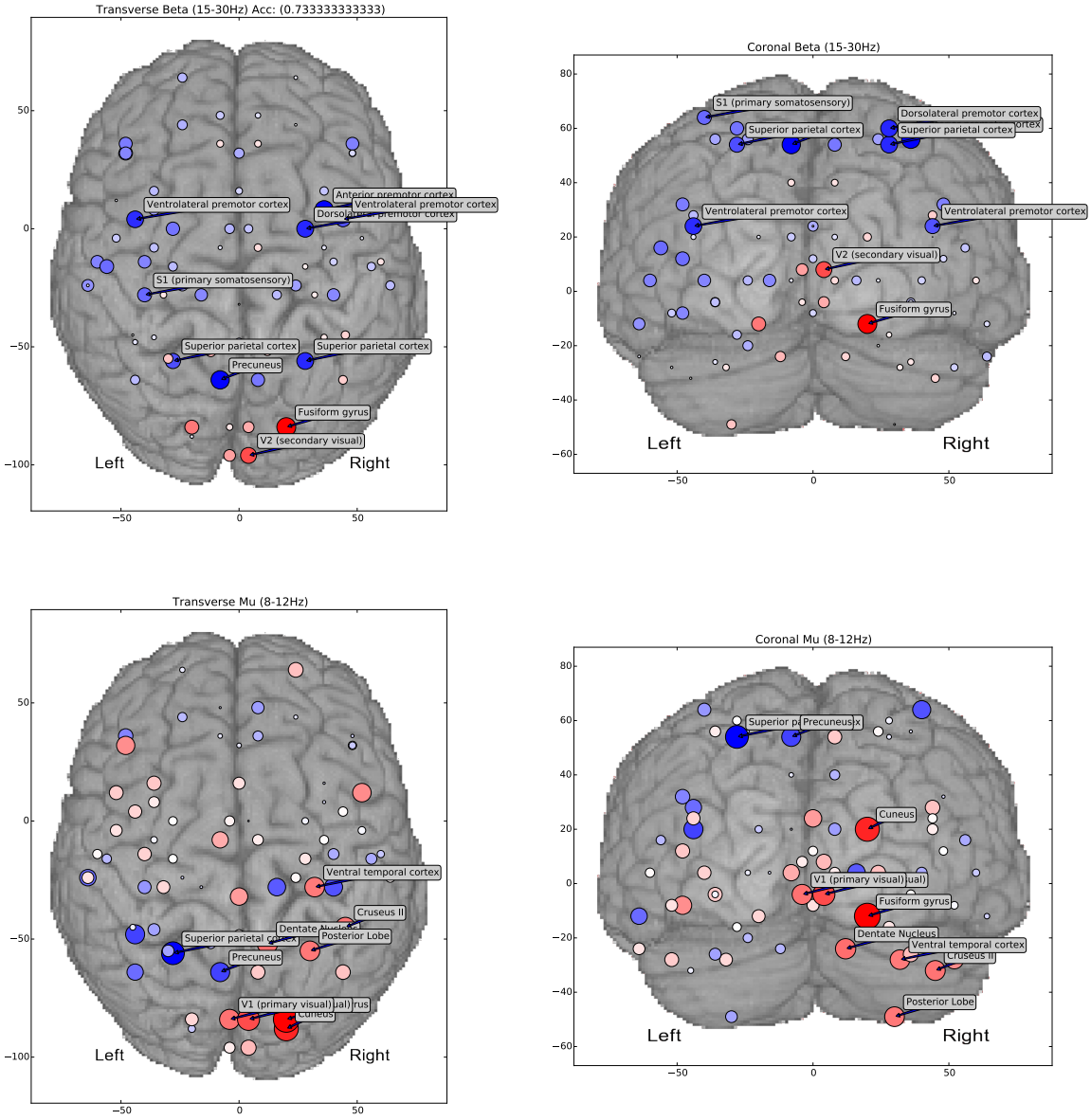

Figure 3.10: Source level model of all participants during ME, participant 4 held back for testing. The size of each source represents the mean absolute value of the weights in the specified frequency range, with colour intensity reflecting the degree to which the mean weight was negative (desynchronization, blue) or positive (synchronization, red). The ten sources with highest mean absolute weighting are labelled. Unlike Kraeutner et al [2], our model found activity in the precuneus, fusiform gurus, and precuneus predictive of ME. They did not examine mu, but our model found mu activity also predictive in the cuneus, precuneus, fusiform gyrus, and V1. Unexpectedly, beta activity in primary motor cortex was not amongst the ten most predictive sources.

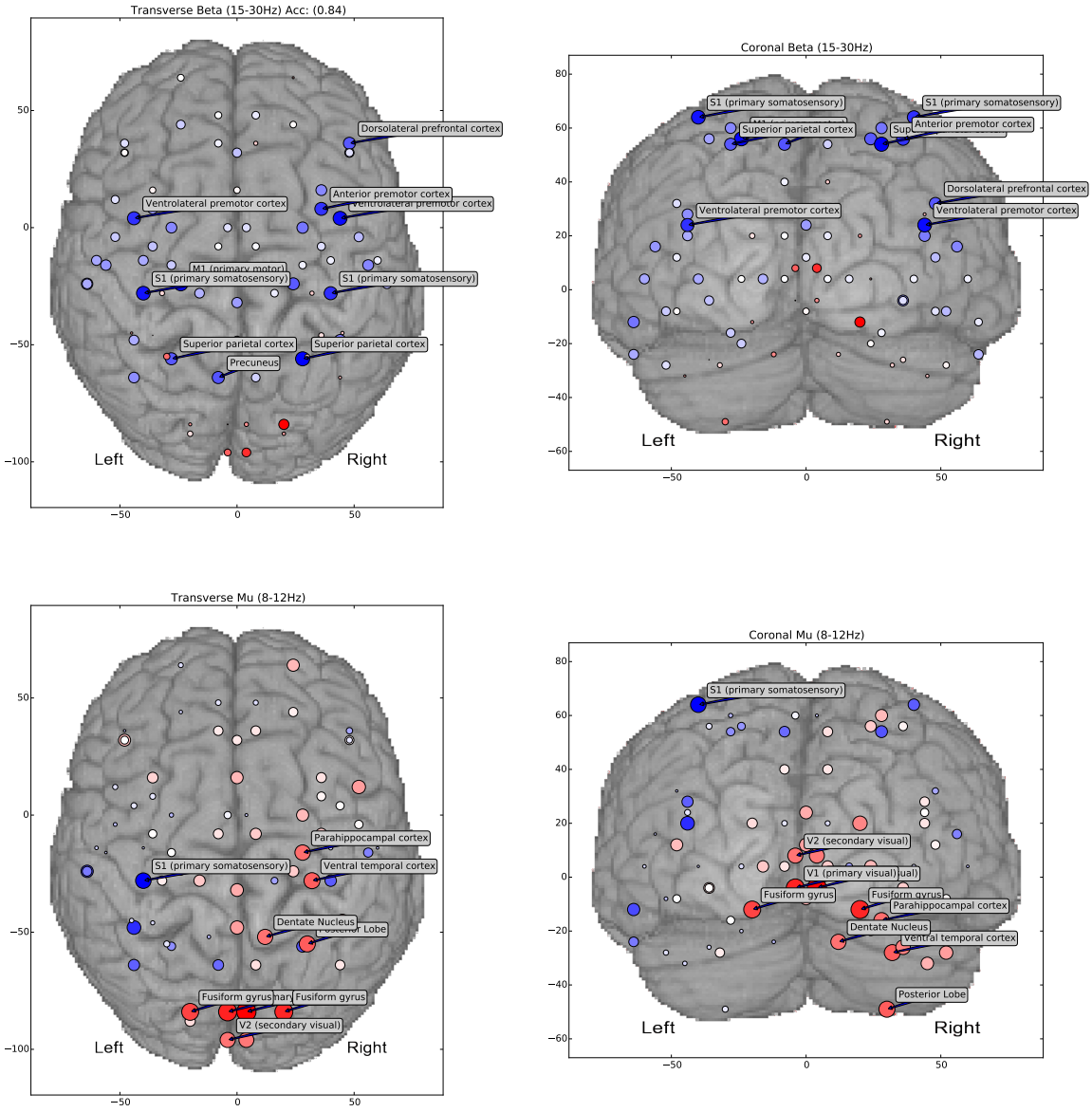

Figure 3.11: Source level model of Feedback group during MI, participant 4 held back for testing. Bilateral premotor, S1, M1, and right hemisphere superior parietal cortex were weighted in beta ERD. Left hemisphere S1 was weighted in alpha ERD, and alpha ERS was weighted in primary visual areas and right hemisphere sources including fusiform gyrus, parahippocampal, and ventral temporal cortex.

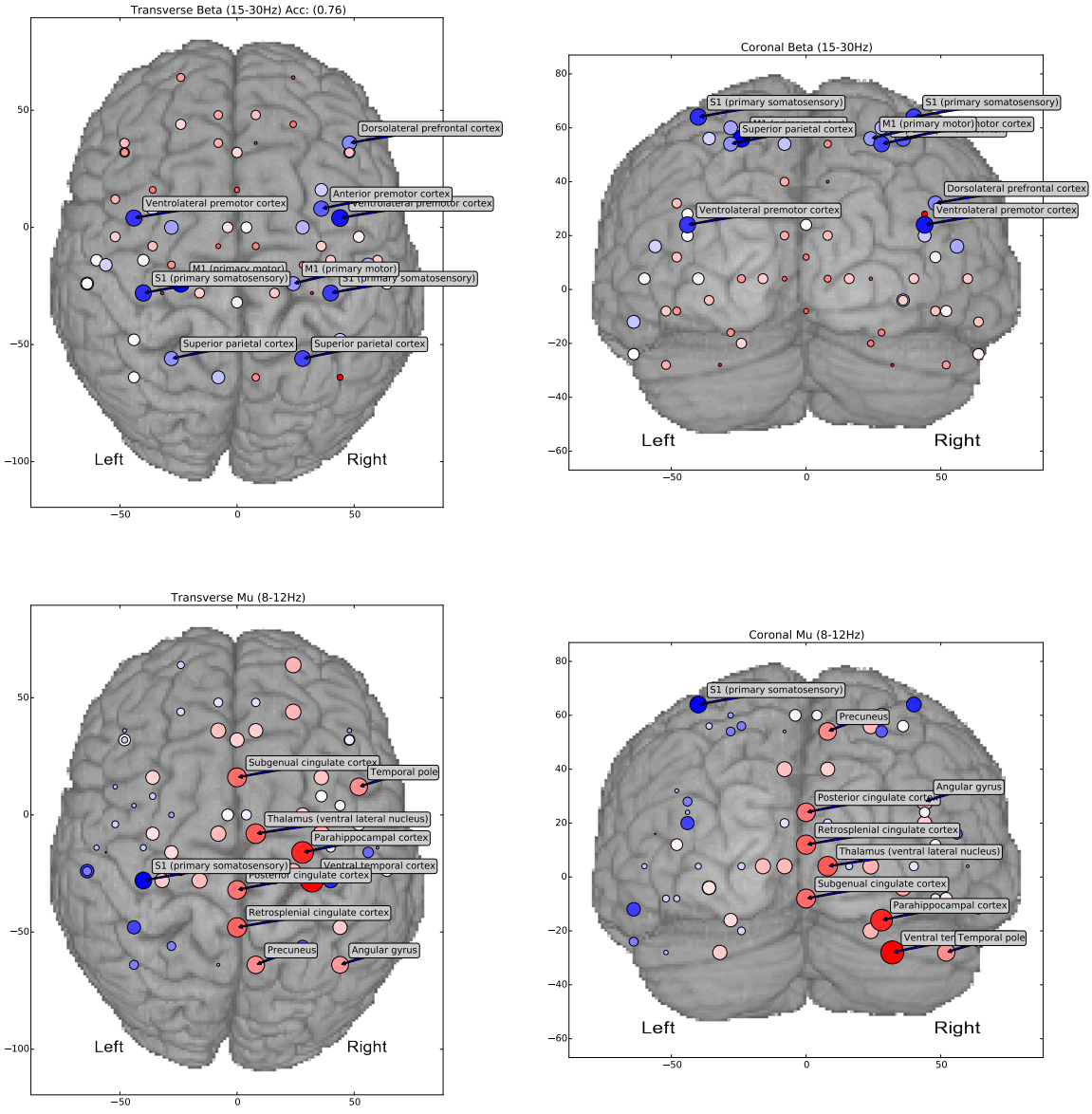

Figure 3.12: Source level model of Feedback group during MI without occipital sources, participant 4 held back for testing. Alpha ERD was primarily weighted in left hemisphere S1 and superior parietal cortex. Alpha ERS was primarily weighted in right hemisphere thalamus, subgenual cingulate, parahippocampal, and ventral temporal cortex. Feedback appears to have shifted the degree to which lateralization of Alpha ERS was predictive of MI.

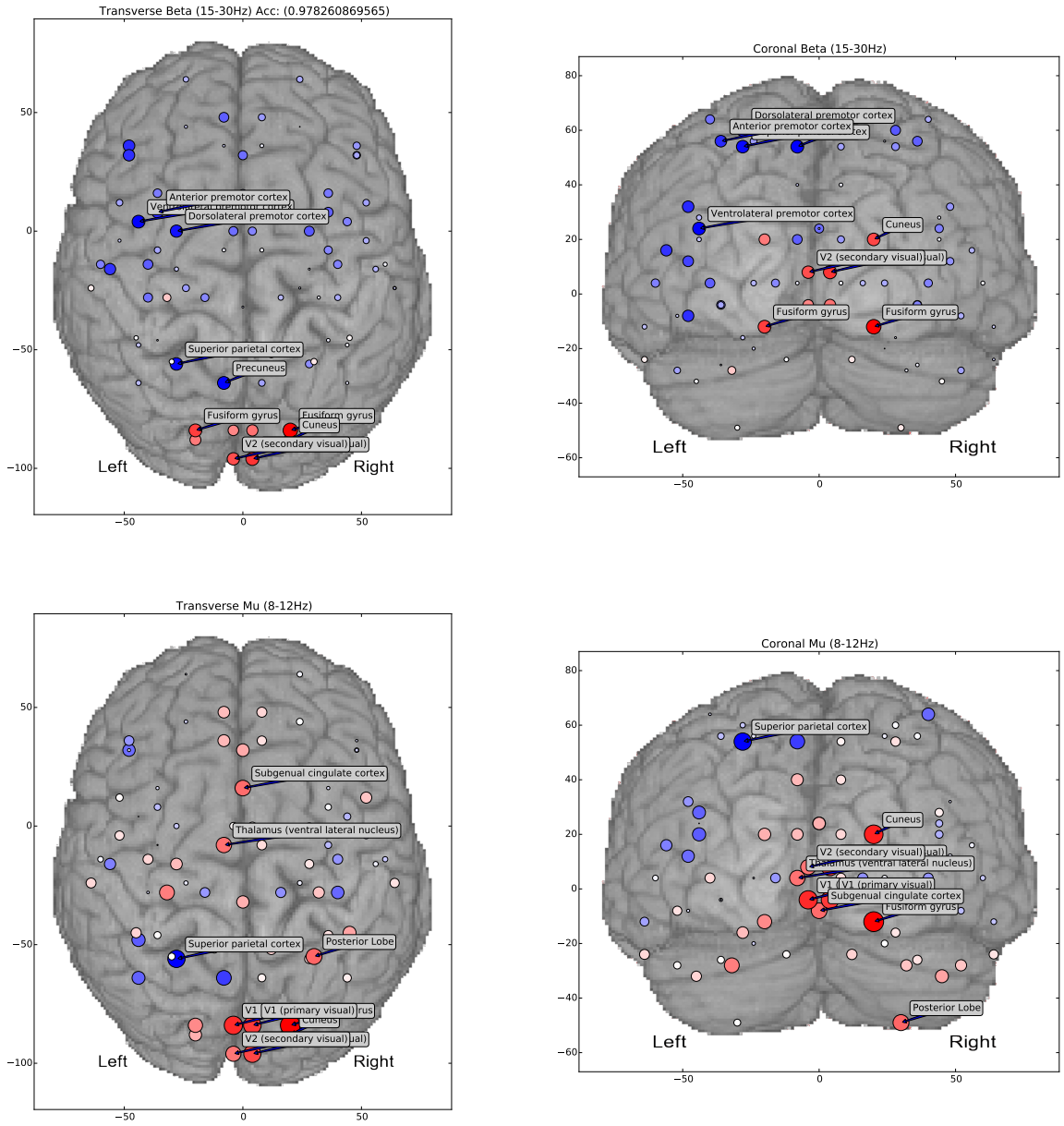

Figure 3.13: Source level model of Control group during MI, participant 13 held back for testing. Left hemisphere premotor and superior parietal cortex were weighted in beta ERD, superior parietal cortex in alpha ERD, and both beta and alpha ERS in primary visual areas.

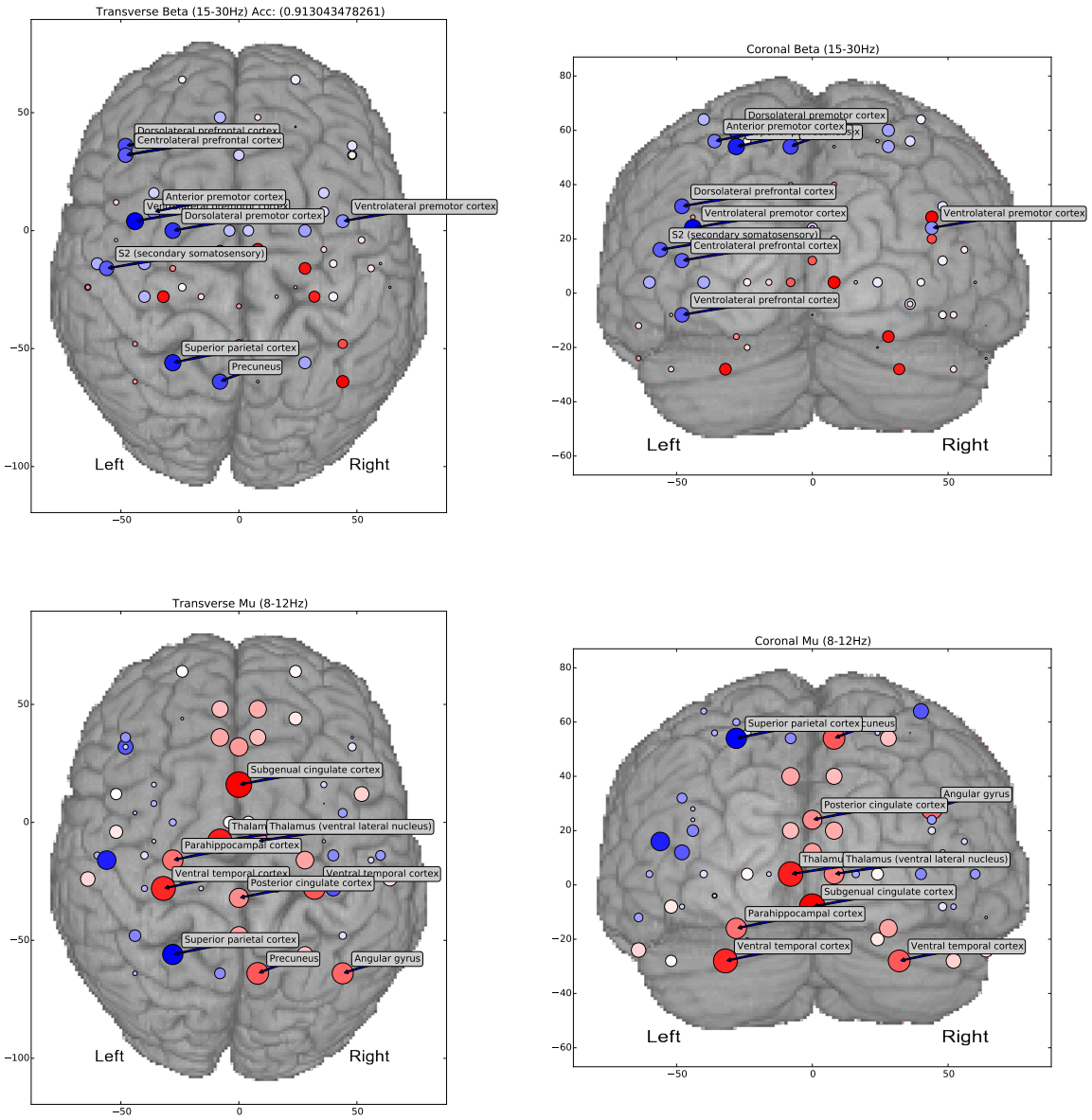

Figure 3.14: Source level model of Control group during MI without occipital sources, participant 13 held back for testing. Beta ERD is primarily weighted in left hemisphere premotor, prefrontal, and superior parietal sources. Alpha ERD was weighted in left superior parietal cortex, and alpha ERS in subgenual cingulate cortex, ventral temporal, and thalamic sources along the midline and left hemisphere.

-0.405±0.044. Neurofeedback increased the degree to which beta desynchronization throughout the sensorimotor and premotor networks in contralateral hemisphere was reliably predictive of MI. Participants were instructed to focus on kinaesthetic imagery, imagining haptic and proprioceptive sensations which would result from the imagined action. S1 plays a key role in processing haptic feedback, and these instructions may have contributed to the high weightings assigned to primary somatosensory cortex after feedback. Left S1: CTRL -0.379±0.079, FB -0.776±0.087, Motor -0.519±0.049; Right S1: CTRL -0.199±0.104, FB -0.653±0.103, Motor -0.420±0.054. The ventrolateral premotor cortex in both hemisphere was consistently amongst the most heavily weighted in all conditions. Areas throughout sensorimotor cortex were also amongst the most heavily weighted. Right hemisphere M1 was important in motor and feedback models, but not control models. Left M1 was important only in feedback models, having relatively low weight in motor and control. The right hemisphere anterior insula, whose neighbour the posterior insula has reciprocal connections to S2 and the thalamus, was weighted heavily in feedback models. This may be a case of over specific labelling beyond the resolution of the beamformer, with activation in posterior insula being included in the estimated source activity for anterior insula. Beta synchronization weights were, as seen in both the sensor and source models, much lower in the feedback group than in control. Feedback seems to have induced dramatic shifts in superior parietal cortex lateralization, with the right hemisphere going from synchronization to desynchronization, and the weighting of left hemisphere desynchronization dropping dramatically. Left Superior Parietal Cortex: CTRL -0.728±0.103, FB -0.081±0.124, Motor -0.647±0.037; Right Superior Parietal Cortex: CTRL 0.211±0.061, FB -0.543±0.105, Motor -0.052±0.028. (Fig. 3.16)

3.3.2 Summary of results for Objective 3

Source level models were able to predict state in unseen participants with high accuracy, and again successfully identified the pattern of lateralization previously found to be associated with neurofeedback. Centrolateral and ventrolateral prefrontal cortex, dorsolateral and ventrolateral premotor cortex, S1, and superior parietal cortex were all identified as potential targets for neurofeedback in the beta band, along with the ventral temporal cortex and thalamus for neurofeedback in the alpha band. The two

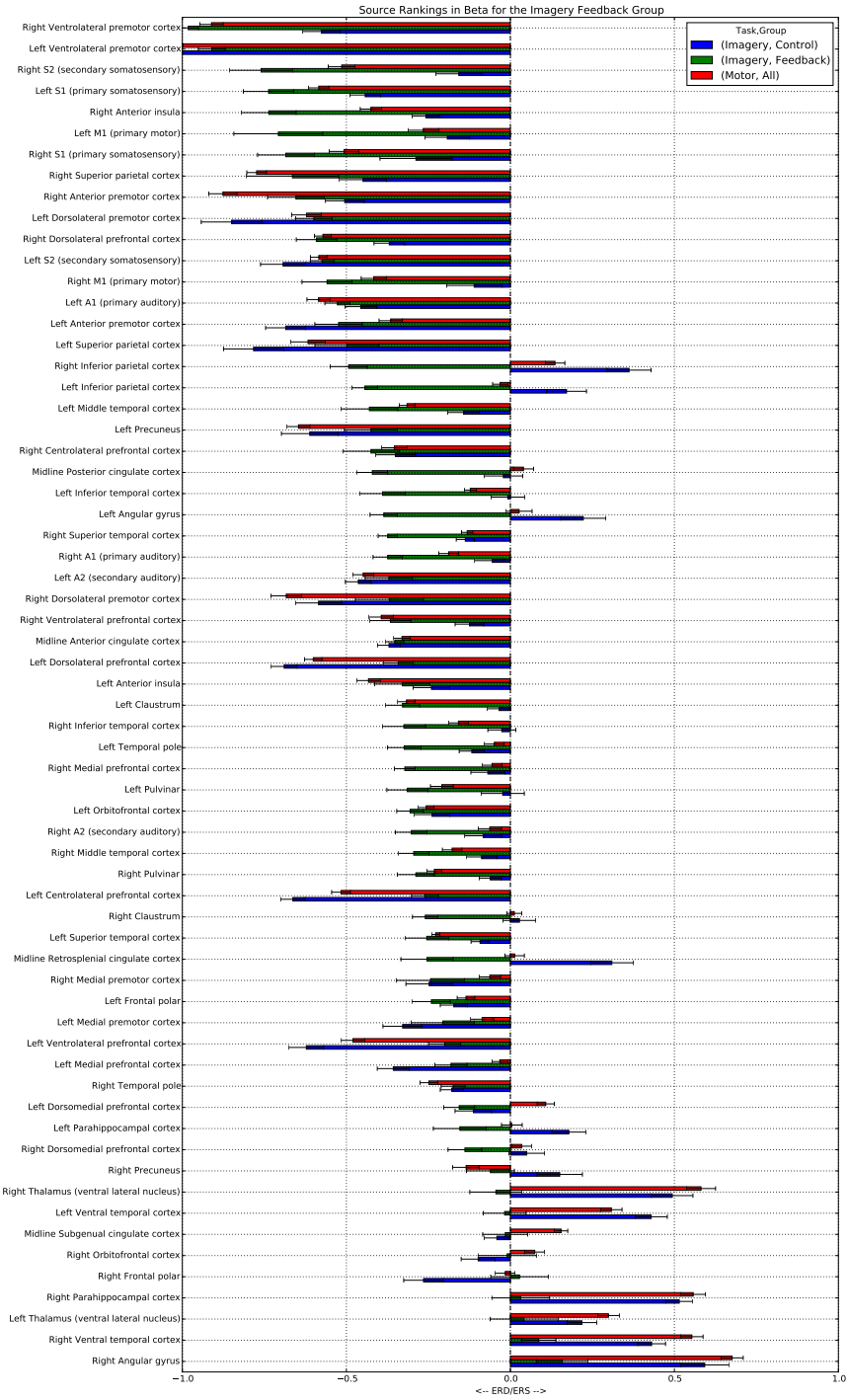

Figure 3.15: Sources ranked by mean normalized weighting in the beta band (15-30Hz) with 95% confidence intervals. Models trained on participants performing MI or ME in Session 3. The Feedback group contained 7 participants, Control 9 participants, and Motor both groups pooled for a total of 16 participants.

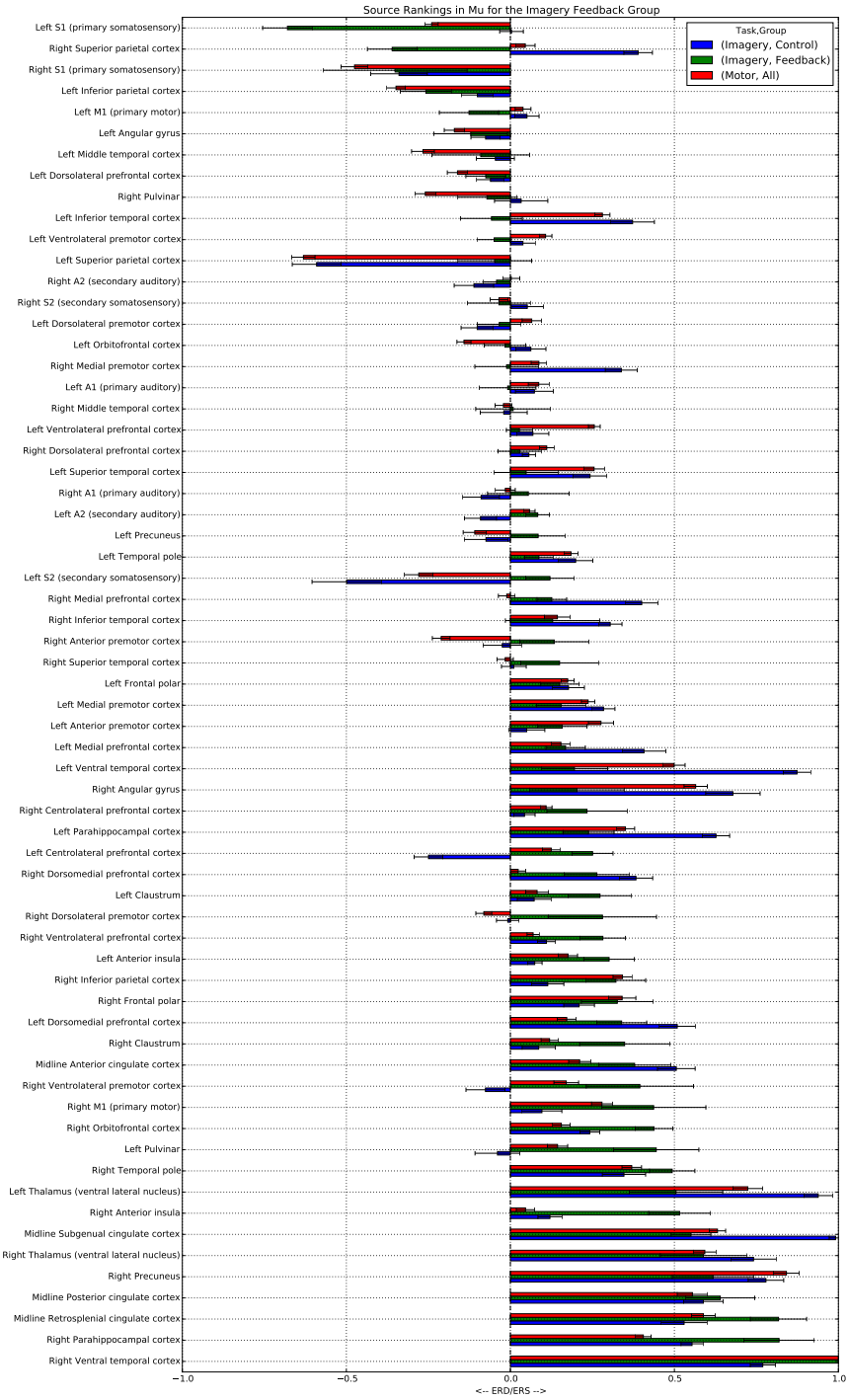

Figure 3.16: Sources ranked by mean normalized weighting in the mu band (8-12Hz) with 95% confidence intervals. Models trained on participants performing MI or ME in Session 3. The Feedback group contained 7 participants, Control 9 participants, and Motor both groups pooled for a total of 16 participants.

groups primarily differed in laterality, with both weighting roughly the same sources but in opposite hemispheres.

Chapter 4

Discussion

We sought to identify whether two ECD estimated sources in right and left sensorimotor cortices localized via median nerve stimulation were sufficient to accurate discriminate MI vs Rest. We hypothesized, based on previous work showing changes in brain activity after beta neurofeedback using those two sources, that MI would be discriminable based on this same two source feature set. We sought to identify which MEG sensors were most discriminative in that same classification problem and which minimal set of MEG sensors was sufficient for reliable classification, hypothesizing that sensors positioned over supplementary and premotor areas, along with primary motor cortex and the superior parietal cortex, being closest to the signals of interest, would be most discriminative. Finally, we sought to identify which frequency bands in which anatomical sources were most discriminative of MI vs Rest. We hypothesized that beta in primary motor, premotor, and superior parietal cortices would be most discriminative, based on their key roles in the motor planning networks as identified in the literature.

Models built on two sources localized via median nerve stimulation could not reliably discriminate MI from rest across our participant pool. This could be due to the variability in source location, making a single model incapable of capturing the different patterns evoked at each location, or it could be a case of insufficient information to reliably overcome noise and patterns from unrelated activity. Accuracy improved dramatically when the model was built using a larger number of consistently placed sources, or all sensors. These larger models had further utility in identifying information rich areas for sensor placement in a proposed clinically practical small electrode EEG system, and in identifying which anatomical regions provide the most discriminative feedback across participants. It is unknown whether provision of feedback based on these areas would improve functional outcomes, but my work provides avenues of inquiry for further MEG based research into the dynamics of MI and

neurofeedback.

As the ultimate goal is to improve provision of neurofeedback, it is encouraging to see the relative ease with which models of MI can transfer to new participants and retain high classification accuracies. The models all learned important differences between the Feedback and Control groups, capturing the same pattern of lateralization of activity reported in Boe et al [1]. Once developed, the system can scale to very large data sets on even moderate hardware using batch learning, processing a small number of trials at a time and iteratively improving the model, allowing hundreds of thousands of participants to be used, improving generalization and effectively avoiding patient training periods which may be clinically impractical.

Online learning and cross-domain adaptation are two machine learning strategies that may have valuable applications in providing neurofeedback, but may also have undesirable drawbacks. Online learning is the process of updating the model with each new training example as it's measured and used to provide feedback, rather than using a static pre-trained model. Cross-domain adaptation, or transfer learning, involves adapting model parameters to better fit a new participant or patient in a way that accounts for some of the inter-patient variability without changing the underlying structure the model has already learned. Both can improve classification accuracy by accounting for specific nuances of a given participant's pattern of activity, but improving functional outcomes and not classification accuracy is the true goal. A moving goalpost for feedback could be frustrating for participants if it drifts away, but a goal that adaptively moves towards the participant's current state is not conducive to driving changes in activity which will in turn drive neuroplasticity in desired areas.

4.1 General Limitations

As rest trials were only used as a baseline in the initial provision of feedback there was no issue with instructing participants to keep their eyes open while resting. This had the unfortunate effect of making it extremely difficult to determine whether there were any genuine effects of MI or provision of feedback on alpha rhythms, and provided a potentially spurious but very predictive feature. The shift in alpha weighting from bilateral in control to primarily right lateralized in feedback suggests there may have been an effect of feedback, but how this is linked to MI is difficult to determine from

this data. It is possible the feedback group are not only better at selectively focusing on kinaesthetic imagery of the left hand, but also at selectively visualizing movement of the left hand and restricting activity to the left visual hemifield.

Beamforming is a powerful technique for source localization, but as with any solution to the inverse problem it comes with assumptions and drawbacks. Due to the way it uses the covariance matrix of the sensors to attenuate signals not originating from the source of interest, two sources which are very tightly correlated may effectively cancel each other out, eliminating output from both. In practice it takes sources that are almost perfectly correlated to cause this effect, and it generally does not cause problems in neuroimaging, but it is possible that pairs of very tightly causally linked sources may not appear in our data. Further, it is not a perfect spatial filter. Signals generated by nearby or intervening regions may bleed in to the signal localized to a specific source. In all cases, it must be kept in mind that the label given to a source may not be the only or even the primary generator of the signal recorded. In conjunction with the use of a template brain in our boundary element model, this likely means there is a small amount of variation in exactly which anatomical region is under each virtual source between participants [69].

Training vectors were averages over 7 seconds of data, drawn from each of the 10 second trials. Reduced time segments with less or no averaging could have been used to generate a much larger number of trials with more noise which would be expected under real use conditions. The longer the epoch averaged over, the greater the degree to which a stochastic signal component cancels itself while the evoked components of the signal remain unattenuated. The preprocessing method was chosen to remain consistent with that used in previous work by Boe et al [1], which found neurofeedback based on similar preprocessing effective in driving changes in brain activity. Since participants are often instructed to keep their eyes closed during imagery, improved temporal response times below the length of the imagery block are of questionable value compared to an improved signal to noise ratio and more representative model for analysis. In an alternative feedback paradigm, reduced sample lengths might be advantageous, and this should be explored in future work.

As the models built were all rest vs task in session 3, they do not directly speak to the effects of feedback or training, but instead only to state specific activity patterns

within each group. A multi class data set containing session and group labelled data would pose a much more difficult classification task, but would generate a model of both feedback and training effects. A more powerful learning algorithm may be required to solve this more difficult classification problem, such as those discussed in future work, but the complexity of the models they generate make them commensurately more difficult to interpret and visualize.

As the models are built with a data driven learning system, it is entirely possible for the model to learn to reinforce undesirable behaviour. On our data set, it learned which features best discriminate the imagery the participants in a group performed from rest, not the features which would yield the best outcomes were they to be provided as feedback. Training the model on first time participants with no MI experience, or on patients in the acute period post stroke, and using those models to provide feedback would almost certainly yield poor results as they were merely taught to continue what they were already doing. Identification of the best feature set for provision of effective feedback requires a data set which includes outcome measures or participants known to be performing skilled MI, and preferably both. We believe the feedback group of participants are performing higher quality MI than the control group, due in large part to the lateralization of activity to the correct hemisphere, and their increased amplitude of beta desynchronization, but in the absence of a quantitative means of evaluating MI quality this remains speculation. A model trained on quantifiably high quality MI data and given a limited ability to adapt certain parameters to new participants - shifting source locations to account for small anatomical differences, or translating weights up or down in frequency space to account for varying centre frequencies of mu or beta activity, for instance - may be the ideal solution.

4.2 Objective 1: Can participant state (MI vs rest) be accurately predicted from beta power in sensorimotor cortex after training?

The simple models generated using two sources were unable to reliably predict participant state, but this does present several important questions. Is it important to be able to predict participant state from the signals used to provide neurofeedback? Is the degree to which imagery can be discriminated from rest indicative of imagery

quality? Is it possible feedback based on less discriminative signals could lead to greater improvements in functional outcome?

If we could be certain that improving beta desynchronization at the site of maximal activity after median nerve stimulation was the best way to improve functional outcomes post-stroke, the fact that it couldn't reliably be used to predict whether a participant was doing imagery would not be relevant. Either reliable discrimination would come with increased dosage, or discriminability would be unrelated to functional outcomes and would never be obtained. It would seem unlikely, though, that feedback from a signal which was often indiscriminable from resting would be useful for training a participant to improve. Discriminability here is defined as the ability to predict the state the participant was instructed to be in, and not necessarily the state they actually reached. It is possible predictive failures are a function of unsuccessful task performance, rather than a failure of the features to discriminate state. If the task is sufficiently difficult, a feedback signal which frequently but correctly reports failure is still valuable. We are forced to rely on subjective reports by participants, and the two previous sessions of training they received, in asserting that they were capable of performing MI when instructed. As Boe et al [1] found significant changes in the patterns of activity evoked when compared to the first session when providing neurofeedback from these sources, discriminability is clearly not required for driving changes in activity with feedback. As we have not tested the effects of provision of feedback based on any of the models we have generated, it remains to be seen whether feedback based on signals with greater discriminability would yield better results. It seems reasonable to hypothesize that signals which are most discriminative might yield the best feedback, but it remains an open question, and one with important implications for the value of our approach.

4.3 Objective 2: Can we identify which MEG sensors are most discriminative between rest and MI?

Identifying a minimal subset of sensor positions which best capture the signals that characterize MI is a key step in realizing a clinically practical aid to MI interventions. The size and complexity of the proposed EEG system, and therefore its setup time, can be dramatically reduced by minimizing the number of sensors required. Further,

this system could be applied to stroke patients to automatically identify new optimal sensor placement sites to account for changes in the topography of the MI signals of interest after cortical damage.

4.3.1 Most Discriminative Sensors

While there was considerable variability in which sensors were selected in the top ten, a smaller set of 3-5 sensors was consistently selected across almost all participants. Given the high accuracy of this reduced sensor set, this suggests it may be worth investigating whether a very reduced set of 5 sensors may be all that is required to build models which provide effective neurofeedback. Sensors along both sides of the midline over sensorimotor cortex and sensors over both lateral extremes of prefrontal cortex were consistently selected. The middle frontal, precentral, and inferior frontal gyri as well as supplementary motor areas have been found to consistently activate during upper limb MI [80], and these are the most likely anatomical sources for the signals these sensors are most sensitive to.

4.3.2 Case Study: Participant 3

Participant 3 showed a number of anomalous traits which made their data worth investigating in greater detail. Models trained on other participants demonstrated poor accuracy when classifying participant 3's imagery, but improved when trained after occipital sensors had been eliminated. Further, models trained after RSE were completely unable to classify participant 3, with accuracy approaching chance. In order to ascertain why this happened, deviation from the training mean in heavily weighted features was used to identify in which sensors and frequency bands this participant differed most from the rest of our sample. By identifying the mean change in brain activity during MI over the participants the model trained on, and comparing it to the pattern of activity actually displayed in the testing set drawn from participant 3, we may be able to identify misleading or spurious features to omit in future.

The deviation from the training mean in each feature was computed and multiplied by the weight assigned by the model to that feature. The five sensors which displayed the largest weighted deviation in mu and beta features were then selected for more in depth analysis, plotting mean feature values over training trials with confidence

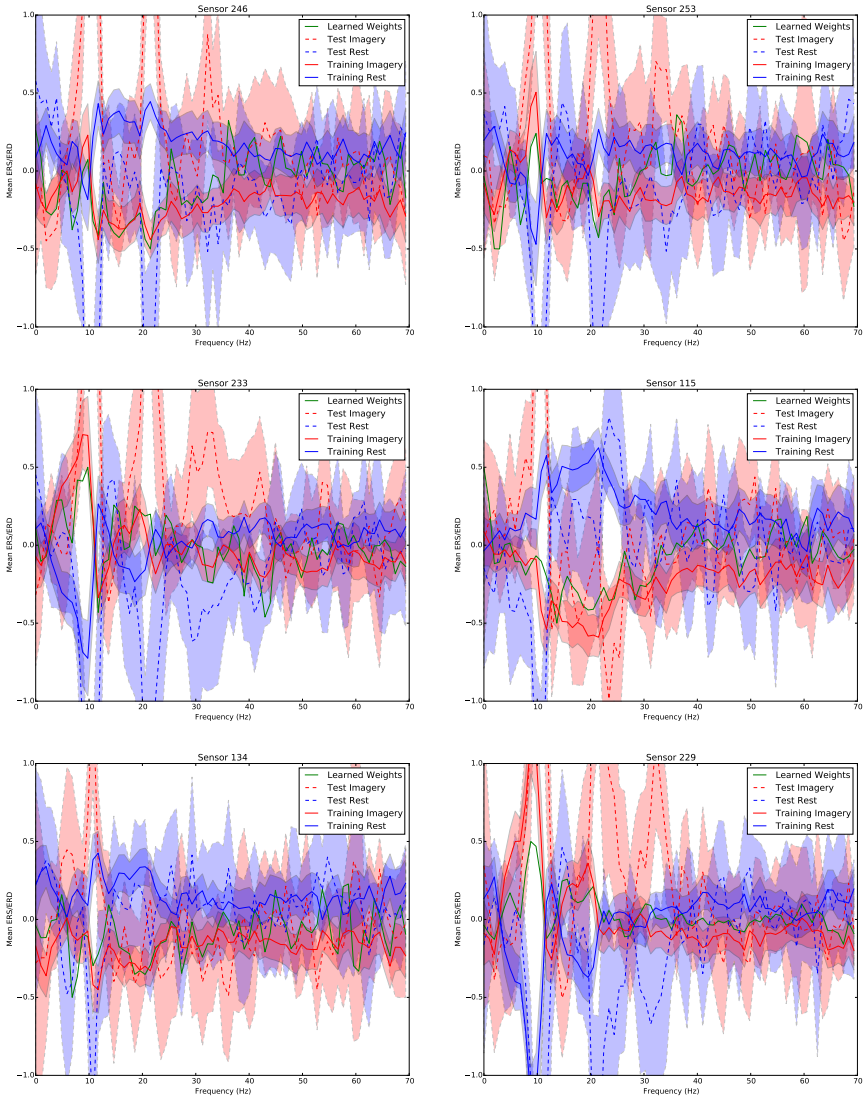

Figure 4.1: The six sensors drawn from the occipital included analysis in which participant 3 deviated most from the training mean. Shaded areas are 95% confidence intervals for the mean. The green line represents the model weights learned on the training data, and in general follow the sign of the solid red line representing mean Imagery activity in the training set. The dotted lines represent the individual testing participant's mean. Areas where the red dotted line and the green line are of opposing sign contribute to misclassification. Participant 3 demonstrated an unexpected large amplitude beta synchronization in the 20-22 Hz band in sensors 246, 253, and 233 during imagery. Unexpected Mu synchronization during imagery was observed in sensor 134 at 11 Hz.

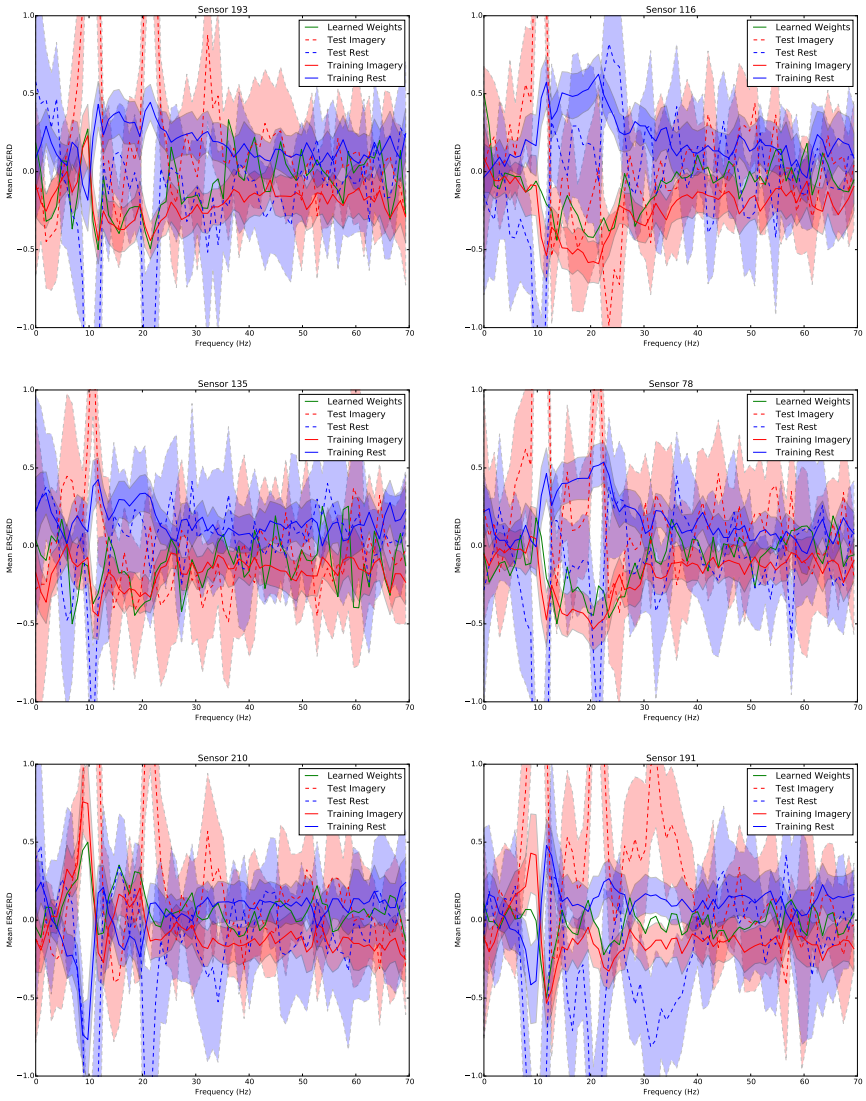

Figure 4.2: The six sensors in which participant 3 deviated most from the training mean. Shaded areas are 95% confidence intervals for the mean. The green line represents the model weights learned on the training data. Sensors 210, 193, 191, and 78, all show unexpected beta synchronization during imagery. In sensor 191, low gamma activity appears highly predictive of MI, though on average these features were uninformative. In sensors 78, 116, and 135 there was strong alpha synchronization during imagery, contrary to the strong desynchronization observed on average.

intervals, mean feature values over test trials with confidence intervals, and finally the weights learned on the training set for imagery including occipital (Fig. 4.1) and excluding occipital (Fig. 4.2).

Participant 3 demonstrated an unexpected large amplitude beta synchronization in the 20-22 Hz band in sensors 246, 253, and 233 during imagery. Unexpected Mu synchronization during imagery was observed in sensor 134 at 11 Hz. It is difficult to determine why participant 3 displayed this reversed pattern of activity, but it seems likely they were not performing the task correctly. The reversal of alpha activity could be explained by confusion with regard to when to keep eyes closed versus open, but this doesn't explain the reversed pattern of beta synchronization. If they were simply performing the task during rest trials and resting during active trials we would expect to see an accuracy well below 0.5. As accuracy was close to chance, it seems the activity patterns displayed by other participants were in no way predictive of state for participant 3.

4.3.3 Towards Clinical Applicability for MI

Practicing physiotherapists often see patients for 20-40 minutes per session. This makes minimization of setup time a critical factor in clinical applicability. Two potentially valuable goals of this research were reducing the number of electrodes and minimizing training time, as the process of EEG cap preparation and initial BCI familiarization form the bulk of preparation time.

We found we could successfully use our model to dramatically reduce the number of sensors with only small losses in accuracy.

LOBO cross validation is both a more strict standard for validating generated models and a means of estimating real-world transfer loss in feedback quality. The generally high accuracy of these models on unseen participants suggests a zero BCI training system would be practical for providing MI neurofeedback. Thought it was not explored here, it may be further improved by online adaptation of the model to the participant during initial trials, though this does pose the risk of learning to reinforce suboptimal imagery activity patterns.

4.4 Objective 3: Using source level modelling of MEG data across the whole brain, can we identify the anatomical regions whose activity best discriminates MI across our participants?

While source localization quickly becomes impractical with a reduced sensor set, source level models can improve our understanding of the the functional roles of the signals we employ for feedback, and thus the potential implications of the changes we drive for recovery. A solid understanding of what neurofeedback changes, the nature and function of the patterns of activity which are evoked by ME and MI, and the changes in activity patterns following stroke are all critical to designing efficacious therapeutic interventions based around MI.

4.4.1 Most Discriminative Sources

When primary visual areas were included they consistently ranked amongst the highest based on reliable changes in alpha synchronization during imagery, but this was confounded by the fact that participants had their eyes open during Rest trials. Beta ERD was most weighted in primary sensory and motor, ventrolateral and dorsolateral premotor, and the superior and inferior parietal cortices, which agrees well with the established key sources in the sensorimotor network. Primary sensory regions and sometimes superior parietal cortex weighted ERD heavily in mu. These source rankings agree well with the meta activation likelihood estimation of 75 fMRI MI studies performed by Hetu et al [80], both showing ventrolateral and dorsolateral premotor cortex, superior parietal cortex, and the precuneus. The dorsolateral prefrontal cortex, which was amongst the top rated sources in some models, maps on to the middle frontal gyrus in their work, though they found that this area was predominantly activated by lower limb MI on the right hemisphere and upper limb MI on the left. This may have been an effect of task laterality in the studies they evaluated, as we found right hemisphere dorsolateral prefrontal cortex activation in this left upper limb imagery task after feedback.

4.4.2 Case Study: Participant 4

In order to ascertain what caused the poor accuracy when classifying Participant 4 in both the Imagery and Motor tasks each trial was multiplied by the learned model weights and the deviation from the training mean was computed to ascertain in which sources and frequency bands this participant differed most. These sources were then selected for more in depth analysis, plotting mean feature values over training trials with confidence intervals, mean feature values over test trials with confidence intervals, and finally the weights learned on the training set in motor (Fig. 4.4) and imagery (Fig. 4.3), respectively. Alpha activity in Thalamus, Ventral Temporal Cortex, Angular Gyrus, and Inferior Parietal Cortex across both imagery and motor tasks was highly misleading in participant 4. Where most participants showed alpha synchronization in these areas during both imagery and motor tasks, participant 4 showed desynchronization. For instance, in the Thalamus in training data there was a mean peak at 11 Hz of 0.6 during ME, while in the testing data there was a mean peak at 10 Hz of -0.5. Similarly, in the imagery data the Thalamus showed a mean peak at 10 Hz in the training set of 0.7, while in testing it showed a peak at 9 Hz of -0.6. It seems likely this participant reversed the instructions for when to open and close their eyes, leaving them open during task and closed during rest. If this is true, it suggests the alpha trends in thalamus, precuneus, ventral temporal cortex, and aubgenial cingulate cortex, among others, are a result of eyes open/closed and not directly related to imagery/rest. A robust clinical system for identifying MI may benefit from resilience to noise introduced by whether the patient opts to perform the task with eyes open or closed. Weights learned on this data set could not account for this, but with a large enough data set, if many participants displayed this pattern of activity, the model would learn these were unreliable indicators and reduce their weighting. Outliers in important features are a major challenge to generalizability and transferability of a model, and the difficulties inherent in verifying compliance in MI tasks for labelling purposes compounds this.

There appears to be conflicting evidence related to alpha lateralization. When occipital sources are removed there is a reliable lateralization of alpha weights to the left hemisphere in the control group and the right hemisphere in the feedback group, but as the case study on participant 4 demonstrated, it's possible these signal

components are generated by whether eyes are open or closed during a trial. If this is an effect of exogenous activation of the visual system, and not imagery, why should neurofeedback affect its lateralization? It seems there may be more than one effect entangled here, and a methodology which more carefully controls for the state of the visual system would be required to separate them.

4.4.3 Comparison with Functional Connectivity Networks

Functional connectivity network analysis is a powerful technique for establishing the role that information transfer between anatomical regions plays in task specific processing. As our technique has generated a task-related source ranking, this is an ideal opportunity to compare our results with several alternate techniques which produce similar results. Bardouille & Boe examined cortico-cortico coherence, a measure of task induced phase-locking and information transfer, to generate task-related functional connectivity maps for bilateral ME vs rest [71]. They found bilateral ventrolateral and dorsolateral premotor cortex showed a high degree of connectivity with other regions, as well as the Frontal Eye Fields (labelled Anterior Premotor Cortex here), Inferior Parietal Cortex, Middle Temporal Cortex, and Precuneus. These sources were all consistently weighted heavily, though they often found the frequency bands of interest were in the gamma band (32-64 Hz) or, in the case of Ventrolateral Premotor Cortex, the delta band (4-8 Hz). These frequency bands were available as features to our models, but were less consistently and heavily weighted than alpha/mu and beta. This may indicate the phase locking of these regions, and not ERS/ERD carries most of their information content, and suggests a possible new set of features with which to augment our models.

Xu et al used the node betweenness centrality metric and fMRI to investigate the differences in the functional connectivity networks of MI and ME [81]. They found supplementary motor area was the key node in the ME network, while the right primary motor cortex and left thalamus were the key nodes in the MI network. That M1 was not a key node in the ME network is unexpected; that it plays a larger role in MI than ME doubly so. This counterintuitive result can also be seen in the weightings assigned to the M1, ventrolateral premotor, anterior premotor, and thalamic sources in our motor and imagery models. M1 was more important for predicting state - or

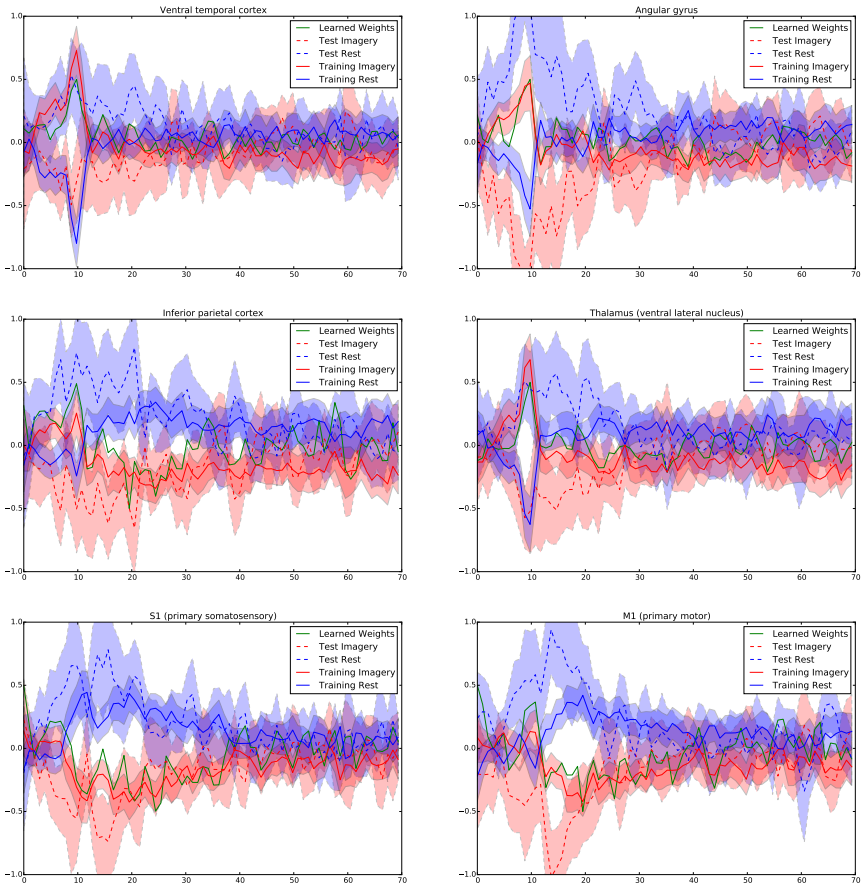

Figure 4.3: The six sources in which participant 3 deviated most from the training mean. Shaded areas are 95% confidence intervals for the mean. The green line represents the model weights learned on the training data, and in general follow the sign of the red line representing Imagery. The dotted lines represent the individual testing participant's mean. Areas where the red dotted line and the green line are of opposing sign contribute to misclassification. This participant shows a pattern of activity in thalamus, ventral temporal cortex, and angular gyrus roughly inverted from the training mean in the alpha band. M1 shows a possible example of overfitting in its weighting of alpha. The weight assigned here despite the overlapping confidence intervals suggests one participant in the training group showed alpha synchronization, and the model adapted to account for this and optimize classification accuracy. In most participants, where there was little to no ERS/ERD here, this had no effect. In this unseen participant, however, it contributed to misclassification.

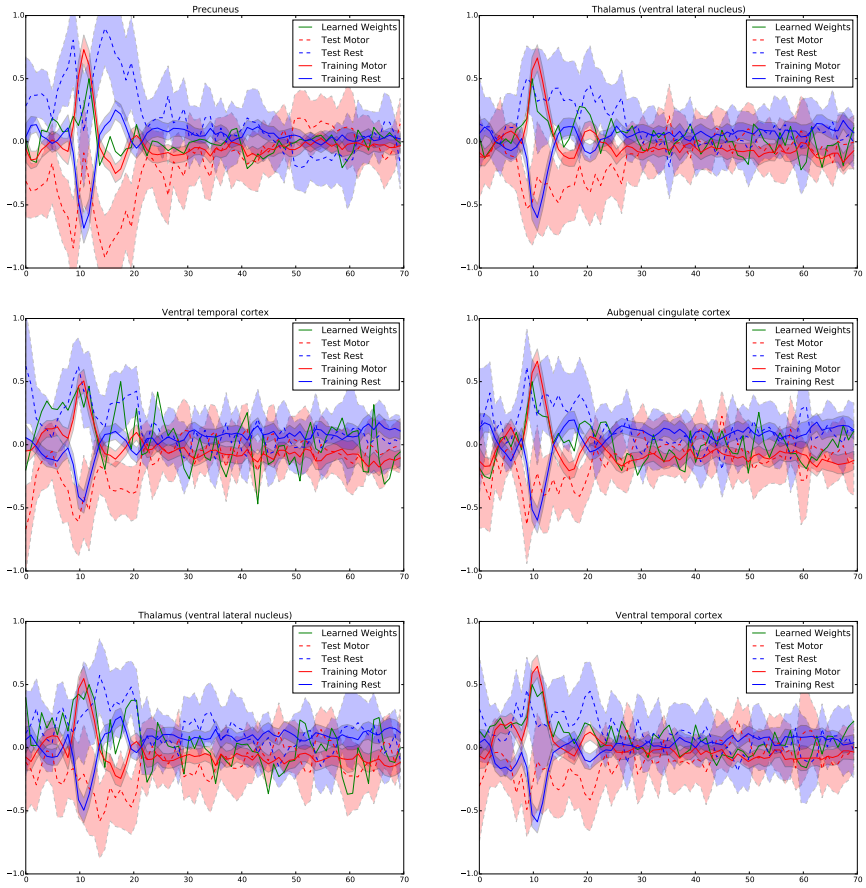

Figure 4.4: The six sources in which participant 4 deviated most from the training mean. Shaded areas are 95% confidence intervals for the mean. The green line represents the model weights learned on the training data, and in general follow the sign of the red line representing motor. The dotted lines represent the individual testing participant's mean. Areas where the red dotted line and the green line are of opposing sign contribute to misclassification. This participant shows a pattern of activity in thalamus, ventral temporal cortex, and the subgenual cingulate cortex in alpha which is roughly the inverse of the training mean.

perhaps other regions less important - in MI after neurofeedback than in ME.

Bajaj et al. [82] explored functional connectivity in the sensorimotor network at frequencies less than 0.1 Hz using spectral granger causality and fMRI. They examined the effects of a course of MI therapy without feedback and MI+RTP in patients between 1 and 54 months post stroke, as well as a healthy control group who received no intervention. Exclusion criteria included a motor activity log score of over 2.5, the inability to extend the affected wrist at least 20°, or the inability to extend two fingers and thumb at least 10°. MI therapy patients performed 8, 30 minute imagery sessions a day guided by an audio tape for a total dosage of 60h over 3 weeks. The MI+RTP group received 3 hours of physical therapy and an hour of MI therapy for 15 days. They found significant impairment in information flow between bilateral M1, premotor cortex, and SMA in the patient group as compared to the control group, with significant improvement in the MI+RTP group but not the MI alone group. This improvement in information flow was correlated with improvement on the Fugl-Meyer Assessment in the MI+RTP group. We would hypothesize the MI group in this study would have shown markedly better improvement if provided neurofeedback while practicing. Providing such large dosages of neurofeedback would require the small portable neurofeedback system previously discussed.

These directional relations between network regions create a complex fingerprint for each state, an information rich representation which could markedly improve separability over largely mutually independent ERS/ERD source models. Brain regions do not operate in isolation, and the coordination of a full network is required for full functionality. We effectively use ERS/ERD as a proxy measure for network activation, but information theoretic measures like cortico-cortico coherence, spectral granger causality, transfer entropy, and conditional mutual information could be used to quantify information flow within the network directly. Further study is required to determine whether reinforcing typical patterns of information flow within the network is more directly correlated with recovery than strengthening ERS/ERD. A feedback methodology based on both would be complex, but tractable with machine learning. Bivariate models, models which examine information flow between pairs of sensors, have been shown highly effective in the complex, network-centric field of seizure prediction where univariate models and traditional statistical models

have proven ineffective [83].

4.5 Future Work

Models built to discriminate first session and third, or later, session trials could directly speak to the axis along which the brain activity of participants is progressing when presented with MI neurofeedback. This could be used to directly quantify progress, allowing neurofeedback efficacy curves to be built for each participant, helping to understand dosage effects and inter-session rest and reinforcement effects on MI ability. By identifying whether there are diminishing returns or a saturation effect of dosage over time we can avoid wasting the time of patients and clinicians. If no such effects are found, we can quantify the tradeoff between session lengths and effect, allowing clinicians to make informed decisions about how to prescribe neurofeedback in practice. These models could also be used to evaluate whether a participant was progressing particularly slowly, potentially signifying damage which precludes MI interventions.

The recursive sensor elimination models should be tested in EEG, and the value of such a sensor selection methodology evaluated in stroke patients with varying extents of damage. By mapping lesion locations and identifying how the sensor selection system adapts, and how this influences outcomes from the resulting neurofeedback, we can determine whether it is beneficial to adapt the feedback to account for the damage. It's possible it would be more beneficial to reinforce limited activity in the damaged tissue, and that encouraging a shift to nearby cortical regions may not result in optimal outcomes. This may heavily depend on the time since stroke. Early in the acute phase driving activity in the damaged area may cause further damage, whereas at the height of plasticity during the 2-8 week period it may encourage recovery of lost functionality. Late in recovery, when plasticity is no longer heightened, it may be impractical to target the damaged regions if they have still not recovered their functional role, and better functional outcomes may be obtained by settling for encouraging those important functions to transfer to adjacent healthy regions.

Augmented with patient outcomes on tests like the Fugl-Meyer, Action Research Arm Test, or similar, a data set similar to this could be used to build an effective patient outcome prediction system. Dong et al [20] explored whether laterality index

could be used to predict patient outcomes using fMRI of MI. A very rich data set could be gathered with MEG from patients prior to beginning a course of standard physiotherapy, including ERS/ERD and information flow in the motor network. If followed up with post-therapy outcome measures, an effective clinical tool could be built to inform clinicians of expected outcomes for patients, allowing them to focus additional effort on patients who might otherwise have a poor prognosis, or attempt alternative therapies.

The structure of neurophysiological data obtained from MEG and EEG lends itself well to several other learning algorithms. Conceptualized as a machine vision problem, with time-frequency response functions forming an image and each sensor acting as a colour layer, the convolutional neural network (convnet) becomes an excellent choice for classifying state. Convnets have achieved state of the art results in many difficult image classification problems [84]. Convnets take advantage of structural assumptions inherent in images, using locally connected neurons in a spatial region instead of layers of totally connected neurons which would generate far too many connections to scale to high resolution data and likely result in overfitting. The kernels used in early layers can combine or leave separate colour channels as necessary. These assumptions hold well in the case of time-frequency response functions, where the two dimensional structure of the data holds rich information, and the manner in which different sensors should be combined may be complex.

Conceptualized as a multi-channel sound recognition problem, with sequences of power spectra from multiple sources arrayed in time as in speech processing, Long Short-Term Memory (LSTM) becomes an excellent choice for capturing these complex temporal sequences. LSTM is a form a recurrent neural network, a technique whereby the output of a layer of neurons is fed as input into itself iteratively to create depth while keeping the number of connections to learn from increasing, allowing very deep hierarchies with low training costs. Graves et al [85] achieved state of the art classification accuracies on the TIMIT speech recognition data set.

These hierarchical, deep learning approaches to feature extraction and classification are robust to noise, generalize well, and can learn new features which are themselves complex conditional combinations of simpler features. Using these learned encodings offers excellent classification accuracy, but can make the model difficult to

interpret. Most immediately accessible are the key features learned, which for images are extremely similar to those encoded in primary visual areas of the primate neocortex. By examining which signal components - potentially in time-frequency, information flow, or other spaces - are captured in these more complex features we might learn the fundamental building blocks of the task-related signal in MEG.

4.6 Conclusion

We investigated which spatiotemporal signal components best discriminate MI from rest at the source and sensor level across participants. Sensor and 80-Source level models were successfully able to discriminate MI vs Rest in unseen participants with an average of greater than 90% accuracy. A likely minimal set of sensor placements was identified, and shown to be capable of classifying MI vs Rest with only a small loss in accuracy. 80-Source level models weighted changes in beta synchronization in bilateral ventrolateral premotor cortex, dorsolateral premotor cortex, anterior premotor cortex, S1, M1, superior parietal cortex, and inferior parietal cortex, along with ipsilateral hemisphere S2. M1 was much more important to classifying MI after feedback. Alpha/mu weightings were more difficult to interpret due to the confounding state related activation of the visual system. The model weighted changes in alpha synchronization in ventral temporal cortex, parahippocampal cortex, retrosplenial cingulate cortex, precuneus, thalamus, and subgenual cingulate cortex. The Control group models heavily favoured ipsilateral hemisphere sources and sensors in both bands, while Feedback group models demonstrated a partial but marked shift to contralateral sources and sensors. Further work is still required to validate the results of this methodology in improving participant outcomes, but it provides avenues for further study in both MEG and EEG, and helps to lay the groundwork for a future clinically useful neurofeedback tool.

Bibliography

[1] Shaun Boe, Alicia Gionfriddo, Sarah Kraeutner, Antoine Tremblay, Graham Little, and Timothy Bardouille. Laterality of brain activity during motor imagery is modulated by the provision of source level neurofeedback. *NeuroImage*, 101C:159–167, July 2014.

[2] Sarah Kraeutner, Alicia Gionfriddo, Timothy Bardouille, and Shaun Boe. Motor imagery-based brain activity parallels that of motor execution: Evidence from magnetic source imaging of cortical oscillations. *Brain Research*, 1588:81–91, 2014.

[3] Health Canada. Tracking Heart Disease and Stroke in Canada. (http://www.phac-aspc.gc.ca/publicat/2009/cvd-avc/index-eng.php), 2009.

[4] Randolph J Nudo. Neural bases of recovery after brain injury. *Journal of communication disorders*, 44(5):515–20, 2011.

[5] Jeffrey a Kleim and Theresa a Jones. Principles of experience-dependent neural plasticity: implications for rehabilitation after brain damage. *Journal of speech, language, and hearing research : JSLHR*, 51(1):S225–39, February 2008.

[6] R. P. Stroemer, T. A. Kent, and C. E. Hulsebosch. Neocortical Neural Sprouting, Synaptogenesis, and Behavioral Recovery After Neocortical Infarction in Rats. *Stroke*, 26(11):2135–2144, November 1995.

[7] P Glees and J Cole. Recovery of skilled motor functions after small repeated leswions in motor cortex in macaque. *J Neurophysiol.*, 13(2):137–148, 1950.

[8] R J Nudo and G W Milliken. Reorganization of movement representations in primary motor cortex following focal ischemic infarcts in adult squirrel monkeys. *Journal of neurophysiology*, 75(5):2144–9, May 1996.

[9] R J Nudo, B M Wise, F SiFuentes, and G W Milliken. Neural substrates for the effects of rehabilitative training on motor recovery after ischemic infarct. *Science (New York, N.Y.)*, 272(5269):1791–4, June 1996.

[10] R J Nudo, G W Milliken, W M Jenkins, and M M Merzenich. Use-dependent alterations of movement representations in primary motor cortex of adult squirrel monkeys. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 16(2):785–807, January 1996.

[11] Christian Xerri, M M Merzenich, B E Peterson, and W Jenkins. Plasticity of primary somatosensory cortex paralleling sensorimotor skill recovery from stroke in adult monkeys. *Journal of neurophysiology*, 79(4):2119–48, April 1998.

[12] Gary W Thickbroom, Michelle L Byrnes, Sarah a Archer, and Frank L Mastaglia. Motor outcome after subcortical stroke correlates with the degree of cortical reorganization. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 115(9):2144–50, September 2004.

[13] Randolph J Nudo. Postinfarct cortical plasticity and behavioral recovery. *Stroke; a journal of cerebral circulation*, 38(2 Suppl):840–5, February 2007.

[14] Scott Barbay, David J Guggenmos, Mariko Nishibe, and Randolph J Nudo. Motor representations in the intact hemisphere of the rat are reduced after repetitive training of the impaired forelimb. *Neurorehabilitation and neural repair*, 27(4):381–4, May 2013.

[15] D Dawson, J Knox, A McClure, Foley N, and R Teasell. Canadian Stroke Best Practice Guidelines, 2013.

[16] Robert Teasell, Norine Foley, Katherine Salter, Sanjit Bhogal, Jeffrey Jutai, and Mark Speechley. Evidence-Based Review of Stroke Rehabilitation. *Topics in stroke rehabilitation*, 16:463–488, 2012.

[17] Julie Duque, Friedhelm Hummel, Pablo Celnik, Nagako Murase, Riccardo Mazzocchio, and Leonardo G. Cohen. Transcallosal inhibition in chronic subcortical stroke. *NeuroImage*, 28(4):940–946, 2005.

[18] Nagako Murase, Julie Duque, Riccardo Mazzocchio, and Leonardo G. Cohen. Influence of Interhemispheric Interactions on Motor Function in Chronic Stroke. *Annals of Neurology*, 55(3):400–409, 2004.

[19] Torunn Askim, Bent Indredavik, Torgil Vangberg, and Asta Hå berg. Motor network changes associated with successful motor skill relearning after acute ischemic stroke: a longitudinal functional magnetic resonance imaging study. *Neurorehabilitation and neural repair*, 23(3):295–304, 2008.

[20] Yun Dong, Bruce H. Dobkin, Steven Y. Cen, Allan D. Wu, and Carolee J. Winstein. Motor cortex activation during treatment may predict therapeutic gains in paretic hand function after stroke. *Stroke*, 37(6):1552–1555, 2006.

[21] J. Liepert, H. Bauder, H R Wolfgang, W H Miltner, E. Taub, and C. Weiller. Treatment-induced cortical reorganization after stroke in humans. *Stroke; a journal of cerebral circulation*, 31(6):1210–6, June 2000.

[22] Manuel Rodríguez, Ramón Muñiz, Belén González, and Magdalena Sabaté. Hand movement distribution in the motor cortex: the influence of a concurrent task and motor imagery. *NeuroImage*, 22(4):1480–91, August 2004.

[23] Franck Di Rienzo, Christian Collet, Nady Hoyek, and Aymeric Guillot. Impact of neurologic deficits on motor imagery: a systematic review of clinical evaluations. *Neuropsychology review*, 24(2):116–47, June 2014.

[24] Stephen J Page, Jerzy P Szaflarski, James C Eliassen, Hai Pan, and Steven C Cramer. Cortical plasticity following motor skill learning during mental practice in stroke. *Neurorehabilitation and neural repair*, 23(4):382–8, May 2009.

[25] A Pascual-Leone, D Nguyet, L G Cohen, J P Brasil-Neto, A Cammarota, and M Hallett. Modulation of muscle responses evoked by transcranial magnetic stimulation during the acquisition of new fine motor skills. *Journal of neurophysiology*, 74(3):1037–45, September 1995.

[26] Francine Malouin, Philip L Jackson, and Carol L Richards. Towards the integration of mental practice in rehabilitation programs. A critical review. *Frontiers in human neuroscience*, 7(September):576, January 2013.

[27] Alessio Faralli, Matteo Bigoni, Alessandro Mauro, Ferdinando Rossi, and Daniela Carulli. Noninvasive strategies to promote functional recovery after stroke. *Neural plasticity*, 2013:854597, January 2013.

[28] Susy M Braun, Anna J Beurskens, Paul J Borm, Thomas Schack, and Derick T Wade. The effects of mental practice in stroke rehabilitation: a systematic review. *Archives of physical medicine and rehabilitation*, 87(6):842–52, June 2006.

[29] K M Newell. Motor skill acquisition. *Annual review of psychology*, 42(1):213–237, 1991.

[30] Benedikt Lauber and Martin Keller. Improving motor performance: Selected aspects of augmented feedback in exercise and health. *European Journal of Sport Science*, 14(1):36–43, 2012.

[31] Han-Jeong Hwang, Kiwoon Kwon, and Chang-Hwang Im. Neurofeedback-based motor imagery training for brain-computer interface (BCI). *Journal of neuroscience methods*, 179(1):150–6, April 2009.

[32] Esther Florin, Elizabeth Bock, and Sylvain Baillet. Targeted reinforcement of neural oscillatory activity with real-time neuroimaging feedback. *NeuroImage*, 88C:54–60, November 2013.

[33] Brian D Berman, Silvina G Horovitz, Gaurav Venkataraman, and Mark Hallett. Self-modulation of primary motor cortex activity with motor and motor imagery tasks using real-time fMRI-based neurofeedback. *NeuroImage*, 59(2):917–25, January 2012.

[34] Mark Chiew, Stephen M LaConte, and Simon J Graham. Investigation of fMRI neurofeedback of differential primary motor cortex activity using kinesthetic motor imagery. *NeuroImage*, 61(1):21–31, May 2012.

[35] Brittany M Young, Zack Nigogosyan, Léo M Walton, Jie Song, Veena a Nair, Scott W Grogan, Mitchell E Tyler, Dorothy F Edwards, Kristin Caldera, Justin a Sattin, Justin C Williams, and Vivek Prabhakaran. Changes in functional brain

organization and behavioral correlations after rehabilitative therapy using a brain-computer interface. *Frontiers in neuroengineering*, 7(July):26, January 2014.

[36] Floriana Pichiorri, Giovanni Morone, Manuela Petti, Jlenia Toppi, Iolanda Pisotta, Marco Molinari, Stefano Paolucci, Maurizio Inghilleri, Laura Astolfi, Febo Cincotti, and Donatella Mattia. Brain-computer interface boosts motor imagery practice during stroke recovery. *Annals of Neurology*, 77:851–856, 2015.

[37] Douglas Owen Cheyne. MEG studies of sensorimotor rhythms: a review. *Experimental neurology*, 245:27–39, July 2013.

[38] Bjørg Elisabeth Kilavik, Manuel Zaepffel, Andrea Brovelli, William a. MacKay, and Alexa Riehle. The ups and downs of beta oscillations in sensorimotor cortex. *Experimental Neurology*, 245:15–26, 2013.

[39] Michael T. Jurkiewicz, William C. Gaetz, Andreea C. Bostan, and Douglas Cheyne. Post-movement beta rebound is generated in motor cortex: Evidence from neuromagnetic recordings. *NeuroImage*, 32(3):1281–1289, 2006.

[40] Lewis Wheaton, Esteban Fridman, Stephan Bohlhalter, Sherry Vorbach, and Mark Hallett. Left parietal activation related to planning, executing and suppressing praxis hand movements. *Clinical Neurophysiology*, 120(5):980–986, 2009.

[41] Wolfgang Omlor, Luis Patino, Ignacio Mendez-Balbuena, Jürgen Schulte-Mönting, and Rumyana Kristeva. Corticospinal beta-range coherence is highly dependent on the pre-stationary motor state. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 31(22):8037–8045, 2011.

[42] Bjorg Elisabeth Kilavik, Adrián Ponce-Alvarez, Romain Trachel, Joachim Confais, Sylvain Takerkart, and Alexa Riehle. Context-related frequency modulations of macaque motor cortical LFP beta oscillations. *Cerebral Cortex*, 22(9):2148–2159, 2012.

[43] Kei Nakagawa, Yoriyuki Aokage, Takashi Fukuri, Yumi Kawahara, Akira Hashizume, Kaoru Kurisu, and Louis Yuge. Neuromagnetic beta oscillation changes during motor imagery and motor execution of skilled movements. *Neuroreport*, 22(5):217–22, March 2011.

[44] G Pfurtscheller, C Neuper, H Ramoser, and J Müller-Gerking. Visually guided motor imagery activates sensorimotor areas in humans. *Neuroscience letters*, 269(3):153–6, July 1999.

[45] Han Yuan, Tao Liu, Rebecca Szarkowski, Cristina Rios, James Ashe, and Bin He. Negative covariation between task-related responses in alpha/beta-band activity and BOLD in human sensorimotor cortex: An EEG and fMRI study of motor imagery and movements. *NeuroImage*, 49(3):2596–2606, 2010.

[46] Andreas K. Engel and Wolf Singer. Temporal binding and the neural correlates of sensory awareness, 2001.

[47] J F Marsden, K J Werhahn, P Ashby, J Rothwell, S Noachtar, and P Brown. Organization of cortical activities related to movement in humans. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 20(6):2307–2314, 2000.

[48] F Cassim, C Monaca, W Szurhaj, J L Bourriez, L Defebvre, P Derambure, and J D Guieu. Does post-movement beta synchronization reflect an idling motor cortex? *Neuroreport*, 12(17):3859–3863, 2001.

[49] Douglas Cheyne, Sonya Bells, Paul Ferrari, William Gaetz, and Andreea C. Bostan. Self-paced movements induce high-frequency gamma oscillations in primary motor cortex. *NeuroImage*, 42(1):332–342, 2008.

[50] Suresh D Muthukumaraswamy. Functional properties of human primary motor cortex gamma oscillations. *Journal of neurophysiology*, 104(5):2873–2885, 2010.

[51] Alice J O'Toole, Fang Jiang, Hervé Abdi, Nils Pénard, Joseph P Dunlop, and Marc a Parent. Theoretical, statistical, and practical perspectives on pattern-based classification approaches to the analysis of functional neuroimaging data. *Journal of cognitive neuroscience*, 19(11):1735–52, November 2007.

[52] Ron Kohavi. A Study of Cross-validation and Bootstrap for Accuracy Estimation and Model Selection. *IJCAI'95 Proceedings of the 14th international joint conference on Artificial intelligence*, 2:1137–1143, 1995.

[53] Roberto Santana, Concha Bielza, and Pedro Larrañaga. Regularized logistic regression and multiobjective variable selection for classifying MEG data. *Biological cybernetics*, 106(6-7):389–405, September 2012.

[54] Pawel Herman, Girijesh Prasad, Thomas Martin McGinnity, and Damien Coyle. Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification. *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, 16(4):317–26, August 2008.

[55] Niels Birbaumer, Ander Ramos Murguialday, Cornelia Weber, and Pedro Montoya. Neurofeedback and brain-computer interface clinical applications. *International review of neurobiology*, 86(09):107–17, January 2009.

[56] Chayanin Tangwiriyasakul, Victor Mocioiu, Michel J a M van Putten, and Wim L C Rutten. Classification of motor imagery performance in acute stroke. *Journal of neural engineering*, 11(3):036001, June 2014.

[57] Jingwei Yue, Zongtan Zhou, Jun Jiang, Yadong Liu, and Dewen Hu. Balancing a simulated inverted pendulum through motor imagery: an EEG-based real-time control paradigm. *Neuroscience letters*, 524(2):95–100, August 2012.

[58] C Cortes and V Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[59] Vladimir Vapnik, Esther Levin, and Yann Le Cun. Measuring the VC-Dimension of a Learning Machine. *Neural Computation*, 6(5):851–876, 1994.

[60] Weibo Yi, Shuang Qiu, Hongzhi Qi, Lixin Zhang, Baikun Wan, and Dong Ming. EEG feature comparison and classification of simple and compound limb motor imagery. *Journal of neuroengineering and rehabilitation*, 10(1):106, January 2013.

[61] Wing-Kin Tam, Kai-yu Tong, Fei Meng, and Shangkai Gao. A minimal set of electrodes for motor imagery BCI to control an assistive device in chronic stroke subjects: a multi-session study. *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, 19(6):617–27, December 2011.

[62] M Spüler, Wolfgang Rosenstiel, and Martin Bogdan. Adaptive SVM-based classification increases performance of a MEG-based Brain-Computer Interface (BCI). *Artificial Neural Networks and Machine Learning*, 7552:669–676, 2012.

[63] Thomas Navin Lal, Michael Schröder, Thilo Hinterberger, Jason Weston, Martin Bogdan, Niels Birbaumer, and Bernhard Schölkopf. Support vector channel selection in BCI. *IEEE Transactions on Biomedical Engineering*, 51(6):1003–1010, 2004.

[64] Pieter Jan Kindermans, Martijn Schreuder, Benjamin Schrauwen, Klaus Robert Müller, and Michael Tangermann. True zero-training brain-computer interfacing - An online study. *PLoS ONE*, 9(7), 2014.

[65] Florin Popescu, Siamac Fazli, Yakob Badower, Benjamin Blankertz, and Klaus R. Müller. Single trial classification of motor imagination using 6 dry EEG electrodes. *PLoS ONE*, 2(7):1–5, 2007.

[66] B.D. Josephson. Possible new effects in superconductive tunnelling. *Physics Letters*, 1(7):251–253, 1962.

[67] PW Anderson and JM Rowell. Probable observation of the Josephson superconducting tunneling effect. *Physical Review Letters*, 10(6):230–232, 1963.

[68] J.E. Mercereau. Superconducting magnetometers, 1970.

[69] Arjan Hillebrand, Krish D. Singh, Ian E. Holliday, Paul L. Furlong, and Gareth R. Barnes. A new approach to neuroimaging with magnetoencephalography, 2005.

[70] A. O. Diaconescu, C. Alain, and A. R. McIntosh. The co-occurrence of multisensory facilitation and cross-modal conflict in the human brain. *Journal of Neurophysiology*, 106(6):2896–2909, 2011.

[71] Timothy Bardouille and Shaun Boe. State-Related Changes in MEG Functional Connectivity Reveal the Task-Positive Sensorimotor Network. *PLoS ONE*, 7(10):e48682, January 2012.

[72] Karl Pearson. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2:559–572, 1901.

[73] Samu Taulu and Riitta Hari. Removal of magnetoencephalographic artifacts with temporal signal-space separation: demonstration with single-trial auditory-evoked responses. *Human Brain Mapping*, 30(5):1524–1534, 2009.

[74] Eric Jones, Travis Oliphant, Pearu Peterson, and Others. SciPy: Open source scientific tools for Python.

[75] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, and Matti Hämäläinen. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7(267), 2013.

[76] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A Engemann, Daniel Strohmeier, Christian Brodbeck, Lauri Parkkonen, and Matti S Hämäläinen. MNE software for processing MEG and EEG data. *NeuroImage*, 86:446–460, February 2014.

[77] John D. Hunter. Matplotlib: A 2D graphics environment. *Computing in Science and Engineering*, 9(3):99–104, 2007.

[78] Wes McKinney. Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 51–56, 2010.

[79] F Pedregosa, G Varoquaux, A Gramfort, V Michel, B Thirion, O Grisel, M Blondel, P Prettenhofer, R Weiss, V Dubourg, J Vanderplas, A Passos, D Cournapeau, M Brucher, M Perrot, and E Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[80] Sébastien Hétu, Mathieu Grégoire, Arnaud Saimpont, Michel Pierre Coll, Fanny Eugène, Pierre Emmanuel Michon, and Philip L. Jackson. The neural network of motor imagery: An ALE meta-analysis. *Neuroscience and Biobehavioral Reviews*, 37(5):930–949, 2013.

[81] L. Xu, H. Zhang, M. Hui, Z. Long, Z. Jin, Y. Liu, and L. Yao. Motor execution and motor imagery: A comparison of functional connectivity patterns based on graph theory. *Neuroscience*, 261:184–194, 2014.

[82] Sahil Bajaj, Andrew J. Butler, Daniel Drake, and Mukesh Dhamala. Functional organization and restoration of the brain motor-execution network after stroke and rehabilitation. *Frontiers in human neuroscience*, 9(March):173, 2015.

[83] Piotr Mirowski, Deepak Madhavan, Yann LeCun, and Ruben Kuzniecky. Classification of patterns of EEG synchronization for seizure prediction. *Clinical Neurophysiology*, 120(11):1927–1940, 2009.

[84] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In F Pereira, C J C Burges, L Bottou, and K Q Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[85] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. Speech recognition with deep recurrent neural networks. *arXiv*, 2013.

Appendix

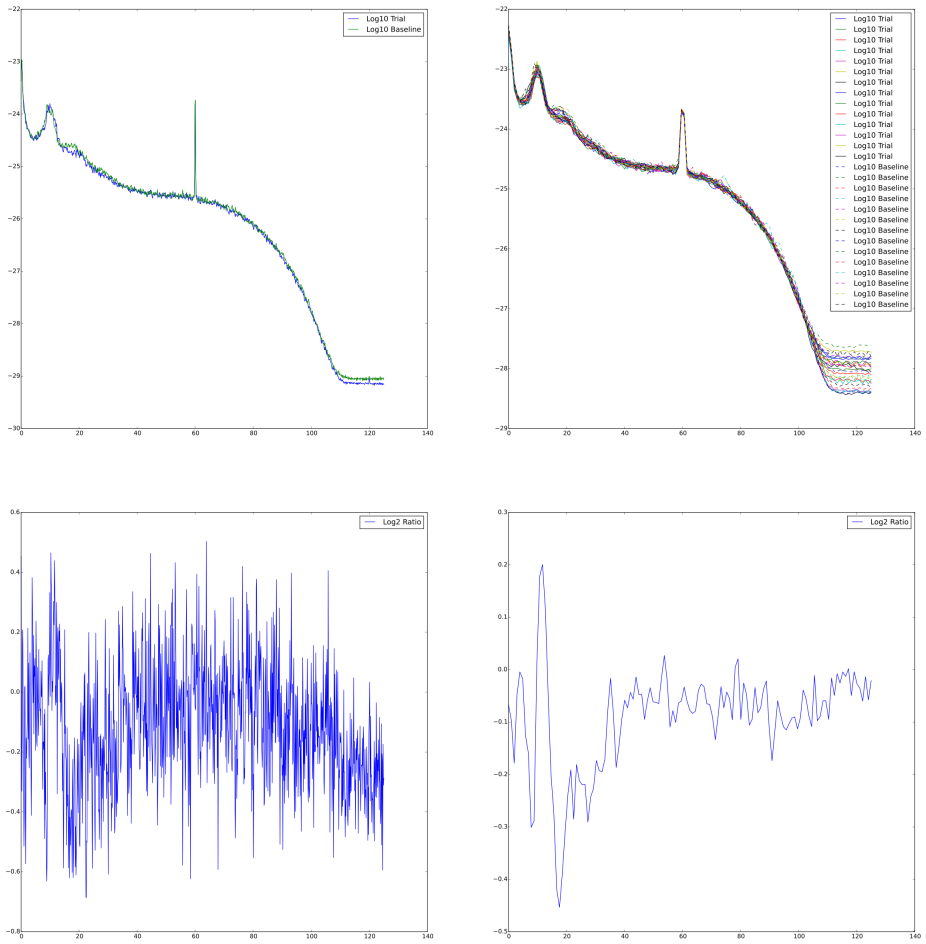

Figure 4.5: Signal from sensor 255 over right hemisphere M1 during ME processed with FFT and STFT demonstrating ratio to baseline signal.

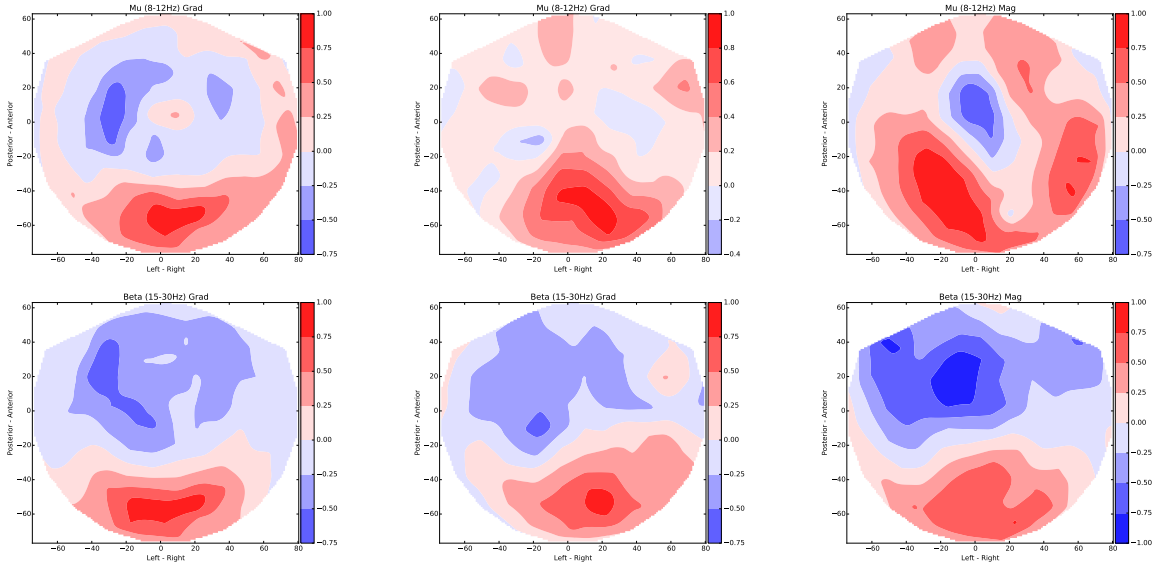

Figure 4.6: Mean weights of all sensor level models of Control group during MI, occipital sources included.

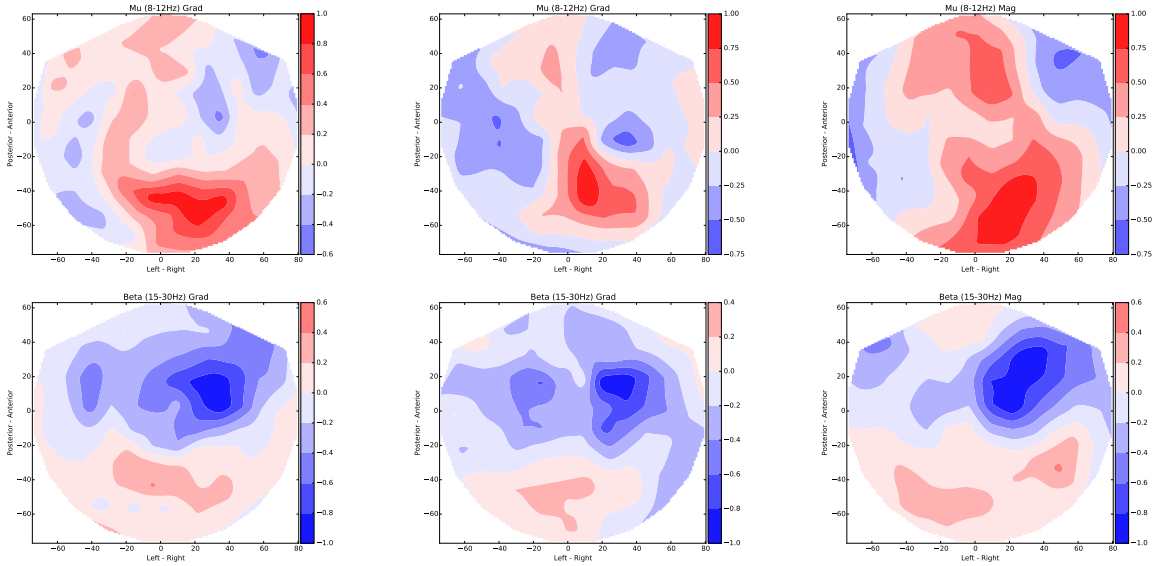

Figure 4.7: Mean weights of all sensor level models of Feedback group during MI, occipital sources included.

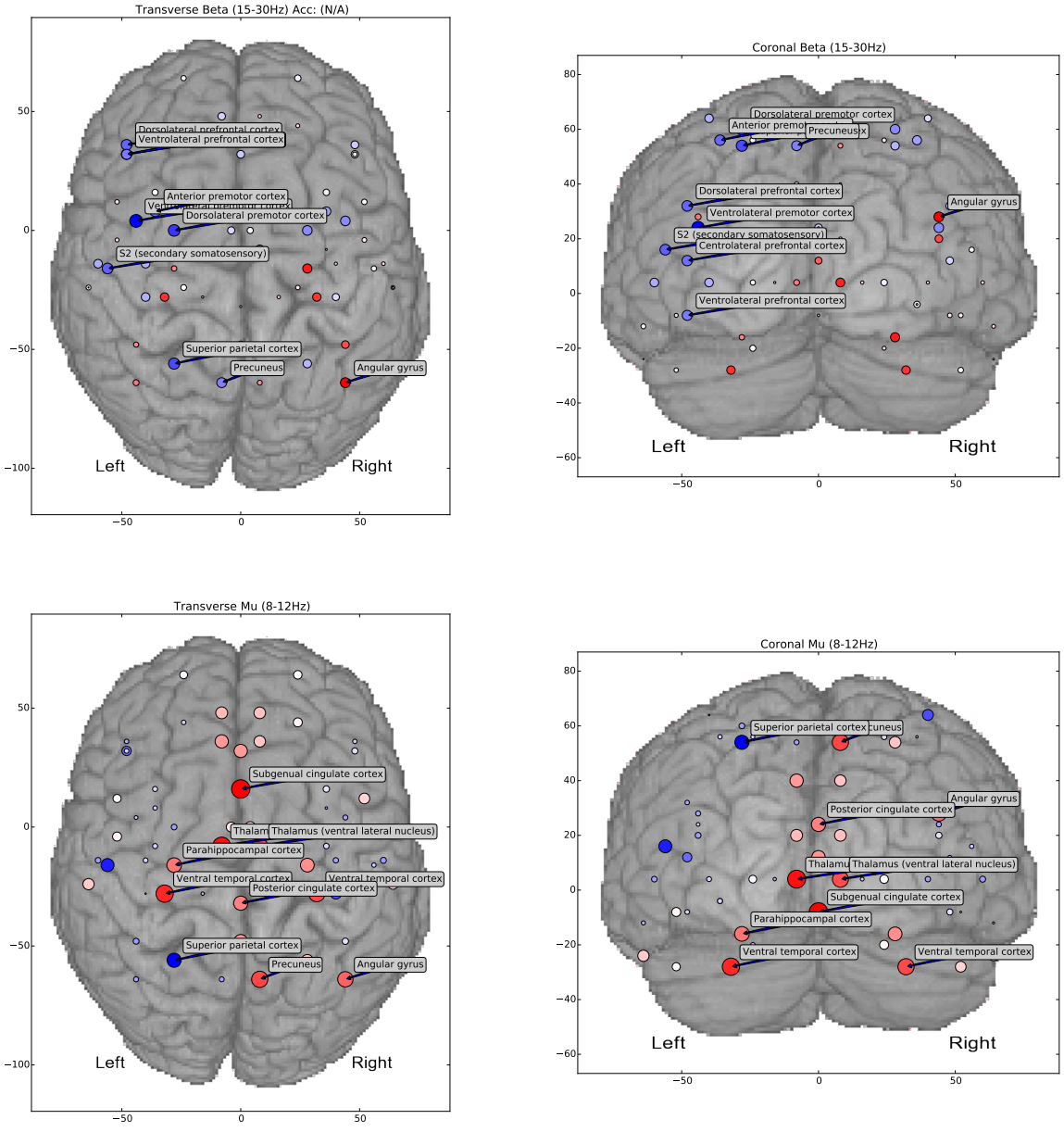

Figure 4.8: Mean weights of all source level models of Control group during MI, occipital sources excluded.

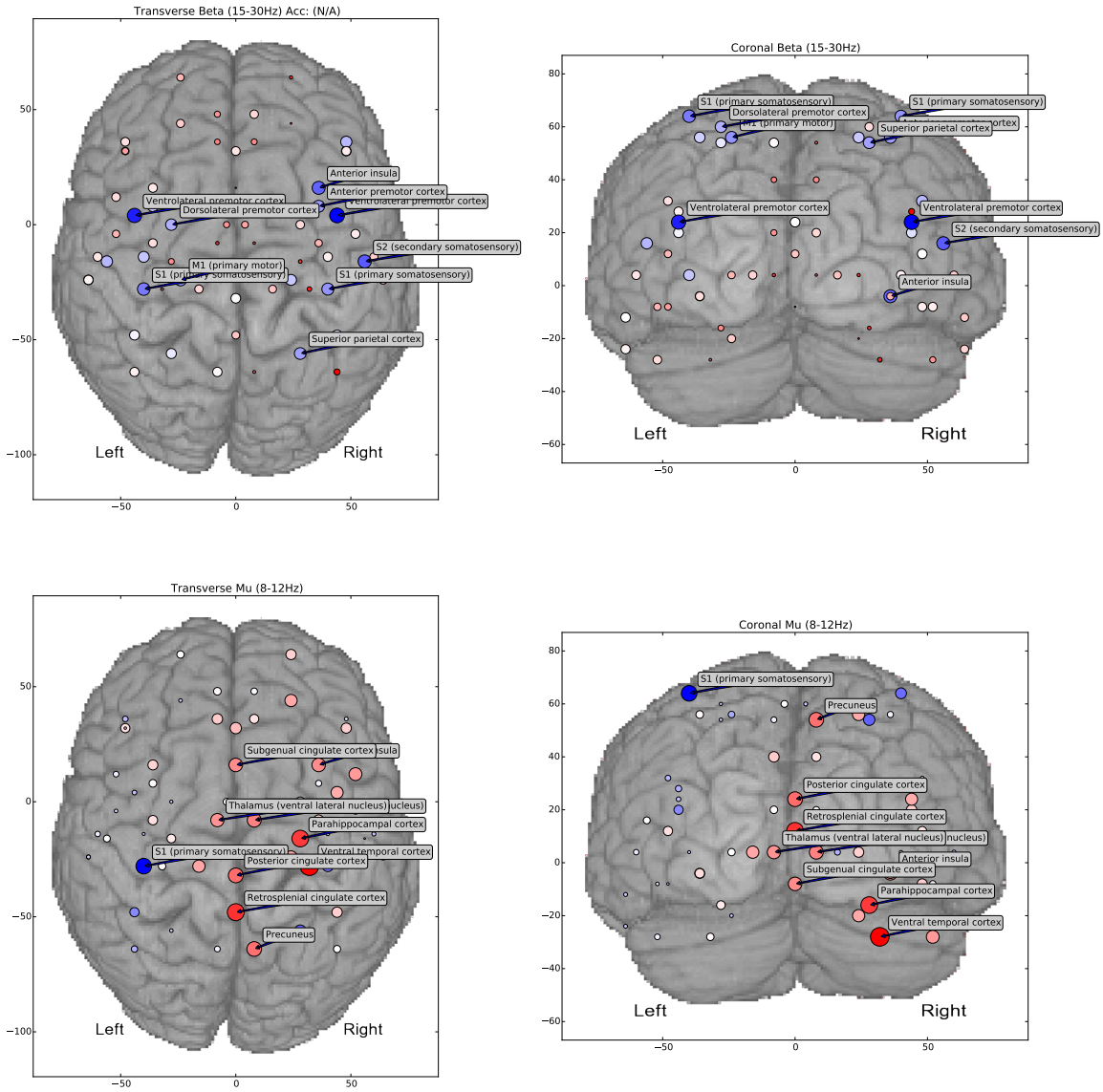

Figure 4.9: Mean weights of all source level models of Feedback group during MI, occipital sources excluded.

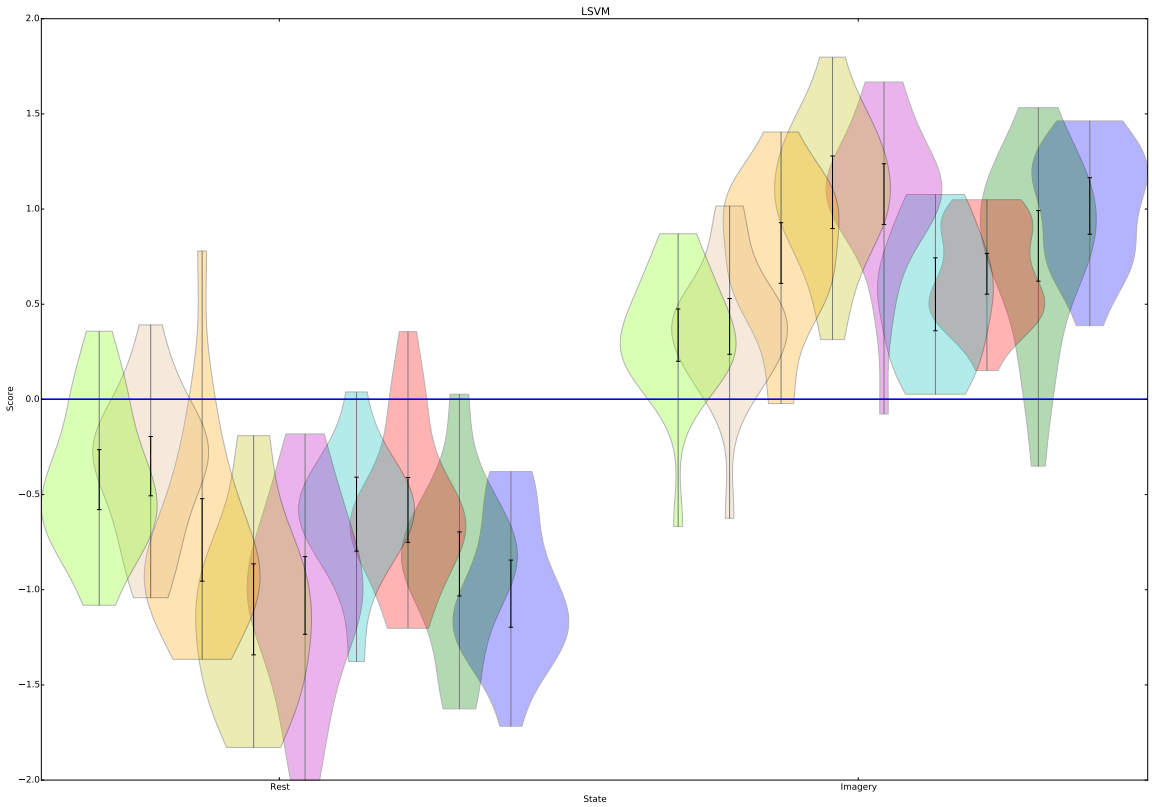

Figure 4.10: Probability density estimates of the feedback which the occipital excluded source model would have provided to each of the 9 participants in the Control group. Each coloured distribution represents one participant, with scores above the blue line representing a classification of imagery and below it a classification of rest. The true class is listed along the bottom axis.