2013

DEVELOPMENT OF AN EEG BRAIN-MACHINE INTERFACE TO AID IN RECOVERY OF MOTOR FUNCTION AFTER NEUROLOGICAL INJURY

Elizabeth Salmon
University of Kentucky, elizabeth.salmon@gmail.com

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained and attached hereto needed written permission statements(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine).

I hereby grant to The University of Kentucky and its agents the non-exclusive license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless a preapproved embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's dissertation including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Elizabeth Salmon, Student

Dr. Sridhar Sunderam, Major Professor

Dr. Abhijit Patwardhan, Director of Graduate Studies

DEVELOPMENT OF AN EEG BRAIN-MACHINE INTERFACE TO AID IN RECOVERY OF MOTOR FUNCTION AFTER NEUROLOGICAL INJURY

THESIS

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science in Biomedical Engineering
in the College of Engineering at the University of Kentucky

By

Elizabeth Ellene Salmon

Lexington, Kentucky

Director: Sridhar Sunderam, Ph.D., Assistant Professor of Biomedical Engineering

Lexington, Kentucky

2013

ABSTRACT OF THESIS

DEVELOPMENT OF AN EEG BRAIN-MACHINE INTERFACE TO AID IN RECOVERY OF MOTOR FUNCTION AFTER NEUROLOGICAL INJURY

Impaired motor function following neurological injury may be overcome through therapies that induce neuroplastic changes in the brain. Therapeutic methods include repetitive exercises that promote use-dependent plasticity (UDP), the benefit of which may be increased by first administering peripheral nerve stimulation (PNS) to activate afferent fibers, resulting in increased cortical excitability. We speculate that PNS delivered only in response to attempted movement would induce timing-dependent plasticity (TDP), a mechanism essential to normal motor learning. Here we develop a brain-machine interface (BMI) to detect movement intent and effort in healthy volunteers (n=5) from their electroencephalogram (EEG). This could be used in the future to promote TDP by triggering PNS in response to a patient's level of effort in a motor task. Linear classifiers were used to predict state (rest, sham, right, left) based on EEG variables in a handgrip task and to determine between three levels of force applied. Mean classification accuracy with out-of-sample data was 54% (23-73%) for tasks and 44% (21-65%) for force. There was a slight but significant correlation (p<0.001) between sample entropy and force exerted. The results indicate the feasibility of applying PNS in response to motor intent detected from the brain.

KEYWORDS: Brain-Machine Interface, EEG, Rehabilitation, Neuroplasticity, Linear Discriminant Analysis

Elizabeth Salmon

April 17, 2013

DEVELOPMENT OF AN EEG BRAIN-MACHINE INTERFACE
TO AID IN RECOVERY OF MOTOR FUNCTION AFTER
NEUROLOGICAL INJURY

By

Elizabeth Ellene Salmon

Sridhar Sunderam, Ph.D.

Director of Thesis

Abhijit Patwardhan, Ph.D.

Director of Graduate Studies

April 17, 2013

Date

ACKNOWLEDGMENTS

Though this thesis is an individual work, I am thankful for the many people who have helped me through the process:

My adviser, Dr. Sridhar Sunderam, who over the past three years has spent many hours with me, explaining concepts, discussing ideas, and serving as my first, and most frequent, subject.

Dr. Lumy Sawaki, one of my committee members, whose involvement really helped give purpose to this project far beyond simply developing a brain-machine interface for the heck of it. Her enthusiasm for rehabilitation is truly inspiring.

Dr. Guoqiang Yu, my other committee member, whose Biomedical Instrumentation lab gave me experience with LabVIEW that has proven invaluable to designing the data acquisition and feedback interfaces for this research.

My wonderful mom, who always had a home cooked meal waiting for me in the evenings, always listened to my rants, and knew when I needed brownies and cupcakes.

My dear boyfriend, Josh, who endured my grumpiness and gave me the encouragement I needed to make it through the past three years.

Last but not least, my grandfather, who gave me the nudge I needed to go back to school for my master's degree.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

1.1 Motivation

Neurological injuries, such as spinal cord injury (SCI) and stroke, affect many people every year throughout the United States and the world. As healthcare has improved, many more people suffering from these events survive, leaving the majority with a variety of disabilities for the remainder of their lives. How exactly an individual is disabled depends on the location and extent of damage to the spinal cord or brain. Perhaps the greatest impact is made on a person's life when the disability includes the loss of function of the upper limbs. In these cases, independent living is often difficult, as assistance may be required for simple, everyday tasks such as feeding oneself. Individuals with neurological injuries may undergo extensive rehabilitation, which can focus on a) how to perform tasks using any residual upper limb function or b) regaining lost function. Lost function can be regained due to the brain's plasticity, or the ability to change as a result of its use. Therefore, an increase in the brain's ability to undergo plastic changes further aids in the recovery of function.

Methods to enhance neuroplasticity are a current topic of research, with somatosensory peripheral nerve stimulation (PNS) being one such method. In a typical PNS protocol, a patient undergoes two hours of stimulation, followed by several hours of guided physical exercise. PNS parameters are tuned to preferentially stimulate sensory afferent nerves, causing a short-term increase in the excitability of the sensorimotor cortex, therefore increasing the potentially beneficial effects of the rehabilitative exercises performed in the hours following stimulation. Although this protocol has demonstrated significant benefits in recent clinical trials, we hypothesize that artificially closing the sensorimotor loop by delivering afferent sensory stimulation only in response to the patient's intent to move will be a more efficient way of upregulating the brain's plasticity.

1.2 Research Plan

The ultimate goal in conducting this research is to test whether delivering PNS in response to motor intent can induce timing-dependent plasticity, leading to greater improvements than that seen when PNS and exercises are performed separately. In order to achieve timing dependence, detection of motor intent must occur immediately after or even in anticipation of movement onset. Additionally, detection of this intent must be accurate in order for the correlation between intent and stimulation to hold. Because early detection is needed, as well as the difficulty in detecting motor signals from electromyogram (EMG) in patients with neurological injury, detection of motor intent must occur at the brain. The primary goal of the current study is to develop a brain-machine interface (BMI) that utilizes electroencephalogram (EEG) signals to accurately detect motor intent and trigger PNS. Prior to studying this in patients, we first developed and tested our system in healthy volunteers without PNS. The proposed BMI will not only classify EEG into “rest” and “task” states, but will also deliver PNS at an intensity matched to the subject’s level of mental effort as well as its duration. A secondary goal of this study is to identify features of EEG that correlate with the level of force exerted during a handgrip exercise.

This thesis will discuss in Chapter 2 what work has been done in the fields of plasticity-inducing physical rehabilitation and brain-machine interfaces for detecting the intent to move. In Chapter 3, our experimental procedures will be introduced, including our data collection and analysis. Chapter 4 will present the results of the study, and Chapter 5 will discuss these results. Finally, conclusions will be drawn in Chapter 6, as well as suggestions for future work.

Chapter 2: Background

2.1 Motor rehabilitation therapy

Incomplete spinal cord injury, stroke, and many other neurological injuries result in a loss of motor function. Because of this, rehabilitation focuses on highly repetitive exercises of the extremities, referred to as “massed practice” (1). Studies have shown that cortical reorganization occurs after both spinal cord injury and stroke regardless of whether or not a patient participates in a rehabilitative program, though rehabilitation capitalizes on the reorganization, leading to better recovery of motor function (2; 3). When the upper extremities are affected, massed practice can include large, general movements of the arm as well as specific movements of the hand or fingers, with a single movement being repeated for an extended period of time (2). These movements can be performed with help from a therapist or with the assistance of a robot that guides the arm in a smooth motion. Alternatively, stimulation of multiple muscles, in the form of functional electrical stimulation (FES), can induce contractions in order to elicit a movement. When utilized only for assistance in executing massed practice exercises, it is referred to as functional electrical therapy (FET) (4). Stimulation is triggered manually, such as the patient or therapist pushing a button, which begins stimulation of muscles to induce a functional movement (4; 5). This method also has been shown to increase cortical excitability when compared to subjects who perform only massed practice without FET (6).

2.2 Somatosensory peripheral nerve stimulation as a therapy

Continuous sensory feedback is important in successful execution of most motor tasks. For instance, think of the difficulty people have speaking while their mouth is numb from a mandibular dental procedure. Although the anesthesia does not directly affect their motor control, due to the lack of proprioceptive feedback, coordinated movements of the lips and tongue are almost impossible. In the therapies discussed above, the emphasis is on self- or externally-induced movement for promoting use-dependent plasticity while the patient receives visual feedback, and in cases where sensory impairment in the affected limb is not complete, proprioceptive feedback. As may be expected, externally applied PNS itself has been shown to enhance cortical excitability when used in conjunction with massed practice as well as when used as a standalone therapy (7).

In several investigated protocols, PNS, a low intensity electrical stimulation, is applied to one or a combination of nerves in the forearm for 2 hours. Each second, five 10 Hz pulses of 1 ms duration are delivered over the 2 hours of stimulation, as is typical of these protocols (7-9). The stimulation is gentle enough to not cause muscle contraction, and rather stimulates afferent pathways, leading to increased excitability of the cortex (7). Studies show that PNS can be delivered prior to or during massed practice in order to increase cortical excitability, therefore enhancing the effects of massed practice (7; 9; 10).

2.3 Induction of timing-dependent plasticity to enhance therapy

While PNS and FET both have proven benefits, there are potential drawbacks that may inhibit their efficacy. In the case of PNS, the subject is receiving stimulation without exerting any effort or without any correlation to their movement. However, Stefan et al. have shown that temporal pairing of PNS and cortical stimulation by transcranial magnetic stimulation (TMS) induces cortical plasticity in healthy subjects only when the two stimuli are delivered less than 100 ms apart (11). In terms of rehabilitation, cortical stimulation is not induced by TMS, but is rather the patient’s cortical activity from performing the massed practice exercises. This leads to the question of whether administering PNS prior to exercises or during the exercises with no timing dependence between stimulation and movement really reaps the maximum benefits of combining the two therapies.

In the case of FET, the external stimulation is the primary or sole driver of the muscle contractions rather than the patient’s volition. While this is very appealing in situations where the patient cannot elicit any meaningful movement of the limb, it doesn’t necessarily require constant mental effort from the patient. Therefore, the subject will experience visual feedback of the hand moving, and if sensory pathways are at least partially intact, proprioceptive feedback. However, if the patient does not attempt to move their limb in conjunction with the stimulator, the sensory information reaching the cortex will lack the critical pairing with activation of the motor cortex. Even if the patient does continue their mental effort while the stimulator works to move their limb, the level of mental effort may vary between repetitions though the stimulus intensity does not. This lack of correlation between the levels of mental effort and stimulation intensity again may not provide the optimal results from a therapeutic perspective.

Both issues raised can be addressed by monitoring a patient's mental activity for signs of attempted movement and level of mental effort in order to trigger stimulation with intensity proportional to the amount of mental effort the patient is putting towards the task. The ability to recognize motor intent is well established, and is achieved with devices known as sensorimotor rhythm brain-machine interfaces. Measurement of EEG correlates of force is a less active field of research, but some have been found.

2.4 Brain-machine interfaces for detection of motor intent

BMIs are devices that record brain activity and use information contained in the signals to control either a computer or a device (12). There are many categorizations of BMIs, the primary one being whether the method of obtaining the signal is invasive or noninvasive. Invasive techniques require surgical procedures to implant electrodes either on the surface of the cortex in electrocorticography (ECoG) or within the cortex with intracortical recording (13). Non-invasive techniques, on the other hand, only require the placement of electrodes on the scalp, in the case of EEG, or the placement of a subject in a large piece of equipment, in the case of functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG). While ECoG and intracortical recordings are superior in terms of signal quality and spatial resolution, they carry all the risks and costs associated with surgery. Of the non-invasive techniques, EEG is most commonly used in BMI development, largely due to the fact that fMRI and MEG systems are large and expensive, while EEG systems are much smaller and less expensive (12).

There are many subcategories of EEG BMIs based on the signal characteristics they are sensitive to, which determines the type of signal processing they perform. There are two features of EEG seen above the motor cortex that are commonly used in BMIs to detect the intent to move: sensorimotor rhythms (SMRs) and movement-related cortical potentials (MRCPs). MRCPs, discovered in 1965, are slow waves that begin 1.5 s before a voluntary movement (14; 15). Until recently, detection of MRCPs typically required averaging EEG over many trials, therefore limiting their usefulness for online BMI control (16). Only recently has detection of single MRCPs been made possible due to the development of new spatial filtering methods that better improve the signal-to-noise ratio of the EEG (17). SMRs are the 8-12 Hz mu rhythm and 18-25 Hz beta rhythm seen on EEG (18; 19). These rhythms have high amplitudes when an individual is as rest, but become much weaker on the contralateral side beginning 2 s before execution of a unilateral movement or during motor imagery (18). During actual movement, suppression of this rhythmic activity is seen on both sides of the cortex (20).

SMR BMIs all use the power contained in the mu and beta bands from channels above the sensorimotor cortex as command signals, or features, for controlling the action of the machine in the brain-machine interface. The mental state of the subject (i.e. motor imagery, no motor imagery) is determined by feeding the features to a classification or regression algorithm. There are many types of these algorithms with varying degrees of complexity, several of which are discussed by Lotte et al. (21), including support vector machines, neural networks, and Hidden Markov Models. One of the most commonly used classifier algorithms is linear discriminant analysis (LDA) because of its low computational requirements, its application to multiclass systems, and its good performance (21).

How well a SMR BMI performs depends largely on the features and type of algorithm used, as well as the ability of the individual subject to modulate their sensorimotor rhythms. In a 2007 study, Blankertz et al. (22) tested a BMI that did not require extensive training periods for the users and instead only required one 20 minute calibration session for training the classifier. During calibration, subjects performed right and left hand and right foot imagery tasks. Of the three tasks, the two tasks with better discriminability were selected to train an LDA classifier. Features were calculated by band pass filtering the raw EEG signals and applying common spatial patterns, a method of spatial filtering that is specific to individual subjects. In online classification sessions, this BMI achieved an average accuracy of 88.4% for two-class classification across nine subjects.

A study by Huang et al. (23) in 2009 looked at the effects of various features and classifiers. Subjects first participated in a calibration session, during which they were cued to start and then either stop or continue right hand or left hand movement or imagery. Various classifiers were used to discriminate between four classes, movement or imagery of the right or left hand, and whether the action continued or stopped. Among the features used was signal power of 4 Hz wide frequency bands in the alpha and beta bands on 14 different channels. A support vector machine was one of the classifiers used and obtained an average accuracy of 88.1% during a 10-fold cross-validation. This was slightly higher than the decision tree classifier (86.4%), which the authors of the study consider the best of the four classification methods they used due to its performance and simplicity.

A 2009 study by Yang et al. (24) compared classification accuracy of right and left hand imagery. The subject responded to the cue of either a right or left pointing arrow by performing imagery for the corresponding hand. Various features were calculated, and results were reported for autoregressive features, spectral peaks, and band power. The data was classified using k-nearest neighbor, LDA, and statistical classifiers. All combinations of features and classifiers obtained greater than 80% accuracy, with the exception of autoregressive coefficients using the k-nearest neighbors classifier. The highest accuracy (>90%) was achieved by using band power as features for LDA.

2.5 Brain-machine interfaces for rehabilitation

The idea to use brain signals to drive the movement of an impaired individual's limb with either FES or a robot is not new, and has in fact been the subject of several small-scale studies. As early as 2000, Pfurtscheller et al. (25) tested a BMI on a tetraplegic individual who controlled the movement of a hand orthosis by performing motor imagery. During this study, the subject underwent 5 months of training in order to learn to control his mu rhythms, which were recorded from bipolar channels anterior and posterior to C3, C4, and Cz. Training sessions consisted of a right or left arrow being shown to the subject for 1.25 s, after which he had 4 s to perform motor imagery of the appropriate limb, such as the right or left hand, and he received feedback in the form of a cursor moving to either the right or left. The performance of a two class linear discriminant analysis classifier using six features was evaluated for different pairings of motor imagery, and it was found that both feet vs. right hand imagery attained the highest accuracy at 95%. Eventually, the subject was able to use this imagery to control the opening and closing of a hand orthosis with almost perfect accuracy, and was even able to grasp and lift objects.

In 2003, Pfurtscheller et al. (26) reported another study which was very similar in nature except that FES of the forearm was used rather than a hand orthosis. The FES allowed for the sequential execution of one of five distinct phases of stimulation for performing a lateral hand grasp. Each time motor imagery of both feet was correctly classified, the next phase in the movement was triggered. It is important to note that motor imagery of the feet was found to be the easiest for the subject to control and produced the greatest modulation in EEG, but the imagery does not correspond with the limb movement being actuated. For rehabilitative purposes, it is important that the motor imagery (or preferably, a motor attempt) is of the limb that is being targeted in the therapy. Otherwise, the sensory feedback is not physiologically correlated with the activation of the motor cortex. Therefore, this application can be seen as an assist device rather than a rehabilitative device, as no functional improvement would be expected once the orthosis or stimulation is removed.

A case study in 2009 by Daly et al. (27) investigated the use of a BMI and FES for the restoration of isolated finger movements in a stroke victim. Before beginning the study, the subject was able to perform limited movement of her fingers together, and was incapable of moving individual fingers. The subject was asked to both attempt extension of the index finger and to perform motor imagery of index finger extension in response to a cue. In both cases, she was asked to attempt to relax the index finger following each extension attempt or imagery. FES to extend the index finger was triggered if band power in particular frequency bands was lower for the current trial than the average of the three previous trials. If the power was higher than the average, the trial was classified as rest and FES was not triggered. The subject underwent 10 weeks of training, during which only the BMI was used in order for the subject to learn to control her brain signals. The subject was given visual feedback in the training sessions when she successfully activated or deactivated her brain signals. Three weeks of treatment followed, during which both the BMI and FES were used. Following treatment, the subject was able to move her index finger 26 degrees independently of her other fingers.

Soekadar et al. (28) in 2011 published results of a BMI study which used MEG to detect suppression at 11-14 Hz during left hand motor imagery and trigger finger extension with an orthosis. The study included 20 healthy volunteers and 4 subcortical stroke patients with an inability to extend one or more fingers. Suppression was detected by comparing spectral features to a reference value, which was calculated in one of two ways. For half the subjects (10 healthy volunteers, 2 stroke patients), the reference value was determined by the power during the rest period preceding an imagery task. In the event that suppression was detected, the orthosis extended the subject's fingers at one of three speeds, determined by the level of suppression. The reference value for the other half of the subjects was determined by the power during the preceding rest period as well as previous imagery tasks. A detection of suppression in this group triggered movement of the orthosis at a constant speed. Healthy volunteers participated in 5 sessions and stroke patients underwent 15 sessions, with sessions occurring on consecutive days in both groups. By the end of the study, performance on the BMI had improved in both healthy volunteers and stroke patients from the group with the reference value determined only during rest and with feedback graded to the level of suppression. This study suggests that graded feedback may be important to motor learning, though it doesn't confirm this, as the results seen could also be due to the differences in determining the reference values.

Niazi et al. (29) developed a BMI in eight healthy subjects that detects MRCPs of the EEG during motor imagery to trigger PNS. Before and after intervention, TMS was used to induce motor evoked potentials (MEPs) of the tibialis anterior muscle as a measure of cortical excitability. To train the BMI, subjects performed, on average, a minimum of 30 self-paced dorsiflexions of their ankles. Subjects were asked to exert a force of 20%-30% of their maximum voluntary contraction (MVC) and were given feedback on a computer monitor to help gauge their level of force. This data was used to train the BMI to detect the initial negative component of the MRCP. The BMI was then tested while the subjects performed motor imagery of the ankle dorsiflexions. Detection of motor imagery triggered PNS of the common peroneal nerve strong enough to be perceived by the subject but not cause a contraction. True positives were indicated by the subject extending their right finger 2 s after performing imagery, while extension of the left finger indicated a false negative. Across all subjects, the BMI had a true positive rate of

67% and a false positive rate of 22%. Offline analysis showed that the BMI detected imagery 125±309 ms before onset. As control experiments, four additional subjects performed only self-paced motor imagery, and another four only received randomly timed stimulation of the common peroneal nerve. The increase in MEP in the BMI subjects was significantly greater than in subjects from either control group, indicating the increased benefit seen when combining exercise and PNS.

All these studies make great contributions toward a BMI that can be used in a clinical setting to enhance rehabilitation of patients with neurological injuries, though none address all the aspects that we ultimately hope to in our research. Recovery of motor function may not be possible in all individuals, but for those with the potential for recovery, rehabilitation could provide more reliable increases in function than an assistive device that must be worn in order to receive any benefit. Logically, for purposes of rehabilitation, the motor imagery used to modulate the control signal for the BMI needs to correspond with the stimulation being applied or movement of the orthosis. For instance, although foot imagery may produce the greatest changes in the sensorimotor rhythm, as in Pfurtscheller's studies, hand imagery must be used if the end goal is to recover volitional control of the hand.

While Daly, Soekadar, and Niazi all address the use of BMIs as rehabilitative tools, there are still gaps we hope to fill. Daly and Soekadar both invoke movement of the disabled hand/finger, either with FES or an orthosis. Niazi's study is the only one that looks at the usefulness of delivering PNS in response to motor intent. However, his study does not require explicit relaxation of the subjects, all of whom were healthy. When applied to a patient population, feedback regarding not only mental effort but also relaxation is needed, because, as Daly (30) points out, these patients often have difficulty with coordinated movements as well as relaxation. Soekadar's study is the only one that provides graded feedback based on the level of effort from the subject. This graded feedback, in the form of the speed of finger extension actuated by an orthosis, is only determined by the relative amount of suppression in the sensorimotor rhythm. There is no evidence provided in the study that correlates the level of sensorimotor rhythm suppression with speed of contraction, however. While Soekadar did show that a combination of graded feedback and less adaptive reference values led to greater improvements in control of the BMI, no attempts were made to measure functional improvements in the stroke patients in the study. Therefore, one can speculate that this particular grading of feedback may show little benefit in functional improvements.

2.6 EEG correlates of exerted force

A study in 1999 by Mima et al. (31) attempted to find a relationship between the force exerted during a sustained finger pinch and EEG power in the alpha, beta, and gamma bands. EEG power in the alpha band during forces at 10% to 60% of MVC was significantly lower than during rest periods. A negative linear correlation was found to exist over the contralateral sensorimotor cortex in this range of forces, though not at 80% of MVC. According to Liu et al (32), the frequency components of EEG which are used for SMR BMIs don't capture the complexity of the EEG signal, and therefore are poor predictors of muscle force, perhaps explaining the inconsistency seen by Mima at 80% of MVC. They instead propose the use of nonlinear measures for this. In their 2005 study, Liu and colleagues investigated fractal dimension as a possible EEG correlate of force, while in 2009, Yao et al (33) looked at the largest Lyapunov exponent. Both of these measures quantify the level of chaos in a deterministic signal. Both studies found that the respective features increased nonlinearly as the exerted force increased. None of these studies attempted to classify the EEG.

Chapter 3: Experimental Design

In the following discussion, the term "session" is used to define a 60 to 120 minute period of data collection, performed once a week. The term "run" is used to describe a sequence of 10 to 25 cue presentations, separated by short rest periods. The term "cue" describes the visual stimulus used to prompt the subject to perform a handgrip task.

In these experiments, three sessions of EEG, EMG, and force data were recorded from each of five subjects. Each session consisted of runs in which the subject was presented with a visual cue and responded by squeezing a hand dynamometer at different levels of force. In the analysis, we address the following questions:

1) Can we distinguish from the EEG signal whether a subject is at rest or engaged in a left/right hand motor task?
2) Can we accurately label segments of a continuously recorded set of EEG signals as periods of rest versus left/right hand movement?
3) Do any EEG features reflect differences in the level of effort exerted by a subject in a motor task?
4) Are any EEG features correlated with varying levels of exerted force?

All experimental procedures were performed with approval from the University of Kentucky's Institutional Review Board (Protocol Number 11-0686-P4S).

3.1 Subjects

Five subjects were recruited from the general public and from the student body at the University of Kentucky, between 18 and 35 years of age, with no known health problems. All subjects indicated no previous experience with BMIs. Prior to participation in the study, individuals were required to give written informed consent. Subjects were compensated for their time.

3.2 Subject preparation

Prior to each data collection, subjects underwent the process of electrode application. To measure EEG, we used active EEG electrodes (g.LADYbird, g.tec, Austria), which contain preamplifiers to reduce noise, affixed in an electrode cap (g.GAMMAcap^2, g.tec, Austria) for ease of application. Electrode placement, shown in Figure 3.1, was based on the International 10-20 system and included the channels C3 and C4, which correspond the right and left hand motor areas, respectively. The electrode at AFz was used as the amplifier ground and the left earlobe as common signal reference. The cap was secured using a chinstrap, tightened to a point that was still comfortable to the subject. To confirm proper placement of the electrodes, Cz was measured at 50% from nasion to inion as well as 50% between preauricular points. An electrolyte gel (g.GAMMAgel, g.tec, Austria) was injected in the space between the scalp and the surface of each electrode, using a syringe with a curved plastic tip. Ideal contact was confirmed by inspecting the EEG trace for alpha waves, observed when a subject is relaxed with eyes closed, eyeblink artifacts, and EMG artifacts from the subject clenching his jaw, as shown in Figure 3.2. If inadequate contact was suspected for an electrode, the tip of the syringe was used to push hair from under the electrode to expose the scalp, and more electrolyte gel was applied as needed.

For the initial sessions, EMG was recorded from the flexor carpi radialis and flexor digitorum superficialis muscles of both the right and left forearms. Prior to electrode application, the forearm and elbow were lightly abraded with an alcohol pad to remove dead skin cells and oil, and then the alcohol was allowed to dry. Two Ag/AgCl recording electrodes (Blue Sensor M, Ambu, Denmark) were placed on each forearm, with the recording electrode placed at one-third the distance from the medial epicondyle of the humerus to the styloid process of the radius, and the reference electrode placed 2 inches distal, as shown in Figure 3.3. The ground electrode was placed on the bony prominence of the elbow.

3.3 Data collection

Refer to Table 3.1 for the sequence of runs during each of the recording sessions.

In each subject's initial session, EMG and grip force were measured in separate trials for the right and left hands. EMG was collected with a 16 channel biosignal amplifier (g.USBamp, g.tec, Austria), which band pass filtered (2-200 Hz) and sampled the signal at 512 Hz. The data was then written to a LabVIEW TDMS file (National Instruments, Texas). The force was measured with a hand dynamometer (Vernier, Oregon) and sampled at 512 Hz with a data acquisition board (NI USB-6251, National Instruments, Texas) and written to a separate TDMS file. After the first session, this data was analyzed to confirm that a correlation exists between the EMG and grip force for the subject. If this was confirmed, only the force was recorded in subsequent sessions instead of EMG. This enabled us to continue recording 16 channels of EEG and also decreased the preparation time for following sessions.

EEG signals were band pass (0.5-60 Hz) and notch filtered (60 Hz), and then digitized and sampled at 512 Hz by a biosignal amplifier (g.USBamp, g.tec, Austria) and written to a TDMS file. Force data was sampled at 512 Hz with a data acquisition board (NI USB-6251, National Instruments, Texas) and stored to a separate TDMS file. A copy of the visual cue state was sampled by both devices and recorded to their respective files to ensure that the data on the two could be properly aligned during offline analysis. A custom LabVIEW program which incorporated g.HIsys software (g.tec, Austria) was used to control the data acquisition and stimulus (visual cue) presentation. All this was performed on a single laptop (Dell Precision 4600) running the Windows 7 operating system.

Subjects were seated upright in a recliner approximately 1 m from a monitor for the duration of the experiment.

3.3.1 Sham stimulus presentation: initial session

In a subject’s initial session, after application of the electrodes, they were shown the stimuli that were later used as a cue to perform the handgrip tasks. At this point, the subjects had never seen the stimuli or been told their purpose before so the stimuli have no meaning to them. The subjects are simply asked to sit quietly and focus on the screen while the stimuli are presented. Therefore, any cortical response in a subject’s EEG seen during stimulus presentation while naïve to the task can be considered a visual evoked potential (VEP), a passive response of the cortex to the stimulus. This trial served as a control to confirm that the cortical responses seen while a subject is performing the handgrip tasks are indeed due to the task and not VEPs.

3.3.2 Measurement of right and left MVC

Each session began with a calibration of the subject's MVC for both the right and left hands. Force data was collected at a rate of 200 samples/second. During this portion of the experiment, the subjects were given a three second countdown to prepare for the five second MVC collection. During MVC collection, subjects were shown a gauge of their exerted force. The 1000 samples collected were smoothed using a 51 sample moving average and the maximum resulting value was taken as the MVC.

3.3.3 Measurement of EMG and grip force: initial session

The initial session also included recordings of 12 channels of EEG, 4 of EMG, and force exerted by each hand on the dynamometer. During these runs, the subjects were shown cues, during which time they were asked to squeeze the hand dynamometer at uniformly distributed percentages between 5 and 75% of MVC. Ten cues were given during this run with each target percentage performed once, cued in random order. This data was analyzed to confirm that EMG from the forearm and grip force are correlated. If this was found to be true for a subject, EMG was not recorded in future sessions, reducing the setup time for sessions and allowing all 16 channels of EEG to be recorded.

3.3.4 Measurement of EEG and grip force

All sessions included the measurement of EEG and force. Runs were performed with a single target force (20%, 35%, or 50% MVC) with 15 repetitions, as well as with varying target force (uniformly distributed between 5 and 75% MVC) with 25 repetitions. The right hand task was always performed first, followed by the left hand. The performance of the task was cued with the aforementioned stimuli, shown in Figure 3.4. During the trials, a fixation cross was visible in the center of the screen during rest periods to aid the subject in focusing their gaze on the center of the screen. The cue was presented as a solid gray dot surrounded by three larger black lines in the center of the screen, prompting the subject to begin performing the hand grip task. As the subject increased their exerted force, the gray dot grew in size. The subject's goal was to exert a force that kept the gray dot as close as possible to the second of the three outer circles, which represented the target force, and within the first and third circles, which represented ±10% of MVC from the target force. In cases where subjects showed fatigue during the variable force runs, the MVC was lowered to 75% of its original value, essentially changing the range of target forces to 3.75% to 56.25% of MVC. This was done in order to ensure that subjects would be able to reach the target force as well as decrease muscular artifacts that can occur in the EEG when subjects strain to reach the target.

3.4 Data analysis

Refer to Table 3.2 for a summary of the data analysis

3.4.1 Correlation between EMG and grip force

Before determining if EMG and force are correlated, both signals were conditioned. EMG was first filtered with a 4th order Butterworth high pass filter, with the cutoff frequency set to 10 Hz. The signal was then notch filtered with a 4th order Butterworth filter with cutoffs at 59.5 and 60.5 Hz to eliminate power line noise. The root-mean-square (RMS) value of the EMG was then calculated using the equation

$$RMS(i) = \sqrt{\frac{1}{n}\sum_{j=i-\frac{n}{2}-1}^{i+\frac{n}{2}} x(j)^2}\,, n = \frac{f_s}{10}$$

where i is one point in time, f_s is the sampling rate in Hz, and x is the filtered EMG signal. Note that this implementation results in an RMS that is $f_s/10$ samples shorter than the filtered EMG signal. In this case, the RMS was zero-padded with $n/2$-1 samples at the beginning and $n/2$ samples at the end. The force signal was smoothed using a 100 ms windowed moving average, and then Pearson's linear correlation coefficient between the two signals was calculated.

3.4.2 Preprocessing of EEG and force data

Prior to any analysis of the EEG and force signals, several steps were followed to first condition the signals and select data representative of rest and task periods. As previously mentioned, the EEG and force data were recorded to separate files because they were sampled by different hardware. Because the sampling was controlled individually by each device, there was the possibility of small sampling rate errors leading to EEG and force files of different lengths. The state signal on each file was paramount in aligning the data from the two files. The first step taken to align the data from the two files was to remove any baseline recording in excess of 2 s before the first change in state cue (from rest to task). In the event that there were 2 s or less of baseline recording, no data was removed. After aligning the beginnings of the data, the lengths of the remaining EEG and force data were evaluated. If one was longer than the other, samples were removed from the end of the longer dataset so that the two datasets were the same length.

After shortening the EEG and force recordings, if necessary, the EEG signals were spatially filtered. If all channels were of acceptable quality, a Laplacian spatial filter was used in order to remove signal common to the surrounding geographical area of the scalp (35). The Laplacian filter approximates the second derivative of the electrical potential and emphasizes current sources (or sinks) directly beneath its spatial center. It is computed using the equation

$$y_j(t) = x_j(t) - \sum_{i=1}^{4} x_i(t)$$

where

i = FC1, FC5, CP1, CP5	when	*j* = C3
i = FC2, FC6, CP2, CP6	when	*j* = C4
i = CP1, CP5, PO7, PO1	when	*j* = P3
i = CP2, CP6, PO8, PO2	when	*j* = P4

This left the spatially filtered C3, C4, P3, and P4 channels for analysis. In the case of Subject 1, who had poor signal quality on at least one channel of the Laplacian neighborhood that affected multiple runs, bipolar derivations (e.g., C3-FC1, C4-FC2, P3-PO1, and P4-PO2) were used to improve spatial resolution without contaminating the signal with a bad channel (36).

At this point, the four spatially filtered EEG channels were analyzed for artifacts. First, the signal was high-pass filtered with a cutoff of 30 Hz and the instantaneous power estimated as the mean-square value calculated over a 250 ms window. The background power was calculated by applying a median filter to the mean-square power, with a subject-specific filter size based on the time scales of artifacts typical of each subject. A dimensionless power ratio was determined by dividing the mean-square power by the background power, which suppressed the background signal and emphasized signal due to artifact. A subject-specific threshold was then applied to the power ratio to locate the occurrence of artifacts. Portions of the signal that crossed the threshold on any channel were marked and excluded from analysis.

3.4.3 Selection of data for analysis

Because the data from the EEG and force files cannot be assumed to have one-to-one correspondence in samples, segments of rest and task data were taken from each file, using the state recording to ensure that the data was aligned to the closest degree possible. For each occurrence of rest and task, segments of 2 s were selected. Each rest period had a randomized duration between 3 and 5 s (uniformly distributed), and each task period had a duration of either 3, 4, or 5 s, depending on the type of task performed. To select data for analysis from rest and sham periods, a 1 s delay from the state change was automatically added. In the case of rest segments, this delay was to allow the subject to return to a resting state following task. An automatic delay of 500 ms from the beginning of a task period was used in order to allow the subject time to react to the stimulus. This delay also reduced or eliminated data the amount of data containing visually evoked potentials in response to the cues, which is discussed in 3.4.4.

The selection of the 2 s segments of the remaining periods was then determined using the force signal. In the case of rest and sham segments during the sham runs, where no force signal was recorded, all segments for analysis were taken with the automatic 1 s delay. For all other runs, the force signal was smoothed with a 100 ms moving average, and then the variances of 2 s long windows were calculated for each possible window during a period, shifting the window by one sample for each calculation. The window that had the lowest variance for each task or rest period was then selected for analysis, and the delay from the state change was recorded. This method for selecting data segments was chosen because it was assumed that a subject’s exerted force would have the least amount of variance when the target force level for a task period was reached. Likewise, it was assumed that during rest periods the force variance would be smallest when the subject was not exerting any force. To control for the possibility that a subject may not have always reached a steady force, segments during task periods with variance which met the criteria for outliers on a boxplot (37) were identified and excluded from analysis. The task indices were then paired with the preceding rest indices, which ensured an equal number of rest and task samples for analysis.

3.4.4 Evaluation of visually evoked potentials

To determine if visually evoked potentials (VEPs) occurred in response to the cues for an individual subject, signal power on C3, C4, P3, and P4 was calculated from 1 s before cue onset to 1 s after cue offset during the sham stimulus run. The power spectrum was estimated over 500 ms windows with 250 ms overlap between windows, resulting in 2 Hz frequency resolution and 250 ms time resolution. Because all further analysis is limited to 8-40 Hz, as explained in section 3.4.5, power data beyond this range was excluded from VEP analysis. To determine if a visual response occurred, paired t-tests were performed for each frequency range by pairing the power values immediately preceding cue onset ($t = -0.5$ to 0 s) with the power values from all other points in time. Each t-test was performed with an alpha of 0.05 and indicates whether a change in power after cue onset was significant in each time segment thereafter. To correct for 16 different t-tests (8 to 40 Hz in 2 Hz wide bands) at each time point, results were considered significant if $p < \alpha/16$.

3.4.5 Comparison of task and rest

To aid in selection of features for classification, analysis of variance (ANOVA) was performed on data that would be used to train the LDA classifier. In order to maximize the amount of training data, the 50% MVC runs from sessions 1 and 2 were pooled together after confirming that the power distributions were similar between the two sessions. Next, the power of the previously selected 2 s long windows of data representing rest, sham, right, and left were calculated for the 4 channels using Thomson's multitaper spectral estimation method (38). The power from 8 to 40 Hz was integrated over 4 Hz bins. Power below 8 Hz was excluded because of the possible contamination by ocular artifacts that were not fully removed with spatial filtering. The upper limit of 40 Hz was chosen as an insignificant amount of relevant EEG information is expected above that point. In addition to calculating band power on each channel, the Teager energy, a nonlinear, instantaneous measure of total power, and sample entropy, a measure of signal complexity, were also calculated as additional features for each of the four channels (39; 40). ANOVA was then performed on the band power, Teager energy, and sample entropy data of the selected segments, comparing the values for rest, right, left, and sham periods across four channels (C3, C4, P3, P4).

3.4.6 Classification of task and rest

Offline classification using LDA was performed on data from rest, right, and left periods from 50% MVC runs, and rest and sham periods from the sham stimulus run. The 50% MVC runs were chosen for this section of analysis, as we theorized that the contrast between the EEG features during rest and task would be greatest for 50% MVC tasks. The classifier was trained on the features from sessions 1 and 2 and tested on the data from session 3. LDA classification was performed using two different sets of features for each subject: the first set used the band power from 8-10, 10-12, 16-24, and 26-30 Hz on all four channels as features, plus each channel's Teager energy and sample entropy, for a total of 24 preset features, and was applied to both the training and testing data sets; the second set of features was the 16 most significant features determined from ANOVA of sessions 1 and 2 and was applied only to the testing data set. To perform LDA classification, each spatially filtered channel was first band pass filtered to the desired frequency bands and then squared to convert the signals to power. Teager energy was calculated for each channel, and then the band power and Teager energies were smoothed with a 250 ms moving average. The signals were then decimated, reducing the time resolution to 250 ms, and band power converted to decibels. The cue state signal was also decimated. Because an accurate estimate of predictability requires more than 250 ms of data, sample entropy was calculated for each channel over 500 ms windows every 250 ms. Finally, the band power, Teager energy, and sample entropy were all smoothed using an 8^{th} order median filter.

At this point, there were 24 (or 16) continuous features in each set, with each sample containing (either all or a combination of) 250 ms of band power data, 250 ms of Teager energy, and 500 ms of sample entropy for each right and left hand 50% MVC run and the sham stimulus run. The corresponding state signals now consisted of 8 consecutive rest segments and 8 consecutive task (or sham) segments, separated by a variable number of segments excluded from analysis for reasons previously stated. These 8 segments corresponded to the 2 s windows established in section 3.4.3. To form the training set for the classifier, the state codes and data segments representing rest and task from each 50% MVC run of sessions 1 and 2 and sham stimulus run of session 1 were extracted and concatenated to form one feature array and one state vector. The LDA was then cross-validated by randomly selecting 90% of the data to train the classifier and testing its performance on the remaining 10%. This process was repeated 10 times and the average results recorded.

The features of the 50% MVC runs from session 3 were then calculated and the 2 s windows extracted to form the testing set. It is important to note that the sham stimulus run was only performed during session 1, as this was the only point in time when the subjects were naïve to the visual cue and motor task. Therefore, the LDA classifier was trained with sham tasks but there were no true sham tasks in the testing set. The training set feature values and corresponding states and testing set feature values were then used as inputs for the Matlab "classify" function, which returns the predicted states of the testing set based on LDA. The performance of the classifier was then evaluated by comparing the true and predicted states of the testing set. To better evaluate the performance of the classifier on the testing set data, the performance expected by chance was also determined. To do this, the true states of the testing data were randomly reassigned to the feature data and the four measures of performance calculated. This was repeated 100 times, and the means for sensitivity, specificity, positive predictive value, and accuracy were taken as the performance expected by chance.

To estimate how the LDA classifier would perform online, classification of session 3 was performed a second time using the continuous signal rather than just the selected segments of data. Features were calculated every 250 ms for the duration of the run, regardless of whether the signals contained artifacts or were during a period of unsteady force. This more closely mimics how the LDA classifier would function during a real-time task.

3.4.7 Comparison of varying levels of force

ANOVA was again performed, this time using power data from runs with 20%, 35%, and 50% MVC target force on a single hand. The same process was followed as previously described in 3.4.5. This iteration of ANOVA indicated possible candidate features for discriminating between rest, 20%, 35%, and 50% MVC on the right or left hand. As before, ANOVA was performed on data pooled from sessions 1 and 2 to identify candidate features for the LDA classifier.

3.4.8 Classification of varying levels of force

LDA classification was performed on data from varying levels of force using the same process as was used for classifying rest, right hand task, left hand task, and sham periods. Performance expected by chance was also calculated. However, in this case, the testing set contained all the states that were in the training set, as sham periods were not included in this analysis.

3.4.9 Correlation between EEG and force

To determine whether a correlation exists between components of EEG and force, data from the randomized MVC runs was used in order to obtain samples ranging from 5 to 75% MVC. Each subject's data was pooled across sessions 1 and 2 and runs 11 and 13 for the right hand, and runs 12 and 14 for the left hand. The 40 features used in ANOVA were log-scaled and plotted in a scatter plot as a function of force. A first order equation was fit to the log-scaled data from task segments and the Pearson correlation coefficient, ρ, and its significance, calculated.

3.4.10 Correlation between sample entropy and force across subjects

Correlation between sample entropy and force was also investigated across subjects. Sample entropy was chosen because it is amplitude invariant and therefore should be the least affected by the differences in amplitude between sessions and subjects. Sample entropy values during rest periods were assigned the force of the following task period. Each subject's task and rest data was separated into 10% MVC-wide bins from 5% to 75% MVC. Percent MVC was used in this portion rather than Newtons as there was a wide range between the subjects' absolute forces. The mean sample entropy of each bin was calculated separately for task and rest periods. The subject means were then averaged across subjects and the results plotted. As in the individual subjects' analyses, a first order equation was fit to the means for each channel, and the correlation coefficient and significance calculated

Table 3.1. Order of runs in experimental sessions

All experiments followed the order listed below. The term "Run" is used to define a continuously recorded group cue of presentations, separated by short rest periods, while the term "Session" is used to define the group of runs recorded in one day. All sessions began with Run 1 (electrode application). Run 2 (sham stimulus) was only performed in the first session. Right and left hand MVC was measured in all three sessions. Runs 3 and 4 (recordings of force with EMG) were only performed during the first session. Runs 5 through 10 required the subject to respond to the cues with one level of force for the entire run, and were performed in all three sessions. Runs 11 through 14 were also performed in all three sessions, and the target force for each cue presentation varied between 5 and 75% of MVC. Task duration indicates the amount of time subjects were given to respond to each cue, and Min ISI is the minimum rest period between two cues.

Run	Hand	% MVC	# cues	Task duration	Min ISI	Sessions
01	Electrode application		-	-	-	1, 2, 3
02	Sham stimulus		15	3s	3s	1only
	Measure R and L hand MVC		-	-	-	1, 2, 3
03	R w/EMG	Random	10	4s	3s	1 only
04	L w/EMG	Random	10	4s	3s	1 only
05	R	50%	15	4s	3s	1, 2, 3
06	L	50%	15	4s	3s	1, 2, 3
07	R	35%	15	4s	3s	1, 2, 3
08	L	35%	15	4s	3s	1, 2, 3
09	R	20%	15	4s	3s	1, 2, 3
10	L	20%	15	4s	3s	1, 2, 3
11	R	Random	25	5s	3s	1, 2, 3
12	L	Random	25	5s	3s	1, 2, 3
13	R	Random	25	5s	3s	1, 2, 3
14	L	Random	25	5s	3s	1, 2, 3

Table 3.2. Order of data analysis

The table below lists the different sections of data analysis performed, what data was used for each step of the analysis, as well as the sections of the text that discuss the analysis process and results. Refer to Table 3.1 for descriptions of each run.

Analysis		Data used in analysis Session(s)	Run(s)	See text section Analysis	Results
Subject-by-subject analysis					
Correlation of EMG and grip force	Right hand	1	Run 3	3.4.1	4.1
	Left hand	1	Run 4		
Evaluation of visual evoked potentials		1	Run 2	3.4.4	4.2
Comparison of task and rest (ANOVA)		1	Run 2, Run 5, Run 6	3.4.5	4.3
		2	Run 5, Run 6		
Classification of task and rest	Classifier training	1	Run 2, Run 5, Run 6	3.4.6	4.4
		2	Run 5, Run 6		
	Classifier testing	3	Run 5, Run 6		
Comparison of varying levels of force (ANOVA)	Right hand	1	Run 5, Run 7, Run 9	3.4.7	4.5
		2	Run 5, Run 7, Run 9		
	Left hand	1	Run 6, Run 8, Run 10		
		2	Run 6, Run 8, Run 10		
Classification of varying levels of force	Classifier training (Right hand)	1	Run 5, Run 7, Run 9	3.4.8	4.6
		2	Run 5, Run 7, Run 9		
	Classifier testing (Right hand)	3	Run 5, Run 7, Run 9		
	Classifier training (Left hand)	1	Run 6, Run 8, Run 10		
		2	Run 6, Run 8, Run 10		
	Classifier testing (Left hand)	3	Run 6, Run 8, Run 10		
Correlation of EEG and grip force	Right hand	1	Run 11, Run 13	3.4.9	4.7
		2	Run 11, Run 13		
	Left hand	1	Run 12, Run 14		
		2	Run 12, Run 14		
Across subject analysis					
Correlation of sample entropy and grip force	Right hand	1	Run 11, Run 13	3.4.10	4.8
		2	Run 11, Run 13		
	Left hand	1	Run 12, Run 14		
		2	Run 12, Run 14		

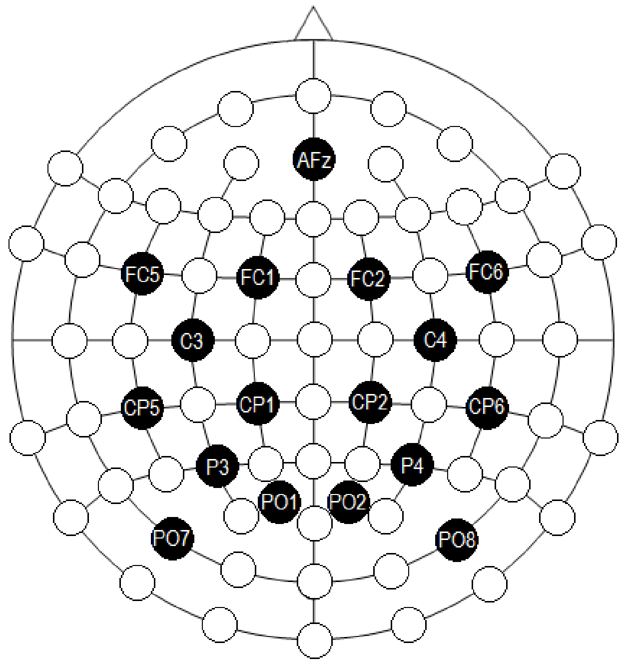


Figure 3.1. Placement of EEG electrodes according to International 10-20 system Sixteen channels plus ground (AFz) and reference (left earlobe) were collected during each of the subjects' three sessions. This configuration was chosen to allow Laplacian spatial filtering of C3 (minus the average of FC1, FC5, CP1, CP5), C4 (FC2, FC6, CP2, CP6), P3 (CP1, CP5, PO1, PO7), and P4 (CP2, CP6, PO2, PO8) in order to improve the spatial resolution of the signals. In cases where one or more channels had poor signal quality, bipolar channels (e.g. C3-FC1) were used rather than Laplacian spatial filtering.

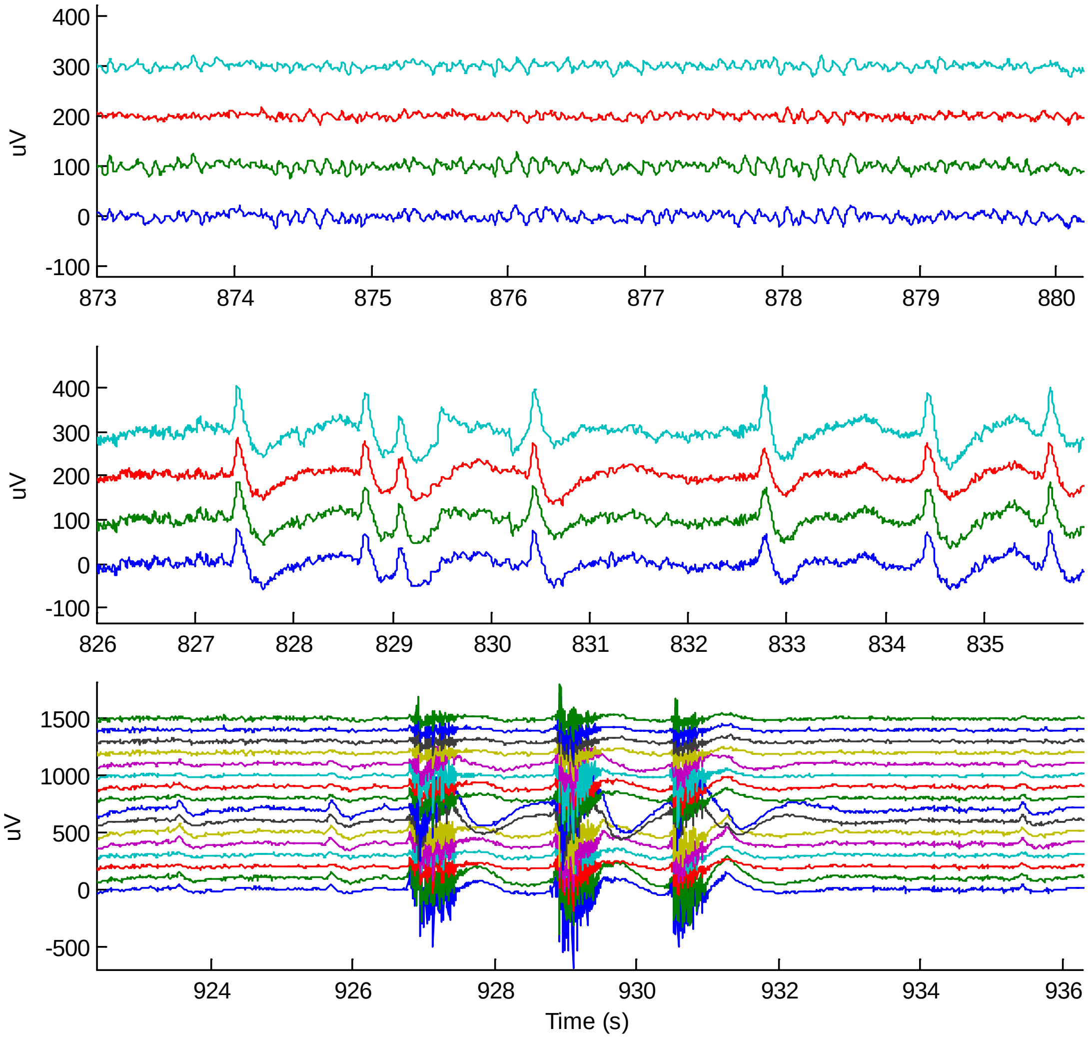


Figure 3.2. EEG components of good scalp-electrode contact

(Top) While a subject is sitting with his eyes closed, strong alpha rhythms are visible across many channels. Here, from bottom, CP1, CP2, CP5, and CP6 are shown. (Middle) Ocular artifacts (from bottom: FC1, FC2, FC5, FC6) should be easily visible on the traces of anterior channels. They are much weaker on posterior channels. (Bottom) EMG artifacts from jaw clenching can be seen across all 16 EEG channels.

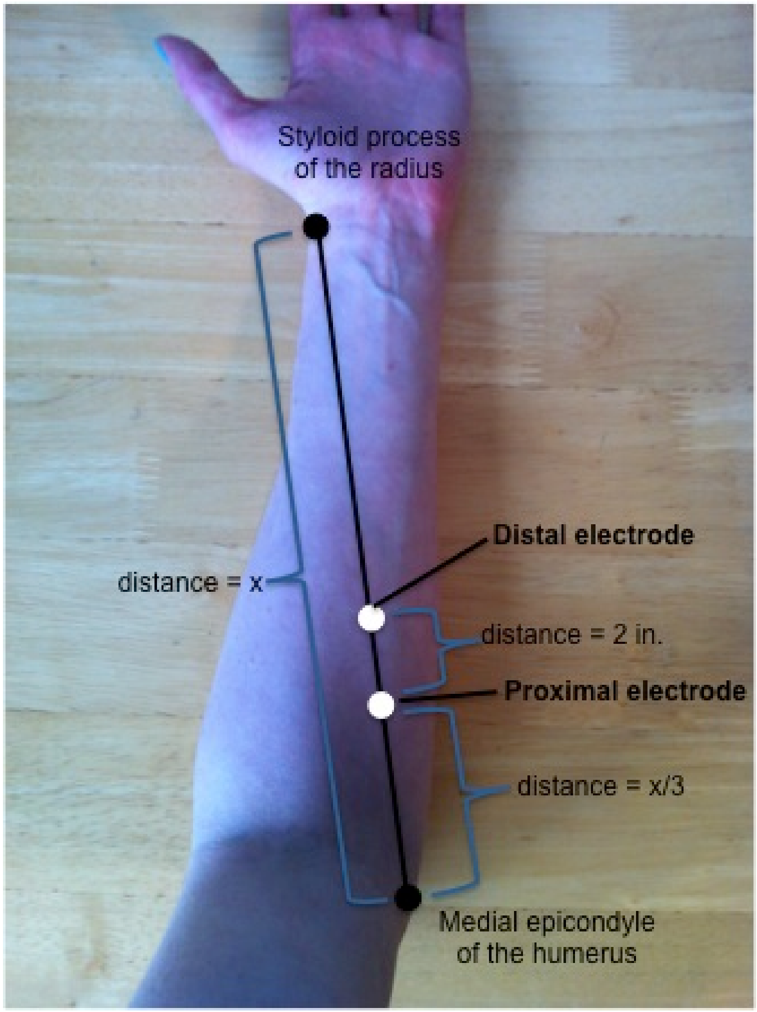


Figure 3.3. Placement of EMG electrodes

EMG was recorded from the flexor carpi radialis and flexor digitorum superficialis muscles of both the right and left forearms. The EMG electrodes were placed 2 inches apart on the forearm, with the ground electrode located on the bony prominence of the elbow.

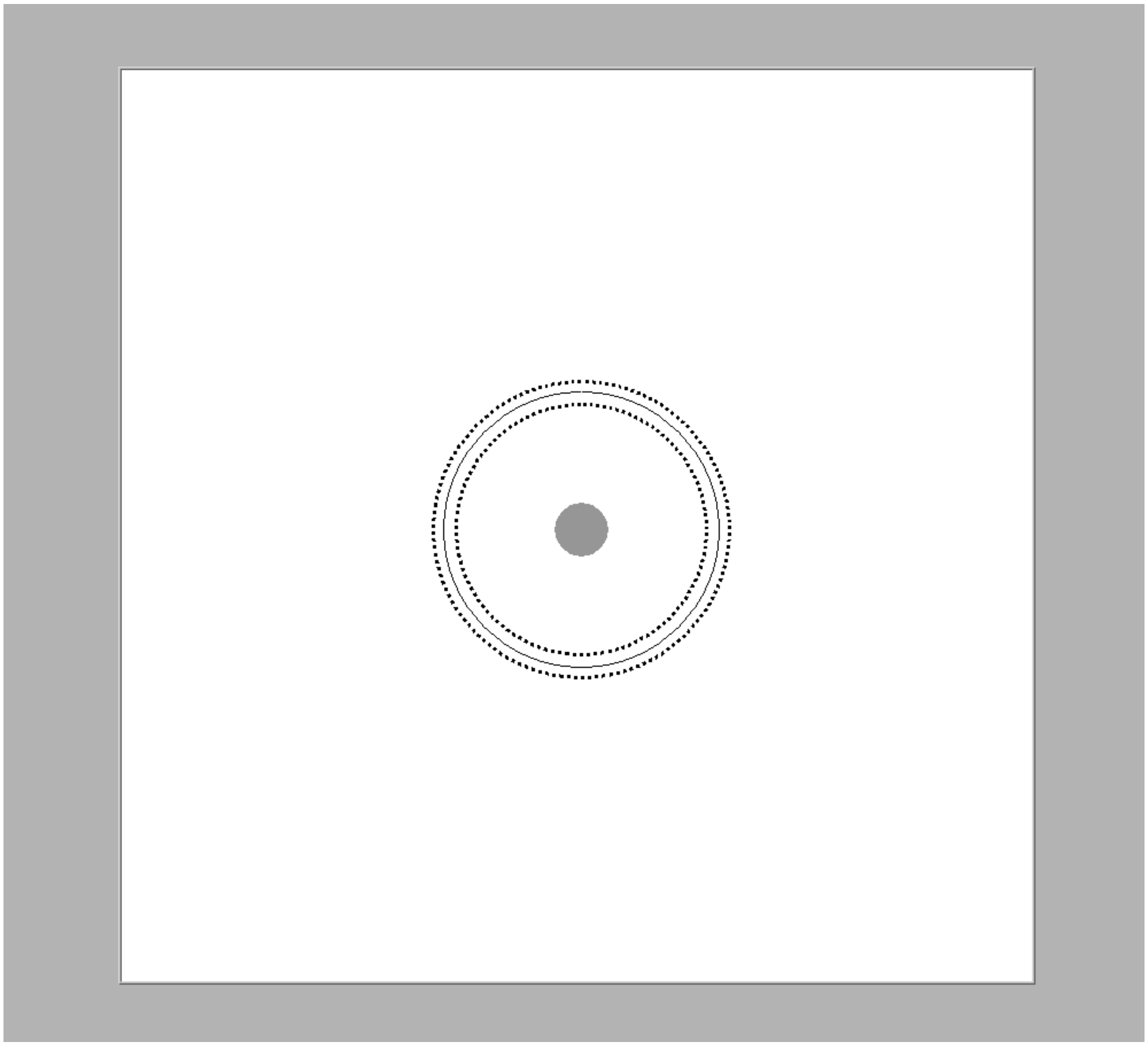


Figure 3.4. Cues shown to subjects

The stimulus presented to subjects is a simple gray dot in the center of three circles. As the subject increases his force on the dynamometer, the gray dot grows in size. The solid middle line represents the target force, while the dotted lines represent +/-10% of the subject's MVC from the target. The subjects are instructed to keep the dot between the two dotted lines, and to try to reach the solid line.

Chapter 4: Results

4.1 Correlation between EMG and grip force

The calculated linear correlation coefficients between the RMS of EMG and grip force are shown in Figure 4.1. The initial sessions gave coefficients greater than 80% for both the right and left arms for all subjects, except for Subject 1. Subject 1 had a correlation of 72% on the right arm for session 1, so EMG and force were recorded again from the right arm in session 2, which gave a correlation of almost 86%. These results suggest that exerted grip force can be used as a reliable indicator of muscle activity when looking for EEG features predictive of motor effort.

4.2 Evaluation of VEPs

VEPs in response to the sham cues were found to vary greatly between subjects as well as between channels. Figure 4.2 shows the responses on C4 of Subject 1 and P3 and P4 of Subject 5. Short-lived, significant changes were seen on C4 of Subject 1 at 1-1.5 s and P4 of Subject 5 at 0.5-1 s. Subject 5 showed longer-lasting changes on P3, reaching significance at 1-2.25 s after cue onset, and again at cue offset at 3-4 s. Though some of the other channels and subjects did exhibit some transient response to the cue, none reached significant levels. The changes that did occur had no discernible consistency in terms of timing or frequency.

4.3 Comparison of task and rest

In this section, signal power from 8 to 40 Hz in 4 Hz-wide bins, as well as Teager energy and sample entropy, were compared for rest, sham, right, and left data from sessions 1 and 2. Post hoc analysis was performed on any features found to be significant in order to determine which pairs of groups (rest, sham, right, left) had significant differences.

The results of ANOVA for Subject 1, shown in Figure 4.3 reveal that P3-PO1 and P4-PO2 discriminate better between at least two states than C3-FC1 and C4-FC2. The best discrimination in the band power features is achieved at 8-24 Hz, with all channels showing significance in this range. Teager energy and sample entropy are most significant on P3-PO1 and P4-PO2. The results of the post-hoc pair wise comparisons show that multiple features allow for discrimination between each of the combinations of tasks.

All channels were significant for Subject 2 at 8-24 Hz, as seen in Figure 4.4. Sample entropy was very significant on all channels, while Teager energy had the greatest significance on P3 and P4. The greatest number of features discriminated between rest and right/left, though all pairs of tasks were found to be significantly different in at least two features.

The results for Subject 3, shown in Figure 4.5, reveal that all channels were significant at 8-24 Hz with the greatest significance on P3 and P4 at 12-20 Hz. Teager energy was more significant than sample entropy, again with the greatest significance on P3 and P4. Pair wise comparisons of the features show that all four channels contain features that can discriminate between rest and right or left. Features of P3 and P4 allow for discrimination between rest and sham, while features of C3 and C4 allow for the discrimination between sham and rest or left.

Figure 4.6 shows that all features for Subject 4 were found to have significant differences between two or more states. The most significant differences in power were primarily at 8-24 Hz. Overall, Teager energy was more significant than sample entropy. A large number of features contained significant differences between rest or sham and right or left. The parietal channels, P3 and P4, both had features with significant differences between rest and sham, while C3 and C4 each contained one feature with significant differences between right and left.

Subject 5 had significant results from ANOVA on all channels at 8-20 Hz, though the significance is much greater on P3 and P4, shown in Figure 4.7. Teager energy and sample entropy were highly significant on P3 and P4, while C3 and C4 showed a lower level of significance for sample entropy, and no significance in Teager energy. At least one feature of each of the four channels contained significant differences between rest and right or left. No features contained significant differences between right and left.

4.3.1 Summary

In general, band power showed a greater difference between states in the lower frequencies than the higher frequencies. The most significant differences were found at 8-20 or 8-24 Hz for all subjects. In addition to signal band power, Teager energy and sample entropy were proven to vary between different states for all subjects. Overall, there were not only more features found to be significant on P3 and P4 than C3 and C4, but these features also had greater significance. Across all subjects, the greatest number of features contained differences between rest and right/left, and the least number of features contained differences between right and left.

4.4 Classification of task vs. rest

The performance of the classifier was quantified with sensitivity, specificity, and positive predictive value (PPV) for each state, as well as overall accuracy across all states. The sensitivity indicates the proportion of state *x* segments which are correctly classified as state *x*, while the specificity indicates the proportion of non-state *x* segments which are correctly identified as not being state *x*. PPV indicates how often segments classified as state *x* are actually state *x*, and the accuracy is the percent correct classifications for all states. All four measures are shown for each state for the cross-validation using the 24 preset features, the testing set using the 24 preset features, and the testing set using the 16 most promising features as determined by ANOVA. The performance expected by chance was also calculated for each of the four measures during classification of the testing set. The cross-validation values represent the mean of the 10 trials, while the error bars represent the standard error of the mean. It is important to note that there can never be a true positive for sham during out-of-sample classification because sham segments were not included in the testing set. Therefore, the sensitivity and PPV for sham are always zero when classifying the testing set. Additionally, the overall accuracy was computed for classification of continuous signals from session 3. An example of the output of a two-class classifier trained on rest and right data and applied to a continuous signal is shown in Figure 4.8.

The results of the classification for Subject 1 are shown in Figure 4.9. Rest was classified well in both training and testing sets, reaching a sensitivity of 70% or greater and very high specificity and PPV, indicating that the classifier was more likely to miss rest segments than incorrectly label segments as rest. Classification of rest in the testing set achieved much higher performance than that expected by chance. Classification of sham had very high specificity. Sensitivity was high as well, though the cross-validation PPV indicates more than 40% of the segments that were classified as sham were not in fact sham. Right was classified well during cross-validation, but classification of the testing set with the preset features yielded a low PPV, while the ANOVA-based feature selection led to poor specificity and PPV for right. This indicates that there were many segments incorrectly labeled as right. The classification of left was good during cross-validation, with high sensitivity and specificity. Sensitivity to left in the testing set varied greatly between the two sets of features, giving a sensitivity of approximately 45% when using the preset features and less than 10% when using the ANOVA-based features. Specificity was high with both sets of features, while PPV was relatively low with preset features but high with selected features. Both sets of features led to a large number of left segments being misclassified as right, with a greater incidence when the selected features were used. Classification accuracy was highest during cross-validation of the training data, reaching 84%. Classification accuracy of the test data yielded slightly better performance with the preset features, achieving 73% accuracy, while the selected features reached 67% accuracy. The difference in performance between the two sets of features is even smaller when applying continuous classification to the testing set, with 58% and 54% accuracy when using the preset and selected features, respectively.

The results of the classification for Subject 2 are shown in Figure 4.10. The classifier's sensitivity to rest decreases slightly from the cross-validation to the test set using the preset features, then increases beyond the cross-validation values to 80% when using the most significant features from ANOVA. The lower values for sensitivity to rest are counterbalanced by very high specificity and PPV, indicating very few false positives. All measures of sham classification perform very well, with the exception of PPV during cross-validation, which indicates a high occurrence of false positives. All three classifications detected right segments well, shown by the high sensitivity, though the relatively low values for specificity and PPV indicate frequent false positives. Finally, left was classified fairly well during cross-validation, but the classifier failed to label any segments from the testing set as left when using either set of features. This led to zero sensitivity and PPV, but perfect specificity. Accuracy was highest during cross-validation, reaching 76%. Use of the selected features led to better overall accuracy during classification of the test data, reaching 72% compared to 63% with the preset features. During continuous classification with the preset features, specificity to rest was much lower, while sensitivity to left remained at zero. The overall accuracy was 41%. The selected features again perform slightly better on the continuous classification with an accuracy of 53%, though a loss of specificity to rest was again seen.

Figure 4.11 shows the results of the classification for Subject 3. During cross-validation slightly more than half of the rest segments were classified as rest, while the sensitivity to sham, right, and left were better, all between 70 and 80%. Specificity for all four states was very satisfactory, with rest being the highest at above 90% and left being the lowest at 86%. The PPV was good for rest, though it was much lower for sham, left, and right, indicating a regular occurrence of false positives for those states. Overall, cross-validation performed with 64% accuracy. When classifying the testing set, the performance is much poorer than the cross-validation and with little difference between the two sets of features. There is low sensitivity for all states, with rest being the worst and left being the best. Specificity had the opposite trend, with rest being perfect and left being the lowest around 50%. PPV was again perfect for rest, but hovered around 20% for both right and left. For classification with both sets of features, rest was overwhelmingly misclassified as right and left. Out of 224 total rest segments in the testing set, 103 and 108 were labeled as right and 89 and 96 were labeled as left when using the preset features and ANOVA features, respectively. Right was also misclassified as left, with 92 out of 120 using the preset features and 74 when using the ANOVA features. The accuracy of classification performs only as well as or worse than chance, with 26% accuracy when using the preset features and 23% accuracy with the selected features. Continuous classification gives the same results.

The performance of the classifier for Subject 4 is shown in Figure 4.12. During cross validation, there is low sensitivity to rest but high specificity and PPV. Classification of sham, right, and left is good in all areas except for PPV, with sham performance being particularly low at 50%. The relatively low sensitivity for rest indicates that many rest segments were misclassified. The 50% PPV for sham indicates that the misclassified rest segments were most likely classified as sham. Despite these misclassifications, the accuracy was fairly high at 77%. During classification of the testing set, classification of rest was slightly poorer. The high specificity for sham indicates that few segments were misclassified as sham. The sensitivity to right increases for classification of the testing set with both sets of features, though specificity drops almost 40% when using the preset features, and 15% when using the selected features. There is also a large decrease in PPV of right for the testing set, with the 16 selected features providing better performance than the 24 preset features. The low specificity and PPV of right is due to the misclassification of rest segments as right. When using the preset features, 102 rest segments and 113 right segments are classified as right, while the classification using the selected features labeled 64 rest segments and 110 right segments as right. Sensitivity to left dropped by almost half when using preset features, but almost achieved cross-validation performance when using the selected features. The specificity and PPV for left were roughly similar to the cross-validation performance for both sets of features. The reason for the very low sensitivity to left when using the preset features is, out of 96 left segments, 25 were mislabeled as rest and 25 as right. The slightly better performance with the selected features is reflected in the 69% accuracy compared to 60% when using the preset features. The selected features again perform slightly better on the continuous classification of the testing set with 56% accuracy compared to 51% with the preset features.

The performance of the classifier for Subject 5 is shown in Figure 4.13. During cross validation, the overall performance is good, with the exception of PPV for sham and right. The large standard deviations of these, shown by the error bars, indicate that there was large variation among the 10 trials. There was much less variation in the accuracy of the cross-validation trials, which averaged 82%. Classification of the testing set had very different results. Sensitivity to rest was only 51% and 45% for the preset and selected features, respectively, though specificity and PPV were almost perfect. Specificity for sham was very good using both sets of features. Classification of right was very poor in terms of sensitivity and PPV, with the preset features giving zero percent for both, and the selected features giving only 3% sensitivity and 11% PPV. However, right had very high specificity for both sets of features. These numbers reflect that very few segments of other tasks were incorrectly labeled as right, but the majority of the 80 right segments were incorrectly labeled as another state. Right was mislabeled as left 78 times when using the preset features, and 73 times when using the selected features. This, along with the misclassification of rest as left, contributes to the low specificity and PPV for left. True left segments were classified very well with the preset features, shown by the high sensitivity, while the selected features led to fewer correct detections of left. Classification accuracy when using the preset features was 48% and only 40% when using the selected features. This difference did not hold with continuous classification, with both sets of features performing around 38% accuracy.

4.4.1 Summary

The classifier performance for all subjects is shown in Figure 4.14. In general, classification performance for Subject 3 was lower than for all other subjects. A mean accuracy of 76% (64%-84%) was achieved across subjects during cross-validation of the training data. Accuracy decreased to 54% for classification of the test set using both the preset (26%-73%) and selected (23%-72%) features. Despite the mean accuracy being equal for the two sets of features, for most subjects, one set of features clearly performed better than the other. Therefore, the decision of whether preset features or subject-specific features should be used should be made on a subject-by-subject basis. For all subjects, the classifier performed well on the testing data in terms of not mistaking non-sham segments as sham, as well as not mistaking non-rest segments as rest. Two common errors of the classifier were labeling left segments as right, or vice versa, as well as misclassifying rest as right and/or left. For most subjects, one type of error was seen much more frequently than the other.

4.5 Comparison of varying levels of force

In this analysis, four states were again compared, with the states being rest, 20%, 35%, and 50% of MVC. Sham was not used in this analysis and the states were compared separately for the right and left hands.

The results of the right hand analysis for Subject 1 are shown in Figure 4.15. All significant power features were below 28 Hz. The most significance was found on P3-PO1 and P4-PO2, which also had very significant Teager energy and sample entropy. The post-hoc pair wise comparisons of the significant features showed that the discrimination was only between rest and force. Significance between rest and 35% of MVC existed on 21 features, followed by 18 features with significance between rest and 50% of MVC. Figure 4.16 shows the results for the left hand, which had significance into higher frequencies, although the greatest significance was again found below 28 Hz. As with the right hand, P3-PO1 and P4-PO2 had the greatest significance. In contrast to the right hand analysis, the multiple comparisons revealed some ability to discriminate between different levels of force, with seven features showing significance between 35% and 50% of MVC, and one feature showing significance between 20% and 50% of MVC.

Figure 4.17 shows the results of the right hand analysis for Subject 2. The differences between levels of force are most significant below 24 Hz and for Teager energy and sample entropy. C3, P3, and P4 contain the most significant features. Pair wise comparisons of the significant features reveal that most differences occur between rest and force. However, two features of C3 had significant differences between 20% and 35% of MVC, and four had significant differences between 20% and 50% of MVC. One feature of P3 contained a difference between 35% and 50% of MVC. The analysis of the left hand, shown in Figure 4.18, reveals that the most significance is again contained below 24 Hz and in the Teager energy and sample entropy. The significant features on P3 and P4 are similar to the right hand analysis, but more significance is found on C4 than on C3 for the left hand. The pair wise comparisons again show the most differences are between rest and force, but one feature of P3 again shows a difference between 35% and 50% of MVC.

Right hand analysis for Subject 3, shown in Figure 4.19, revealed that the most significance occurred on C3, P3, and P4. The significance was found in band power features below 28 Hz and in the Teager energy and sample entropy. Pair wise comparisons indicate that the significance was only between rest and force. The left hand analysis is shown in Figure 4.20. The results for P3 and P4 appear similar to those for the right hand. However, there are many more significant features of C4 than C3, though Teager energy and sample entropy of C4 have low significance. Pair wise comparisons again show that the significant differences are between rest and force, with no significance between different levels of force.

The results of the right hand analysis for Subject 4 are shown in Figure 4.21. The significance is fairly widespread across the power features, though the greatest significance is contained below 24 Hz on all channels. While the number of very significant features (blue) on P3 and P4 are equal, C3 contains more significant features than C4. The majority of the differences are between rest and force. However, features of C4 and P4 have significant differences between 20% and 35% of MVC as well as 20% and 50% of MVC. Figure 4.22 shows the results of the left hand analysis. Teager energy and sample entropy are very significant on all channels. The greatest significance on P3 and P4 is contained below 24 Hz. Fewer features of C3 are found to be significant compared to the right hand analysis, while many more significant features are found on C4. Pair wise comparison shows that the majority of significant features contain a difference between rest and force. However, features of P4 allow discrimination between 20% and 35% of MVC as well as 20% and 50% of MVC.

Figure 4.23 shows the results of the right hand analysis for Subject 5. The greatest significance in the power features occurs below 20 Hz, with the greatest significance on P3 and P4. All channels have significance in the sample entropy, while P3 and P4 have significance in the Teager energy as well. The pair wise comparisons revealed that the significance only occurred between rest and force and not between the different levels of force. Figure 4.24 shows the results of the left hand analysis, which is similar to the right hand, though significant differences are found in a broader range of frequencies. Again, the greatest significance is found on P3 and P4. As with the right hand analysis, the significant differences were most often found between rest and force. One feature of C3 revealed a significant difference between 20% and 50% of MVC.

4.5.1 Summary

In general, the most significance was seen in the lower frequencies, typically at 8-24 Hz, and in the Teager energy and sample entropy, with more features of greater significance found on P3 and P4 than C3 and C4. These results are similar to those seen when comparing rest, sham, right, and left. Post-hoc analysis of the significant features revealed that most of the differences were between rest and the three levels of force. Differences between the different levels of force were found for some subjects, with significant differences between 20% and 50% MVC found most often.

4.6 Classification of varying levels of force

The performance of the LDA classifier was again quantified by the sensitivity, specificity, and PPV for rest and each of the three levels of force, and overall accuracy. In this classification, all four states contained in the training set are present in the testing set, unlike during the classification of the four tasks, for which sham was only contained in the training set. Additionally, classification was performed separately for the right and left hands.

The performance of the LDA classifier for the right hand of Subject 1 is shown in Figure 4.25. During cross-validation, its performance on rest is good by all measures. Good specificity is achieved for 20%, 35%, and 50% of MVC, but sensitivity and PPV are fairly low for the three levels. Of the three, 50% MVC is classified the best, with 75% sensitivity and 64% PPV. Classification of 20% MVC is the worst, with 55% sensitivity and 54% PPV. In general, those trends hold true for both classifications of the testing set, though the performance is usually lower than that of the cross-validation. When using the preset features, the sensitivity and PPV for each of the levels of force drops by 20-30%, with drops of less than 10% for specificity, with the specificity of 35% MVC increasing by 3%. The selected features provide better performance in most measures than the preset features. The sensitivity to rest drops, though it increases for all three levels of force. The sensitivity to 35% MVC and 50% MVC increase by approximately 25% compared to the preset features, with 50% MVC exceeding the performance during cross-validation. The selected features also provide better performance on the test set with regards to PPV for the different levels of force. By using the selected features, the PPV for 20% of MVC increases by almost 30%, with smaller improvements in performance for 35% and 50% of MVC. As expected, overall accuracy is highest during cross-validation at 77%, while the preset features yield 60% accuracy and the selected 65%. Application of the classifier to the continuous data from the testing set yielded 47% accuracy with the preset features and 50% accuracy with the selected features.

Figure 4.26 shows the results of the left hand classification for Subject 1. As with the right hand, the classifier performed well on rest in all measures during cross-validation and when using both sets of features on the testing set. During cross-validation, specificity was high for all three levels of force, but sensitivity ranged from 53% to 64% and PPV only reached 53%. Overall accuracy reached 72%. When classifying the testing set using the preset features, the sensitivity to all three levels of force dropped, with 50% MVC achieving only 3%. Specificity to 20% MVC dropped by 15%, while 35% and 50% MVC were similar to their cross-validation values. PPV was similar to sensitivity, with 50% MVC falling to 5%. Accuracy fell by a greater margin to 60% when applied to the non-continuous data, and only reached 49% when applied to the continuous data. By performing classification using the selected features, performance on 20% MVC and 35% MVC dropped by varying degrees, though modest improvements were seen for 50% MVC. Classification accuracy however, was lower when using the selected features, falling to 55% and 45% when classifying continuous data. For both the right and left hands, the classifier is able to correctly detect the majority of the rest segments without incorrectly labeling many segments as rest. The difficulty for the classifier is assigning the correct labels to the different levels of force.

The results of the right hand analysis for Subject 2 are shown in Figure 4.27. The classifier performance during cross-validation is good for most measures, generally staying above 80%, and with an overall accuracy of 84%. However, the PPV for 35% and 50% MVC only reach 66%. When using the preset features to classify the testing set, the sensitivity to rest increases to 95%, though the specificity for rest falls to 22%. There is also a steep decrease in the PPV for rest. The low values for specificity and PPV are due to a large number of 20% MVC segments getting misclassified as rest. No 20% MVC segments are classified correctly, explaining the 0% sensitivity and PPV, but specificity to 20% MVC is perfect. Performance for 35% MVC is worse compared to the cross-validation performance, with 50% sensitivity and only 11% PPV. There is little difference in specificity. The classifier also fails to correctly label any 50% MVC segments, leading to 0% sensitivity and PPV. Specificity is very high at 98%. The performance when using the selected features is very similar to using the preset features. The only notable difference is that sensitivity to 35% MVC is much better at 75%. The classification accuracy is fairly low for both sets of features, at 49% with preset and 48% with selected features. The classification accuracy is very similar when applied to continuous data, with 49% with the preset features and 45% with the selected features.

The left hand analysis results for Subject 2 are shown in Figure 4.28. During cross-validation, sensitivity is approximately 75% and specificity is 90% or above for all groups. PPV is high for rest, though lower for the three levels of force, being highest for 35% MVC at 75% and lowest for 20% MVC at 50%. Overall classification accuracy is 75%. Classification of the testing set is much less consistent across tasks when using both the preset features and the selected features. When using the preset features, sensitivity to rest drops to 20% though specificity and PPV are perfect. The classifier fails to correctly label any 20% MVC segments, resulting in 0% sensitivity and PPV but perfect specificity. All 35% MVC segments are correctly labeled, leading to perfect sensitivity. However, segments of other groups are misclassified as 35% MVC, which leads to its low specificity and PPV. The classifier also fails to correctly label any 50% MVC segments, resulting in 0% sensitivity and PPV. Only two segments are incorrectly labeled as 50% MVC, which explains the high specificity. By training and testing the LDA with the selected features, the results were dramatically different for rest and 35% MVC, while there was little difference for 20% and 50% MVC. The sensitivity to rest increased almost to cross-validation levels, though its specificity and PPV dropped to 50% and 60%, respectively, indicating an increase in the number of segments mislabeled as rest. The sensitivity to 35% MVC was again perfect, while the specificity increased to 71% and PPV to 33%. The increases in specificity and PPV are a reflection of a decrease in the number of rest and 20% MVC segments misclassified as 35% MVC. These increases are responsible for the improvement in accuracy from 22% with the preset features (23% when applied to continuous data) to 48% with the selected features (44% when applied to the continuous data).

Figure 4.29 shows the results of the right hand analysis for Subject 3. During cross-validation, classification of rest shows specificity and PPV of over 90%, indicating that few segments of other groups are misclassified as rest. The sensitivity, approximately 70%, indicates that over 30% of rest segments are missed. Classification of 20% MVC results in high specificity but low sensitivity and PPV. This indicates that many of the 20% MVC segments were misclassified, and that only half of the segments labeled as 20% MVC are correct. The results for 35% MVC and 50% MVC were similar to that for 20% MVC, with slightly worse performance in all measures. Overall accuracy was 64%. Classification of the test set using the preset features results in a sharp decrease in sensitivity to rest and a maintenance of the specificity and PPV, indicating that more rest segments are missed in this classification of the test set. Sensitivity to 20% MVC falls to only 2%, with a PPV of 7%, though specificity is high. The majority of 20% MVC segments are mislabeled as 35% MVC and 50% MVC, while most of the segments labeled as 20% MVC are rest. The best sensitivity of this classification is to 35% MVC, though it only reaches 56%. Specificity for 35% MVC is similar at 52%, while the PPV is only 17%. Sensitivity and PPV for 50% MVC are also extremely low, with marginal specificity. The poor performance in classification of the individual states is reflected in the accuracy at 25%, which is only slightly higher than that expected by chance. Accuracy remains at 25% when applying the classifier to the continuous data from the testing set. Performance of the classifier using the selected features was similarly poor, again only reaching 25% accuracy, and drops even further to 22% when applied to the continuous data. The greatest change in performance by using the selected features was an increase in the sensitivity and PPV for 20% MVC, though both were still very low. An improvement was also seen in the sensitivity and PPV for 50% MVC, which reached 42% and 21%, respectively. In both classifications of the test set, rest was often misclassified as 35% MVC or 50% MVC, while the different levels of force were mislabeled as one another.

The results of the classification of the left hand for Subject 3 are shown in Figure 4.30. Cross-validation of the training set resulted in 75% sensitivity to rest and over 90% specificity and PPV. The performance of the classifier for 20%, 35%, and 50% MVC is similar for the three levels of force, with all having lower sensitivity, specificity, and PPV than rest. Overall classification accuracy was 64%. When using the preset features to classify the test set, there is very low sensitivity to rest, though specificity and PPV are perfect. This is because the classifier does not mislabel any force segments as rest, but most of the rest segments are mislabeled as a level of force. Sensitivity and PPV for 20% MVC are low, but specificity is high. In this case, more 20% MVC segments get mislabeled as a higher force level than get labeled correctly, and almost 75% of the segments labeled as 20% MVC are mislabeled. Performance on 35% MVC is worse by all measures compared to 20% MVC. Sensitivity is the highest for 50% MVC, though it only reaches 51%. Specificity and PPV are also low due to the tendency of rest and 35% MVC segments to get mislabeled as 50% MVC. The frequent mislabeling of segments leads to the low accuracy of 21%, which is only slightly higher than would be expected by chance. When applied to the continuous data, the accuracy drops even further to 19%. When using the selected features to classify the test set, little difference is seen in the performance regarding rest. However, the sensitivity to 20% MVC decreases while the specificity increases slightly, indicating an overall decrease in the number of segments labeled as 20% MVC. The sensitivity and PPV for 35% MVC increase, while the specificity drops by 10%. These numbers reflect the increase in the number of segments labeled as 35% MVC from all classes. There is a slight increase in all measures of performance for 50% MVC. Accuracy improves slightly over that seen when using the preset features, but still only slightly exceeds that expected by chance, at 25%, when applied to the selected data segments, and falls below chance to 20% when applied to the continuous data.

Figure 4.31 shows the results of the right hand analysis for Subject 4. Cross-validation of the training set resulted in fairly good performance on rest, with sensitivity being the lowest measure at 78%. Sensitivity to 20% MVC was slightly higher at 82%. Specificity for 20% MVC was high, but PPV was less than 70%. 35% MVC only obtained a sensitivity of 68% and a PPV of 59%, though specificity was high. The worst performance was for 50% MVC with 55% sensitivity and 46% PPV. Overall, 71% of segments were classified correctly. By classifying the test set using the preset features, half of the rest segments were incorrectly labeled, though few force segments were mistaken as rest. The sensitivity to all the force groups was at or below 40%, while specificity was much better with a minimum of 70%. PPV was low, with all three levels of force reaching approximately 20%. In addition to rest segments getting mislabeled as force, each of the three levels of force were incorrectly labeled as the other two levels of force more often than they were correctly labeled, leading to an accuracy of only 39%. This accuracy dropped slightly to 35% when applying the classifier to the continuous data. By classifying the test set using the selected features, there is a slight increase in overall performance, reaching 49% accuracy for selected segments and 43% for continuous data. This is shown by the increase in most measures for the three levels of force, though it comes at the expense of slightly lower specificity and PPV for rest, as more force segments get misclassified as rest.

Figure 4.32 shows the results of the left hand analysis for Subject 4. Cross-validation on the training set performed well on rest for all measures. Performance on 20% MVC performed well also, though the PPV was slightly below 80%. There was a lack of sensitivity to 35% MVC, with only 63% of the segments being labeled correctly. PPV was below an ideal level at 76%, but specificity was very high. Finally, sensitivity to 50% MVC was acceptable at 72% and had good specificity. However, PPV was less than 50%, indicating that more than half of the segments labeled as 50% MVC were actually from another state. Overall, 77% of the segments were classified correctly. Classification of the test set using the preset features yielded good results for rest, though the sensitivity was slightly below the desired level at 78%. Sensitivity to 20% MVC was very low, only labeling 38% of the segments correctly. However, specificity was very high. The classifier was the most sensitive to 35% MVC, correctly detecting 80% of the segments. The relatively lower specificity and low PPV indicate that the classifier also labeled a large number of segments from other states as 35% MVC. Sensitivity is the lowest for 50% MVC at only 24%. A combination of missing many of the 50% MVC segments and mislabeling other states as 50% led to a low PPV. However, the specificity was high. The preset features led to an overall classification accuracy of 63% when applied to the selected data segments and 52% when applied to the continuous data. When classifying the test set using the selected features from ANOVA, performance dropped, sometimes dramatically, for all states, causing a drop in accuracy to 48%. Classification accuracy of the continuous data decreased to 43%.

The results of the right hand analysis for Subject 5 are shown in Figure 4.33. Cross-validation of the training set resulted in a classifier most sensitive to rest at 81%. All other measures of performance for rest were also high. The classifier was the least sensitive to 20% MVC, though the large error bars indicate there was a large amount of variation between the 10 cross-validation trials. The variation was also high for PPV which had an average of 73%. Specificity for 20% MVC was very high at 95%. Sensitivity and PPV for 35% MVC were marginal at 70% and 67%, respectively, while specificity was high. Sensitivity to 50% MVC was relatively high, though PPV was the lowest for this class, reaching only 57%. Overall accuracy of classification during cross-validation was 76%. When applying the classifier to the test set using the preset features, the performance between classes was highly varied, correctly classifying 48% of the selected data segments, and 42% when classifying continuous data. Sensitivity to rest was very low, only detecting 26% of the rest segments. The specificity and PPV for rest were fairly high due to a small number of force segments being misclassified as rest. The opposite results were seen for 20% MVC, which was detected correctly 80% of the time, but had low specificity and PPV due to 184 rest segments and 77 of the 20% MVC segments being labeled as 20% MVC. Sensitivity to 35% MVC was lower at 69%, though much fewer segments were incorrectly labeled as 35% MVC, leading to a better PPV of 63%. Specificity was high. Sensitivity to 50% MVC was the lowest of the three levels of force, with 56% of the segments detected correctly. Very few segments were incorrectly labeled as 50% MVC, however, which led to the high specificity and relatively high PPV. By using the selected features to classify the test set, drastic differences were seen in the performance, with some measures improving and others doing much worse, leading to a slight drop in accuracy down to 41% when applied to the selected data segments, and 36% when applied to the continuous data. There was little difference in the sensitivity to rest, but much fewer segments were misclassified as rest, leading to an increase in specificity and PPV. Almost all sensitivity to 20% MVC was lost, with only 4% of segments being correctly labeled. Specificity increased, primarily due to only 78 rest segments getting mislabeled as 20% MVC. PPV decreased, largely due to the inability to detect the 20% MVC segments. Sensitivity for 35% MVC was almost perfect, with 103 out of 104 segments getting labeled correctly. However, segments from other states were also labeled as 35% MVC with fairly high frequency, resulting in low specificity and PPV. Changing the features of the classifier had the smallest effect on the performance of 50% MVC. A 12% increase in sensitivity was seen, along with a 12% decrease in PPV, while specificity saw almost no change.

Figure 4.34 shows the results of the left hand analysis for Subject 5. Cross-validation of the training set resulted in high performance for rest. Similar performances were seen for all three levels of force, except for slightly lower sensitivity and PPV for each. The greatest difference was seen for 35% MVC, which only had a PPV of 59%. Overall, however, classification of the left hand data was better than the right hand, reaching 82% accuracy. Consistent with the findings on the right hand analysis of Subject 5, the performance of the classifier on the testing set was very different for the four states. The preset features led to a sensitivity of 23% for rest, but perfect specificity and PPV. This is because many rest segments were mislabeled as force, but no force segments were mislabeled as rest. There was very high sensitivity for 20% MVC, but there was low specificity and PPV. Although 73 of 80 total 20% MVC segments were correctly labeled, 178 segments were incorrectly labeled as 20% MVC. Sensitivity and PPV to 35% MVC were the worst for all four states at 14% and 15%, but specificity was high. The classifier was second most sensitive to 50% MVC, correctly detecting 74% of the segments. While PPV was only 60%, specificity was high. Using the selected features for the classifier once again resulted in very different performance which was neither consistently better nor worse. The selected features led to increased sensitivity to rest by 21%, with little change to other measures. The sensitivity to 20% MVC decreased by 30%, but an increase in specificity to almost 80% was seen. A large increase in sensitivity to 35% MVC occurred, along with an increase in PPV. As with the right hand analysis, changing the features of the classifier made little difference in the classification of 50% MVC. Though neither set of features resulted in better performance for all states, the preset features only reached an overall accuracy of 41%, while the selected features achieved an overall accuracy of 52% and more consistent performance between the four groups. The slightly better accuracy of classification with the selected features than with the preset features was also seen when classifying the continuous data, with 46% with the former and 35% with the latter.

4.6.1 Summary

The classifier performance for right hand and left hand tasks of varying levels of force across all subjects is shown in Figure 4.35and Figure 4.36, respectively. For both right and left hand classifications, much better specificity is achieved than sensitivity. Classification accuracy during cross-validation averaged 74% for all subjects for both the right (64%-84%) and left hands (64%-82%). Mean accuracy dropped approximately 30% for classification of the testing set, with the selected features only performing slightly better. For classification of right hand data, the preset features reached a mean accuracy of 44% (25%-60%) while the selected features gave a mean accuracy of 46% (25%-65%). Left hand analysis showed similar results, with a mean accuracy of 42% (21%-63%) when using the preset features and 45% (25%-55%) when using the selected features. Overall, the classifiers performed best at not mislabeling force segments as rest. However, the reverse was not true. Additionally, there was difficulty in distinguishing the different levels of force from one another.

4.7 Correlation between EEG and force

Visual inspection of the 40 log-scaled features plotted as a function of force revealed correlation for some features. There was wide variability in the features during rest, which seemed to decrease as force increased. Fitting of a linear equation to the log-scaled task data resulted in relatively small, but many significant, correlation coefficients. Subject 4 had the highest correlation coefficients, with $\rho = -0.6191$ and $p = 3.45 \times 10^{-10}$ on C4 at 8-12 Hz for the right hand, shown in Figure 4.37, and $\rho = 0.6153$ and $p = 4.15 \times 10^{-11}$ in Teager energy on P4 for the left hand. These correlation coefficients were much higher than those calculated for most of the other features of Subject 4 as well as for other subjects.

4.8 Correlation between sample entropy and force across subjects

The average sample entropy across subjects as a function of force during right hand grip is shown in Figure 4.38. There appears to be a slight upward trend in sample entropy with force during the task segments, which is confirmed by the positive slopes of the linear equations fit to the data. C4 shows the strongest relationship between force and sample entropy in terms of the correlation coefficient and its significance, while P4 had the weakest relationship by both measures. A similar relationship between sample entropy and force was not found during the rest segments, as would be expected since force is not actually being exerted during these periods.

Figure 4.39 shows the sample entropy during left hand grip averaged across subjects. None of the slopes during left hand task were as steep as during right hand task on C4, but a better and more significant fit to the data was achieved on P4. There are no significant correlations between force and sample entropy on any of the channels during rest periods.

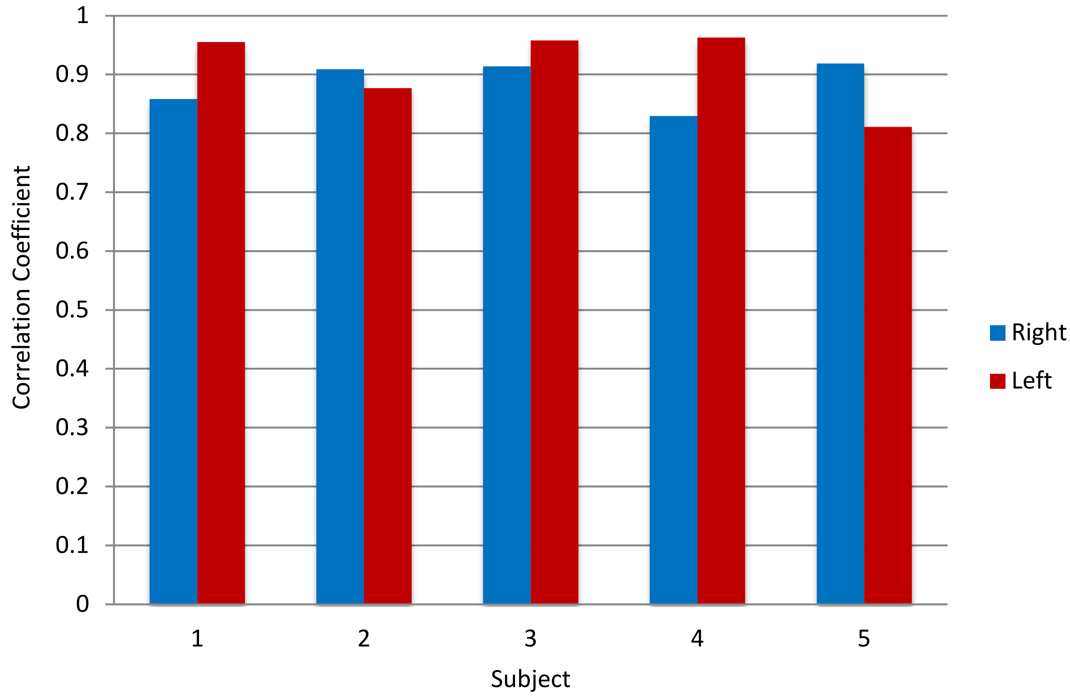


Figure 4.1. Correlation between EMG and force measurements

The correlation coefficients between RMS power of EMG and grip force measured by a hand dynamometer were calculated to confirm that grip force is a reliable measure of muscle activity. Grip force was much easier to measure than EMG, as it didn't require skin preparation and placement of electrodes, and was therefore preferred over EMG. Correlations between RMS power of EMG and grip force were found to be 80% or greater for all subjects, indicating that hand force can be used as a reliable measure of muscle activity when looking for predictive EEG features.

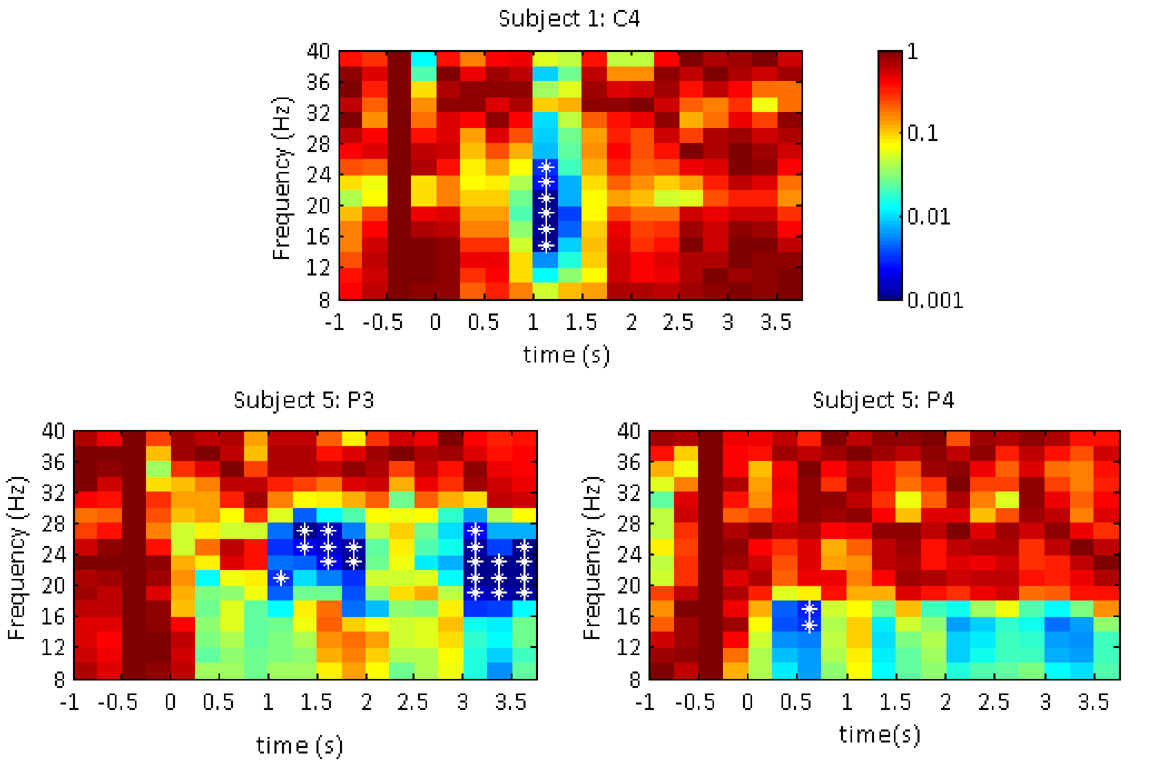


Figure 4.2. Visual evoked potentials in response to cue presentation

Visual evoked potentials due to cue presentation were measured by showing the subjects the cue while they were naïve to its meaning. Therefore, any changes in EEG can be attributed to the passive visual response to the cue. The figures above show the significance levels of paired t-tests comparing the power in 2 Hz wide frequency bands during the reference segment (t = -0.5 to 0 s) with the power estimates from all other half second time segments. The white stars indicate a significant change in power from baseline, with significance at $p < 0.003125$, using the Bonferroni correction in order to achieve an overall significance of $p < 0.05$. Subjects 1 and 5 were the only subjects that showed significant changes associated with the cue presentation. (Top) Subject 1 demonstrated a significant change in signal power at 1-1.5 s across a 12 Hz wide band. (Bottom left) Subject 5 experienced changes from baseline on P3 at t = 1-2.25 s after cue presentation and again at cue offset at t = 3 s. (Bottom Right) Subject 5 showed a significant response earlier on P4 at t = 0.5-1 s and at lower frequencies than on P3.

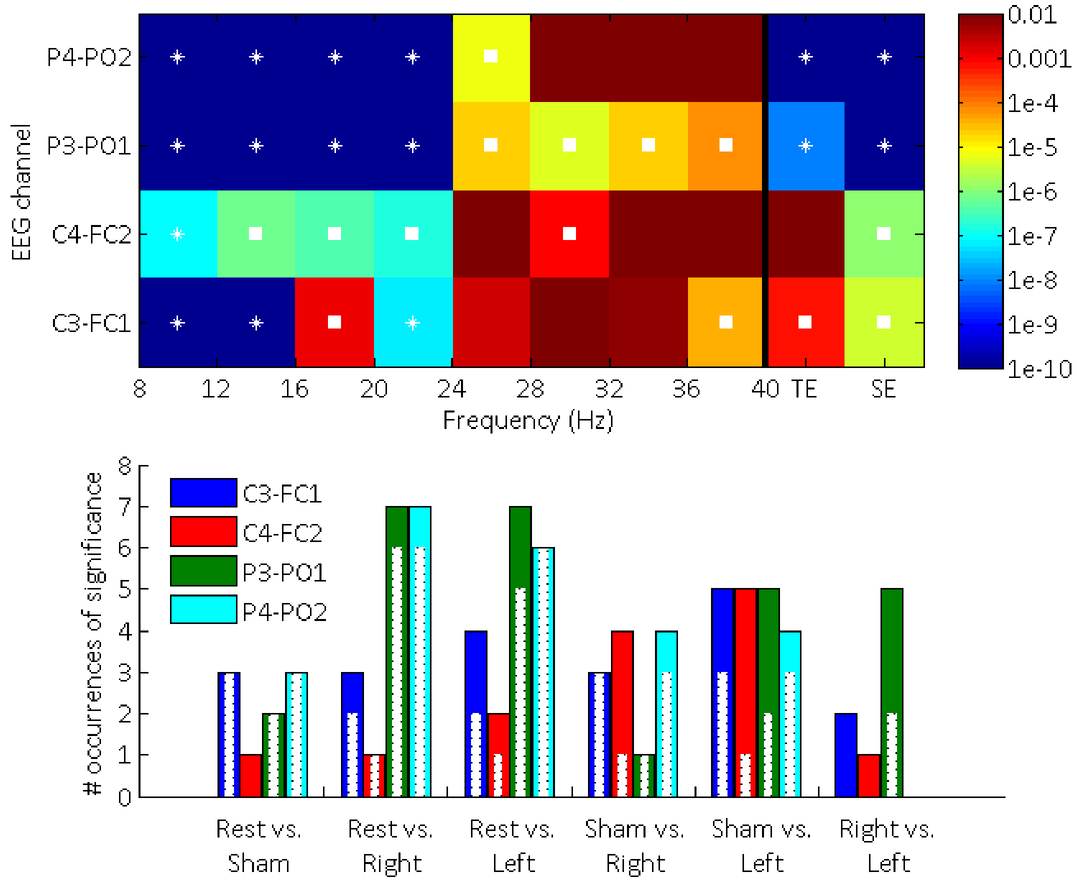


Figure 4.3. Subject 1 results of ANOVA comparing rest, sham, right, and left (Top) Results are shown as the p-value from ANOVA comparing rest, sham, right, and left states. To achieve an overall significance of $p < 0.05$, the Bonferroni correction was applied, and only $p < 0.0013$ was considered significant. The 16 most significant features are marked with white stars and the remaining significant features are marked with white squares. It was found that most of the significant differences for Subject 1 were contained on P3-PO1 and P4-PO2 at 8-24 Hz and Teager energy and sample entropy. (Bottom) Post-hoc pair wise comparisons were performed for significant features from ANOVA to identify which pairs of tasks were different. The number of features that showed significance between a pair of tasks was tallied for each channel. The white bars inside the colored bars represent the 16 most significant features from ANOVA which were used in the second LDA classifier. All pairs of tasks had significant differences in multiple features. Among the16 most significant features, significant differences were found between each pair of tasks.

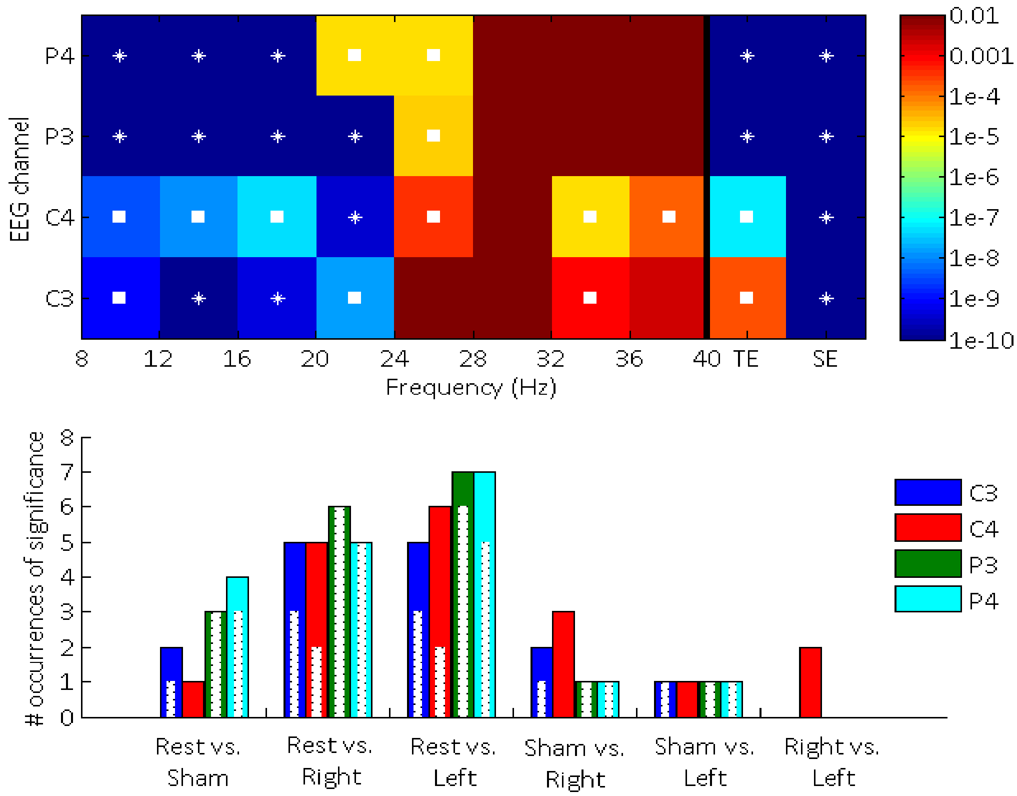


Figure 4.4. Subject 2 results of ANOVA comparing rest, sham, right, and left (Top) Results are shown as the p-value from ANOVA comparing rest, sham, right, and left states. To achieve an overall significance of $p < 0.05$, the Bonferroni correction was applied, and only $p < 0.0013$ was considered significant. The 16 most significant features are marked with white stars and the remaining significant features are marked with white squares. All channels showed significance in band power at 8-24 Hz as well as on sample entropy. The greatest significance was found on P3 and P4, on which Teager energy was also highly significant. (Bottom) Post-hoc pair wise comparisons were performed for significant features from ANOVA to identify which pairs of tasks were different. The number of features which showed significance between a pair of tasks was tallied for each channel. The white bars inside the colored bars represent the 16 most significant features from ANOVA which were used in the second LDA classifier. Discrimination between rest and right or left was found on the greatest number of features. However, each pair of tasks had at least two features that contained significant differences between the tasks. Among the 16 most significant features selected for the LDA classifier, none contained a significant difference between right and left.

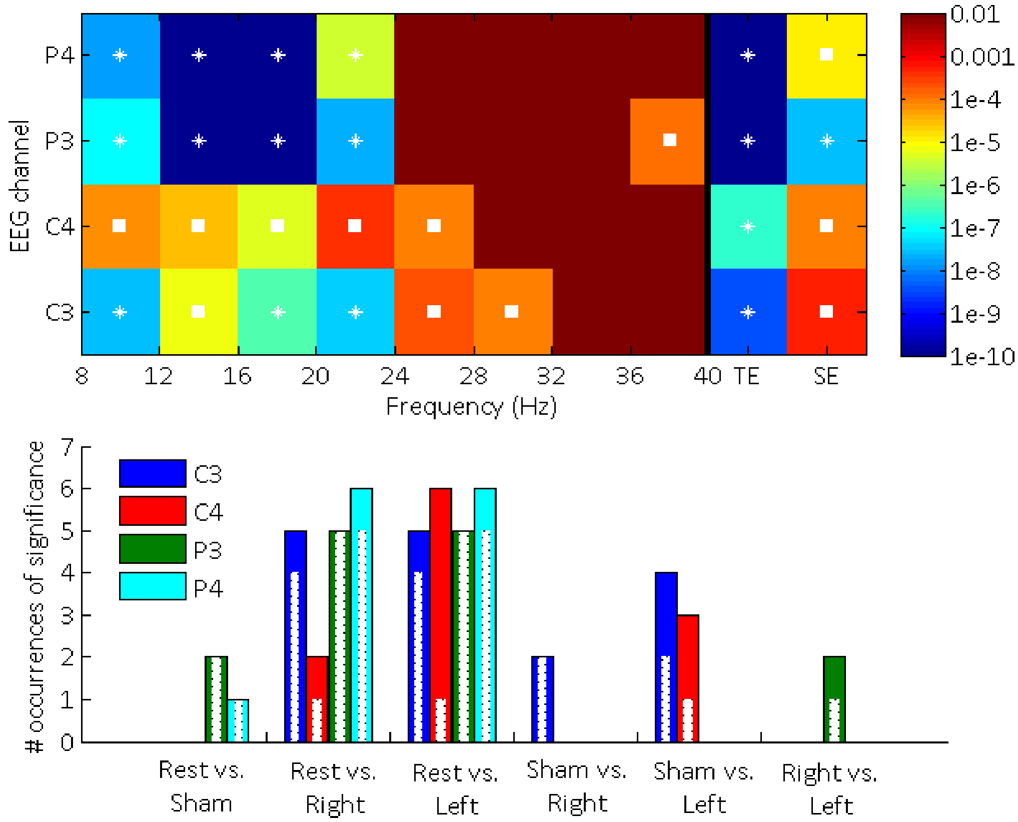


Figure 4.5. Subject 3 results of ANOVA comparing rest, sham, right, and left

(Top) Results are shown as the p-value from ANOVA comparing rest, sham, right, and left states. To achieve an overall significance of $p < 0.05$, the Bonferroni correction was applied, and only $p < 0.0013$ was considered significant. The 16 most significant features are marked with white stars and the remaining significant features are marked with white squares. All channels showed significant differences in band power at 8-24 Hz, though the levels of significance were much greater on P3 and P4 than on C3 and C4. Little to no significance was found at 28-40 Hz. (Bottom) Post-hoc pair wise comparisons were performed for significant features from ANOVA to identify which pairs of tasks were different. The number of features which showed significance between a pair of tasks was tallied for each channel. The white bars inside the colored bars represent the 16 most significant features from ANOVA which were used in the second LDA classifier. Significant differences between rest and right or left were found in the greatest number of features, with differences being found on all four channels. P3 and P4 were the only channels to contain differences between rest and sham, while differences between sham and right or left were found on C3 and C4.

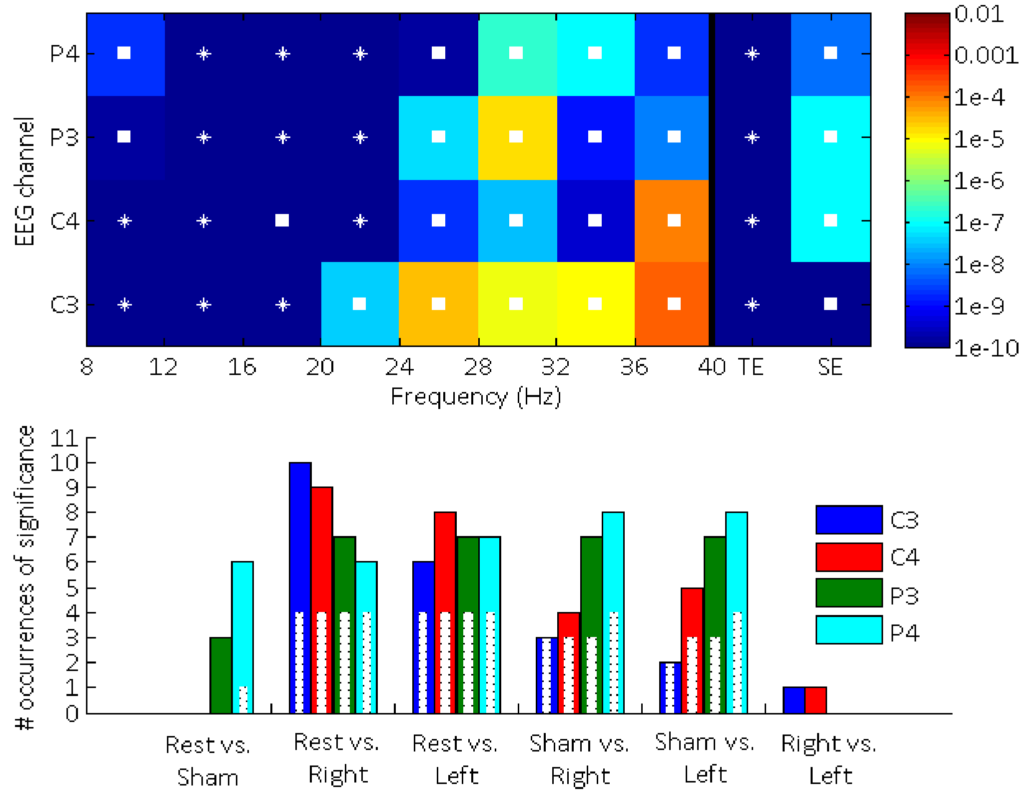


Figure 4.6. Subject 4 results of ANOVA comparing rest, sham, right, and left (Top) Results are shown as the p-value from ANOVA comparing rest, sham, right, and left states. To achieve an overall significance of $p < 0.05$, the Bonferroni correction was applied, and only $p < 0.0013$ was considered significant. The 16 most significant features are marked with white stars and the remaining significant features are marked with white squares. All features were found to contain significant differences between the four tasks. The greatest significance was focused on 8-24 Hz band power as well as Teager energy. (Bottom) Post-hoc pair wise comparisons were performed for significant features from ANOVA to identify which pairs of tasks were different. The number of features which showed significance between a pair of tasks was tallied for each channel. The white bars inside the colored bars represent the 16 most significant features from ANOVA which were used in the second LDA classifier. At least one feature was found to contain a significant difference between each of the pairs of states, with the greatest number of features containing differences between rest and right and rest and left. The two features with significant differences between right and left were not among the 16 most significant selected for the LDA classifier.

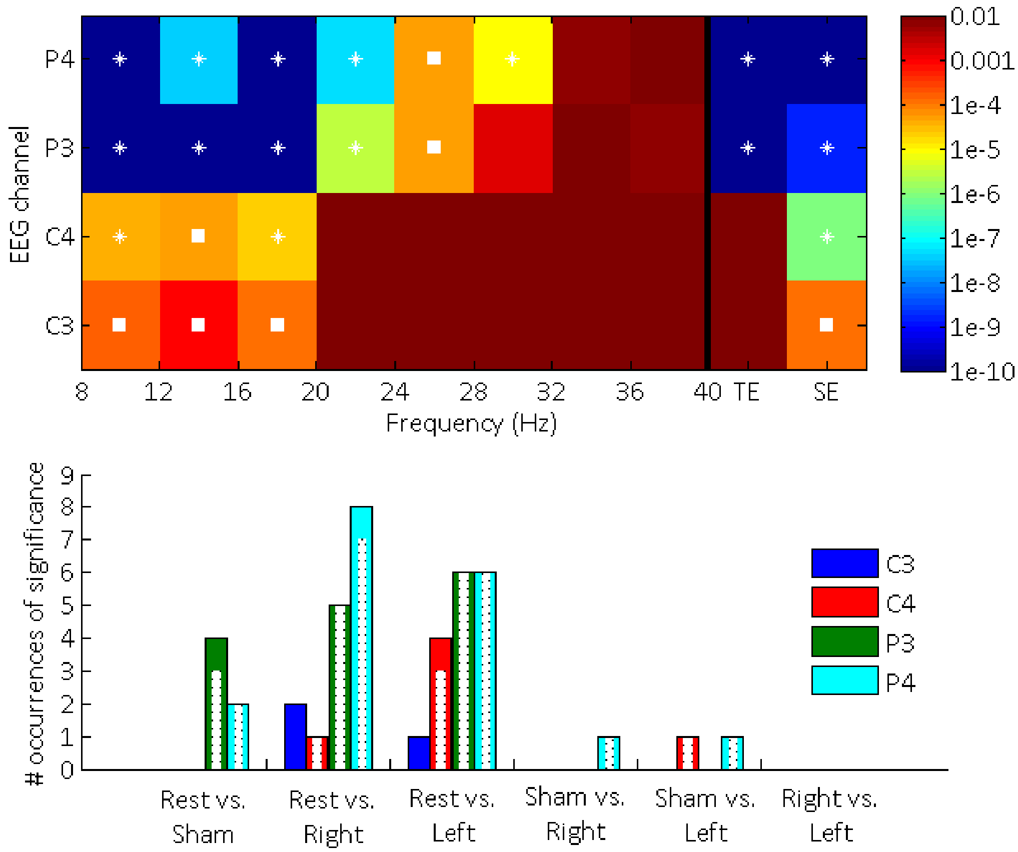


Figure 4.7. Subject 5 results of ANOVA comparing rest, sham, right, and left (Top) Results are shown as the p-value from ANOVA comparing rest, sham, right, and left states. To achieve an overall significance of $p < 0.05$, the Bonferroni correction was applied, and only $p < 0.0013$ was considered significant. The 16 most significant features are marked with white stars and the remaining significant features are marked with white squares. The majority of significance was found on P3 and P4 in the lower frequencies and on Teager energy and sample entropy. (Bottom) Post-hoc pair wise comparisons were performed for significant features from ANOVA to identify which pairs of tasks were different. The number of features which showed significance between a pair of tasks was tallied for each channel. The white bars inside the colored bars represent the 16 most significant features from ANOVA which were used in the second LDA classifier. Pair wise comparison revealed that at least one feature contained significant differences between each of the paired states except for right and left.

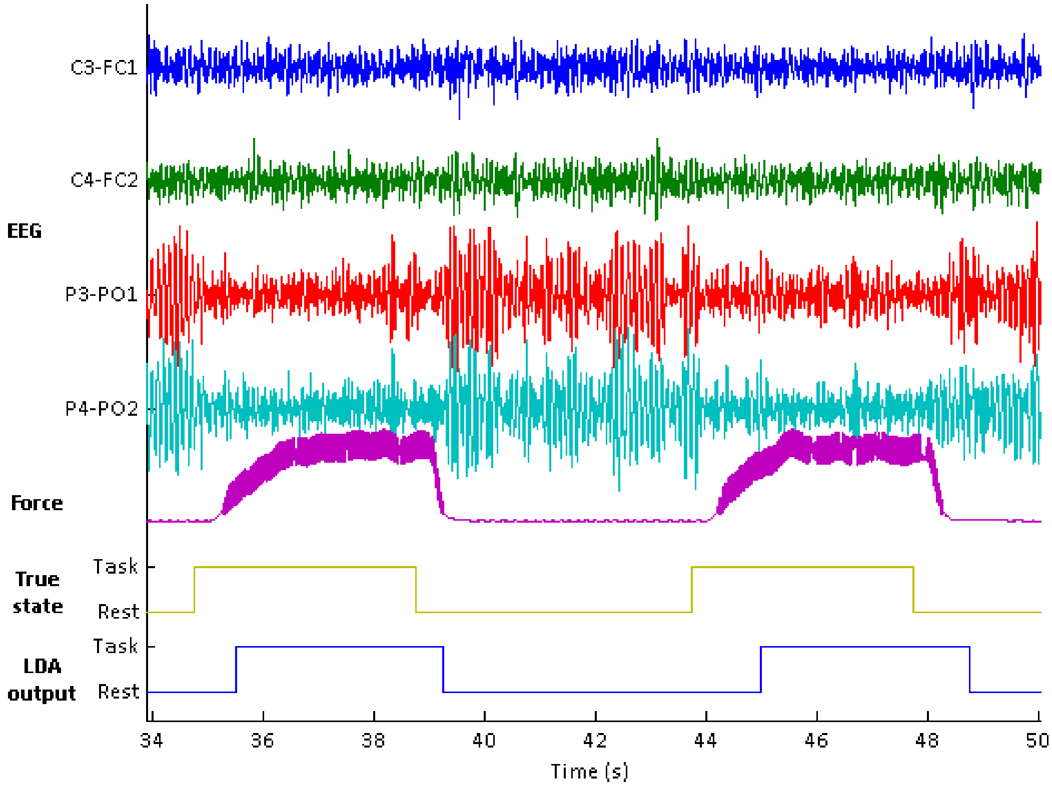


Figure 4.8. Classifier output when applied to a continuous signal

The LDA classifier was trained on EEG data from the 50% MVC right hand runs of sessions 1 and 2 and tested on the corresponding run from session 3. Preset features of each of the four channels (8-10 Hz, 10-12 Hz, 16-24 Hz, and 26-30 Hz band power estimates, as well as Teager energy and sample entropy) were used in the classifier. The figure above shows the results applied to a continuous signal from Subject 1. The force signal was not used for classification, though the classifier output shows a high level of correlation between the two.

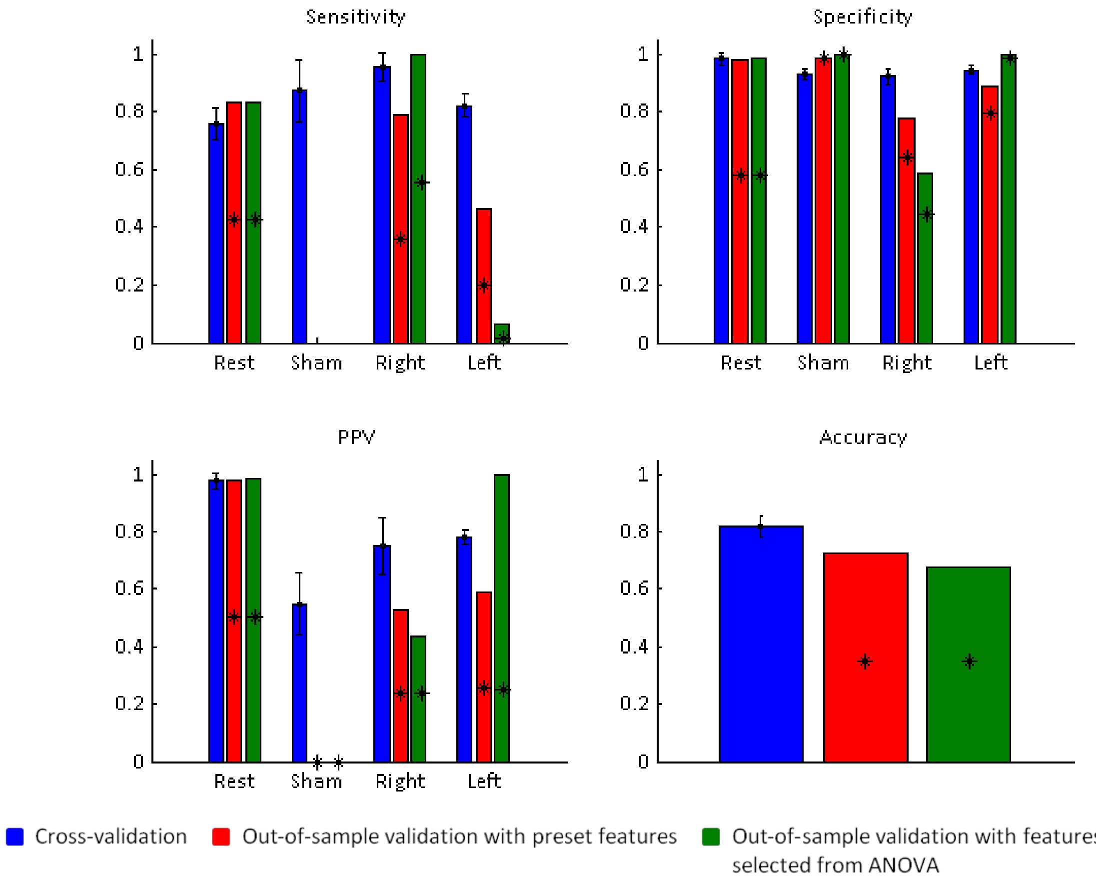


Figure 4.9. Subject 1 rest, sham, right, and left classification performance
Classifier performance was measured in terms of sensitivity, specificity, and positive predictive value for each of the four states, as well as the overall accuracy. Mean results of the 10 cross-validation trials are shown in blue, with the error bars representing the standard deviation. Also shown is the performance on the testing set using the preset features (in red) and the 16 most significant features selected from ANOVA (in green). Expected chance performance was estimated for classification of the testing set, and is marked by the black stars. The classifier performed fairly well on most states during cross-validation, while classification of the testing set gave less consistent results between states. The classification using the preset features performed with slightly higher accuracy, with 73% of the classifications being correct, while the classification with the selected features only reached an accuracy of 67%. All measures of performance when using either set of features for classifying the testing set were at or above the level expected by chance.

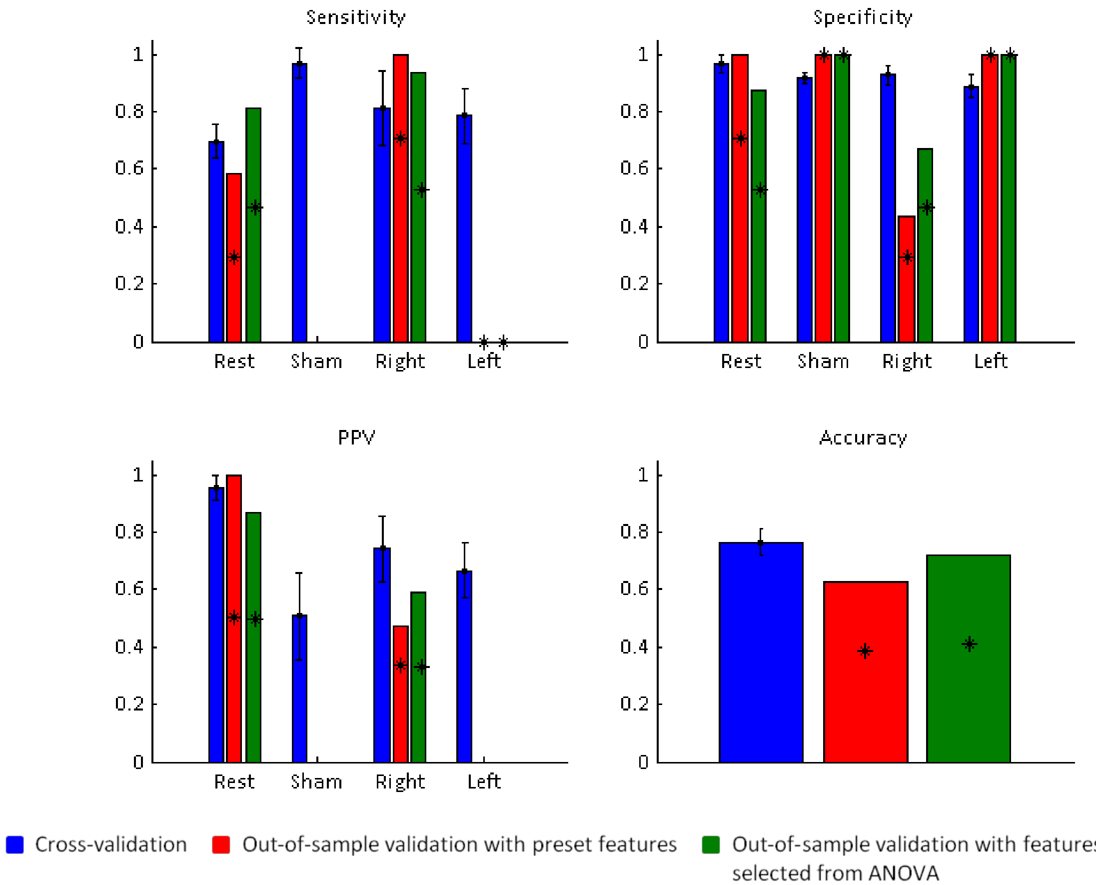


Figure 4.10. Subject 2 rest, sham, right, and left classification performance
Classifier performance was measured in terms of sensitivity, specificity, and positive predictive value for each of the four states, as well as the overall accuracy. Mean results of the 10 cross-validation trials are shown in blue, with the error bars representing the standard deviation. Also shown is the performance on the testing set using the preset features (in red) and the 16 most significant features selected from ANOVA (in green). Expected chance performance was estimated for classification of the testing set, and is marked by the black stars. Cross-validation of the training set resulted in poor sensitivity to rest, while the PPV for sham was low. Both classifications of the testing set failed to detect any left hand segments, with many left segments getting misclassified as right. Both sets of features used to classify the testing set resulted in classification performance at or above that expected by chance. The overall accuracy was higher when using the selected features, reaching 72%, compared to 63% when using the preset features.

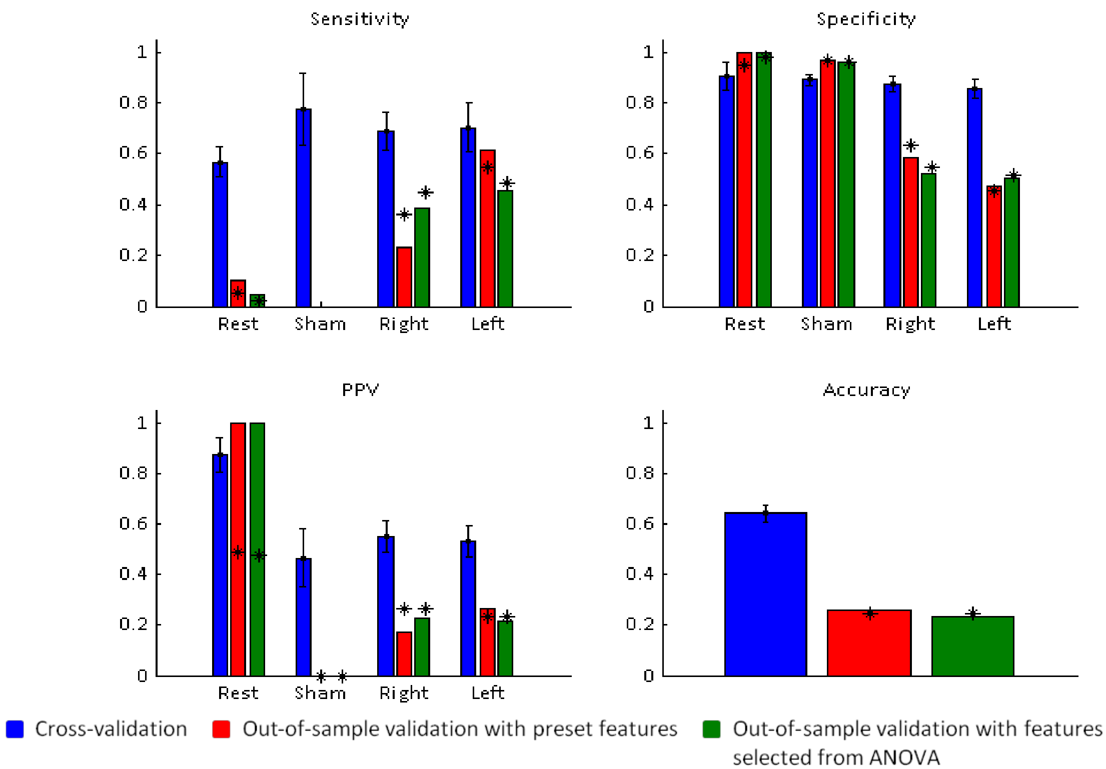


Figure 4.11. Subject 3 rest, sham, right, and left classification performance
Classifier performance was measured in terms of sensitivity, specificity, and positive predictive value for each of the four states, as well as the overall accuracy. Mean results of the 10 cross-validation trials are shown in blue, with the error bars representing the standard deviation. Also shown is the performance on the testing set using the preset features (in red) and the 16 most significant features selected from ANOVA (in green). Expected chance performance was estimated for classification of the testing set, and is marked by the black stars. Cross-validation resulted in marginal to poor sensitivity to all four groups. The PPV was high for rest but low the three other states. Sensitivity was low for all four states when using both sets of features to classify the testing set. Both sets of features led to a high number of rest segments being misclassified as right and left. Sensitivity and specificity to rest in the testing set was similar to that expected by chance, while PPV was almost twice the chance level. Both sets of features gave lower sensitivity, specificity, and PPV for right than expected by chance, and left performed near chance. The accuracy was poor for both classifications of the testing set, reaching 26% with the preset features, the level expected by chance, and 23% with the selected features, slightly below that expected by chance.

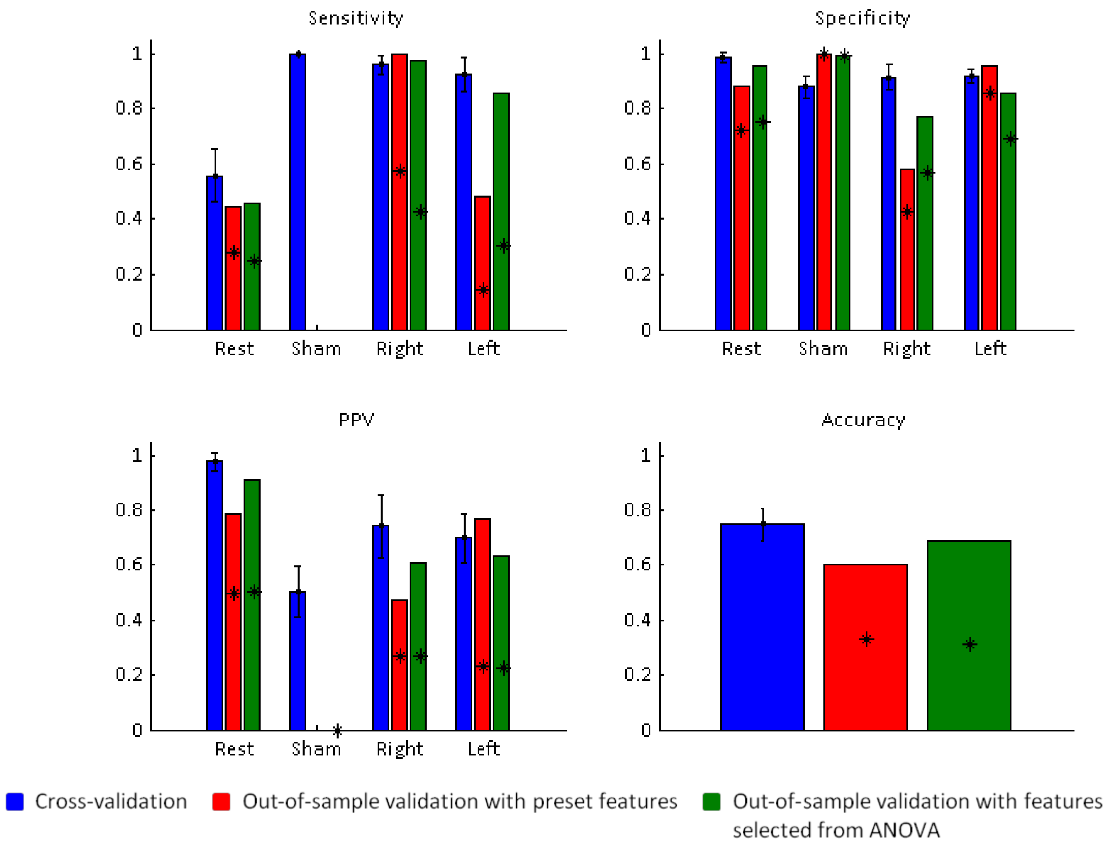


Figure 4.12. Subject 4 rest, sham, right, and left classification performance Classifier performance was measured in terms of sensitivity, specificity, and positive predictive value for each of the four states, as well as the overall accuracy. Mean results of the 10 cross-validation trials are shown in blue, with the error bars representing the standard deviation. Also shown is the performance on the testing set using the preset features (in red) and the 16 most significant features selected from ANOVA (in green). Expected chance performance was estimated for classification of the testing set, and is marked by the black stars. Cross-validation of the training set resulted in good classification of right and left. However, it lacked sensitivity to rest, and only half the segments labeled as sham were correct. There was again a lack of sensitivity to rest when classifying the testing set with both sets of features, though it was better than that expected by chance. All measures of performance for right and left, as well as overall accuracy, were better than chance for both sets of features when classifying the testing set.

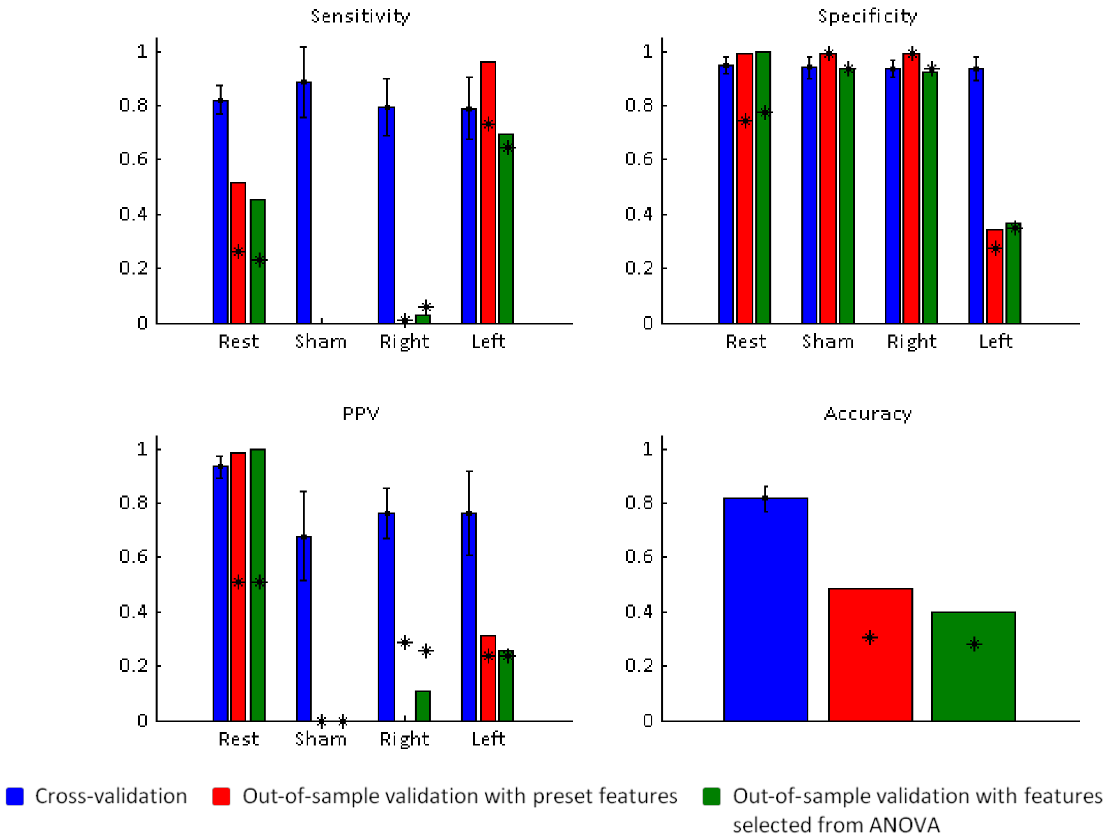


Figure 4.13. Subject 5 rest, sham, right, and left classification performance
Classifier performance was measured in terms of sensitivity, specificity, and positive predictive value for each of the four states, as well as the overall accuracy. Mean results of the 10 cross-validation trials are shown in blue, with the error bars representing the standard deviation. Also shown is the performance on the testing set using the preset features (in red) and the 16 most significant features selected from ANOVA (in green). Expected chance performance was estimated for classification of the testing set, and is marked by the black stars. The classifier performed well on cross-validation for all measures, with an overall accuracy slightly above 80%. Classification of the testing set using the preset features resulted in good sensitivity but poor specificity to left. Specificity was greater than chance, however. There were no correct detections of right segments. By using the selected features, 2 of 80 right hand segments were detected, resulting in sensitivity and PPV below that expected by chance. Overall, using the preset features resulted in an accuracy of 48% while classification using the selected features only reached 40% accuracy. Both these are greater than accuracy expected by chance.

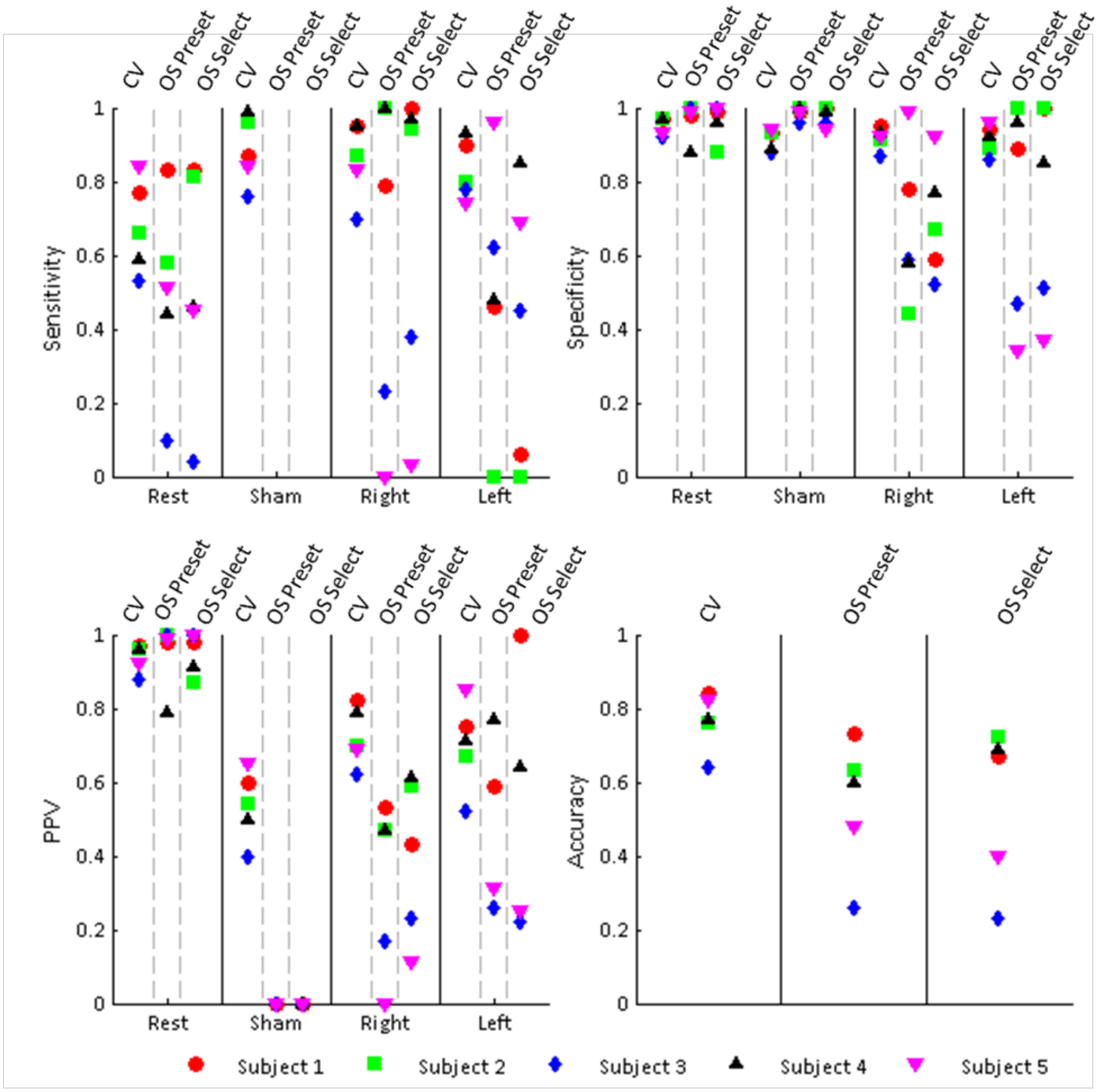


Figure 4.14. Classifier performance for rest, sham, right, and left across all subjects Classifier performance during cross validation (CV), out-of-sample validation on the testing set with preset features (OS Preset), and out-of-sample validation on the testing set with the selected features (OS Select) is shown above for all subjects. There is wide variation between the results seen for the subjects except for specificity to rest and sham and PPV for rest. In general, the best results were seen during classification of data from Subject 4, while the worst performance was seen for data from Subject 3. Poor performance was also often seen for data from Subject 5.

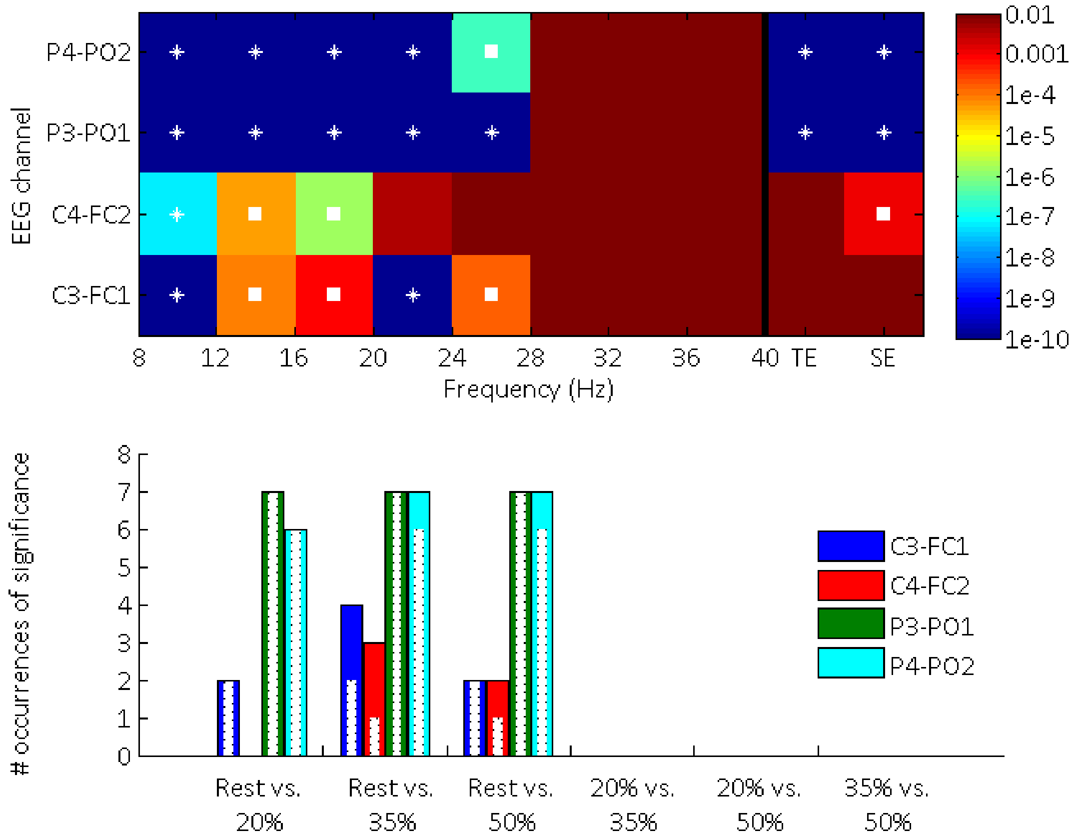


Figure 4.15. Subject 1 results of ANOVA of varying levels of force on the right hand (Top) Results are shown as the p-value from ANOVA comparing rest, 20%, 35% and 50% MVC data. To achieve an overall significance of $p < 0.05$, the Bonferroni correction was applied, and only $p < 0.0013$ was considered significant. The 16 most significant features are marked with white stars and the remaining significant features are marked with white squares. A greater number of significant features were found on P3-PO1 and P4-PO2 than on C3-FC1 and C4-FC2. Significant differences were found at 8-28 Hz band power as well as Teager energy and sample entropy. (Bottom) Post-hoc pair wise comparisons were performed for significant features from ANOVA to identify which pairs of tasks were different. The number of features which showed significance between a pair of tasks was tallied for each channel. The white bars inside the colored bars represent the 16 most significant features from ANOVA which were used in the second LDA classifier. It was found that all significance occurred between rest and the different levels of force, but no significant differences were found between different levels of force.

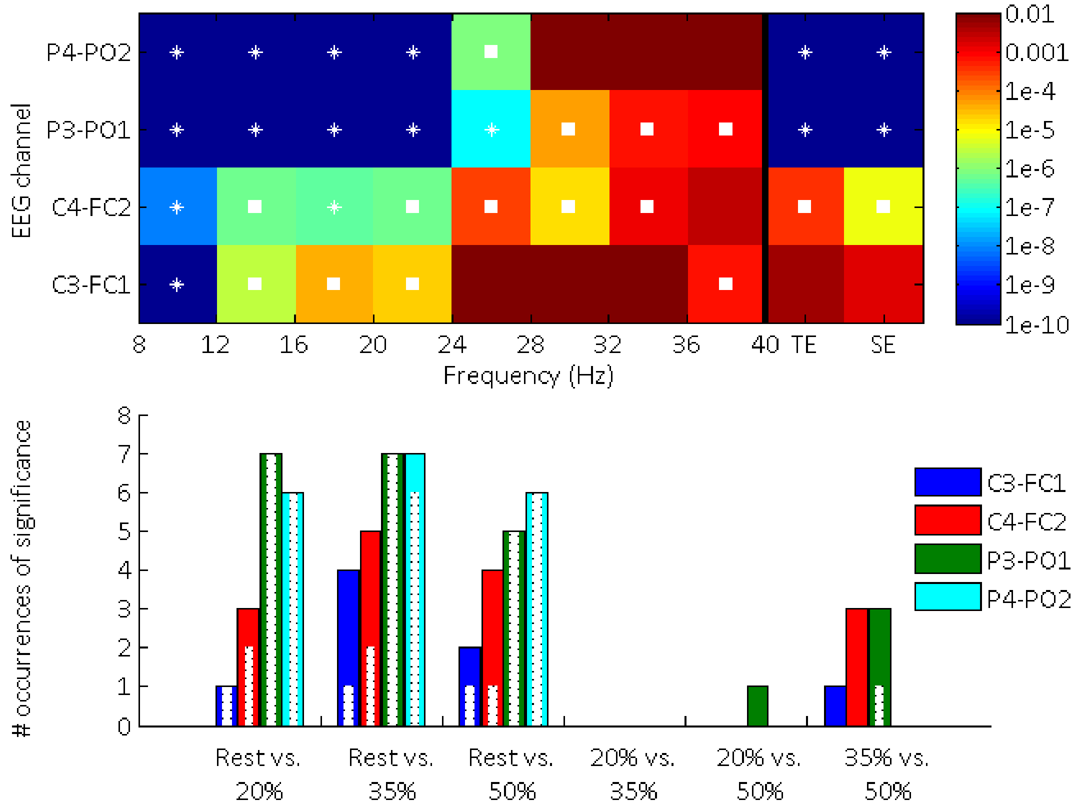


Figure 4.16. Subject 1 results of ANOVA for varying levels of force on the left hand (Top) Results are shown as the p-value from ANOVA comparing rest, 20%, 35% and 50% MVC data. To achieve an overall significance of $p < 0.05$, the Bonferroni correction was applied, and only $p < 0.0013$ was considered significant. The 16 most significant features are marked with white stars and the remaining significant features are marked with white squares. The majority of significance was contained on P3-PO1 and P4-PO2 in the band power at 8-24 Hz and in Teager energy and sample entropy. (Bottom) Post-hoc pair wise comparisons were performed for significant features from ANOVA to identify which pairs of tasks were different. The number of features which showed significance between a pair of tasks was tallied for each channel. The white bars inside the colored bars represent the 16 most significant features from ANOVA which were used in the second LDA classifier. It was found that at least one feature allowed for discrimination between each pair of tasks, with the exception of 20% vs. 35% MVC. In the 16 most significant features, only one feature contained a significant difference between two levels of force (35% and 50% MVC). The rest of the significant differences were between rest and a level of force.

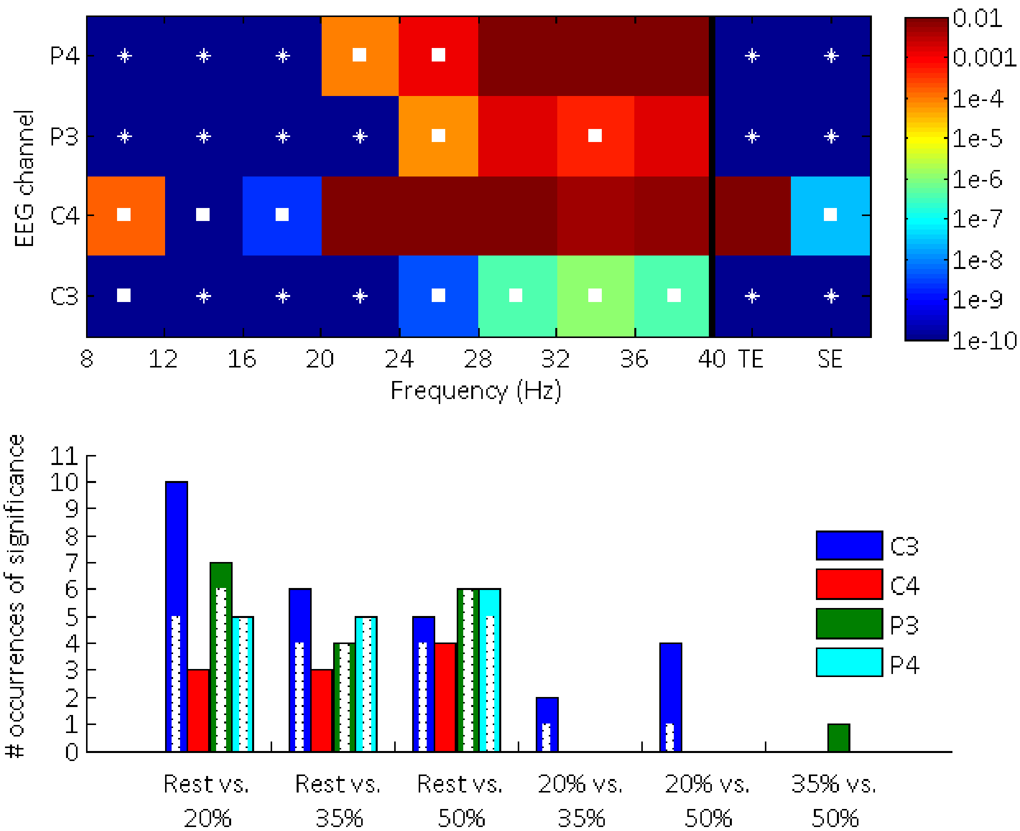


Figure 4.17. Subject 2 results of ANOVA for varying levels of force on the right hand (Top) Results are shown as the p-value from ANOVA comparing rest, 20%, 35% and 50% MVC data. To achieve an overall significance of $p < 0.05$, the Bonferroni correction was applied, and only $p < 0.0013$ was considered significant. The 16 most significant features are marked with white stars and the remaining significant features are marked with white squares. Several highly significant features were found on C3, P3, and P4. (Bottom) Post-hoc pair wise comparisons were performed for significant features from ANOVA to identify which pairs of tasks were different. The number of features which showed significance between a pair of tasks was tallied for each channel. The white bars inside the colored bars represent the 16 most significant features from ANOVA which were used in the second LDA classifier. It was found that at least one feature contained a significant difference for each of the pairs of states. The 16 most significant features contained differences between each pair of states except for 35% MVC and 50% MVC.

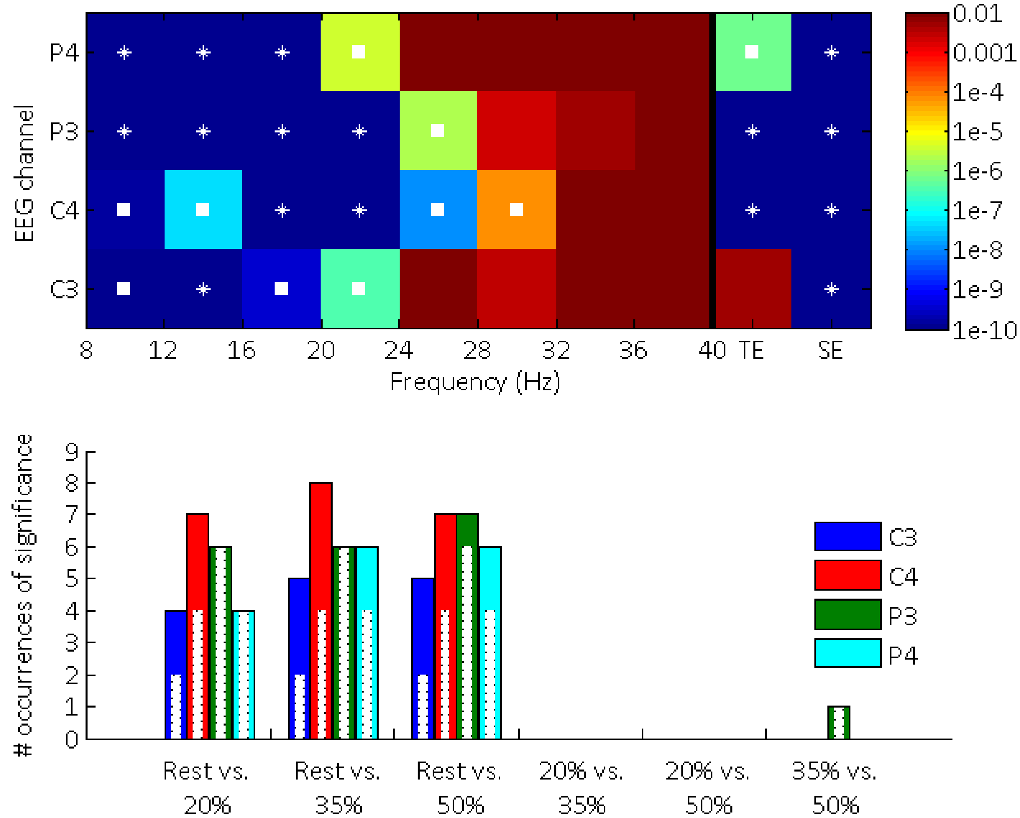


Figure 4.18. Subject 2 results of ANOVA for varying levels of force on the left hand (Top) Results are shown as the p-value from ANOVA comparing rest, 20%, 35% and 50% MVC data. To achieve an overall significance of $p < 0.05$, the Bonferroni correction was applied, and only $p < 0.0013$ was considered significant. The 16 most significant features are marked with white stars and the remaining significant features are marked with white squares. High levels of significance were found on all channels in the lower frequencies and in sample entropy. (Bottom) Post-hoc pair wise comparisons were performed for significant features from ANOVA to identify which pairs of tasks were different. The number of features which showed significance between a pair of tasks was tallied for each channel. The white bars inside the colored bars represent the 16 most significant features from ANOVA which were used in the second LDA classifier. Significant differences were most frequently found between rest and one of the three levels of force. One feature of P3, which was among the 16 most significant features of ANOVA, contained a difference between 35% and 50% MVC.

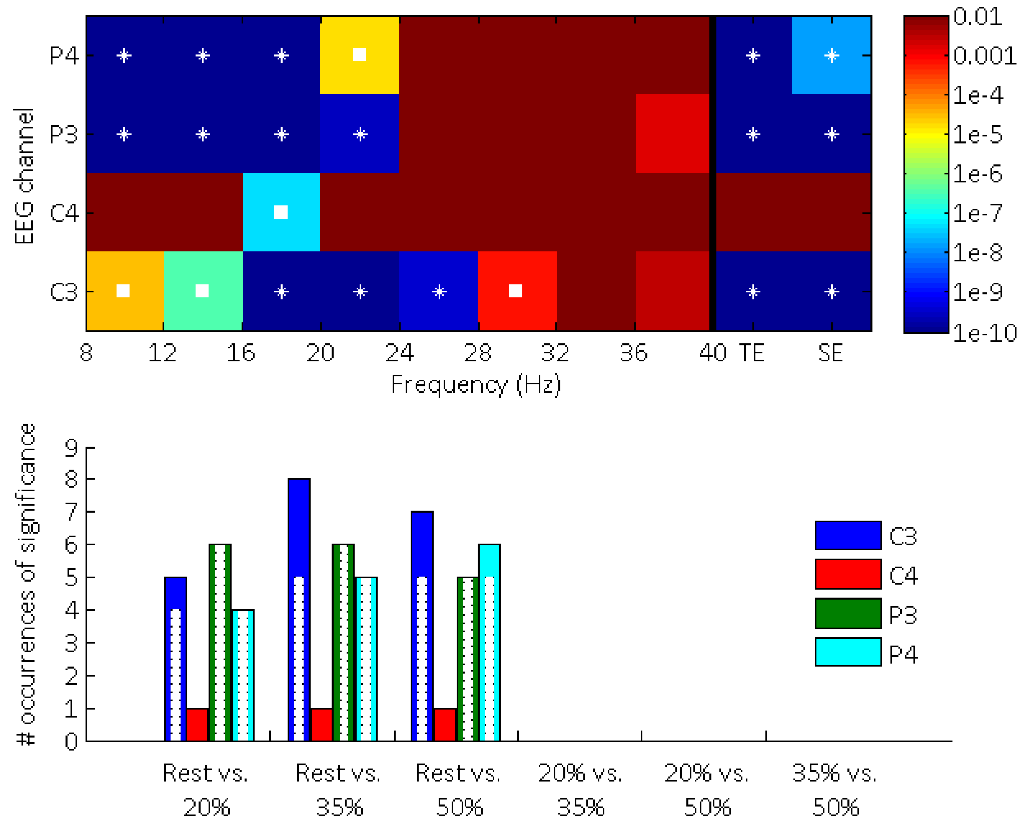


Figure 4.19. Subject 3 results of ANOVA for varying levels of force on the right hand (Top) Results are shown as the p-value from ANOVA comparing rest, 20%, 35% and 50% MVC data. To achieve an overall significance of $p < 0.05$, the Bonferroni correction was applied, and only $p < 0.0013$ was considered significant. The 16 most significant features are marked with white stars and the remaining significant features are marked with white squares. ANOVA revealed multiple significant features on C3, P3, and P4, with the significance falling in a lower frequency range on P3 and P4 than on C3. (Bottom) Post-hoc pair wise comparisons were performed for significant features from ANOVA to identify which pairs of tasks were different. The number of features which showed significance between a pair of tasks was tallied for each channel. The white bars inside the colored bars represent the 16 most significant features from ANOVA which were used in the second LDA classifier. It was found that all the significance was between rest and force pairings, with no significant differences between the different levels of force.

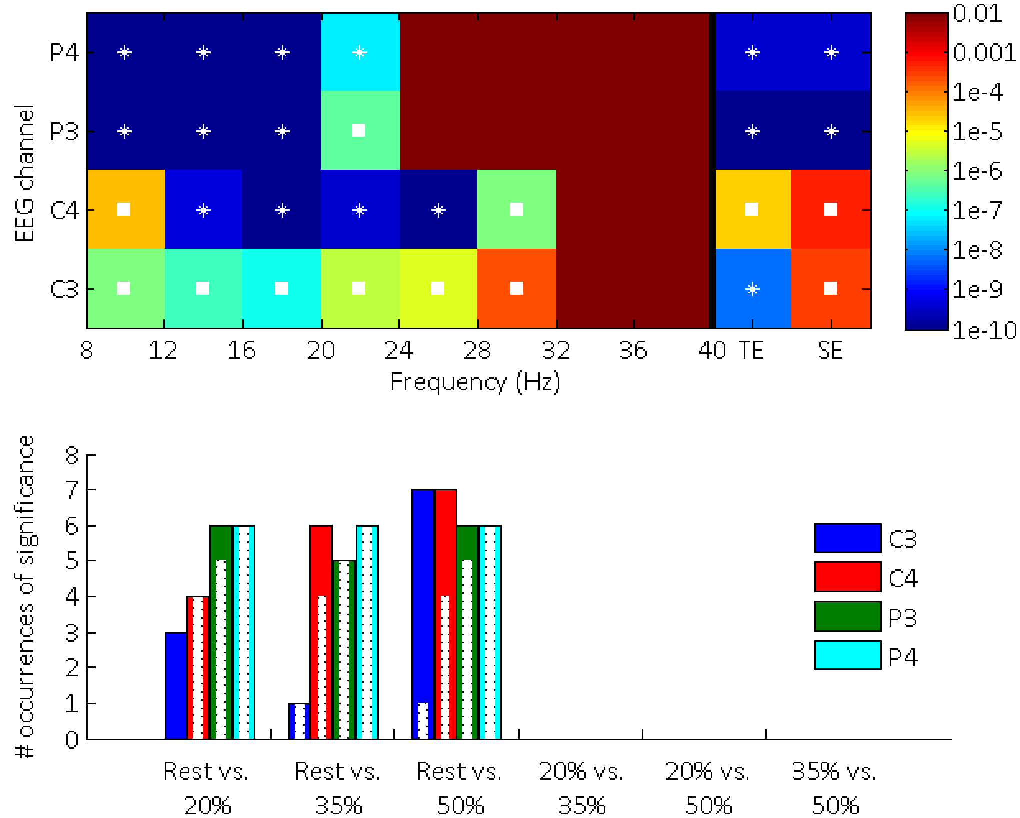


Figure 4.20. Subject 3 results of ANOVA for varying levels of force on the left hand (Top) Results are shown as the p-value from ANOVA comparing rest, 20%, 35% and 50% MVC data. To achieve an overall significance of $p < 0.05$, the Bonferroni correction was applied, and only $p < 0.0013$ was considered significant. The 16 most significant features are marked with white stars and the remaining significant features are marked with white squares. ANOVA revealed that the greatest significance was found on C4, P3, and P4. Significance was also found on C3, but not as strong as on the other three channels. (Bottom) Post-hoc pair wise comparisons were performed for significant features from ANOVA to identify which pairs of tasks were different. The number of features which showed significance between a pair of tasks was tallied for each channel. The white bars inside the colored bars represent the 16 most significant features from ANOVA which were used in the second LDA classifier. It was found that none of the features contained significant differences between levels of force, though significant differences were found between rest and each of the three levels of force.

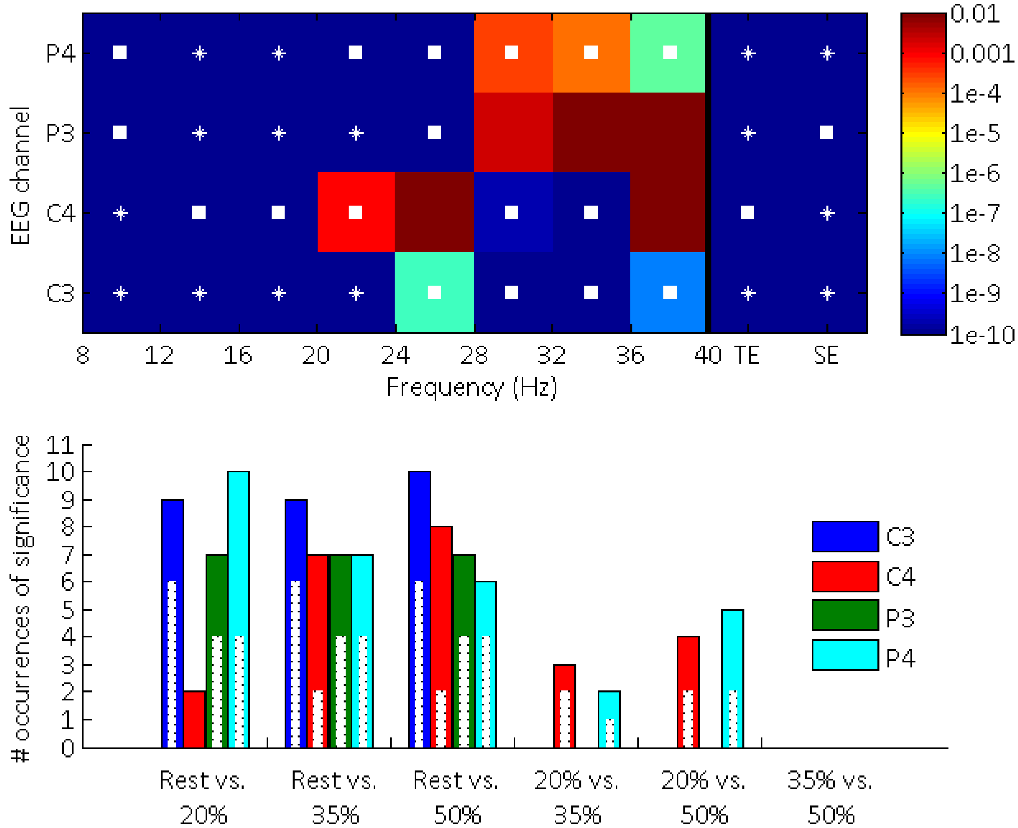


Figure 4.21. Subject 4 results of ANOVA for varying levels of force on the right hand (Top) Results are shown as the p-value from ANOVA comparing rest, 20%, 35% and 50% MVC data. To achieve an overall significance of $p < 0.05$, the Bonferroni correction was applied, and only $p < 0.0013$ was considered significant. The 16 most significant features are marked with white stars and the remaining significant features are marked with white squares. High levels of significance were found on all channels. The 16 most significant features were in Teager energy and sample entropy as well as in band power at 8-24 Hz. (Bottom) Post-hoc pair wise comparisons were performed for significant features from ANOVA to identify which pairs of tasks were different. The number of features which showed significance between a pair of tasks was tallied for each channel. The white bars inside the colored bars represent the 16 most significant features from ANOVA which were used in the second LDA classifier. Significant differences were found between all pairs of states except for 35% and 50% MVC.

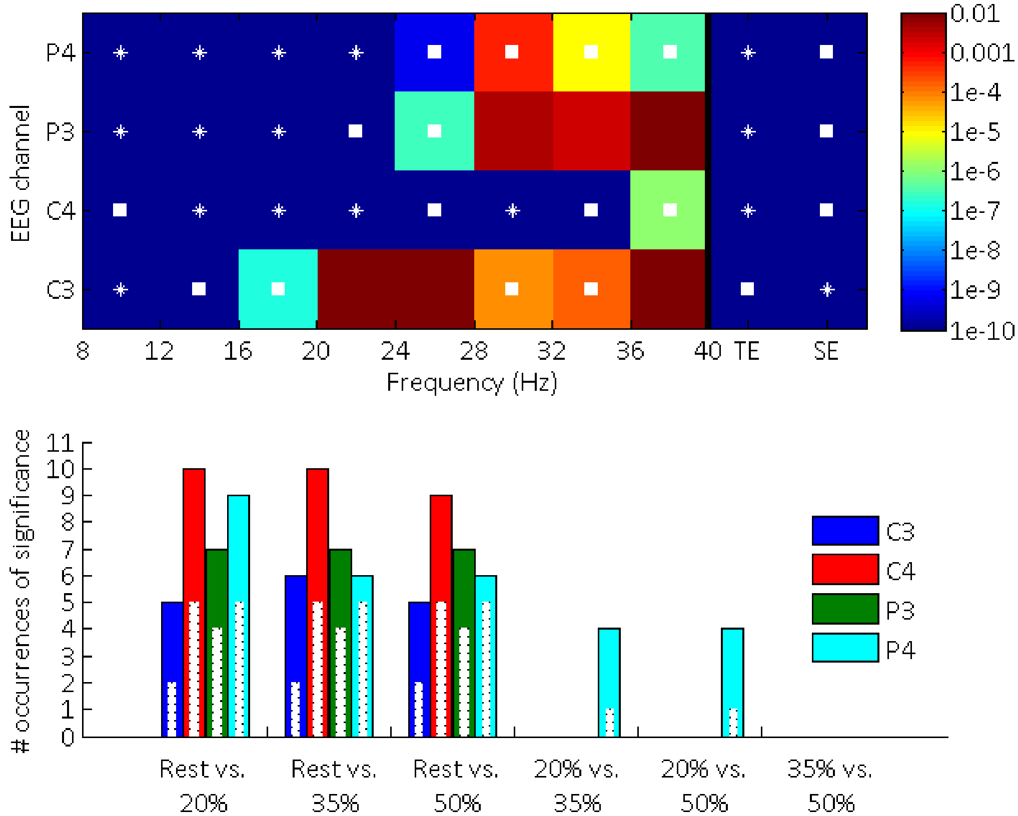


Figure 4.22. Subject 4 results of ANOVA of varying levels of force on the left hand (Top) Results are shown as the p-value from ANOVA comparing rest, 20%, 35% and 50% MVC data. To achieve an overall significance of $p < 0.05$, the Bonferroni correction was applied, and only $p < 0.0013$ was considered significant. The 16 most significant features are marked with white stars and the remaining significant features are marked with white squares. Highly significant differences were found on all channels, with the most occurring on C4. (Bottom) Post-hoc pair wise comparisons were performed for significant features from ANOVA to identify which pairs of tasks were different. The number of features which showed significance between a pair of tasks was tallied for each channel. The white bars inside the colored bars represent the 16 most significant features from ANOVA which were used in the second LDA classifier. It was found that the greatest number of features contained differences between rest and force. P4 was the only channel to contain significant differences between levels of force, with differences between 20% and 35% MVC as well as 20% and 50% MVC being found on four features each.

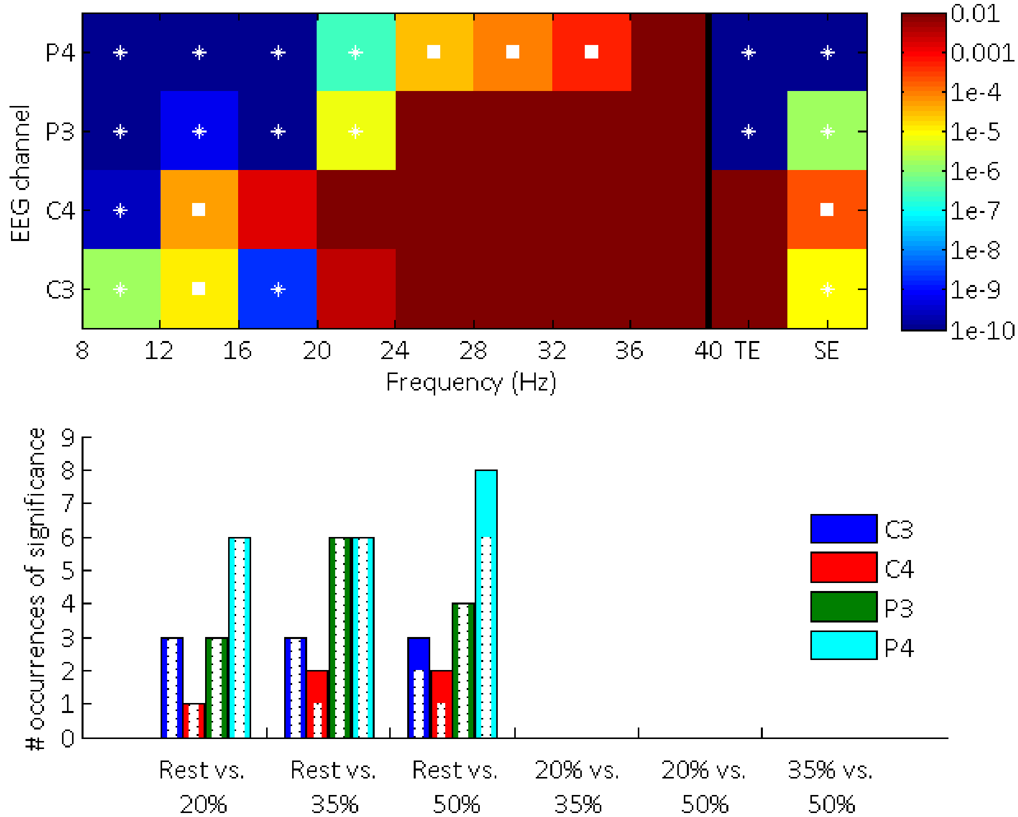


Figure 4.23. Subject 5 results of ANOVA for varying levels of force on the right hand (Top) Results are shown as the p-value from ANOVA comparing rest, 20%, 35% and 50% MVC data. To achieve an overall significance of $p < 0.05$, the Bonferroni correction was applied, and only $p < 0.0013$ was considered significant. The 16 most significant features are marked with white stars and the remaining significant features are marked with white squares. The greatest significance was found on P3 and P4 at 8-20 Hz, as well as on Teager energy and sample entropy. (Bottom) Post-hoc pair wise comparisons were performed for significant features from ANOVA to identify which pairs of tasks were different. The number of features which showed significance between a pair of tasks was tallied for each channel. The white bars inside the colored bars represent the 16 most significant features from ANOVA which were used in the second LDA classifier. It was found that significant differences only existed between rest and the three levels of force, with no significant differences between the levels of force themselves.

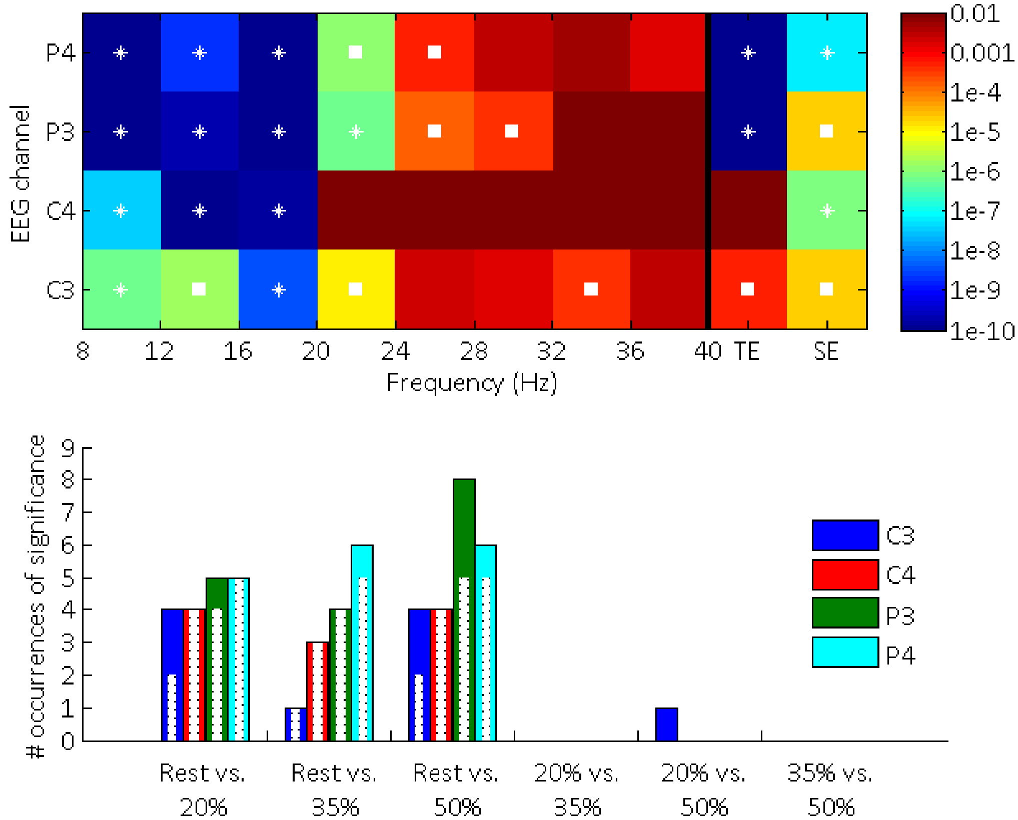


Figure 4.24. Subject 5 results of ANOVA for varying levels of force on the left hand (Top) Results are shown as the p-value from ANOVA comparing rest, 20%, 35% and 50% MVC data. To achieve an overall significance of $p < 0.05$, the Bonferroni correction was applied, and only $p < 0.0013$ was considered significant. The 16 most significant features are marked with white stars and the remaining significant features are marked with white squares. The greatest significance was found in band power at 8-20 Hz on C4, P3, and P4, and in Teager energy on P3 and P4. (Bottom) Post-hoc pair wise comparisons were performed for significant features from ANOVA to identify which pairs of tasks were different. The number of features which showed significance between a pair of tasks was tallied for each channel. The white bars inside the colored bars represent the 16 most significant features from ANOVA which were used in the second LDA classifier. It was found that most of the significance was between rest and force. One feature of C3 showed a difference between 20% and 50% MVC, but this feature was not among the 16 most significant which were later used in the subject-specific LDA classifier.

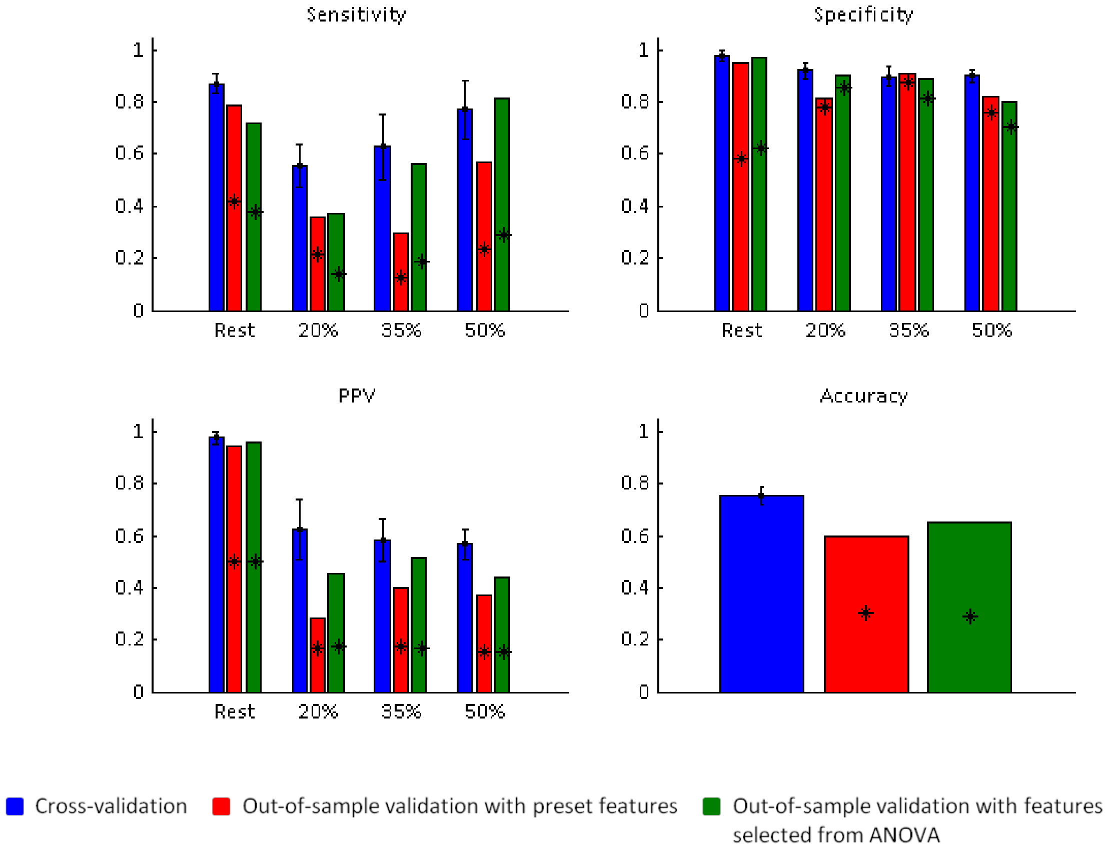


Figure 4.25. Subject 1 classification performance for varying levels of force on the right hand

Classifier performance was measured in terms of sensitivity, specificity, and positive predictive value to rest and each of the three levels of force, as well as the overall accuracy. Mean results of the 10 cross-validation trials are shown in blue, with the error bars representing the standard deviation. Also shown is the performance on the testing set using the preset features (in red) and the 16 most significant features selected from ANOVA (in green). Expected chance performance was estimated for classification of the testing set, and is marked by the black stars. During cross-validation, the best performance was achieved for rest. Application of the classifier to the testing set using the preset features gave slightly worse performance than when using the selected features. Both sets of features gave better performance than that expected by chance.

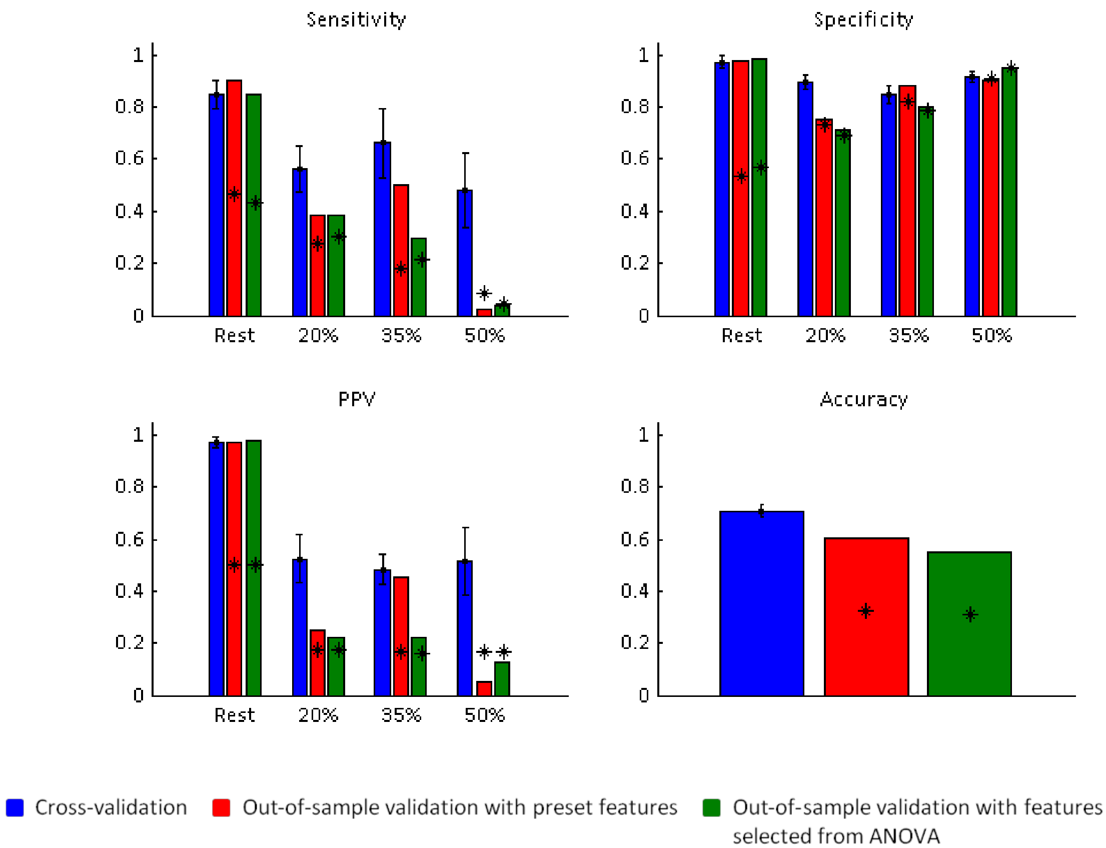


Figure 4.26. Subject 1 classification performance for varying levels of force on the left hand

Classifier performance was measured in terms of sensitivity, specificity, and positive predictive value for rest and each of the three levels of force, as well as the overall accuracy. Mean results of the 10 cross-validation trials are shown in blue, with the error bars representing the standard deviation. Also shown is the performance on the testing set using the preset features (in red) and the 16 most significant features selected from ANOVA (in green). Expected chance performance was estimated for classification of the testing set, and is marked by the black stars. Cross-validation of the training data gave the best performance. Performance during classification of the testing set was better than chance when using both sets of features, except for sensitivity and PPV for 50% MVC. Slightly better accuracy was achieved with the preset features.

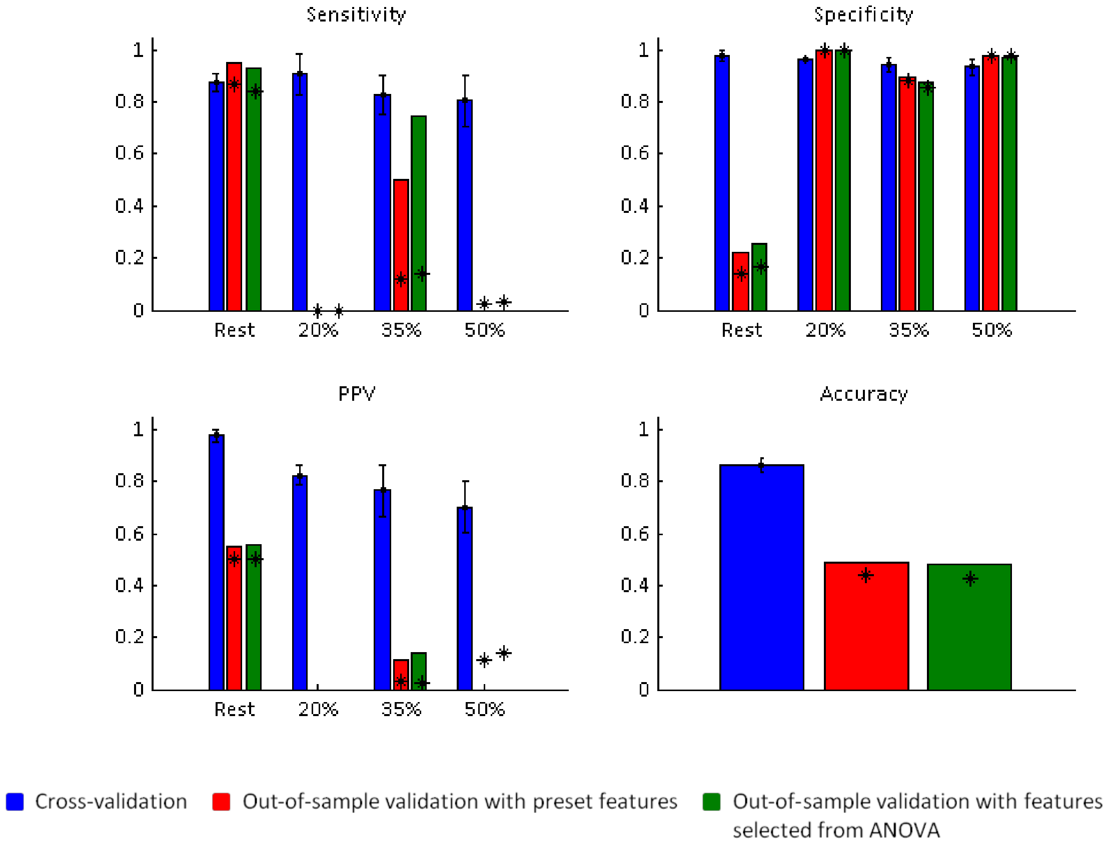


Figure 4.27. Subject 2 classification performance for varying levels of force on the right hand

Classifier performance was measured in terms of sensitivity, specificity, and positive predictive value for rest and each of the three levels of force, as well as the overall accuracy. Mean results of the 10 cross-validation trials are shown in blue, with the error bars representing the standard deviation. Also shown is the performance on the testing set using the preset features (in red) and the 16 most significant features selected from ANOVA (in green). Expected chance performance was estimated for classification of the testing set, and is marked by the black stars. Performance was good in all measures during cross-validation. Both sets of features had good sensitivity to rest in the testing set, though specificity and PPV were low. No 20% or 50% MVC segments were detected in the testing set when using either set of features, leading to zero sensitivity and PPV, but perfect specificity for both force levels. Overall accuracies were similar for both sets of features, and were only slightly higher than that expected by chance.

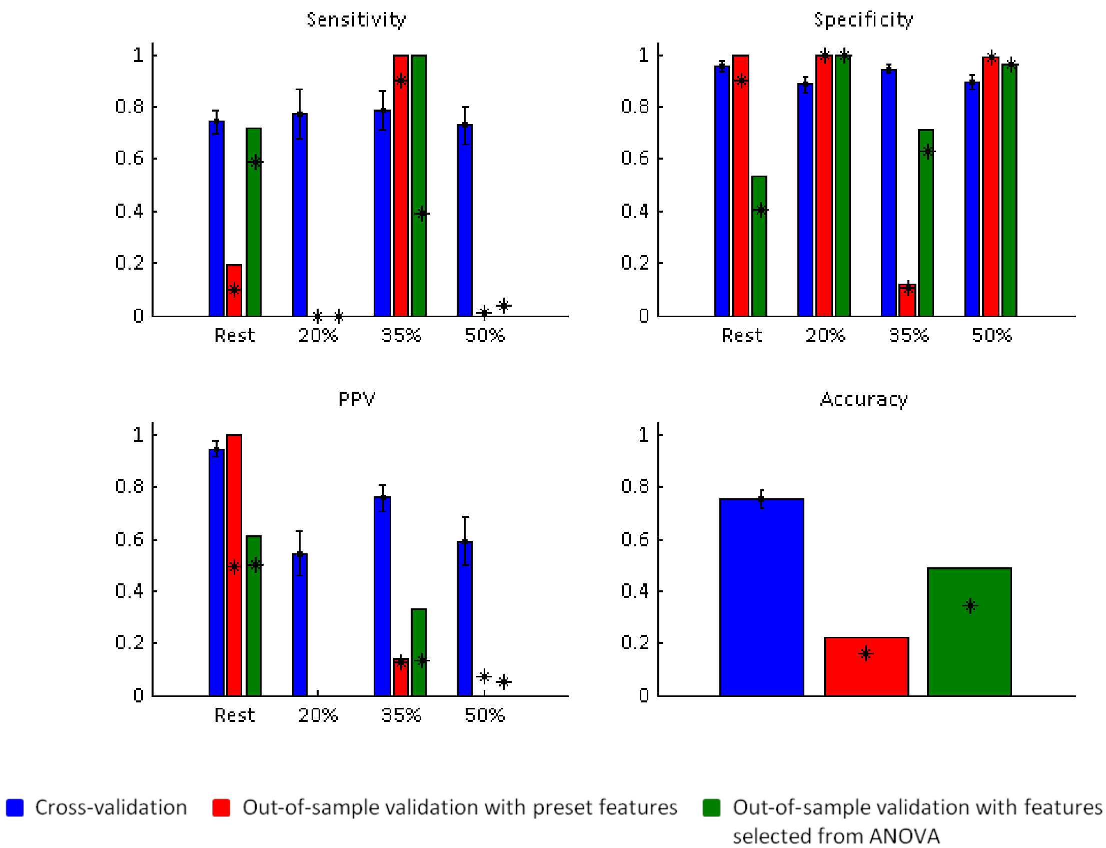


Figure 4.28. Subject 2 classification performance for varying levels of force on the left hand

Classifier performance was measured in terms of sensitivity, specificity, and positive predictive value for rest and each of the three levels of force, as well as the overall accuracy. Mean results of the 10 cross-validation trials are shown in blue, with the error bars representing the standard deviation. Also shown is the performance on the testing set using the preset features (in red) and the 16 most significant features selected from ANOVA (in green). Expected chance performance was estimated for classification of the testing set, and is marked by the black stars. Cross-validation gave the best performance overall. During classification of the testing set, both sets of features failed to detect any 20% or 50% MVC segments. Both sets of features gave performance better than or equal to that expected by chance, with the exception of sensitivity and PPV for 50% MVC. Overall higher accuracy was achieved when classifying the testing set with the selected features from ANOVA.

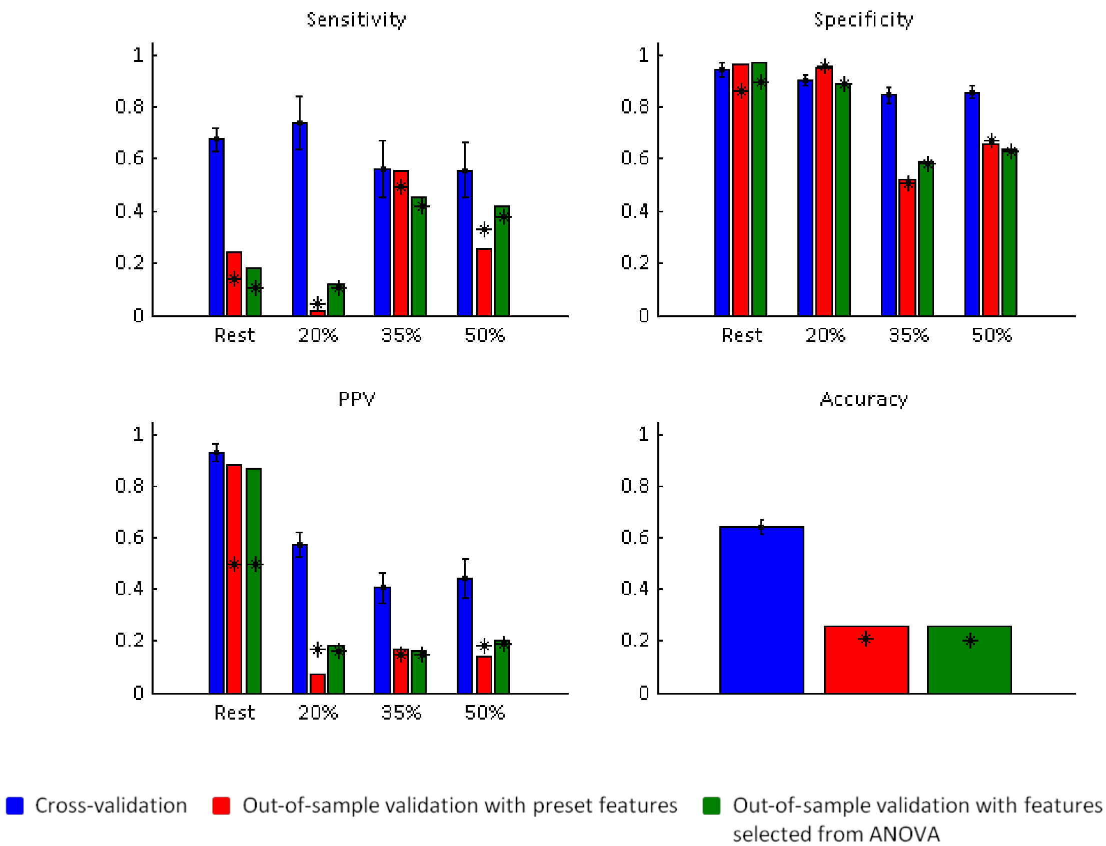


Figure 4.29. Subject 3 classification performance for varying levels of force on the right hand

Classifier performance was measured in terms of sensitivity, specificity, and positive predictive value tfor rest and each of the three levels of force, as well as the overall accuracy. Mean results of the 10 cross-validation trials are shown in blue, with the error bars representing the standard deviation. Also shown is the performance on the testing set using the preset features (in red) and the 16 most significant features selected from ANOVA (in green). Expected chance performance was estimated for classification of the testing set, and is marked by the black stars. Cross-validation performance was marginal, though it was much better than the performance on the testing set. Both sets of features used in classifying the testing set gave low sensitivity to all four states, while the PPV was low for all three force levels but not rest. Overall accuracy for classifying the testing set was slightly above that expected by chance for both sets of features.

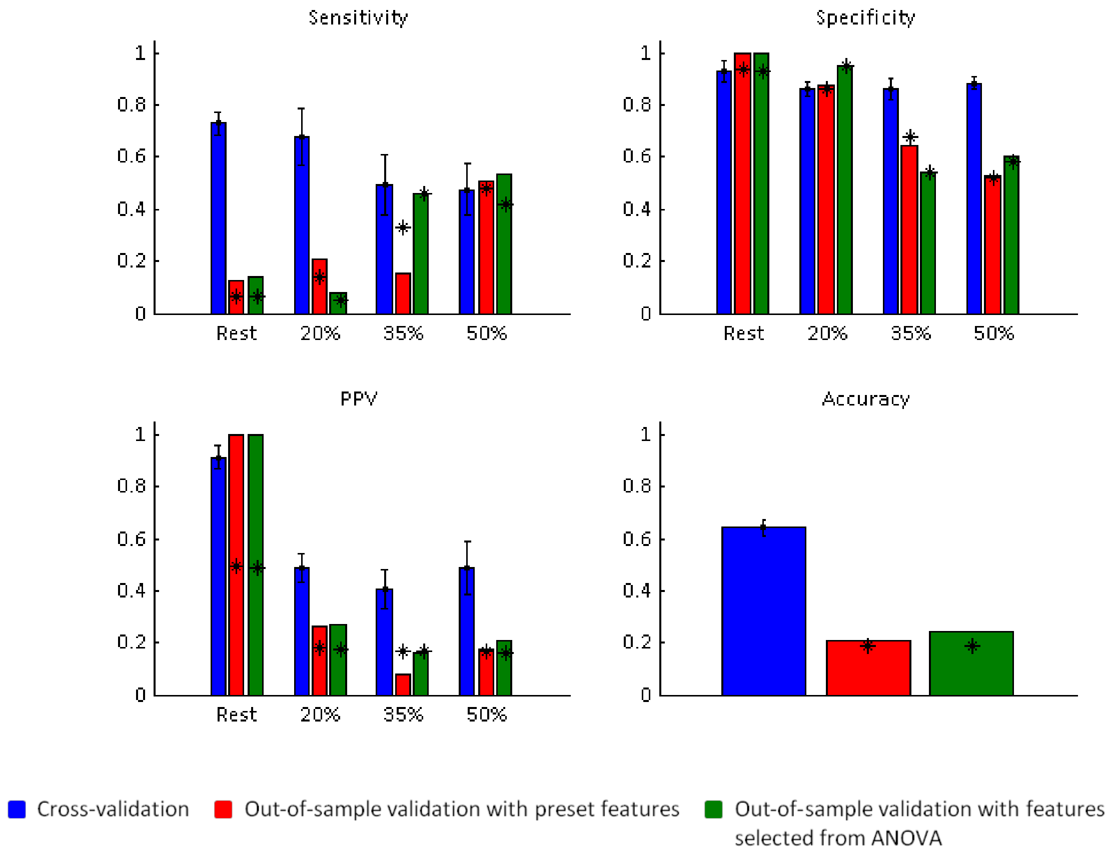


Figure 4.30. Subject 3 classification performance for varying levels of force on the left hand

Classifier performance was measured in terms of sensitivity, specificity, and positive predictive value for rest and each of the three levels of force, as well as the overall accuracy. Mean results of the 10 cross-validation trials are shown in blue, with the error bars representing the standard deviation. Also shown is the performance on the testing set using the preset features (in red) and the 16 most significant features selected from ANOVA (in green). Expected chance performance was estimated for classification of the testing set, and is marked by the black stars. Classification was generally best during cross-validation. Both sets of features gave overall low sensitivity when classifying the testing set. Most performance was still above that expected by chance, with the exception of sensitivity, specificity, and PPV for 35% MVC when using the preset features. Accuracy was slightly higher than chance for both sets of features, with the selected features giving slightly better performance.

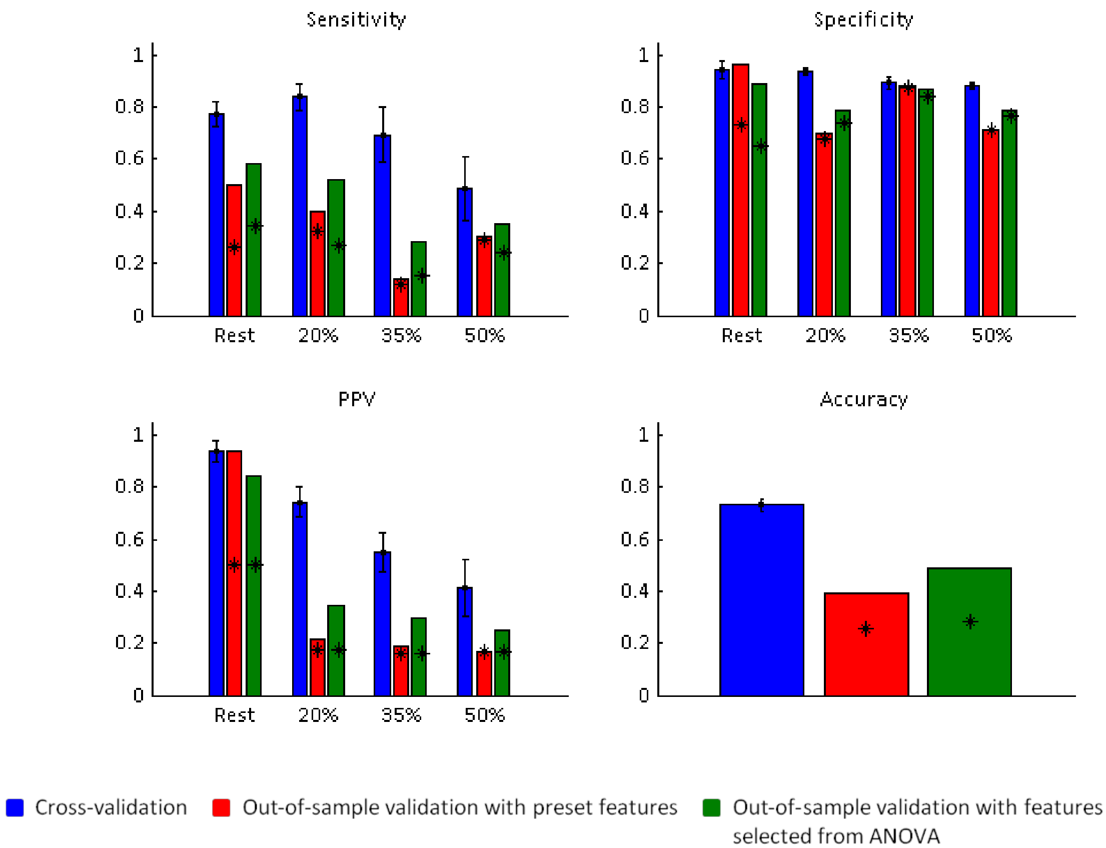


Figure 4.31. Subject 4 classification performance for varying levels of force on the right hand

Classifier performance was measured in terms of sensitivity, specificity, and positive predictive value for rest and each of the three levels of force, as well as the overall accuracy. Mean results of the 10 cross-validation trials are shown in blue, with the error bars representing the standard deviation. Also shown is the performance on the testing set using the preset features (in red) and the 16 most significant features selected from ANOVA (in green). Expected chance performance was estimated for classification of the testing set, and is marked by the black stars. Cross-validation of the training data gave the best overall results. While both sets of features gave better than chance performance when classifying the testing data, performance was slightly better with the selected features, as shown by the higher overall accuracy.

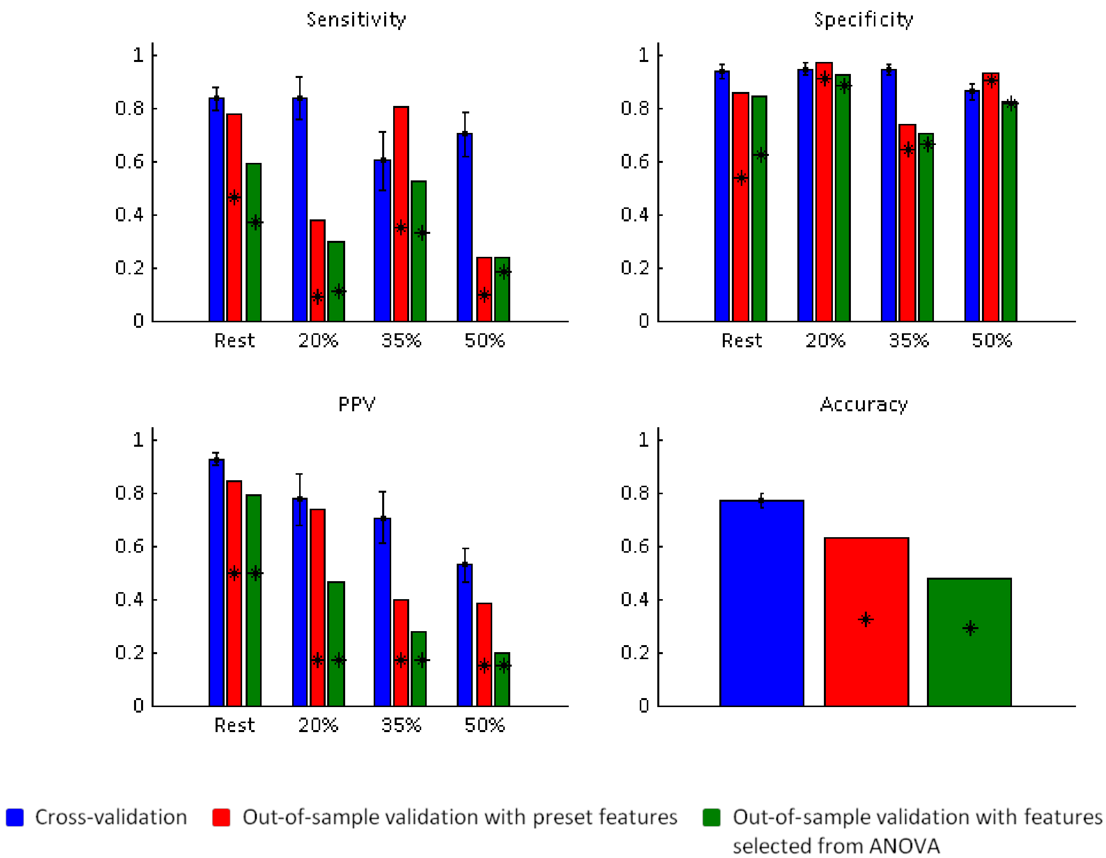


Figure 4.32. Subject 4 classification performance for varying levels of force on the left hand

Classifier performance was measured in terms of sensitivity, specificity, and positive predictive value for rest and each of the three levels of force, as well as the overall accuracy. Mean results of the 10 cross-validation trials are shown in blue, with the error bars representing the standard deviation. Also shown is the performance on the testing set using the preset features (in red) and the 16 most significant features selected from ANOVA (in green). Expected chance performance was estimated for classification of the testing set, and is marked by the black stars. Cross-validation of the classifier on the training data gave the best performance. Both sets of features gave better than chance performance when classifying the testing data, though performance was better overall when using the preset features.

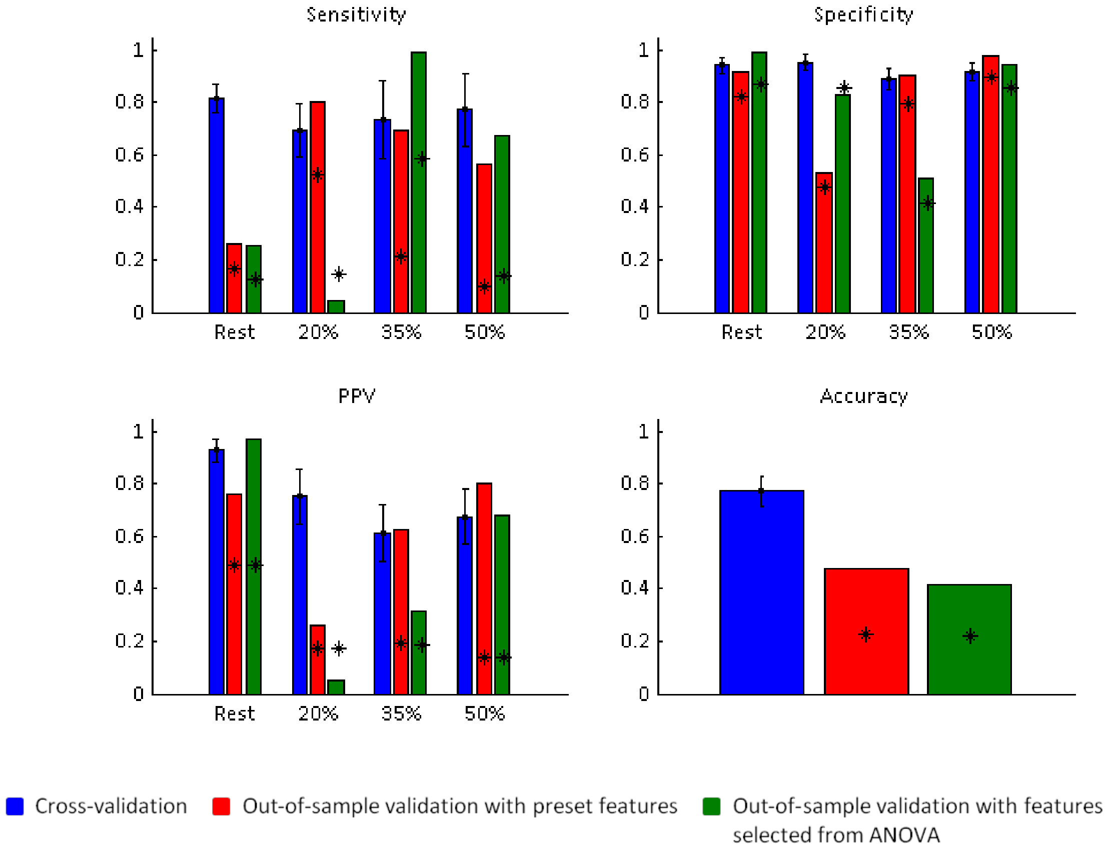


Figure 4.33. Subject 5 classification performance for varying levels of force on the right hand

Classifier performance was measured in terms of sensitivity, specificity, and positive predictive value for rest and each of the three levels of force, as well as the overall accuracy. Mean results of the 10 cross-validation trials are shown in blue, with the error bars representing the standard deviation. Also shown is the performance on the testing set using the preset features (in red) and the 16 most significant features selected from ANOVA (in green). Expected chance performance was estimated for classification of the testing set, and is marked by the black stars. Cross-validation gave the best results and most consistent performance across the four states. During classification of the testing data, very different performance was seen between the two sets of features for 20% and 35% MVC. The overall accuracy was slightly better when using the preset features than when using the selected features.

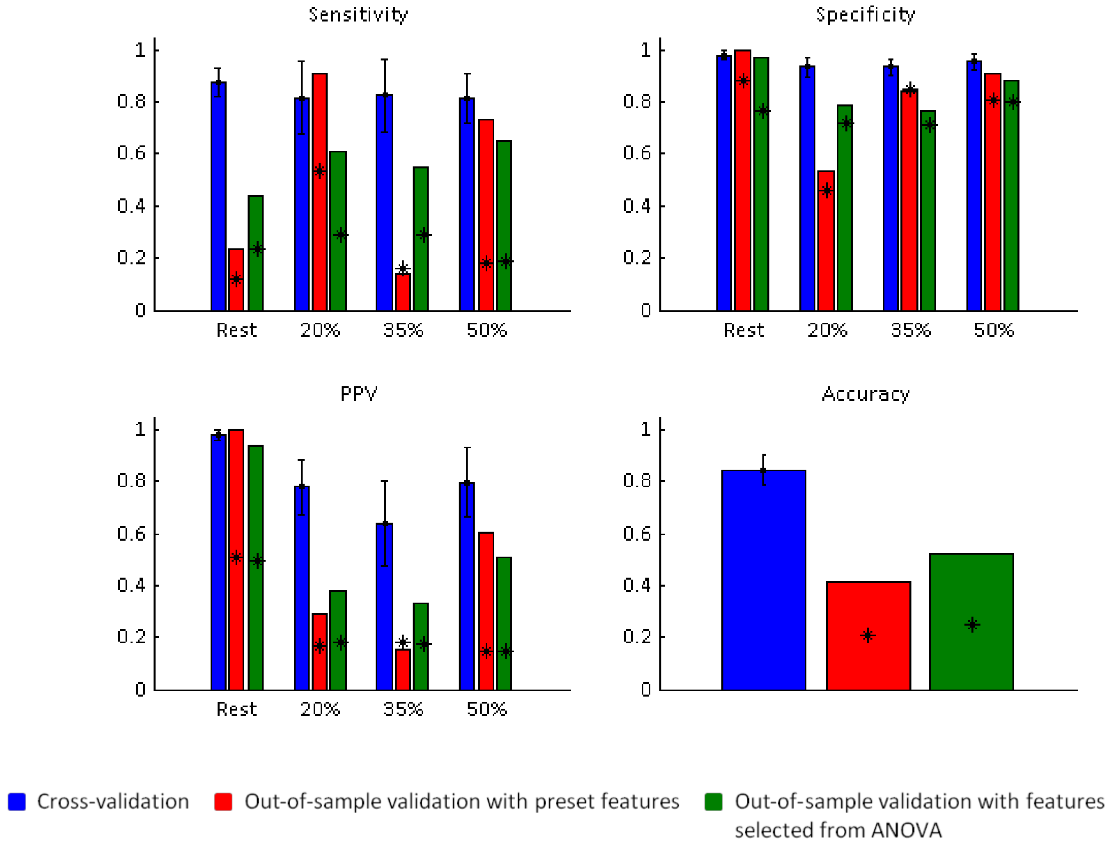


Figure 4.34. Subject 5 classification performance for varying levels of force on the left hand

Classifier performance was measured in terms of sensitivity, specificity, and positive predictive value for rest and each of the three levels of force, as well as the overall accuracy. Mean results of the 10 cross-validation trials are shown in blue, with the error bars representing the standard deviation. Also shown is the performance on the testing set using the preset features (in red) and the 16 most significant features selected from ANOVA (in green). Expected chance performance was estimated for classification of the testing set, and is marked by the black stars. Cross-validation of the training set gave very good performance with consistent results across the four states. During classification of the testing set, much variability in performance for the four states was seen between the two sets of features, particularly for rest, 20%, and 35% MVC. The selected features gave slightly higher overall accuracy than the preset features.

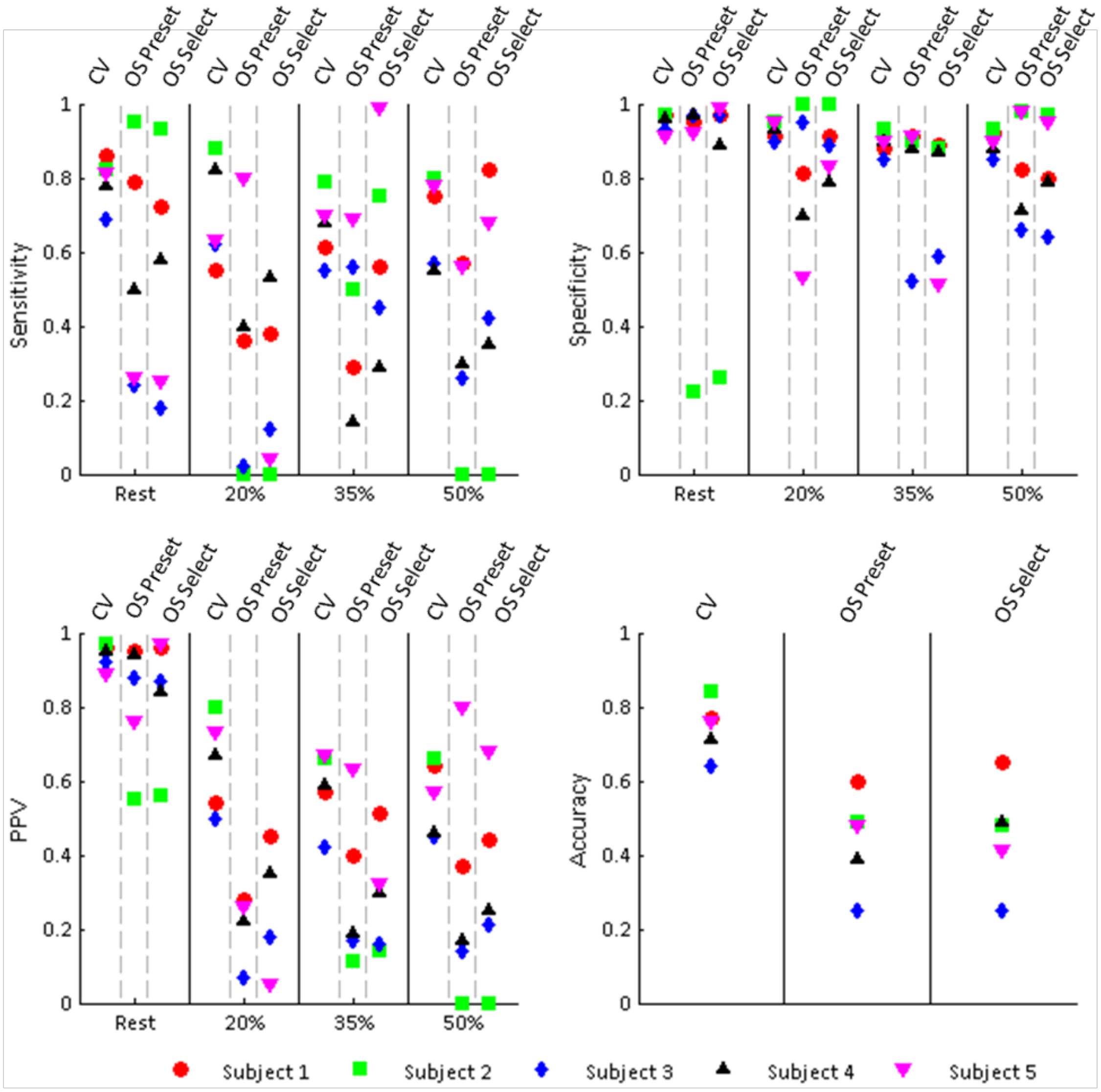


Figure 4.35. Across-subject classification performance for varying levels of force on the right hand

Classifier performance during cross validation (CV), out-of-sample validation on the testing set with preset features (OS Preset), and out-of-sample validation on the testing set with the selected features (OS Select) is shown above for all subjects. There is wide variability among the subjects for all measures of performance. Overall, the lowest performance is seen for Subject 3. However, Subjects 2 and 5 also have low performance on particular measures for certain states.

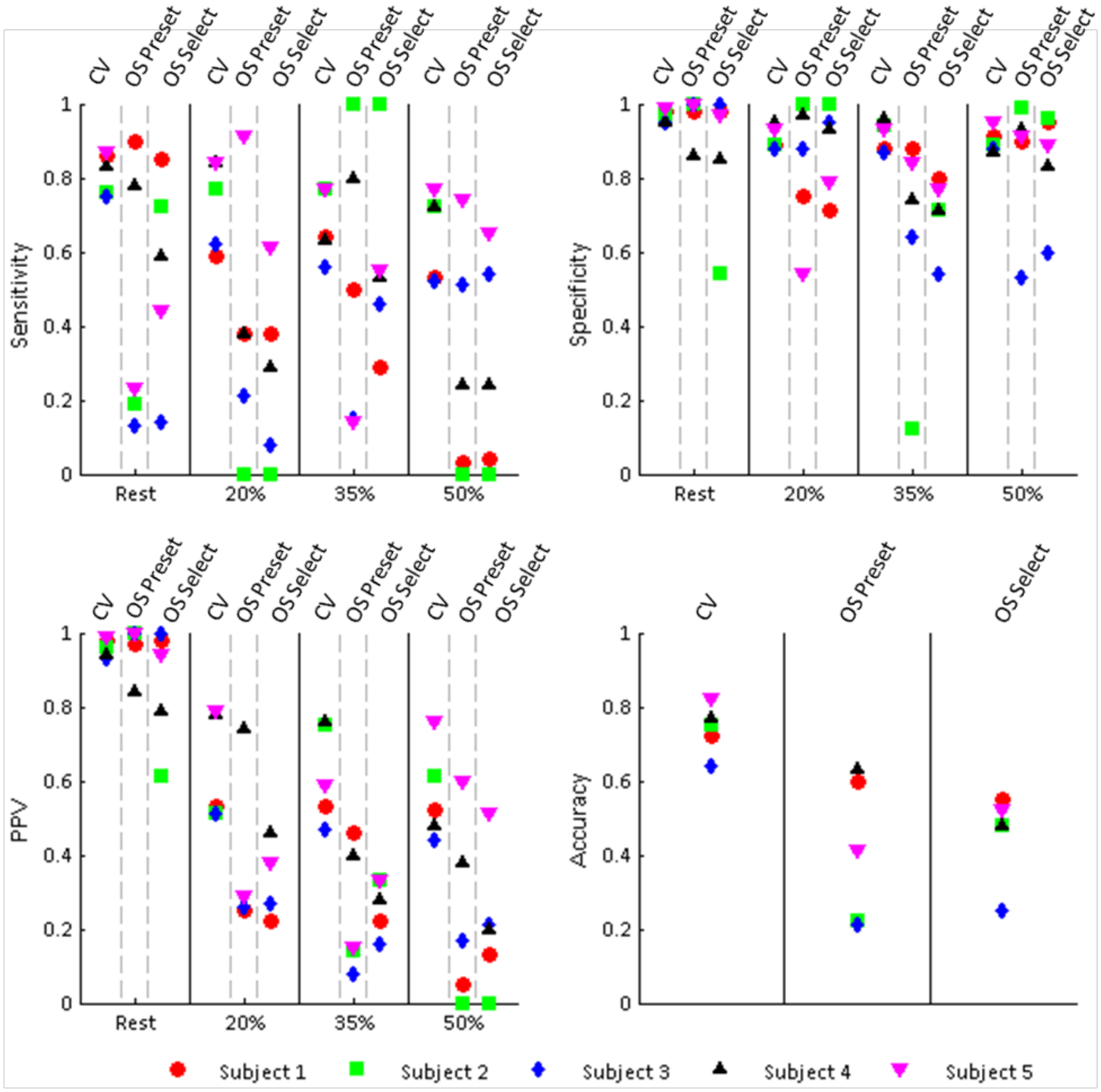


Figure 4.36. Across-subject classifier performance for varying levels of force on the left hand

Classifier performance during cross validation (CV), out-of-sample validation on the testing set with preset features (OS Preset), and out-of-sample validation on the testing set with the selected features (OS Select) is shown above for all subjects. Very wide variability is seen across subjects for sensitivity, specificity, and PPV. Specificity shows much less variation across subjects, with the low specificity to rest and 35% for Subject 2 appearing to be an outlier.

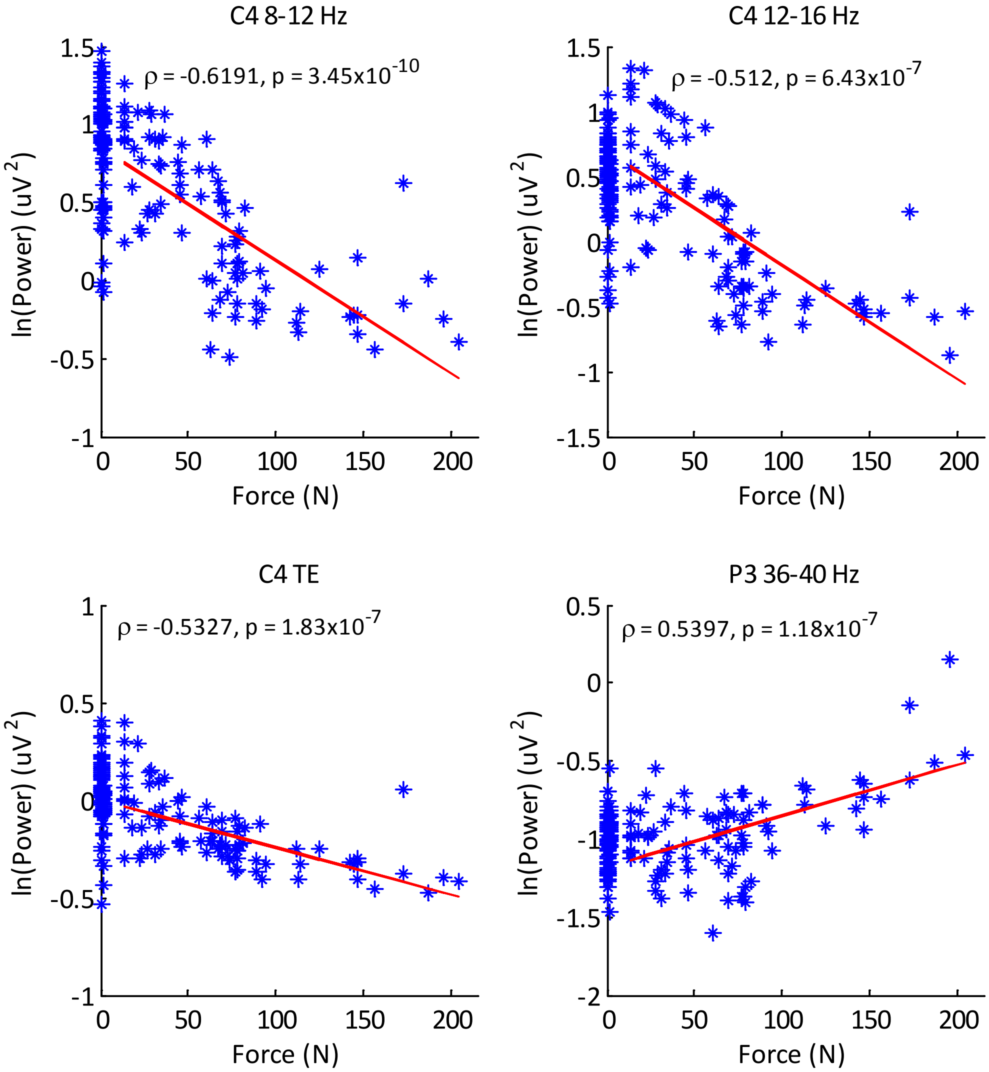


Figure 4.37. Force vs. feature value for Subject 4 right hand

The scatter plots show force vs. EEG feature values from Subject 4 with the four most significant correlations. Each blue asterisk is the feature value and the corresponding force measurement during a single task or rest segment. The red line is the first order linear equation fit to the data from during the task periods. The correlation coefficients for the EEG feature and force, as well as the significance of the correlations, are reported on each plot as ρ and p. All four plots show high variance during rest and a general decrease in variance as force increases. The three features of C4 all show a decrease as a function of force, while power at 36-40 Hz on P4 shows an increase with force.

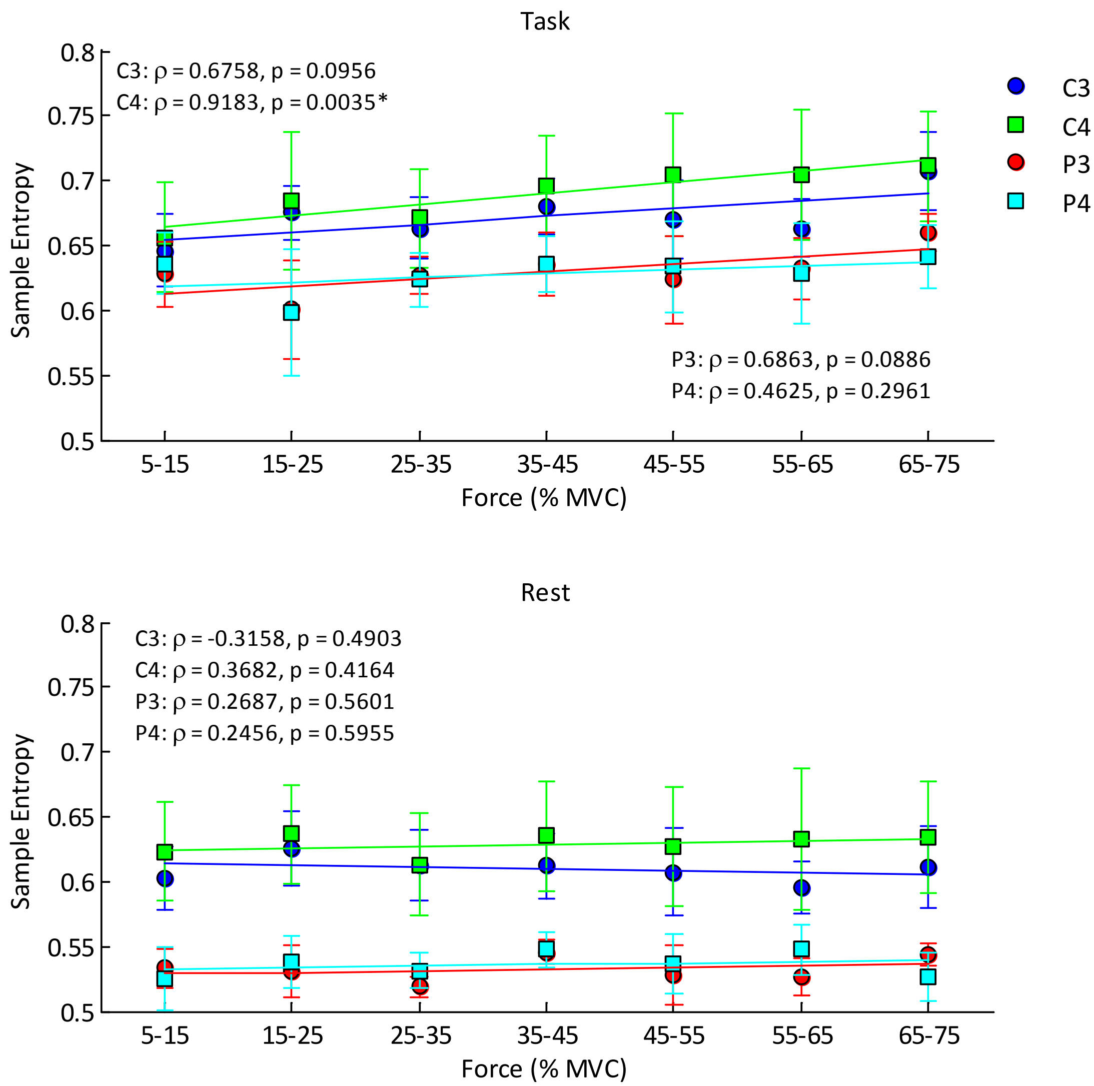
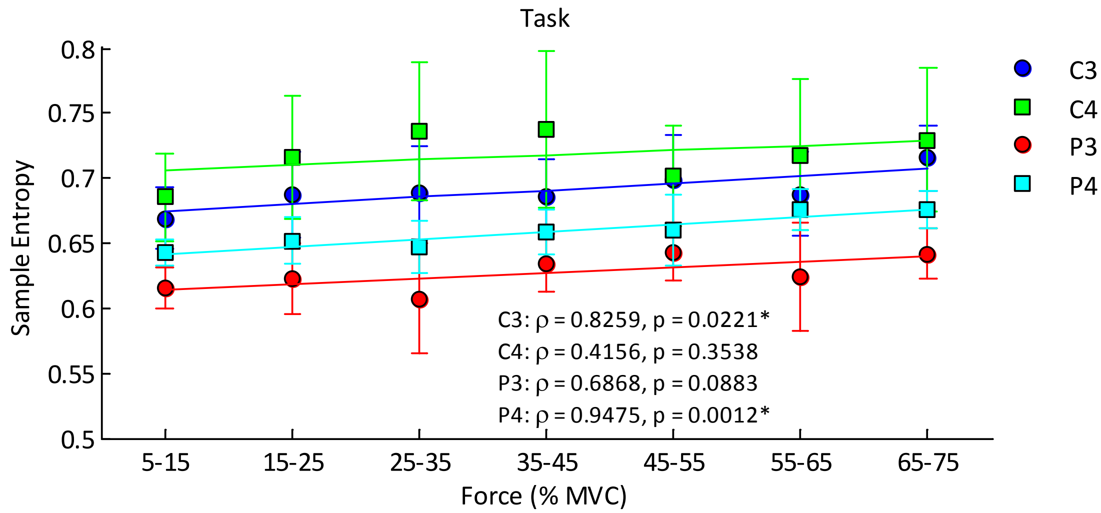


Figure 4.38. Sample entropy average across subjects during right hand runs Markers indicate the mean sample entropy of all five subjects, and the error bars represent the standard error of the mean. The lines represent the first order linear equation fit to the data from each channel. The linear correlation coefficients between sample entropy and force on each channel, and their significances, are shown on the plots as ρ and p. Any $p < 0.05$ is considered significant. During task, C4 has the best fit and is the only channel that is significant. C3 and P3 tend towards significance, while P4 shows no indication of significance. Linear regression of the rest samples shows no trend between force and sample entropy.



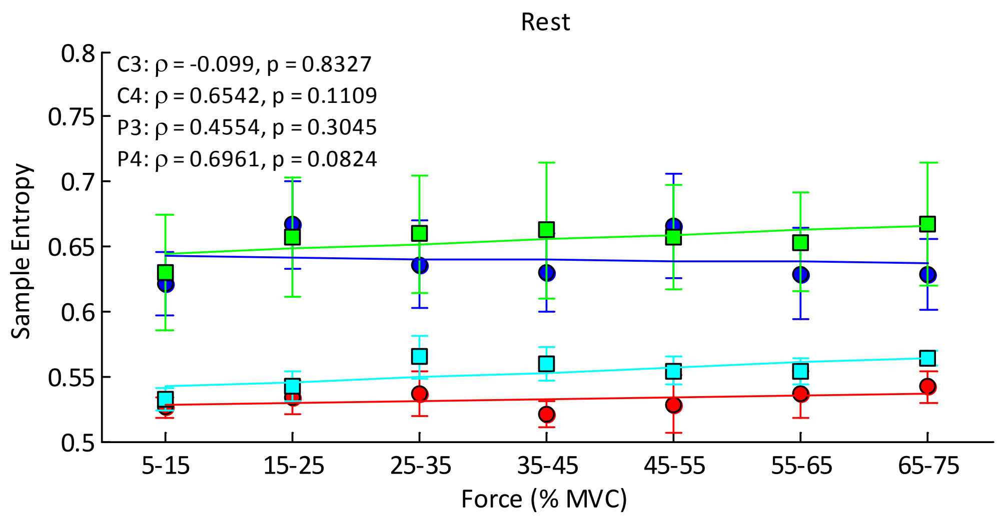


Figure 4.39. Sample entropy averaged across subjects during left hand runs Markers indicate the mean sample entropy of all five subjects, and the error bars represent the standard error of the mean. The lines represent the first order linear equation fit to the data from each channel. The linear correlation coefficients between sample entropy and force for each channel, and their significances, are shown on the plots as ρ and p. Any $p < 0.05$ is considered significant. During task, P4 has the strongest, most significant correlation. The correlation on C3 is also significant, though less so. During rest, no significant correlations were found, though P4 shows results trending toward significance.

Chapter 5: Discussion

The goals of this study were to accurately detect periods of rest, non-response (sham), and right and left hand movements from features of EEG, as well as determine whether any of these features also reflect the level of force being exerted by a subject during a motor task. We discovered that we are able to label segments with a four-state classifier with accuracy better than chance for four of five subjects who participated in the study. Performance for each of the four states varied between subjects, and was often influenced by which features were used for the classifier. Right and left hand tasks were often mistaken for one another, but rarely mistaken for rest. Rest, on the other hand, was frequently mistaken as task for some subjects. Application of the classifier to continuous signals, rather than brief, labeled sample segments, generally resulted in decreased performance. However, this is not of great concern for the proposed application to rehabilitation, which would involve cue-based activities similar to those used in the present study. We found that discrimination between 20%, 35%, and 50% MVC is fairly difficult with the set of features used, particularly when applying the classifier to data from the testing set. Again, all three levels of force were rarely mistaken for rest. Particular features of the EEG appeared to exhibit a correlation with force during task, though no single feature alone appears to provide a strong enough correlation that would predict the level of force with acceptable accuracy.

5.1 Design of visual cue and feedback interface

The visual cue was designed specifically for this study and in many ways is unlike any we have encountered in other BMI studies. The configuration of the gray dot, which increased in size as exerted force increased, in the center of the three outer circles, was chosen in order to combine the cue and feedback in a manner that wouldn’t induce ocular movement biased along a particular plane. This is in contrast to several other studies that used a bar that originated from the middle of the screen and extended to one side to indicate the amount of force being exerted or the level of SMR suppression detected. The diameter of the outer circles was limited so that overt ocular movements were not necessary for following the visual feedback. The outer circles and gray dot during zero exerted force were the same size regardless of the target force associated with a particular cue presentation in order to 1) constrain the size of the circles so that ocular movements aren't required for tasks with high target forces, and 2) keep from indicating to the subject the magnitude of the target force during randomized target force runs. In an earlier version of the cue program, the two dotted circles (representing ±10% MVC from the target force) were not present, and instead the dot was green when within the target range and red when either above or below the target range. During testing of the cue program, the cue often flickered between red and green, which was found to be distracting. Additionally, red flickering stimuli have been shown to preferentially induce seizures in photosensitive patients (41). Although subjects in our study were healthy volunteers, we considered it a possibility that red stimuli could evoke stronger visual responses to the cue than if we used a gray, monochromatic color scheme. Therefore, the dot was changed to a constant gray color and the dotted circles were added to indicate the upper and lower limits of the target force.

5.2 Inclusion of sham state

The inclusion of a sham state is a novel aspect of this project. Including a sham state is important in order to confirm that changes seen on the EEG during presentation of a visual stimulus are due to performance of a motor task and not simply a passive cortical response to the stimulus (i.e., a visual evoked potential). With the healthy volunteers of this study, there was no question of whether or not the subjects were responding to the cues because they were performing actual motor tasks, but in future clinical studies with patients who are unable to translate motor intent into a physical movement, this becomes very important. Of the five subjects in this study, three exhibited no significant changes in EEG signal power from rest when shown the sham stimulus. However, because of the small sample size of this study, it cannot be inferred that 60% of the population will show no significant changes in EEG signal power in response to a visual stimulus. Therefore, inclusion of a sham stimulus should be a part of all cued BMI systems in order to determine whether discrimination between response and non-response to a cue is possible for each subject.

The most similar study in terms of inclusion of a sham state was performed by Shibata et al. (42) in 1999. The primary purpose of this study was to investigate the changes that occur in the gamma band of the EEG during a "Go/NoGo" protocol. The experiment included a control run, during which the subjects were shown the cues from the Go/NoGo protocol but were instructed not to respond. In contrast to our study, this portion was performed after the subjects were aware of the meaning of the cues. ANOVA was performed comparing the increase in power for Go, NoGo, control Go, and control NoGo for eight subjects. The results showed that the increase in gamma band power during Go/NoGo was significantly greater than either of the control states during two periods within the first 200 ms of cue presentation. This "sham" state was most likely included in order to demonstrate the immediate (< 250 ms) differences between non-response and active inhibition during NoGo task, while our sham was included to differentiate between states of rest, non-response, and response over a more prolonged period of time.

5.3 Identification of rest, sham, right, and left

To the best of my knowledge, there are no other BMIs that attempt to classify rest, sham, left hand, and right hand movement. Most studies using cued BMIs only attempt to discriminate tasks, though there are a small number that classify rest, left hand, and right hand movement/imagery. In a 2005 study by Akrami et al. (43), a single subject was cued to perform motor imagery of circling either the right or left arm in the shoulder or to rest, with each imagery/rest segment lasting 15 s. Each task was repeated for 10 trials in a session over 2 sessions. Data from each trial was broken into 2 s windows for analysis, and a combination of features and classification algorithms were used. One set of features used included logarithmic power spectral density components of five frequency bands which were calculated for C3, C4, and Cz; P3, P4, and Pz; and F3, F4, and Fz, for a total of 15 features for each set of channels. These features are the most similar to the features used in our study. Classification with a linear classifier achieved 75% accuracy when using C3, C4, and Cz, and 71% when using P3, P4, and Pz. In the analysis of our study, the highest classification accuracy of 73% was achieved when using six features from each of the four channels for a total of twenty-four features. While our study used nine more features than Akrami's, we were classifying four classes instead of only three, hence our error rate per class is effectively smaller. It is also important to note that while in our study the rest state was sampled in the periods preceding hand grips, Akrami's rest data was sampled from a prolonged period during which the subject was only resting. It is known that the transition of the sensorimotor rhythms from an activity state to a relaxed state is not immediate (44). While the relatively short interstimulus intervals in our study may not allow complete rebounding of the sensorimotor rhythms, in theory it will enable better classification of rest when prolonged rest periods are not possible, such as in self-paced BMI control.

The importance of including a rest state is supported by the results published by Daly (30) in 2008 which used an EEG-BMI to trigger FES or a robot in rehabilitation of stroke survivors. When asked to perform actual movement or imagery of a shoulder/elbow movement, the subjects achieved fairly high accuracy in controlling their brain signals when performing the attempted/imagined movement, but had great difficulty reaching the relaxed condition. Therefore, including monitoring and feedback for not only imagery/attempted movements, but also relaxation (e.g., unclenching the hand), is needed for a BMI with neurorehabilitation applications.

The application of a four-class system is a more rigorous test of classification capability than that actually required of the BMI in the targeted clinical application, in which discrimination between right and left hand tasks will likely not be needed. Rehabilitation most often focuses on either the right or left hand, thus requiring a three-class system that discriminates between rest, non-response (sham), and either right or left hand task. Because our BMI will use a cued protocol, classification can be further broken down into two different two-class classifiers. During rest periods, triggering of the stimulator will be disabled, and the classifier will only have to determine whether the subject has reached a state of rest. Output of the classifier will be used to provide feedback, either visually on the computer monitor or verbally from the therapist, to the patient regarding their relaxation as determined from their sensorimotor rhythms. The mislabeling of rest segments as task, which occurred with some frequency in our study when classifying out-of-sample data, becomes mitigated in this case. In the event of a false detection of motor intent, the subject will receive inaccurate visual or verbal feedback but will not get stimulated. Upon presentation of the cue, the classifier must determine whether the subject is not responding (rest/sham) or is attempting a movement (right OR left). The most common error made in classification of task segments was mistaking right for left or vice versa. Very few task segments were mislabeled as rest or sham. Therefore, by removing the need to distinguish between right and left hand tasks, further improvements in performance are expected.

5.4 Identification of varying levels of force

As mentioned in section 2.6, the relationship between EEG and force is not a highly researched area. To the best of our knowledge, no studies have attempted to predict force based on EEG. In the study by Mima et al. (31) a negative linear correlation was found to exist between power in the alpha band (on channels contralateral to the hand performing the task) and force from 10-60% of MVC during a pinch task, though no details of the goodness of fit or significance of differences between force levels are reported. The use of simple band power is an advantage of their study, however, the correlation between alpha band power and force does not hold at 80% of MVC, demonstrating a weakness in the use of alpha band power. Additionally, similar results were not consistently found in our study when comparing log-scaled band power at 8-12 Hz and force from the randomized percent MVC trials.

The use of sample entropy (SE) in this study is similar to the use of the largest Lyapunov exponent (L1) by Yao et al. (33), as both features reflect the complexity of the signal. Yao's subjects performed hand grips at 20, 40, 60, and 80% of MVC, while in our study the data for the linear regression model was taken from 5 to 75% of MVC, and binned into 10% intervals for analysis. Both L1 and SE increase with force, but comparison of the two indicates that L1 is more closely correlated with force level than SE. Yao does not report the slope of the line fit to the data, but the R^2 values indicate a much better fit than that given with SE, with four out of five channels being significant, compared to one (right hand) or two (left hand) in our analysis. When comparing the averaged values of L1 for each level of force, significance was only found between 80% MVC and each of the lower levels. This, in a way, validates the difficulty of classifying the different levels of force in our study's analysis, presented in section 4.6. Rest generally had the best classification rates while the classifier had difficulty distinguishing the different levels of force. Perhaps if the protocol of our experiment had included a higher level of force, classification of that group would be better than that seen for 20%, 35%, and 50% MVC.

Although the linear regression of section 4.7 was performed on a single feature at a time, perhaps the small slopes found using SE, and the likely small slope found by Yao with L1, indicate that discrimination between different levels of force will be difficult despite a strong correlation with the EEG feature.

5.5 Possible explanations for poor performance

Very poor performance was seen for Subject 3 in classification of rest vs. task as well as rest vs. varying levels of force. Cross-validation performance was lower than for the other subjects, but the discrepancy is even more apparent during classification of the testing set, for which the accuracy was at or below chance. Raw signal amplitudes for this subject were much lower than for the four other subjects, and mu and alpha waves were difficult to detect when the subject relaxed with eyes closed. The subject’s dense hair may have contributed to poor contact with the scalp and consequently the low signal amplitudes. The low amplitude in turn may have caused the difficulty in discerning the mu and alpha waves. It is also possible that this subject would not have exhibited large mu and alpha waves even if good contact with the scalp had been made. This is sometimes referred to as BMI “illiteracy”, which was found in approximately 20 percent of subjects in a study by Popescu (45). Because the subject was performing an actual movement rather than imagery, there is no question of whether he was performing the imagery correctly, which can sometimes lead to poor BMI control.

5.6 Limitations

The primary limitation of this study was the small sample size. With data from only five subjects, there is very low statistical power to this study. This precludes our ability to perform across-subject analyses with the needed level of confidence. While most of the analysis was performed separately for each subject, the results from an individual are much less informative than the results from a sample of a population. However, many published BMI studies use small sample sizes. For instance, among all the BMI studies cited in this work, five studies, excluding the case studies, only had data from five or fewer subjects, while only three studies had more than 10 subjects. Therefore, our small sample size is not unusual in the field of BMIs.

One major drawback of this study was that segments of data containing artifacts were excluded from analysis. The number of segments remaining for analysis varied greatly between subjects. In the case of Subject 2, only eight samples from 35% MVC on the right hand were retained in the test set, leaving very few possibilities for correct detections. Additionally, the need to exclude segments with artifacts will likely limit the accuracy of the classifier during online use, particularly in situations with frequent artifacts.

Another limitation of this study was that task data were selected for analysis during periods of the least variance in force, with an automatic delay of 500 ms. Therefore, the classifier was not trained on task data prior to movement onset, leading to delayed detection of task periods seen when classifying continuous data. This will need to be investigated before the BMI can be used to trigger stimulation.

Chapter 6: Conclusions

The primary goals of this study were to:

1) Identify features of the EEG signal with significant differences when a subject is at rest, performing a right hand grip, and performing a left hand grip
2) Identify features of the EEG signal with significant differences when a subject is performing a hand grip at varying levels of effort
3) Accurately label portions of the EEG signal as rest/right hand grip/left hand grip or as the appropriate level of effort
4) Use features of the EEG signal to accurately model varying levels of effort

To achieve these goals, EEG was collected from five subjects while performing cued hand grip tasks with a dynamometer. Each subject participated in three sessions (one session/week). Band power and two nonlinear features were calculated for two central and two parietal EEG channels, and analysis (ANOVA) of these features from sessions 1 and 2 was performed to identify which ones contained significant differences between tasks (rest, sham, right, and left) and varying levels of force (rest, 20%, 35%, and 50% MVC). The 16 most significant features identified by ANOVA, as well as 24 preset features (8-10 Hz, 10-12 Hz, 16-24 Hz, 26-30 Hz band power, Teager energy, and sample entropy from each of the four channels), were used to classify the data, training the classifier on data from sessions 1 and 2, and testing its performance on session 3. To model varying levels of effort, sample entropy from each subject during runs with varying force (5-75% MVC) was binned into 10% MVC wide bins and averaged. A grand average across subjects was then calculated and a linear model fit to it.

Classification of rest, sham, right, and left data resulted in better accuracy than classification of the different levels of force. Cross-validation of the classifier on sessions 1 and 2 yielded much better results than classification of session 3. However, with the exception of Subject 3, accuracy was much better than chance for all classifications of the test data. A significant relationship between force and sample entropy was found on C4 during right hand tasks and on C3 and P4 during left hand tasks. However, the slopes of the lines fit to the data were very shallow, indicating that an accurate approximation of force would be difficult using sample entropy alone. Also, the low statistical power due to the small sample size of this study prevents us from making definite conclusions about our results.

One limitation of this study is that segments containing artifacts were excluded from analysis. This will likely decrease the performance of the classifier during online tasks, during which either artifact-containing segments will be fed to the classifier, which may result in misclassifications, or they will be excluded, leading to periods with no control signal to the BMI. Also leading to classification difficulties during online control is the fact that task data was taken during periods of the least variability in force, with an automatic delay of 500 ms from cue onset to sampling. The exclusion of data during motor planning and early stages of execution from analysis fails to address the need for early detection of intent for triggering stimulation.

Before this study can be moved to a clinical setting, several issues will first need to be addressed. The first issue that will need to be addressed is the small sample size. Ideally, more subjects will be enrolled in this study in order to increase the statistical power. Another primary issue will be investigation of easily extracted features that are less affected by presence of artifacts in order to create a more robust system for online classification. The early detection of intent will also need to be explored, with possible solutions including calculation of band power in smaller windows, or the employment of a more robust classifier (e.g. support vector machines) or dynamical models (e.g., a hidden Markov Model) for classification.

Appendix

List of commonly used abbreviations

Abbreviation	Stands for
ANOVA	analysis of variance
BMI	brain-machine interface
ECoG	electrocorticography
EEG	electroencephalogram
EMG	electromyogram
FES	functional electrical stimulation
FET	functional electrical therapy
LDA	linear discriminant analysis
MEG	magnetoencephalography
MEP	motor-evoked potential
MRCP	movement-related cortical potential
MVC	maximal voluntary contraction
PNS	peripheral nerve stimulation
PPV	positive predictive value
SCI	spinal cord injury
SMR	sensorimotor rhythm
TMS	transcranial magnetic stimulation
VEP	visual-evoked potential

References

1. Dobkin BH. 2009. Motor rehabilitation after stroke, traumatic brain, and spinal cord injury: common denominators within recent clinical trials. *Current opinion in neurology* 22:563-9

2. Backus D. 2010. Exploring the Potential for Neural Recovery After Incomplete Tetraplegia Through Nonsurgical Interventions. *PM&R* 2:S279-S85

3. Beekhuizen KS, Field-Fote EC. 2005. Massed practice versus massed practice with stimulation: effects on upper extremity function and cortical plasticity in individuals with incomplete cervical spinal cord injury. *Neurorehabilitation and Neural Repair* 19:33-45

4. Popovic MR, Thrasher TA, Adams M, Takes V, Zivanovic V, Tonack MI. 2005. Functional electrical therapy: retraining grasping in spinal cord injury. *Spinal cord* 44:143-51

5. Popovic MR, Thrasher TA, Zivanovic V, Takaki J, Hajek V. 2005. Neuroprosthesis for Retraining Reaching and Grasping Functions in Severe Hemiplegic Patients. *Neuromodulation: Technology at the Neural Interface* 8:58-72

6. Tarkka IM, Pitkänen K, Popovic DB, Vanninen R, Könönen M. 2011. Functional electrical therapy for hemiparesis alleviates disability and enhances neuroplasticity. *The Tohoku journal of experimental medicine* 225:71

7. Kaelin-Lang A, Luft AR, Sawaki L, Burstein AH, Sohn YH, Cohen LG. 2002. Modulation of human corticomotor excitability by somatosensory input. *The Journal of Physiology* 540:623

8. Hamdy S, Rothwell JC, Aziz Q, Singh KD, Thompson DG. 1998. Long-term reorganization of human motor cortex driven by short-term sensory stimulation. *Nat. Neurosci.* 1:64-8

9. Sawaki L, Wu CWH, Kaelin-Lang A, Cohen LG. 2006. Effects of somatosensory stimulation on use-dependent plasticity in chronic stroke. *Stroke* 37:246-7

10. Beekhuizen KS, Field-Fote EC. 2008. Sensory Stimulation Augments the Effects of Massed Practice Training in Persons With Tetraplegia. *Archives of Physical Medicine and Rehabilitation* 89:602-8

11. Stefan K, Kunesch E, Cohen LG, Benecke R, Classen J. 2000. Induction of plasticity in the human motor cortex by paired associative stimulation. *Brain* 123:572-84

12. Rothschild RM. 2010. Frontiers: Neuroengineering Tools/Applications for Bidirectional Interfaces, Brain–Computer Interfaces, and Neuroprosthetic Implants–A Review of Recent Progress. *Frontiers in Neuroengineering* 3

13. Daly JJ, Wolpaw JR. 2008. Brain-computer interfaces in neurological rehabilitation. *The Lancet Neurology* 7:1032-43

14. Wang Y, Makeig S. 2009. Predicting Intended Movement Direction Using EEG from Human Posterior Parietal Cortex. In *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience*, ed. D Schmorrow, I Estabrooke, M Grootjen, 5638:437-46: Springer Berlin Heidelberg. Number of 437-46 pp.

15. Shibasaki H, Barrett G, Halliday E, Halliday AM. 1980. Components of the movement-related cortical potential and their scalp topography. *Electroencephalography and clinical neurophysiology* 49:213-26

16. Deecke L, Grozinger B, Kornhuber HH. 1976. Voluntary finger movement in man - cerebral potentials and theory. *Biological Cybernetics* 23:99-119

17. Niazi IK, Jiang N, Tiberghien O, Nielsen JF, Dremstrup K, Farina D. 2011. Detection of movement intention from single-trial movement-related cortical potentials. *Journal of neural engineering* 8:066009

18. Bai O, Rathi V, Lin P, Huang DD, Battapady H, et al. 2011. Prediction of human voluntary movement before it occurs. *Clinical Neurophysiology* 122:364-72

19. McFarland DJ, Miner LA, Vaughan TM, Wolpaw JR. 2000. Mu and beta rhythm topographies during motor imagery and actual movements. *Brain Topography* 12:177-86

20. Pfurtscheller G, Neuper C, Flotzinger D, Pregenzer M. 1997. EEG-based discrimination between imagination of right and left hand movement. *Electroencephalography and Clinical Neurophysiology* 103:642-51

21. Lotte F, Congedo M, Lécuyer A, Lamarche F, Arnaldi B. 2007. A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of neural engineering* 4:R1

22. Blankertz B, Dornhege G, Krauledat M, Müller KR, Curio G. 2007. The non-invasive Berlin Brain-Computer Interface: Fast acquisition of effective performance in untrained subjects. *NeuroImage* 37:539-50

23. Huang D, Lin P, Fei DY, Chen X, Bai O. 2009. Decoding human motor activity from EEG single trials for a discrete two-dimensional cursor control. *Journal of neural engineering* 6:046005

24. Yang R, Gray DA, Ng BW, He M. Comparative analysis of signal processing in brain computer interface, *2009*:580-5: IEEE

25. Pfurtscheller G, Guger C, Müller G, Krausz G, Neuper C. 2000. Brain oscillations control hand orthosis in a tetraplegic. *Neuroscience letters* 292:211-4

26. Pfurtscheller G. 2003. [] Thought'-control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia. *Neuroscience letters* 351:33-6

27. Daly JJ, Cheng R, Rogers J, Litinas K, Hrovat K, Dohring M. 2009. Feasibility of a new application of noninvasive brain computer interface (BCI): a case study of training for recovery of volitional motor control after stroke. *Journal of Neurologic Physical Therapy* 33:203

28. Soekadar SR, Witkowski M, Mellinger J, Ramos A, Birbaumer N, Cohen LG. 2011. ERD-Based Online Brain–Machine Interfaces (BMI) in the Context of Neurorehabilitation: Optimizing BMI Learning and Performance. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* 19:542-9

29. Niazi IK, Mrachacz-Kersting N, Jiang N, Nielsen K, Farina D. 2012. Peripheral Electrical Stimulation Triggered by Self-Pace Detection of Motor Intention Enhances Motor Evoked Potentials. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* PP:1-

30. Daly JJ, Cheng R, Hrovat K, Rogers JM, Litinas K, Dohring ME. 2008. Development and Testing of Non-Invasive BCI+ FES/Robot Sys-tem For Use in Motor Re-Learning After Stroke. *Stroke* 1314:c1339plas

31. Mima T, Simpkins N, Oluwatimilehin T, Hallett M. 1999. Force level modulates human cortical oscillatory activities. *Neuroscience letters* 275:77-80

32. Liu J, Yang Q, Yao B, Brown R, Yue G. 2005. Linear correlation between fractal dimension of EEG signal and handgrip force. *Biological cybernetics* 93:131-40

33. Yao B, Liu JZ, Brown RW, Sahgal V, Yue GH. 2009. Nonlinear features of surface EEG showing systematic brain signal adaptations with muscle force and fatigue. *Brain Research* 1272:89-98

34. Davis J. 1959. Manual of Surface Electromyography, Wright Patterson Air Force Base, Ohio

35. Tao W, Jie D, Bin H. Classification of motor imagery EEG patterns and their topographic representation. *Proc. Engineering in Medicine and Biology Society, 2004. IEMBS '04. 26th Annual International Conference of the IEEE, 2004,* 2:4359-62:

36. Nunez PL, Srinavasan R. 2006. *Electric fields of the brain*. New York; Oxford: Oxford University Press

37. Frigge M, Hoaglin DC, Iglewicz B. 1989. Some Implementations of the Boxplot. *The American Statistician* 43:50-4

38. Thomson DJ. 1982. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE* 70:1055-96

39. Yadav R, Agarwal R, Swamy MNS. Detection of epileptic seizures in stereo-EEG using frequency-weighted energy. *Proc. Circuits and Systems, 2007. MWSCAS 2007. 50th Midwest Symposium on, 2007*:77-80:

40. Richman JS, Moorman JR. 2000. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology - Heart and Circulatory Physiology* 278:H2039-H49

41. Parra J, da Silva FHL, Stroink H, Kalitzin S. 2007. Is colour modulation an independent factor in human visual photosensitivity? *Brain* 130:1679-89

42. Shibata T, Shimoyama I, Ito T, Abla D, Iwasa H, et al. 1999. Event-related dynamics of the gamma-band oscillation in the human brain: information processing during a GO/NOGO hand movement task. *Neuroscience research* 33:215-22

43. Akrami A, Solhjoo S, Motie-Nasrabadi A, Hashemi-Golpayegani M-R. EEG-based mental task classification: linear and nonlinear classification of movement imagery. *Proc. Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the, 2006*:4626-9: IEEE

44. Pfurtscheller G, Lopes da Silva FH. 1999. Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical neurophysiology* 110:1842-57

45. Popescu F, Blankertz B, Müller K-R. 2008. Computational challenges for noninvasive brain computer interfaces. *IEEE Intelligent Systems* 23:78-9

Vita

Elizabeth Ellene Salmon

Education:

Bachelor of Engineering, May 2008
Vanderbilt University, Nashville, Tennessee

Research positions:

August 2010 - present
Research Assistant, Neural Systems Lab
University of Kentucky, Lexington, Kentucky

September 2010 – January 2011
Research Assistant, Cardiovascular Lab
University of Kentucky, Lexington, Kentucky

Professional publications:

Cuaycong M, Engel M, Weinstein SL, SALMON E, Perlman JM, Sunderam S, Vannucci SJ. A novel approach to the study of hypoxia-ischemia induced clinical and subclinical seizures in the neonatal rat. Dev Neurosci, 2011 (special issue "Frontiers in Neuroprotection and Neurorepair").