Spring 5-15-2016

Development and Investigation of Sparse Co-Adaptive Algorithms in ECoG based Closed-Loop Brain Computer Interface

Piyush Karande
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

School of Engineering & Applied Science
Department of Biomedical Engineering

Dissertation Examination Committee:
Daniel W. Moran, Chair
Dennis L. Barbour
Baranidharan Raman
Lawrence H. Snyder
Kilian Q. Weinberger

Development and Investigation of Sparse Co-Adaptive Algorithms in ECoG based Closed-Loop Brain Computer Interface

by

Piyush Karande

A dissertation presented to the
Graduate School of Arts & Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2016
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

At the end of my graduate studies, I would like to thank a few people. Firstly, I would like to extend my utmost gratitude to my mentor and advisor Dr. Daniel Moran for giving me an opportunity to be a part of his research group. In the past few years, he has guided me through my research not by telling me what to do or pushing me to meet deadlines but by letting me find my own path and nudging me in the right direction when I felt lost. In the process, he helped me to develop a way of thinking that proved to be crucial for my doctoral research and its culmination with this dissertation.

I would also like to thank all the members of my thesis committee for their timely feedback. Their involvement gave me new ideas and a different set of perspectives. I believe it eventually led my doctoral research to be significant not only in the field of brain computer interfaces but also neuroscience. For that, I am truly thankful.

The members of the Moran lab have been a big part of my research in the past year as well. Adam, Jordan, Tom, and Jesse helped me start in the lab by guiding me through their work and helping me learn things that now seem like the back of my hand. In the time that followed, they and Jonathan became the people that I discussed my ideas and experiments on an everyday basis. Without all of their inputs and constructive criticism, I wouldn't have been able to accomplish all this in past few years. I am also incredibly grateful to Donna for helping me with my experiments and putting up with my tardiness from time to time.

I have been lucky to have made some great friends in St. Louis over the past few years. Karthik, Phani, and Adhira helped me transition into a new country so smoothly that I didn't even realize it was happening. Going forward they, Kshamata, Vivek, Afreen, Venkat, and Shradha became my family away from home. We had a lot of great times together and they all stuck by my side and helped me get through some difficult phases of my Ph.D. and life in general. I have also had many other friends back home and a few spread across different countries who have a played a crucial role in shaping my personality. I would specifically like to thank Sibani for being a true friend since we were a couple of 10-year-olds and believing in me during the times I did not believe in myself. To all these people, I am truly in your debt.

Finally, I would like to thank my parents and family. They have been very encouraging throughout my life and pushed me to achieve certain goals that I didn't even think to consider. They have also been very understanding and supportive over the past few years I have been so far away from them. Though we have been able to meet only a few times in these years, they have never complained and have always wished for my success. I dedicate this dissertation to my parents, Ranjana and Ajit, for bringing me in this world and making me the person I am.

Piyush Karande

Washington University in St. Louis

May 2016

To Mummy & Pappa.

ABSTRACT OF THE DISSERTATION

Development and Investigation of Sparse Co-Adaptive Algorithms

in ECoG based Closed-Loop Brain Computer Interface

by

Piyush Karande

Doctor of Philosophy in Biomedical Engineering

Washington University in St. Louis, 2016

Research Advisor: Professor Daniel W. Moran

Electrocorticography (ECoG) has gained a lot of momentum and has become a serious contender as a recording modality for the implementation of Brain-Computer Interface (BCI) systems in the last few years. ECoG signals provide the right balance between minimal invasiveness and robust spectral information to accomplish a BCI task. However, all the BCI studies until now have used signals recorded from a large number of implanted electrodes and a larger number of spectral features. The recording and processing of these signals uses a lot of electrical power and thus hinders its use outside the research setting. To translate this research to the clinic as a chronic recording modality for neural prosthesis, minimizing the number of features and thus, the power consumption to record and process them, is of prime importance. This thesis develops and investigates two different techniques to minimize the feature space required to obtain a robust BCI control in a virtual environment setting. ECoG electrodes embedded in thin-film polyimide or Silastic were implanted in the epidural space over pre-motor, primary motor and parietal cortical areas in non-human primates. Subjects

tested this thesis had had their electrode arrays implanted at least 1-2 years before the beginning of these experiments. Monkeys were trained to perform a classic 2D center out task using the recorded signals and one of two new BCI decoding algorithms developed in this thesis. Both the algorithms used for BCI control updated the decoding model using data from the previous trials. The parameters of the decoding algorithms were varied every 1-2 weeks to gradually reduce the number of features being used for control. A robust BCI control was obtained using only 30-40% of the available feature set. Post hoc analysis of the reduced feature set revealed a significant presence of mid-gamma (75-115Hz) band followed by the beta band (15-30 Hz). A novel, 1D Up-Down BCI task was used to study the modulation frequency of these two bands and the differences between them. It was observed that though subjects gradually increased the frequency of modulation in both the bands over a few weeks, they were able to modulate the mid-gamma band at a faster rate. Finally, as a proof concept, two previously trained subjects were used to perform the 2D center-out task with features recorded from only 4 ECoG electrodes. The laboratory recording system and a low power recording system were used in different sessions of experiments, and a robust control was obtained in both the cases. The overall observations and results of these studies provide with a strong basis for ECoG as a low power recording modality that can be chronically used for neural prosthesis.

Chapter 1
Introduction

1.1 Motivation

Movement is how humans interact with the world. Walking around to explore the surroundings, building a shelter, growing and cultivating food, reaching out to feel the environment, or even vocalizing to interact with other fellow humans, all require the coordinated use of muscles. The nervous system enables us to produce all the movements. The Central Nervous System (CNS), comprising of the brain and spinal cord in humans, produces and processes the commands. These are then carried by the motor nerves innervating different parts of the body. The Peripheral Nervous System (PNS) that the nerves are a part of, acts as a relay between the musculature and the CNS to communicate the desired command.

What happens, though, when the commands cannot be relayed? The human brain that can think and feel complex emotions like happiness, sadness, anger and compassion, all of a sudden loses the channel to express them. Degenerative motor neuron diseases like Amyotrophic Lateral Sclerosis (ALS) affect the peripheral nervous system and cause a disruption of commands being relayed from the brain to the muscles. Though patients with ALS can think and feel all the emotions, their ability to express them is severely compromised.

Also, the inevitable dependence on family or caretakers for day to day activities often leads to emotional damage.

Brain-Computer Interfaces (BCI) are devices that circumvent the peripheral nervous system and directly decode the commands from the brain. A BCI could be controlling something as simple as a cursor on a screen or something as complex as a prosthetic limb. The main goal is to provide the patients with ALS, spinal cord injuries, or locked in syndrome, a channel to turn their thoughts into action.

The human brain is composed of billions of neurons that use electrochemical signals to communicate with each other and control different functions of the body. To decipher the motor commands produced in the brain, BCI devices have to record these signals in one form or another. The electrochemical activity of an individual neuron produces electrical fields that can be recorded using a small single-unit electrode (20 μm tip) that is placed in close proximity (100 μm) to the cell body. If a group of closely spaced neurons are simultaneously active and processing similar information, a larger electrode can be used to record the local field potential (LFP) of the ensemble.

Different recording modalities ranging from intra-cortical single-units to scalp based local field potentials (electroencephalography or EEG) have been studied over the past few decades to record and interpret brain activity for use in a BCI. In the past decade, however, a new intermediate modality known as electrocorticography or ECoG has gained significant recognition and has had much success. This dissertation argues for the significance of ECoG in the field and backs it up with experimental data. It uses simple algorithms to select the most

reliable signals from the available set and uses them for robust BCI control. It also shows how certain spectrums of ECoG signals manifest themselves across experiments as the most significant features.

1.2 Aims and Organization of Dissertation

The initial chapters of the dissertation provide the foundation that this doctoral research was built on. Specifically, Chapter 2 discusses decades of research in motor neuroscience and the field of BCI that has provided the basis for the theory and design of the studies in this dissertation. Chapter 3 describes the subjects, surgical and recording procedures, experimental design, and other general methodology used throughout the dissertation.

The main goal of this dissertation was to develop algorithms that optimally select the smallest subset of available control signals that can produce a robust BCI. The ECoG signal recorded from each of the implanted electrodes was separated into five different frequency bands (alpha: 8-15 Hz, beta: 15-30 Hz, low-gamma: 30-55 Hz, mid-gamma: 70-115 Hz, and high-gamma: 130-175 Hz). Much like independent radio stations on the AM spectrum, it was assumed that each frequency band produced independent control information. Likewise, the information in each frequency band was assumed to be encoded in its overall power – again mimicking the amplitude modulation seen in the AM radio spectrum. Each electrode/frequency band pair formed a "control feature" that constituted the smallest independently controllable signal obtained from the electrocorticographic recordings.

Chapter 4 describes the experimental observations from using a feature pruning algorithm for a classic 2D center-out BCI task. This algorithm was designed to select a subset

of features while preserving their distribution in the 2D space. On selection of a percentage of features to be pruned, the algorithm set a variable threshold across the 2D space on the magnitudes of features. Only the features above this threshold were used for BCI control. The percentage of features being pruned was increased, gradually, and post hoc analysis was conducted to study the reduced feature set and its effect on BCI control metrics (*e.g.* speed, accuracy).

Chapter 5 discusses another set of feature pruning experiments where the most consistent features in 2D space were pruned. The goal of this experiment was to a) study the characteristics of features that were most important for BCI task and b) to evaluate if the subject can adapt and learn to modulate the features that were previously less significant for the task.

Chapter 6 uses a different approach to reducing the feature space. A regularized least squares regression algorithm was used to gradually decrease the number of features used for control. With the use of a 2D center-out BCI task and gradually increasing the regularization parameter, subjects were trained to use a subset of features for control. The method of regularization used was a combination of ℓ_1 and ℓ_2 norms and was designed to select a single subset of features to control the movement in both the dimensions of motion. Post hoc analysis was conducted to study the changes in performance and the characteristics of the features selected.

Chapter 7 uses a novel 1D Up-down path task and a 1D Target-tracking task to study a subject's ability to modulate power in different ECoG frequency bands. These experiments

were designed to determine the highest rate of modulation in both the beta and mid-gamma frequency bands during real-time BCI cursor control: two frequency bands that were observed, from previous experiments, to be used most frequently used.

As a proof of concept, Chapter 8 demonstrates the use of signals from only four recording electrodes for the control of 2D cursor movements in a center-out task. Using the laboratory system and a low power recording system in separate sessions, this set of experiments argues that a small number of ECoG signals can be used to efficiently control a 2D BCI.

Finally, Chapter 9 concludes this dissertation with an overall discussion of the observed results and their implications for the field of BCI, and more specifically, for ECoG in the coming years. It also provides some new directions that can be evaluated in the future studies.

Chapter 2
Background

Impairment of motor function due to injury or disease pathology leads to a severe reduction in the quality of life. It inhibits the patients from carrying out everyday tasks and makes them dependent on caregivers and family. The ALS Association approximates there are 30,000 people living with ALS at any given time, with 5600 new cases diagnosed every year, just in the United States [1]. The numbers for patients with spinal cord injuries (SCI) is almost an order of magnitude higher, at about 276,000 [2]. Assistive devices that use residual movement to perform some basic functions like pressing a button using a mouth stick offer limited functionality using unnatural movements. Rehabilitation techniques can be successfully used in a few cases where the disability is limited. However, even these require extremely long training periods and have almost no affect in cases like the last stages of ALS or paraplegia caused due to injury at higher cervical levels of the spinal cord.

With both spinal cord injury and motor degenerative diseases not affecting the brain, neural activity in the motor cortex is mostly intact. Rerouting the signals from the brain directly to a prosthetic device (bypassing the affected sections) can provide the patients with a reliable pathway to improve the quality of life. Devices that enable to do so are aptly named Brain-computer interfaces. Decoding brain commands that would naturally produce movement can help the patients learn and internalize the model to control BCIs with ease.

The human brain is a complex organ with billions of neurons interacting with each other to produce complex thoughts and movements of human body. Decoding such complex signals might seem a daunting task. However, centuries of curiosity regarding the nervous system and more particularly the last few decades of research in motor neurophysiology have tremendously increased our understanding of the field. Equipped with this knowledge and the plasticity of the brain, neural signals have been translated into simple movements of a cursor or even complex control of multiple degrees of freedom in a prosthetic limb.

This chapter provides a background of motor neuroscience that has helped establish BCI as a reliable method for helping various patient populations. It further investigates BCI research using different recording modalities and decoding algorithms. Finally, it discusses how all this background has led to the studies in this dissertation.

2.1 Neurophysiological Basis

One of the first evidence of neurophysiological investigations of the nervous systems dates back to late 1700s when Luigi Galvani demonstrated that nerve cells and muscles produce electricity [3]. The electrical interactions between the adjacent nerve cells were demonstrated by Hermann von Helmholtz in the nineteenth century; marking the beginning of modern electrophysiology [3]. Since then several advances have been made in the understanding of how the brain functions and produces complex behaviors. Since motor functions are the primary behavioral output of the brain, observed in everyday activities, they have received special importance.

The motor and somatosensory cortices are situated anterior and posterior to the central sulcus on both left and right hemispheres of the brain. Early electrical stimulation studies in dogs, non-human primates, and humans showed that different portions of the body are mapped in these cortices [3], [4]. Since then, neurophysiological studies from these areas have provided a lot of insight on how humans move and feel the environment around them. The three most important findings that are most relied upon by BCI systems are:

a) Cosine tuning characteristics of motor cortical neurons,

b) Modulation of cortical activity during imagined movements, and

c) Cortical plasticity exhibited by the brain.

2.1.1 Tuning characteristics

Georgopoulos *et al.* demonstrated in 1982 that spiking activity of neurons in the motor cortex was modulated with hand movements [5]. It used the then novel and elegant center-out task (used in this dissertation). The task involved the subjects moving their arms from the center of a circle to one of the eight equally spaced targets placed on its circumference. This task enabled them to demonstrate that the spiking rate of the recorded neuron was cosine tuned with the direction of the motion of the arm. That is, the neuron fired with the highest rate when the arm was moved in one particular direction (called the preferred direction of the neuron) and fell off as the cosine of the angle between the preferred direction and the direction of movement. Several studies [6]–[11] since then have demonstrated the tuning of spiking activity from the motor cortex with different parameters and types of movement. In the recent

years tuning characteristics have been also observed in intra-cortical and cortical surface field potential recordings [12], [13].

2.1.2 Cortical Activity and Imagined Movements

Researchers have discovered that imagination or mental rehearsals of movements also produces modulations in the motor cortex as real movements do. Previous studies with humans have used techniques like magnetoencephalography and regional cerebral blood flow (rCBF) to demonstrate the similarity. Jean Decety reviewed several studies [14] in 1996 that used PET, SPECT and other imaging techniques to demonstrate the similarities between the neural substrates of real and imagined movements. Schnitzler *et al.* demonstrated the involvement of the primary motor cortex (M1) in motor imagery in 1997 [15]. They used whole-scalp magnetoencephalography to show modulation of 20 Hz signal during imagined and real finger movements. A study by Hochberg *et al.* in 2006 has demonstrated the modulation of firing rates in several simultaneously recorded M1 neurons when a subject with tetraplegia was instructed to imagine different movements in hands, arms, and legs [16]. Electrocorticography (ECoG) activity modulation during motor imagery in humans was demonstrated by Leuthardt *et al.* in 2004 [12].

2.1.3 Cortical Plasticity

The third and arguably the most important characteristic of sensory-motor cortex for the purpose of BCI systems is its plasticity. Decades ago Merzenich *et al.* demonstrated changes in somatosensory cortical maps in owl monkeys after the amputation of the third digit

of hand [17]. Since then studies have also provided evidence of cortical plasticity in M1 following pathological changes. Sanes [18] and Donoghue [19] demonstrated, with the help of electrical stimulation and mapping experiments, functional reorganization in the motor cortex in rats after transection of facial motor nerves. Ungerleider *et al.* have used fMRI to demonstrate the reorganization in M1 produced by learning sequential finger movements [20].

Previous BCI studies in the lab have demonstrated cortical plasticity through ECoG signals. Subjects performed a circle drawing task by controlling the vertical movement of the cursor with one electrode and horizontal movement with another. To improve the performance in this task, a decorrelation of the control signals was necessary. A steady decrease in the correlation between the two control electrodes, specifically in spectrum controlling the cursor was observed, demonstrating targeted changes in the signals to increase performance in the task [21].

The knowledge acquired from these studies, and their major implications have encouraged and enabled researchers to make tremendous advances in developing brain-computer interfaces. Over the past couple of decades BCI systems have been developed using different modalities and decoding algorithms, described in Section 2.2 and 2.3. These sections also demonstrate the importance that ECoG has gained in the field and the basis of algorithms used in this dissertation.

2.2 Recording modalities

The cerebral cortex is composed of layers of neurons interacting with each other through electrical and chemical signals. These neurons generate action potentials that produce

an extracellular electrical field potential. Different recording techniques have been used to capture the electrical activity to understand the how the brain processes information. Spiking activity of single neurons can be recorded using intra-cortical electrodes embedded in the parenchyma and is traditionally termed "single unit activity" [22]–[24]. The electrical summation from several simultaneously spiking neurons generates field potentials that can be recorded using similar intra-cortical electrodes (LFPs) or surface electrodes (ECoG or EEG) [25], [26]. BCI researchers have mainly used three recording modalities to record and decode brain activity, a) Electroencephalography (EEG) [27]–[30], b) Intra-cortical single or multi-unit activity [16], [31]–[36], and c) Electrocorticography (ECoG) [12], [21], [37]–[40].

EEG signals are recorded from the surface of the skull and come with the big advantage of being non-invasive. However, these signals have been observed to exhibit only nonspecific modulations with different movement parameters [41]. Also, the distance from the source and the electrical characteristics of the tissue effectively "low-pass filter" the underlying cortical activity, greatly reducing the information recordable by EEG electrodes (typically only below 70 Hz) [42]. Moreover, the amplitude of EEG signal is very small making it highly susceptible to noise [41]. Even with all these limitations, EEG-based BCI studies have been able to demonstrate two [27]–[29] and three [30] dimensional control, albeit requiring multiple lengthy training sessions.

Intra-cortical signals have been the most successful modality for implementing BCI. Recording spiking activity from multiple neurons simultaneously gives direct access to the building blocks of electrical activity in the brain. Research groups have used single and multi-

unit activity recorded from arrays of intra-cortical electrodes to control prosthetic limbs with multiple degrees of freedom [16], [30], [32], [34], [41], [42]. Studies have also been conducted in both non-human primates [32], [36] and humans with motor disabilities[16], [34], [43], [44] with great success. However, using penetrating metal electrodes to record from the brain has its drawbacks. Brain tissue is soft and metal electrode penetrating it with pointed tips cause damage to the vasculature [45]. The process of implantation also increases the risk of CNS infections [45]. In addition, the presence of a stiff, foreign body in the cortex leads to tissue response from the microglial and astrocyte cells [45]. Accumulation of these cells and their reactive response eventually leads to electrode encapsulation and degradation of signal quality [46], [47]. For these reasons use of intra-cortical recordings is not a feasible long-term option. To enable the clinical use of BCI or even chronic use in patients with motor disabilities, researchers need to use recording modalities that are less susceptible to infection and can provide with stable and reliable recording of the neural activity.

The short-comings of EEG and intra-cortical recordings have led to the interest in Electrocorticography (ECoG) as a recording modality for BCI. ECoG provides with a balance between EEG and intra-cortical recordings. Though semi-invasive, ECoG electrodes are placed on the surface of the cortex or the dura. The elimination of thick skull tissue reduces the distance between the electrode and the brain and increases the recordable bandwidth from about 70 Hz in EEG to approximately 200 Hz in ECoG recordings[12], [48]. As described in section 2.1.2, ECoG signals also exhibit specific modulation with overt and imagined movements. Also, high gamma frequency bands (>75 Hz) have been popularly theorized to

represent synchronous spiking activity from the cortical columns under the recording electrode.

Subdural ECoG grids have been extensively used by neurosurgeons in the treatment of intractable epilepsy. A majority of human ECoG-based BCI experiments have been conducted in such patients during the period of the implant. The drawback of such experiments has been that the location of electrodes is governed by the clinical needs of the patients and that the grids can only be implanted for a short duration of a few weeks. However, seeing initial success with ECoG in human studies [12], [37], [38], [49], researchers have started using non-human primates to develop ECoG-based BCI [21], [39], [50]–[52].

Our lab has investigated the use of ECoG electrodes in the epidural space of non-human primates over the last decade with a lot of success. Epidural ECoG electrodes are implanted over the surface of the dura as opposed to the subdural electrodes. As the outside layer of the dura is part of the peripheral immune system, these epidural electrodes are less susceptible to initiating infections in the CNS. Also, the electrical conductivity of the dura is similar to that of cerebral spinal fluid, therefore causing negligible degradation of the signal due to its presence between the electrode and the brain. Thus, epidural ECoG provides an ideal balance between fidelity of recorded signals and the risk of infection for the purpose of BCI systems.

2.3 BCI Implementation and Algorithms

Most of the work in the field of BCI has been in designing and implementing decoding algorithms that help capture the most information embedded in the recorded signal. The

algorithms are used to make it easier for the subjects to gain efficient control of the effector with the least effort. Different algorithms have been used in the past, to name a few, Population Vector Algorithm (PVA); Optimal Linear Estimator (OLE) and Velocity Kalman filter (VKF). Population vector algorithm was a successor to Georgopoulos' study demonstrating cosine tuning of single M1 neurons with hand movement and was proposed by him and his colleagues in 1983. This algorithm predicts the instantaneous velocity of movement by summing the vectors, representing the preferred direction of multiple neurons recorded from M1, scaled by their instantaneous firing rate [53]. Further development in the PVA came with OLE, which assigned weights to the individual vectors in a way that minimized the least squares error between the final estimate of the velocity and the true observed velocity. OLE was proposed by Salinas and Abbott, and they demonstrated, on the data obtained from Georgopoulos *et al.*, that it performed better than PVA [54].

Another newer algorithm that has been widely used is the Velocity Kalman filter. First proposed by Black *et al.* in 2003 [55], this algorithm uses a state space representation of the hand positions and velocities. The prediction of the state of the system at any given time step is dependent on the recorded firing rates and the previous state of the system. The matrix that transforms one state to the other and the prior estimate of state from the recorded signal are constructed with the intention to minimize the mean squared error of reconstruction.

Though the above-mentioned algorithms and several others using neural networks [56], Hidden Markov Model [57], *etc.* have been proposed for use in implementing BCI, the most important factor in the success of BCI is the real-time visual feedback (closed-loop

configuration) to the subject and the resulting cortical changes (*i.e.* neural plasticity). Studies have shown marked difference between performances of various algorithms when used in open-loop and closed-loop configurations [58]–[60]. Chase *et al.* demonstrated an equivalent closed-loop performance of OLE and PVA algorithms [58]. Cunningham *et al.* demonstrated the difference in time of integration of the recorded signal that produced optimal results and overall smaller errors when the same algorithm was used in closed-loop [60]. Using a Kalman filter decoder updated every 1-2 minutes during the experiment, Orsborn *et al.* showed rapid improvements in closed-loop performance [61]. These observations have led researchers to believe that algorithms that incorporate feedback and update the parameters iteratively would be more successful in efficient and easier control of BCI devices.

2.4 Significance of Sparse Algorithms

This dissertation aims to study the use of algorithms that result in a sparse control feature set for 2D BCI. A sparse set of features that can produce robust BCI control requires less number of signals to be recorded and processed. To translate BCI systems from the lab to a clinical setting with implantable devices, the number of recorded signals is a key factor. Initial recording and filtering of a large number of extremely low voltage neural signals is extremely power consuming [62]. Hence, a smaller set of recorded signals would lead to lesser power consumption and longevity of the implant.

The other significant outcome of using sparsity producing algorithms is the low-dimensionality of the output. A sparse set of features is easier to interpret and would give an insight into the strategies used for control. The characteristics of the sparse feature set can be

studied to determine which features are more significant in control. Also, using different approaches to produce a subset of the available feature space, and studying the similarities across them will lead to further strengthening of conclusions from individual approaches. For instance, the features used in the experiments for this dissertation are the spectral estimates of five different frequency bands of ECoG signals. The raw ECoG signals themselves are recorded from different locations on the sensory-motor cortex, using electrodes of different recording surface areas. Thus, by studying the distribution of the sparse feature set with respect to their placement on the cortex and their frequency band, will help in deciphering the spatial and spectral characteristics of ECoG signals important for BCI control.

Though the use of such algorithms introduces another level of complexity into the BCI system, they are ideal candidates to be used in a research setting. The understanding of ECoG based BCI that would result from such algorithms can be later used by simpler algorithms to produce robust BCI control with a small number of recorded signals.

Two approaches to reduce the number of control features have been used in the experiments for this dissertation. The first approach emphasizes on maintaining a uniform distribution of the features in the 2D space. By doing so, this approach ensures that the BCI control is not biased in any direction. The second approach uses the regularization principles from machine learning to obtain a single subset of features that controls movement in both the dimensions. The parameters of the algorithms in both the cases were varied in such a way that the size of control feature set was gradually reduced with time. This was done with the

intention of enabling the subjects to slowly adapt to the changes. The experiments were carried out to a point where only 30-40% of the initial feature set was being used for control.

2.5 Summary

Previous sections of this chapter have discussed several different studies that have tremendously increased our understanding of motor neuroscience and applied their findings in the field of BCI in the recent decades. Different recording modalities have been explored, and several algorithms and techniques have been developed to control BCI devices. The field has seen a lot of success in the past few years. However, most of the research in the field has been in developing better strategies or complex algorithms with more accurate performance. The main drawback of these approaches is that they are not practically feasible to be used outside of a research setting.

To translate BCI from its research phase to being used by patients in need, it is necessary to investigate approaches that use simple algorithms, exploit neuroplasticity through closed-loop training, and use a small set of signals that can be reliably recoded by an implantable system for a long period. The experiments in this dissertation are a step in that direction. As has been discussed in section 2.2, epidural ECoG recordings provide the best tradeoff between invasiveness and the reliable information required for BCI control. The algorithms developed here are based on OLE and are simple to implement. The different feature selection approaches are designed with the intention of developing an understanding of the characteristics of ECoG signals that are essential for BCI control.

Finally, the goal of this dissertation is to not only solve an engineering problem but also to use the experimental data and BCI as a tool to study basic neuroscience. The results of the experiments have provided us with insights into the strategies used for BCI control. As described earlier, different techniques have been used to solve the same problem to get a picture of the underlying neurophysiological mechanisms important for BCI. An additional set of experiments was carried out where the most consistent and reliable features were gradually eliminated instead of finding the best feature set for control. By performing this experiment, BCI was used as a tool to compel the subject to use features that were less reliable to accomplish the task. Another set of experiments has been performed with individual frequency bands of ECoG to determine which one of them is a stronger contender in the field. Through all these studies the following chapters of this dissertation attempt to provide the field of ECoG BCI with a base to translate out of a laboratory setting and into the clinic.

Chapter 3
General Experimental Methods

This chapter explains the general methods used during the different BCI experiments carried out for this dissertation. It describes the subjects used for the experiments, the electrodes and devices used to record the ECoG signals, the BCI tasks, and the virtual environment used to display them. Any deviations from the general methodology are outlined and described in the later chapters.

3.1 Subjects

All the experiments described in this dissertation use male Rhesus macaques (*Macaca mulatta*, monkeys G, I, K, P, R, T) weighing 6-10 kilograms. All the subjects, except Monkey G, were chronically implanted with ECoG electrodes in their epidural space over the motor and/or sensory cortices. Monkey G was implanted with an array of subdural ECoG electrodes. Monkey I, K, R, and T had arrays implanted on both the hemispheres (bilateral) whereas, monkey P and G had only unilateral, left hemisphere implants. Monkey I was used in experiments described in Chapters 4, 5 and 7. Monkeys K, P were used in experiments in Chapter 4. Monkey R was used during experiments for both Chapters 5 and 6. Monkey G and T were used only for experiments in Chapter 7 and 6 respectively. All the subjects had been implanted a minimum of 1-2 years for previous studies and were carried over to the studies in this dissertation after their completion. Most subjects (G, I, K and P) were previously

trained in one or more BCI tasks. Monkeys R and T, although implanted about a year prior to my experiments, were experimentally naïve, before carrying out BCI tasks for this study.

3.2 ECoG Arrays

Two different types of ECoG arrays were implanted in the subjects used in this dissertation. Monkey K, P, R, and T were implanted with thin film polyimide arrays with a foldable design. Monkey G and I were implanted with electrodes encased in Silastic. The details of these electrodes and their placement are described in the following sub-sections.

3.2.1 Polyimide arrays

The thin-film polyimide arrays (Figure 3.1 A) were designed and manufactured in collaboration with Dr. Justin Williams' lab in University of Wisconsin, Madison. Each array consisted of two halves that folded over each other before being implanted. Half of the array consisted of 32 circular recording electrodes arranged in a 4-by-8 grid. The other half of the array was printed with a large "H" shaped ground pad and two reference electrodes similar in size to the recording electrodes. The thin-film polyimide material was flexible and allowed the array to be folded before being inserted into the epidural space under the skull (Figure 3.1 B). By folding the array, the recording electrodes were placed on the dura mater covering the brain whereas, the ground pad and reference electrodes were placed facing the skull. Such a placement prevented the reference and ground electrodes from picking up a significant amount of cortical field potentials but, allowed access to the noise and artifacts from the recording area.

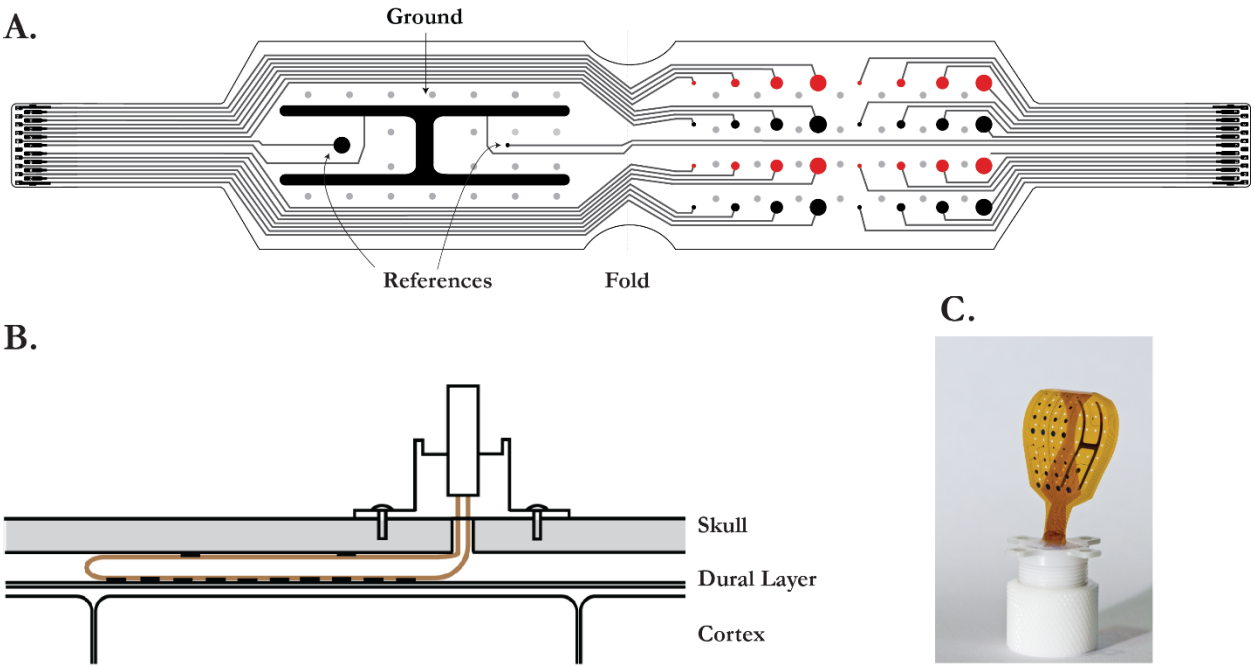

Figure 3.1: Polyimide electrode design. A. Polyimide thin film ECoG electrode array layout with one half housing 32 channels of four different sizes of exposed electrode surface and separated by 3mm (center to center). The other half is covered with an "H" shaped ground and reference electrodes. B. Schematic of implanted electrode with its relative position to the skull, dura, and brain. The array was folded at the mark shown in A and placed under the skull with ground and reference facing up. C. Picture of the folded array before implantation. The white Delrin housing contains the Zero Insertion Force (ZIF) connector and is exposed outside after the implant.

All the electrodes on the flexible polyimide array were connected to metal pads on each end of the array (16 recording electrodes and one reference on each side) through thin printed metal connections. After folding the array, these pads were inserted in a Zero Insertion Force (ZIF) connector (Tucker-Davis Technologies, Alachua FL). The ZIF connector was then placed in a Delrin® Plastic (DuPont) chamber with screw-on cap (Figure 3.1 C) and was secured using dental acrylic.

The polyimide electrode arrays implanted in all the monkeys were exactly the same except for the sizes of recording electrodes. Monkey K and P were implanted with arrays containing electrodes with 300μm, 425μm, 520μm, and 600μm diameters. Monkey R and T

were implanted with electrodes sizes of 300μm, 600μm, 900μm and 1200μm. In both cases, the electrodes were separated by 3mm center-to-center. Approximate electrode locations on the surface of the brain will be provided in the following chapters.

3.2.2 Silastic Arrays

Monkey G and I were implanted with Silastic electrode arrays manufactured by Ad-tech Medical and PMT Corporation (Figure 3.2 A). Each array consisted of fourteen Platinum (Pt) recording electrodes and three skull-facing Pt reference electrodes. The Ad-tech array consisted of thirteen electrodes with 1.5mm diameter and one electrode with 3mm diameter separated by 3.75mm spacing center-to-center. The PMT arrays consisted of thirteen electrodes with 2mm diameters and one electrode with 3mm diameter separated by 3.75mm spacing center-to-center. Monkey G was implanted with one PMT array on the left hemisphere in the subdural space. This array had only 7 active recording electrodes. Monkey I was implanted bilaterally with one Ad-tech array on the left hemisphere and one PMT array on the right hemisphere, both in the epidural space and with all recording electrodes active.

All the electrodes and references on each array were connected to wires wound into a thin cable going out of the array. In monkey G, this cable was directly connected to a male Omnetics Connector (18 Position Nano-Miniature Connector, Omnetics Connector Corp.) placed in a small stainless steel chamber with cap and secured with acrylic. A skull screw was used as ground and was connected to the ground cable of the connector. In the case of monkey I, the two thin cables from the left and the right implant were placed and secured with acrylic

in a stainless steel chamber (Figure 3.2 B). The ends of each cable were connected to a male Omnetics Connector similar to the one in monkey G. However, these connectors were placed on a circular printed circuit board (PCB) that resided snugly in the metal chamber. The PCB enabled the choice of any one of the reference electrodes during the experiment. The stainless steel chamber connected to externally accessible skull screws, through alligator clips, was used as a ground during the experiments. The stainless steel chamber was also permanently secured over the skull using acrylic.

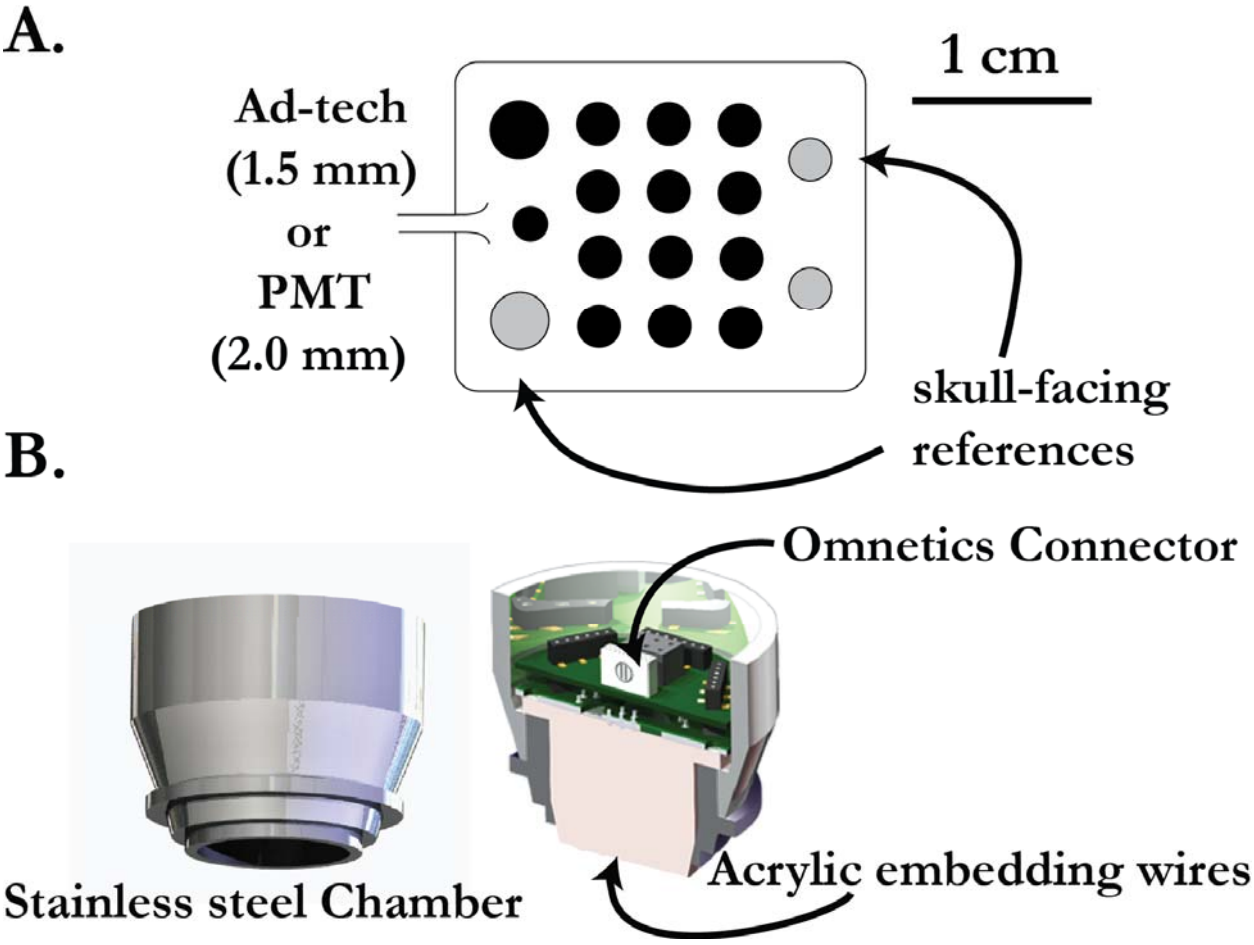

Figure 3.2: Electrode Design and placement. A. ECoG electrode array with 14 active Pt electrodes and three skull facing references embedded in Silastic. B. Left: Schematic of the stainless steel chamber housing the connecting wires and the PCB. Right: Cross-section of the chamber showing the PCB and one male Omnetics connector. (Drawings of chamber courtesy of JJ Wheeler).

3.3 Surgical Implantation Procedure

All the surgeries were performed at the primate facility in the Washington University School of Medicine. All the surgeries, except one, were performed by Dr. Daniel W Moran with one or more graduate students assisting in the procedure. Surgery for Monkey G, with a subdural implant, was performed by neurosurgeon Dr. Eric C Leuthardt. Subjects were placed under general anesthesia, and the facility's veterinary staff was available to assist if required. All the surgeries were performed in accordance with an approved set of protocols. The general steps of each surgery after placing the animal under anesthesia were as follows:

1. A 19 mm craniotomy was performed using a hand trephine centered at the point determined using stereotactic maps of the Rhesus macaque brain. The dura mater covering the brain was kept intact, and the bone plug was retained.

2. The array to be implanted was slid between the skull and the dura through the craniotomy. Plastic spatulas were used to separate the dura from the skull as well as to gently depress the dura against the brain to ensure the array remained flat and was placed as desired.

3. With the array placed and the recording chamber as close to the skull as possible, the bone plug was placed back on the craniotomy hole. Any gaps between the plug and skull were filled with Gelfoam® (Pfizer) moistened with saline, to facilitate bone growth. The bone plug was secured to the rest of the skull using 2 cm titanium straps.

4. If a subject had a bilateral implant, steps 1-3 were repeated on the other side.

5. A few more titanium straps were screwed on at different places around the craniotomy. This was done to provide a greater adhering surface area for the acrylic covering the skull and create a more stable acrylic head-cap.

6. A head-cap was built by laying one thin layer of dental acrylic at a time, over the exposed skull, straps and the base of the chamber.

7. Metal head-posts were placed at the top of the head-cap to secure the subject's head in a fixed position during the experiments.

Steps were modified or changed according to the requirements of a particular subject with respect to the stereotactic coordinates of the craniotomy, type of recording electrodes and chambers, and presence or absence of external skull screws to be used for ground.

3.4 ECoG Recording and Signal Processing

ECoG signals were recorded using the Tucker-Davis Technologies (TDT, Alachua FL) system. The raw signals at the connector were routed using a short cable (30 cm) to a differential head-stage and pre-amplifier assembly. The head-stage provided an initial amplification gain of 5 and then bandpass filtered the raw signals between 3 and 500 Hz. The signals were then routed to a pre-amplifier with a gain of 10. The pre-amplifier also digitized these signals at a 6 kHz sampling rate and converted them to optical signals to be carried to the TDT base station. The use of optical fiber cables from the pre-amplifier in the recording room to the adjacent control room prevented any 60Hz electrical noise being picked by the long cables.

The control room consisted of the TDT base-station and a computer running Microsoft Windows 7 OS that was used to control the TDT system and also to display the virtual BCI environment to the subject in another room. The TDT base-station was programmed to down-sample and store the ECoG signals at approximately 2 kHz.

Several studies have shown modulation of different frequency bands in field potentials recorded from the brain with movement activity or imagined movement [12], [13]. EEG BCI studies have used signals from low-frequency alpha and beta bands for achieving BCI control [30]. ECoG based studies have used information from higher frequency gamma bands [39], [63]. Previous work in the lab has separated the available bandwidth of the recorded ECoG signals into five different bands; alpha, beta, and three gamma bands [52]. The experiments in this dissertation take a similar approach.

The signals recorded from all the electrodes were separated into alpha (8-15 Hz), beta (15-30 Hz), low-gamma (30-55 Hz), mid-gamma (70-115 Hz), and high-gamma (130-175 Hz). The gamma band was split into three individual bands to avoid noise from the 60Hz mains power supply and its harmonics (*i.e.* 120 Hz). Voltage amplitude estimates in each of these bands were used as control features in various BCI algorithms. A basic amplitude demodulation technique, using the digital signal processors (DSPs) in the TDT base-station, was used to process the signals as described above. Signals were first band pass filtered using an 8^{th} order Butterworth filter with the desired low and high cutoff frequencies of each band. The output from these filters was then full-wave rectified before being low-pass filtered at 2 Hz using a 1^{st} order Butterworth filter, to get an amplitude estimate in the particular band. All

the filters were implemented in the DSP hardware of the TDT base-station. This enabled fast near real-time computations and reduced the delays in the system.

After the initial filtration, separation into individual bands, and amplitude estimation, all the features were relayed to the computer that controlled the virtual environment of the task. The amplitude distribution of the features at this point was skewed towards the positive values (Fig 3.3 A). To equalize the range of amplitudes below and above the mean, a log transform of the amplitude estimation was carried out. This compressed the higher positive values and produced an output with a near-normal distribution (Fig 3.3 B).

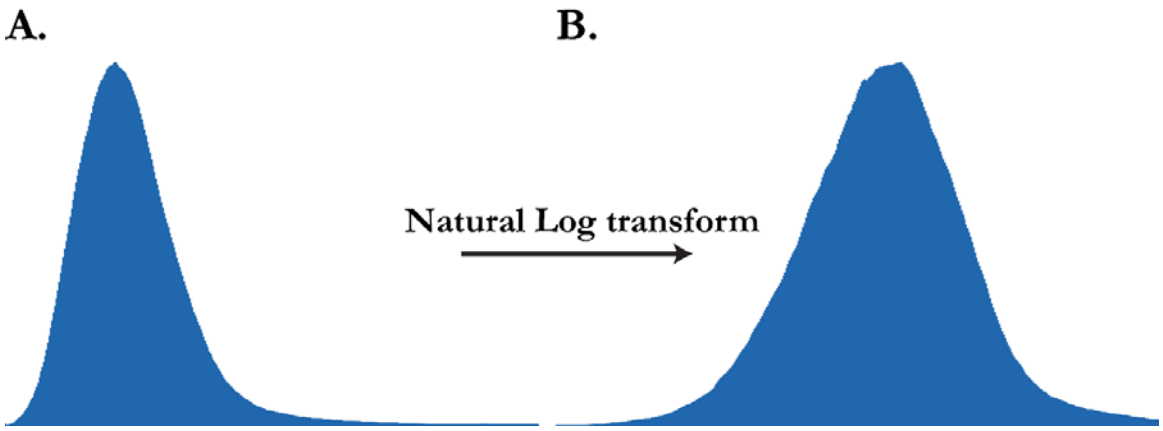

Figure 3.3: Feature histograms. A. Histogram of the amplitude estimate of the band-pass filtered signal is skewed towards right. B. Histogram of the same signal after a natural log transform is closer to a normal distribution.

The algorithms used in this dissertation are based on linear regression. As the spectral power of ECoG signals falls off approximately inversely proportional to the frequency, features from the different frequency ranges had to be normalized to avoid any incorrect weighing by the decoding algorithm. All the experiments used a real-time computation of the

z-score of the features. The mean and the standard deviation were calculated at every time point using the following equations [64]:

$$m_k = m_{k-1} + (x_k + m_{k-1})/k \quad \textbf{(3.1)}$$

$$s_k = s_{k-1} + (x_k + m_{k-1}) * (x_k + m_k) \quad \textbf{(3.2)}$$

$$std_k = \sqrt{s_k/(k-1)} \quad \textbf{(3.3)}$$

Where, m_k and std_k are the mean and standard deviation at time k, x_k is the k^{th} measurement sample, and s_k is an intermediate variable. The above equations are used for $k \geq 2$ with m_1, s_1, and std_1 initialized to x_1, 0 and 0, respectively. The z-score for each feature was then calculated using:

$$z_k = (x_k - m_k)/std_k \quad \textbf{(3.4)}$$

This z-scored value of each feature was used by the decoder to form the control signal of the BCI task. Figure 3.4 shows a comprehensive schematic of all the steps in the basic signal processing chain.

3.5 BCI Task Virtual Environment

All the tasks used in this dissertation were programmed using an in-house application framework with an acronym NERVE (Neuroscience and Engineering Reconfigurable Virtual Environment), designed using Visual Studio and C++. NERVE was created in the lab by TM Pearce and JJ Wheeler and allowed subsequent users to design, customize and run tasks for

the BCI experiments by providing a basic framework for the virtual environment. With an open access to all the libraries of NERVE, certain changes to framework were also made for the tasks designed for this dissertation.

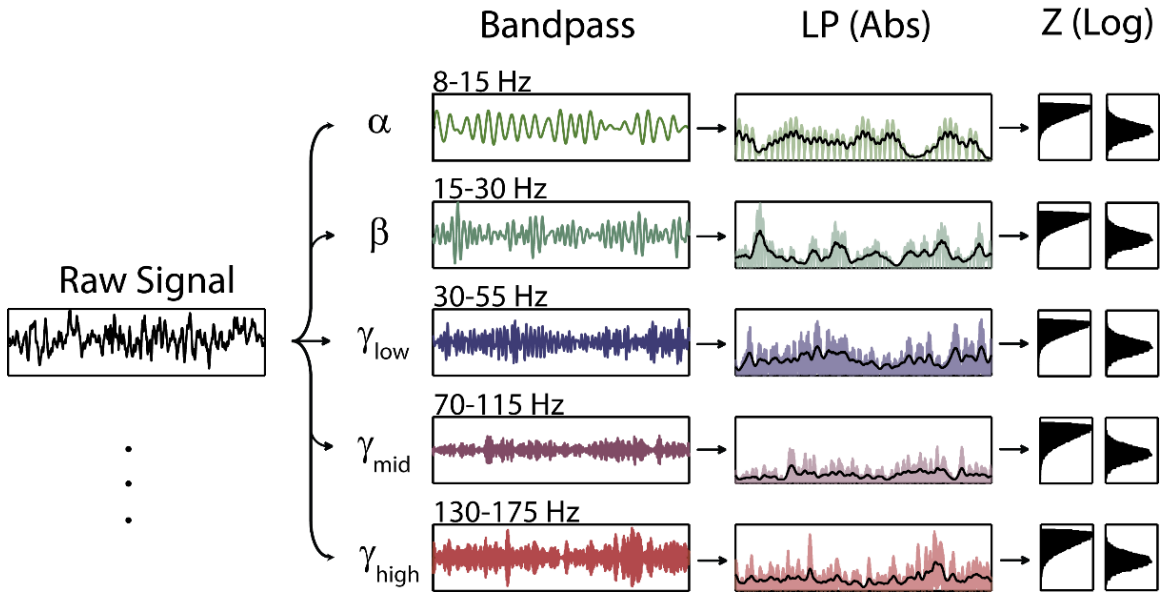

Figure 3.4: Signal processing chain. After initial filtering and amplification the raw signals are separated into 5 different frequency bands; 8-15Hz (alpha), 15-30Hz (beta), and three gamma bands; 30-55Hz (low gamma), 70-115Hz (mid gamma), and 130-175Hz (high gamma). Band separation is followed by amplitude estimation by rectification and low pass filtering at 2 Hz. The signal is then log transformed to make the distribution closer to Gaussian before calculating a z-score. Taking a z-score helps eliminate the difference in absolute magnitudes of signals from different bands and thus preventing them from being weighted differently by the decoding algorithm. (Figure courtesy of JJ Wheeler).

Each task environment consisted of Graphical User Interface (GUI) designed in Qt, to interact with and change the parameters of the current setup. Two graphical windows were spawned by the GUI at the beginning of each experiment. One window was presented to the subject in full-screen mode on a 20-inch monitor with a 4:3 aspect ratio. The other window was for the convenience of the experimenter and reflected what was presented in the subject window. All the graphical objects presented to the subject, during the task, were programmed with the OpenSceneGraph.

The state sequence of the task was programmed in a C++ script that interacted with the virtual environment described above. This script read the parameters described in the GUI and controlled the displayed objects on the graphical window based on it. The script also received signals from the TDT base station after they had been separated into five different bands. The features formed by a combination of recording electrode and frequency band were then log-transformed and a real time z-score was computed as described in equation 3.4. A MATLAB (MathWorks, Inc., Natick, MA) engine was used to interact with a MATLAB script that consisted of the decoding algorithm for the BCI task. The C++ script supplied the data required by the algorithm and used its output to combine the different features to form the control signals. The control signal was then used to move the elements in the virtual environment based on the rules of the task.

Data from TDT base station was read at 100 Hz and was used to update the virtual environment at the same rate. The recorded and processed features were stored at 20 Hz in a text file. All the other details of the task and specifics about each trial were stored in different text files. Features were stored at a lower rate to keep the size of files created during each session in the range of 1-3 gigabytes. Also, no changes were expected in the signals at a rate higher than 5 Hz. The TDT base station stored the raw ECoG signals received from the head-stage and pre-amplifier assembly, the current state of the task, and movement of the element controlled in the task.

3.6 BCI Tasks

Three types of BCI tasks were used in the experiments for this dissertation. Experiments in Chapters 4, 5, and 7 use the classic 2D Center-out task. In Chapter 6, a 1D Up-down path task and a sinusoidal tracking task were designed and used.

3.6.1 2D Center-out task

A classic 2D Center-out task was used for most of the experiments carried out during this study [5]. All the experiments used 8 spherical targets placed around the periphery of a circle, equally spaced at 45º starting from 0º to 315º. Each trial started with the spherical cursor, controlled by the subject, placed at the center of the circle. One of the targets appeared, and the subject was allowed a maximum time limit to move from the center to the displayed target. If the subject was successful in reaching the target in the specified time, the trial was marked as correct. If the subject did not reach the target, it was marked incorrect. The distance of the target from the center was set between 6 and 13 times the radius of the target/cursor spheres, depending on the task and the subject. The specifics for each experiment are provided in the respective chapters.

All the BCI experiments used small blocks of trials for every update of the decoding algorithm. Data from incorrect trials was ignored by the decoder and the targets for these trials were presented until all targets were reached correctly. Every target was presented in a pseudorandom order an equal number of times. Once 2 or 4 correct reaches to each target (16 or 32 total correct reaches to 8 targets) were completed, the decoder was updated, and a new block was initiated.

3.6.2 Up-down Path Task

The experiment in Chapter 6 used a novel 1D Up-Down path task to train the decoder using features from only one frequency band at a time. The task was designed to study the modulation frequency of the band used. The goal was to investigate the frequency of modulation that could be achieved with a band and if there were any differences between different bands.

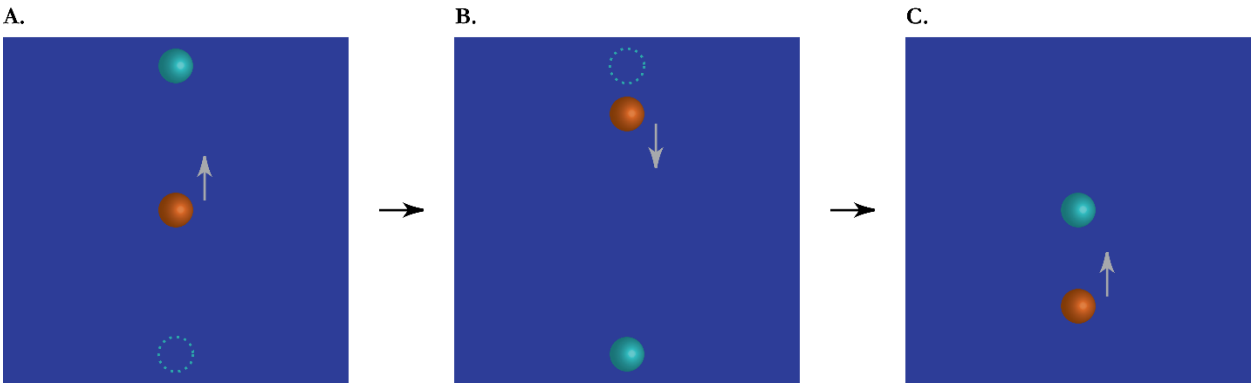

Figure 3.5: Up-down Path Task. A. First target is presented either vertically above (solid sphere) or below (dotted circle) at the beginning of the task. B. After the cursor reaches the first target, a second target is presented. C. A third target is presented at the center after the second target is reached. The trial ends when the cursor reached back to its starting position.

At the beginning of each trial, the spherical cursor controlled by the subject was placed at the center of the screen. The first target appeared either vertically above (Fig 3.5 A., Blue sphere) or below (Fig 3.5 A. Blue dotted circle) the center. Once the subject moved the cursor from the center to the displayed target, a second target, opposite in direction and at equal distance from the center was presented (Fig 3.5 B). After the subject successfully reached this target, a third target was presented at the center, instructing the subject to return the cursor to its initial position and complete a path starting and ending at the center. To ensure the subject did not just memorize the task and paid attention to each trial, the first target was set to either

Up or Down in a pseudorandom order. Movement of the cursor was restricted to the vertical dimension. The subject was given a set amount of time to complete the whole trial. Any trial that was not completed in that time was marked incorrect. No time restrictions were placed on reaching individual targets.

The two changes in the direction of movement during each trial required the subject to modulate the signals controlling the cursor around the mean during each trial. The goal of the task was to encourage the subject to complete a trial as fast as possible and thus requiring an increase in the modulation frequency of the control band in order to do so. For this reason, subjects were given rewards that were inversely proportional to the time taken for each trial with the faster completion of trials earning the highest reward. Subjects were also given an auxiliary reward for reaching individual trials when encouragement was required.

Subjects used in this experiment (Monkey R and T) were naïve and had not performed any behavioral tasks before. After going through the process of a general training in the lab to use a joystick, they were trained specifically to perform this task for 1-2 weeks. After this initial training of the task, subjects were switched to brain control. During the BCI trials, the decoder was updated every four correct trials, two with the Up target presented first and two with the down target presented first, all arranged in a pseudorandom order.

3.6.3 Tracking Task

The tracking task was interlaced with the Up-down path task in the experiments for Chapter 6. The goals of the tasks were similar to that of the Up-down task. However, the movement of the cursor in this task was restricted to a sinusoidal wave with a set frequency.

Each trial began with the cursor at the center of the screen. A target appeared at the center and moved up, down and back to the center in a sinusoidal form. The frequency of the sinusoid traced by the target was set by the experimenter and varied during the days of experiments. The goal of the subject was to track the sinusoidal movement of the target with the cursor. The subject had to be in contact with the target during the time of its movement. The size of the target and cursor were varied to allow more (bigger spheres) or less (smaller spheres) tolerance. A tolerance of about 20% of the movement time was also allowed to make the task easier, where the subject did not have to touch the target. For instance for a target frequency of 0.1Hz or a 10 second movement time, the subject had to touch the target for a total of only 8 seconds.

Subjects were trained using a joystick for 1-2 weeks before starting the brain control trials. During the days of BCI experiments, Up-Down path, and tracking trials were interlaced with one block (4 correct trials) of each following the other. The experiment began with a block of Up-down path trials and the model created by the decoder was then used to control the next block of tracking trials. This pattern was repeated throughout the experiment. The control model was updated only from the data obtained during Up-down path trials.

Chapter 4
Selective Feature Pruning: Preserving 2D Distribution

The main focus of BCI studies until now has been on controlling multiple degrees of freedom and developing sophisticated algorithms to do so. Researchers have had great success with humans and non-humans primates controlling multiple degrees of freedom in either a virtual environment or a prosthetic arm. These studies have spanned all the different recording modalities including intra-cortical recordings, ECoG, and EEG. All the studies to date, however, have used the whole set of signals available from the recording arrays. In most cases this involves recording and preprocessing neural activity from tens or even hundreds of electrodes. In the experiments using intra-cortical recordings, the processing steps also involve using complex real-time spike sorting algorithms. These factors make BCI systems highly impractical to be used on a day to day basis in a clinical setting. There is a need for simple algorithms that require a small number of reliable features to control a BCI system. This chapter outlines such a method that can determine an optimal subset of features using a closed-loop BCI task that adapts over time.

The experiment proposed in this chapter studies the possibility of using a small number of features to control a cursor in 2D space. A pruning algorithm was designed with the intention of preserving the overall distribution of features as they were gradually reduced in

number. This enabled the subjects to efficiently control the cursor with a small subset of features with a 2D distribution that was similar to the distribution of all the features together. The details of the algorithm are described in Section 4.1.6.

The pruning algorithm was used in conjunction with a decoding algorithm based on the Optimal Linear Estimator. Previous studies in the lab have demonstrated the use of a regularized least squares regression algorithm to assign weights to the recorded features (a combination of frequency band and recording channel). This previously used algorithm analyzed data collected from 200 consecutive trials over the block of time with the highest percentage of correct trials. The decoder thus constructed was used in an experimental session on the next day [51]. In contrast, the decoding algorithm described here adapted in real-time to update the decoder every 2-3 minutes. The decoding algorithm is outlined in Section 4.1.4.

This chapter is organized in three different sections. Section 4.1 describes the methodology used in the experiments. Post hoc analysis of the experimental data has been outlined in Section 4.2. Finally, Section 4.3 discusses the observation and draws conclusions to lead to the experiments in Chapter 5.

4.1 Methods

4.1.1 Subjects

Two Rhesus Macaques (*Macaca mulatta*, monkeys P, R) weighing 6-10 kilograms were used as subjects in the experiments described in this chapter. Monkey P had a unilateral implant on the left hemisphere of his brain whereas monkey K had a bilateral implant on both

left and right hemispheres. Both the subjects had been implanted for over a year before starting the experiments for this study. Both the monkeys had also been previously trained in using a joystick or ECoG signals for a 2D center-out task. It can be safely assumed that though a different control algorithm was used in this study, both the subjects did not need any initial behavioral training in the task performed.

4.1.2 ECoG Arrays: Design and Placement

Both Monkey P and K were implanted with a 32 electrode thin-film polyimide arrays designed in collaboration with Dr. Justin Williams' Lab in University of Wisconsin, Madison. The details of the array design and the electrical connectors are described in Section 3.2.1. Monkey P and K were implanted with one array placed in the rostral-caudal direction on the left hemisphere. The center of the 19mm craniotomy was located at the Stereotaxic coordinates (Anterior 15mm, Lateral 15mm). This placed the center of the most anterior row of the array at approximately (Anterior 20mm, Lateral 15mm). The array approximately covered areas spanning from premotor to S1. Monkey K was implanted with another array placed in the medial-lateral direction with the craniotomy centered at (Anterior 15mm, Lateral 9mm). The medial most row, in this case, was centered at (Anterior 15mm, Lateral 4mm). This array mostly followed the pre-motor and primary-motor strip. The approximate location on both the hemispheres is depicted in Figure 4.1.

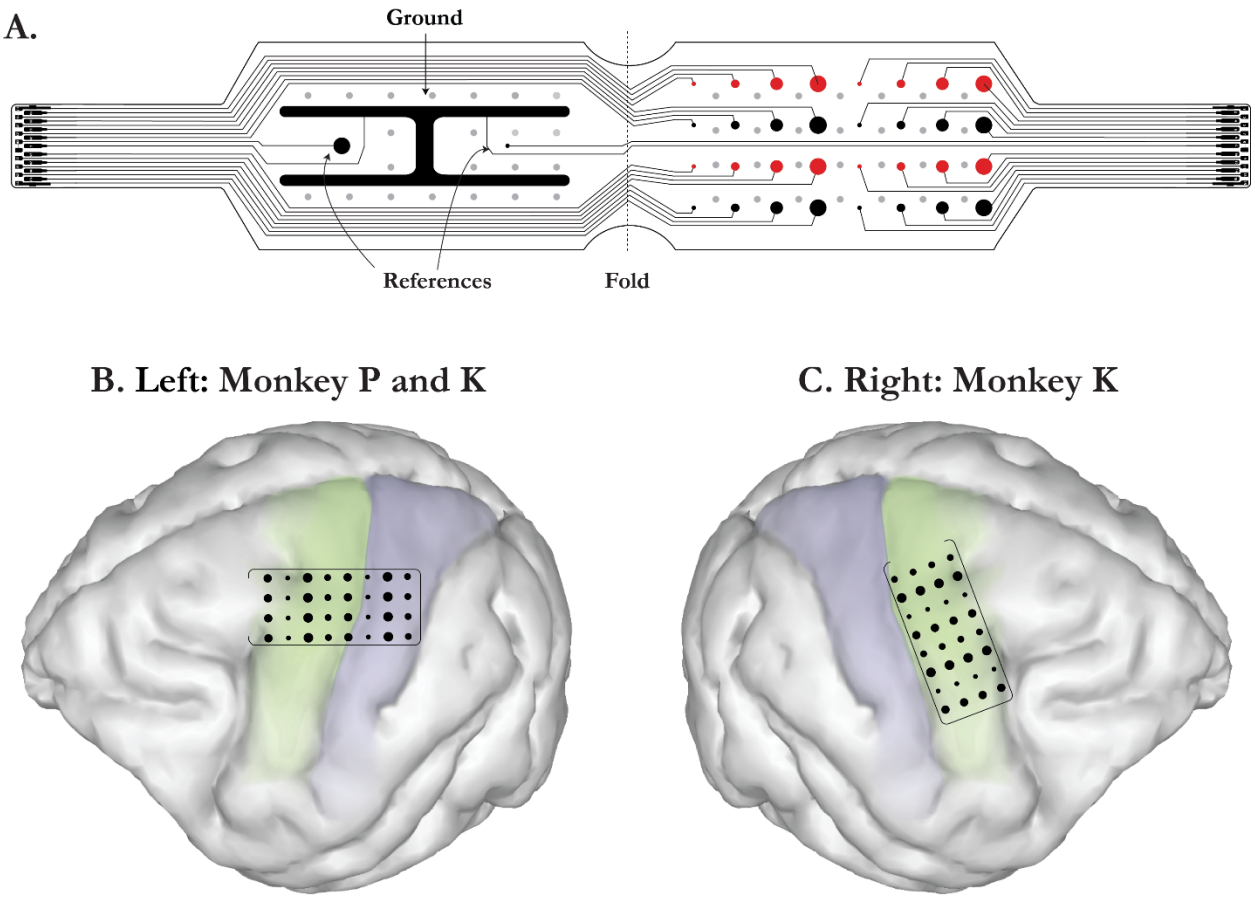

Figure 4.1: Electrode Design and placement. A. Polyamide thin film ECoG electrode array layout with one half housing 32 channels of four different sizes (300µm, 425µm, 520µm, and 600µm) of exposed electrode surface and separated by 3mm (center to center). The other half is covered with an "H" shaped ground and reference electrodes. B. Left hemisphere placement in both monkeys K and P, implanted anterior to posterior, covering the motor and sensory cortex. C. Right hemisphere placement in monkey K, implanted medial to lateral, covering the motor cortex.

4.1.3 Behavioral Task and Experimental Protocol

The subjects performed a classic center-out task with eight targets. The targets were separated by 45° starting from 0° and were placed on the circumference of the circle with a radius of 20 screen units. Both the target and the cursor were spheres with a radius of 1.5 screen units. Each trial started with the cursor at the center of the circle. One of the eight targets appeared on the screen after a 500 ms hold period. The subject then had to move the

cursor from the center to the presented target within a certain maximum movement time. If the subject was successful in doing so, the trial was marked as correct and a small water reward was given. In any other case, the trial was marked incorrect and no reward was given.

Experimental sessions were divided into blocks of trials. During each block, targets were presented in a pseudorandom order. As has been stated earlier, the BCI decoder was updated every 2-3 mins. The ECoG data collected during each block was used to do the same. In this set of experiments, each block consisted a total of 16 correct reaches (2 to each target). All the incorrect trials were ignored and a block of trials ended only after each target was correctly reached twice.

Each session of the experiment began with a block of watch task and a naïve decoder. That is, the movement of the cursor was completely controlled by the computer as the monkey watched and got rewarded every time the cursor touched the presented target. The cursor was moved at a constant speed by the computer in a straight line towards the target. This constant velocity was termed as "Bias" and had a gain associated with it. The details of the bias and how it worked in conjunction with the ECoG signals are described in Section 4.1.4.

Beginning a few minutes before the first watch task block, signals were recorded from the implanted ECoG arrays. As has been described in detail in Section 3.4, each raw ECoG signal was separated into five different frequency bands (alpha: 8-15 Hz, beta: 15-30 Hz, low-gamma: 30-55 Hz, mid-gamma: 70-115 Hz, and high-gamma: 130-175 Hz). An amplitude estimate of each frequency band, followed by a log transform and z-score, was used to create a "control feature" (electrode/frequency band pair) for the decoding algorithm. The data

collected during the watch task was used to construct an initial decoded model. At the end of each block the decoded model was updated and used for the following block. This happened throughout the session.

4.1.4 Control Signal

Subjects controlled the velocity of the cursor in the center-out BCI task for this study. A weighted linear sum of the ECoG control features along with a constant speed bias (during the first few blocks of each session) was mapped to the velocity of the cursor. This velocity signal was integrated using Euler Integration to give the position of the cursor on the screen. The two-dimensional velocity V of the cursor at a time t is given by:

$$V(t) = E(t) + B(t) \tag{4.1}$$

Where $E(t)$ and $B(t)$ represent the ECoG control signal and the constant speed bias respectively and are given by:

$$E(t) = G_f \mathbf{W}^{\mathrm{T}} F(t) \tag{4.2}$$

$$B(t) = G_b U(t) \tag{4.3}$$

$\mathbf{W}$ is the weight matrix (Nx2, N = number of features) used during the current block, $F(t)$ is a vector of z-scored features (length N) at time t, and $U(t)$ is the two-dimensional unit direction vector from the cursor to the current target at time t. The superscript T denotes

a matrix transpose. Terms G_f and G_b represent the gain applied to the ECoG control signal and the bias signal, respectively.

The weight matrix $\mathbf{W}$ was initialized to zero at the beginning of each session and updated after each block of trials using the decoding and pruning algorithms described in Sections 4.1.5 and 4.1.6. Therefore, during the first block of trials (watch task block) the velocity of the cursor was only controlled by the constant speed bias B. The ECoG feature gain term G_f was kept mostly constant throughout the days of experiment. The bias gain term G_b on the other hand started out equal to 100 during the watch task block of experiment and was gradually decreased to 0 over 8-12 blocks. This resulted in a gradual decrease in the computer assistance for the cursor's velocity control. Starting from complete computer control of the cursor during the watch task block, the control was thus slowly and completely transferred to the subject from the time G_b was set to 0. The use of this assistive bias strategy helped encourage the subjects to perform the task and transition to unassisted BCI control with ease.

4.1.5 Decoding Algorithm

The decoding algorithm (Figure 4.2 A) was designed with an intention of updating the weights based on the modulation of features, at the end of each block of trials while not causing any localized overfitting. The data during the movement period of the correct trials in the block was used to calculate a new set of weights. A fair assumption that the monkey intended to move the cursor to the target in a straight line, was made. Thus, the instantaneous

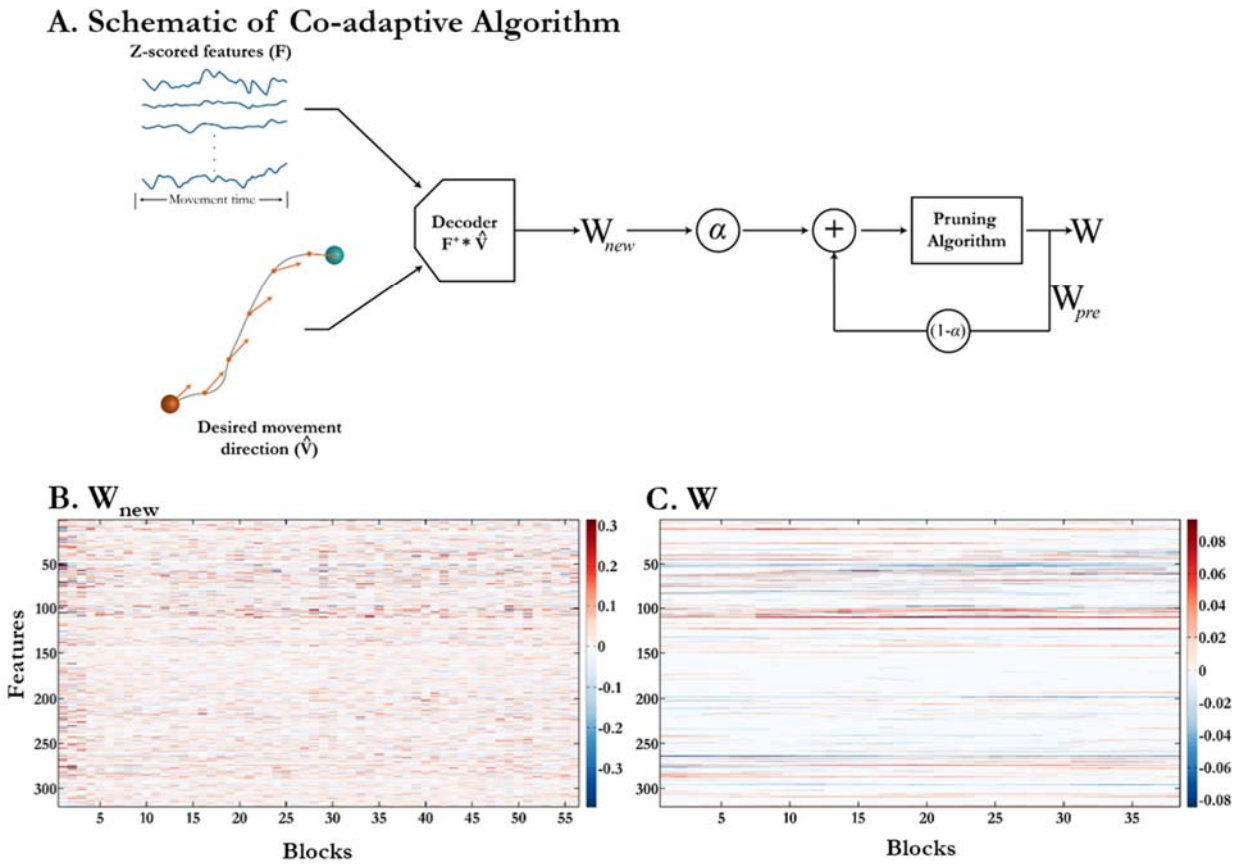

Figure 4.2: Schematic of decoding algorithm and effect of smoothing. A. Steps following the end of each block of 16 correct trials. The z-score of features recorded during the movement time of each trial and the desired cursor movement direction is supplied to the decoding algorithm to produce the weights for that block. These weights are then combined with the previously used weights using a smoothing factor α. These smoothed weights are then pruned using the pruning algorithm before being used for the next block of trials. B. $\mathbf{W}_{new}$, weights assigned from the current block of trials. C. $\mathbf{W}$, weights produced using the smoothing factor before being used for the following trials. Use of smoothing factor reduces the variations in the weights from one block to another, thus preventing overfitting and sudden changes from one block to another.

unit vector in the desired movement direction was used as the response variable ($\widehat{\mathbf{V}}$). The modulation of features ($\mathbf{F}$) during the movement period was used as the independent variable that predicted the movement direction. Weights were calculated using the following equation:

$$\mathbf{W}_{new} = \mathbf{F}^{+}\widehat{\mathbf{V}} \tag{4.4}$$

Where $\mathbf{W}_{new}$ is the Nx2 matrix of new weights calculated based on previous block, and + denotes the Moore-Penrose inverse.

To prevent any overfitting and sudden changes from one block to another, a smoothing technique similar to one previously described in intra-cortical BCI studies was implemented [32], [61]. An update smoothing factor α was used to blend the new weights with the previously used weights. Alpha was set to 0.1 or 0.2 in this sets of experiments.

$$\mathbf{W} = \alpha\mathbf{W}_{new} + (1 - \alpha)\mathbf{W}_{pre} \tag{4.5}$$

$$\mathbf{W}_{pre} = \mathbf{W} \tag{4.6}$$

Where, $\mathbf{W}_{pre}$ is the matrix of previously used weights and $\mathbf{W}$ is the matrix of weights to be used for the upcoming block. Figure 4.2 B shows the large variations in weights of individual features from one block to another. These variations are greatly reduced and the weights are stabilized over time when the above equation is used with alpha set to 0.2 (Figure 4.2 C).

4.1.6 Pruning Algorithm

The schematic of the steps carried out by the BCI system to update the weights after each block of trials is shown in Figure 4.2 A. The decoding algorithm described in the previous section has been previously used in other experiments. The pruning algorithm described in this section was designed to work in conjunction with this decoder.

As described earlier, each session of experiment began with a naïve decoder and a watch task block. For the initial blocks with computer assistance and 1-2 blocks after that, the decoder was used without pruning any weights. This allowed the weights to stabilize for the current session of experiment. After this period the pruning algorithm was introduced in the chain as shown in Figure 4.2 A. Starting at 10% pruning in the first few weeks, this percentage was increased in steps of 10% every 1-2 weeks of experiments. Experiments were carried out until both subjects had reached 70% pruning; that is, using only 30% percent of the available feature space for control.

Features were pruned after they were smoothed using Eq. 4.5. Each feature was assigned a weight in the cardinal directions. The first step of the pruning algorithm was to calculate the magnitude and direction of each feature using the following equations:

$$r = \|\mathbf{W}\| \tag{4.7}$$

$$\theta = tan^{-1}\left(\frac{W_y}{W_x}\right) \tag{4.8}$$

Where r and θ are the vectors of magnitude and direction of each feature, respectively. $\|\ \|$ represents the Euclidean or ℓ_2 norm and $\mathbf{W}$ is the weight matrix obtained from Eq. 4.5, with W_x and W_y as its components in the cardinal directions.

Each feature was then segregated into one of the 8 sectors spanning the 2D space. Each of these sectors were centered at the directions of the targets used in the task and occupied an angular region of $\pi/4$.

$$I_j = \left\{\forall i : \left\{\theta(i) \leq j * {}^{\pi}/_{4} - {}^{\pi}/_{8}\right\} \cap \left\{\theta(i) > (j-1) * {}^{\pi}/_{4} - {}^{\pi}/_{8}\right\}\right\}, 1 \leq j \leq 8 \quad \textbf{(4.9)}$$

I_j in the above equation is a set of indices of features that lie in the sector j. The conditional statement in the set builder marks the boundaries of each sector. A vector of segregated feature magnitudes (M_j) was created and sorted in increasing order. The sort function produced an output of indices of the sorted feature set from each sector (given by ind_j). Eq. 4.10 and 4.11 were used for the same.

$$M_j = r(I_j) \quad \textbf{(4.10)}$$

$$[ind_j, M_j] = sort(M_j) \quad \textbf{(4.11)}$$

Depending on the pruning percentage, the number of features to be pruned from each sector were selected using the following equation:

$$ind_j = ind_j(k), \qquad 1 \leq k \leq \lfloor f * num(M_j) \rceil \quad \textbf{(4.12)}$$

Where, f is the fraction of feature to be pruned (*e.g.* 0.3 for 30% pruning), and function $num(\,)$ gives the number of elements in the input vector. The integer parameter k is the index and $\lfloor\,\rceil$ represents a closest integer function. The above equation produced the indices of vectors with magnitudes in the bottom f fraction of the distribution in sector j. The magnitude of these features were set equal to 0 (Eq. 4.13) in the magnitude vector obtained in Eq. 4.7. The magnitude and direction vectors were then combined and converted into cardinal

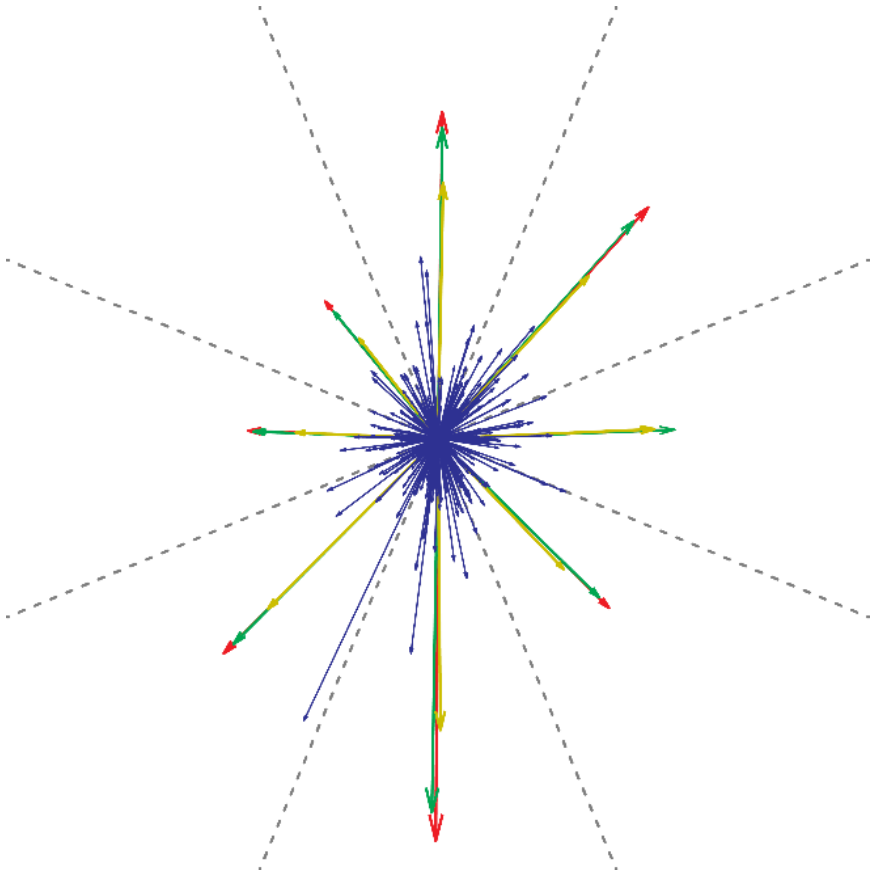

Figure 4.3: Distribution of feature weight vectors in 2D space. The 2D space was divided into 8 sectors centered at each target and zero weights were assigned to the smallest x% of the weights in each sector, thus preserving the distribution of features. The co-adaptive algorithm used the whole feature space for each update and same percent of features were pruned for 1-2 weeks of experiment. The red, green, and yellow vectors in each sector represent the sum of features at 0, 20, and 50% pruning.

weights assigned to the matrix $\mathbf{W}$, to be used in the following block of trials. The $\mathbf{W}_{pre}$ weight matrix was updated using Eq. 4.6.

$$r\left(I_j\left(ind_j\right)\right) = 0 \tag{4.13}$$

$$\mathbf{W} = \langle r, \theta \rangle \tag{4.14}$$

Figure 4.3 shows a representative distribution of weights (blue arrows) and the resultant vectors in each sector. Red arrows represent a sum of all the features in a given sector. Green

and yellow arrows represent the sum of features at 20% and 50% pruning respectively. A decrease in the magnitude of the resultant vector in each sector is seen with increasing pruning percentage. However, the directions of the resultant do not vary a lot. This is expected as the resultant vectors are a linear combination of vectors from a small sector.

4.2 Results

4.2.1 Psychophysics and Task Performance

The performance in a 2D center out task can be quantified by the movement times to target and the trajectories taken to reach it. To maximize the water reward subjects would try to reach the target as fast as possible and take the shortest path to get there. Therefore, shorter movement times and straight line trajectories are an optimal choice.

Each correct trial was segmented into 40 time-bins starting from center to the target and a mean cursor position in this time bin was calculated. This produced 40-point trajectories for all the trials during the experiment. Trials were grouped by the target presented. Mean 40-point trajectories were calculated for each group. A distribution of perpendicular deviations from the straight line trajectory was created at each of the 40 time points for all the trials in a group. The standard deviation of these distributions was used to plot the trajectory deviations.

Figure 4.4 shows the differences in performance with feature sets used at different percentages of pruning. Each row in Figure 4.4 represents the performance during one session of the experiment at a pruning percentage between 0 and 70. Black lines in each plot show the mean trajectories to the 8 center-out targets. The color of the patch around each mean

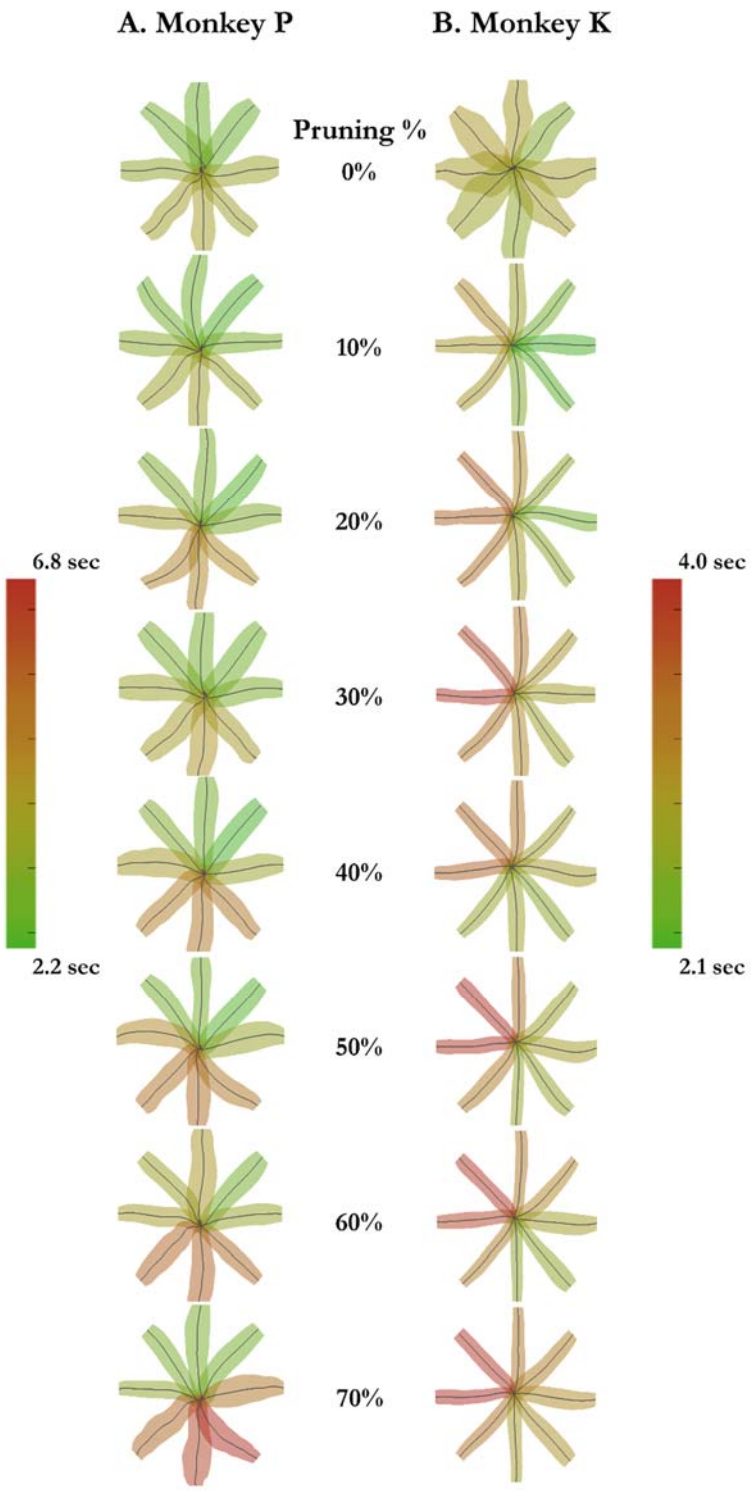

Figure 4.4: Performance during 2D BCI Task A. Monkey P, B. Monkey K. Mean and Standard deviations of trajectories to each target at different levels of pruning. Mean trajectories are represented with black lines. The patches around each trajectory represent the standard deviations and are colored to show mean movement times to each target.

trajectory is given by the mean movement time to a particular target. The patches were created using the distributions of perpendicular deviations from straight line trajectories. Each point on the patch is one standard deviation (above and below) away from the mean trajectory.

It is clear from Figure 4.4 that both monkeys P and K perform near straight line reaches to each of the targets. The trajectory standard deviations are low and do not change as the number of features are reduced from 0 to 70% pruning. However, changes in movement times are observed. Both the subjects tend to get slower with the pruning percentage. A near monotonic increase in movement times is observed for each of the targets. It is interesting to see that the variation of movement times with the targets is maintained. That is, targets that were relatively slower to reach at 10% pruning are slower to reach at 70% pruning as well and vice-versa.

4.2.2 Distribution of Selected Bands

Post hoc analysis was conducted to determine the distribution of features selected for control. Features were grouped by their frequency band to analyze each band's relative contribution as the pruning percentage was changed. Figure 4.5 shows the variations in the feature space with the pruning percentage. As each raw signal was divided into five different frequency bands, the recorded feature space consisted 20% each of alpha, beta, low-gamma, mid-gamma, and high-gamma features. Data collected during 1-2 weeks of experiments for each pruning percentage was used to calculate the average contribution of frequency bands. The total height of each stacked bar in Figure 4.5 is a representative of the fraction of feature space used for control. Sizes of individual bars within each stack represent the relative

contribution of each band at that pruning percentage. This contribution is also represented in percent contribution by the numbers in each bar.

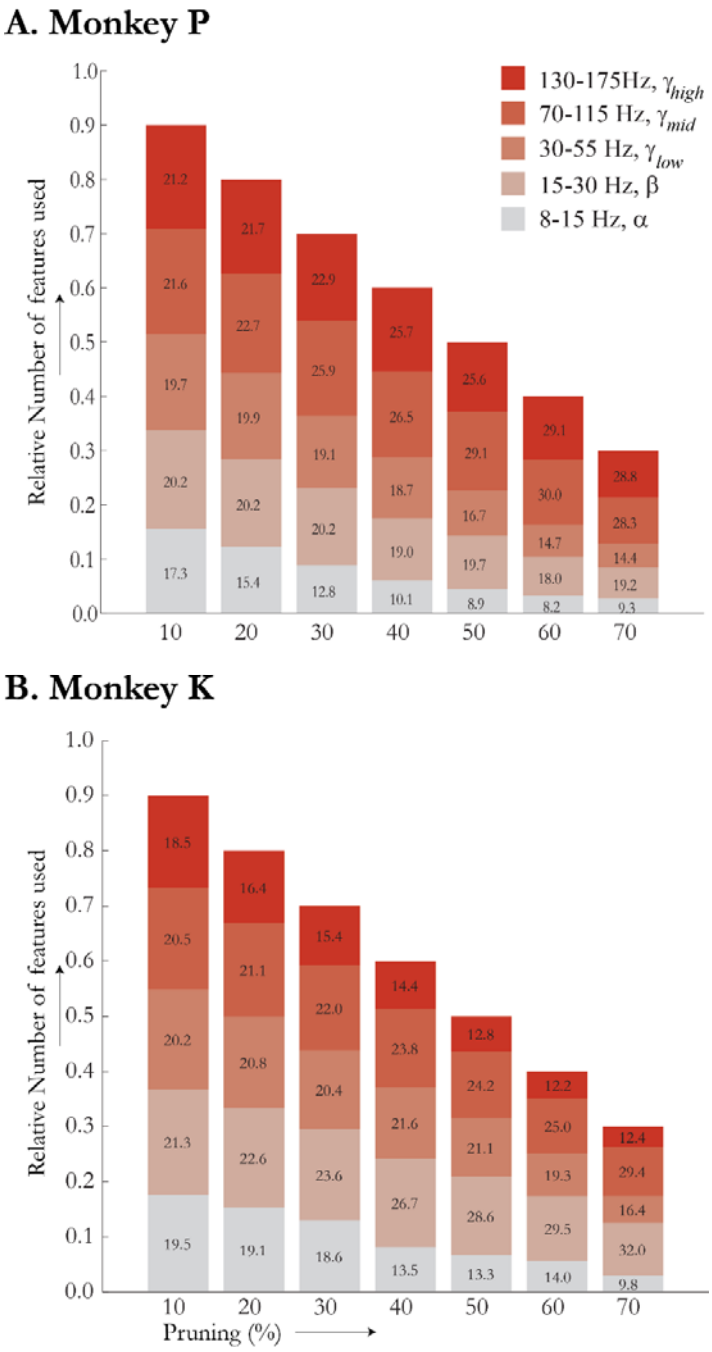

Figure 4.5: Relative usage of frequency bands at different percentages of pruning A. Monkey P, B. Monkey K. Each bar gives the distribution of features used by the subject at different levels of pruning. The numbers for each band represent the percentage of times a feature from band was picked. Both the subjects prefer features from the mid-gamma band. On one hand subject P is more inclined towards features from high-gamma band whereas subject K prefers beta band.

The first observation evident from Figure 4.5 is that, in both the subjects, some bands were pruned more than the others. An almost monotonic increase in contribution from mid-gamma band is observed going from 10% to 70% pruning, in both Monkey P and K. It is interesting to see, however, that besides mid-gamma band Monkey P prefers high-gamma band but Monkey K does not. On the contrary, Monkey K shows preference to beta band which appears to be Monkey P's third choice. Both the subjects also show the least preference to alpha band which sees a steep decrease in its contribution from the beginning.

4.2.3 Distribution of Selected Electrodes

Features were further grouped by their recording electrodes to determine the differences in contributions from regions covered by the ECoG arrays. Due to a large number of electrodes from both the subjects (31 and 60), their preference becomes clear only when most of the recorded features are eliminated. Figure 4.6 shows the relative contribution of each electrode at 70% pruning. Electrodes missing on the arrays had very noisy signals and were not used through the course of experiments.

If features from each electrode were equally selected, the relative contributions from each electrode would be 0.032 (1/31) or 3.2% in case Monkey P and 0.017 (1/60) or 1.7% for Monkey K. It is seen in the figure that individual electrodes contribute towards the selected features both more and less than those values. This indicates that certain electrodes were more significant in control than others. The approximate position of the arrays on the brain was known from when the implantation surgery was carried out. Using that knowledge, the

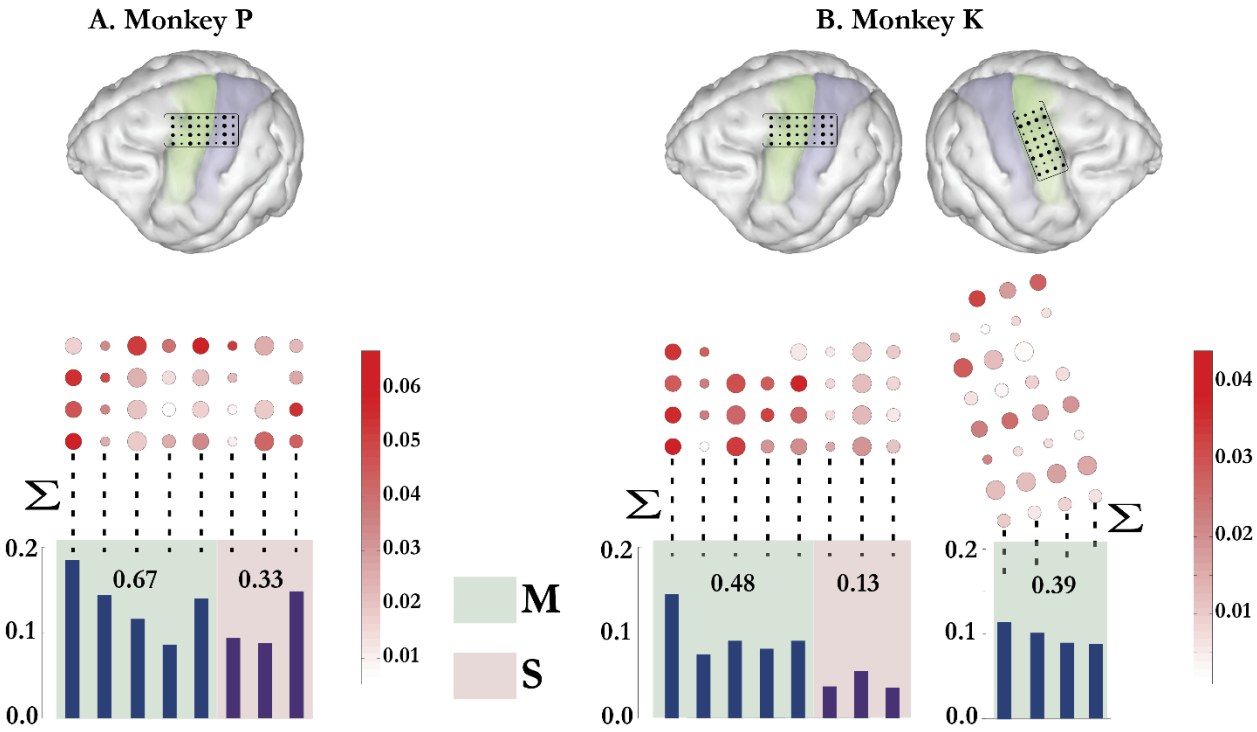

Figure 4.6: Contribution from each recording electrode at 70% pruning A. Monkey P, B. Monkey K. The intensity of color of each electrode represents the relative number of times features from that electrode were picked by the pruning algorithm. Equal contribution from all electrodes would lead to 0.032 in A, (1/31) and 0.017 in B (1/60). The electrodes are separated into motor and sensory based on their approximate location. Contributions from each row of electrodes was combined to plot the bar plots. The numbers in each section (green or red) of the plot gives a sum of contribution from the rows in motor or sensory electrodes. No difference is observed in the feature contribution from motor or sensory electrodes with percent contribution approximately equal to the fraction of electrodes over the two areas.

electrodes were separated into motor and sensory groups to determine the differences in their contribution. The bar plots at the bottom of the figure show the total contribution from each row of electrodes. These rows were then separated into motor and sensory, and the total contributions from the two regions was calculated.

Monkey P had 11 of his 31 ($11/31 \approx 0.35$) electrodes and Monkey K had 12 of his 60 ($12/60 = 0.20$) electrodes on the sensory cortex. The numbers in the bar plot indicate that the features recorded from the sensory are slightly less than their relative number in the whole set.

However, there isn't a dramatic difference between the relative number of sensory electrodes and their contribution to the control feature set. In any case, a contribution from the features recorded from the sensory side, even when only 30% of whole feature set was selected, indicates these features were reliable for BCI control.

To further dissect the selected features during the experimental sessions with 70% pruning, each electrode was segregated into the five different bands. Figure 4.7 shows the relative number of times each band from a given electrode was chosen. As expected from the observations in Figure 4.5, alpha and low-gamma bands are very rarely selected from any of the electrodes in both monkeys except for some motor electrodes on Monkey K's left hemisphere implant. Mid-gamma band is selected often in both subjects. However, mid-gamma band is almost never selected from the sensory electrodes in Monkey K. High-gamma band in Monkey P shows a lot of overlap with mid-gamma and is not selected often in Monkey K. Beta band is observed to be selected very often from almost all the electrodes in Monkey K. On the other hand in Monkey P where only a few beta band features were selected, sensory electrodes are observed to be the predominant contributors. All these observations point towards differences in modulation of sensory and motor features with respect to the recorded band. The analysis in the following section investigates this difference further.

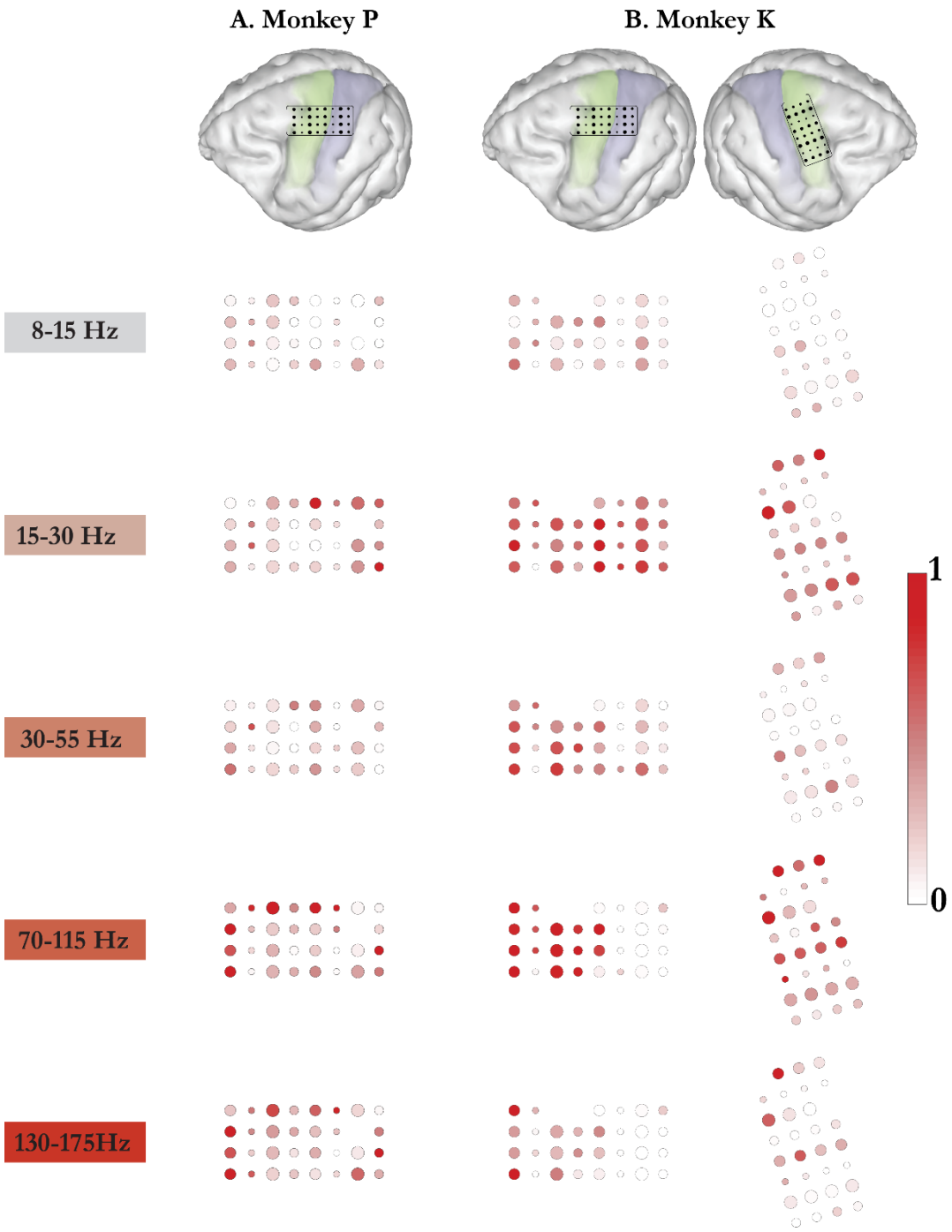

Figure 4.7: Segregation of features selected at 70% pruning into individual frequency bands. A. Monkey P, B. Monkey K. The intensity of each electrode represents how often that particular feature was picked. In both the subjects alpha and low gamma bands play very little role and very rarely selected for control. Subject P is observed to prefer the features over motor areas from mid and high gamma bands and features over sensory area from beta band. Subject K on the other hand prefers almost all beta band features from the array on the left hemisphere and most of features over the motor areas from the mid gamma band.

4.2.4 Tuning of Individual Features

This section studies the tuning of individual features with the movement of the cursor in the 2D plane. The modulation (z-score of amplitude estimate) of each feature during the movement of the cursor was used as the independent variable. This variable was fit to the velocities in the cardinal directions using multiple linear regression. Data from all the correct trials was combined, and an R^2 statistic was determined for the fit for each block. The mean R^2 value for each feature for one session of the experiment was used in the following figure.

Figure 4.8 shows the mean R^2 values for each feature separated by frequency bands and different percentages of pruning. Within each frequency band, the features were rank ordered based on their R^2 value. The color of each bar was set to red if there was an above chance probability of selecting the feature through the experiment session. The color was set to blue if this probability was below chance. The horizontal lines in each frequency band represent the mean R^2 value of the red and blue features. The chance level was determined by the pruning percentage in any given session. For instance, at 30% pruning 70% percent of the features were selected for control. Therefore, if a feature was selected in more than 70% of all updates during the experiment session, it was termed as an above chance feature and assigned the color red. On the other hand, any feature that was selected in less than 70% of the updates was colored blue.

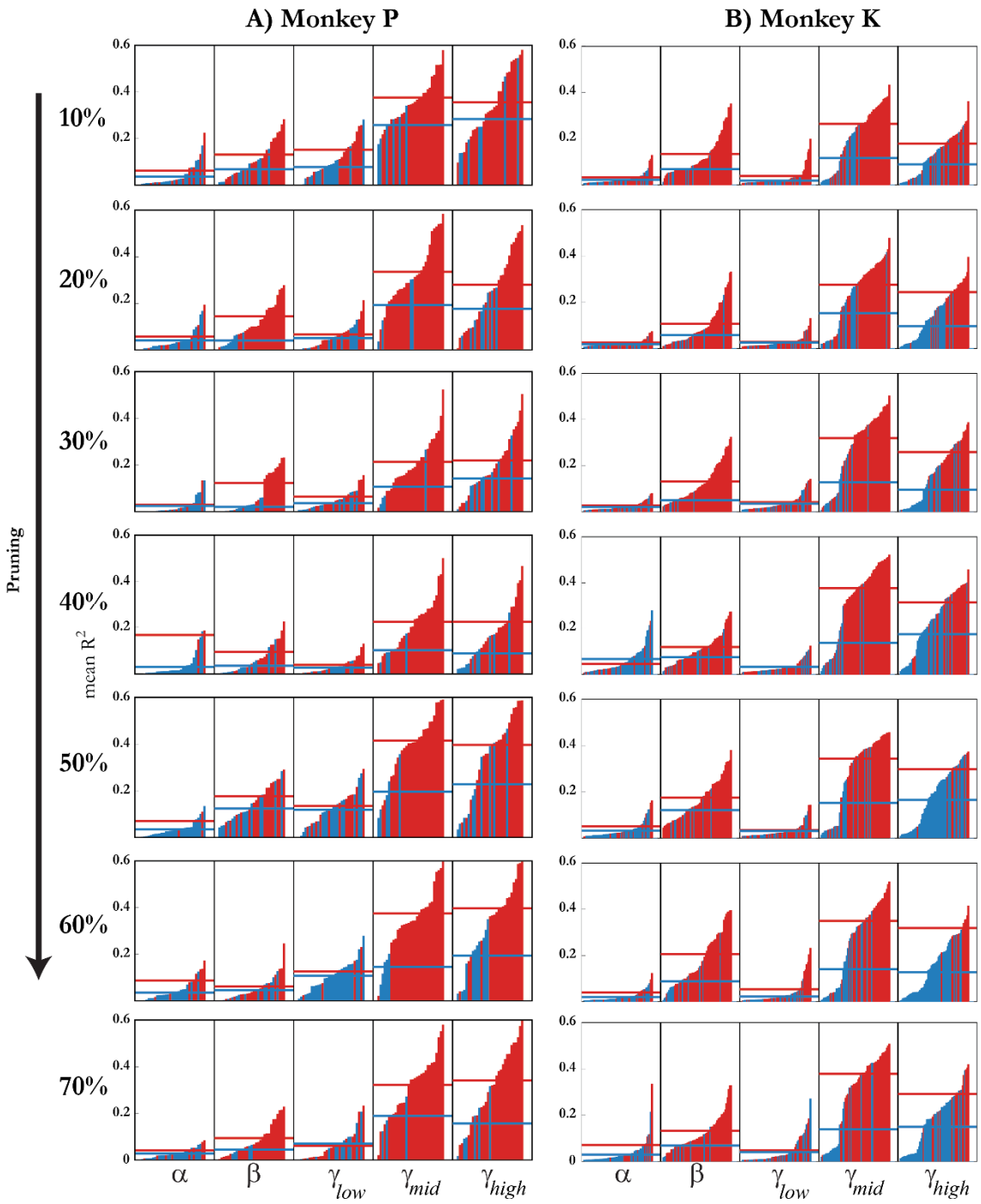

Figure 4.8: Mean R^2 values of features for different percentages of pruning, grouped by frequency bands and rank ordered. A. Monkey P, B. Monkey K. Modulation of each features during cursor moment periods of each correct trial was fit to the cursor velocity in *x* and *y* direction during every block of experiment. This figure shows the R^2 values of that fit. The color of each feature represents if it was picked for control below (blue) or above (red) chance level. The horizontal lines represent the means of red and blue features. Higher values are observed in mid ad high gamma band followed by beta band. Beta band values for Monkey K are observed to be higher than those for Monkey P.

For both the monkeys, it can be seen that the mean R^2 values of above chance (Red) features were higher than that of the below chance (Blue). Individual feature R^2 are higher in mid-gamma and high-gamma bands. Monkey K was observed to show preference towards beta band in Figure 4.5, which is also reflected in higher R^2 values compared to that of Monkey P. Alpha and low-gamma band show low values in both the cases.

As in the case of electrode selections, each feature was separated into motor and sensory and the distributions of R^2 values in each band determined. Figure 4.9 shows the mean (lines) and standard deviations (patch) of these distributions at 70% pruning. Features recorded from the motor electrodes are colored in green and features from sensory electrodes are in red. In both the subjects the motor features are observed to show higher R^2 values in mid-gamma and high-gamma bands; the differences in the distributions are however only significant in Monkey K. On the other hand, sensory features in the beta band show higher values compared to the motor features; with significant differences in Monkey P.

It was observed from Figure 4.7 B. for Monkey K that, the sensory features from mid-gamma and high-gamma bands were almost never selected from control. The huge difference in the R^2 values in these two bands for Monkey K strengthens that observation. Looking at the beta band, the differences are significant only in Monkey P. This is also reflected in the Figure 4.7 A where sensory features were most likely to be selected for control. Finally, R^2 values lower than the other bands are observed in alpha and low-gamma bands from both the subjects.

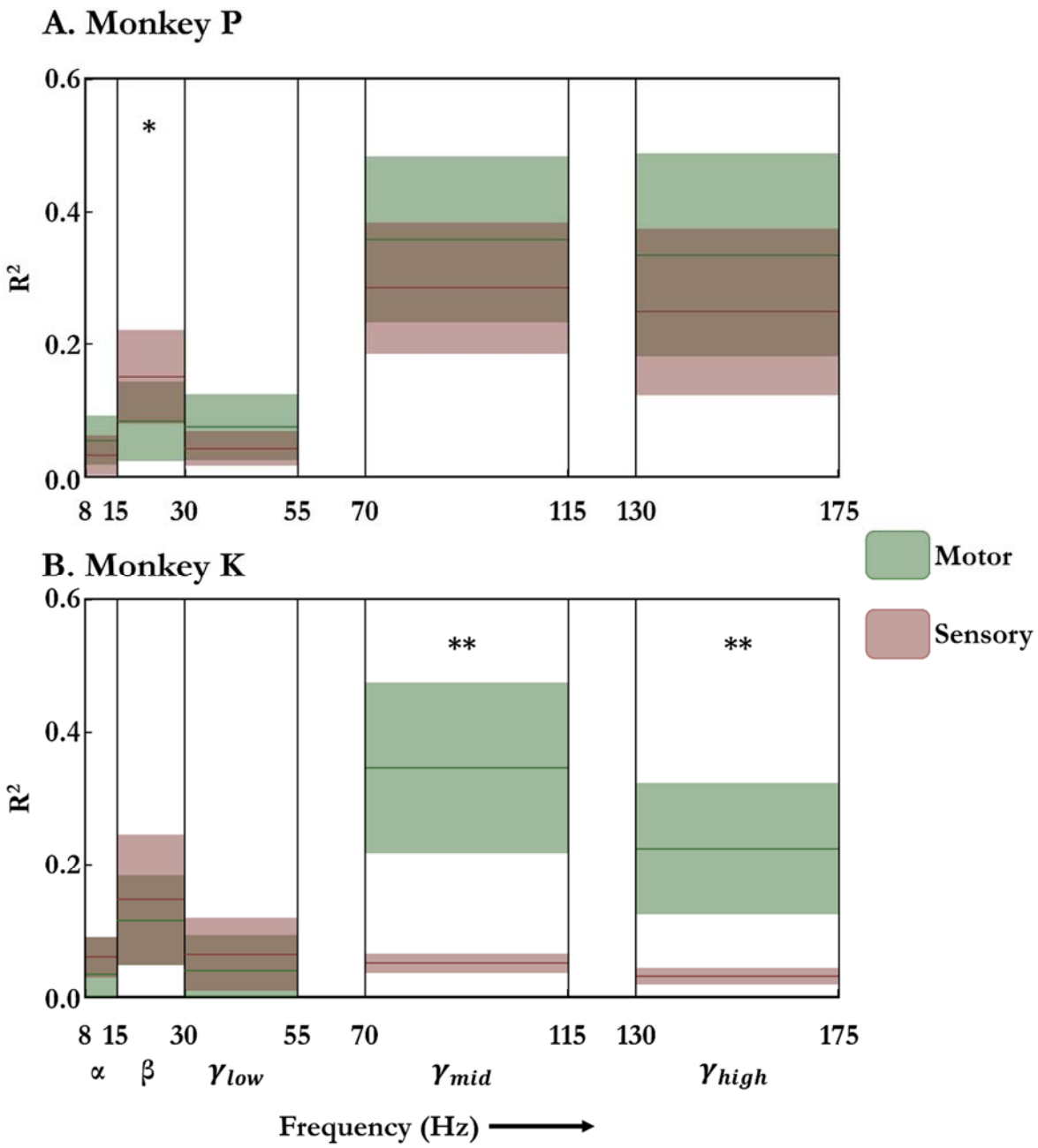

Figure 4.9: Mean and standard deviations of R^2 values of electrodes over motor and sensory areas at 70% pruning. A. Monkey P B. Monkey K. Darker lines in the plot represent the mean with the patch representing one standard deviation around the mean. Significant differences between motor and sensory R^2 distributions are marked with * ($p<0.05$) and ** ($p<0.01$). Both subjects show a low tuning of alpha and low gamma features. Subject P is observed to have higher tuning in mid and high gamma bands with significant overlap between the motor and sensory electrodes. Subject K on the other hand shows huge differences between the motor and sensory electrodes in mid and high-gamma bands. The distributions of R^2 values in these plots are similar to the distribution of features picked from each band shown in Figure 4.7, confirming that the features with higher individual R^2 values were chosen for BCI control.

4.3 Conclusions

From the performance results (Figure 4.4) it was observed that the pruning algorithm was successfully used in conjunction with the decoder to obtain robust control of 2D BCI. Using only 30% of the available features, both the subjects were able to control the movement of the cursor in an efficient straight-line trajectory towards the desired target. The movement times were observed to increase with the pruning percentage. As has been noted earlier the gain term on the ECoG features was kept fairly constant throughout the experiments carried out for this chapter. The control signal produced by a smaller feature set would likely be of a lower total magnitude and thus, result in smaller velocities of the cursor. This was probably the reason the changes in movement times were observed. However, these changes were relatively small compared to the changes in the feature set. This is an indication of most eliminated features being noisy and not significantly contributing towards the movement.

Proceeding towards the frequency distribution of features (Figure 4.5), it is observed that both the subjects tend to prefer two of the five available bands for control. A monotonic, almost linear increase in the two preferred bands is observed as the pruning percentage is increased. The other three bands see a monotonic decrease in their contribution. The differences in the two subjects with respect to the preferences towards the band are an interesting observation. One possible explanation could be the noise in the recorded ECoG signals. ECoG signals from Monkey K were relatively noisier, and as the higher frequency signals are the first to get affected by noise due to their lower spectral power, the monkey chose to use a frequency band that was more reliable than the high-gamma band. This could

have resulted in a preference towards mid-gamma and beta band features. An in-depth analysis of the noise in the signals is required to enforce this hypothesis.

The observations from the selected electrode distribution (Figures 4.6, 4.7) and individual feature tuning (Figures 4.8, 4.9) are not only relevant for ECoG BCI but also produce some interesting insights about the underlying neurophysiology. The most important observation from the electrode distributions is that though the contribution of sensory features was a little lower than their relative number, they were still present at 70% pruning. This is a very important observation because it shows that even though sensory cortex does not naturally control performed movements, it can be used for BCI control. This can be attributed, to some extent, to the plasticity of the cortex.

The second significant observation is the differences observed in the modulation of motor and sensory features in different bands. Sensory features were observed to have better tuning than the motor features in the beta band of both the monkeys (Figure 4.9). Even in Monkey P where very few features were used from the beta band, the ones that were used were from sensory electrodes (Figure 4.7). In the higher gamma bands, the difference in tuning between the motor and sensory features is ambiguous. On one hand where Monkey K shows a huge difference, Monkey P doesn't.

To draw an overall conclusion, the experiments in this chapter demonstrate that a small subset of the available features can be used for BCI control. The post hoc analysis of the selected features also determines the frequency spectrums most significant in BCI control.

The results from both the subjects point towards the gamma frequencies above 70 Hz and the beta band between 15-30 Hz.

The observations from these experiments raise some interesting questions. What would happen if the most reliable features were eliminated and the monkeys were only allowed to use the features that were initially unreliable? How would the control in such experiment change? Can the subjects learn to modulate features that were initially not used? With the intentions of finding answers to these questions, the experiments in Chapter 5 were carried out.

Chapter 5
Selective Feature Pruning of Consistent Features

Chapter 4 described a feature pruning method to find the optimal feature subset for BCI control. By gradually eliminating the noisy, unreliable features subjects were trained to use a small subset of features with the largest and most consistent weights. The experiments in this chapter take on an "opposite" approach of eliminating the features that were most effective in controlling the cursor movement. The goal of this chapter is to reinforce the findings from Chapter 4 and get a better understanding of the strategies used by the subjects to accomplish the BCI task. By eliminating the "good" control features, subjects are compelled to recruit features that were previously noisier and did not significantly contribute to the movement of the cursor.

This chapter is divided into three different sections. Section 5.1 describes the methodology used in for the experiments performed. Section 5.2 describes the results obtained from the experiments. Post hoc analysis of the collected data was used to determine the variations in the BCI control, distribution of features, and changes in modulation of individual features with the gradual reduction in the control set. Finally, Section 5.3 concludes the chapter with a discussion of the observed results and their overall implication towards ECoG BCI.

5.1 Methods

5.1.1 Subject

One male Rhesus Macaque (*Macaca mulatta*, monkey I) was used as a subject in the experiments described in this chapter. Monkey I had a bilateral implant on both left and right hemispheres of his brain. The experiments described here were carried out almost three years post implant. The subject had performed several behavioral and BCI experiments during this three year period. Therefore, for all intents and purposes, it can be safely assumed that the subject was fully trained in the behavioral task. The only novelty experienced by the subject, while performing the task, was the pruning algorithm described in this chapter.

5.1.2 ECoG Arrays: Design and Placement

Monkey I was implanted with two Silastic electrode arrays manufactured by Ad-tech Medical (Left hemisphere) and PMT Corporation (Right hemisphere). The details of the array design and the electrical connectors are described in Section 3.2.2. Both arrays were implanted in the rostral-caudal direction. The center of the 19 mm craniotomy was located at the Stereotactic coordinates (Anterior 6 mm, Lateral 14 mm). This places the array partially over the motor and sensory cortices of both the left and right hemisphere. The layout of the electrode array and their approximate location is shown in Figure 5.1. Both the arrays are observed to have several electrodes on both sides of the central sulcus.

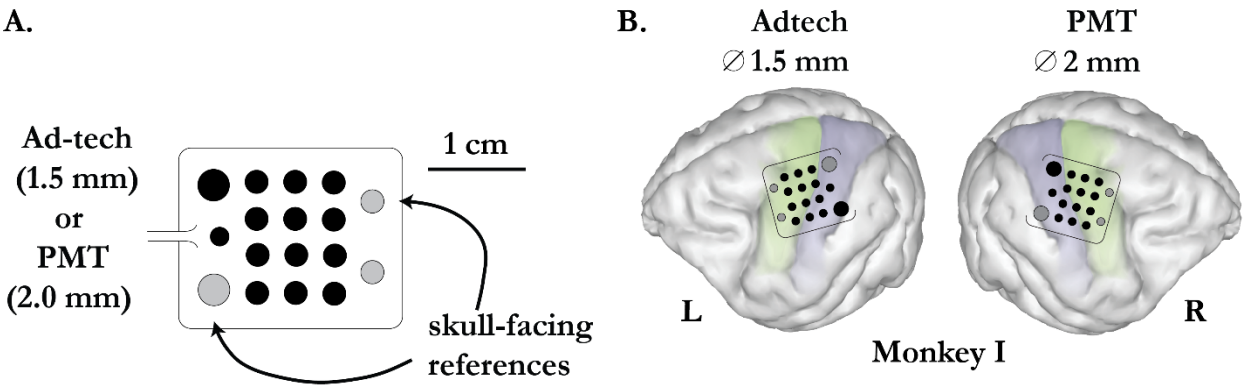

Figure 5.1: Electrode Design and placement. A. ECoG electrode array with 14 active Pt electrodes and three skull facing references embedded in silastic. B. Left and right hemisphere in monkey I were implanted with one array each with 1.5mm diameter electrodes (L, Ad-tech) and 2mm diameter electrodes (R, PMT), respectively. Both the arrays covered a part of motor and sensory cortices as shown in the picture.

5.1.3 Behavioral Task and Experimental Protocol

The subject performed a classic 2D center-out task with eight targets. The targets were equally spaced at 45° from each other, starting at 0°, on the circumference of a circle. The radius of this circle was set at 20 units. Both the target and the controlled cursor were spheres with a radius of 1.6 units. Each trial started with the cursor at the center of the circle. After a hold period of 500 ms, one of the eight targets appeared and the monkey had to move the cursor to the target in a certain amount of time. If successful, the monkey was rewarded with a drop of water and the trial was marked as correct. If the target was not reached in the allowed maximum time, the trial was marked as incorrect and no reward was given.

As in Chapter 4, experimental sessions were divided into blocks of 16 correct trials (2 to each target). Each session started with a naïve decoder that was updated after every block of trials. The incorrect trials in the block were ignored. The constant speed assistive bias approach used in the previous chapter was also used in these set of experiments. The first

block of each session was a watch task block with complete computer control of the cursor. As the decoder was updated after each block of trials, the gain associated with the assistive bias was gradually reduced to completely transfer the control to the subject. The details of the bias and how it was employed in conjunction with the ECoG signals, presented in Section 4.1.4 of Chapter 4, also applies to experiments in this chapter. The recorded ECoG signals were separated in the five different frequency bands and the signal processing steps outlined in Section 3.4 of Chapter 3 were applied. These signals formed the "control features" of the BCI system and the decoding algorithm described in Section 4.1.5 of Chapter 4 was used to assign weights to each one of them.

5.1.4 Pruning Algorithm

The pruning algorithm for this chapter was designed to eliminate the features that were most consistently weighted by the decoding algorithm and hence were the most reliable for control. Unlike the algorithm in Chapter 4, this algorithm was used to eliminate the features before they were received by the decoder. The rationale behind this was that because the most reliable features were eliminated, the subject needed to be given an opportunity to modulate the other features that were previously noisy. If the features were pruned after the decoder had generated a model, the subject would not be able to use the remaining feature set. The schematic of the decoding algorithm along with the pruning step is shown in Figure 5.2. As it can be seen, the pruning algorithm is used before the decoded model was created, unlike in Figure 4.2 where features were pruned after they were blended using the smoothing factor α.

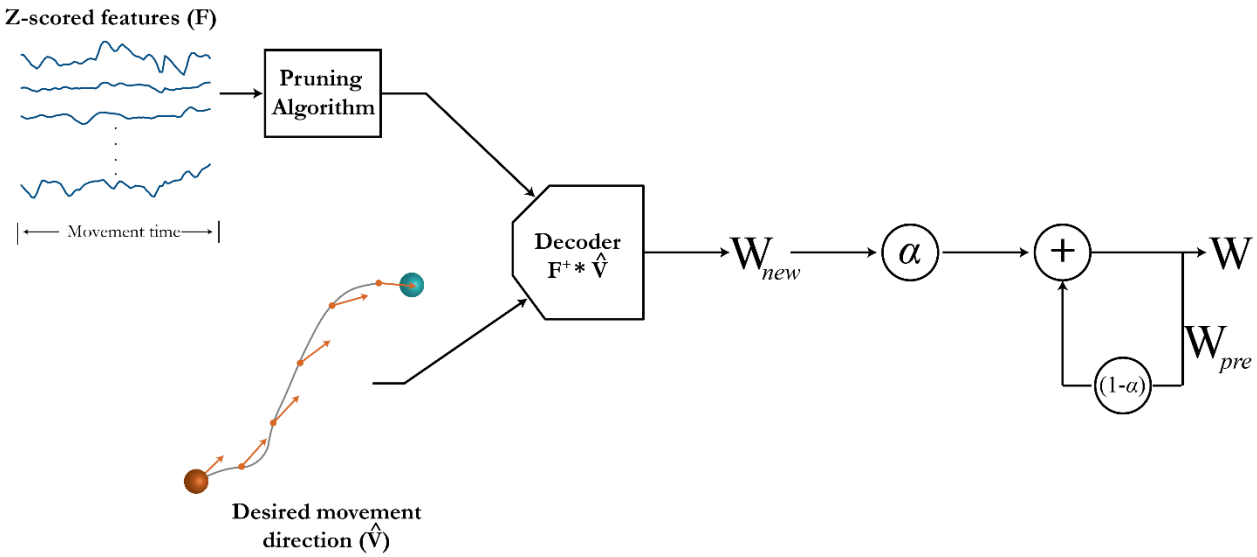

Figure 5.2: Schematic of decoding algorithm. The same decoding algorithm described in Chapter 4 was used here. The pruning algorithm that was used to gradually eliminate the most consistent features was, however, applied before the decoder. This was done to enable the subject to learn to modulate the features that did not significantly contribute to the 2D movement of the cursor. The consistency of a feature was determined using a circular resultant vector metric which produced an output of 0 for a feature assigned preferred directions randomly distributed in 2D space and an output closer to 1 for a feature assigned preferred directions close to each other across updates.

After 1-2 weeks of conducting experiments with the whole feature set, the pruning algorithm was used offline to determine which features were most consistent. The top 20% of consistent features were picked and were eliminated from the control set for the following 1-2 weeks. This method was employed three consecutive times, each time using the data from previous 1-2 weeks of experiments and eliminating another 20% of the total number features. Thus, 20%, 40%, and 60% of features were pruned and experiments were conducted for each of those values for 1-2 weeks. At the end of 60% pruning, another 2 weeks of data was collected with the whole set of control features.

The pruning algorithm in this chapter was fundamentally different from the one used in Chapter 4. Instead of preserving the 2D distribution of features, this pruning algorithm

eliminated features that were the most reliable for control and were consistently assigned similar preferred directions every time the decoder updated the model. To quantify the "consistency" of a given feature, a mean resultant vector length (given by function circ_r) metric from the CircStat toolbox for MATLAB was used [65].

As described earlier, experiments were carried out at a given percentage of pruning for 1-2 weeks; starting from all features (0% pruning) to 60% percent pruning and back to all features. The decoder model was updated several times during this period of 1-2 weeks. A distribution of the preferred direction of each feature was created from $\mathbf{W}_{new}$ combined across all days, neglecting the updates when any computer assistance was present. This distribution was used to calculate the mean resultant vector length metric from the CircStat toolbox. The function calculated the mean resultant vector of unit vectors pointing in the directions present in the distribution. Thus, a feature that was assigned the same preferred direction had a mean resultant vector length longer than the one that was noisy and assigned random directions across decoder updates.

After calculating the resultant vector length for each feature, the longest 20% of features were eliminated from the control set for the following 2 weeks. Going from one pruning percentage to next, the number of features eliminated was 20% of the total number of features and not the number of features in the current set. Also, distributions were created using $\mathbf{W}_{new}$ and not $\mathbf{W}$ as the values in $\mathbf{W}$ were dependent on one another due to the use of smoothing factor α.

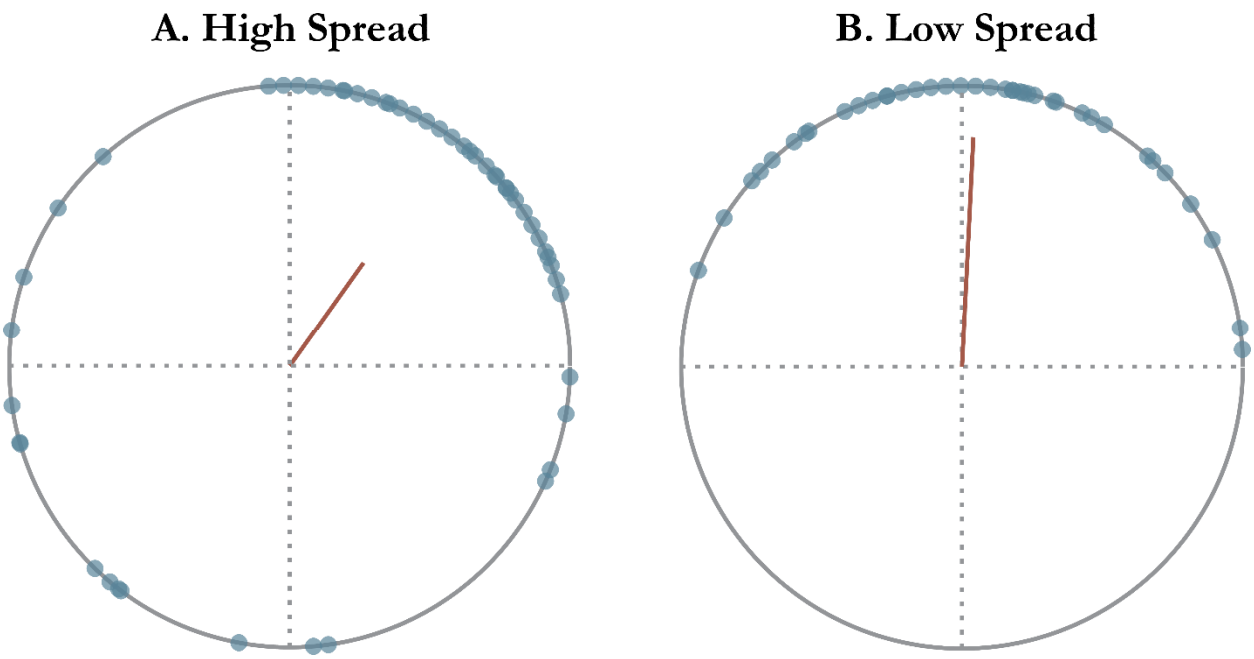

Figure 5.3: Example results using the Resultant vector length metric. A. Feature with a high circular spread of assigned preferred directions. B. Feature with a low circular spread of assigned preferred directions. Each blue dot represents the assigned preferred direction on a unit circle. The red line in each plot is mean resultant vector calculated using all the points on the unit circle. The length of this resultant vector was used to quantify how "consistent" each feature was.

Figure 5.3 shows the resultant vectors for distributions of two representative features. The feature in Figure 5.3 A. has a high circular spread and hence the resultant vector (in red) is short. On the other hand, feature in Figure 5.3 B. has a smaller circular spread and, therefore, a longer resultant vector.

5.2 Results

5.2.1 Psychophysics and Task Performance

The performance in the 2D center-out task was quantified using three different parameters: number of blocks, mean movement times and deviations from straight line trajectories. The variation of each of these parameters is shown in Figure 5.4.

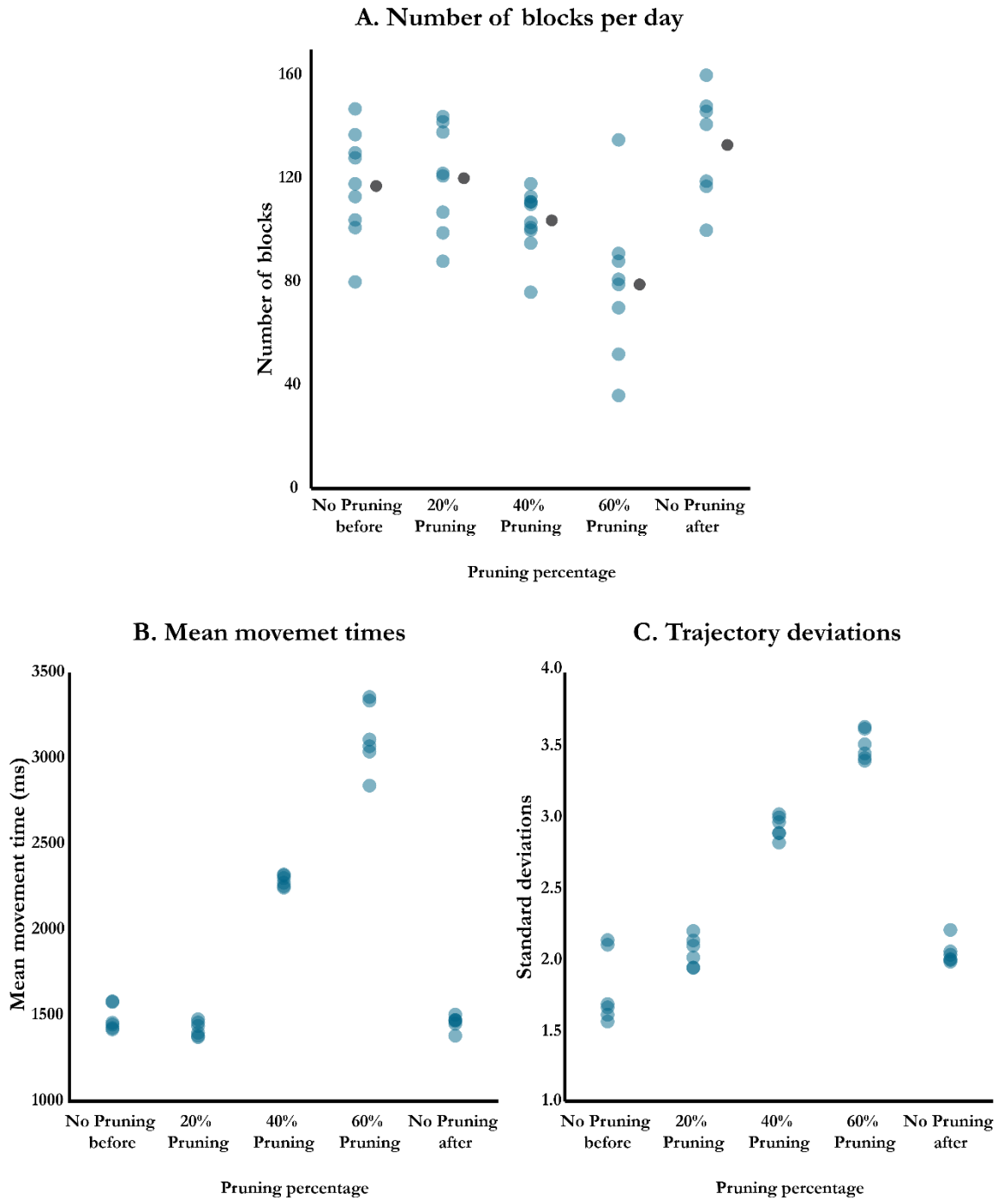

Figure 5.4: Performance during experiments with different levels of pruning. A. Number of correct trial blocks performed. Each blue point is the number of blocks from one day of experiments. The gray dot next to each type of experiment is the mean number of blocks performed during those days. A decrease in the number of blocks performed is an indicator of increased difficulty. B. Mean movement times during each day of experiment. A significant and linear increase is observed in the movement times after experiments with 20% pruning. C. Standard deviations from straight line trajectories. Similar to the movement times a significant and linear increase in observed in the trajectory deviations after 20% pruning.

The variation of the number of blocks performed per day with the pruning percentage is shown in Figure 5.4 A. As described earlier, each session was divided into blocks of 16 correct reaches. Each blue point on the plot corresponds to a single session of experiment. The gray points represent the mean number of blocks for each type of experiment. No significant difference is observed between "No pruning before" days and "20% pruning" days. However, an almost linear decrease in the number of blocks performed is seen after that. The numbers go back up again for "No pruning after" days, which used all the control features and were carried out after the 60% pruning experiments. A decrease in the number of trials performed is an indicator of increased difficulty in performing the task resulting in the subject getting frustrated and refusing to work.

Trends similar to Figure 5.4 A are observed both in 5.4 B and C. The average movement times in Figure 5.4 B are fairly constant for "No pruning before" and "20% pruning" days and linearly increase after that. They are again seen to reduce to their initial values for the "No pruning after" days. For Figure 5.4 C. each trial trajectory was divided into 40 time segments starting from center to the target. Perpendicular deviation from the straight line trajectory was calculated in each time segment for each trajectory. Each point in the plot represents the mean standard deviations of the 40 distributions created from one session of experiment. Again the values are fairly close to one another during "No pruning before" and "20% pruning" days. The deviations increase linearly from this point and go back to the lower values for "No pruning after" days. This is an interesting trend and can be explained by investigating the characteristics of control feature set in the following sections.

5.2.2 Distributions of Control Features

The distribution of control features was first studied without grouping them by frequency bands or recording electrodes. Figure 5.5 shows the mean 2D weight vectors of the features on a polar plot. Figure 5.5 A shows the distribution of the whole feature set available from the electrode array. This distribution, by the virtue of being not uniform, is interesting in itself. It is clear that a handful of features have bigger mean weight vectors and are primarily oriented in three directions. The rest of the features have smaller vector lengths and seem to be uniformly distributed in the 2D space. One way to interpret this is that the subject was using only a small set of dominant features for control. Also, the orientation of these features

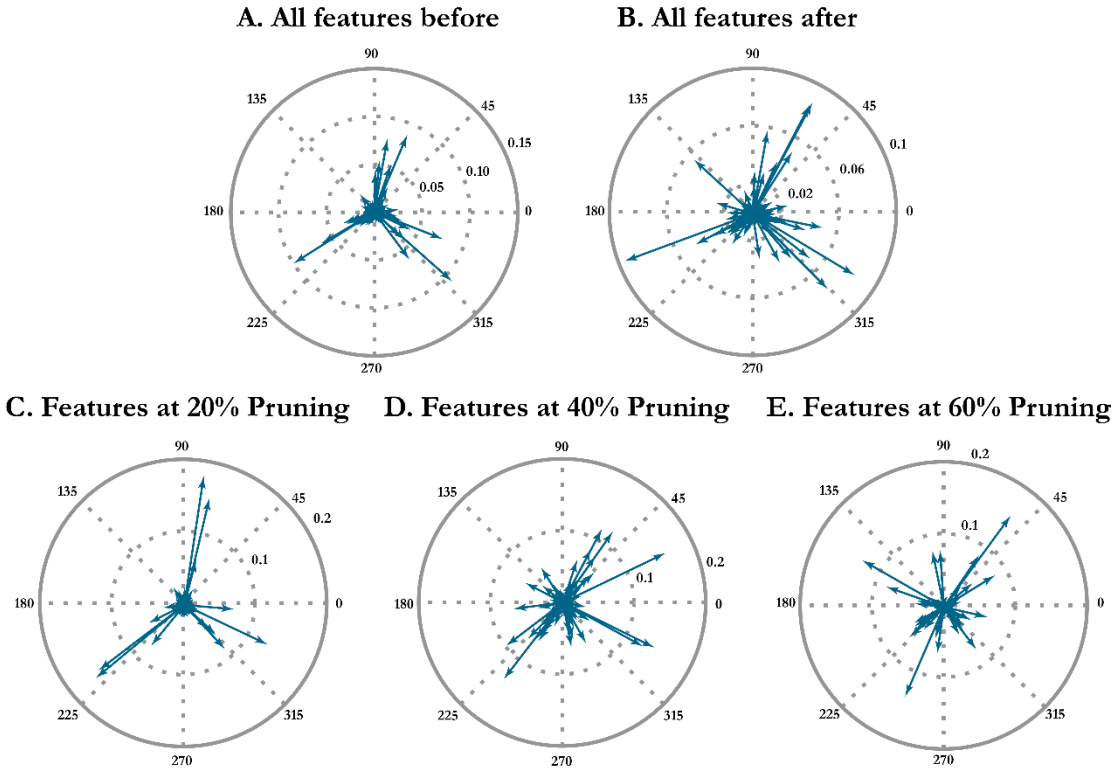

Figure 5.5: Distribution of features in 2D space. This figure shows the distribution of mean feature weights in 2D space during experiments with different levels of pruning. A. All features at the beginning of the experiments. B. All features towards the end of the experiments. C. Features at 20% Pruning D. Features at 40% Pruning E. Features at 60% Pruning. A few features are observed to be assigned larger weights during each set of experiments with most of the features assigned relatively low weights.

is observed to be in three directions, approximately 120º apart from each. From a purely mathematical point of view, a distribution like this could be easily used to reach any of the presented targets by "up-modulating" the feature in one or two of the three directions. The subject could have chosen only two orthogonal directions, instead of three, to control the movement in the 2D plane, however, this would require "down-modulating" features to reach certain targets.

Figure 5.5 C shows the distribution of features at 20% pruning and it is worth noting that it looks very similar to the one in A. As 20% of the most reliable features were eliminated going from A to C, the prominent features in A are no longer available for control in C. However, a striking amount of similarity is still observed between the two. This has two implications; a) the newly recruited "dominant features" at 20% percent pruning have very similar preferred directions to the ones prominent at no pruning (Figure 5.5 A) and it is important to investigate the relationship between the two, and b) the lack of changes in performance observed from Figure 5.4 could be a result of the similarity in between the two distributions.

As more and more features are eliminated (Figure 5.5 D and E) the distributions become random and probably more uniform in the 2D plane. Figure 5.5 B, for the days where all the recorded features were again used for control, presents one more interesting finding. The distribution here is more uniform than the one in A. The prominent directions from A are still present with additional prominent features in other directions. The subject seems to have retained some features that were recruited during the days of pruning sessions. The

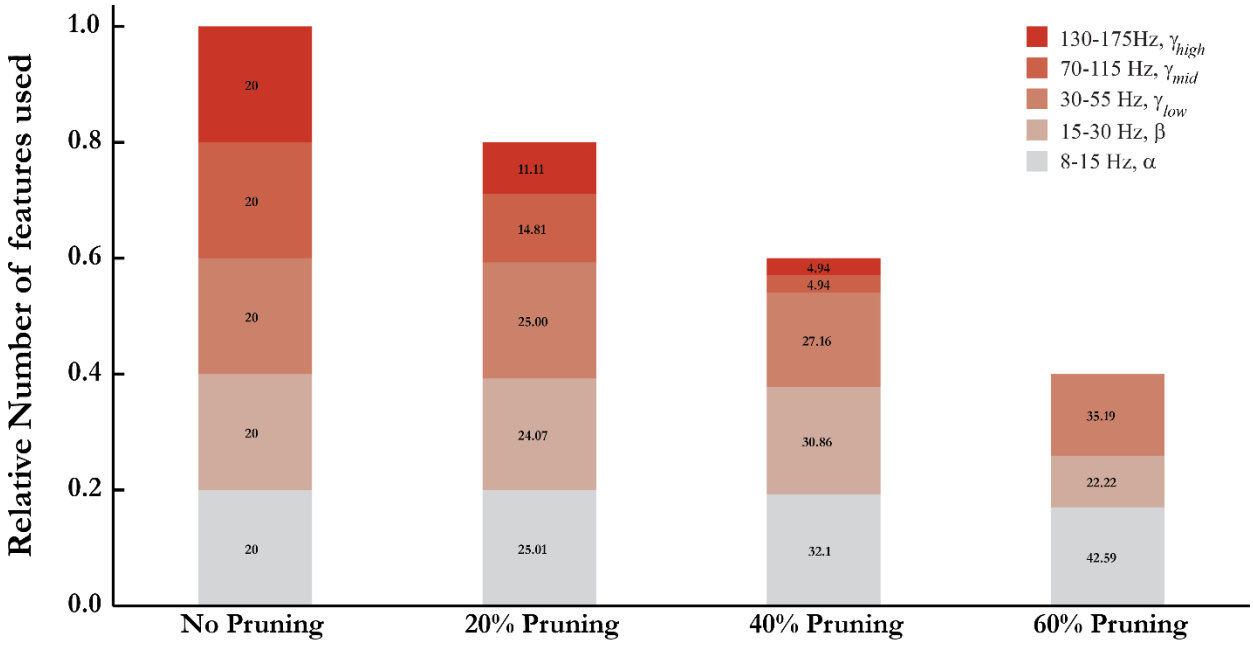

Figure 5.6: Relative number of features used from each frequency band at different percentages of pruning. Each bar gives the distribution of features used by the subject at different levels of pruning. The numbers for each band represent the percentage of times a feature from band was picked. Features in mid-gamma and high-gamma bands are quickly pruned with none of them present at 60% pruning. Beta band features are eliminated next.

directions of the newly recruited features are, however, determined by the decoder which like the optimal linear estimator (OLE) spreads the prominent features as uniformly as possible.

After looking at the overall distribution of features in 2D space, it is important to group the features by their frequency bands and recording electrodes to determine the characteristics of the consistent features that were eliminated. Figure 5.6 shows the frequency band distributions of the control features at different pruning percentages. It is clear from the figure that mid-gamma and high-gamma band features were the most consistent and were pruned by the algorithm first. Monkey I had 27 active recording electrodes and therefore 135 control features in total. At 40% percent pruning, only 8 of 81 control features were from the mid or high-gamma band. All the features above 70 Hz were absent during 60% pruning days.

Looking at the bars for 40% and 60% pruning, a reduction in the beta band features is seen. 20% (27 of 135) of the features were eliminated by the pruning algorithm between the two sets. When none of the higher gamma features were available, beta band features were found to be most consistent and that is why a significant reduction in their number was seen.

Figure 5.6 shows the changes in feature space with respect to the frequency bands. However, to study the relationship between the eliminated and newly recruited control features, recording electrode location of individual features have to be used as another grouping parameter. Figure 5.7 combines the two groups to form a contingency table of mean feature weights as a function of both frequency band and recording electrode. Each row represents the mean weight magnitudes of the features at a given pruning percentage. The mean weights are represented on a grid of electrodes arrays similar to the ones implanted in Monkey I. Each electrode is further divided into five different sectors, one for each frequency band. The mean weights are represented by the intensity of the colors in each sector. Higher color intensity represents bigger relative values (arbitrary units). The grayed out sectors represent the features that were eliminated and not used in the control feature set.

As seen from Figure 5.6, Figure 5.7 also shows that the features from mid-gamma and high-gamma band were eliminated first. Looking at the intensity of the mean weights, it is also evident that the features with larger weights were also the most consistent and were eliminated first. This reinforces the interpretation made from Figure 5.5 A and C that, the prominent features in the 2D distribution were eliminated first. Apart from these two observations,

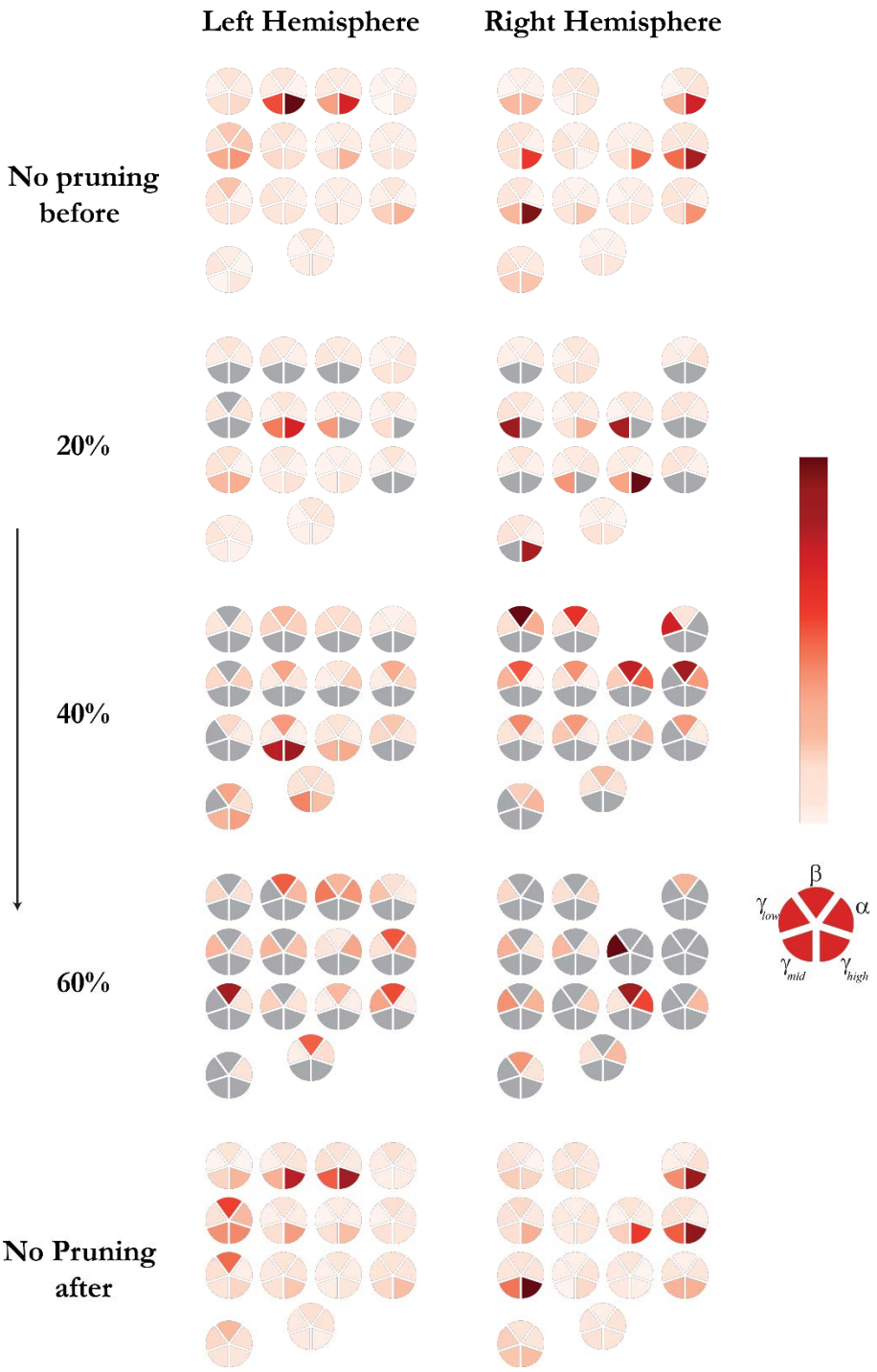

Figure 5.7: Distribution of features used during different levels of pruning. Each row is a representation of each feature used during a given set of experiments, conducted chronologically from top to bottom. Each electrode on the arrays is divided into 5 sectors representing the five bands starting with alpha at the top right and going anti-clockwise to beta, low-gamma, mid-gamma, and high-gamma. The grayed out sectors represent the features that were pruned out and the intensity of the color in each sector represents the mean magnitude (AU) of the weight assigned to it.

Figure 5.7 also helps with understanding the relationship between the eliminated features and the features that were newly recruited. As it can be seen, going from one level of pruning to next (one row to next), features that are closer to the newly eliminated features were assigned higher weights. The new features were either closer with respect to the recording electrode or the frequency band of the eliminated feature.

The mean weights of features that were used to make Figure 5.7, were further analyzed to quantify the relationship between the newly recruited and eliminated features. The first step in this process was to mark a feature as either a newly recruited feature or, a feature that didn't change much going from one pruning level to the next. For this, a distribution was created using the changes in weights of all control features at a given pruning percentage from the previous pruning percentage, for example, "20% pruning" from "No pruning before". Features were termed as new recruits if their difference in weight was more than one standard deviation away from the median of the distribution. After all the new features were determined, the first step was to quantify their spatial distance from the eliminated features. This is depicted through Figure 5.8.

The figure shows the distance between the recording electrodes of the new features and the recording electrodes of 20% features that were eliminated going from one pruning percentage to next. Each point in the plot represents a newly recruited control feature. The *x*-axis of the plot shows the three transitions; "All features-20% Pruning", "20% Pruning-40% pruning", and "40% Pruning-60% Pruning". The *y*-axis is the distance to the closest electrode in which a feature was eliminated. A distance of 0 means that the new feature was on the same

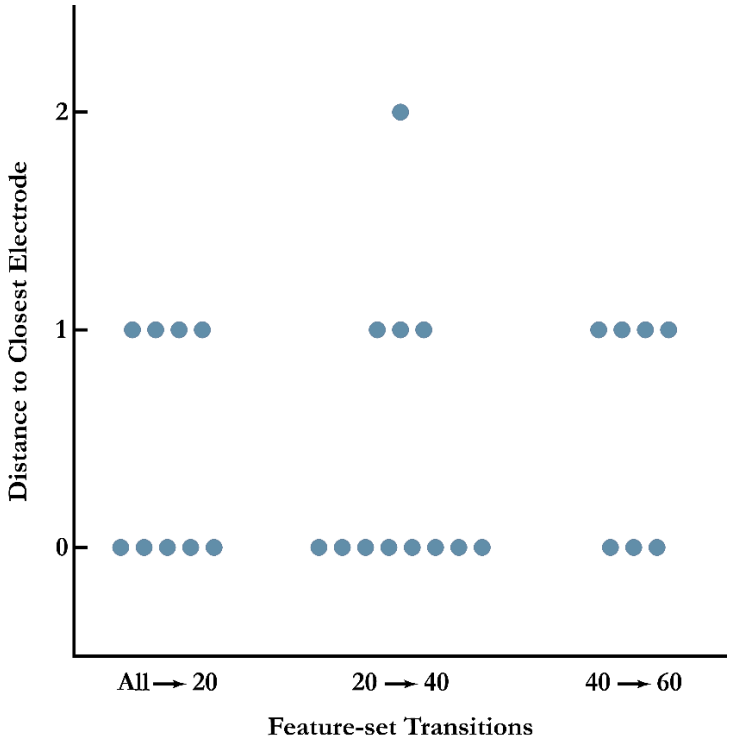

Figure 5.8: Spatial distance of new features from the closest eliminated feature. Each point on the plot represents a feature that was newly recruited for control going from one pruning percentage to next. The *x*-axis represents the three transitions. The *y*-axis is the spatial distance of the new feature from the closest feature that was eliminated after the previous pruning percentage sessions.

electrode as one of the eliminated features. It can be seen from the plot that most of the eliminated features are either on the same electrode as one of the eliminated features or on adjacent electrodes.

Going one step further, for the features that were recruited from the same electrode as one of the eliminated features, Figure 5.9 shows the spectral distance between these features. Each point depicts a newly recruited feature. The *x*-axis is the same as in Figure 5.8. Frequency bands were assigned values 1 to 5 going in order alpha (1), beta (2), low-gamma(3), mid-gamma (4), and high-gamma (5). The *y*-axis is the absolute difference between the frequency bands of newly recruited and eliminated features. For instance, if a mid-gamma feature was eliminated

and a high-gamma feature was recruited from the same electrode, the difference would be 1. It is seen from the figure that all the newly recruited features in the first column were spectrally adjacent to the eliminated features. In the second column, going from 20% pruning to 40% pruning, where most of the features above 70 Hz were eliminated, most of the newly recruited features were 2 units away. That is, a lot of beta band or low-gamma band features were recruited due to the elimination of mid-gamma and high-gamma band features. The last column has very few features as more features were recruited from adjacent electrodes.

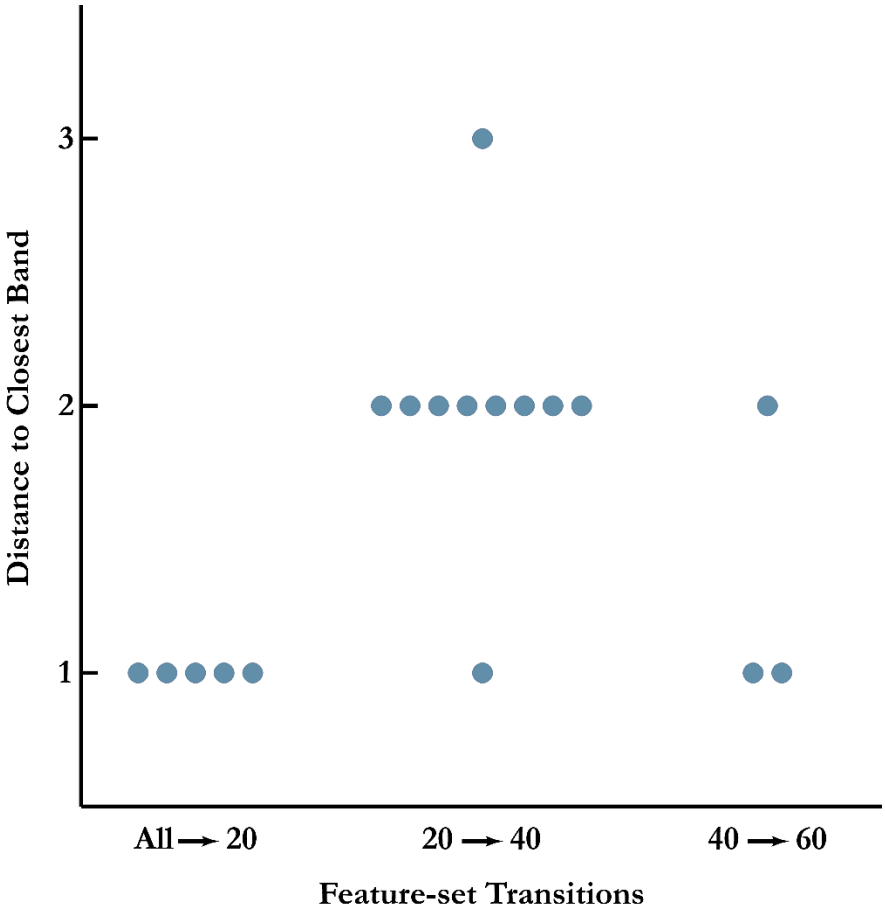

Figure 5.9: Spectral distance of new features from the features eliminated on the same electrode. Each point on the plot represents a feature that was newly recruited for control going from one pruning percentage to next. The *x*-axis represents the three transitions. The *y*-axis is the distance between the frequency bands of the new and eliminated features.

5.2.3 Changes in Feature Modulations

In the experiments conducted for this chapter, features were completely eliminated from the control set as the pruning percentage was increased. This provided a unique opportunity to study the changes in feature modulation when they were part of a control set and when they were not. Also, due to the elimination of reliable, consistent features, the subject was required to use previously noisy and unreliable features to accomplish the task. This could only be accomplished if the subject was able to change the modulations of noisy features to make up for the lost ones. Though the results from Figure 5.4 clearly show a degradation in task performance, even with the "worst" 40% of the recorded features, the subject was able to perform hundreds of correct reaches to each of the eight targets with an average movement time close to 3 seconds. Figures in Section 5.2.2 also show higher mean weight magnitudes assigned to features that were previously quite low. Though this could be accounted for by the decoding algorithm trying to find the best fit with the available feature space, it could also result from the changes in feature modulations with the changes in the control set. With these observations in mind, this section studies the modulation of recorded features and their changes with the pruning percentage.

The first step in the process of studying the changes in feature modulations with changes in control feature set is to quantify the relative modulations of each feature during movement of the cursor compared to the baseline activity. Instead of using the baseline activity that was recorded before starting an experimental session, the activity between the movement periods of consecutive trials was used as the baseline. This provided a more robust measure

of in-task modulation as it accounted for any changes in the overall activity during the multi-hour session.

It has been described previously that each trial began with a 500 ms hold period at the center of the cursor. Prior to the initial hold period, there was also a setup period which took a few milliseconds and was used to select a target for the upcoming trial. Also, at the end of each trial a few milliseconds of time was spent in marking a trial as correct or incorrect and appropriately rewarding the subject. So, the activity starting from the end of movement period of the previous trial to the beginning of movement period of current trial was used as the baseline for that trial. An root mean square (RMS) value for each feature was computed during this period and termed the rest RMS, $RMS_{rest,f}$. Another RMS value was computed during the movement period of the trial and was termed as the movement RMS, $RMS_{mov,f}$. The percentage change from baseline for each correct trial was given by:

$$PC_f = 100 * (RMS_{mov,f} - RMS_{rest,f})/RMS_{rest,f} \quad \textbf{(5.1)}$$

Distributions were created for each feature's percent change from baseline and were grouped by the target direction. As the targets were equally spaced at 45° on a circle, each feature distribution was fit with a cosine function. Data from one session of experiments were used to find the cosine fit. Cosine fits were determined for every recorded feature, for all sessions (*i.e.* days). Figure 5.10 shows the distributions and cosine fit for two representative features with high and low R^2 values of the fit.

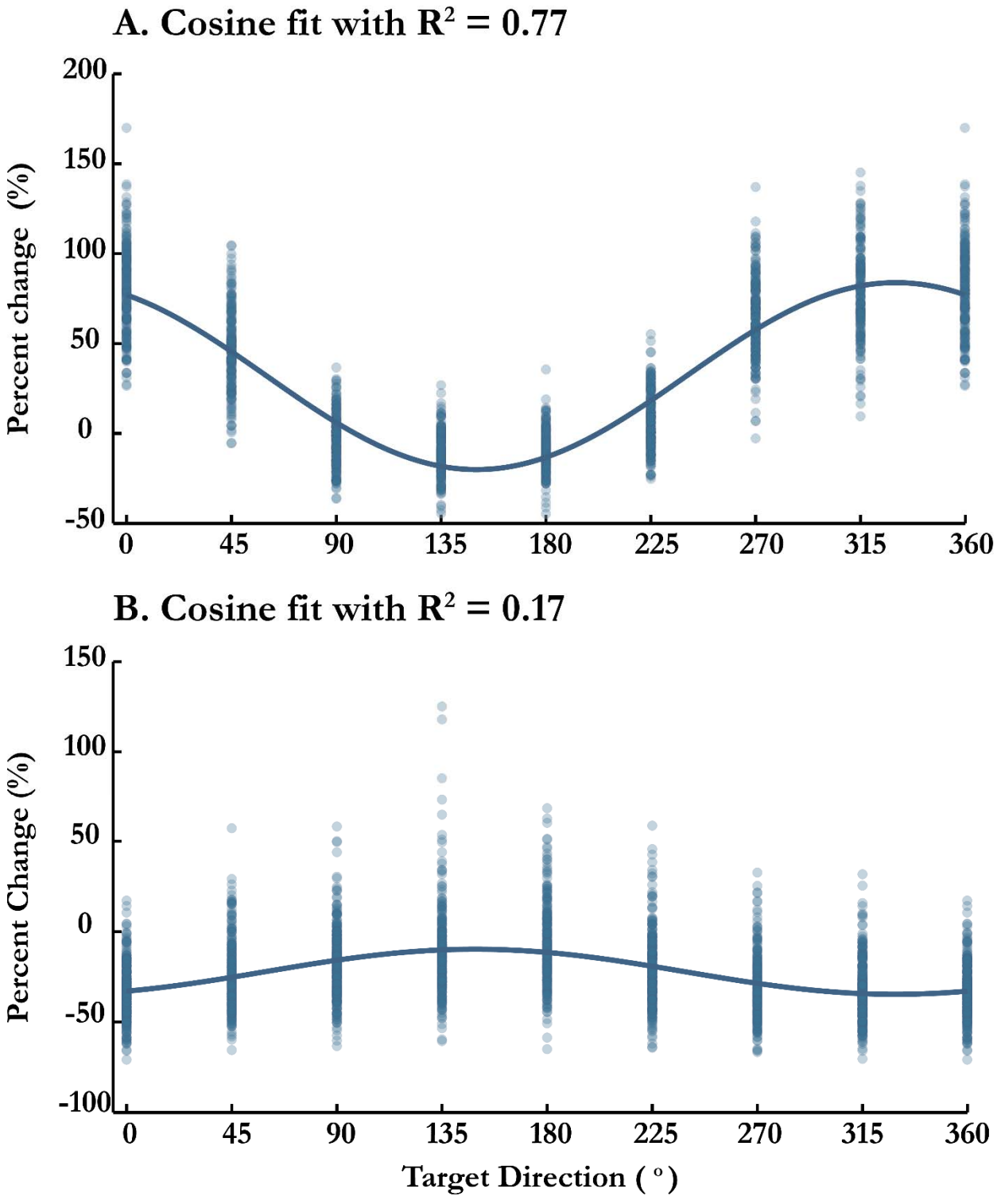

Figure 5.10: Cosine fits to the percent change in RMS from baseline. A. Reliable feature with low noise and a high R^2 value for the cosine fit. B. Noisy feature with a low R^2 value for the cosine fit.

After the cosine fit for a session of experiment was determined for each feature, a depth of modulation (DM) metric was assigned to it. The depth of modulation was simply equal to peak to peak value of the cosine fit. Features from each frequency band, on all days of experiments, were grouped together. These groups were further divided by the pruning percentage. Figure 5.11 A shows the changes in mean DM of different frequency bands as the pruning percentage was varied. All the recorded features, irrespective of them being a part of the control set or not, were used to plot this figure. As a quick reminder of the observations from previous results, features above 70 Hz were eliminated first followed by the beta band. As the pruning percentage was increased, features below the 55 Hz frequency range dominated the control feature set. As can be seen from Figure 5.11 A, an overall decrease in modulation is observed in both mid-gamma and high-gamma frequency bands. The DM in these bands goes back to its higher value when the whole recorded feature set is again used for control at the end. A small, but significant, increase in the modulation of beta and low-gamma band features is observed with the increase in the pruning percentage. This indicates that the subject was trying to accommodate for the loss of reliable features.

The changes in DM become clearer when the features are separated into control (used) feature sets (Figure 5.11 B) and pruned feature sets (Figure 5.11 C). As the features eliminated by the pruning algorithm were reliable and consistent, a decrease in mean DM of used features is expected. This is seen for mid-gamma and high-gamma band features in Figure 5.11 B. However, the more important observation from this figure is the increase in mean DM of the beta band and low-gamma band with the increase in pruning percentage. As was inferred from

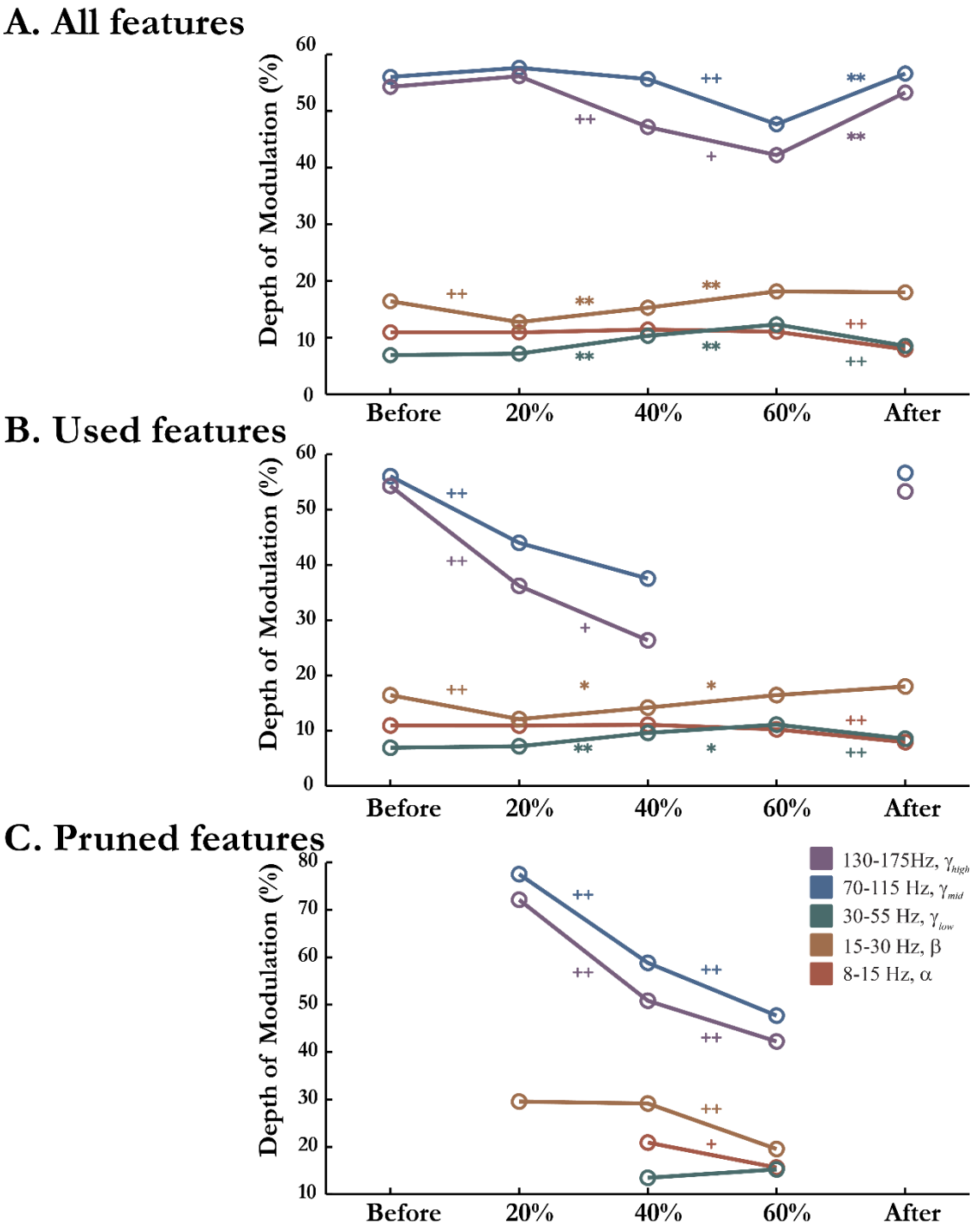

Figure 5.11: Variation of depth of modulation of different frequency bands. A. Depth of modulation (DM) of all the features from five frequency bands. B. DM of the features that were used during the specified task C. DM of the features that were pruned out. No data was available at the places where values are missing. + indicates a significant decrease in DM (+ $p<0.05$, ++ $p<0.01$) and * indicates the significant increase in the DM, (* $p<0.05$, ** $p<0.01$) from left to right. A decrease in the DM is expected with elimination of features with highest DM. An increase or no significant change in DM is an indicator of intrinsic changes employed by the subject to accomplish the BCI task with features that previously had a lower DM.

the Figure 5.11 A, this increase in DM is an indicator of changes in modulation of control feature set, employed to compensate for the lost features. It is also interesting to see that the mean DM values of all of the frequency bands almost return back to the initial state. This shows that the subject quickly realized that the whole feature space was available for control, and went back to using the most reliable features from the set. Figure 5.11 C just shows a decrease in DM for the pruned features, as expected.

Like in Figure 5.10, an R^2 value of the cosine fit was calculated for each feature, for every session of experiment. This value is an indicator of how noisy a given feature is; higher R^2 values indicate less noise in percent change of RMS in a given direction whereas, lower R^2 values indicate that the percent change of RMS in a given direction was relative random. This is seen in the two representative features of Figure 5.10 as well. Plots in Figure 5.11 clearly showed significant changes in the mean DM values with changes in pruning percentage. As pointed out before, DM values were calculated from the cosine fits. However, there is no indicator of how reliable the cosine fits used for the figure were, in other words, what were the R^2 values for the fits and did these change as well? Figure 5.12 answers this question. The three parts of the figure A, B, and C are exactly the same except they use the R^2 values to calculate and plot the means instead of DM value. Interestingly patterns very similar to Figure 5.11 emerge with R^2 values as well. That is, the beta band and low-gamma band features used for control at 40% and 60% pruning didn't only increase their mean DM, but also became more cosine tuned. Similarly, for features above 70 Hz in conjunction with the decrease in DM, a decrease in cosine tuning fit is also observed. It is to be noted that, even though the

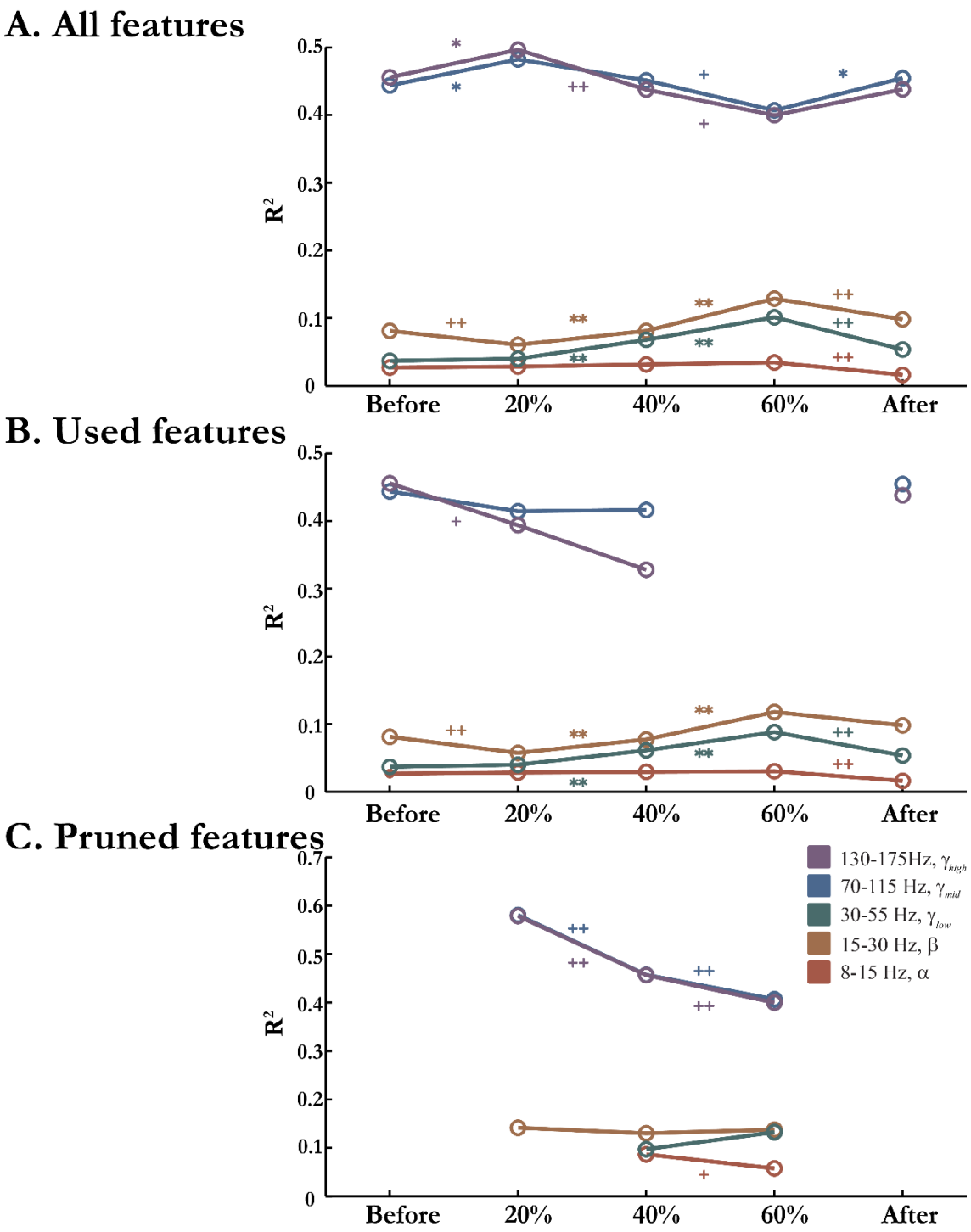

Figure 5.12: Variation of cosine fit R^2 of different frequency bands. A. R2 of all the features from five frequency bands. B. R^2 of the features that were used during the specified task C. R^2 of the features that were pruned out. No data was available at the places where values are missing. + indicates a significant decrease in R^2 (+ $p<0.05$, ++ $p<0.01$) and * indicates the significant increase in the R^2, (* $p<0.05$, ** $p<0.01$) from left to right. An increase in the R^2 values indicates a decrease in the randomness of modulation in any particular direction and vice-versa.

changes in Figure 5.12 are small, they are significant. Also, it is important to note that even with an increase in some DM values, the cosine fits in features below 55 Hz are still poor.

5.3 Conclusions

The experiments described in this chapter were designed to get a better understanding of the optimal ECoG frequency bands used for BCI control. By eliminating features with robust tuning characteristics, the experiments also intended to the train the subjects to use the features that did not show significant tuning. Much like a physician putting an eye patch over a patient's good eye to force the lazy eye to improve, we wanted to determine whether we could train weaker neural features to become more robust and whether that training persisted after the pruned features were returned to the control pool (i.e. the eye patch was removed). Promising results were obtained from these experiments and their implications have been outlined in this section.

The first thing that becomes clear from the frequency distribution (Figure 5.6) is that frequencies above 70 Hz show robust tuning characteristics. When this result is combined with the performance changes during the experiment another important conclusion can be made. As has been described earlier, practically no degradation of performance is observed going from all features to 20% percent pruning. Therefore, it can be inferred that as long as a few significant features from mid-gamma and high-gamma band are present, subjects would be able to obtain a robust BCI control. Similar results were seen from Chapter 4 pruning experiment as well. The trajectory deviations were fairly constant throughout the different pruning percentages due to the presence of gamma band features in the control set. The only

degradation in performance was with respect to the movement times. As has been explained before, this was likely due to the constant gain term and a reduced number of features. The reason an increase in movement time was not observed going from all features to 20% pruning in this chapter was how the decoding model was created. Even though the gain term was kept constant and the feature space was reduced, the decoding model was created after the elimination of features. On the other hand, in the previous chapter, the features were eliminated after updating the decoding model. Due to this difference, the absolute values of the weights in the two cases were different and thus resulted in longer movement times in the former experiments and no changes in the current experiment. This argument is, however, valid only when significant features from the higher frequencies are present in the control set.

Going beyond the frequency distributions of the control set, the results from the relative spatial and spectral locations of the features show how new features were recruited to accomplish the BCI task. From a pure neurophysiological point of view, ECoG electrodes record electrical field potentials produced by synchronized activity of an ensemble of neurons in the cortex. There is a lot of debate on which type of activity produces what type of signal and what is the spatial resolution of ECoG. However, it can be safely assumed that ECoG activity on adjacent electrodes is more correlated than activity from electrodes further away from each other. Also, the ensemble of neurons closer to the recording site would very likely affect the whole spectrum of signals recorded from that site. With these two rationales, the spatial and spectral distances of the newly recruited features (Figure 5.8 and 5.9) seem almost trivial. The subject chose to recruit the features that were the closest to eliminated features, in

both spatial and spectral dimensions, because, given the underlying neurophysiology, it was the easiest and most efficient approach.

The experiments also produced a subtle effect that was not the initially expected. To be able to control the BCI, the subject had to increase the modulation of features that were previously quiet or just random. Some of the mid-gamma, high-gamma and beta band features that were initially assigned low weights were recruited in the process of the experiments. The interesting by-product of going through the pruning process was that some of the newly recruited features were retained even after the whole feature space was again made available for control. This can be seen by comparing the 2D distributions in Figure 5.5 A and B and the first and last rows of Figure 5.7. This result though not very significant highlights the importance of closed-loop feedback in the field of BCI. It shows that forcing subjects to use unreliable features (i.e. putting on an eye patch) in a closed-loop BCI experiment can help train these unreliable features to become more robust and accurate, which persists even after the originally dominant features are returned to the control pool. . Thus, this method could be used to train up all recorded features so that they contribute equally to the overall control algorithm. This will ensure overall robustness in BCI control as a loss of any given feature will have minimal effect on the overall control since the remaining features would provide reliable information.

Chapter 6
Regularized Least Square Feature Selection in BCI

The previous two chapters used feature selection algorithms in conjunction with a co-adaptive decoder. This chapter describes a regularized least square regression based decoding algorithm that produces a sparse control feature set. The algorithm was designed with the intention of inducing sparsity in the control feature set and obtain 2D BCI control with them. Like the previously used decoding algorithms (Chapters 4 & 5) this algorithm uses data collected from a few trials to iteratively update the decoding model through the duration of the experiment. The parameter controlling the sparsity of the control feature set was varied over weeks to gradually reduce the number of control features. After a few weeks of recording and obtaining a sparse decoding model, experiments were conducted with the fixed model for two weeks, to study the performance changes. At the end, one week of experiments were conducted using a random set of features of size equal to the obtained sparse set. A simple least square regression-based decoder was used during this week.

This chapter is organized in three sections. Section 6.1 describes the general methodology including the subjects, experimental protocols, and the decoding algorithm implementation. Section 6.2 compiles the results from the experiments describing the performance of the decoder and the characteristics of the control set. It also compares the

performances using a continually updating decoder with a fixed decoder and a decoder using a random set of features. Finally, Section 6.3 concludes the chapter and discusses the significance of the observations from the experiments in this chapter.

6.1 Methods

6.1.1 Subjects

Two male Rhesus Macaques (*Macaca mulatta*, monkey I and R) weighing 6-10 kilos were used for the experiments in this chapter. Both the monkeys had bilateral implants on the left and right hemispheres of their brain. The experiments described here were carried out two years after the implantation surgery. Monkey I had previously been trained on multi-dimensional (2 or more) BCI tasks for more than two years. Monkey R, on the other hand was relatively naïve to 2D BCI. He had been trained in 2D center-out task using a joystick and 1D BCI task for approximately a year before starting these experiments. He was first exposed to 2D BCI only for 2-3 weeks before collecting the data used in this chapter. Neither monkey (I or R) had previously performed a BCI task using the decoding algorithm specifically designed for this study.

6.1.2 ECoG Arrays: Design and Placement

Monkey I was implanted with two Silastic electrode arrays manufactured by Ad-tech Medical (Left hemisphere) and PMT Corporation (Right hemisphere). Monkey R was implanted with thin-film polyimide arrays designed in collaboration with Dr. Justin Williams'

Lab in University of Wisconsin, Madison. The details of the array design and the electrical connectors are described in Section 3.2 of Chapter 3.

All the arrays in the two monkeys were implanted in a rostral-caudal direction. The 19mm craniotomy in monkey I was centered at the Stereotactic coordinates (Anterior 6 mm, Lateral 14 mm) on both the hemispheres. This placed the arrays partially on the motor and sensory cortices. In monkey R, the craniotomy was centered at the coordinates (Anterior 15 mm, Lateral 15 mm) on both hemispheres. The arrays, in this case, spanned from premotor cortex to S1. The approximate locations of all the arrays with the different electrode sizes are shown in Figure 6.1.

6.1.3 Behavioral Task and Experimental Protocol

A classic eight target 2D center-out task was performed by the subjects. The targets were equally spaced at 45° on the circumference of a circle. The radius of the circle was set at

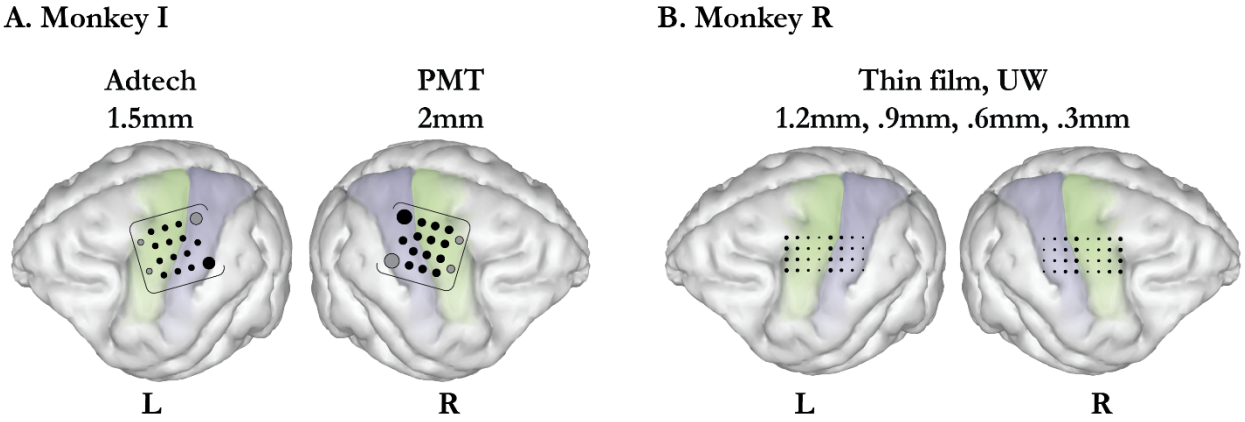

Figure 6.1: Electrode Design and placement. A. Left and right hemisphere in monkey I were implanted with one array each with 1.5mm diameter electrodes (L, Ad-tech) and 2mm diameter electrodes (R, PMT). Both the arrays covered a part of motor and sensory cortices as shown in the picture. B. A polyamide thin film ECoG electrode array was implanted in each hemisphere on monkey R. Each array consisted of 32 channels of four different sizes (Diameters 300µm, 600µm, 900µm, and 1200µm) of exposed electrode surface. Both the arrays covered a part of motor and sensory cortices as shown in the picture.

20 screen units for monkey I and 15 screen units for monkey R. Both the target and cursor were spheres of radius 1.6 screen units for monkey I and 2.5 screen units for monkey R. Each trial began with the cursor at the center of the circle. At the end of a 500ms initial hold period, one of the targets was presented. The subjects had to move the cursor and touch the presented target in a set amount of maximum movement time. If they were successful in doing so, the trial was marked correct and a small water reward was awarded. In any other case, the trial was marked incorrect. The differences in cursor and circle dimensions between the two subjects were to make the task relatively easier for the naïve monkey R.

A constant speed assistive bias, same as Chapter 4 and 5 experiments, was also used. Each session of recording started with a naïve decoder and a watch task block. The movement of the cursor in this block was completely controlled by the assistive bias. As described in Section 4.1.4, the gain term associated with the bias was gradually reduced and the control was completely transferred to the subject in a few blocks.

Similar to the previously described experiments, each recording session was divided into blocks of trials. In this case, each block consisted of 32 correct reaches (4 to each target). Targets were presented in a pseudorandom order during a block and all the incorrect trials were ignored. The data collected during a block were used to update the decoding model used in the following block. Updates to the decoding model were carried out through the end of the experimental session. No memory of the model was carried over from the previous session and the initial model was constructed from the data recorded during the watch task block.

During the days where a single fixed model was used, no updates were carried out and the fixed model was loaded into the system at the beginning of each session. The assistive bias was still used during a few initial blocks of these sessions to help the subject to start.

6.1.4 Control Signal

Subjects controlled the velocity of the cursor in center-out BCI task. Similar to previous studies the raw ECoG signals were divided into five frequency bands; alpha, beta, low-gamma, mid-gamma, and high-gamma. Signals from each frequency band/electrode combination went through a few steps of processing before being used as a "control feature" (details in Section 3.4). A weighted linear sum of the ECoG features in combination with the constant speed assistive bias was mapped to the velocity of the cursor. This velocity signal was integrated using Euler Integration to give the position of the cursor on the screen. The two-dimensional velocity V of the cursor at a time t is given by:

$$V(t) = E(t) + B(t) \tag{6.1}$$

Where $E(t)$ and $B(t)$ represent the ECoG control signal and the constant speed bias respectively and are given by:

$$E(t) = G_f \mathbf{W}^{\mathrm{T}} F(t) \tag{6.2}$$

$$B(t) = G_b U(t) \tag{6.3}$$

$\mathbf{W}$ is the weight matrix (Nx2, N = number of features) produced by the decoding algorithm to be used during the current block, $F(t)$ is a vector of ECoG control features (length N) at time t, and $U(t)$ is the two-dimensional unit direction vector from the cursor to the current target at time t. The superscript T denotes a matrix transpose. Terms G_f and G_b represent the gain applied on the ECoG control signal and the bias signal, respectively.

For the watch task block, weight matrix $\mathbf{W}$ was initialized to all zeros and the bias gain term G_b was set to a 100. Thus, the cursor was controlled only by the constant speed assistive bias. After the initial model was constructed at the end of this block, the cursor was controlled both by the subject and the assistive bias. Similar to the previous experiments the bias gain was gradually reduced to 0 in 5-10 blocks and the control was completely transferred to the monkey. Unlike the previous experiments the ECoG feature gain term G_f was dependent on the weight matrix. This relation is given by:

$$G_f = \frac{G_d}{\sum_j^N \sqrt{\sum_k^2 \left(w_{j,k}\right)^2}} \qquad \textbf{(6.4)}$$

Where G_d is the desired gain and $w_{j,k}$ denotes the (j,k) entry of $\mathbf{W}$, the weight matrix (Nx2, N = number of features). This dependence of gain on the sum of feature weight magnitudes was used with the intentions of keeping the overall gain constant as the number of features were varied by the decoding algorithm. Desired gain G_d was initialized to a low value at the beginning of the session and gradually increased to a maximum value by the time

the control was transferred to the subject. This maximum value was kept the same across weeks of experiment as the regularization parameter was varied.

6.1.5 Decoding Algorithm and Implementation

The decoding algorithm was designed with the goal of producing a sparse control model for 2D BCI. To achieve this, a squared loss function was chosen. The introduction of the least absolute shrinkage and selection operator, or *lasso*, a technique popularized by Tibshirani *et. al* [66], was intended for producing a sparse solution. Also known as basis pursuit [67] in the signal processing literature, this technique constrains the ℓ_1 norm of the weight vector to set some coefficients exactly equal to 0. The decoding algorithm for a 2D BCI task needs to produce two weight vectors, one for each dimension. The *lasso* technique could have been used to produce a separate sparse weight vector for each dimension. However, this would not have resulted in same features being selected for each dimension, defeating the goal of selecting a single subset of features for the task. An extended version of the *lasso* technique, proposed by Obozinski *et. al.* [68], [69], was employed to obtain the desired result. This method solves the problem by penalizing the ℓ_1 norm of the vector formed by the ℓ_2 norms of the feature weights in each dimension (ℓ_1/ℓ_2-norm). By doing so, it produces a single sparse feature set to predict both the spatial dimensions.

The input to the decoding algorithm was, as in the previous decoder, unit vector in the desired movement direction, denoted by $\widehat{\mathbf{V}}$, and the modulation of ECoG features given by $\mathbf{F}$. The squared loss function combined with the ℓ_1/ℓ_2-norm constraint is given by:

$$\mathcal{L} = \left\|\mathbf{FW} - \widehat{\mathbf{V}}\right\|_2^2 + \lambda\left[\sum_j^N \sqrt{\sum_k^2 \left(w_{j,k}\right)^2}\right] \quad \textbf{(6.5)}$$

Where $\mathbf{W}$ is the desired weight matrix (Nx2, N = number of features), λ is the Lagrange multiplier and $w_{j,k}$ denotes the (j,k) entry of $\mathbf{W}$. The desired decoding model was obtained by minimizing the above loss function. To minimize the function with respect to $\mathbf{W}$ the non-differentiability of the square root had to be overcome. The lemma below was used for the same [70], [71].

Lemma 1. Given any $g(x) > 0$, the following hold:

$$\sqrt{g(x)} = \min_{z>0} \frac{1}{2}\left[\frac{g(x)}{z} + z\right] \quad \textbf{(6.6)}$$

That is, $\sqrt{g(x)}$ minimizes the right hand side of expression. This can be easily proved by taking a partial derivative with respect to z. Thus, using this lemma the square root term in equation 6.5 can be replaced with the following:

$$\sum_j^N \sqrt{\sum_k^2 \left(w_{j,k}\right)^2} \longrightarrow \sum_j^N \frac{1}{2}\left[\frac{\sum_k^2 \left(w_{j,k}\right)^2}{z_j} + z_j\right] \quad \textbf{(6.7)}$$

With this substitution the optimal solution for $\mathbf{W}$ was found by solving:

$$\min_{\mathbf{W}} \left\|\mathbf{FW} - \widehat{\mathbf{V}}\right\|_2^2 + \frac{\lambda}{2}\sum_j^N \left[\frac{\sum_k^2 \left(w_{j,k}\right)^2}{z_j} + z_j\right] \quad \textbf{(6.8)}$$

Note that an auxiliary variable z has been introduced in the expression. For now, this variable was assumed to be constant and the partial derivative of the above expression with respect to $\mathbf{W}$ was set to zero.

$$\frac{\partial}{\partial \mathbf{W}}\left(\left\| \mathbf{FW} - \widehat{\mathbf{V}} \right\|_2^2 + \frac{\lambda}{2}\sum_j^N \left[\frac{\sum_k^2 \left(w_{j,k}\right)^2}{z_j} + z_j \right] \right) \Longrightarrow 0 \tag{6.9}$$

Converting the above equation to matrix form and solving for $\mathbf{W}$ we get the following:

$$2\mathbf{F}^{\mathrm{T}}\left(\mathbf{FW} - \widehat{\mathbf{V}}\right) + \lambda \mathbf{ZW} = 0 \tag{6.10}$$

$$2\mathbf{F}^{\mathrm{T}}\mathbf{FW} + \lambda \mathbf{ZW} = \ 2\mathbf{F}^{\mathrm{T}}\widehat{\mathbf{V}} \tag{6.11}$$

$$\mathbf{W} = (\mathbf{F}^{\mathrm{T}}\mathbf{F} + \frac{\lambda}{2}\mathbf{ZW})^{-1}\mathbf{F}^{\mathrm{T}}\widehat{\mathbf{V}} \tag{6.12}$$

Where $\mathbf{Z}$ is a diagonal matrix of terms $\frac{1}{z_j}$. To minimize the loss function $\boldsymbol{\mathcal{L}}$, $\mathbf{W}$ was solved using equation 6.12 assuming $\mathbf{Z}$ constant, followed by computing $\mathbf{Z}$ while keeping $\mathbf{W}$ fixed. From equations 6.6 and 6.7, each diagonal term of $\mathbf{Z}$ is given by:

$$\mathbf{Z}_{\boldsymbol{jj}} = \frac{1}{\sqrt{\sum_k^2 \left(w_{j,k}\right)^2}} \tag{6.13}$$

Thus, by alternating between the two steps of solving $\mathbf{W}$ and $\mathbf{Z}$ the decoding model was derived. Because Lemma 1 is not defined at 0, a threshold of 10^{-3} was set on $\mathbf{W}$ and any value below this was set to 0. The alternating calculations were carried out until all the

following four convergence criteria were achieved or a maximum of 500 iterations (calculation of $\mathbf{W}$ followed by the calculation of $\mathbf{Z}$) reached:

1. Sum of new weights was smaller than previous weights:

$$\sum_{j}^{N}\sum_{k}^{2}(w_{j,k} > 10^{-3}) \leq \sum_{j}^{N}\sum_{k}^{2}(prev\_w_{j,k} > 10^{-3}) \quad \textbf{(6.14)}$$

2. Sum of weights assigned in the current iteration was less than five thousandth of sum of all weights:

$$\left\{\sum_{j}^{N}\sum_{k}^{2}\left[(w_{j,k} > 10^{-3})\ != \ (prev\_w_{j,k} > 10^{-3})\ \right] / \sum_{j}^{N}\sum_{k}^{2}(w_{j,k} > 10^{-3})\right\} < \frac{1}{5000} \quad \textbf{(6.15)}$$

3. Current value of the loss function was less than its previous value:

$$\boldsymbol{\mathcal{L}} \leq prev\_\boldsymbol{\mathcal{L}} \quad \textbf{(6.16)}$$

4. Relative change in the loss function was less than 10^{-5}

$$prev\_\boldsymbol{\mathcal{L}} - \boldsymbol{\mathcal{L}}/\boldsymbol{\mathcal{L}} < 10^{-5} \quad \textbf{(6.17)}$$

The algorithm was implemented in MATLAB and a weight matrix was calculated between blocks of experiments. Due to the sparse nature of the decoded model, a smoothing filter, like the one used in previous experiments (Chapter 4, Equation 4.5) could not be used. If the weights were blended across updates, features that were selected during certain updates and not others, would be a part of control set and the sparsity of the decoder would be lost.

The drawback of not using any memory from previous updates was that the decoded model only depended on one block of trials. For this reason four reaches to each target (32 total) instead of previously used two, (16 total) were put together into one block. This gave more data for the algorithm to work with and find a model for control.

The Lagrange multiplier λ controlled the sparsity of the decoded model. Therefore, the value of lambda was gradually increased every 1-2 weeks to reduce the number of features in the control set. This gradual change was expected to help the subjects adapt and learn to control the BCI with a smaller subset of recorded features. Post hoc-analysis was conducted to study the changes in performance and the distribution of selected features.

6.2 Results

6.2.1 Feature Selection

Experiments were conducted by varying the value of λ every 1-2 weeks. Figure 6.2 shows the changes in the number of selected features for both the subjects. Each point on the figure represents the average number of features selected during a session of experiment. Monkey I started with 135 features and Monkey R with 240. As expected with the increase in the value of the Lagrange multiplier in the ℓ_1/ℓ_2-norm regularized square loss function, there is an increase in the sparsity of the model or, a reduction in the number of features selected. The value of λ was increased from 10 to 120 in monkey I and 5 to 90 in monkey R. By the end of experiments, a 40-50% reduction in the control set was seen in both the subjects.

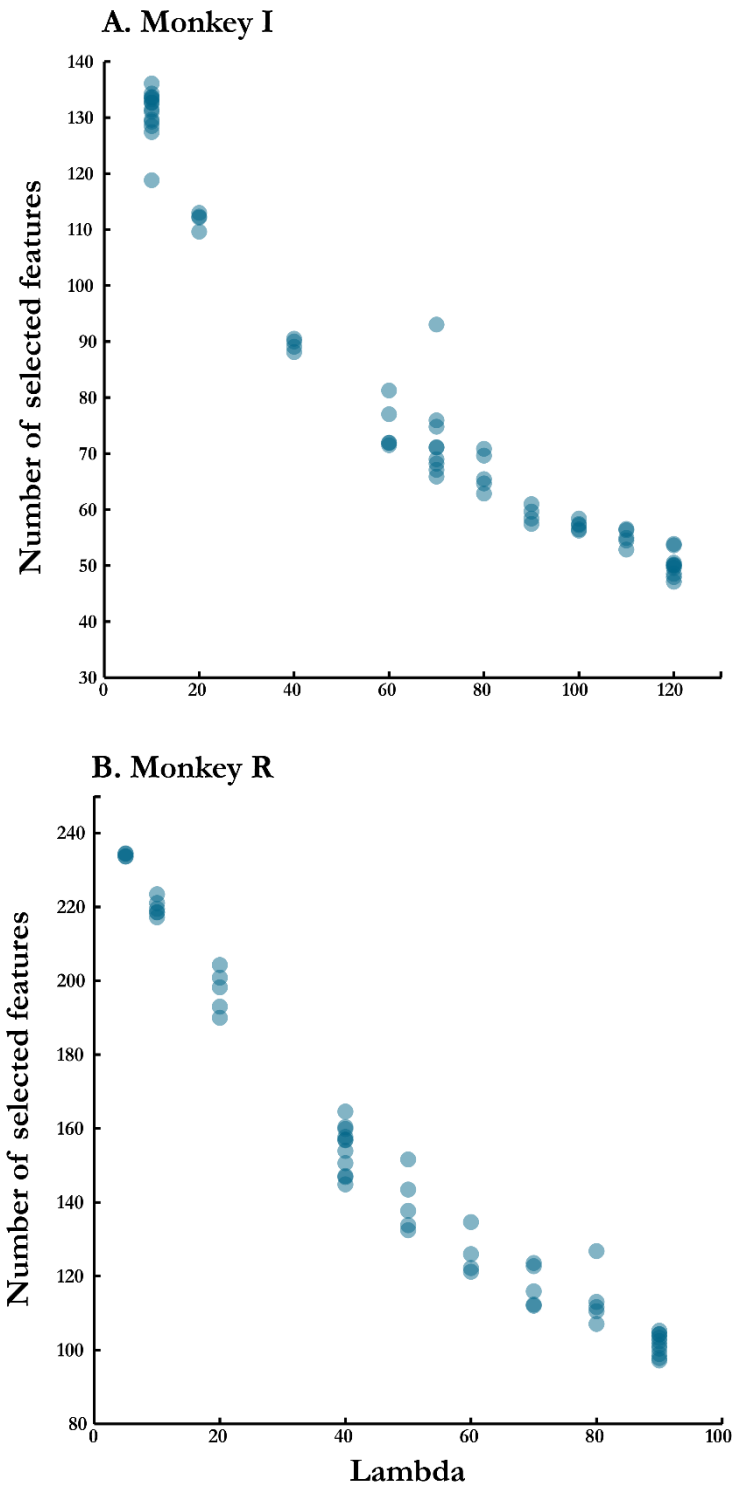

Figure 6.2: Number of selected features for a given value of the Lagrange multiplier lambda. A. Monkey I, B. Monkey R. Lambda was increased every 1-2 weeks of experiment. The figure shows the size of the sparse feature set obtained for different values of lambda. Each point represents the average number of features selected through one session (day) of experiment. As expected, the number of selected features reduces with the increase in the value of lambda, which penalizes the ℓ_1/ℓ_2-norm of the features.

6.2.2 Psychophysics and Performance

The performance in the 2D center-out BCI task was quantified using the movement time to target and the standard deviations from straight line trajectories. The changes in their values with the variation of λ were studied. Figure 6.3 shows the variation of median movement times at different values of lambda. Each point on the plots represents a mean calculated from one session.

A steep decrease in the median movement times is observed with the increasing value of λ. This implies that the subjects got faster as the number of control features were reduced (Figure 6.3, 6.2). When compared to the results in Figure 4.4 of Chapter 4, this observation follows an opposite trend. In the selective feature pruning study, the movement times were observed to increase as a smaller number of control features were used. The difference between the two studies lies in the gain term associated with the ECoG features. In the previous study, a constant gain was applied to the ECoG features, irrespective of the number of features selected. In the current experiment, the gain was scaled by the sum of feature weight magnitudes. Keeping the desired gain (G_d) constant, resulted in the increase in the value of ECoG feature gain (G_f) as the restraint on the ℓ_1/ℓ_2-norm was increased. This increase in the gain, most likely, overcompensated the decrease in the control signal magnitude, resulting higher cursor speeds and hence lower movement times.

To further understand the changes in the BCI performance, the trajectories of the center-out reaches were studied. Each correct center-out reach was segmented into 40 time

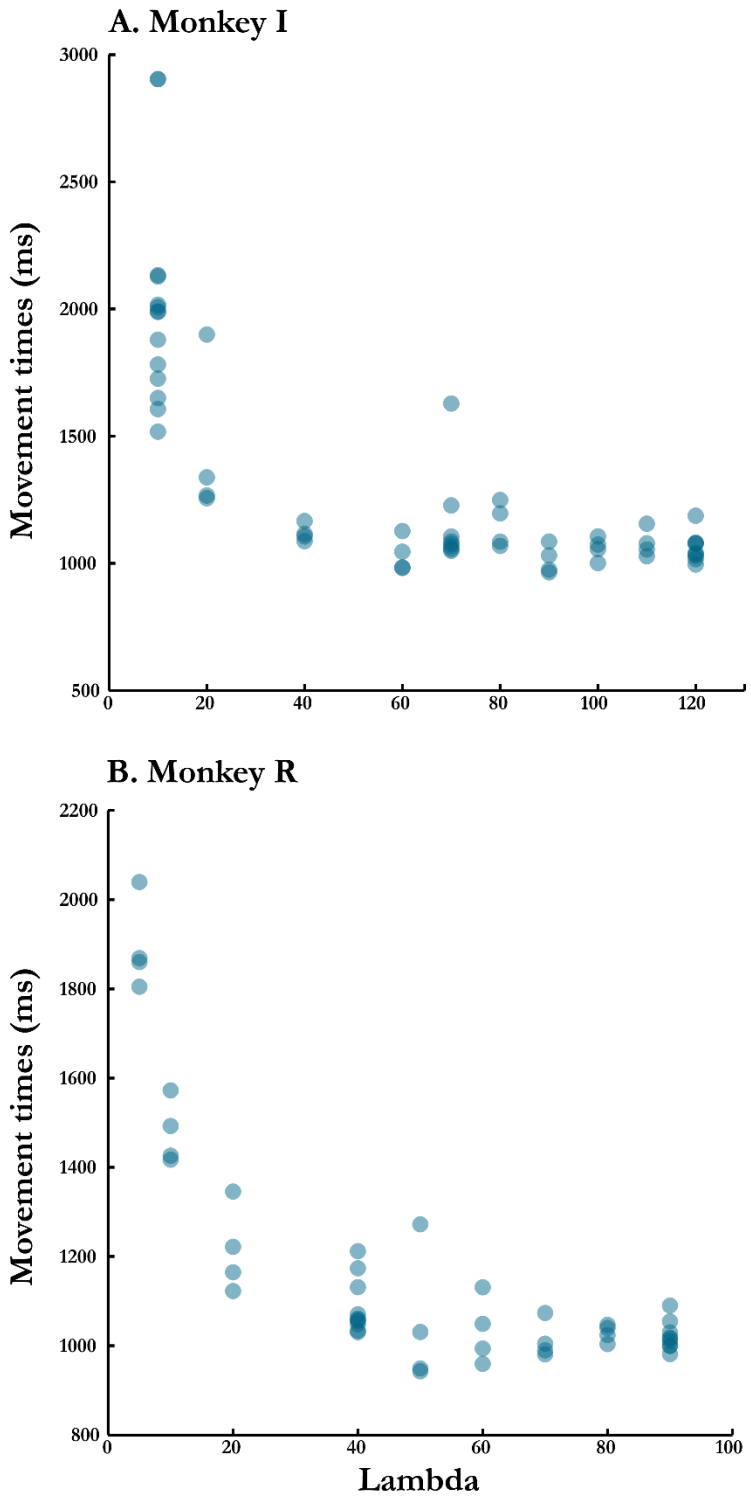

Figure 6.3: Variation in movement times with the Lagrange multiplier lambda. A. Monkey I, B. Monkey R. Movement times were calculated for correct reaches during an 8-target 2D center out BCI task. Only the trials with complete BCI control were used. Each point on the above graphs represents the median movement time during one session (day) of experiment. Both subjects were observed to perform faster reaches with the increase in the value of lambda that eventually plateau at around 1 second per reach.

bins, starting from the center to the target. Perpendicular deviations from the straight line trajectory were calculate in these time bins, for each trial. Forty distributions of perpendicular deviations were created from one session of experiment. Figure 6.4 shows the variation in mean standard deviations of these distributions as the value of λ was varied. Each point on the plot represents one session of recording.

A near linear increase in the mean standard deviations is observed with an increase in λ or reduction in the number of control features. This again is different from what was observed in Chapter 4, where the standard deviations did not change with the control feature set. Comparing the figures 6.4 and 6.3 a classic speed-accuracy trade-off becomes evident. It can be argued, as before, that the changes in ECoG feature gain G_f overcompensated for the changes in the magnitude of the control signal. If the decrease in movement times was a result of subjects getting better at performing the task, they would be expected to show straighter trajectories as well. Also, as the subjects were rewarded every time they successfully completed a trial and not for maintaining straight line trajectories, it was favorable to them to finish as many reaches as possible. Thus, faster movement times, irrespective of the nature of the cursor trajectory, were ideal for maximizing the reward during a session. To put it in layman terms, all the monkeys cared about was satiating their thirst as quickly as possible.

6.2.3 Distribution of Selected Frequency Bands

Figure 6.2 showed the changes in the feature space as the value of λ was varied. This section segregates the features in their respective frequency bands and studies the changes in

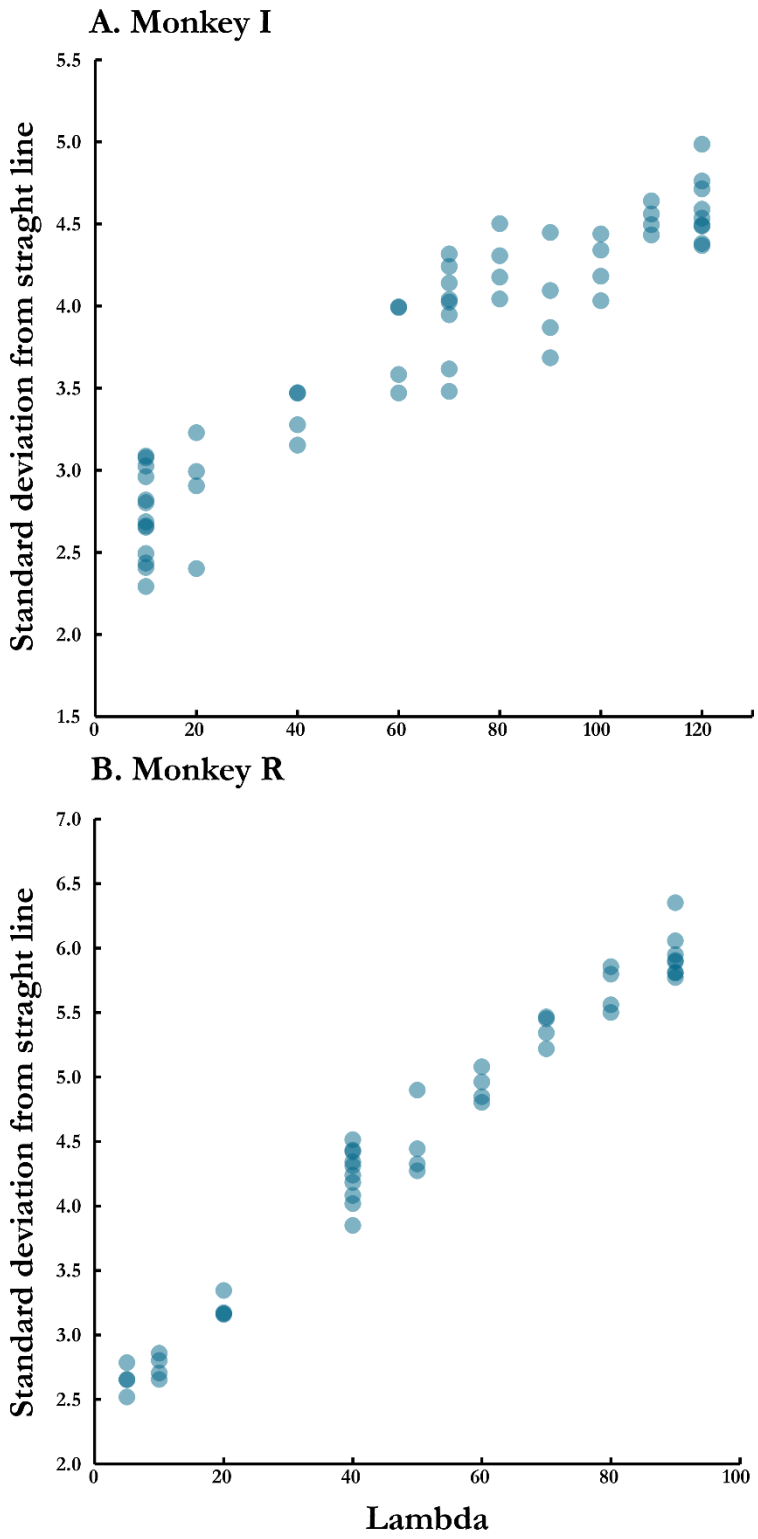

Figure 6.4: Deviation from straight-line trajectory with the Lagrange multiplier lambda. A. Monkey I, B. Monkey R. In this figure each point shows the mean standard deviation from a straight like trajectory in one session (day) of experiment. Units of standard deviation are arbitrary. In both the subjects the deviation is observed to increase with the value of lambda.

the distribution as the control feature set changed. Figure 6.5 shows the relative contribution of each frequency band at a given value of λ. The total height of each stacked bar represents

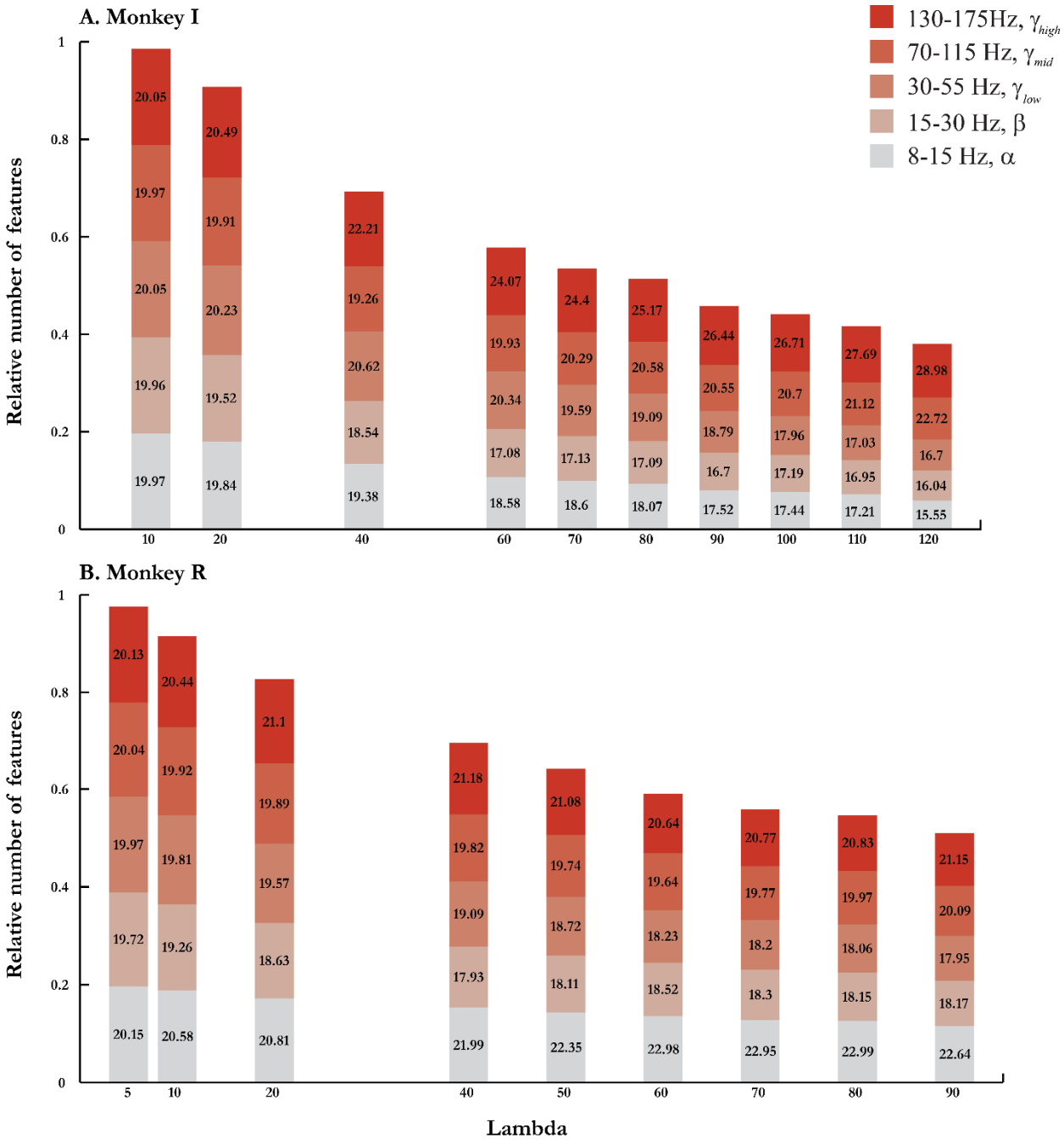

Figure 6.5: Relative number of features used from each frequency band at different values of lambda A. Monkey I, B. Monkey R. Each bar gives the distribution of features used by the subject at increasing values of lambda. The numbers for each band represent the percentage of times a feature from that band was picked. Subject I prefers features from the mid and high gamma bands. On one hand subject R shows very little preference to high gamma and alpha bands.

the total of fraction of the features in the control set, relative to the available feature space. Each bar in the stack gives the fractional contribution of one of the five bands towards the control set. The numbers inside the bars represent the percentage of total features from that band.

In the case of monkey I a monotonic increase in the contribution from mid-gamma and high-gamma band and a decrease in the contribution from the other bands is observed. Monkey R, on the other hand, shows small increases in alpha band and high-gamma band contributions. The mid-gamma band is relatively constant. The final distribution in the case of monkey I shows bigger differences between bands compared to the final distribution in monkey R.

As has been stated earlier, the decoding algorithm used in this chapter uses only the data from one single block to construct the model. There is no memory from one update to another. Therefore, if the subject does not follow a consistent control strategy, the distributions plotted in the above figure are expected to be random; that is, close to 20% contribution from each frequency band. With that rationale and the plots in Figure 6.5, it can be said that monkey I was more consistent in his control strategy or chose features that were consistently weighted, compared to monkey R.

To reinforce this, a consistency metric was calculate for each feature when it was selected by the decoding algorithm. This metric called the Resultant Vector Length (RVL) metric was computed using the CicrStat toolbox [65] and has already been described in Chapter 5. Using the updates from one session of experiment, each feature was assigned a

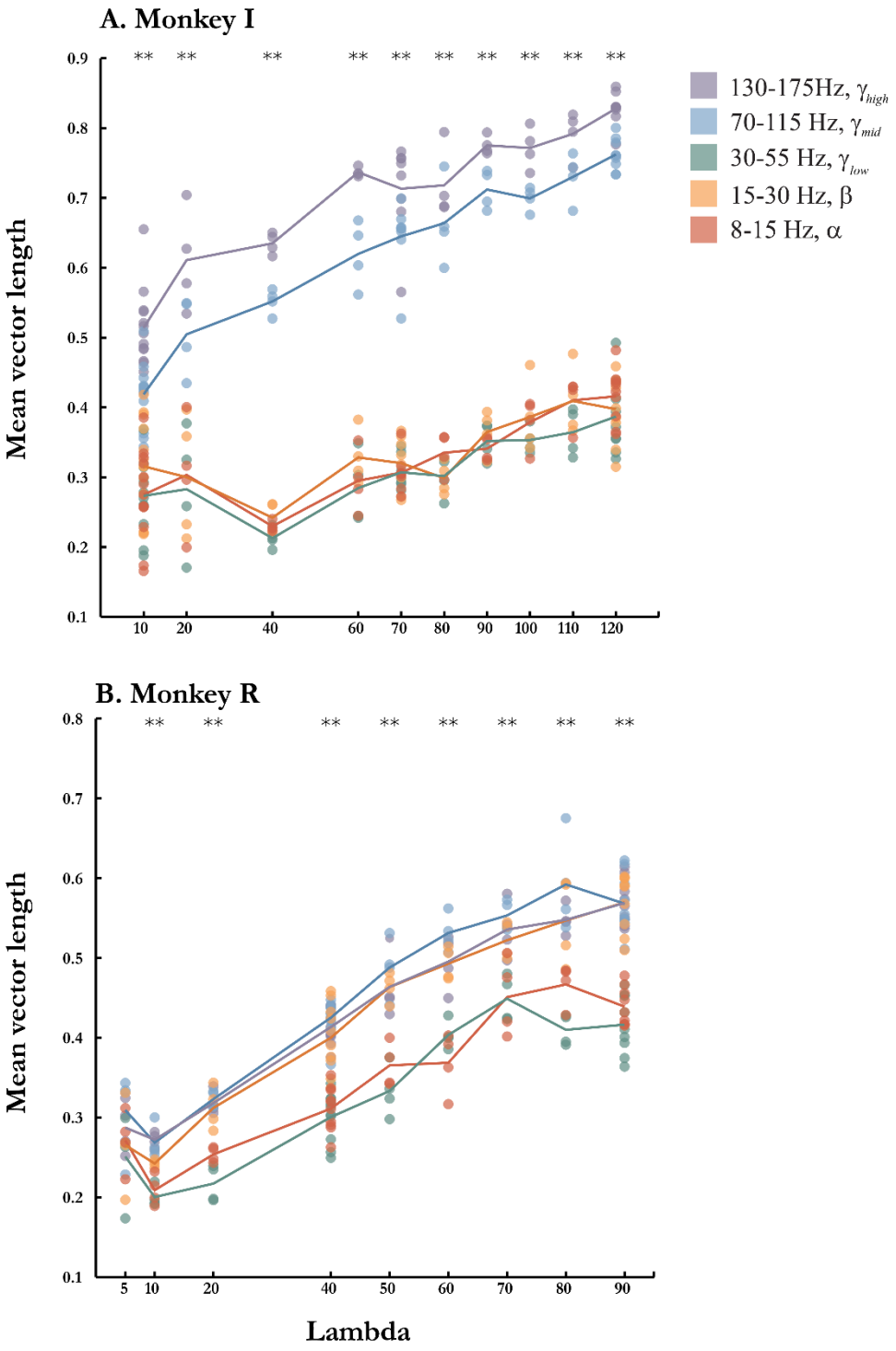

Figure 6.6: Consistency of features from different bands. A. Monkey I, B. Monkey R. Monkey I was clearly using consistent weights for mid and high gamma bands with the consistency increasing with lambda. ** denote significantly higher values for mid and high-gamma compared to other 3 bands with $p<0.01$. Monkey R though chose features from alpha band (Fig 5), the consistency of features in alpha band is lower than that for beta, mid gamma and high gamma. ** denote significantly higher values for beta, mid-gamma and high-gamma compared to other 2 bands with $p<0.01$.

resultant vector length for the distribution of preferred directions assigned to it. Data from the blocks where a given feature was not a part of the control set was ignored. The mean resultant vector for each frequency band was calculated on each day. Figure 6.6 plots these mean values segregated by the value of λ.

The first observation from Figure 6.6, for both the monkeys, is that the consistency of the feature set as a whole increases with the increase in the value of λ. That is, the decoding algorithm is choosing consistent features to construct the control feature set and the consistency of the features selected is higher when the control set is sparser.

Studying the plot for monkey I (Figure 6.6 A) it is evident that mid-gamma and high-gamma features are more consistent and hence more reliable to be used for BCI control. The other three frequency bands, show small increases with λ and have much smaller resultant vector lengths. The difference is highlighted by the conducting a paired t-test between the distribution formed by the combination of mid-gamma and high-gamma, and the second distribution formed by the other three bands. The ** on top of each λ value represent the distributions were significantly different with a p-value < 0.01. The differences are observed to be significant for all the λ values in monkey I.

Looking at Figure 6.6 B. for monkey R, the differences in resultant vector are not as clear as for monkey I. It is interesting to point out that, even though from Figure 6.5 B, monkey R is observed to select relatively more alpha band features, the resultant vector lengths for these features is lower than beta, mid-gamma and high-gamma bands. Again, to highlight the difference, data from alpha and low-gamma bands was combined into one distribution and

compared against the other three. The ** on top of each λ value represents that the distributions were significantly different with a p-value < 0.01.

6.2.4 Distribution of Selected Electrodes

This section analyzes the spatial distribution of features in sparse control set. The goal of this analysis is to determine if features from certain electrodes were more likely to be chosen compared to others. Because of the large number of electrodes in both the subjects (27 and 48), the differences in the contributions from individual electrodes become clear when only a small number of features are present in the control set.

Figure 6.7 shows the relative contribution from each electrode at the highest λ value used for each subject (120 for monkey I, and 90 for monkey R). Each column first shows the approximate location of the array and then the contribution of each electrode, given by the intensity of the color. The black lines under the arrays represent the approximate location of the central sulcus. For each hemisphere percentage contributions from the motor and sensory area electrodes are stated in the figure. It is to be noted that equal contributions from each electrode would lead to a value close to 3.70% (100*1/27) in monkey I and 2.08% (100*1/48) in monkey R.

The first observation from the figure is that features from certain electrodes are more likely to be chosen for the control set. Monkey I had bigger electrodes on the right hemisphere and the relative sizes of electrodes in monkey R are represented in the plot. Interestingly, in both the cases, larger electrode sizes are observed to contribute a little more than the smaller

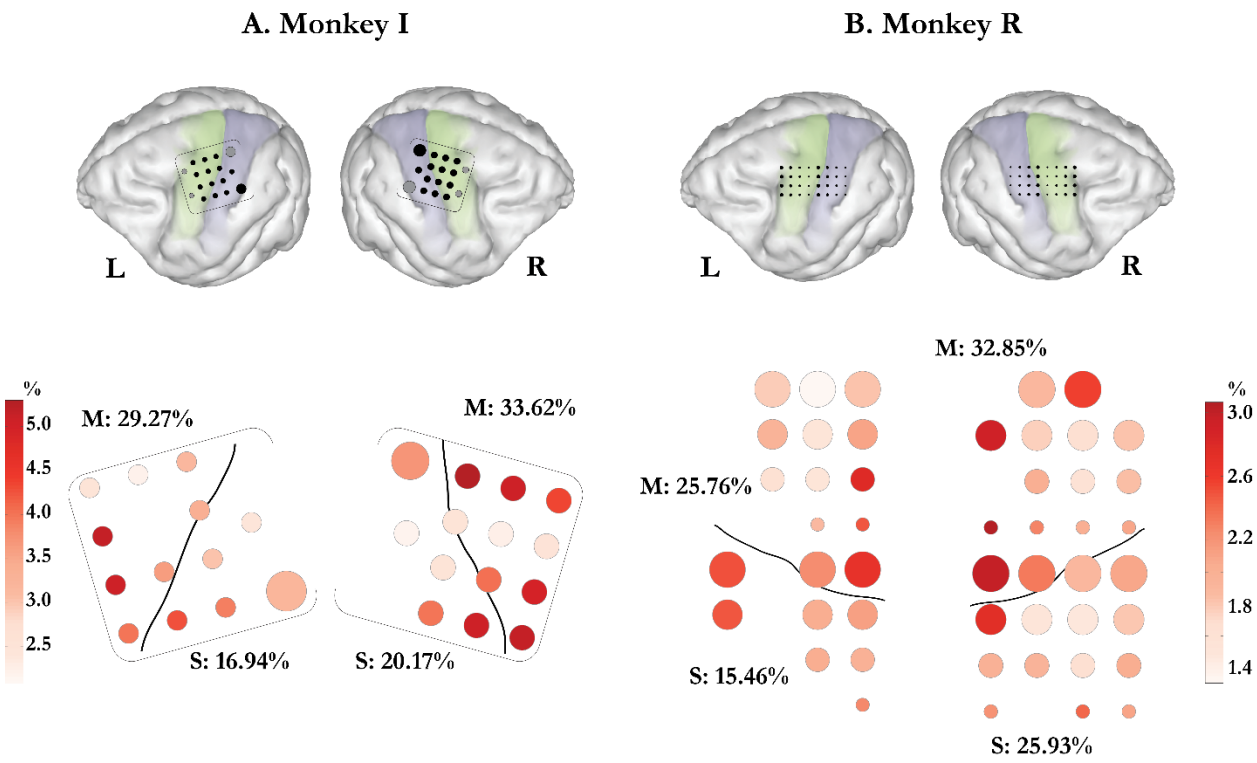

Figure 6.7: Electrode contributions at highest value of lambda for A. Monkey I at $\lambda = 120$, B. Monkey R $\lambda = 90$. The intensity of color of each electrode represents the relative number of times features from that electrode were picked by the algorithm. Equal contribution from all electrodes would lead to 3.70% (100*1/27) in monkey I, and 2.08% (100*1/48) in monkey R. The lines below the arrays in each column show the approximate location of the central sulcus with respect to it. Contribution of electrodes over motor and sensory areas were calculated.

ones. For both the monkeys, the relative contribution from the motor and sensory areas is also calculated and shown in the figure. In monkey I electrodes on the motor side of the sulcus comprise of approximately 63% percent of the control set and in monkey R that number is about 59%. Counting the electrodes on each side of the sulcus it is observed that 17 of the 27 electrodes in monkey I and 28 of the 48 electrodes in monkey R are placed on the motor side. These numbers are almost exactly same to the relative contribution of these electrodes towards the control set (17/27 ~ 63% and 28/48 ~ 58%). Looking at these results, it can be safely said that both motor and sensory electrodes contributed equally towards the control set.

To further investigate how often each frequency band was picked from each electrode at the highest value of λ, the array plots in Figure 6.7 were separated into the five individual bands. This is depicted in Figure 6.8. Each row in the figure represents one of the five frequency bands. The feature counts from each frequency band are plotted on the implanted electrode array. The intensity of each electrode on the array represents how often that particular feature was picked to be a part of the control set. The values range from 0 (never picked during an update) to 1 (picked by all the updates).

It is clear from the figure that monkey I prefers a few of the mid and high-gamma band features more than any other. These features are consistently picked by the decoding algorithm. There is also a high contrast in the plot for these two bands in monkey I. That is, some electrodes are almost always picked whereas others are almost never picked. This kind of contrast is missing in the other three bands where all the electrodes are picked with a low probability (randomly). Looking at the column for monkey R, higher contrast is again seen in the mid and high-gamma band. Alpha band features in monkey I was picked more often at $\lambda = 90$, as seen from the last bar in Figure 6.5 B. This is also apparent from the first row in Figure 6.8 where the relative values are higher for almost all the electrodes. Again, the lack of contrast in the plot shows that though the alpha band features were overall picked more often than others, within the band they were picked randomly.

The high and low contrasts in the plots also reinforce the observations from Figure 6.6. Features that were almost always picked to be a part of the control set were also consistently weighted. Due to this reason the average resultant vector values in mid-gamma

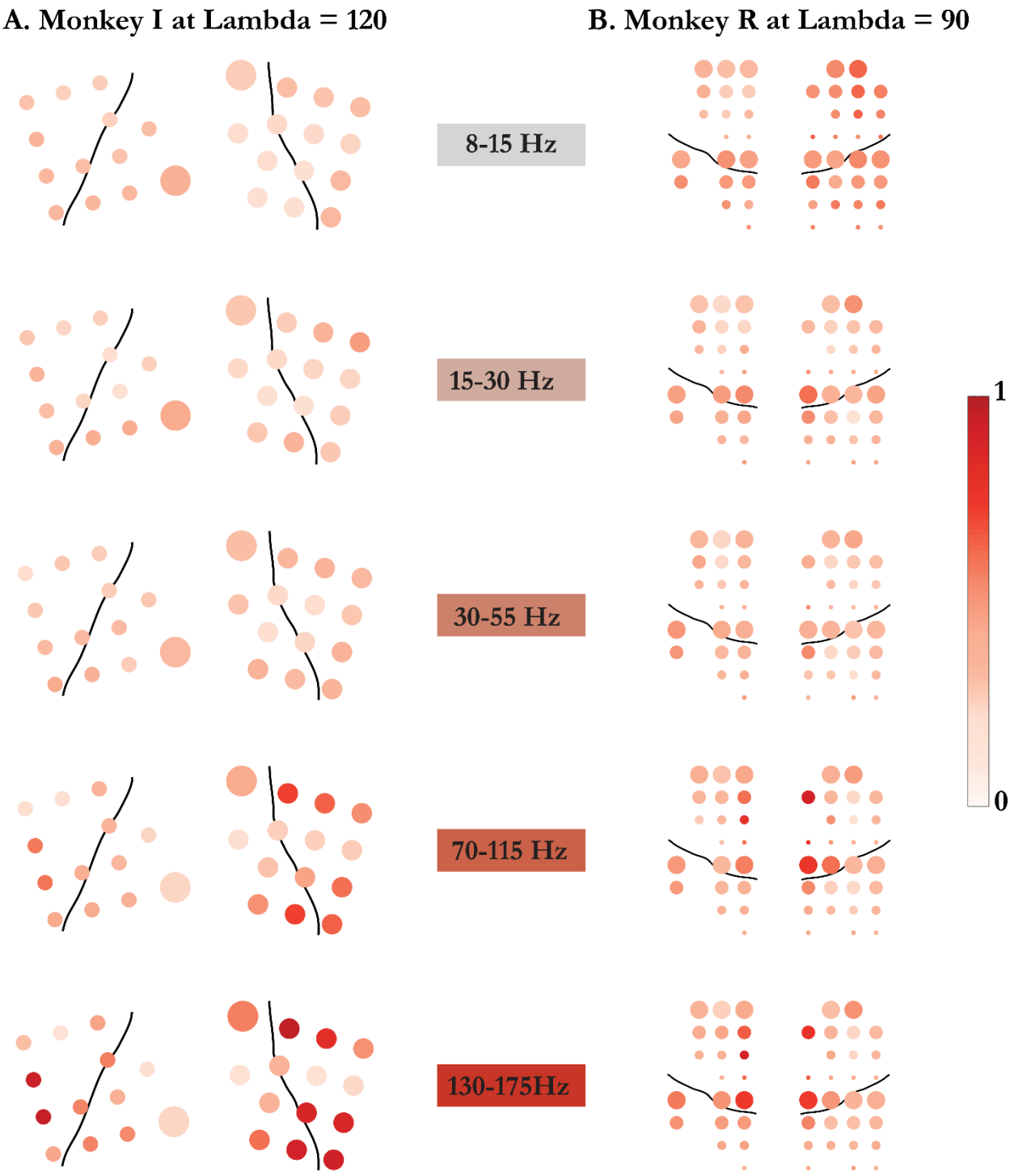

Figure 6.8: Features selected at highest lambda value separated into individual bands and represented over the implanted electrode grid. A. Monkey I, at $\lambda = 120$, B. Monkey R, at $\lambda = 90$. The intensity of each electrode represents how often that particular feature was picked. Subject I is observed to prefer the features over motor areas from mid and high gamma bands. Other bands are observed to play very little role in control. Subject R on the other hand doesn't show strong preference to bands except for a few electrodes in mid and high gamma.

and high-gamma band, in particular for monkey I, are much higher than other bands. This is also the reason that even though the alpha band features in monkey R were picked more often, they showed low values for the mean resultant vector due to the randomness within the band.

6.2.5 Performance Using a Fixed Model

At the end of the experimental sessions with the highest value of lambda, experiments were carried out for two weeks with a fixed decoding model. During this time, no updates to the model were carried out and the fixed model was loaded in the system at the beginning of the session. Constant speed assistive bias was used during the initial few blocks of experiment to help the subjects start the task.

The weight matrix for the fixed model was selected from the set of updates carried through the sessions with the maximum value of λ. The selected fixed model was intended to be the best representative of all the updates. To find this the ℓ_2-norm of all the weight matrices was computed to get the magnitude of the weight vector for each feature. Cross correlation between these magnitude vectors was computed. A mean of the correlation values was computed for each magnitude vector. The magnitude vector that had the highest mean correlation with all the other magnitude vectors was termed as the best representative of the all the updates. The weight matrix $\mathbf{W}$ corresponding to this magnitude vector was used as the fixed model for two weeks of BCI experiments. The total number of features in the fixed model was 38 for monkey I and 93 for monkey R. The frequency band distribution of these

features is given in Table 6.1 with the number of features from each frequency band and their relative percentage given in brackets.

Table 6.1: Frequency band distribution of the fixed model for monkey I and R

Fixed Model	**8-15 Hz (α)**	**15-30 Hz (β)**	**30-55 Hz (γ_{low})**	**70-115 Hz (γ_{mid})**	**130-175 Hz (γ_{high})**
Monkey I	1 (2.63%)	2 (5.26%)	8 (21.05%)	15 (39.48%)	12 (31.58%)
Monkey R	19 (20.43%)	18 (19.36%)	16 (17.20%)	21 (22.58%)	19 (20.43%)

During these two weeks of experiments the value of the desired gain G_d was kept the same as before for monkey I (50). On the other and, due to high variations in monkey R's trajectories at $\lambda = 90$ (Figure 6.4), the desired gain in his case was reduced to 50 from 75. Figure 6.9 shows the performance variations in the two monkeys going from the sessions of experiment using the maximum λ value to the days using the fixed model in Table 6.1. Each point on the plot in Figure 6.9 A represents the median movement time from one recording session. In Figure 6.9 B, each point represents the mean standard deviations from a straight line trajectory.

In the case of monkey I, where the desired gain was same between the two tasks, a small but significant improvement in performance was observed. Monkey I performed faster reaches that were closer to a straight line trajectory while using the fixed decoding model. In the case of monkey R, a big reduction in the trajectory deviations was observed with the slowing down of the cursor. As described earlier, the desired gain term for monkey R was reduced during the days of using the fixed decoding model. This change in the gain term, as

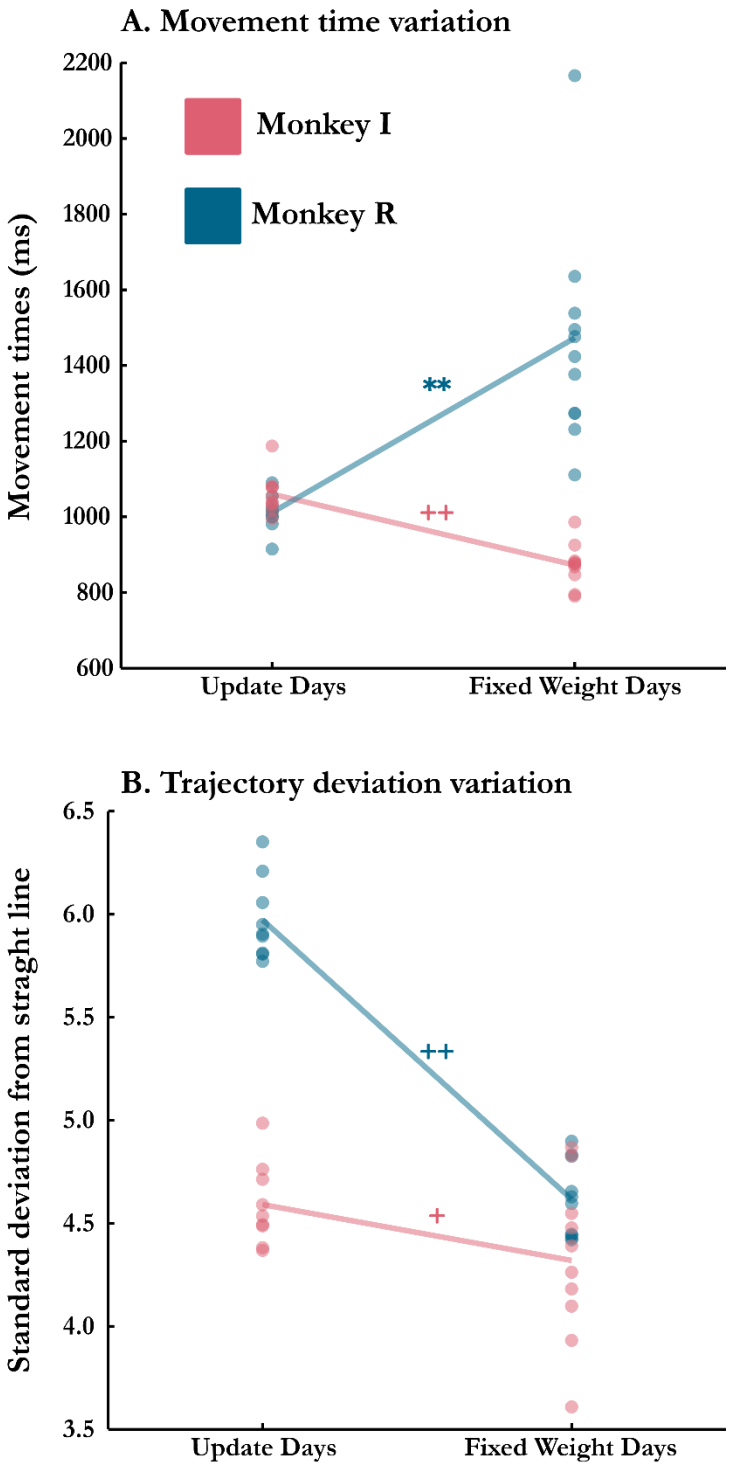

Figure 6.9: Performance comparison between update days and days using a fixed decoding model. A. Movement time variation. Monkey I decreased median movement times during the days of fixed weights whereas Monkey R seemed to have increased median.

expected, resulted in the observed performance change. The observations from this comparisons show that though an improvement in performance can be achieved by practicing with the same model, the gain term associated with the control signal also plays a significant role in the cursor movement.

6.2.6 Performance Using a Random Feature Set

The last week of experiments carried out for this chapter used a random set of features. The goal was to study how a simple decoding algorithm using squared loss function and randomly selected set of features compares against the decoding algorithm in this chapter. The number of randomly selected was same as the number of features in the fixed model in the previous section (38 for monkey I, and 93 for monkey R). The frequency band distribution of these features is shown in Table 6.2.

Table 6.2: Frequency band distribution of the randomly selected features for monkey I and R

Random Features	**8-15 Hz (α)**	**15-30 Hz (β)**	**30-55 Hz (γ_{low})**	**70-115 Hz (γ_{mid})**	**130-175 Hz (γ_{high})**
Monkey I	8 (21.05%)	10 (26.32%)	7 (18.42%)	5 (13.16%)	8 (21.05%)
Monkey R	21 (22.58%)	21 (22.58%)	18 (19.36%)	19 (20.43%)	14 (15.05%)

A squared loss function without the constraint of any regularization term was used for one week of experiments. Same experimental protocol of starting each session with a naïve decoder and using the constant speed computer assistance for the initial blocks of trials was used. The desired gain term G_d was set to 50 for both the subjects.

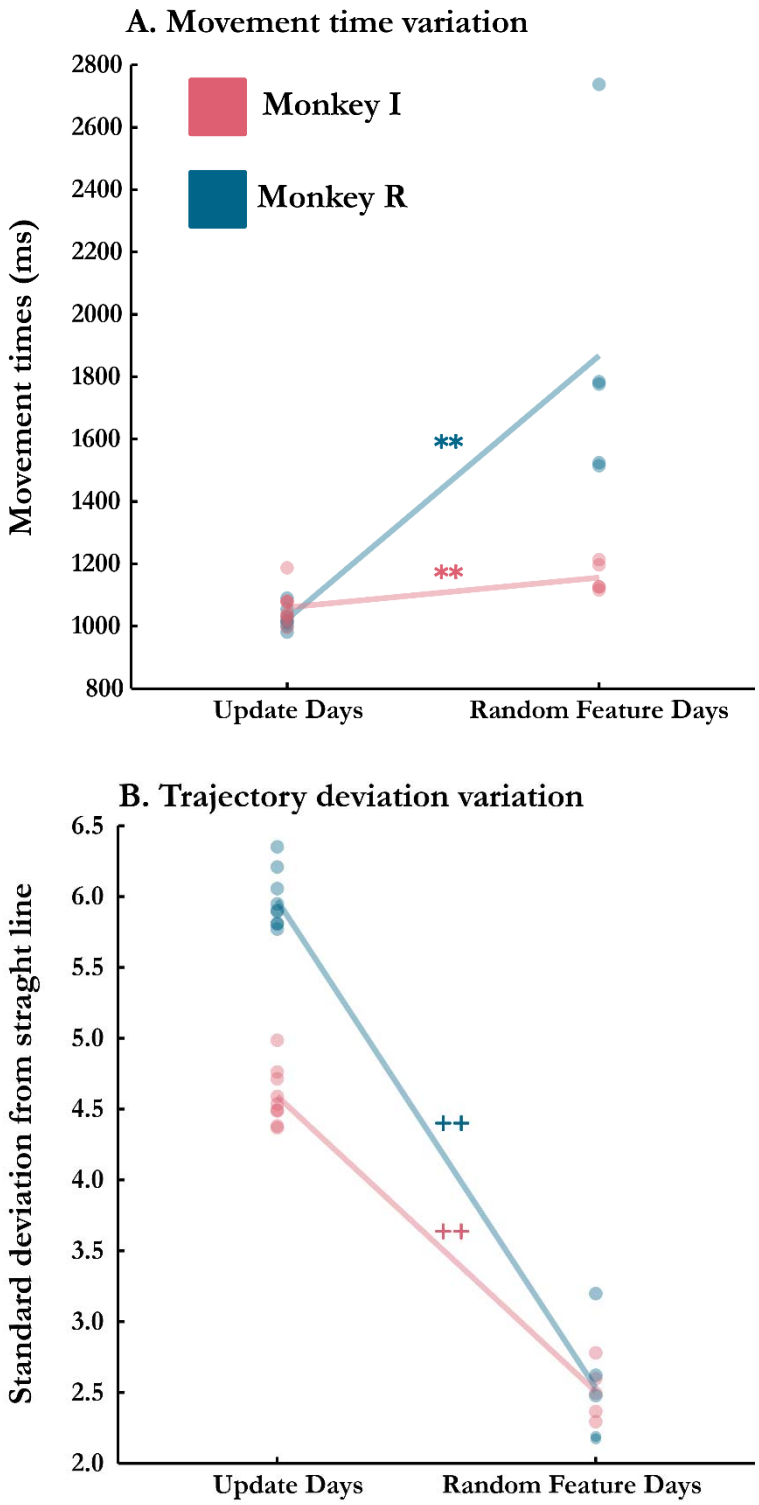

Figure 6.10: Performance comparison between update days and days using a random set of features. A. Movement time variation. Both the monkeys exhibited higher median movement times using the random set of features. Trajectory deviation variation. Both the monkeys showed a smaller mean deviation in the trajectory from the straight line. Significant increase is denoted by ** ($p<0.01$) and significant decrease is denoted by ++ ($p<0.01$).

Figure 6.10 shows the performance comparison between the days of experiments using the highest value of λ and the one week of experiments using the random set of features. The median movement times to targets (Figure 6.10 A) and deviation from the straight line trajectory (Figure 6.10 B) were used to measure the performance. Each point in the two plots represents one session of experiment.

In the case of monkey I, the desired gain G_d was same for both the update days and random feature days. Figure 6.10 A shows a small but significant increase in the movement time while using the random features. In the case of monkey R, the movement time increase is much higher. However, the desired gain, in this case, was set at 75 for update days and 50 during the days using random features. The reduction in gain is, most likely, the cause of such a big difference in the values. It is to be noted that monkey R also shows a lot of variability during the random feature days.

From Figure 6.10. B it is clear that irrespective of the desired gain value both the subjects show a large and significant decrease in the mean standard deviations from the straight line trajectories. One of the explanations for such well-restrained trajectories during the random feature days could be the decoder used during these days. As mentioned earlier, no regularization term was used with the random feature set and hence the optimal solution during each update was only dependent on the data collected from the previous block. In the other case, while using a ℓ_1/ℓ_2-norm constraint, the output weights were dependent both on the data and the constraint placed on it. At the highest value of the Lagrange multiplier λ, the constructed model was more biased towards producing a sparse model than reducing the

squared loss. Thus, it is likely that the models constructed using the random features were very well fitted to the localized data and produced better trajectories. However, seeing the high variability in the movement times for monkey R, it can also be said that the models constructed without the regularization term were very different across the recording sessions.

6.3 Conclusions

Using a sparsity inducing decoding algorithm, this chapter demonstrated a method of feature selection for ECoG BCI. Several weeks of experiments were conducted while gradually decreasing the control feature set while post hoc analysis was carried out to study the resulting performance and control set characteristics. Experiments were also conducted with a fixed model and a random set of features. This section discusses the implications of the results observed in Section 6.2.

The first important observation is that the decoding algorithm was able to reduce the control feature set that produced faster reaches during the task. However, a steep increase in the variability of the trajectories was also seen. As discussed in brief previously, the ℓ_1/ℓ_2-norm constraint on the decoder biased it to produce an output that was sparse. With the increase in the value of λ, this bias was increased and was possibly one of the causes of variability in the trajectories. This is depicted in Figure 6.4, where a nearly linear increase in the trajectory deviations with the increase in the value of λ is observed.

The upside of using the regularization term, and increasing the bias towards it, is evident from Figure 6.6. By biasing the model towards producing a sparse feature set while reducing the squared loss, the decoding algorithm chose features that were more consistent

and reliable. The results in this figure showed that mid-gamma and high-gamma features were more consistent in both the subjects. In monkey R, the mean resultant vector for the beta band was also observed to be close to the mid-gamma and high-gamma bands. The overall values of mean resultant vector for monkey R were lower than those for monkey I. This can be attributed to the difference in the experience of the two subjects. Monkey I was a seasoned BCI subject with over 3 years of experience. The experiments carried out in that time though used different behavioral task and/or decoding algorithms, the recorded set of features was always the same. By using the same features for all that time, monkey I knew what the best way was to perform a BCI task. This knowledge manifested into the reliable modulation of certain features and randomness in others. Monkey R, on the other hand, had only performed a 1D BCI task with a subset of features for a few months, before starting the experiments for this study. He was most likely still in the process of figuring out the best strategy he could use for BCI control.

Going ahead with the observed results, the performance using a random feature set raises a few very important questions that cannot be ignored. A relatively complex decoding algorithm was designed for this chapter to determine a small subset of features for robust BCI control. Several weeks of experiments were conducted to determine this control set. However, a random set of features and a simple decoding algorithm was also able to demonstrate almost equivalent performance. With respect to the straight line trajectories, the random feature set even performed better. There could be a few explanations to this observation. It is to be noted that the goal of the regularized decoding algorithm was to obtain a small subset of features

that were most reliable for control. Once an optimal subset is obtained, it is probably important to construct a decoding model using only a squared loss function. Secondly, due to the nature of the output of the regularized decoder, it cannot be used to blend with the previous models and hence maintain a memory. This problem can be solved by using a decoding algorithm similar to the one described in Section 4.1.5 of chapter 4, but only with the optimal subset of features determined by the regularized decoder. Training the subject using the subset of features and the simple algorithm with a smoothing factor could produce the desired results.

Finally, the distribution of the randomly selected features in Table 6.2, shows the presence of mid-gamma and high-gamma features in both the subjects. The presence of these features, which have been observed to be more reliable in control from different experiments in Chapter 4, 5, and 6, could also explain the reason for the performance observed. It shows that good BCI control can be achieved with the presence of a few reliable features. This is investigated further in Chapter 8.

Chapter 7
Modulation of ECoG Signals

Previous chapters of this dissertation have described ECoG BCI experiments that used different feature selection and decoding algorithms to obtain a small subset of control features. These studies were also used to determine which frequency bands from the recordable ECoG spectrum are most likely to be used for control. The experiments from Chapters 4, 5, and 6 have all demonstrated the importance higher frequency ECoG spectrum. 70-115 Hz band (mid-gamma) was always observed to be most reliable. The high-gamma band (130-175Hz) showed very similar characteristics to mid-gamma when the recorded noise was lower. While the two higher frequency bands were typically the best control bands, the beta band (15-30 Hz), was usually the third best control band (with alpha and low gamma being the worst two control bands) Since the power in each band varies inversely with frequency, the beta band is less susceptible to noise than the high gamma bands. In situations where there is a lot of external noise affecting the recordings, the beta band may yield the most reliable control signal for BCI.

The experiments in this chapter were designed to compare the modulation of mid-gamma versus the beta band while they were individually used for BCI control. As mid-gamma and high-gamma bands had earlier shown similar tuning characteristics, it is believed that they are generated by the same neuronal source; thus, only the mid-gamma was chosen for this

analysis. A novel 1D BCI task that required both positive and negative modulation of control signals was designed. The first goal of this study was to determine that, given an incentive, could the subjects be trained to increase the modulation frequency of a particular spectrum of ECoG signals? Secondly, the two bands, mid-gamma and beta, were compared against each other to determine if there were differences in their modulation frequency; namely, can one band be modulated faster than the other band?

This chapter is organized in three sections. Section 7.1 describes the subjects and their recording electrodes, experimental tasks and the BCI control signal and algorithm. Section 7.2 discusses the observations from the experiments and the results obtained from post hoc analysis. Section 7.3 concludes the chapter.

7.1 Methods

7.1.1 Subjects

Two male Rhesus Macaques (*Macaca mulatta*, monkey R, and T) weighing 6-10 kilograms were used as subjects for the experiments in this chapter. Both the monkeys had bilateral implants on the left and right hemispheres of their brains. The experiments described in this chapter were carried out a little less than two years after the recording electrode arrays were implanted. In the period before this study, subjects had only gone through a general behavioral training in the laboratory that included simple 2 or 4 target center-out tasks using a joystick. For 1-2 weeks preceding the study, subjects were trained to perform the 1D task described in Section 7.1.3 using a joystick.

7.1.2 Electrodes Arrays: Design and Placement

Both monkey R and T were bilaterally implanted with thin-film polyimide arrays designed in collaboration with Dr. Justin Williams' Lab at University of Wisconsin, Madison. The details of the array design (Figure 7.1) and electrical connectors are described in Section in Section 3.2 of Chapter 3. All the arrays were implanted in the rostral-caudal direction. The 19 mm craniotomy center was located at the stereotactic coordinates (Anterior 15 mm, Lateral 15 mm) on both the hemispheres. This placed the center of the anterior-most row of

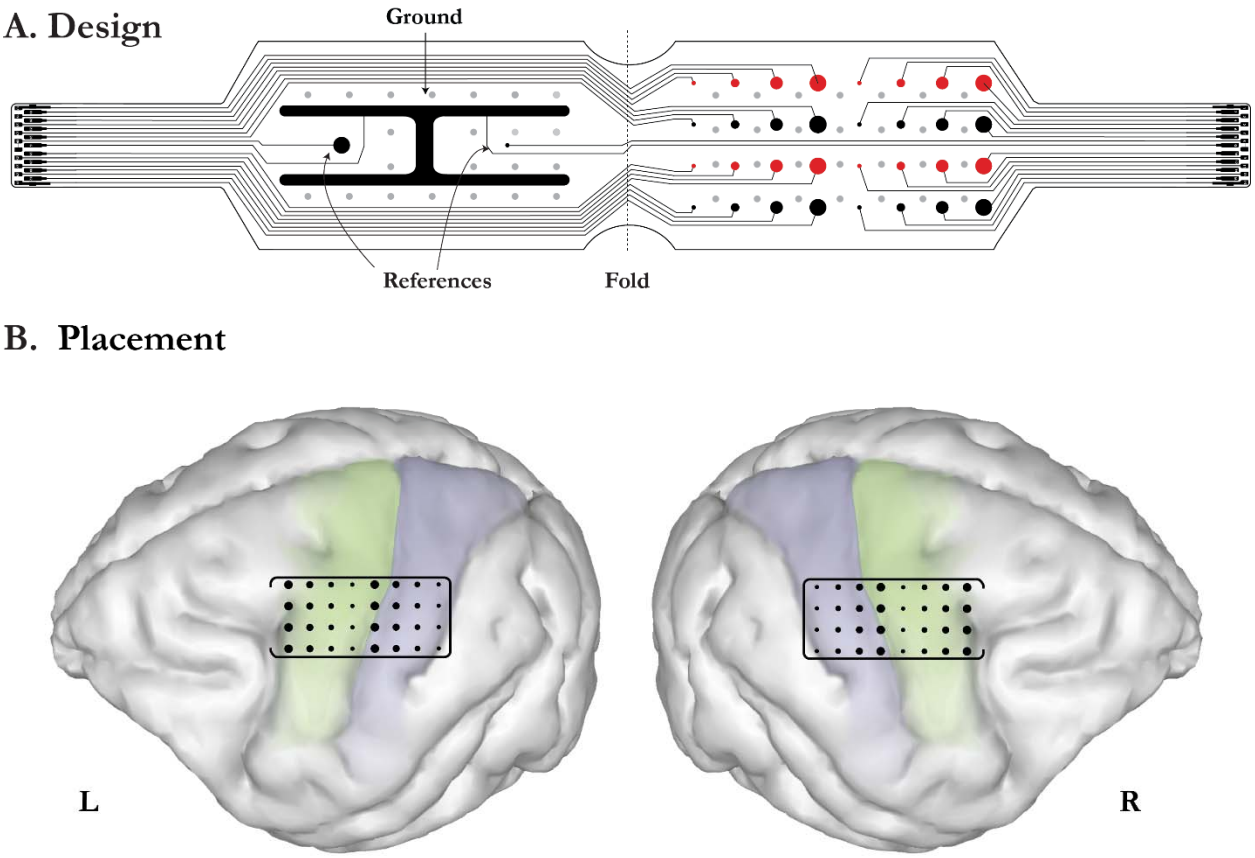

Figure 7.1: Electrode Design and placement. A. Polyamide thin film ECoG electrode array layout with one half housing 32 channels of four different sizes (300µm, 600µm, 900µm, and 1200µm diameters) of exposed electrode surface and separated by 3mm (center to center). The other half is covered with an "H" shaped ground and reference electrodes. B. Placement in both left and right hemisphere covering areas ranging from pre-motor to S1.

electrodes at approximately (Anterior 20 mm, Lateral 15 mm). The arrays thus covered areas spanning from pre-motor to S1. The approximate locations of the implanted arrays are shown in Figure 7.1 B.

7.1.3 Behavioral Task and Experimental Protocol

A novel 1D Up-down path task (Figure 7.2) was used in the experiments in this chapter. Each trial started with the cursor at the center. After an initial 200 ms hold period, the first target was presented either vertically above or below (Figure 7.2 A). Once the subject touched the first target with the cursor, a second target in the opposite direction appeared (Figure 7.2 B). After the subject was successful in reaching this target, a third target at the center was presented (Figure 7.2 C). The trial ended when the subject reached back to the center target. If the subject was able to finish all the steps in a set maximum allotted time, the trial was marked as correct. In any other case, the trial was marked as incorrect. To ensure the subject did not perform the task just by memory, the position of the first target was pseudo-randomly selected to be either up or down target. The second target was presented in the vertically

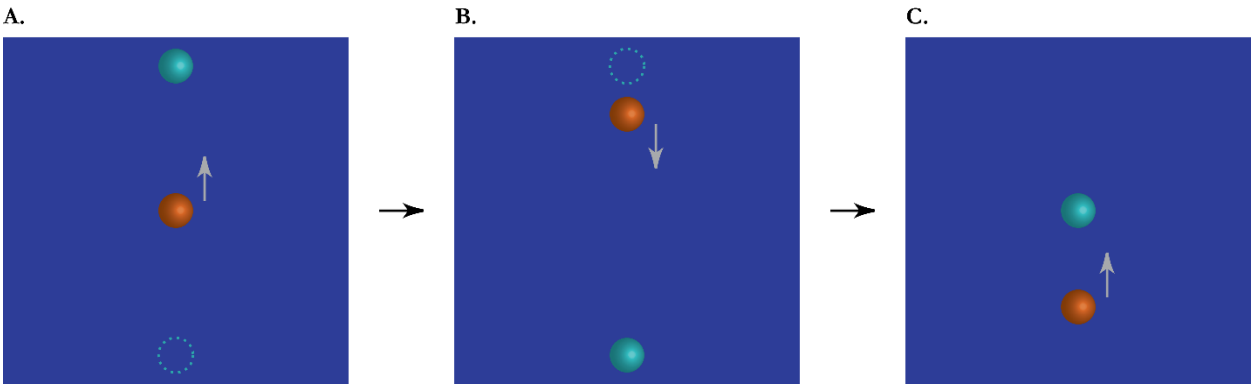

Figure 7.2: Up-down Path Task. A. First target is presented either vertically above (solid sphere) or below (dotted circle) at the beginning of the task. B. After the cursor reaches the first target, a second target is presented. C. A third target is presented at the center after the second target is reached. The trial ends when the cursor reaches back to its starting position. The dotted circles show the location of targets if the first target was a down target.

opposite direction to the first target. All the targets and the cursors were spheres of radius 3 screen units. The distance of the targets from the center in either direction was set at 15 screen units.

The completion of each trial required two changes in the direction of the cursor movement. To accomplish this, the control signal needed to modulate around the mean. The goal of the task was to encourage the subjects to maximize the modulation frequency of the spectrum of ECoG contributing to form the control signal. For this reason the amount of reward given at the end of each trail was inversely proportional to the duration of the trial. The water reward was linearly varied between a minimum and maximum value depending on how long it took to finish the trial, with shorter duration trials awarded larger rewards.

Similar to previously described experiments, each recording session was divided into blocks of trials. In this experimental setup, each block consisted of 4 correct trials. The position of the first target was selected in a pseudo-random order with each up and down targets chosen 2 times. A constant speed bias was also used for a few blocks at the beginning of the session. Each session started with a naïve decoder and a watch-task block where the cursor was completely controlled by the assistive constant speed bias. Using the data collected during this block, an initial model was created with the decoding algorithm described in Section 7.1.4. Similar to previous experiments, this model was updated after each block of trial and the cursor control was completely transferred to the monkey in 5-10 blocks.

As mentioned earlier ECoG signals from two frequency bands (mid-gamma: 70-115 Hz, and beta: 15-30 Hz) were individually used as control signals for these experiments. Four

weeks of experiments were carried out using each frequency band. To avoid learning effects, these weeks of experiments were interlaced with each other. The experiments started with Week 1 of using mid-gamma band followed by Week 2 of using beta band. Following the same sequence, each odd numbered week of experiment used mid-gamma band for control and each even numbered week used the beta band.

A second target tracking task was interlaced with the up-down path task during each session of experiment. The task involved tracking a target moving in a sinusoidal fashion at a set frequency. The task has been described in detail in Section 3.6.3 of Chapter 3. Each block of the up-down task was followed by a block (4 correct trials) of target tracking task. The decoder model was updated following each up-down path block and used for control in the following target tracking block and up-down path block. The decoder was not updated using the data collected during target tracking blocks. The observations from this part of the experiments are presented at the end of Section 7.2.

7.1.4 Control Signal and Decoding Algorithm

Subjects controlled the velocity of the cursor in the 1D BCI tasks used in this study. As mentioned earlier, signals from a single frequency band were used to control the movement of the cursor. Raw ECoG signals were separated into five different frequency bands and went through a few processing steps to form "control features". The details of the processing steps have been outlined in Section 3.4 of Chapter 3. A weighted linear sum of control features from the chosen frequency band (either mid-gamma or beta) along with the constant speed bias was used to determine the 1D velocity of the cursor. The velocity signal was integrated using Euler

Integration to plot the position of the cursor on the screen. The movement of the cursor was restricted to the vertical dimension. The vertical velocity V of the cursor at a time t is given by:

$$V(t) = E(t) + B(t) \quad \textbf{(7.1)}$$

Where $E(t)$ and $B(t)$ represent the ECoG control signal using the chosen frequency band and the constant speed bias, respectively and are given by:

$$E(t) = G_f \mathbf{W}^{\mathrm{T}} F(t) \quad \textbf{(7.2)}$$

$$B(t) = G_b U(t) \quad \textbf{(7.3)}$$

$\mathbf{W}$ is the weight vector (N, N = number of recording electrodes) used during the current block, $F(t)$ is a vector of z-scored features (length N) at time t, and $U(t)$ is the relative direction of the current target from the cursor at time t. The direction is equal to 1 when the target is above the cursor and -1 when it is below the cursor. The superscript T denotes a transpose. Terms G_f and G_b represent the gain applied on the ECoG control signal and the bias signal, respectively. The ECoG signal gain G_f was kept constant through the days of experiments. The bias gain term G_b started out equal to a 100 during the watch-task block at the beginning of the session and was then gradually reduced to 0 in 5-10 blocks, transferring the control of the cursor completely to the subject.

The weight vector $\mathbf{W}$ was initialized to all zeros at the beginning of the recording session. A decoding algorithm similar to the one described Section 4.1.5 of Chapter 4 was used to update $\mathbf{W}$ after each block of up-down path task. The modulation of the signals from the frequency band chosen for control during the movement of the cursor was recorded in the matrix $\mathbf{F}$. The values of the relative direction of the target from the cursor was also recorded during the movement of the cursor as $\widehat{\mathbf{V}}$. The value of $\widehat{\mathbf{V}}$ was equal to 1 when the target was above the cursor and -1 when it was below. A new set of weights $\mathbf{W}_{new}$ from the current block of trials was determined using the following equation:

$$\mathbf{W}_{new} = \mathbf{F}^{+}\widehat{\mathbf{V}} \tag{7.4}$$

Where the $+$ symbol denotes a Moore-Penrose inverse. The weight vector $\mathbf{W}$, to be used for the following block, was calculated by blending $\mathbf{W}_{new}$ with the previously used weight vector.

$$\mathbf{W} = \alpha\mathbf{W}_{new} + (1 - \alpha)\mathbf{W}_{pre} \tag{7.5}$$

$$\mathbf{W}_{pre} = \mathbf{W} \tag{7.6}$$

Where α, is the smoothing factor used to blend the weight vectors. This smoothing factor was used to prevent any local overfitting to the data and was set at 0.2 for these experiments.

7.2 Results

7.2.1 Performance and Learning during Up-Down Path Task

The main goal of the up-down path task was to encourage the subjects to maximize the modulation frequency of the control features. As each trial required both up and down modulation of the frequency band chosen for control, the total time taken to move the cursor in each trial could be used to quantify the modulation frequency. Figure 7.3 shows the changes in the mean movement time with the weeks of experiment.

As stated earlier, mid-gamma band features were used for control in odd numbered weeks and beta band features were used in even numbered weeks. The first important observation is that both subjects gradually reduced the movement time as the weeks of experiments progressed. Secondly, a difference in the two control frequency bands is observed. Though both the mid-gamma band features were able to decrease the movement times with each passing week, mid-gamma band was always observed to be faster than the beta band.

This is a new and important finding in the field of ECoG based BCI. The rate of modulation of the control signals is a crucial factor responsible for efficient control of a BCI system. Signals that can change the direction of modulation at a higher rate would be more

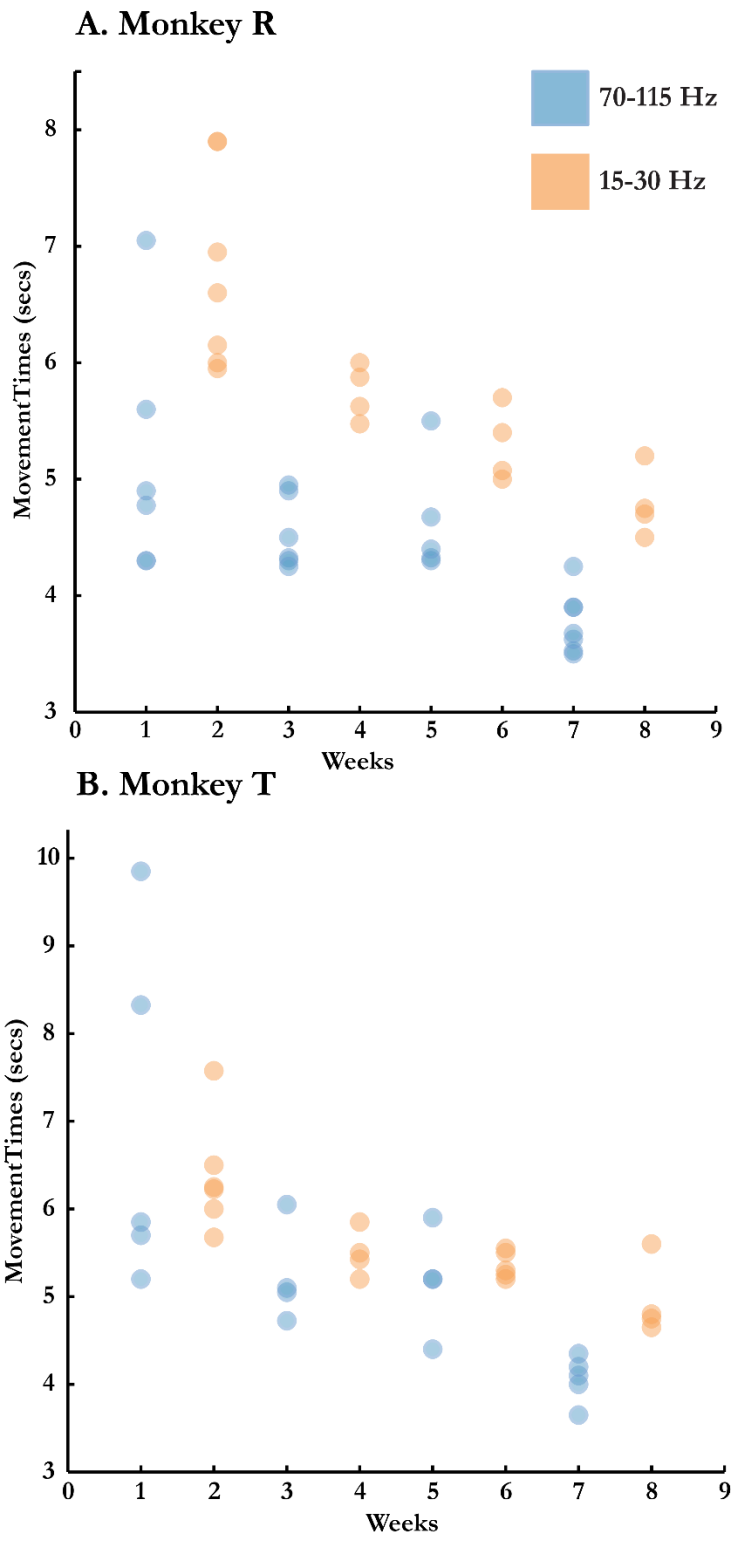

Figure 7.3: Movement times in the up-down path task A. Monkey R, B. Monkey T. The task was performed by two monkeys using all the features from either beta (Weeks 2, 4, 6, and 8) or the mid-gamma band (Weeks 1, 3, 5, and 7). A decrease in the movement times was observed with training in both the bands. However, mid-gamma band was observed to be faster than the beta band.

reliable in controlling the fast changes in the desired movement direction. Therefore, the results from the figure suggest that mid-gamma band signals are faster for BCI control than the beta band. In both the control bands, however, a plateau in the movement time was not reached. To determine the maximum rate of modulation of each band, further experiments need to be carried out for longer durations and with higher values of the gain term G_f.

7.2.2 Feature Modulations

The mid-gamma or beta band features recorded from all the active electrodes were used to control the 1D movement of the cursor. These features were weighted and linearly combined to form the velocity control signal. Though this combined signal was required to change directions from positive to negative, this may not have been the case for the individual signals. The subject could have modulated only some features to form the positive half of the control signals and others to form the negative half. The results presented in this section determine if the individual control features were also changing the modulation direction.

Starting from the initial center position of the cursor to the time it reached back to the center, each trial was divided into 50 equal time bins. The mean of the z-scored values of individual features was calculated during each of these time bins. The 50 point vectors obtained for every feature were then averaged across all the correct trials from one session of experiment.

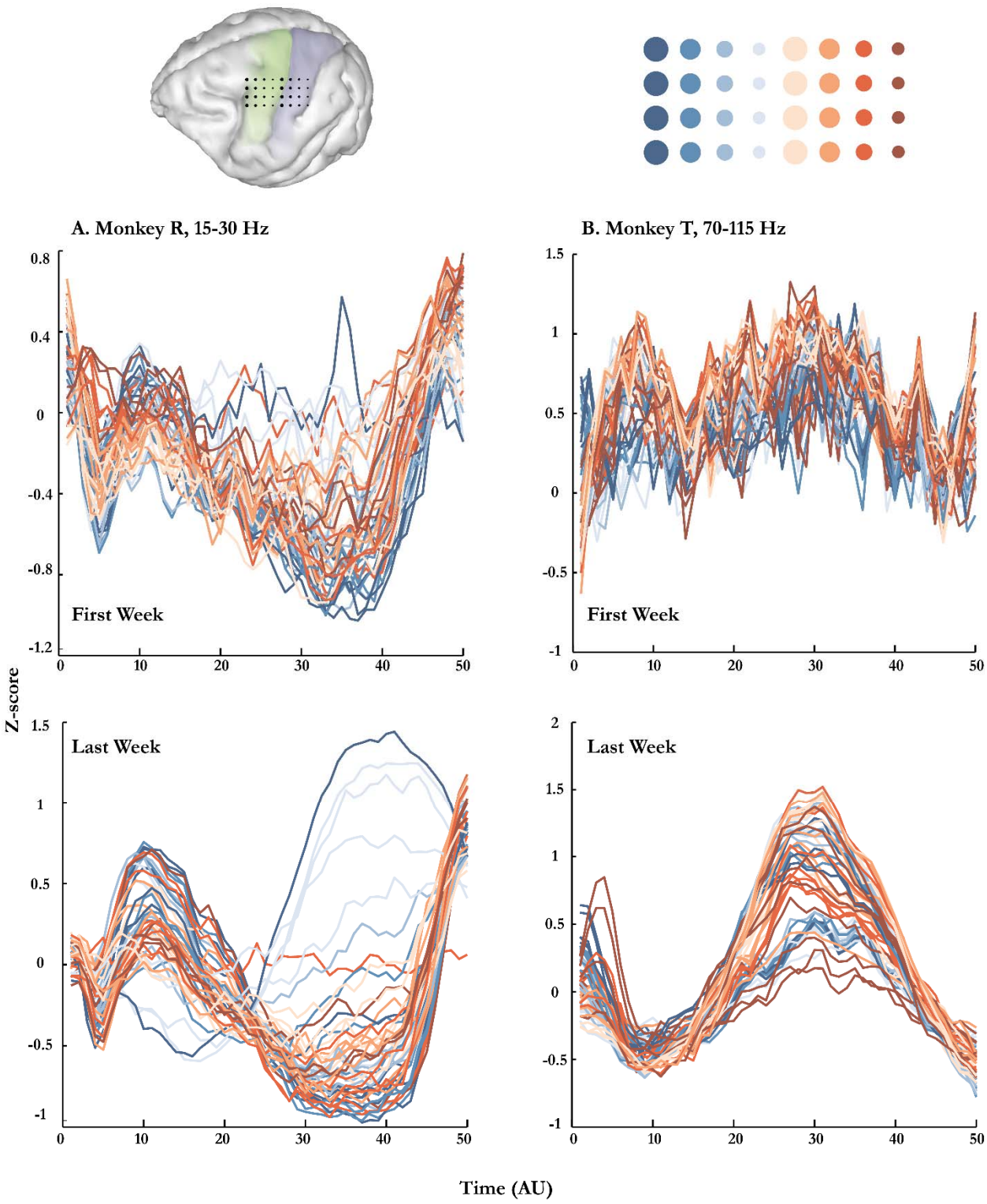

Figure 7.4: Average modulation of individual features during movement of the cursor. A. Monkey R, beta band, 15-30 Hz. B. Monkey T, mid-gamma band, 70-115 Hz. The top row plots were constructed using the data from one session of experiments in the first week. The bottom row used data collected during a session in the last week of experiments.

Figure 7.4 shows the plots of averaged mid-gamma and beta band features from two days of experiments. The top row shows the plots during a session at the beginning of the experiments and the bottom row shows the plots from a session in the last week of experiments, for the respective bands. The features are color coded to represent the row of recording electrodes as shown on the top of the figure. Rows with blue hues were approximately on the motor side of the sulcus and red were on the sensory side.

The plots clearly show that the individual features were very noisy and random during the beginning of the experiments. The plot from the last week, however, show a modulation of all the features around 0. This shows that even the individual features were changing the direction of modulation to control the cursor movement. Some features are observed to show a larger depth of modulation than others and some even show modulation that is opposite in direction. Though there is some grouping with respect to the rows of recording electrodes, no clear trend is visible.

To further study the changes in the feature modulations indicated by Figure 7.4, correlations were computed between individual features and cursor velocities. Similar to the features, recorded cursor positions during the trials were segmented into 50 time bins. The position vector was differentiated using a two-point slope method to calculate the velocity vector. Correlation between this velocity vector and individual features was calculated for every correct trial. Figure 7.5 shows the absolute correlation values averaged across the trials and control frequency band for each session of experiment. Blue dots represent mean

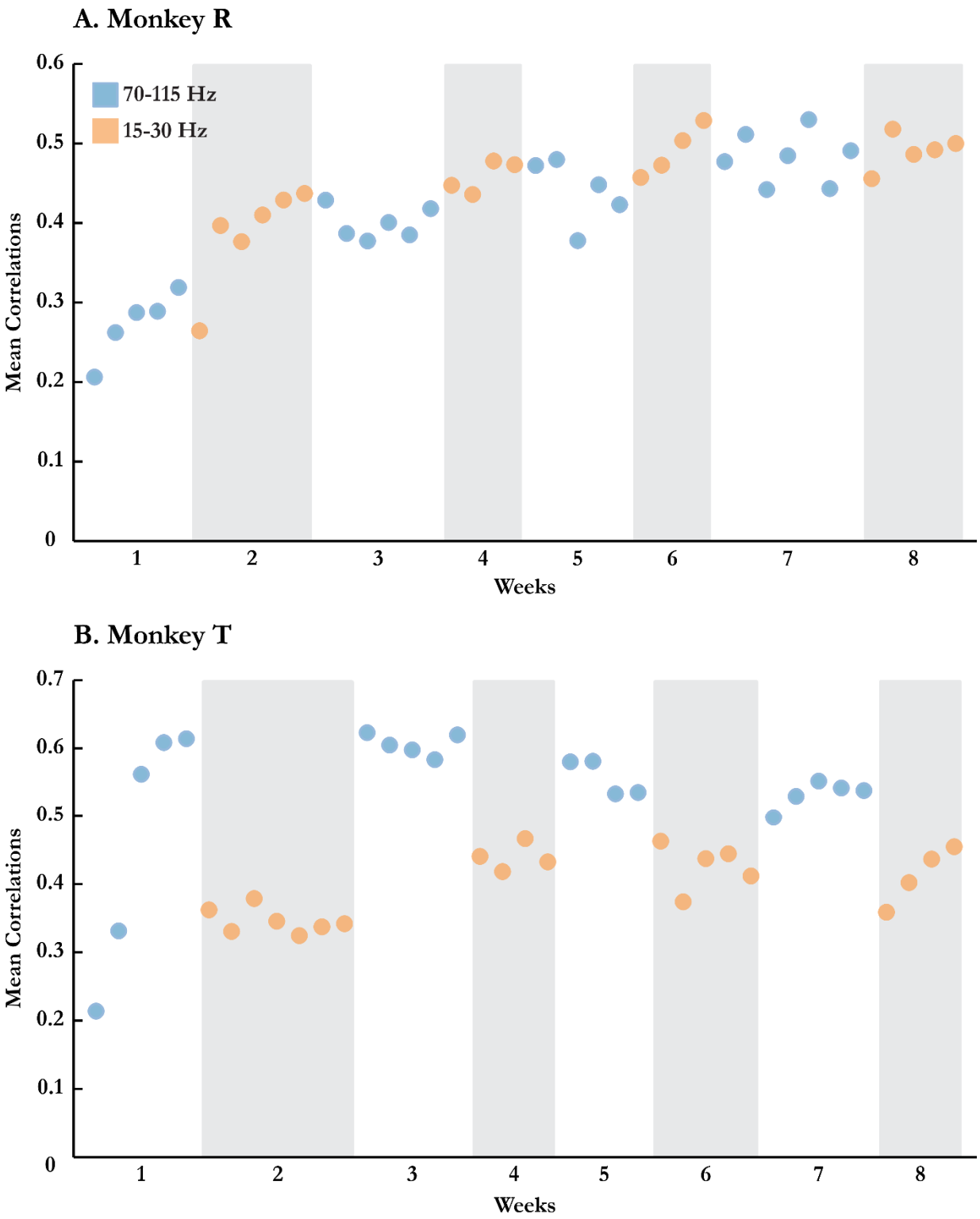

Figure 7.5: Mean correlations between cursor velocities and individual features from the control frequency band. A. Monkey R, B. Monkey T. Plots show the mean correlations of the control bands during 8 weeks of recordings. The even numbered weeks where beta band was used for control are shaded with gray background for clarity.

correlation for mid-gamma band and orange dots represent the values for beta band. Even numbered weeks, where beta band was used for control, are shaded with a gray background.

Monkey R is observed to show a gradual increase in the mean correlations with each week of experiment. The increase is steeper in the first 2-3 weeks and then comes to a relative plateau. It is interesting to note that a gradual increase in the means is seen in week 1 for the mid-gamma band features. On the other hand, in week 2, the correlation values see a relatively big increase going from first recording session to second and then don't change much after that. It is possible that the learning from week 1 using mid-gamma band was quickly incorporated into using beta band in week two.

Unlike monkey R, monkey T shows a steep increase in the mean correlation for mid-gamma band features in the first week of experiments. The values do not change much after the first week. Beta band in monkey T shows smaller values compared to mid-gamma band and show a slow increase with the weeks of experiments.

Only one band of ECoG signals out of five was used for control during each week of experiments. However, all five frequency bands were recorded. After studying the correlations for the control bands during the weeks they were used, similar mean correlations values were computed for all the five frequency bands, irrespective of which band was used for control. Figure 7.6 shows the variation of these values with each week of experiments. The even numbered weeks, where beta band was used for control, are marked with gray backgrounds for clarity of separation.

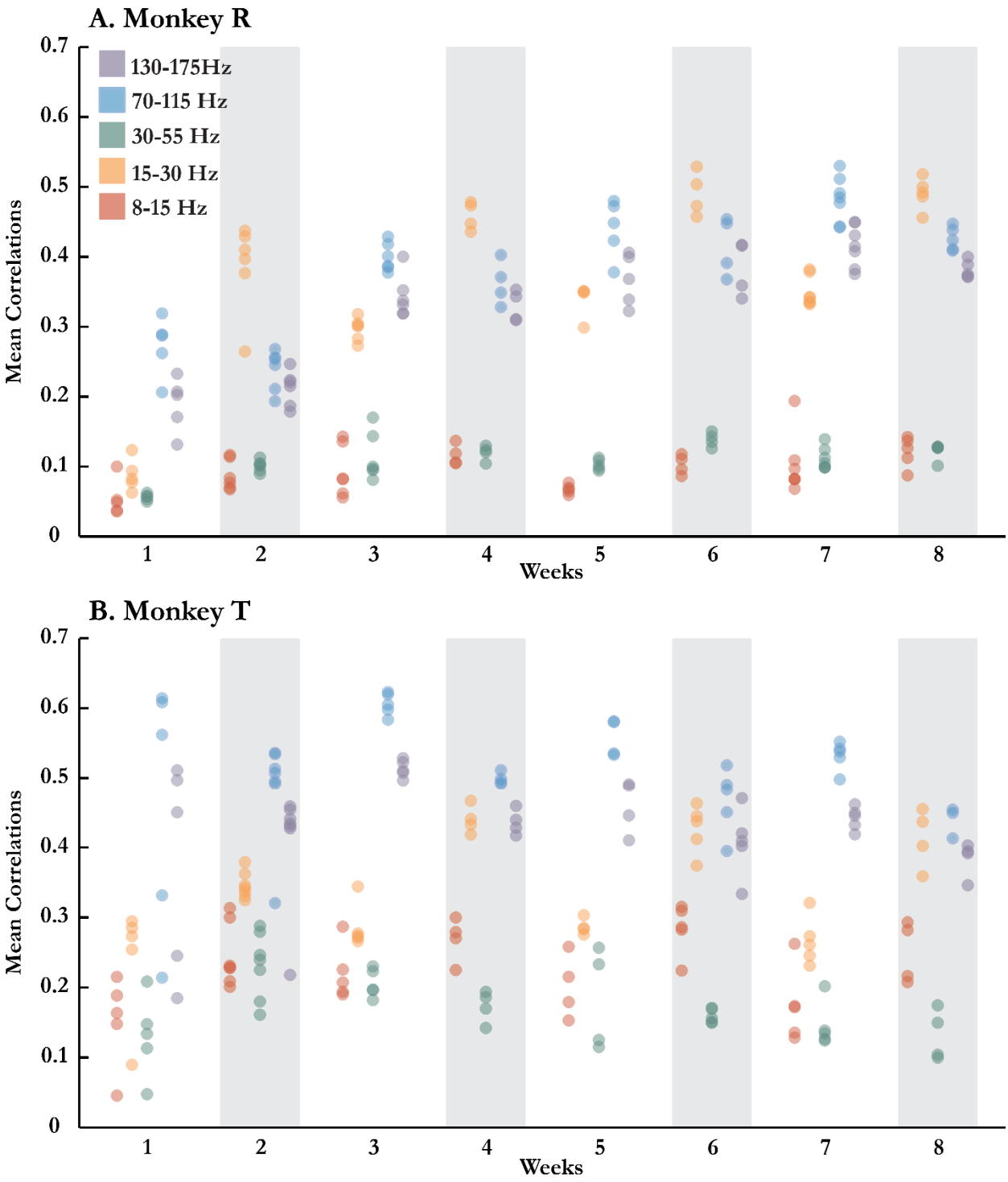

Figure 7.6: Mean correlations between cursor velocities and individual features from all five frequency bands. A. Monkey R, B. Monkey T. Plots show the mean correlations of all frequency bands during 8 weeks of recordings. The even numbered weeks where beta band was used for control are shaded with gray background for clarity.

A few interesting observations come forward from the Figure 7.6. In both the subjects, correlation values for alpha (8-15 Hz) and low-gamma (30-55 Hz) are always lower than other bands and close to zero. Comparing the correlation values of the two control bands, it is seen that the values are higher when the band is controlling the cursor movement, compared to when it's not. This is expected since the movement of the cursor and the control band share a causal relationship. However, lower correlation values during the week where a different frequency band was used for control, also shows that the modulation of ECoG signals had some local spectral differences. Finally, it is evident from the plots that the mean correlations for high-gamma band follow the increase and decrease in mid-gamma band correlation with a slightly lower mean for the distribution.

7.2.3 Tracking Task Performance

As has been stated earlier, a target tracking task was interlaced with the up-down path task during the 8 weeks of experiments. The frequency of the moving target was set for each session of recording and increased gradually every few sessions. To make the task easier, the radius of the moving target was set at 5 units and a time tolerance of about 20% was allowed. For instance, at 0.1 Hz target frequency where one sinusoidal cycle took 10 seconds, the subject had to be touching the target for at least a total of 8 seconds. The mean center to center distance between the cursor and target during the movement time was used to quantify the performance in the task (Figure 7.7). It is to be noted that with 3.5 screen units of cursor radius and 5 screen units of target radius the center to center distance could be up to 8.5 units for the cursors to be touching. Also, using time as the tolerance made it possible for the

subjects to be further away from the cursor on average. The subjects were expected to improve the performance in the task with time as full rewards were given for

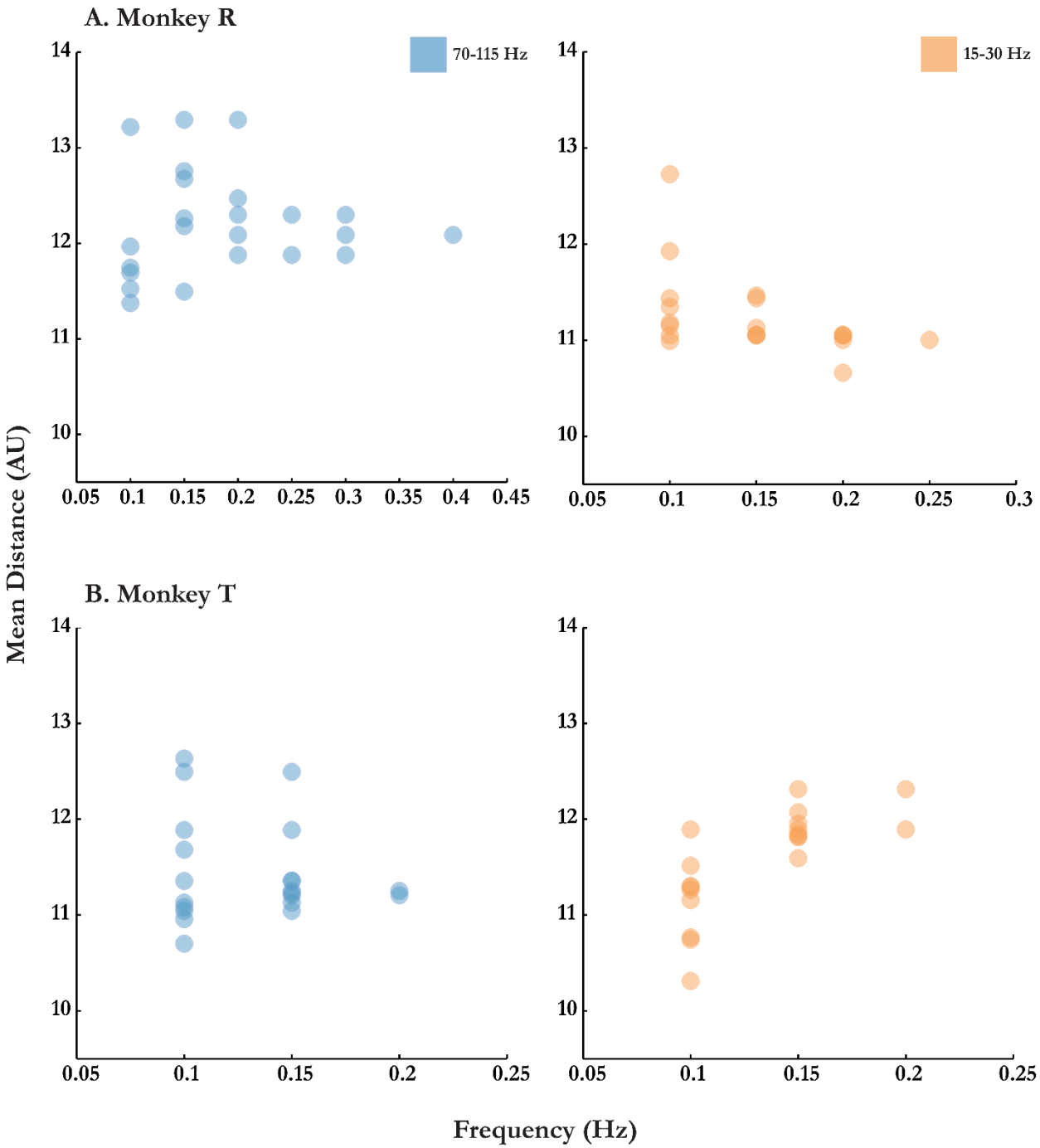

Figure 7.7: Mean distance between cursor and target during tracking task. A. Monkey R, B. Monkey T. Plots show the center to center distance between the cursor and moving target at different frequencies of the target. Each point represents the value calculated from one session of experiment.

completion of trials and other auxiliary rewards were given once every few seconds for touching the targets. However, the relative difficulty of the tracking task compared to the up-down task interlaced with it did not provide enough incentive for the subjects to improve their performance. Most of the blocks of tracking trials were observed to have large number of incorrect trials and a few of them that were marked correct due to the high tolerance that was set to make the trials easier. For the sake of completeness the mean distance metric for the two control bands at different target frequencies are plotted in Figure 7.7.

7.3 Conclusions

The results from the up-down task in this chapter reinforced a few observations made in the previous chapters. Figure 7.3 demonstrated that mid-gamma band features could achieve a higher modulation rate compared to the beta band features. As discussed earlier, this finding is important for the field of ECoG BCI because it emphasizes the importance of higher frequency bands in control. The maximum attainable rate of modulation for any signal is a marker of the information that can be carried by that signal. A signal that can modulate at a faster rate can encode independent information at a faster rate. The information rate of signals in BCI plays an important role as a higher value can result in better performance.

The faster modulation observed in higher frequency bands is by virtue of the underlying physiology that is responsible for its generation. Gamma band signals above 70 Hz are believed to be representation of the faster synchronized spiking activity of an ensemble of neurons [13], [72]–[74]. On the other hand, lower frequency potentials in the beta band are caused by the slower synaptic activity. Due to these signals forming the building blocks of

different spectrums of recorded signals, the difference in modulation rates is expected. It is to be noted, however, that a more elaborate study involving variation in gain parameters and longer recording periods is necessary to decipher the upper limit of the modulation rate in the different bands.

The two other important observations from correlation plots (Figure 7.6) are a) Low correlations demonstrated by alpha and low-gamma bands, b) Similarity of correlations between mid-gamma and high-gamma bands. These two results have also been previously seen in Chapters 4, 5 and 6. The recurrence of these observations from different experiments suggests that they are significant and should be incorporated into the initial steps of recording ECoG signals and separating them into control features.

Chapter 8
2D BCI using Four ECoG Electrodes

For ECoG to become a clinically accepted BCI modality, it has to overcome the hurdles that make it feasible only in a research environment. The laboratory system (TDT system) used for all the experiments described in this dissertation uses a head-stage/pre-amp assembly to record raw ECoG signals. This assembly uses about 10-20 mWatts of power per channel. Significantly more power is used by the DSP base stations for further processing of the signals. A recording system that uses low power in conjunction with a BCI setup that works with a small number of signals is required to translate ECoG to clinical BCI applications.

Chapters 4, 5 and 6 of this dissertation have discussed decoding and feature selection algorithms to obtain robust control with a small subset of features. These chapters also demonstrated that parts of ECoG spectrum are more important for BCI control than others. This chapter describes a proof of concept study that uses a simple decoding algorithm and signals recorded from only four ECoG electrodes to obtain 2D BCI control. The study further goes on to use a low power recording system developed by Medtronic to conduct the same experiments. The goal of the experiments was to demonstrate that it is possible to use ECoG signals, recorded from a system feasible for long term implants, for clinical application of BCI.

This chapter is organized in three different section. Section 8.1 describes the general methods used across the two recording systems and provides some specifics of the low power

recording system. Section 8.2 describes the results from the two systems and compares them with each other. Finally, Section 8.3 concludes the chapter with a discussion of significance of this study.

8.1 Methods

8.1.1 Subjects and Recording Electrodes

Two male Rhesus Macaques, (*Macaca mulatta*, monkey G, and I) weighing 6-10 kilograms were used as subjects for the experiments in this chapter. Both the monkeys had over two years of BCI training experience before this study. However, they had performed BCI tasks using only 4 electrodes for only 2-3 days before the experiments in this chapter. Both the monkeys were implanted with Silastic arrays manufactured by PMT Corporation (Section 3.2.2, Chapter 3). Monkey G had a subdural implant whereas, monkey I had an epidural implant. The four electrodes selected for each monkey had 2mm exposed electrode diameters and came from a single implanted array. All the electrodes were referenced to a skull facing electrode. Selection of the recording electrodes was made through a visual inspection of weights assigned to the features originating from previous experiments.

8.1.2 ECoG Signal Processing

The raw ECoG signal recorded from each of the electrodes was separated into five different frequency bands. The 20 electrode/frequency band pairs were used as "control features" by the decoding algorithm to control the movement of the cursor. The five frequency bands used during the experiments with the TDT system were alpha (8-15 Hz), beta

(15-30 Hz), low-gamma (30-55 Hz), mid-gamma (70-115 Hz), and high-gamma (130-175 Hz). The details of band separation and processing have been described earlier in Section 3.4 of Chapter 3.

The low power recording system from Medtronic was designed record four raw ECoG signals and compute their spectral power. The signals were recorded at a sampling rate of 500 Hz. A 256 point FFT was used to extract the spectral estimate of five different bands. The five bands programmed into the Medtronic system were closely matched to the ones used with the TDT system. Details of these bands are shown in Table 8.1. All the recording and processing was carried out on an application-specific integrated circuit (ASIC) and used about 10μWatts of power per recording channel, which is three orders of magnitude lower than the laboratory recording system.

Table 8.1: Frequency bands computed using 256 point FFT in the low-power system

Band	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\boldsymbol{\gamma_{low}}$	$\boldsymbol{\gamma_{mid}}$	$\boldsymbol{\gamma_{high}}$
Center Frequency (Hz)	11	22	42	95	152
Bandwidth (Hz)	8	16	24	40	44
Pass Band (Hz)	7-15	14-30	30-54	75-115	130-174

The Medtronic system used a serial port interface to communicate with the computer controlling the virtual environment of the BCI task. Once the spectral estimates from the system were received by this computer, they were processed exactly the same way as the signals from the TDT system were. This involved a log transform and a computation of a real-time z-score of each feature. This too has been described in Section 3.4 of Chapter 4.

8.1.3 Behavioral Task and Experimental Protocol

A classic center-out task with eight targets was used for the experiments in this chapter. The targets were placed on the circumference of a circle of radius 15 screen units, spaced at 45° starting at 0°. The target and cursor were both spheres with a radius of 2.5 screen units. Each trial started with the cursor at the center of the circle. After a 500 ms initial hold period, one of the targets was displayed. The subjects had to move the cursor from the center to the displayed target in a specified maximum movement time. If successful in doing so, the trial was marked correct and a small water reward was presented. In any other case the trial was marked incorrect.

Similar to the previous experiments, each session was divided into blocks of trials and started with a naïve decoder and a watch-task block. A constant speed assistive bias was used during the first 5-7 blocks to help the subjects to get started. In this 5-7 blocks the bias gain was gradually reduced and the cursor control was completely transferred to the subjects.

As this was a proof of concept study, only four days of experiments were carried out to test BCI using 4 electrodes and compare the two recording systems. The low-power Medtronic system was used during Day 1 and 2. TDT system was used for recording on the following two days (3 and 4).

8.1.4 Control Signal and Decoding Algorithm

As in previous experiments, subjects controlled he velocity of the cursor in the center-out task. The velocity signal was integrated using Euler Integration to give the position of the

cursor on the screen. A simple decoding algorithm that used the modulation of features during the movement of the cursor and the desired movement direction as the input, was designed to assign weights to the 20 features. Weights were updated after every block of trials and blended using a smoothing factor to avoid any overfitting. The details of the control signal and the decoding algorithm have been described in section 4.1.4 and 4.1.5 of Chapter 4.

8.2 Results

8.2.1 Distribution of Control Features

Each of the 20 control features were assigned weights in the two control dimensions. Figure 8.1 shows the mean weight vectors compute from weights assigned during a session of recording, using the TDT system. As there were only 4 recording electrodes, individual features from each electrode are shown on a polar plot. The weight vectors are color coded by the recording band.

As seen from the plot, mid-gamma and high-gamma bands are observed to be assigned the highest weights. These high-frequency features from different electrodes are also seen to be distributed in different directions to cover the 2D space. It is interesting to see that the two frequency bands recorded from the same electrode share similar preferred directions with some variation in the magnitude. The features from the three other bands are mostly small with only some of them showing larger weight magnitudes.

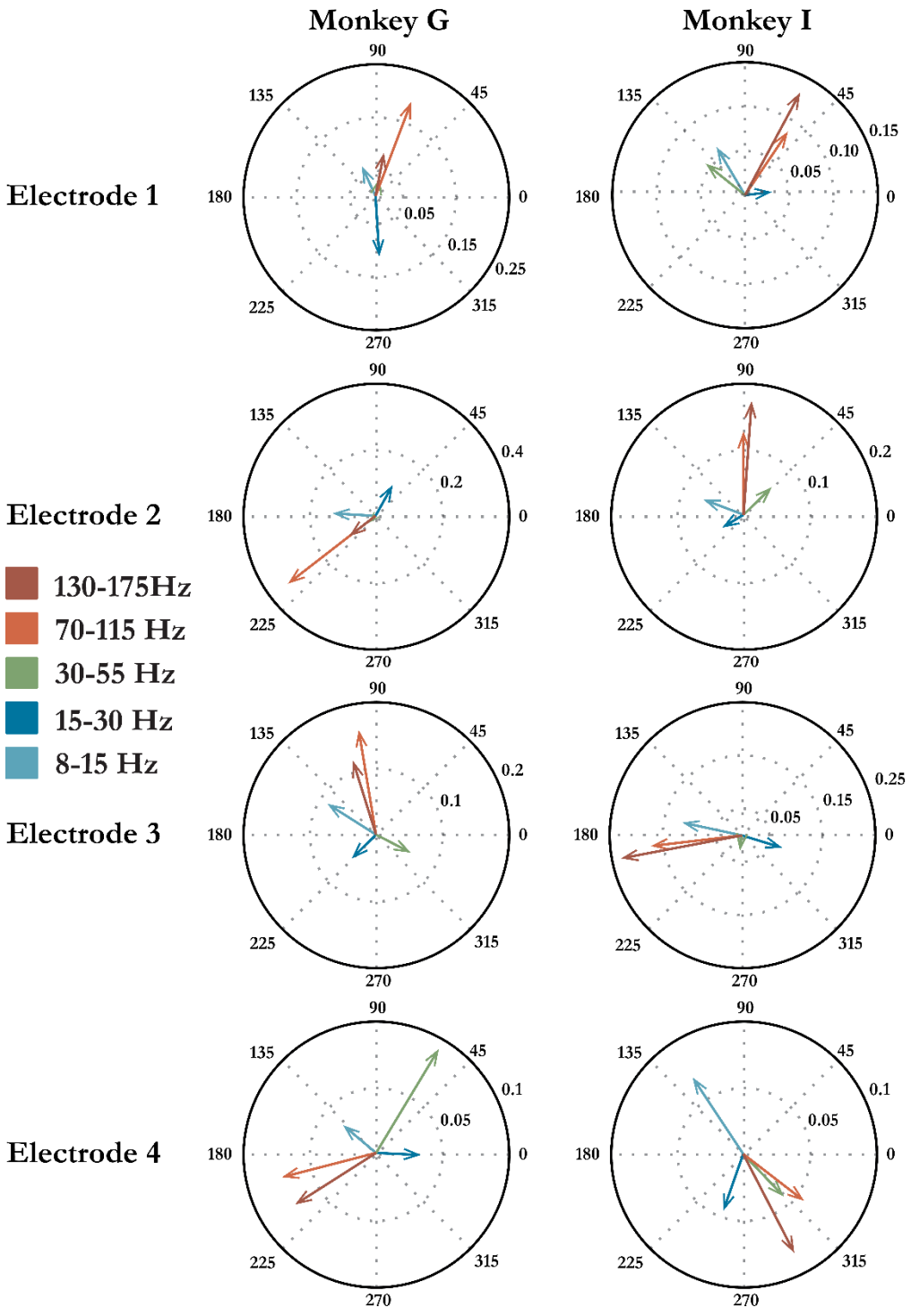

Figure 8.1: Distribution of feature weights in 2D space. Left column-Monkey G, Right column-Monkey I. An average weight vector was calculated from the weights assigned to each band recorded from the 4 electrodes, during a session of experiment. Each vector in the polar plots above is color coded with the features frequency band. By looking at the average vectors it is seen that monkeys preferred the mid and high gamma bands. The distribution of weights also shows a high degree of similarity between the weights assigned to these two bands.

8.2.2 Performance Comparison

The BCI performance while using the two systems was quantified by the percentage of correct trials during each session and the average movement time to the individual targets. Each subject performed over 800 correct reaches (100 to each target) on all the days of experiments using Medtronic and TDT systems. This is close to the average number of reaches seen on any day of experiment. Table 8.2 gives the percentage of correct trials on each day for the two subjects. The percentages were calculated for reaches performed under complete brain control in the absence of assistive bias. The number in the brackets represents the day on which a particular experiment was carried out relative to the beginning of the study.

Table 8.2: Percentage of correct trials during different sessions of experiment using TDT and Medtronic recording systems.

Monkey	Medtronic		TDT	
	Session 1 (1)	Session 2 (2)	Session 1 (3)	Session 2 (4)
G	80.04	98.60	96.45	98.41
I	90.54	93.14	92.02	97.40

Figure 8.2 shows the average movement time to each target on Day 2 (Medtronic, in red) and Day 3 (TDT, in blue) of experiment, on a polar plot. Even though the monkeys had a maximum of 10 seconds to reach the targets, the movement times were observed to be less than 3 seconds on an average. It is interesting to see that the monkeys show similar target preferences across the two recording systems in both the monkeys. This indicates that similar feature modulation was used in each case. This is investigated further using concepts from signal detection theory.

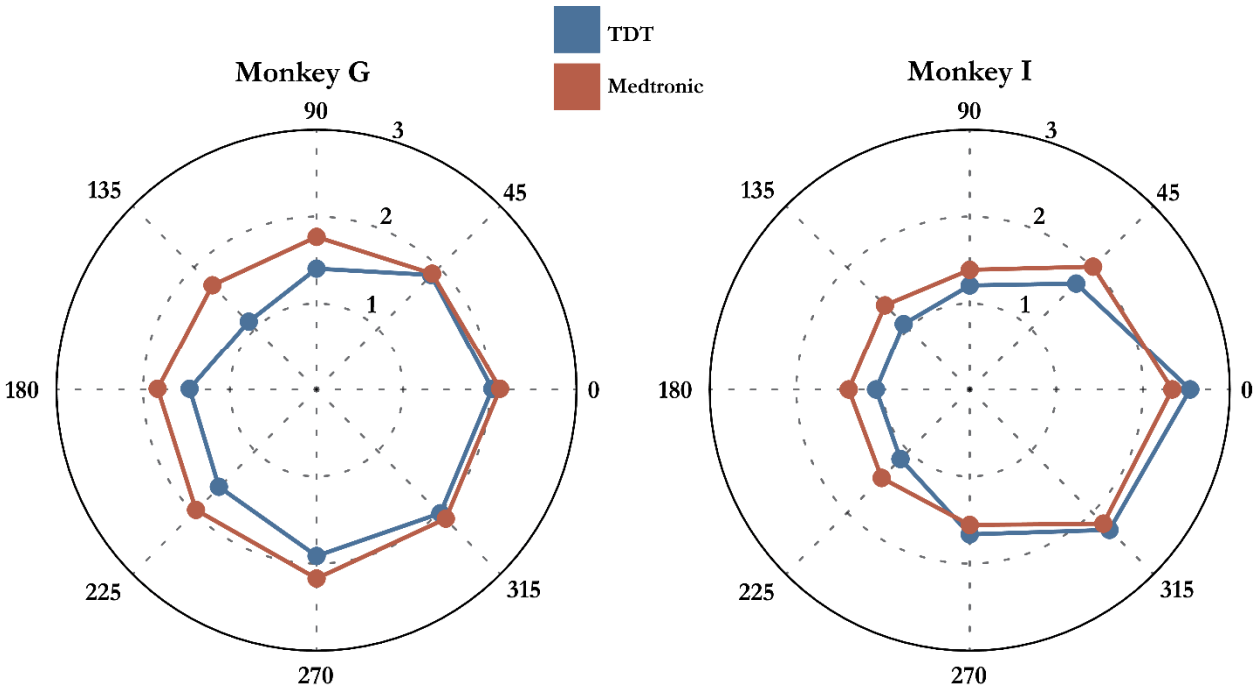

Figure 8.2: Average movement times for each target. Left-Monkey G, Right-Monkey I. Average movement times for each target during the BCI task are shown on a polar plot, for each monkey. Even though each monkey had a maximum of 10 secs to get to the target all the correct reaches on an average were less than 3 seconds long.

8.2.3 Receiver Operating Characteristic (ROC) Analysis of Feature Modulation

The average weight vectors calculated for each feature can be affectively quantified as that features' preferred direction (PD), similar to the ones calculated in intra cortical population vector studies [5], [53]. The target closest to this direction was assigned as its preferred target.

To test the discriminability of the signal recorded by the system, a ROC analysis was conducted between the preferred target and the target diametrically opposite from it. A root mean squared (RMS) value of the feature amplitudes was calculated during each correct reach

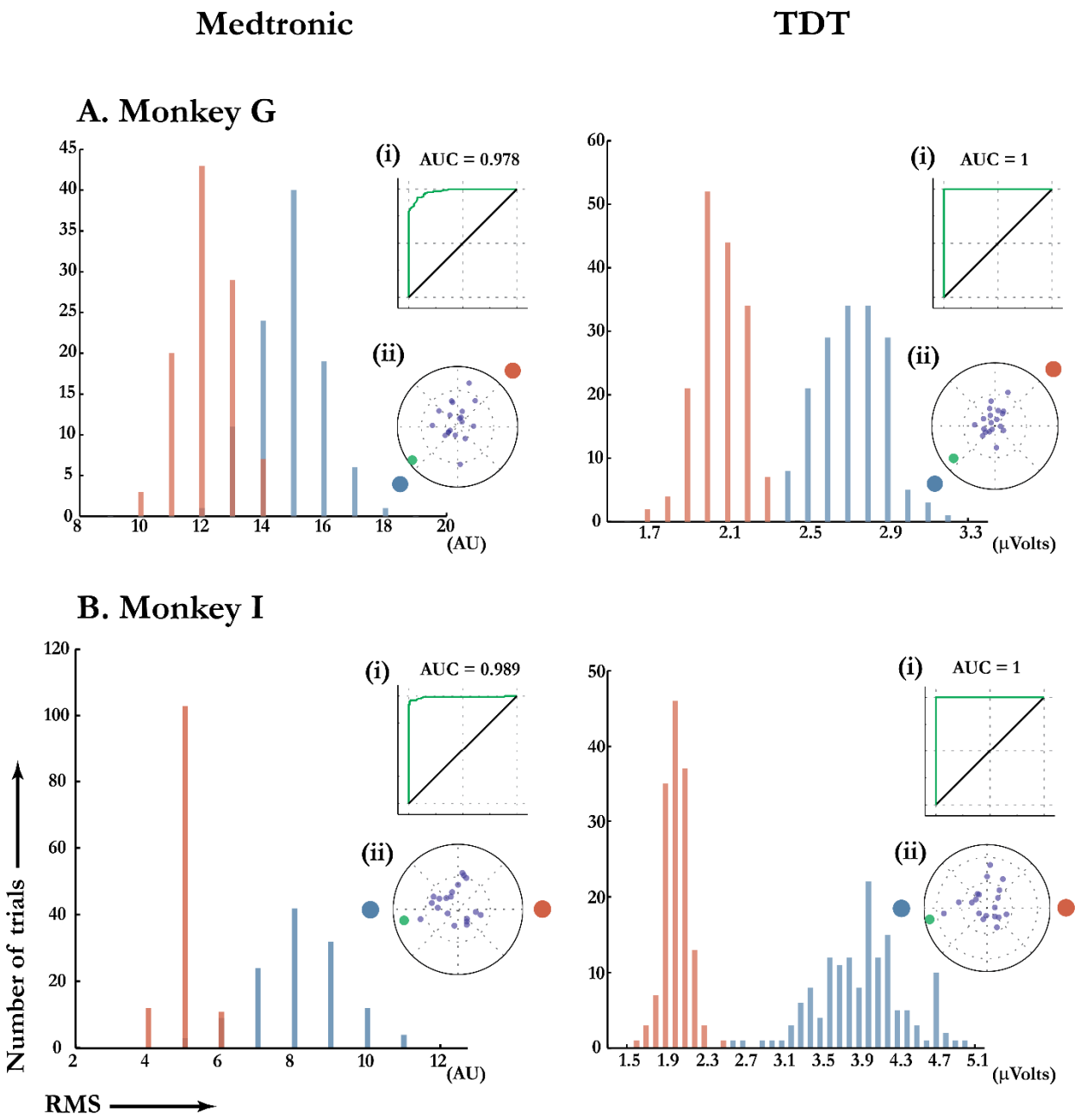

Figure 8.3: ROC analysis of feature modulation. Left column- Medtronic system, Right column-TDT system. Top row-Monkey G, Bottom row-Monkey I. Each figure shows the distributions of RMS values of the feature for two diametrically opposite targets during the movement time of trial. Each plot has two insets. Inset (i) is the ROC curve for the distributions with the AUC on top. Inset (ii) shows the distribution of features in 2d space (similar to figure 2) with the feature used to create the ROC curve marked in green. The *For* and *Against* targets are marked in blue and red, respectively, with their corresponding distributions using the same colors. Same feature and target pairs were used while making these plots for two recording systems in the same monkey.

towards the preferred and the opposite target. Distributions created by these RMS values were used for ROC analysis.

Figure 8.3 shows the distributions of RMS values for the same feature recorded during TDT and Medtronic sessions. Each panel has two insets; (i) ROC curve for the distributions shown in the main figure and (ii) distribution of mean weights vectors in a polar plain with the feature being studies marked in green. Inset (ii) also shows the preferred (blue) and opposite (red) targets used to create the distributions. The area under the ROC curve (AUC) gives a measure of discriminability between the two distributions. It represents the probability that the RMS value of a randomly selected preferred target trial will be greater than the RMS value of a randomly selected opposite target trial. The AUC value for each ROC curve is also shown in inset (i) of each panel. The high values of AUC in all the panels represented in Figure 8.3 indicates high discriminability in the feature during movement towards the two targets. The similarity between the feature distributions is notable.

To further understand the modulation of features resulting in the movement of cursor in 2D space, the ROC analysis was expanded to all the eight targets. For every feature, a ROC curve was constructed for each target assigned as the *For* and the target diametrically opposite to it assigned as the *Against* target. Theoretically, for any given feature, the RMS signal distributions created using the preferred target and the target diametrically opposite should be most discriminable. The high values of AUC in Figure 8.3 serves as proof for that hypothesis. Extending that notion, it can be said that the discriminability of the distributions should

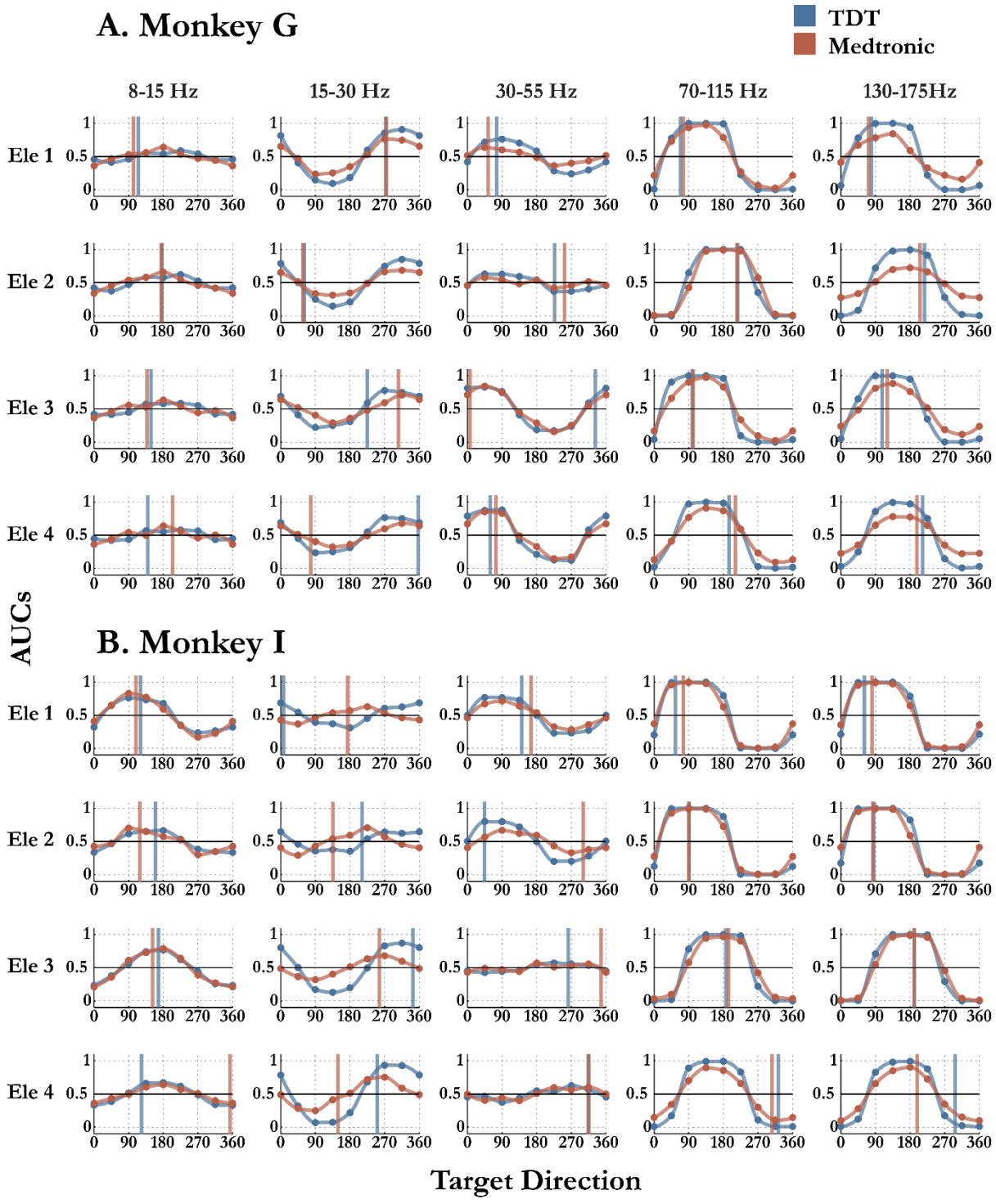

Figure 8.4: Variation of AUC with *For* target direction. A. Monkey G, B. Monkey I. Extending the ROC analysis to each pair of targets, AUC was plotted as a function of *For* target direction. A cubic interpolate was used to connect values. Each row in the figure represents a recording electrode and each column the frequency band. Preferred directions are observed to be aligned to the pair with highest AUC value. AUC curves across recoding system are aligned to each other, indicating use of same control strategy in both cases.

gradually decrease as the *For* target moves away from the features' preferred target. A plot of AUC as a function of *For* target direction, for each feature, was created to test this hypothesis.

Figure 8.4 compares the AUC plots for all the features computed from data collected using the TDT system (blue) and Medtronic system (red). The first noticeable characteristic of the plot is the AUC values being less than 0.5. As mentioned earlier, ROC curves were created for each target as the *For* target. A value of less than 0.5 for the AUC thus signifies that the particular *For* target had lower distribution values than the *Against* target. Also, this is equivalent to calculating the AUC value for the curve created using a flipped assignment of targets and then subtracting it from 1.

The plots in Figure 8.4 also qualitatively demonstrate the modulation of features in the 2D space. The vertical lines representing the preferred directions of the features, are aligned with the highest AUCs. These values stay fairly close to 1 for a couple of targets around the preferred target and then gradually decrease as the targets moves away from this direction. This confirms the earlier hypothesis that the discriminability of distribution worsens as we move away from the preferred direction. The plots also show similarities in the modulation of mid-gamma and high-gamma bands, as was observed from the weight assignments in Figure 8.1. A few features from alpha, beta or low-gamma bands showed discriminability between targets and some changes in the AUC values. However, none of these features had AUC values close to 1.

8.3 Conclusions

The results from this short study demonstrated that ECoG signals recorded from only four electrodes can be used to reliable control a 2D BCI. The experiments used two subjects that had negligible experience in using such small number of features. Both the subjects were able to efficiently control the cursor within the first recording session and were able to perform like during any other recording session using several recording electrodes. In addition to using only four electrodes, the experiments also demonstrated the use of a low power system to record the signals. Comparison of performance between the two systems showed no significant difference.

The results presented in figures 8.3 and 8.4 depict the robustness of the low-power recording system designed by Medtronic. The AUC variation curves created using the data from the Medtronic system demonstrate that it was able to record the small changes in the signals that led to the accurate BCI control of the cursor in the 2D space. A comparison of these curves between the two recording systems shows that the discriminability was practically equivalent for both. The alignment of the curves to each other and also of the preferred directions shows that the monkeys were using the same strategy for BCI control using TDT or Medtronic systems. This is expected as the data was recorded on consecutive days. However, the signal-to-noise in the recorded signals plays a huge role in how the monkeys are able to use them for control. Seeing similar features weights and modulations while using the two systems shows that the low power recording system also had low noise characteristics

similar to the laboratory recording system that used 3 orders of magnitude more power to record the same signals.

In conclusion, the observations from this comparison study provide a strong basis for ECoG signals as a feasible modality for clinical BCI. A system that consumes power in the range of a few tens of micro watts is feasible for a chronic implant of several years. Combining that with the stability of ECoG signals, small number of recording electrodes, and the ability of simple algorithms to obtain robust BCI control, gives rise to an optimal solution to overcome the hurdles preventing the patient population from getting the much required help.

Chapter 9
Conclusions and Future Directions

9.1 Feature Selection Algorithms

9.1.1 Summarization and Interpretation of Results

Chapters 4 and 6 described and used different feature selection algorithms to obtain a sparse control feature set. Through post hoc analysis it was also determined that gamma frequency bands above 70 Hz were most significant in ECoG BCI. Instances where the recording signals were relatively noisy, the beta band was found to be the next best contender. This was a secondary effect from the distribution of spectral power of ECoG signals. The power spectrum of ECoG signals (or any field potential recordings) falls off at a rate proportional to $1/f$, where f is the frequency. Therefore, broadband white noise in the system effects the higher frequency signals more since they have lower baseline power (*e.g.* beta band ~ 15 μVolts RMS vs mid-gamma band ~ 2 μVolts RMS). This was likely the primary reason a difference in the selected features were observed between monkeys K and P in Chapter 4.

From the performance standpoint, experiments in Chapter 4 showed highly promising results, with little changes in movement time and no observable changes in the trajectory deviations. In Chapter 6, though faster cursor speeds were observed, they were also linked to an increase in the trajectory deviations. The variation in the gain term with the number of

features selected has been provided as one possible explanation for this observation. It is also important to point out the other drawbacks of the decoding algorithm used in this chapter and how they can be overcome.

Firstly, the decoding model in Chapter 6 was built and updated after every block (32 trials), without any memory from the previous blocks. Secondly, as the value of λ was increased, the tradeoff between squared loss and the regularization term was slowly tilted in the direction of the latter. That is, the algorithm was more inclined towards producing a sparse output than reducing the least squared error. It is likely that due to these two reasons, an almost linear increase in the trajectory deviations was observed. The solution to this, however, is relatively straight forward. Once a sparse subset of control features is determined, it is important to use that subset of features and only squared loss function to build models for the subsequent session of experiment. It is also important to use a smoothing factor to blend weights from one update to another. Using only a squared loss function on the obtained feature set would lead to a better fit between the data and features. Blending the weights across updates will prevent local overfitting due to any abrupt changes during a block of experiments. These advantages have already been demonstrated in Chapter 4 and can also be employed to a sparse feature set obtain from the decoder in Chapter 6.

9.1.2 Future Directions

In the future, the algorithms described in chapters 4 and 6 can be expanded to control higher degrees of freedom. To incorporate a third control dimension in the pruning algorithm described in Chapter 4, the 3D space can be segmented into 8 equal volumes covered by each

octant. By setting and varying the threshold on features spanning each segment a reduced control feature set can be determined, without significantly affecting the overall distribution of the features. Beyond three-dimensional control, the segmentation of multidimensional space becomes highly dependent on the type of control dimension (for example controlling 1D rotation versus opening and closing for grasp) and requires further investigation when implemented. The regularized least square algorithm in Chapter 6 produced a single sparse feature set across the two dimensions and, is much easier to expand to multidimensional space. This can be accomplished just by incorporating the weights from all dimensions in ℓ_2-norm part of the ℓ_1/ℓ_2-norm constraint.

The algorithm in Chapter 6 can also be modified to select a subset of recording electrodes for control instead of just selecting a subset of features that could span across all the active electrodes. To select a subset of control electrodes for a single dimension, a vector can be formed by computing ℓ_2-norms of the weights assigned to each frequency band from a single recording electrode. The ℓ_1-norm of this vector can then be used as a constraint on the squared loss to use all frequency bands but from a sparse set of electrodes. This method has been previously used with human ECoG signals for differentiating between wakeful and sleep like states [75]. The obvious problem with a method like this would be that a sparse set of electrodes will need to be determined for each control dimension. This could again result in a selection of different control electrodes for different degrees of freedom. To overcome this problem the new constraint has to be either incorporated into the regularization term used earlier or a second regularization term needs to be added to the loss function. In both these

cases, optimization of the algorithm would be more complicated and would need further investigation of machine learning literature.

9.2 Spectral Modulations of ECoG Signals

As mentioned briefly in the previous section, the higher gamma band spectrum in ECoG recordings was found to be a significant contributor towards BCI control. This is demonstrated by the different frequency band distribution figures from Chapters 4, 5, 6, and 8 (Figure 4.5, 5.6, 6.5, 6.6, and 8.1). It was also observed from a couple of these figures that beta band features were used in recordings with higher noise (Figure 4.5) or when the higher gamma band features were scarce (Figure 5.6). The modulation of the two frequency bands (mid-gamma and beta) was also compared against each other to show that mid-gamma features were able to change the direction of modulation at a faster rate (Figure 7.3). The post hoc analysis of the data collected in Chapter 7 and the distributions of frequency bands seen in Chapters 4, 5, 6, and 8 all show a correlation between mid-gamma and high-gamma band features. Thus, on a whole, the same pattern of spectral modulations has manifested itself during different experimental recordings, providing a strong evidence of an inherent characteristic of ECoG signals.

With similar observations from all the experiments, one observation in the data collected for Chapter 4 was unexpected and novel. In Figure 4.9, higher R^2 values were observed for beta band features originating from sensory electrodes compared to those from motor electrodes. On the other hand, motor features were observed to have higher R^2 values in mid-gamma and high-gamma band. This is an interesting result that has not been observed

or investigated before with field potential recordings. Lower R^2 values of sensory features are expected as the BCI task provides an alternate path to produce movement, which is the primary function of motor cortex and not sensory. Both higher gamma bands and beta band spectrum has previously been observed to show modulations with overt and imagined movement. For these two reasons the higher R^2 values of beta band sensory features are surprising. Thankfully, ECoG BCI can be used as an experimental tool for an in-depth investigation of this phenomenon. Feature selection algorithms used in the different chapters of this dissertation can all be used with only beta band features to determine if there are any significant differences between motor and sensory cortices and if these differences reinforce the results from Figure 4.9. The 1D UP-Down path could also be used to compare the modulations of motor and sensory beta band features.

The target tracking task designed to study the differences in spectral modulations of ECoG signals was the one part of this dissertation that could not be successfully implemented. The two possible reasons were the difficulty of accurately tracking a moving cursor and the interleaving of the task with an easier, high reward Up-Down task. The parameters of the tracking task were very restrictive as the subject had to match the frequency set by the experimenter. Also, interleaving the task with the Up-Down task to construct a decoding model lead to the subjects putting just enough efforts to finish the tracking task block and move on to the Up-Down task block. To properly investigate if monkeys can be trained to do a BCI task like this, it is probably important to carry it out independently. A decoding model developed in advance from a 1D BCI task like the Up-Down task or just a simple two target

center-out task can be used before exclusively using the tracking task. This way, the subjects would have an incentive to improve their performance.

The Up-Down path task, in this dissertation, was used to study modulation of a single frequency band. However, the task has a potential to be expanded to a 2D tracing task that can be used to compare or even enforce different modulation patterns in the control features. For instance, individual ECoG signals have previously been used in a Circle Drawing task to analyze the ability of a subject to independently modulate the signals [63]. Similarly, using different frequency bands from the ECoG spectrum to independently control the two dimensions, subjects can be trained to trace different paths from the family of Lissajous curves that demand differences in modulation phase and frequency of the contributing signals. These type of curve tracing tasks can also be used in the future to train subjects in using prosthetic devices for naturalistic smooth movements compared to the more frequently used, ballistic movements.

9.3 Final Thoughts on ECoG BCI

In Chapter 2 of the dissertation, it was described how epidural ECoG provides an ideal balance between signal fidelity and risk of infection. In all the experiments carried out for this dissertation, the subjects had been implanted for at least a year with most experiments performed over two years of the implant. Some subjects were used for multiple experiments that were separated by several months. During this duration, all the recordings from different subjects stayed relatively stable and the issues encountered with any subject were mainly due to the wear and tear of the external connectors or in some instances subcutaneous infections

caused to the acrylic head-caps. One example of the stability of recording in a subject can be seen by comparing the distribution of features in monkey I from figures 5.5 A. and 8.1 (left column). Few of the large features seen pointing in three directions in Figure 5.5 A are the same as the mid-gamma and high-gamma features of monkey I seen in Figure 8.1. The two figures were created from data collected one year part and three (Figure 5.5) or four (Figure 8.1) years after the implant.

Beyond the stability of the recordings, different experiments in this dissertation demonstrated that a small set of ECoG signals can be used to control a BCI. The experiments also demonstrated that basically two frequency bands: gamma (>70 Hz) and beta are important for BCI applications. Finally, the study in Chapter 8 showed that a low power recording system, feasible for a long term implant, is just as good in recording and processing ECoG signals as the laboratory recording system. With all these observations, this dissertation has provided a strong base for wider adoption of ECoG signals in the field of BCI and its translation into clinical studies with the patient population.

References

[1] "Amyotrophic Lateral Sclerosis (ALS) Fact Sheet: National Institute of Neurological Disorders and Stroke (NINDS)." [Online]. Available: http://www.ninds.nih.gov/disorders/amyotrophiclateralsclerosis/detail_ALS.htm. [Accessed: 25-Sep-2015].

[2] "Spinal Cord Injury (SCI) Facts and Figures at a Glance." [Online]. Available: https://www.nscisc.uab.edu/Public/Facts%202015%20Aug.pdf. [Accessed: 25-Sep-2015].

[3] E. Kandel, J. Schwartz, and T. Jessell, *Principles of Neural Science, Fourth Edition*. McGraw-Hill Companies,Incorporated, 2000.

[4] D. Purves, *Neuroscience*, 5th ed. Sunderland, Mass: Sinauer Associates, 2012.

[5] A. P. Georgopoulos, J. F. Kalaska, R. Caminiti, and J. T. Massey, "On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex," *J. Neurosci.*, vol. 2, no. 11, pp. 1527–1537, Nov. 1982.

[6] A. P. Georgopoulos, A. B. Schwartz, and R. E. Kettner, "Neuronal population coding of movement direction," *Science*, vol. 233, no. 4771, pp. 1416–9, 1986.

[7] A. B. Schwartz, "Motor cortical activity during drawing movements: population representation during sinusoid tracing," *J. Neurophysiol.*, vol. 70, no. 1, pp. 28–36, Jul. 1993.

[8] D. W. Moran and A. B. Schwartz, "Motor cortical representation of speed and direction during reaching," *J. Neurophysiol.*, vol. 82, no. 5, pp. 2676–2692, Nov. 1999.

[9] D. W. Moran and A. B. Schwartz, "Motor cortical activity during drawing movements: population representation during spiral tracing," *J. Neurophysiol.*, vol. 82, no. 5, p. 2693, 1999.

[10] G. A. Reina, D. W. Moran, and A. B. Schwartz, "On the relationship between joint angular velocity and motor cortical discharge during reaching," *J. Neurophysiol.*, vol. 85, no. 6, pp. 2576–2589, Jun. 2001.

[11] W. Wang, S. S. Chan, D. A. Heldman, and D. W. Moran, "Motor cortical representation of position and velocity during reaching," *J Neurophysiol*, vol. 97, no. 6, pp. 4258–70, 2007.

[12] E. C. Leuthardt, G. Schalk, J. R. Wolpaw, J. G. Ojemann, and D. W. Moran, "A brain-computer interface using electrocorticographic signals in humans," *J. Neural Eng.*, vol. 1, no. 2, pp. 63–71, Jun. 2004.

[13] D. A. Heldman, W. Wang, S. S. Chan, and D. W. Moran, "Local field potential spectral tuning in motor cortex during reaching," *IEEE Trans. Neural Syst. Rehabil. Eng. Publ. IEEE Eng. Med. Biol. Soc.*, vol. 14, no. 2, pp. 180–183, Jun. 2006.

[14] J. Decety, "Do imagined and executed actions share the same neural substrate?," *Cogn. Brain Res.*, vol. 3, no. 2, pp. 87–93, Mar. 1996.

[15] A. Schnitzler, S. Salenius, R. Salmelin, V. Jousmäki, and R. Hari, "Involvement of Primary Motor Cortex in Motor Imagery: A Neuromagnetic Study," *NeuroImage*, vol. 6, no. 3, pp. 201–208, Oct. 1997.

[16] L. R. Hochberg, M. D. Serruya, G. M. Friehs, J. A. Mukand, M. Saleh, A. H. Caplan, A. Branner, D. Chen, R. D. Penn, and J. P. Donoghue, "Neuronal ensemble control of prosthetic devices by a human with tetraplegia," *Nature*, vol. 442, no. 7099, pp. 164–71, 2006.

[17] M. M. Merzenich, R. J. Nelson, M. P. Stryker, M. S. Cynader, A. Schoppmann, and J. M. Zook, "Somatosensory cortical map changes following digit amputation in adult monkeys," *J. Comp. Neurol.*, vol. 224, no. 4, pp. 591–605, Apr. 1984.

[18] J. N. Sanes, J. Wang, and J. P. Donoghue, "Immediate and delayed changes of rat motor cortical output representation with new forelimb configurations," *Cereb. Cortex N. Y. N 1991*, vol. 2, no. 2, pp. 141–152, Apr. 1992.

[19] J. P. Donoghue, S. Suner, and J. N. Sanes, "Dynamic organization of primary motor cortex output to target muscles in adult rats II. Rapid reorganization following motor nerve lesions," *Exp. Brain Res.*, vol. 79, no. 3, pp. 492–503, Mar. 1990.

[20] J. D. Leslie G Ungerleider, "Imaging brain plasticity during motor skill learning.," *Neurobiol. Learn. Mem.*, vol. 78, no. 3, pp. 553–64, 2002.

[21] A. G. Rouse and D. W. Moran, "Neural adaptation of epidural electrocorticographic (EECoG) signals during closed-loop brain computer interface (BCI) tasks," *Conf Proc IEEE Eng Med Biol Soc*, vol. 2009, pp. 5514–7, 2009.

[22] B. Renshaw, A. Forbes, and B. R. Morison, "Activity of Isocortex and Hippocampus: Electrical Studies with Micro-Electrodes," *J. Neurophysiol.*, vol. 3, no. 1, pp. 74–105, Jan. 1940.

[23] C.-L. Li, "Action and resting potentials of cortical neurones," *J. Physiol.*, vol. 130, no. 1, pp. 96–108, Oct. 1955.

[24] D. H. Hubel and T. N. Wiesel, "Receptive fields of single neurones in the cat's striate cortex," *J. Physiol.*, vol. 148, no. 3, pp. 574–591, Oct. 1959.

[25] P. D. H. Berger, "Über das Elektrenkephalogramm des Menschen," *Arch. Für Psychiatr. Nervenkrankh.*, vol. 87, no. 1, pp. 527–570, Dec. 1929.

[26] H. Jasper and W. Penfield, "Electrocorticograms in man: Effect of voluntary movement upon the electrical activity of the precentral gyrus," *Arch. Für Psychiatr. Nervenkrankh.*, vol. 183, no. 1–2, pp. 163–174, Jan. 1949.

[27] J. R. Wolpaw and D. J. McFarland, "Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 51, pp. 17849–17854, Dec. 2004.

[28] A. Kubler, F. Nijboer, J. Mellinger, T. M. Vaughan, H. Pawelzik, G. Schalk, D. J. McFarland, N. Birbaumer, and J. R. Wolpaw, "Patients with ALS can use sensorimotor rhythms to operate a brain-computer interface," *Neurology*, vol. 64, no. 10, pp. 1775–7, 2005.

[29] D. J. McFarland, D. J. Krusienski, W. A. Sarnacki, and J. R. Wolpaw, "Emulation of computer mouse control with a noninvasive brain-computer interface," *J Neural Eng*, vol. 5, no. 2, pp. 101–10, 2008.

[30] D. J. McFarland, W. A. Sarnacki, and J. R. Wolpaw, "Electroencephalographic (EEG) control of three-dimensional movement," *J. Neural Eng.*, vol. 7, no. 3, p. 036007, Jun. 2010.

[31] P. R. Kennedy and R. A. Bakay, "Restoration of neural output from a paralyzed patient by a direct brain connection," *Neuroreport*, vol. 9, no. 8, pp. 1707–11, 1998.

[32] D. M. Taylor, S. I. Tillery, and A. B. Schwartz, "Direct cortical control of 3D neuroprosthetic devices," *Science*, vol. 296, no. 5574, pp. 1829–32, 2002.

[33] M. D. Serruya, N. G. Hatsopoulos, L. Paninski, M. R. Fellows, and J. P. Donoghue, "Instant neural control of a movement signal," *Nature*, vol. 416, no. 6877, pp. 141–142, Mar. 2002.

[34] J. P. Donoghue, A. Nurmikko, M. Black, and L. R. Hochberg, "Assistive technology and robotic control using motor cortex ensemble-based neural interface systems in humans with tetraplegia," *J Physiol*, vol. 579, no. Pt 3, pp. 603–11, 2007.

[35] S. P. Kim, J. D. Simeral, L. R. Hochberg, J. P. Donoghue, and M. J. Black, "Neural control of computer cursor velocity by decoding motor cortical spiking activity in humans with tetraplegia," *J Neural Eng*, vol. 5, no. 4, pp. 455–76, 2008.

[36] M. Velliste, S. Perel, M. C. Spalding, A. S. Whitford, and A. B. Schwartz, "Cortical control of a prosthetic arm for self-feeding," *Nature*, vol. 453, no. 7198, pp. 1098–101, 2008.

[37] J. A. Wilson, E. A. Felton, P. C. Garell, G. Schalk, and J. C. Williams, "ECoG factors underlying multimodal control of a brain-computer interface," *IEEE Trans Neural Syst Rehabil Eng*, vol. 14, no. 2, pp. 246–50, 2006.

[38] G. Schalk, K. J. Miller, N. R. Anderson, J. A. Wilson, M. D. Smyth, J. G. Ojemann, D. W. Moran, J. R. Wolpaw, and E. C. Leuthardt, "Two-dimensional movement control using electrocorticographic signals in humans," *J Neural Eng*, vol. 5, no. 1, pp. 75–84, 2008.

[39] A. G. Rouse, "Neural Adaptation and the Effect of Interelectrode Spacing on Epidural Electrocorticography for Brain Computer Interfaces," Electronic Theses and Dissertations, Washington University in St. Louis, 2012.

[40] W. Wang, J. L. Collinger, A. D. Degenhart, E. C. Tyler-Kabara, A. B. Schwartz, D. W. Moran, D. J. Weber, B. Wodlinger, R. K. Vinjamuri, R. C. Ashmore, J. W. Kelly, and M. L. Boninger, "An electrocorticographic brain interface in an individual with tetraplegia," *PloS One*, vol. 8, no. 2, p. e55344, 2013.

[41] E. C. Leuthardt, G. Schalk, J. Roland, A. Rouse, and D. W. Moran, "Evolution of brain-computer interfaces: going beyond classic motor physiology," *Neurosurg Focus*, vol. 27, no. 1, p. E4, 2009.

[42] A. B. Schwartz, X. T. Cui, D. J. Weber, and D. W. Moran, "Brain-controlled interfaces: movement restoration with neural prosthetics," *Neuron*, vol. 52, no. 1, pp. 205–220, 2006.

[43] L. R. Hochberg, D. Bacher, B. Jarosiewicz, N. Y. Masse, J. D. Simeral, J. Vogel, S. Haddadin, J. Liu, S. S. Cash, P. van der Smagt, and J. P. Donoghue, "Reach and grasp by people with tetraplegia using a neurally controlled robotic arm," *Nature*, vol. 485, no. 7398, pp. 372–375, May 2012.

[44] J. L. Collinger, B. Wodlinger, J. E. Downey, W. Wang, E. C. Tyler-Kabara, D. J. Weber, A. J. McMorland, M. Velliste, M. L. Boninger, and A. B. Schwartz, "7 degree-of-freedom neuroprosthetic control by an individual with tetraplegia," *Lancet*, vol. 381, no. 9866, pp. 557–564, Feb. 2013.

[45] C. S. Bjornsson, S. J. Oh, Y. A. Al-Kofahi, Y. J. Lim, K. L. Smith, J. N. Turner, S. De, B. Roysam, W. Shain, and S. J. Kim, "Effects of insertion conditions on tissue strain and vascular damage during neuroprosthetic device insertion," *J. Neural Eng.*, vol. 3, no. 3, pp. 196–207, Sep. 2006.

[46] J. C. Williams, J. A. Hippensteel, J. Dilgen, W. Shain, and D. R. Kipke, "Complex impedance spectroscopy for monitoring tissue responses to inserted neural implants," *J. Neural Eng.*, vol. 4, p. 410, 2007.

[47] J. Williams, "Performance of chronic neural implants: measurement, modeling and intervention strategies.," 2001.

[48] G. Pfurtscheller, B. Graimann, J. E. Huggins, S. P. Levine, and L. A. Schuh, "Spatiotemporal patterns of beta desynchronization and gamma synchronization in corticographic data during self-paced movement," *Clin Neurophysiol*, vol. 114, no. 7, pp. 1226–1236, 2003.

[49] E. A. Felton, J. A. Wilson, J. C. Williams, and P. C. Garell, "Electrocorticographically controlled brain–computer interfaces using motor and sensory imagery in patients with temporary subdural electrode implants," *J. Neurosurg. Pediatr.*, vol. 106, no. 3, pp. 495–500, 2007.

[50] Z. C. Chao, Y. Nagasaka, and N. Fujii, "Long-term asynchronous decoding of arm motion using electrocorticographic signals in monkeys," *Front Neuroengineering*, vol. 3, p. 3, 2010.

[51] J. J. Williams, "ECoG correlates of visuomotor transformation, neural plasticity, and application to a force based brain computer interface," Washington University in St. Louis, St. Louis, 2012.

[52] J. J. Williams, A. G. Rouse, S. Thongpang, J. C. Williams, and D. W. Moran, "Differentiating closed-loop cortical intention from rest: building an asynchronous electrocorticographic BCI.," *J. Neural Eng.*, vol. 10, no. 4, p. 046001, 2013.

[53] A. Georgopoulos, R. Caminiti, J. Kalaska, and J. Massey, "Spatial coding of movement: a hypothesis concerning the coding of movement direction by motor cortical populations," *Exp. Brain Res.*, vol. Suppl. 7, pp. 327–336, 1983.

[54] E. Salinas and L. F. Abbott, "Vector reconstruction from firing rates," *J. Comput. Neurosci.*, vol. 1, no. 1–2, pp. 89–107, Jun. 1994.

[55] W. W. M. J. Black, Y. Gao, M. Serruya, A. Shaikhouni, E. Bienenstock, and J. P. Donoghue, "Neural decoding of cursor motion using a Kalman filter," in *in Advances in Neural Information Processing Systems 15*, 2003, pp. 133–140.

[56] D. Sussillo, P. Nuyujukian, J. M. Fan, J. C. Kao, S. D. Stavisky, S. Ryu, and K. Shenoy, "A recurrent neural network for closed-loop intracortical brain–machine interface decoders," *J. Neural Eng.*, vol. 9, p. 026027, 2012.

[57] C. Kemere, G. Santhanam, B. M. Yu, A. Afshar, S. I. Ryu, T. H. Meng, and K. V. Shenoy, "Detecting Neural-State Transitions Using Hidden Markov Models for Motor Cortical Prostheses," *J. Neurophysiol.*, vol. 100, no. 4, pp. 2441 –2452, Oct. 2008.

[58] S. M. Chase, A. B. Schwartz, and R. E. Kass, "Bias, optimal linear estimation, and the differences between open-loop simulation and closed-loop performance of spiking-based brain-computer interface algorithms," *Neural Netw. Off. J. Int. Neural Netw. Soc.*, vol. 22, no. 9, pp. 1203–1213, Nov. 2009.

[59] S. Koyama, S. M. Chase, A. S. Whitford, M. Velliste, A. B. Schwartz, and R. E. Kass, "Comparison of brain-computer interface decoding algorithms in open-loop and closed-loop control," *J. Comput. Neurosci.*, vol. 29, no. 1–2, pp. 73–87, Aug. 2010.

[60] J. P. Cunningham, P. Nuyujukian, V. Gilja, C. A. Chestek, S. I. Ryu, and K. V. Shenoy, "A closed-loop human simulator for investigating the role of feedback control in brain-machine interfaces," *J. Neurophysiol.*, vol. 105, no. 4, pp. 1932–1949, Apr. 2011.

[61] A. L. Orsborn, S. Dangi, H. G. Moorman, and J. M. Carmena, "Closed-Loop Decoder Adaptation on Intermediate Time-Scales Facilitates Rapid BMI Performance Improvements Independent of Decoder Initialization Conditions," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 4, pp. 468–477, Jul. 2012.

[62] A. G. Rouse, S. R. Stanslaski, P. Cong, R. M. Jensen, P. Afshar, D. Ullestad, R. Gupta, G. F. Molnar, D. W. Moran, and T. J. Denison, "A chronic generalized bi-directional brain-machine interface," *J. Neural Eng.*, vol. 8, no. 3, p. 036018, Jun. 2011.

[63] A. G. Rouse and D. W. Moran, "Neural adaptation of epidural electrocorticographic (EECoG) signals during closed-loop brain computer interface (BCI) tasks," *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Conf.*, vol. 2009, pp. 5514–5517, 2009.

[64] D. E. Knuth, *Art of Computer Programming*, 3rd ed., vol. 2: Seminumerical Algorithms. Addison-Wesley Professional., 1998.

[65] P. Berens, "CircStat: a MATLAB toolbox for circular statistics," *J. Stat. Softw.*, vol. 31, no. 10, pp. 1–21, 2009.

[66] R. Tibshirani, "Regression Shrinkage and Selection Via the Lasso," *J. R. Stat. Soc. Ser. B*, vol. 58, pp. 267–288, 1994.

[67] S. S. Chen, D. L. Donoho, Michael, and A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, pp. 33–61, 1998.

[68] G. Obozinski and B. Taskar, "Multi-task feature selection," in *Workshop of structural Knowledge Transfer for Machine Learning in the 23rd International Conference on Machine Learning (ICML)*, 2006.

[69] G. Obozinski, B. Taskar, and M. Jordan, "Joint covariate selection and joint subspace selection for multiple classification problems," *Stat. Comput.*, vol. 20, no. 2, pp. 231–252, 2010.

[70] Z. Xu, M. Kusner, K. Weinberger, and M. Chen, "Cost-Sensitive Tree of Classifiers," presented at the Proceedings of The 30th International Conference on Machine Learning, 2013, pp. 133–141.

[71] Z. (Eddie) Xu, M. J. Kusner, K. Q. Weinberger, M. Chen, and O. Chapelle, "Classifier Cascades and Trees for Minimizing Feature Evaluation Cost," *J. Mach. Learn. Res.*, vol. 15, pp. 2113–2144, 2014.

[72] J. Zhuang, W. Truccolo, C. Vargas-Irwin, and J. P. Donoghue, "Decoding 3-D reach and grasp kinematics from high-frequency local field potentials in primate primary motor cortex," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1774–1784, Jul. 2010.

[73] K. J. Miller, "Broadband spectral change: evidence for a macroscale correlate of population firing rate?," *J. Neurosci. Off. J. Soc. Neurosci.*, vol. 30, no. 19, pp. 6477–6479, May 2010.

[74] S. Ray, N. E. Crone, E. Niebur, P. J. Franaszczuk, and S. S. Hsiao, "Neural correlates of high-gamma oscillations (60-200 Hz) in macaque local field potentials and their potential implications in electrocorticography," *J. Neurosci. Off. J. Soc. Neurosci.*, vol. 28, no. 45, pp. 11526–11536, Nov. 2008.

[75] M. Pahwa, M. Kusner, C. D. Hacker, D. T. Bundy, K. Q. Weinberger, and E. C. Leuthardt, "Optimizing the Detection of Wakeful and Sleep-Like States for Future Electrocorticographic Brain Computer Interface Applications," *PLoS ONE*, vol. 10, no. 11, p. e0142947, Nov. 2015.