Facilitating the incorporation of neuroscience methods and knowledge into brain-computer interfaces

Mark Wronkiewicz

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Washington

2017

Reading Committee:

Adrian KC Lee, Chair

Adrienne Fairhall

Raj Rao

Kurt Weaver

Program Authorized to Offer Degree:

Neuroscience

University of Washington

Abstract

Facilitating the incorporation of neuroscience methods and knowledge into brain-computer interfaces

Mark Wronkiewicz

Chair of the Supervisory Committee:
Adrian KC Lee, Sc.D.
Speech and Hearing Sciences
Institute for Learning and Brain Sciences
Graduate Program in Neuroscience

This body of work is focused on brain-computer interfaces (BCIs) – devices that interpret brain activity in order to generate artificial output. BCIs have a great deal of promise in both rehabilitative and commercial domains. However, there is still a list of obstacles that must be solved before BCIs can advance out of controlled research settings and into real world scenarios. Some obstacles arise due to a growing separation between neuroengineering and neuroscience; many modern neuroscience publications report findings at the cortical level. Therefore, it is difficult to leverage this basic science research in BCIs that use non-invasive recordings (e.g., electroencephalography or magnetoencephalography) because those data are recorded from outside the head. To alleviate this disconnect, we borrow a technique from neuroimaging called "source imaging," which allows the estimation cortical activity (on the surface of the brain) from non-invasive data (recorded on or above the surface of the scalp). We used source imaging in an attempt to address three separate issues facing the BCI field. First, we showed that targeting activity from a specific region of cortex known to be important

for auditory spatial attention led to improved performance, compared to a traditional sensor-based approach, when classifying if a subject switched attention. Second, we developed a new transfer learning technique aimed at reducing the 20-30 minute calibration period required for most non-invasive BCIs. This method used source imaging to partially normalize variance due to head anatomy, brain anatomy, and electrode positioning, and could train a subject-independent BCI with higher accuracy than the traditional subject-dependent model despite using no training data from the subject of interest. Third, we explored functional connectivity to generate features for a BCI attempting to classify resting state. While performance did not surpass traditional methods, we outline future directions and the importance of continued research. Overall, these three aims are focused on using tools and research findings from neuroscience as a principled way to advance BCI methodology.

Table of Contents

Chapter 1: Background and introduction **1**

1.1 Introduction to brain-computer interfaces 1

1.2 Mapping non-invasive signals to the brain 6

1.3 Application: targeting regions of interest 12

1.4 Application: transfer learning 13

1.5 Application: detecting network activity 15

1.6 References 17

Chapter 2: Incorporating modern neuroscience findings to improve brain-computer interfaces: tracking auditory attention **23**

2.1 Abstract 23

2.2 Introduction 23

2.3 Methods 27

2.4 Results 34

2.5 Discussion 39

2.6 Conclusion 46

2.7 References 47

Chapter 3: Leveraging anatomical information to improve transfer learning in brain-computer interfaces **53**

3.1 Abstract 53

3.2 Introduction 54

3.3 Methods 56

3.4 Results 62

3.5 Discussion 69

3.6 Conclusion 74

3.7 References 75

Chapter 4: Incorporating modern neuroscience findings to improve brain-computer interfaces: detecting network activity **79**

4.1 Introduction 79

4.2 Methods 82

4.3 Results 84
4.4 Discussion 86
4.5 References 88

Chapter 5: Outlook and Future Directions 81

5.1 Next steps in BCIs 81
5.2 Alternative and individualized electrode montages 82
5.3 Real-time implementation: the proof is in the pudding 83
5.4 Human connectome project, network BCIs, and deep learning 86
5.5 Overpromising 87
5.6 References 89

List of Figures

Figure 1.1 Comparison of sensor space and source space 4

Figure 1.2 Illustration of the EEG forward modeling process 9

Figure 1.3 Comparison of standard subject-specific training and subject-independent training 14

Figure 2.1 Illustration of source and sensor space in a cross section of one subject's head reconstruction 25

Figure 2.2 Illustration of the two most common conductivity models in source imaging 28

Figure 2.3 Overview of the experimental organization 32

Figure 2.4 Intersubject variability of sensor space topology from identical cortical activations 35

Figure 2.5 Single-trial classification accuracy of *simulated* RTPJ activation data heavily contaminated with eye blink artifacts using sensor space features and two forms of source space features (generic and individualized) 37

Figure 2.6 Single-trial classification accuracy of *experimentally recorded* attention switching data using sensor space features and two forms of source space features (generic and individualized) 38

Figure 3.1 Schematic of source-based transfer learning method 60

Figure 3.2 Three different weighting strategies for an exemplar ROI 62

Figure 3.3 An LDA classifier trained with source-based transfer learning performed comparably to the traditional subject-specific classifier 63

Figure 3.4 Difference in accuracy between the (Gaussian mixture weighted) transfer learning approach and the traditional BCI classifier at 72 ROIs distributed across the cortical surface 64

Figure 3.5 The classifier trained with transfer learning achieved higher classification accuracy than the traditional classifier for all tested ROIs related to motor imagery 65

Figure 3.6 The transfer learning classifier achieved higher classification accuracy than the traditional classifier for all tested ROIs thought to be related to the P300 ERP 66

Figure 3.7 The classifier trained with transfer learning achieved higher classification accuracy than the traditional classifier in one of the primary auditory and visual ROIs tested 67

Figure 3.8 The transfer learning classifier achieved incremental performance gains in classifying real data from an attentional switching task and eventually surpassed the subject-specific benchmark 69

Figure 4.1 Accuracy for 20 subjects when classifying rest and task data from a motor task using support vector machines. 85

Figure 4.2 Accuracy for 20 subjects when classifying rest and task data from a motor task using random forests. 86

Acknowledgements

When I was trying to make the biggest decision of graduate school – picking a lab – I heard from multiple people that KC always puts his trainees first in all decisions. I realized soon after joining just how important that was. He constantly helped tailor my projects and tasks so that the skills I was acquiring prepared me for a career after grad school. He was also fully supportive of pursuing extra internships and workshops that weren't going to have an immediate impact on my academic career, but would become highly valuable as I progressed to future projects and opportunities. As my goals evolved throughout my years in the lab, he was always quick to encourage me to explore computational/mathematical classes through Coursera or read books about scientific writing on the side. Through all of this support, I learned how to enjoy learning again – something that was partially tamed by prior years of pre-selected course loads and by-the-book teaching coming to UW. While at UW, KC also helped me gain a critical eye when analyzing the work of others, and taught me how to view research with a wider lens in order to seek out research challenges worth tackling – even if they weren't popular problems. KC also possesses an innate ability to ask sharp questions well-outside of the standard mentor-mentee playbook. I was on the receiving end of many of those questions, and (after plenty of difficult self-reflection) he really helped me polish my skills as a critical thinker in and outside of science.

I also need to thank a number of others in the lab for their mentorship. First and foremost, Eric Larson, who helped me grow as a researcher and person beyond what could have imagined when starting graduate school. He taught me how to be principled in my science, and helped me develop the practical skills to support better work habits. Because of him, I now know how to program in Python, program much better in Matlab, develop software in a large collaborative environment on Github, build a computer, survive on the Linux command line, and dial in on an optimal work-life balance. He also elevated the quality of my published work by giving me crucial feedback on every piece of peer-reviewed writing I put together.

Ross Maddox, Dan McCloy, and Nick Foti also all pushed me outside my comfort zone in a number of ways. Ross was a conspirator in many a hair-brained project like our bitcoin mining phase, 3D modeling,

and my attempt at machine learning to fantasy football, which he was able to do successfully. I also need to thank him and Julia for being the first people to welcome me to Seattle by introducing me to (and organizing) Nerd Nite. Dan was a dependable office mate who helped me on many stats and Github questions. He also taught me how to be more purposeful in designing figures and more conscious of the audience. Finally, I need to thank Nick for indulging me in my recent obsession with deep learning. Through lots of discussions in lab (and over beers), I gained a much more pragmatic view of this technology, its best use cases, and problems worth attacking. He also had the tall task of helping me stay focused on the few most important nails to go after with this new found hammer when there are so many interesting problems that need solving.

I also want to thank the rest of the lab for their help and support along the way. Perhaps most importantly, the environment they created never felt judgmental or competitive. The supportive and energizing environment they created (c.f., success scotch, SHACS, nose goes, happy hours) was vital to keeping my motivation and moral high. Outside the lab, I also want to credit my committee members including Raj Rao, Adrienne Fairhall, Kurt Weaver, and Eric Chudler for the advice and support along the way. Their insight kept me on track throughout the process, and helped me avoid the rabbit holes too deep to explore during a graduate degree. Specifically, I want to credit Adrienne for all the opportunities she provided me. She was the co-creator of the Allen Institute's Summer Workshop on the Dynamic Brain takes place in Friday Harbor in the San Juan Islands. Because of her efforts, I was further honed my programming skills, began a memorable long-term collaboration with a couple other graduate students, and heard a truly captivating lecture on deep learning by her husband Blaise that changed my career path. Jumping into bioluminescent-algae-filled 50-degree ocean water, attracting seals with underwater lamps, and enjoying multiple meals on the back porch of the cafeteria weren't bad experiences either and made for one of my favorite stretches of grad school.

Finally I would like to thank my family and my friends in the program. None of the breaks or holidays would have been nearly as fun or refreshing without your support. And thanks to Aiva Ievina for constant

support throughout the taxing period of academia as well as her unending drive to seek out adventures. The unending stream of adventures and shenanigans made all the tough times easy to forget.

Dedication

This work is dedicated to my grandpa, Daniel Michalek. When I was a kid, we spent a lot of time talking about and practicing gymnastics together – even at rest stops in the desert on family road trips. His dedication to the sport (and the fact that he would regularly get into hand stand battles with his brother until his early 60s) made him a natural role model for me. While physically tough and energetic, he has always possessed herculean mental strength. He is the kind of person that cared for so many family members while in illness and hard times – often multiple at once and usually over decades. And yet, he's the kind of person that also never complained or asked for anything in return. He still shoos the family away from the dirty dishes after the holiday dinner even though he is usually already off the hook for cooking it. All in all, he's the kind of person that I'm still striving to emulate and the kind of person that I will continue to think of when facing momentous life decisions.

Chapter 1. Background and introduction

Significant parts of this chapter have been submitted by Mark Wronkiewicz, Eric Larson, and Adrian KC Lee as a book chapter for the book titled "Brain-Computer Interfaces Handbook: Technological and Theoretical Advances" with the title "Facilitating the integration of modern neuroscience into non-invasive BCIs" scheduled to be published in 2017.

1.1 Introduction to brain-computer interfaces

The notion of using BCIs to liberate the mind from the physical constraints of the body is a captivating technological idea – an idea that, despite being over three-quarters of a century old (Wolpaw & Wolpaw, 2012), continues to shape facets of both science fiction stories and many scientists' careers. Regardless of their long-standing promise, however, the transition from fiction to reality has been sluggish; BCIs have yet to make a meaningful impact on clinical or commercial fields. Compared to other grand scientific accomplishments like mapping the human genome or establishing an international space station, it is puzzling as to why conventional BCI paradigms have not found success outside controlled laboratory settings.

Three conventional BCI paradigms – all published over a quarter-century ago – continue to comprise a large majority of non-invasive BCI research. Systems based on sensorimotor rhythms (SMRs) related to movement, for instance, were first published in 1991 (Wolpaw, et al., 1991). Other BCIs like those relying on the P300 event-related potential or steady state visually evoked responses (SSVEPs) were first reported even earlier in 1988 and 1977, respectively (Farwell & Donchin, 1988; Vidal, 1977). The literature first reporting the underlying neurophysiology of the signals behind these BCIs is even older; SMRs related to motor imagery were reported as reliably detectable in 1979 (Pfurtscheller & Aranibar, 1979), the P300 ERP was characterized in a pair of papers from 1964 and 1965 (Chapman & Bragdon, 1964; Sutton, et al., 1965), and visually-evoked potentials were first outlined in 1973 (Vidal, 1973).

These pillars of human BCI work were all established just as the field of cortical mapping (using both MRI and electromagnetic methods) was about to begin its crescendo. Even motor-imagery BCIs – the youngest of the main paradigms – was published the year before fMRI was first used in humans. This is important because those same three BCI paradigms still dominate today's research, 25+ years later, and they still rely on the same neuroscience foundation that existed at the time of the original publications. Given that researchers are still working today to build a solid footing for BCIs in real-world applications outside the laboratory, it is fair to question whether continued iterative research on these conventional paradigms represents the most encouraging path forward. With decades of neuroscience research since the original BCI publications, perhaps it is worth striving to incorporate novel neuroscience methods and findings (or at least use updated versions of the original research) to further the BCI field.

In this chapter, we start with a high level explanation of some of the limitations that continue to hinder the advancement of BCIs. We then cover some effective methodologies that could accelerate the transition of BCIs into the commercial and clinical spaces. Here, we focus mostly on non-invasive EEG-based methods because they make up the majority of BCI applications, are portable, and are currently suitable for healthy users. In Section 1.2, we cover the neuroimaging tools that allow estimation of cortical activity from recordings made outside the head. We then discuss how infusing both knowledge (Section 1.3 and 1.5) and techniques (Section 1.4) from neuroimaging into BCIs addresses some of the important obstacles that must be overcome in BCI research. Each of these sections explain the basic concepts and motivations for the primary research that follows in the Chapters 2, 3, and 4.

To understand why it is worthwhile to port knowledge and tools from neuroimaging into BCIs, an understanding of the complex role of machine learning in BCI research is needed. In standard machine learning problems like classifying handwritten digits with the MNIST dataset, the objective is clear. There are ten different handwritten digits and the task is to distinguish the digits as accurately as possible. The end goal is known, ground truth labels are available for each data point, and the experimenter is primarily tasked with finding an algorithm that produces a set of discriminable features.

Currently, BCIs do not fit this mold. The overall goal – to interpret brain activity – is not clearly defined, so researchers typically cue the subject to perform a very precise action at regular intervals in hopes of constraining brain activity. Typically, choosing a paradigm is difficult because there is no strong objective basis for choosing a task given the range of tasks humans can accomplish (Schalk, et al., 2008). Even then, electrophysiologic brain data is known to be non-stationary and show strong variability within trials of the same type making it difficult to assign ground truth labels to each data sample. Regardless, the researcher must still apply algorithms tasked with generating some set of discriminable features (as in classic supervised learning problems). In other words, BCI researchers must define a task and assign data labels *in addition to* generating features to discriminate the data samples. Consider if the handwritten dataset was flawed in this manner; perhaps a sizable subset of digits might actually be letters mislabeled as numbers (so the sample labels are sometimes inaccurate) or maybe the digits came from the first week of an elementary school penmanship class (causing wide variability). In BCI work, therefore, any method informs the choice in task (and therefore, the data produced) or the features used to classify those data, is worth exploring.

Basic neuroimaging research offers one method of tackling this problem. Often, neuroimaging involves localizing a brain region that shows changes in activity during cognitive processing, which can offer insights into both tasks and features that may be useful for a BCI. One of the central limitations facing BCI research, however, is the difficulty of incorporating these novel advances in neuroscience into new BCI systems. For example, recent insights into resting state networks or the functional brain areas involved in recognizing faces have been explored in neuroscience but not neuroengineering. A core reason for this divide is that non-invasive BCIs most often record electric potentials at a set of EEG electrodes on the scalp; more formally, this domain is often referred to as the *sensor space* because it is where the EEG sensors reside (Figure 1.1, top row). However, recordings made in the sensor space are only an external proxy to the cortical activity that is truly of interest for much of the neuroscience community.

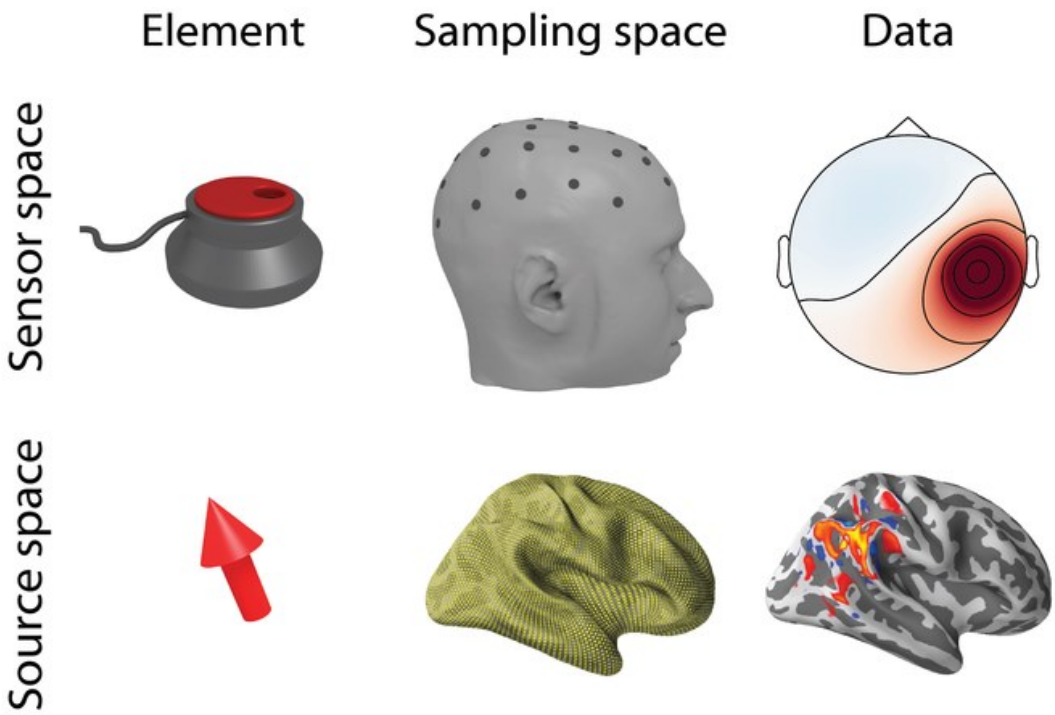

Figure 1.1 Comparison of sensor space and source space. Top row: The sensor space is traditionally where non-invasive EEG recordings are processed and visualized. Electric potentials are recorded at tens of electrodes (left) positioned against the scalp (middle) and data is analyzed at the scalp surface (right). Bottom row: In source imaging, the source space provides an alternative reference frame to process EEG data that better aligns with modern human neuroscience. Data is projected from the sensor space to the source space (here using source imaging) to estimate activity at model current dipoles (left). Several thousand dipoles are modeled at the grey/white matter boundary of each hemisphere (middle) and data is analyzed on the cortical surface (right).

In contrast, modern human neuroscience uses a number of methods in addition to EEG to acquire brain signals and map them to the brain itself. Therefore, it is difficult to develop new BCI paradigms or signal processing strategies from these advancements because of the difficulties in translating these discoveries into an EEG context. For example, functional MRI (fMRI), electrocorticography, and single-unit recordings all represent major families of neuroscience research where the collected data is fundamentally different

from EEG recordings. More specifically, a key difference with these techniques is that they operate in a cortical reference frame instead of the sensor space. Therefore, any methods that can port EEG data in this cortical reference frame can tighten the connection between EEG and other modern neuroscience techniques.

A neuroimaging tool called "source estimation" offers an alternative to the sensor space. Source estimation attempts to calculate the sources of brain activity in the *source space* from EEG recordings made in the sensor space. In contrast to the sensor space, the source space refers to the domain where brain activity originates. Throughout this chapter, we discuss a popular form of source estimation called source imaging – the topic of Section 1.2. The underlying goal of source imaging is to estimate cortical sources of activity at several thousand discrete points distributed across the entire cortex. In other words, it estimates whole-brain cortical activity from recorded EEG data (Figure 1.1, bottom row).

In certain situations, working with data in the source space carries important advantages over the sensor space. The benefits we examine in this chapter arise because the field of human neuroimaging also operates primarily in the source space. Therefore, collections of knowledge and techniques from neuroimaging become useful for improving BCIs after converting the sensor-based EEG data into a source space representation. Three specific examples will be discussed in this dissertation. First (in Chapter 2), functional imaging literature has localized numerous cortical regions where activity is modulated during certain behavioral tasks. With source imaging, we can focus on processing EEG activity originating only a specific cortical region-of-interest (ROI) and avoid expending resources on less relevant data. Second (in Chapter 3), neuroimaging often involves normalizing brain data across subjects for processes like group-level statistics and visualization. These same algorithms are useful for tackling noteworthy goals in BCIs like transfer-learning, where the aim is to recycle training data across subjects to expedite BCI calibration. Third (in Chapter 4), relatively recent neuroimaging work has begun to characterize how certain cortical networks – broadly called resting state networks – are involved in particular tasks. Tracking the synchronization in activity between these sets of regions is explored as a

feature in a BCI classifier. We will formally review the source imaging technique in Section 1.2 before detailing these three specific opportunities that the source space offers BCIs.

1.2 Mapping non-invasive signals to the brain

Source estimation is the area of study concerned with bridging the sensor and source spaces. Specifically, it is used to estimate the sources of neural activity that gave rise to signals we observe outside the head with noninvasive measurements like EEG and MEG. In this section, we specifically describe *source imaging* – a form of source estimation – and provide a mathematical overview of this procedure.

EEG and MEG largely reflect the summed activity from large populations of pyramidal neurons (Hämäläinen, et al., 1993), which are the largest and most common excitatory neurons in the cortex. Pyramidal neurons are generally oriented with the "trunk" of their large dendritic tree (called the apical dendrite) oriented normal to the cortical surface. Therefore, when pyramidal neurons are influenced by excitatory or inhibitory postsynaptic potentials (PSPs), charged ions flow within the neuron primarily along an axis aligned with the apical dendrite. This charge flow is often referred to as the "primary current." Since charge cannot accumulate in the brain, "secondary" (or volume) current loops also flow extracellularly throughout the head to compensate for the primary current (Lopes da Silva & Van Rotterdam, 2011). Therefore, an activation pattern resembling a source-sink configuration arises that mimics the characteristics of a current dipole – a fact that simplifies source imaging (Hämäläinen, et al., 1993; Lopes da Silva & Van Rotterdam, 2011; Dale & Sereno, 1993) as discussed later. These PSP activations (lasting tens to hundreds of milliseconds) are much slower than action potentials (lasting ~1 ms), so the PSP activation of a population of neurons is more likely to temporally overlap (Lopes da Silva, 2010). Combined with the aligned arrangement of these neurons, the field potentials of simultaneously active neurons add constructively and become detectible outside the head (Lopes da Silva & Van Rotterdam, 2011).

MEG and EEG sensors are separated from the cortex by a few centimeters of tissue. The neural activity is therefore undesirably transformed before reaching these sensors. Part of this transformation arises due to volume conduction; activity from any given neural source is detected by a number of MEG and/or EEG sensors, or equivalently, the voltage trace for any single electrode or magnetometer is the sum of many sources of activity. A second transformation results because the tissues in the head all have different conductivity profiles (e.g., the skull is much less conductive than the scalp). These inhomogeneities between tissue layers cause further distortion as electromagnetic activity propagates to the EEG electrodes (Lopes da Silva, 2010). At the group level, a related issue results from the substantial geometrical differences in the shape of head and brain tissues between subjects. Since these anatomical differences heavily influence spatial profile of brain activity on the scalp, even activations of hypothetically identical brain regions will appear different in the EEG recordings because of the geometric anatomical variability across subjects. In a similar problem, the specific location of each EEG sensor, including small deviations from the ideal placement, affect the profile of recorded activity. All of these components act to obscure the relationship between source space activity and recordings in the sensor space on a unique subject-by-subject basis.

Source estimation is a modeling approach that attempts to account for all components of this undesirable transformation when estimating the active neural sources. As non-invasive recordings do not allow us to directly observe the sources of cortical brain activity, we can view this problem as a member of a broader class of physics and mathematics questions often referred to as "inverse problems." To solve this particular inverse problem, we first construct a forward model to capture how neural signals are distorted by this transformation before reaching the EEG electrodes. We then generate an inverse solution that attempts to account for (or "undo") this collective transformation. In the end, solving this inverse problem provides an estimate of the unobservable cortical activity from the observed EEG data.

There exist several major methods of carrying out source estimation, and each is based on different assumptions. Our work uses source imaging, which is a technique for estimating neural activity across the entire cortex simultaneously. This approach entails creating a cortical model of several thousand

model current dipoles distributed evenly across the cortical surface. These dipoles are oriented orthogonal and outward to reflect the biophysical properties of the cortex, and each dipole can be thought to represent the primary current due to one cortical macrocolumn of pyramidal neurons. The goal in source imaging is to estimate the activity of many distributed elements that sample the entire cortical space (analogous to estimating the intensity of pixels in a digital image).

Putting these principles into practice, we first quantitatively characterize how neural EM signals propagate to the M/EEG sensors with the forward model. This model consists of two main components. First, a source space model describes the location and orientation of each simulated current dipole. As this consists of a set of 3D locations and normal vectors that define the dipoles' geometry, it can often be conveniently expressed as a 3D mesh or volumetric grid. Typically, this requires either a brain reconstruction segmented from a structural MRI scan or a generic brain model (like Freesurfer's "fsaverage"; Fischl, 2012). Second, the head model characterizes how EM activity is transformed when traversing the different head layers (like the skull and scalp). The changes in conductivity at each layer boundary act to spatially smear the signals of interest (but note that this is less of a problem in MEG as the magnetic permeability of all biological tissues is essentially the same as free space). While a handful of different head models exist, there are two used most commonly: spherical head models, which are mathematically efficient but simplistically approximate the head as a sphere, and boundary element models, which are MRI-reconstructed anatomically accurate representations of the head layers (Ermer, et al., 2001). For further review of the forward model, see Hämäläinen, et al., 1993 or Baillet, 2010.

Given the forward model, we can estimate the influence of each dipole on each MEG or EEG sensor. In source imaging, these weights make up the entries of the forward (or "gain") matrix $\mathbf{A}$. We describe a pattern of cortical activity (i.e., the current at all dipoles) using the vector $\mathbf{j}$, and relate this cortical activity to the vector of sensor measurements $\mathbf{x}$ (i.e., the electric potential at all electrodes) as:

$$\mathbf{x} = \mathbf{A}\mathbf{j} + \mathbf{n} \tag{1.1}$$

Here, $\mathbf{n}$ is Gaussian distributed white noise. See Figure 1.2 for an illustration of the forward modeling problem for EEG electrodes.

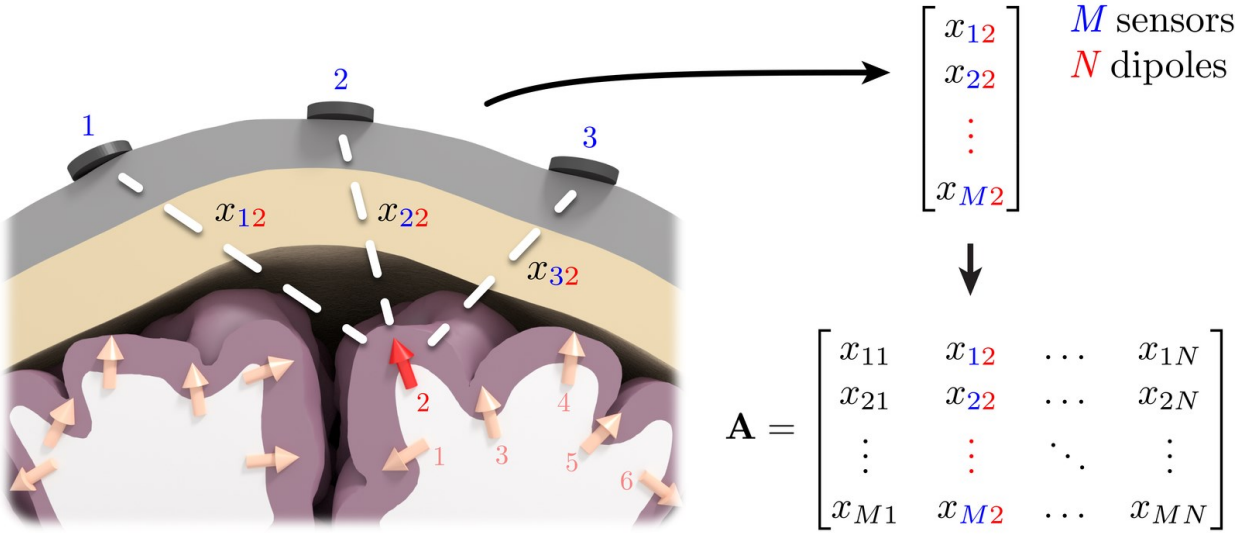

Figure 1.2 Illustration of the EEG forward modeling process. Left: The relationship between an exemplar dipole (red) and the EEG electrodes (black) is calculated (white dotted lines) by solving for the electric potential of the propagating electromagnetic activity at the interface of each pair of conducting layers. These weights are compiled into a vector that quantifies how electrical current at one dipole in the source space influences the electric potential at each electrode in the sensor space. The weight vectors are computed for all dipoles and concatenated as columns to form the forward (or gain) matrix $\mathbf{A}$, which allows calculation of the electric potential $\mathbf{x}$ at all M electrodes given the activity $\mathbf{j}$ at all N dipoles. Head layers depicted here include the scalp (grey), skull (off-white), grey matter (pinkish grey) and white matter (white).

Source imaging is aimed at estimating the active brain regions from non-invasive measurements, i.e., the reciprocal to the forward problem. This challenge is one example of an "ill-posed" problem, which are a class of problems that satisfy one of the following properties: there are infinite solutions, there is no solution, or the solution is non-continuous such that a small amount of noise in the observation can lead to drastic changes in the estimated model parameters – here, the model parameters represent the activity of the model current dipoles. To transform an ill-posed problem into a well-posed one, we require

regularization, which introduces additional assumptions to reduce the model's variance. There are a number of different regularization methods, all of which have the overarching goal of ensuring a unique and stable solution.

The commonly used minimum-norm estimate (MNE) source imaging approach (Gramfort, et al., 2014; www.martinos.org/mne/) attempts to minimize the ℓ_2-norm using Tikhonov regularization (Baillet, et al., 2001; Tikhonov & Arsenin, 1977). A solution with these properties was obtained by calculating the pseudoinverse matrix $\mathbf{M}$ of the gain matrix using MNE, which balances error between the measured and modeled data with the Euclidean (or ℓ_2) norm of the solution vector (regularized by λ; Baillet, 2010; Hämäläinen & Ilmoniemi, 1984):

$$\mathbf{M} = \mathbf{R}\mathbf{A}^T(\mathbf{A}\mathbf{R}\mathbf{A}^T + \lambda^2\mathbf{C})^{-1} \qquad (1.2)$$

This method of calculating $\mathbf{M}$ also requires an estimate of the covariance matrices $\mathbf{C}$ and $\mathbf{R}$, which are the spatial covariances of vectors $\mathbf{n}$ and $\mathbf{j}$, respectively, when generalized across time. One can obtain a noise covariance $\mathbf{C}$ for each subject by concatenating all intertrial intervals or using several minutes of an empty room recording (for MEG only). The source covariance $\mathbf{R}$ is typically assumed to be the identity matrix but is sometimes used to incorporate other priors, e.g., from fMRI (Dale, et al., 2000). $\mathbf{M}$ can also be interpreted as the maximum *a posteriori* estimate given the sensor and source covariance matrices and assuming that the source amplitudes are normally distributed (Hämäläinen, et al., 2010). For a full mathematical derivation, see Baillet, et al., 2001 or Hämäläinen, et al., 2010. In the end, the pseudoinverse matrix $\mathbf{M}$ in (2) embodies a unique and linear mapping to provide an estimate the source activity $\hat{\mathbf{j}}$ from the EEG sensor recordings $\mathbf{x}$:

$$\hat{\mathbf{j}} = \mathbf{M}\mathbf{x} \qquad (1.3)$$

Thus, $\mathbf{M}$ is composed of a set of spatial filters where each component estimates current at a particular current dipole (representing a specific patch of cortex).

The MNE source imaging approach does come with some drawbacks that are worth reviewing. First, the underlying ℓ_2 regularization scheme is known to produce spatially diffuse cortical estimates even if the

true brain activity is focal. There is also little biological evidence supporting any particular regularization technique, including the minimum-norm. The brain is not known to minimize the energy of the total cortical current. Often, regularizers are chosen because they are intuitive, yield some desired activity distribution, and have a linear solution making the calculation computationally convenient (Baillet, 2010). Finally, we often assume time-invariance in regard to the sensor and source covariance matrices, which are an important component in source imaging that act to normalizing the spatial coactivation patterns in M/EEG sensors and current dipoles, respectively. This assumption that is not supported in theoretical (Friston, 1997) or empirical studies (de Pasquale, et al., 2010; Brookes, et al., 2014) though recent work has proposed source imaging methods that are amenable to nonstationary brain signals (Gramfort, et al., 2013; Castaño-Candamil, et al., 2015; Woolrich, et al., 2013).

Source imaging aside, a common task throughout many types of neuroimaging is to spatially normalize neural activity estimates across brains for statistical analysis and visualization. As mentioned previously, this is needed to account for the substantial intersubject anatomical variability. Often, brain morphing is accomplished with either a surface- or volume-based normalization algorithm. Surface-based methods more accurately morph cortical regions (Fischl, et al., 1999), making them appropriate for transferring cortical source estimates. The spherical morphing procedure is one such example of surface-based morphing. This procedure starts by first transferring data onto a high-resolution version of the cortical surface mesh and smoothing it using an isotropic diffusion process (Gramfort, et al., 2014). The mesh is then mapped onto a sphere using the procedure described in Fischl, et al. 1999. The convexity of sulcal/gyral regions, defined using a curvature threshold, are maximally aligned to a second spherical brain from a different subject. This alignment is crucial as cortical folding varies considerably between subjects causing variation in how electromagnetic signals propagate to the sensor space (even for identical ROIs; Chapter 2). The cortical activation themselves are then transferred, the target brain is deflated, and finally the data are smoothed onto a standard resolution mesh. The entire process is often compiled into one morphing matrix, to allow data morphing with a single matrix multiplication. In BCIs, we can extend this tool to facilitate transfer learning – a method of recycling data across subjects described in Section 1.4 and Chapter 3.

1.3 Application: targeting regions of interest

When comparing sensor- and source-based BCIs, the source domain carries with it a few theoretical advantages over the conventional sensor space paradigms. The sensor-based approach very often uses spatial filtering algorithms like independent component analysis (ICA) or the common spatial pattern (CSP) algorithm to spatially filter the brain activity into an algorithmically optimal data space. This strategy has proven useful empirically but has some limiting drawbacks. First, these spatial filtering algorithms, which linearly recombine EEG signals, are difficult to interpret neuroscientifically; by projecting data into an algorithmic space, they further complicate the (already loose) spatial relationship between scalp electrodes and the cortex. Second, it is difficult to directly compare these spatial projections (e.g., EEG activity projected onto subject-specific CSP components) across subjects. These filters are often recomputed for each recording session such that the projected data from one session will not share an easily-identifiable common reference frame with the next (i.e., the projected recordings do not share a common vector space). This contributes to the challenges with sensor-based transfer learning as discussed in the Sections 1.4 and 3. Finally, many BCI signal processing algorithms require that useful signal features are punctate in space and/or time. While this is an appropriate assumption in event-related paradigms (e.g., P300 or SSVEP), this approach is not likely to generalize to more nuanced brain signals. As a few groups have suggested (Vansteensel, et al., 2010), future BCIs may instead work better with distributed cognitive activity or some form of network synchronization – two topics at the center of the neuroimaging community's attention.

Source imaging projects brain activity directly to the cortex thereby mitigating many of the shortcomings associated with the sensor space. Most importantly, it allows a tighter link with modern neuroscience research since most modern human neuroscience experiments collect data in the source domain. The practical implication for BCI research is that novel neuroscience findings are more readily applicable in the source space. In functional neuroimaging experiments for example, the spatial extent of a cortical region under investigation is often well-defined (e.g., as a discrete set of dipoles on a template brain).

Therefore, these discoveries are readily translated to enable ROI-based targeting in a BCI. This targeted source approach is also more likely to generalize because the ROI originates from an independent study that has often statistically confirmed its spatial extent across a cohort of subjects. This represents a more scientifically informed and repeatable signal processing approach; it does not rely on hand-picking electrodes or recomputed spatial filters for every recording session. This also largely sidesteps any trial-and-error process to find the optimal electrode(s) or signal processing technique for a given task-related activation. Additionally, in algorithms where spatial projections are difficult to rank in order of importance (like in ICA), this also avoids the need for an expert to subjectively choose which projection(s) to continue processing. Overall, the source space remains a principled way to inject insights from neuroscience research into the non-invasive BCI development process.

To test these assertions, we tested a technique that attempted to track users' spatial attention with the same data but in sensor and source space representations. We hypothesized that by targeting a specific region in the source space, we could achieve higher classification accuracy of when a subject switched auditory spatial attention relative to the sensor space. Importantly, the targeted region was shown in an independent study to have statistically higher activity during switching periods. In this work, we found that in targeting the most relevant activity in the source space performed better than standard sensor-based BCI techniques. This work is explained further in Chapter 2.

1.4. Application: transfer learning

Unlike early "operant-conditioning" BCIs where the processing pipeline was static, modern BCI systems now use machine learning classifiers to rapidly adapt to the statistics of a user's brain activity. This transition was a critical advancement in BCI methodology as it shortened BCI training periods from several weeks to 20-30 minutes (Krauledat, et al., 2007; Guger, et al., 2003) for some paradigms. However, a further reduction in calibration time is vital in order for BCIs to gain acceptance outside laboratories (Lotte & Guan, 2010). In healthcare, any therapy that requires a half hour of system calibration before each treatment session is financially prohibitive. For BCIs aimed at healthy users, this lengthy calibration period is a serious marketing and usability concern. Imagine attempting to

commercialize a car that required a 20-30 minute startup period to calibrate the steering wheel and pedals before *every single* trip.

A natural question to ask is whether existing BCI recordings can be recycled to speed up calibration for an upcoming BCI session. While conceptually straightforward, it is challenging to gain information from a previous recording because EEG signals are affected by many sources of intersubject (and intersession) variability. A few examples of this variation include changes in electrode number or positioning, head anatomy, cortical folding, fatigue, and mental strategy. Thus, the reuse of training data across subjects or sessions is significantly constrained; these inconsistencies violate the implicit assumption of standard machine learning algorithms that the training and test data share both an identical feature space and probability distribution. This hurdle forces most BCI systems to obey a subject-specific scheme and be recalibrated from scratch for every recording session (Figure 1.3A).

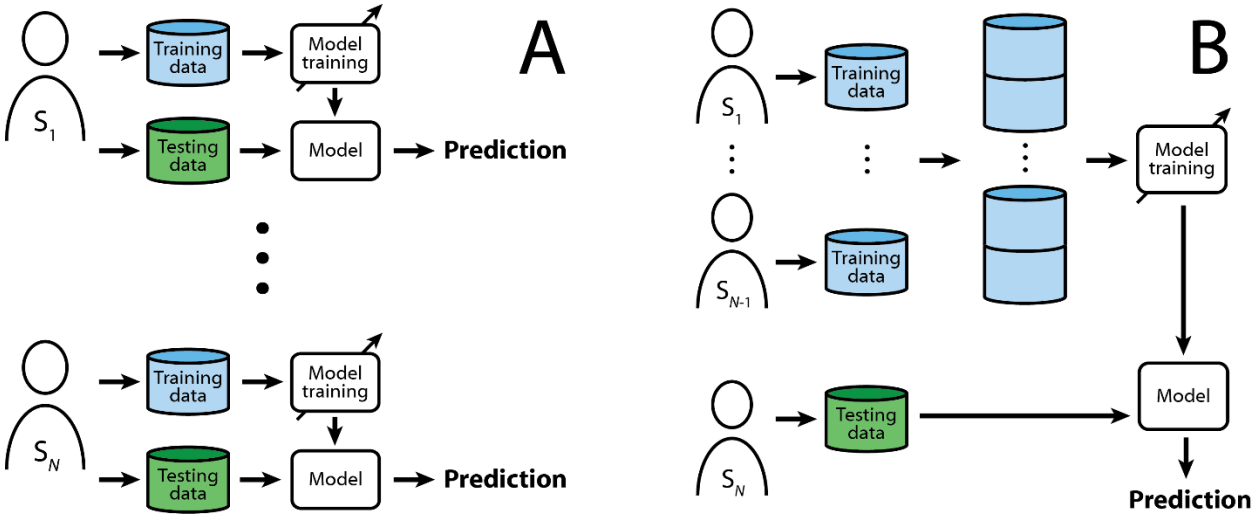

Figure 1.3 Comparison of standard subject-specific training (A) and subject-independent training (B). **A**: In most BCIs, a machine learning algorithm is trained and tested separately for every BCI recording session because the data's underlying statistics are highly subject- and session-dependent. Therefore, each classification model is typically only effective for one recording session. **B**: Subject-independent BCIs

attempt to account for intersubject variability with transfer learning techniques. In this way, pre-recorded data from N-1 subjects is used to train a model that can quickly begin to make predictions on testing data from the N^{th} subject. Here, the transfer of training data is depicted, but an equally valid approach is to transfer knowledge from a set of classifiers (each trained on one subject) into an aggregate classifier.

Within machine learning research, the field of "transfer learning" aims to address this problem of reusing training data even when the features or their distributions differ (Pan & Yang, 2010). In BCIs specifically, an attractive application of this idea is to pool prerecorded EEG data from N-1 subjects to train the upcoming (N^{th}) subject's classifier (Figure 1.3B). A system that requires no training data for a new subject is referred to as a subject-independent BCI (Fazli, et al., 2009). A number of studies have carried out transfer learning in the sensor space (see Lotte, 2015 for a review), but directly transferring EEG activity neglects several sources of substantial variability related to the electrode positioning, head anatomy, and brain anatomy. Therefore, we conducted an experiment to study a source-based transfer learning technique, and the outcome supported these theoretical advantages in reducing variance. We found that in both simulation and real data, this approach often surpassed the accuracy of a conventional subject-specific BCI even when trained only on data morphed from a few other subjects (i.e., in a subject-independent manner). This work is explained in Chapter 3.

1.5 Application: detecting network activity

As outline in Sections 1.2 and 1.3, the source space provides a path to more directly target the activity in cortical regions for BCIs (i.e., leveraging neuroimaging research in neuroengineering work). Like the large majority of BCIs (including SSVEP, P300, and motor-imagery paradigms), this approach usually assumes a change in an anatomically localized brain activity and uses this information to modulate a BCIs output accordingly. The neuroimaging community, however, has explored many concepts beyond functional localization of activity. Over the past 15 to 20 years, network neuroscience, which often involves

estimating functional connectivity between anatomically separated regions (rather than activation of a single cortical area), has gained significant attention in the neuroimaging community. This evidence suggests that a finite number (7-10) large-scale networks exist (Damoiseaux et al., 2006; van den Heuvel and Pol 2010), which each consist of functionally linked regions that can be identified from resting state activity. While the majority of this research was achieved using fMRI and PET, electrophysiological analogs of these initial network findings have been found over the past decade. Therefore, network activity may offer a fundamentally new type of feature to explore in BCIs (Besserve et al., 2011; Vansteensel et al., 2010).

Two of these networks are particularly relevant to neuroimaging and worth exploring in BCIs. The first is the default mode network (DMN), which tends to show higher functionally connectivity between its regions during rest (i.e., when a subject is not engaged in a cognitive task; (Hellyer, et al., 2014). The second is the dorsal attention network (DAN), which is believed to aid in top-down direction of attention in goal-directed tasks (Corbetta and Shulman, 2002). The functional connectivity in these networks may provide a window into the high-level state of the user. For example, detecting increased synchronization between the DMN regions could serve as a signal that the user is not cognitively engaged with the world and a BCI system should pause making decisions. The DAN may provide a second independent source of complementary information as increased functional activity suggest interaction with the world. In Chapter 4, we explore the use of source imaging to examine DMN and DAN activity. We hypothesize that this strategy could offer a new set of features that are fundamentally different from those used in standard BCIs.

Acknowledgments

This work was funded by the National Science Foundation Division of Graduate Education, Graduate Research Fellowship Program (DGE-1256082 to MW), The Department of Defense Air Force Office of Scientific Research Young Investigator Program (FA9550-12-1-0466 to AKCL), and The Department of Defense Office of Naval Research Young Investigator Program (N00014-15-1-2124 to AKCL).

1.6 References

Ahn, M., Hong, J. H. & Jun, S. C., 2012. Feasibility of approaches combining sensor and source features in brain-computer interface. *Journal of Neuroscience Methods,* 204(1), pp. 168-78.

Baillet, S., 2010. The Dowser in the Fields: Searching for MEG Sources. In: P. C. Hansen, M. Kringelbach & R. Salmelin, eds. *MEG: An Introduction to Methods.* New York: Oxford University Press, Inc., pp. 83-123.

Baillet, S., Mosher, J. C. & Leahy, R. M., 2001. Electromagnetic Brain Mapping. *IEEE Signal Processing Magazine,* 18(6), pp. 14-30.

Besserve, M., Martinerie, J. & Garnero, L., 2011. Improving quantification of functional networks with EEG inverse problem: Evidence from a decoding point of view. *NeuroImage,* 55(4), pp. 1536-47.

Brookes, M. J. et al., 2014. Measuring temporal, spectral and spatial changes in electrophysiological brain network connectivity. *NeuroImage,* Volume 91, pp. 282-99.

Castaño-Candamil, S. et al., 2015. Solving the EEG inverse problem based on space-time-frequency structured sparsity constraints. *NeuroImage,* Volume 118, pp. 598-612.

Chapman, R. M. & Bragdon, H. R., 1964. Evoked Responses to Numerical and Non-Numerical Visual Stimuli while Problem Solving. *Nature,* Volume 203, pp. 1555-7.

Cincotti, F. et al., 2008. High-resolution EEG techniques for brain-computer interface applications. *Journal of Neuroscience Methods,* 167(1), pp. 31-42.

Corbetta, M. & GL, S., 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience,* Volume 3, pp. 201-15.

Dale, A. M. & Sereno, M. I., 1993. Improved Localization of Cortical Activity by Combining EEG and MEG with MRI Cortical Surface Reconstruction: A Linear Approach. *Journal of Cognitive Neuroscience,* 5(2), pp. 162-76.

Damoiseaux, J. et al., 2006. Consistent resting-state networks across healthy subjects. *Proceedings of the National Academy of Sciences,* 103(37), pp. 13848-13853.

Darvas, F., Ermer, J. J., Mosher, J. C. & Leahy, R. M., 2006. Generic Head Models for Atlas-Based EEG Source Analysis. *Human Brain Mapping,* Volume 27, pp. 129-43.

de Pasquale, F. et al., 2010. Temporal dynamics of spontaneous MEG activity in brain networks. *PNAS,* 107(13), pp. 6040-45.

Debener, S., Emkes, R., De Vos, M. & Bleichner, M., 2015. Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear. *Scientific Reports,* Volume 5, pp. 1-11.

Do Valle, B., Cash, S. S. & Sodini, C. G., 2014. *Wireless behind-the-ear EEG recording device with wireless interface to a mobile device (iPhone/iPod touch).* Chicago, IL, 2014 36th Annual International Conference on the IEEE EMBS.

Ermer, J. J., Mosher, J. C., Baillet, S. & Leahy, R. M., 2001. Rapidly recomputable EEG forward models for realistic head shapes. *Physics in Medicine and Biology,* Volume 46, pp. 1265-81.

Farwell, L. A. & Donchin, E., 1988. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogray and Clinical Neurophysiology,* 70(6), pp. 510-23.

Fazli, S., Danóczya, M., Schelldorfer, J. & Müller, K.-R., 2011. ℓ1-penalized linear mixed-effects models for high dimensional data with application to BCI. *NeuroImage,* 56(4), pp. 2100-8.

Fazli, S. et al., 2009. Subject-independent mental state classification in single trials. *Neural Networks,* 22(9), pp. 1305-12.

Fischl, B., 2012. FreeSurfer. *NeuroImage,* 62(2), pp. 774-81.

Fischl, B., Sereno, M. & Dale, A. M., 1999. Cortical Surface-Based Analysis II: Inflation, Flattening, and a Surface-Based Coordinate System. *Neuroimage,* Volume 9, pp. 195-207.

Fischl, B., Sereno, M. I., Tootell, R. B. & Dale, A. M., 1999. High-Resolution Intersubject Averaging and a Coordinate System for the Cortical Surface. *Human Brain Mapping,* 8(4), pp. 272-84.

Friston, K. J., 1997. Transients, Metastability, and Neuronal Dynamics. *NeuroImage,* Volume 5, pp. 164-71.

Fuchs, M. et al., 2002. A standardized boundary element method volume conductor model. *Clinical Neurophysiology,* Volume 113, pp. 702-12.

Golub, M. D., Chase, S. M., Batista, A. P. & Yu, B. M., 2016. Brain-computer interfaces for dissecting cognitive processes underlying sensorimotor control. *Current Opinion in Neurobiology,* Volume 37, pp. 53-8.

Gramfort, A. et al., 2014. MNE software for processing MEG and EEG data. *NeuroImage,* Volume 86, pp. 446-60.

Gramfort, A. et al., 2013. Time-frequency mixed-norm estimates: Sparse M/EEG imaging with non-stationary source activations. *NeuroImage,* Volume 70, pp. 410-22.

Grosse-Wentrup, M., Liefhold, C., Gramann, K. & Buss, M., 2009. Beamforming in Noninvasive Brain-Computer Interfaces. *IEEE Transactions on Biomedical Engineering,* 56(4), pp. 1209-19.

Grosse-Wentrup, M. & Schölkopf, B., 2014. A brain-computer interface based on self-regulation of gamma-oscillations in the superior parietal cortex. *Journal of Neural Engineering,* 11(5), pp. 1-13.

Guger, C. et al., 2003. How Many People are Able to Operate an EEG-Based Brain-Computer Interface (BCI)?. *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* 11(2), pp. 145-147.

Hämäläinen, M. et al., 1993. Magnetoencephalography - theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics,* 65(2), pp. 413-97.

Hämäläinen, M., Lin, F.-H. & Mosher, J., 2010. Anatomically and Functionally Constrained Minimum-Norm Estimates. In: *MEG: An Introduction to Methods.* New York: Oxford University Press, Inc., pp. 186-215.

Hellyer, P. J. et al., 2014. The Control of Global Brain Dynamics: Opposing Actions of Frontoparietal Control and Default Mode Networks on Attention. Journal of Neuroscience, 34(2), pp. 451-61.

Jasper, H. H., 1958. Report on the Committee on Methods of Clinical Examination in Electroencephalography: 1957. *Electroencephalography and Clinical Neurophysiology,* 10(2), pp. 370-5.

Kamousi, B., Liu, Z. & He, B., 2005. Classification of Motor Imagery Tasks for Brain-Computer Interface Applications by Means of Two Equivalent Dipoles Analysis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* 13(2), pp. 166-71.

Kamousi, B., Nasiri, A. & He, B., 2007. Classification of motor imagery by means of cortical current density estimation and Von Neumann entropy. *Journal of Neural Engineering,* 4(2), pp. 17-25.

Krauledat, M., Schroder, M., Blankertz, B. & Müller, K.-R., 2007. Reducing Calibration Time for Brain-Computer Interfaces: A Clustering Approach. *Adv. Neural Inf. Process. Syst.,* pp. 753-60.

Lopes da Silva, F. H., 2010. Electrophysiological Basis of MEG Signals. In: P. C. Hansen, M. Kringelbach & R. Salmelin, eds. *MEG: An Introduction to Methods.* New York: Oxford University Press, Inc., pp. 1-23.

Lopes da Silva, F. H. & Van Rotterdam, A., 2011. Biophysical Aspects of EEG and Magnetoencephalogra Generation. In: D. L. Schomer & F. H. Lopes da Silva, eds. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields.* Philadelphia: Lippincott Williams & Wilkins, pp. 91-110.

Lotte, F., 2015. Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based Brain-Computer Interfaces. *Proceedings of the IEEE,* 103(6), pp. 871-90.

Lotte, F. & Guan, C., 2010. Learning from other subjects helps reducing brain-computer interface calibration time. *2010 IEEE International Conference on Acoustics Speech, and Signal Processing,* Volume 1, pp. 614-7.

Lotte, F., Lécuyer, A. & Arnaldi, B., 2009. FuRIA: An Inverse Solution Based Feature Extraction Algorithm Using Fuzzy Set Theory for Brain-Computer Interfaces. *IEEE Transactions on Signal Processing,* 57(8), pp. 3253-63.

Mattout, J., 2012. Brain-computer interfaces: a neuroscience paradigm of social interaction? A matter of perspective. *Frontiers in Human Neuroscience,* Volume 6, pp. 1-5.

O'Connell, R. G. et al., 2009. Uncovering the Neural Signature of Lapsing Attention: Electrophysiological Signals Predict Errors up to 20 s before They Occur. *Journal of Neuroscience,* 29(26), pp. 8604-11.

Pan, S. J. & Yang, Q., 2010. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering,* 22(10), pp. 1345-59.

Penfield, W. & Rasmussen, T., 1950. *The cerebral cortex of man.* New York, New York: The Macmillan Company.

Pfurtscheller, G. & Aranibar, A., 1979. Evaluation of event-related desynchronization (ERD) preceding and following voluntary self-paced movement. *Electroencephalography and Clinical Neurophysiology,* 46(2), pp. 138-46.

Qin, L., Ding, L. & He, B., 2004. Motor imagery classification by means of source analysis for brain-computer interface applications. *Journal of Neural Engineering,* 1(3), pp. 135-41.

Ramoser, H., Müller-Gerking, J. & Pfurtscheller, G., 2000. Optimal Spatial Filtering of Single Trial EEG During Imagined Hand Movement. *IEEE Transactions on Rehabilitation Engineering,* 8(4), pp. 441-446.

Schalk, G. et al., 2008. Brain-computer interfaces (BCIs): Detection instead of classification. *Journal of Neuroscience Methods,* 167(1), pp. 51-62.

Sutton, S., Braren, M. & Zubin, J., 1965. Evoked-Potential Correlates of Stimulus Uncertainty. *Science,* 150(3700), pp. 1187-8.

Tikhonov, A. & Arsenin, V., 1977. *Solutions of Ill-Posed Problems.* Washington D.C.: Winston & Sons.

van 't Ent, D., de Munch, J. & Kaas, A. L., 2001. A Fast Method to Derive Realistic BEM Models for E/MEG Source Reconstruction. *IEEE Transactions on Biomedical Engineering,* 48(12), pp. 1434-43.

van den Heuvel, M. P. & Hulshoff Pol, H. E., 2010. Exploring the brain network: a review on resting-state fMRI functional connectivity. *Eur. Neuropsychopharmacol.,* 20(8), pp. 519-34.

Vansteensel, M. J. et al., 2010. Brain-Computer Interfacing Based on Cognitive Control. *Annals of Neurology,* 67(6), pp. 809-16.

Vatta, F. et al., 2010. Realistic and Spherical Head Modeling for EEG Forward Problem Solution: A Comparative Cortex-Based Analysis. *Computational Intelligence and Neuroscience,* Volume 2010, pp. 1-11.

Vidal, J. J., 1973. Toward direct brain-computer communication. *Annual Review Biophysics and Bioengineering,* Volume 2, pp. 157-80.

Vidal, J. J., 1977. Real-time detection of brain events in EEG. *Proceedings of the IEEE,* 65(5), pp. 633-41.

Wander, J. & Rao, R. P., 2014. Brain-computer interfaces: a powerful tool for scientific inquiry. *Current Opinion in Neurobiology,* Volume 25, pp. 70-5.

Wolpaw, J. R., McFarland, D. J., Neat, G. W. & Forneris, C. A., 1991. An EEG-based brain-computer interface for cursor control. *Electroencephalography and Clinical Neurophysiology,* 78(3), pp. 252-9.

Wolpaw, J. R. & Wolpaw, E. W., 2012. Brain-computer interfaces: something new under the sun. In: J. R. Wolpaw & E. W. Wolpaw, eds. *Brain-Computer Interfaces: Principles and Practice.* New York: Oxford University Press, pp. 3-12.

Woolrich, M. W. et al., 2013. Dynamic state allocation for MEG source reconstruction. *NeuroImage,* Volume 77, pp. 77-92.

Yuan, H., Doud, A., Gururajan, A. & He, B., 2008. Cortical Imaging of Event-Related (de)Synchronization During Online Control of Brain-Computer Interface Using Minimum-Norm Estimates in Frequency Domain. *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* 16(5), pp. 425-31.

Zich, C., De Vos, M., Kranczioch, C. & Debener, S., 2015. Wireless EEG with individualized channel layout enables efficient motor imagery training. *Clinical Neurophysiology,* 126(4), pp. 698-710.

Chapter 2. Incorporating modern neuroscience findings to improve brain-computer interfaces: tracking auditory attention

This chapter was published in the Journal of Neural Engineering by Mark Wronkiewicz, Eric Larson, and Adrian KC Lee with the title "Incorporating modern neuroscience findings to improve brain–computer interfaces: tracking auditory attention" here: http://dx.doi.org/10.1088/1741-2560/13/5/056017.

2.1 Abstract

Objective. Brain-computer interface (BCI) technology allows users to generate actions based solely on their brain signals. However, current non-invasive BCIs generally classify brain activity recorded from surface electroencephalography (EEG) electrodes, which can hinder the application of findings from modern neuroscience research. *Approach.* In this study, we use source imaging – a neuroimaging technique that projects EEG signals onto the surface of the brain – in a BCI classification framework. This allowed us to incorporate prior research from functional neuroimaging to target activity from a cortical region involved in auditory attention. *Main Results.* Classifiers trained to detect attention switches performed better with source imaging projections than with EEG sensor signals. Within source imaging, including subject-specific anatomical MRI information (instead of using a generic head model) further improved classification performance. This source-based strategy also reduced accuracy variability across three dimensionality reduction techniques – a major design choice in most BCIs. *Significance.* Our work shows that source imaging provides clear quantitative and qualitative advantages to BCIs and highlights the value of incorporating modern neuroscience knowledge and methods into BCI systems.

2.2 Introduction

Brain-computer interfaces (BCIs) attempt to interpret patterns of brain activity in order to generate artificial output. A great deal of research is directed toward realizing their potential to rehabilitate or

restore communication after nervous system damage (Wolpaw, et al., 2002) and permit augmentation of healthy individuals (Blankertz, et al., 2012). However, BCIs face a number of challenges related to reliability, accuracy, and limited exploration of new paradigms that continue to stifle their progress in clinical and commercial applications (see (Wolpaw & Wolpaw, 2012) for review). In electroencephalography (EEG) BCIs, several of these issues stem from how the brain recordings are analyzed; recorded signals are typically processed and interpreted in the "sensor space" (i.e., at the electrode level on the surface of the scalp).

There are a number challenges associated with the sensor space that impede the advancement of BCI research. First, multiple electrodes record activity from any one source – a phenomenon attributed to volume conduction. In practice, this causes spatial averaging (i.e., smearing) of the recorded potentials across electrodes, which hinders the classification of brain activity because of the induced signal correlations. This issue is exacerbated by the variety of tissues in the human head – each with different conductivity properties (Nunez, et al., 1997; Schoffelen & Gross, 2009; van den Broek, et al., 1988). Second, there is variability in head tissue morphology (e.g., skull and scalp), cortical geometry (e.g., brain folding; Friston, et al., 1995), and location of electrodes across subjects and experiments. For these reasons, it is difficult to choose a control signal that will generalize across subjects for any given BCI task. Third, implementing BCI systems that operate on sensor space signals makes it difficult to incorporate the wealth of functional localization information that is continuously emerging from brain mapping research. Discoveries from this area could serve as prior knowledge to help guide the construction of features for BCI classification tasks. These findings, however, are often reported in the cortical domain, so their relationship to the sensor space is not straightforward.

Source imaging is a widely used neuroimaging technique applicable to EEG data (see Baillet, et al., 2001; Baillet, 2010; Hämäläinen, et al., 1993 for review) that alleviates the aforementioned problems, yet remains surprisingly underutilized in BCIs. It estimates the sources of activity on the cortex (also referred to as the "source space") using noninvasive electrophysiological data recorded in the sensor space. Although it provides lower spatial resolution than invasive methods like electrocorticography or

implantable multielectrode arrays, source imaging does not carry surgical risks. Source imaging works by first modeling the "forward problem," which estimates how signals from distributed cortical sources are distorted as they pass through the different head layers to the sensors (Figure 2.1). With the forward problem modeled, an inverse operator is computed, which attempts to account for this distortion and allows an estimation of the active cortical areas that produced the observed EEG signals. Since source imaging incorporates a head model and 3D electrode locations, it also partially normalizes variations in head anatomy, brain anatomy, and electrode positioning across subjects and/or sessions. Finally, source imaging provides a principled path to include prior knowledge from modern neuroscience research; for example, any neuroimaging study that has localized a task-specific activation pattern could be used to objectively choose a cortical region-of-interest (ROI) for BCI control. Findings from systems-level electrophysiology in humans or animals are also more readily applicable in the source space. Thus, source imaging may represent an opportunity to choose priors from an ever-growing body of neuroscience literature thereby better informing choices that depend on a BCI expert.

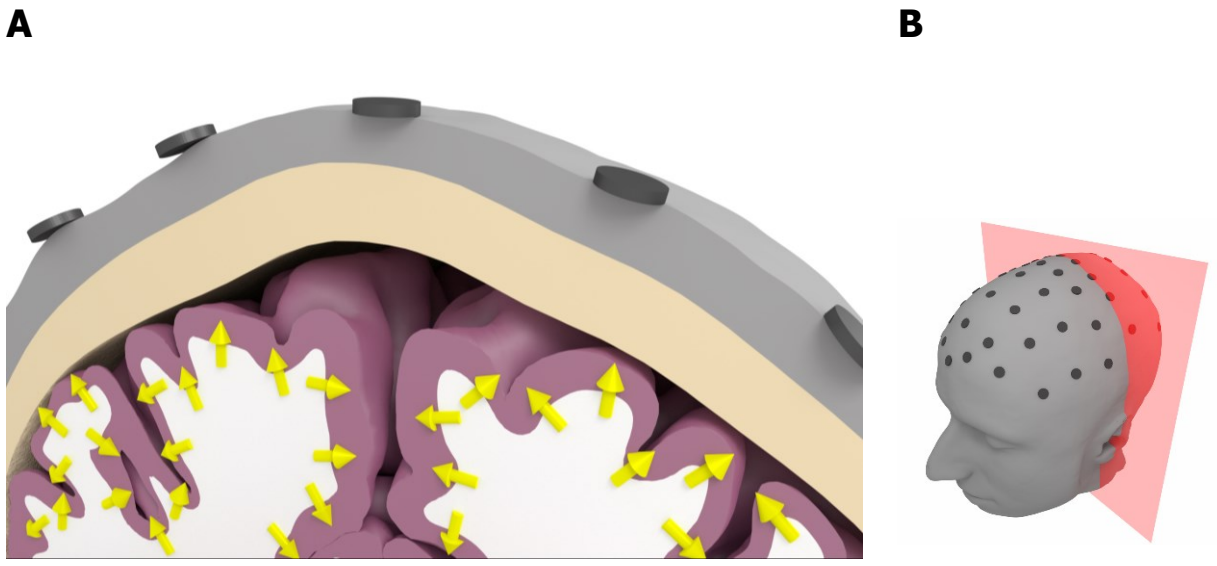

Figure 2.1 Illustration of source and sensor space in a cross section of one subject's head reconstruction. **(A)** In source imaging, current is estimated at modeled current dipoles (yellow arrows)

that are distributed in the source space from data recorded by electrodes in the sensor space (black disks). The head layers depicted here include the scalp (gray), skull (off-white), gray matter (pink), and white matter (white). The current dipoles are typically positioned at (and normal to) the gray/white matter boundary and are thought to reflect activity in patches of apical dendrites belonging to pyramidal neurons. **(B)** Diagram showing the plane of the coronal cross section used in **(A)**.

While previous BCI studies classified brain activity in the source space using both offline (Qin, et al., 2004; Menendez, et al., 2005; Congedo, et al., 2006; Yuan, et al., 2008; Kamousi, et al., 2005; Kamousi, et al., 2007; Edelman, et al., 2015; Noirhomme, et al., 2008) and online analysis (Cincotti, et al., 2008; Grosse-Wentrup, et al., 2009; Grosse-Wentrup & Schölkopf, 2014), these source estimation studies represent a very narrow portion of the total BCI literature. Further, they all focus on hand-motor activity – a well-studied signal within the BCI community. This is understandable given the usefulness of sensorimotor activity for rehabilitation and our relatively deep understanding of primary motor neurophysiology. However, exploratory work that examines non-invasive BCI control utilizing more cognitive regions or network activity is scarce (Besserve, et al., 2011; Grosse-Wentrup & Schölkopf, 2014). Moreover, many insights gained in functional localization research have not been incorporated into new BCI paradigms. These neuroimaging findings could therefore aid in the development of BCIs that are not restricted to the field of hand-motor rehabilitation; however, it is still unclear how this type of information might be exploited in a BCI context.

Toward this goal of broadening BCI paradigms via the source space, we first conducted a simulation study to demonstrate the variability and diffuse nature of sensor space recordings. We then focused on detecting shifts in attention with both simulated and experimental data collected in a previous auditory attention study (Larson & Lee, 2014), using the prior knowledge that the right temporoparietal junction (RTPJ) is significantly more active when switching auditory spatial attention (Larson & Lee, 2013). Given this research prior, we selected cortical activity from only this ROI when classifying offline whether or not a subject voluntarily switched attention. We explored the RTPJ because a quantitative measure of

attention is relevant in a number of situations including BCI-based hearing aids or tracking the attention of someone in a classroom, driving a car, conversing in a café, etc. Compared to a naive sensor-based approach, we hypothesized that a source-based approach using this neuroimaging information would simplify the signal analysis pipeline and improve classification accuracy. Toward the aim of pursuing new BCIs, the RTPJ also falls outside the set of regions typically used for BCI control. Finally, we assessed the practicality of this source-based BCI approach by classifying source space activity obtained from a generic head model and an individualized head model that incorporated each subject's anatomical MRI information.

2.3 Methods

The code for this analysis is available at https://github.com/LABSN-pubs/2016-JNE-source-based-BCI

2.3.1. Source imaging: spherical vs. boundary element model

This study used source imaging according to the MNE-Python algorithm described in Sections 1.1 and 1.2. For the source-based approach in this work, we tested two popular approaches to the source imaging forward model: generic or individualized, which exclude or include individualized structural MRI information, respectively. Forward models are comprised of two parts: a conductivity model (which defined how the different head tissues conduct neural activity) and a distributed source model (which defined the location and orientations of simulated current dipoles). Comparing generic and individualized forward models allowed us to evaluate any advantages that individual anatomical MRI information might confer in a source-based BCI.

Spherical conductivity models represent a generic method of formulating the forward conductivity model and assume that the head tissues are comprised of concentric spherical shells each with a different conductivity (Figure 2.2A). Here, the spheres were aligned to the 3D electrode positions and anatomical points mentioned previously, and MNE-Python's default values for the brain, CSF, skull, and scalp relative

shell radii (0.9, 0.92, 0.97, 1.0, respectively) and relative conductivities (0.33, 1.0, 0.004, 0.33, respectively) were used. Along with this forward model, we used a template brain (FreeSurfer's "fsaverage") when modeling the distributed current dipoles at the gray/white matter boundary. Spherical head models do not rely on MRI information, so this method is often chosen when an MRI scan is impractical or too costly. In general, this strategy is the most straightforward method to implement source imaging and was used in the majority of previous source-based BCIs (Grosse-Wentrup, et al., 2009; Qin, et al., 2004; Edelman, et al., 2015; Kamousi, et al., 2005; Kamousi, et al., 2007; Lotte, et al., 2009; Menendez, et al., 2005; Congedo, et al., 2006; Ahn, et al., 2012; Noirhomme, et al., 2008).

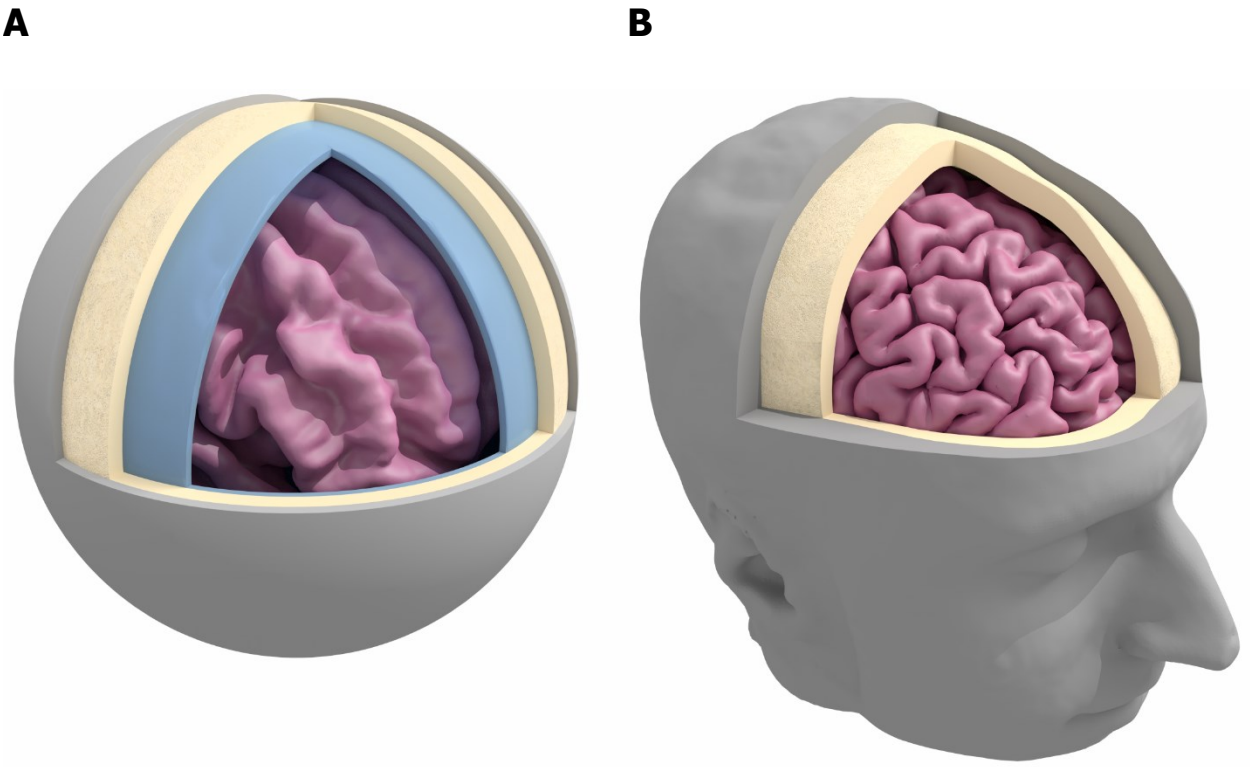

Figure 2.2 Illustration of the two most common conductivity models in source imaging. **(A)** Generic model: The spherical conductivity shells (starting with the outermost) include the scalp (gray), skull (off-white and textured), and cerebrospinal fluid (blue) centered on the subject's head. The innermost shell,

which models brain conductivity and encompasses the current dipoles, is not shown so that cortical surface of the template brain (FreeSurfer's "fsaverage") is visible (pink), which is used to define the distributed source model. **(B)** Individualized model: The boundary element model (BEM) conductivity shells (starting with the outermost) include the scalp and skull. They were reconstructed from a structural (T1-weighted) MRI scan allowing an anatomically accurate head model. Again, the innermost shell is not shown so that the individual's brain reconstruction, which defines the distributed source model, is visible. In both the generic and individualized source imaging approaches, dipoles were modeled at the gray/white matter boundary.

The boundary element model (BEM) is an individualized approach for constructing a conductivity model to compute the forward matrix (Figure 2.2B). Because it incorporates a subject's 3D head morphology, this approach allows a more accurate estimation of cortical activity than the generic spherical model (Baillet, et al., 2001). BEMs are generally built from structural (T1-weighted) MRI information; they require an anatomically accurate segmentation of the head into different layers (accomplished here with FreeSurfer, http://surfer.nmr.mgh.harvard.edu/; Dale, et al., 1999) and were used in some previous source-based BCIs (Yuan, et al., 2008; Cincotti, et al., 2008; Besserve, et al., 2011). The same relative conductivities were used as in the spherical model (0.33, 0.004, and 0.33) for the brain, skull, and scalp layers, respectively. In this study, MRI information was obtained using one structural multi-echo magnetization prepared rapid gradient echo scan as well as two fast low-angle shot scans at 5 and 30 degrees obtained with a 3T Philips scanner. We also reconstructed each individual's gray and white matter surfaces from this MRI information to place the modeled current dipoles normal to gray/white matter boundary facing outward.

2.3.2.1 Simulation study: anatomical ROIs

We conducted a controlled simulation study to quantify the electrical influence of a few important cortical regions on the sensor space (i.e., the EEG electrodes). Regions were chosen based on relevance to the task here or other BCI paradigms, their known localization, and because they illustrate the range of ways

cortical activity can manifest in the sensor space. Two of these anatomically defined regions were the RTPJ (using the anatomical definition from Larson & Lee, 2013 based on previous studies; Mitchell, 2008) and primary auditory cortex (anterior transverse temporal gyrus; Destrieux, et al., 2010). We also included the hand-motor region, which was defined by taking the center from a hand motor functional localization meta-study (Witt, et al., 2008) and selecting all current dipoles within 10 mm of this centroid (along the cortical surface) that were also part of FreeSurfer's precentral gyrus ROI (Destrieux, et al., 2010). For each ROI tested, dipoles were assigned unit current (flowing outward and normal to the gray/white matter boundary) if they were within the ROI and zero current otherwise. Using the forward solution, we calculated the normalized influence of each ROI on the EEG sensors. This provided a map of activity in the sensor space for each of 10 subjects. From these maps, we calculated the mean activity map for each ROI and the equipotential contour corresponding to 50% of the maximum sensor value for each subject.

2.3.2.2 Simulation study: detecting RTPJ activity with eye blink artifacts

To compare sensor- and source-based classification strategies, we simulated elevated and baseline levels of activity within the RTPJ and carried out a binary classification task. In trials with elevated activity, dipoles within the RTPJ were assigned positive current, and the current magnitude was obtained by normalizing 10 nA·m of current by the number of dipoles in RTPJ (separately for each subject) to achieve a uniform current density. In baseline trials, all dipoles were assigned zero current. For the elevated and baseline trial classes, 40 trials worth of activity were generated with each consisting of 600 time points (to match the experimentally recorded data outlined in Section 2.3). EEG activity was then obtained by multiplying the cortical activity by the forward matrix and adding temporally white Gaussian noise for each subject with spatial covariance equivalent to that of the EEG data recorded during the inter-trial intervals of the task described in Section 2.3. The noise level was scaled to obtain an SNR of -10 dB, which was consistent with similar studies (Wronkiewicz, et al., 2015) and yielded a performance comparable to the experimentally recorded data (Section 2.3). Here, dB SNR is defined as in the MNE-Python toolbox (Gramfort, et al., 2013) as $10 \times \log_{10}\left(\sigma^2_{\text{signal}}/\sigma^2_{\text{noise}}\right)$. Since non-stationary physiological

noise is a key concern in BCI signal processing, we also added eye blink artifacts to the EEG data (and purposefully avoided removing them during preprocessing). Blink times were drawn from an inhomogeneous Poisson process where the blink rates oscillated between 9 and 21.5 blinks/minute – a relatively high but physiologically plausible range (Bentivoglio, et al., 1997).

2.3.3 Experimental study: tracking auditory attention

We explored a source imaging framework with experimental data collected as part of a previous study investigating cortical activity during voluntary (i.e., top-down) modulation of spatial attention (Larson & Lee, 2014). All subjects gave informed consent according to the procedures approved by the University of Washington. In this task, 10 subjects listened to simultaneously spoken target and a masker auditory streams each of which consisted of a series of spoken letters. Both streams were made spatially distinct to induce a leftward or rightward spatial percept (using an interaural time delay of ± 300 ·s). Subjects were cued before each trial began whether to maintain attention to a single stream throughout the trial or switch to the other stream halfway through (during a 600 ms gap period in the spoken letters). EEG data from this switching period were used to classify whether or not a subject switched attention. Subjects were also instructed to visually fixate on a central dot at all times. At the end of each trial, they were required to report the number of times a particular letter occurred in the target stream, and only trials with the correct response were analyzed here. For this reason, the number of trials varied for each subject (mean=38, SD=10.92).

2.3.4 Preprocessing, feature construction, and classification

The goal of both the simulation and experimental studies was to detect the maintenance or switching of spatial auditory attention in single trials. In each study, we used the same data preprocessed in the same way for both sensor and source space approaches as shown in Figure 2.3. In only the study using experimental data, EOG and ECG artifacts were first removed using signal space projection (Uusitalo & Ilmoniemi, 1997). In both studies, data from the 600 ms switching period was selected for all ~60 electrodes and downsampled to 50 ms bins since previous RTPJ research indicates a relatively stable increase in current density during switching (Larson & Lee, 2013). For the sensor space approach,

dimensionality reduction was applied directly to this sensor space data, which is typical of many BCI experiments. For the source space approaches, we used inverse imaging to estimate cortical activity using both generic and individualized forward models. Once in source space, we subselected current dipoles from the RTPJ region and concatenated their estimated currents to create a feature vector prior to dimensionality reduction. This dipole selection step reflects the functional neuroimaging prior that the RTPJ region is significantly more active when switching attention (Larson & Lee, 2013). The RTPJ ROI was defined anatomically from previous non-auditory studies of the RTPJ (Mitchell, 2008; see Larson & Lee, 2013 for details).

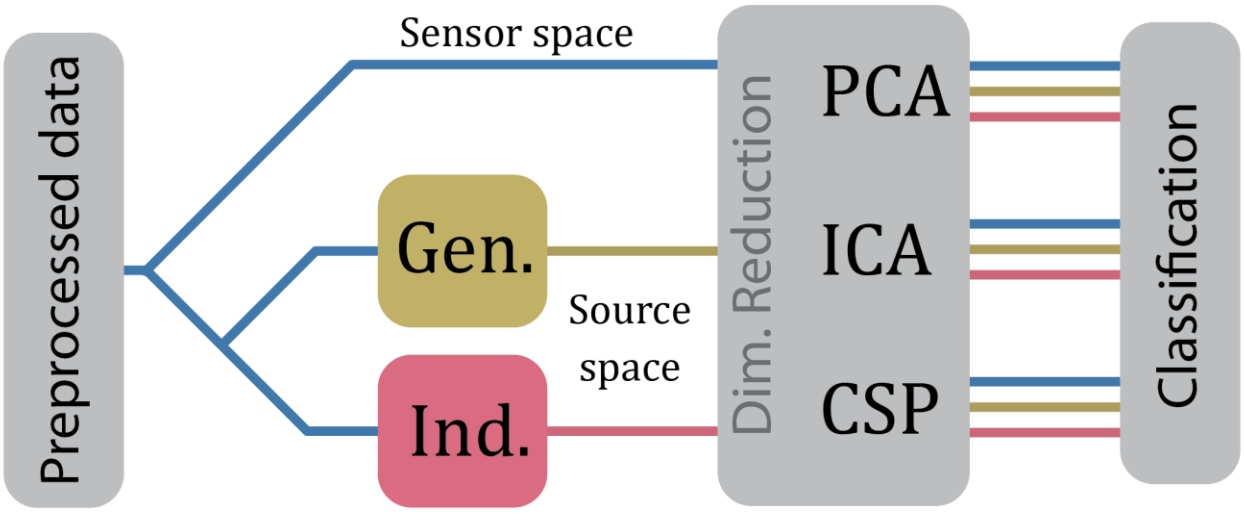

Figure 2.3 Overview of the experimental organization. The same preprocessed data was analyzed in the sensor space (blue) as well as in the source space with generic and individualized source imaging approaches (yellow and pink, respectively). In the source imaging approaches, we selected only dipoles belonging to the RTPJ for further processing (reflecting our neuroimaging prior concerning the RTPJ as a locus of activity during attention switches). For all three approaches, the same dimensionality reduction techniques (principal component analysis, independent component analysis, and common spatial

patterns; PCA, ICA, and CSP, respectively) were applied before support vector machine (SVM) classification resulting in nine experimental conditions.

We automated the parameter selection in the dimensionality reduction and classification schemes to establish a fair comparison across all signal analysis approaches. For both sensor and source space strategies, we applied dimensionality reduction with three techniques (Figure 2.3): principal component analysis (PCA), extended infomax independent component analysis (ICA), and common-spatial patterns (CSP). We swept over a wide range of parameters identically across source and sensor space methods. For PCA, we tested the inclusion of the top 1 to 5 components after ranking them by variance captured. For ICA, we used PCA to reduce the rank of our data down to components containing a cumulative proportion of explained variance (0.9, 0.95, 0.995 for data with simulated RTPJ activation; 0.85, 0.9, 0.95, and 0.99 for the experimental data) before computing the unmixing matrix. This is a common approach to reduce the data dimensionality prior to applying ICA (Krusienski, et al., 2012) and avoids the problem of subjectively picking the best component(s). For CSP, we tested 1 to 5 components corresponding to the most extreme eigenvalue pairs as well as a range of standard covariance regularization methods. These included Ledoit-Wolf shrinkage (Ledoit & Wolf, 2004), Oracle Approximating Shrinkage (Chen, et al., 2010), and fixed shrinkage values of 0.15, 0.2, and 0.25 in simulated data and 0.05, 0.1, 0.15, 0.2, and 0.25 using the experimental data. The exact ranges for ICA and CSP parameters differed slightly between simulation and experimental studies to ensure convergence of the respective algorithms. Features were classified using a support vector machine (SVM) classifier because it is widely used and relatively robust to high dimensional data. Here, we varied SVM parameters C and ・ over a wide range of reasonable values (in order of magnitude steps) and used a radial basis function kernel. C, which controls the complexity of the decision boundary surface, was varied in orders of magnitude from 10^{-4} to 10^{5} while ・, which controls the radius of influence for the support vectors, was varied from 10^{-7} to 10^{3}. We used 10-fold cross validation to stabilize accuracy results. To obtain final classification accuracies for each subject in each of the nine experimental conditions, we first averaged

cross validation scores and then selected the maximum score across the two SVM parameters on a subject-by-subject basis, typical of online BCI experiments. Next, we selected the dimensionality reduction parameters that gave the highest mean group accuracy to ensure a unified dimensionality reduction technique in each condition.

2.3.5 Statistics

We used a linear mixed effects regression model with two categorical predictors to statistically compare the classification approaches using the Statsmodels open source software package (www.statsmodels.org; Seabold & Perktold, 2010). The two categories were the data formulation (sensor, generic source, and individualized source) and dimensionality reduction method (PCA, ICA, CSP; Figure 2.3) with a by-subject random intercept. Both predictors were dummy-coded with "individualized source" and "PCA" as baselines. Here, the null hypothesis was that the data formulation approaches and decomposition methods did not affect classification accuracy (i.e., that model coefficients of all categorical predictors were zero). In the experimental study, we included the interaction between the data formulation and dimensionality reduction method categories to test if the dimensionality reduction techniques statistically modulated the effect of the data formulation. Note that this statistical model provides the same information as a 2-way within-subjects repeated measures ANOVA but has preplanned *t*-tests to the baseline categories built in.

2.4 Results

2.4.1 Simulation study: anatomical ROIs

We first investigated how volume conduction, head anatomy, and brain anatomy affect the composition of sensor space recordings. To accomplish this, we simulated unit activity for 10 subjects in three anatomically defined regions relevant to BCIs (Figure 2.4, top row). The activity in these regions was projected to the sensor space (without adding simulated noise) using anatomically accurate boundary element conduction models and individualized source models (as described in the Methods) and then

normalized to the maximum absolute electrode potential. The mean sensor activations illustrate the distributed pattern of potentials that arises in the sensor space even for small punctate ROIs (Figure 2.4, middle row). This simulation assumes that the rest of the brain is silent and ignores other possible noise sources (such as electromyography or electrooculography artifacts, bad electrode contacts, etc.), and yet there is still clear variability in the activity pattern across subjects at the sensor level (Figure 2.4, bottom row). This has implications for BCI studies where features are constructed from activity at a single pre-selected electrode (or small set of electrodes), as discussed in Section 4.1.

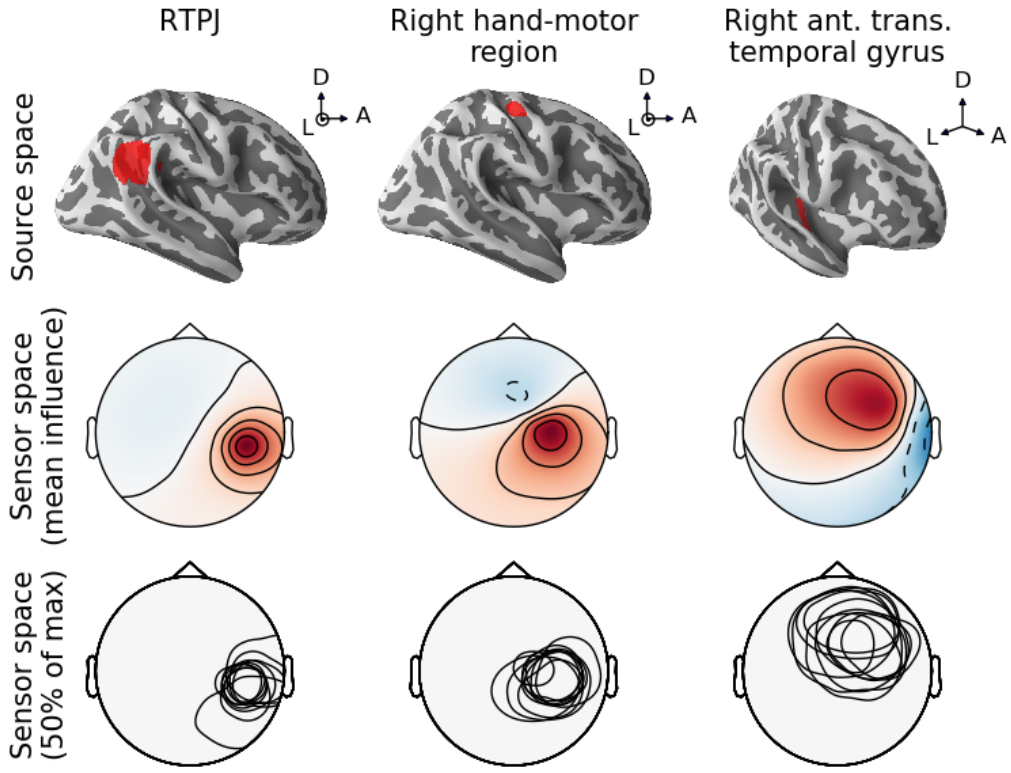

Figure 2.4 Intersubject variability of sensor space topology from identical cortical activations. Activity was simulated for 10 subjects in three BCI-relevant ROIs: RTPJ, right hand-motor area, and right primary auditory area (left, middle, and right columns, respectively). **Top row:** Exemplar ROIs plotted on a template brain (FreeSurfer's "fsaverage", partially inflated; see Methods for details). Coordinate frames indicate lateral, anterior, and dorsal directions. **Middle row:** The mean sensor space activation across 10 subject-specific brain models normalized to the maximum value. **Bottom row:** For each subject, an

equipotential line shows the boundary where sensor level influence was 50% of the maximum (absolute) value. The simulations here assumed unit current at dipoles within the ROI, zero current at all other dipoles, and noiseless conditions. Abbreviations: A-anterior, D-dorsal, L-lateral. Circle with center dot denotes axis orthogonal to and out of the page.

2.4.2 Simulation study: detecting RTPJ activity with eye blink artifacts

To compare sensor- and source-based approaches, we used both strategies to classify whether RTPJ activity was elevated or at baseline in simulated data. In addition to simulating activity in RTPJ, we included Gaussian sensor noise with spatial covariance from experimental measurements (see Section 2.2.3) and heavily contaminated the data with eye blink artifacts. We did not attempt to remove the eye blinks in order to assess the robustness of source imaging to nonstationary and improperly suppressed artifacts. On average, we found that the individualized source imaging approach performed 5.4% higher than the sensor space approach ($p = 0.00161$) and 1.7% higher than the generic source imaging approach ($p = 0.332$; Figure 2.5).

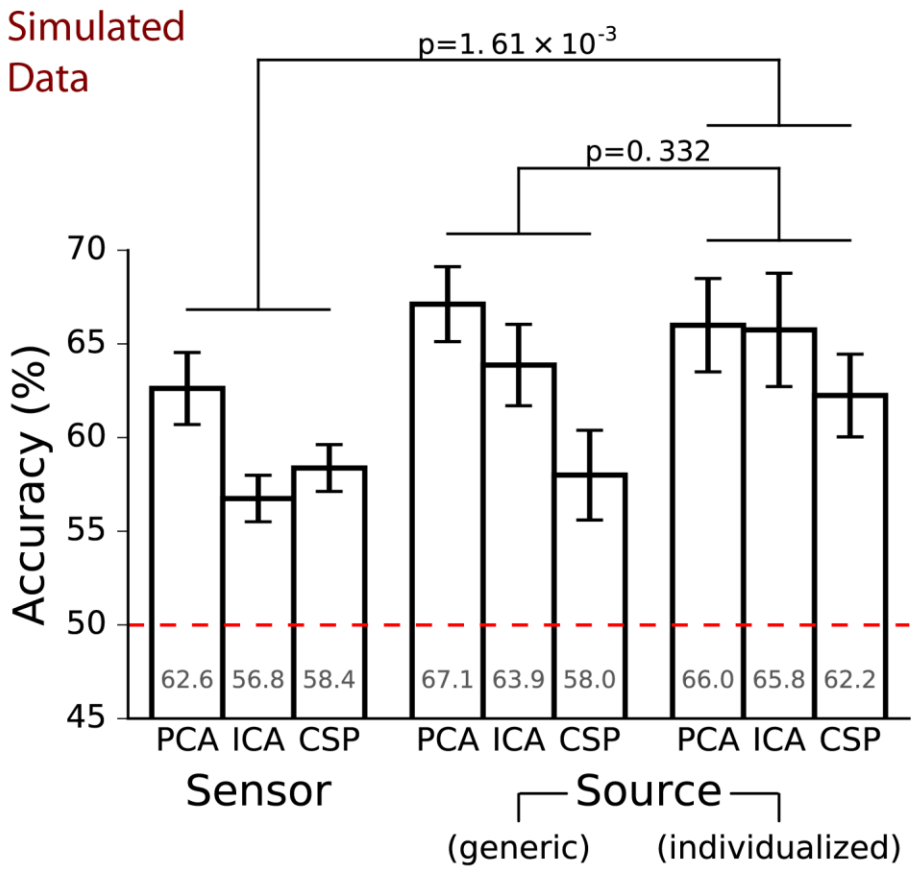

Figure 2.5 Single-trial classification accuracy of *simulated* RTPJ activation data heavily contaminated with eye blink artifacts using sensor space features and two forms of source space features (generic and individualized). Within each category, three dimensionality reduction techniques (PCA, ICA, and CSP) were tested over an identical range of parameters. The individualized source imaging approach gave statistically higher accuracy than the sensor approach, but not the generic source imaging approach (see the Methods for details). Individualized source imaging was selected as the statistical baseline, which prevented a direct statistical comparison between the sensor and generic source imaging approaches. Error bars are ± SEM and the red dashed line indicates the theoretical chance accuracy.

2.4.3 Experimental study: tracking auditory attention

To compare the sensor- and source-based approaches using real EEG data, we tested both strategies when classifying whether or not a subject switched attention during a previously recorded auditory attention experiment. On average, the classification accuracy for the individualized source imaging approach was 7.2% higher than a sensor approach ($p = 1.49 \times 10^{-6}$) and 4.1% higher than a generic source imaging approach ($p = 3.66 \times 10^{-3}$; Figure 2.6). When comparing the highest performing dimensionality reduction strategy in each category, the best individualized source imaging score (using PCA) was 5.2% higher than the best sensor space approach (using CSP) and 3.4% higher than the best generic source imaging approach (using CSP). The trend of CSP outperforming the other dimensionality reduction techniques was neutralized when using individualized source space analysis (i.e., the interaction between the CSP dimensionality reduction method and the individualized source model was significant, p=0.025).

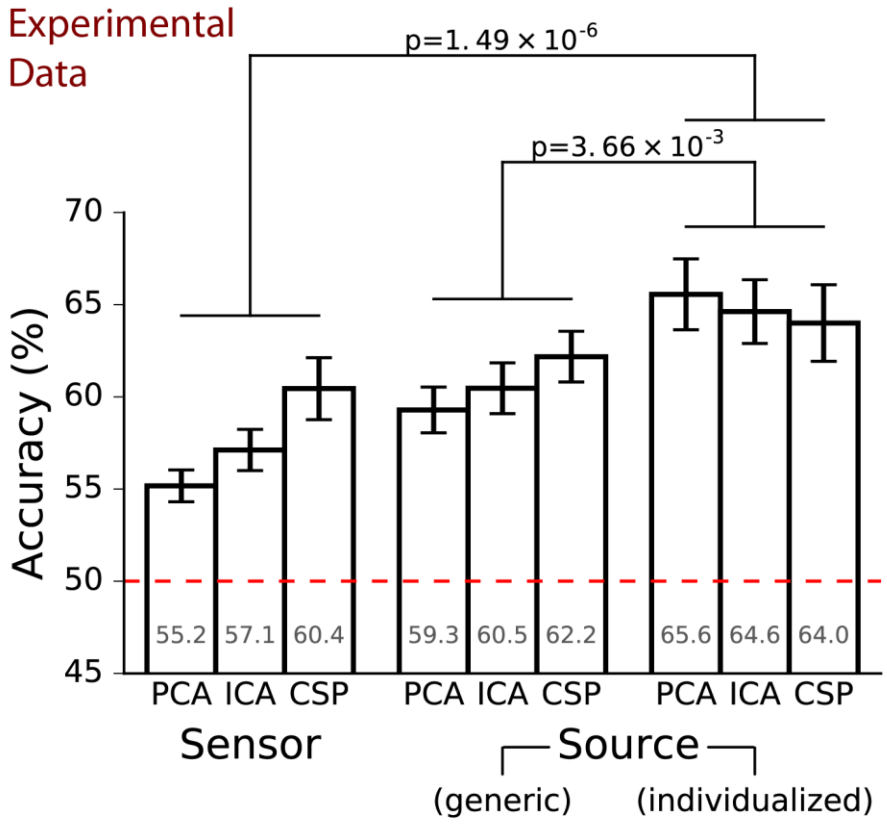

Figure 2.6 Single-trial classification accuracy of *experimentally recorded* attention switching data using sensor space features and two forms of source space features (generic and individualized). Within each category, three dimensionality reduction techniques (PCA, ICA, and CSP) were tested over an identical range of parameters. The individualized source imaging approach gave statistically higher accuracy than both the sensor and generic source imaging approaches (see the Methods for details). Individualized source imaging was selected as the statistical baseline, which prevented a direct statistical comparison between the sensor and generic source imaging approaches. Error bars are ± SEM and the red dashed line indicates the theoretical chance accuracy.

2.5 Discussion

2.5.1 Current state of BCIs

The field of noninvasive BCIs continues to face a number challenges that have slowed the field's recent progress, which is particularly evident when reviewing the age of the three canonical BCI paradigms still widely used today. The first BCI controlled by sensorimotor rhythms was published 25 years ago by Wolpaw, et al., 1991, and the underlying event-related desynchronization was characterized as reliably detectable over 35 years ago by Pfurtscheller & Aranibar, 1979. Prior to even these results, the foundations for primary sensory and motor neuroanatomy were outlined in the Penfield homunculus in 1937 and the more popular 1950 rendition that is still taught today (Penfield & Boldrey, 1937; Penfield & Rasmussen, 1950). The "P300" BCI, named for the positive deflection in parietal electrodes that peaks about 300 ms after a desired oddball stimulus, was originally published by Farwell & Donchin, 1988, but the P300 event-related potential was identified over 20 years prior (Chapman & Bragdon, 1964; Sutton, et al., 1965). BCIs based on visually evoked potentials were outlined in Vidal, 1973 and subsequently tested in Vidal, 1977. While these studies represent the bedrock of noninvasive BCI research, the BCI field has yet to capitalize on this technology's clinical or augmentative potential despite decades of revision and optimization.

One core reason for this hindrance is rooted in the analysis of brain data in the sensor space – a common strategy in the BCI literature. Our simulation (Figure 2.4) shows that activity from the same cortical region can manifest very differently in the sensor space – even when sources are spatially punctate and in noiseless conditions – due to variability across subjects' head and brain anatomy. This should serve as a cautionary finding for BCI work where an electrode or set of electrodes is preselected (e.g., choosing C3/C4 for motor-imagery or Pz for P300). Put simply, the composition of brain activity recorded at each electrode is likely to differ especially when comparing across subjects. Moreover, brain activity – even when spatially focal – is measurable across a wide area of the scalp, so choosing data only from the maximally sensitive electrode will discard useful information.

When instead analyzing data from a large number of electrodes, dimensionality reduction techniques like PCA, ICA, and CSP are often used to distill sensor-level data, which suffers from intersubject spatial variability and noise, into more informative signals. These dimensionality reduction techniques project a data set so that it conforms to some statistical property. The specific property depends on the algorithm, but they all produce a vector or matrix linearly remapping the data (see Parra, et al., 2005 for a unified discussion). Because the computation is data-driven, the projections will differ across each study and typically, even across each subject's recording within that study. Given the engineering goal of best classifying activity from each independent set of data, this approach has proven effective. However, it limits the generalizability of any BCI experiment because the projection for one subject tells us little about the next – the statistically optimal projection will differ across subjects making it difficult to theoretically or neuroscientifically ground outcomes between studies. In other words, we can expect these experiments to validate a dimensionality reduction technique or a signal-processing pipeline, but it is much more difficult to capitalize on the rapid progress in systems neuroscience with this approach. Any BCI strategy that incorporates knowledge from advances in neuroscience is likely to produce a more efficient and accurate system – a point identified by Makeig et al., 2012.

2.5.2 Source imaging for BCIs

Through source imaging, noninvasive EEG data is projected into the same space (or reference frame) used by many areas of neuroscience. In contrast to algorithmic projections (e.g., PCA, ICA, and CSP), this facilitates the ability to exploit both findings and methodology from related research in a number of ways. First, source imaging provides a principled path to deepen the connection between BCIs and neuroscience; it allows BCI systems to operate directly on estimates of cortical activity, rather than representations of sensor activity projected into mathematically efficient (but anatomically irrelevant) spaces. We explored here how this permits the application of neuroimaging research – an ever-growing source of *a priori* information – to guide feature construction for new BCI paradigms and target the most informative brain activity. Note that this prior neuroimaging information comes from independent studies with explicit goals of, for example, identifying task-relevant cortical ROIs. Therefore, it is reasonable to expect that a strategy incorporating this information may generalize better than recalculating data-driven projections of sensor space data for each subject. Second, it should also be possible to leverage findings from electrophysiological research (including invasive work) obtained through both humans and animals models. Again, by projecting EEG data into the source space, we can systematically take advantage of recent findings to improve BCI systems (e.g., when detecting localized high gamma activity in EEG; Ball, et al., 2008; Darvas, et al., 2010). Third, source imaging can also be used to morph BCI training data across subjects. This is a common task in neuroimaging studies when displaying average activity from multiple subjects on a single template brain. Applied to BCI research, however, this "transfer learning" technique is focused on eliminating lengthy calibration session needed in most BCIs (for review, see Lotte, 2015). Source-based transfer learning has been shown to better account for intersubject variability and could train a BCI classifier using only data transferred from other subjects with better accuracy than the standard subject-specific approach (Wronkiewicz, et al., 2015). Here, we explored some theoretical and practical advantages of a source space approach to BCIs.

Our goal was to validate a source imaging framework for BCIs. We found that a source-based approach could capitalize on a recent finding from human neuroimaging to detect voluntary attention modulation better than a sensor-based approach. This was true for simulated data contaminated with eye blink

artifacts (Figure 2.5) and experimentally recorded data during an attention switching task (Figure 2.6). The accuracy increase was observed when using both generic and individualized forward models (i.e., regardless of whether structural MRI information was available or not). While most previous source-based BCI studies focused on hand-motor imagery, they report gains over the standard sensor-based approach (from low single digits to over ten percent) consistent with our accuracy improvement of about five and seven percent in the simulated and experimental data, respectively (Besserve, et al., 2011; Cincotti, et al., 2008; Grosse-Wentrup, et al., 2009; Edelman, et al., 2015; Menendez, et al., 2005; Ahn, et al., 2012; Lotte, et al., 2009; Grosse-Wentrup & Schölkopf, 2014; Noirhomme, et al., 2008). Therefore, this demonstrates one strategy of utilizing modern functional neuroimaging to guide the development of new BCI paradigms and boost accuracy.

The classification advantages we saw when using source imaging were likely due to two factors. First, source imaging permits the objective inclusion of human neuroimaging information; in this case, the recent finding that RTPJ has a higher cortical current density while a subject is switching spatial auditory attention (Larson & Lee, 2013). Second, source imaging eliminates some spatial smearing and intersubject variability that affect the composition of neural activity when it is recorded in the sensor space (Figure 2.4). As mentioned previously, the use of a head model brings additional information (e.g., concerning variability in head anatomy, cortical folding, and electrode positioning) that complements the EEG recordings. Source imaging thus works to simplify the feature selection step and improve classification accuracy by bolstering the discriminability of the chosen signal feature. Moreover, our simulation experiment suggests that source imaging maintained its advantage even when the data were heavily contaminated with eye blink artifacts. The strategy employed here also uses a neuroscience-driven selection of a cortical region to reduce the impact of subjective choices common to BCIs (e.g., picking the best electrode(s) or signal feature), which lessens the need for a dedicated BCI expert and facilitates reproducibility.

Source imaging also appeared to equalize the effect of different dimensionality reduction techniques (Figure 2.6, right) observed in the sensor space (Figure 2.6, left). Source imaging may thus circumvent

the discussion concerning the best method for condensing high dimensionality electrophysiological data in BCIs (despite necessitating other parameters related to source imaging as discussed in Section 2.5.3). This equalizing effect was not as clear in the simulation study (Figure 2.5) possibly because of the heavy eye blink artifacts, to which CSP is particularly vulnerable (Bashashati, et al., 2007). Furthermore, we found a statistically significant interaction between CSP and the different signal processing approaches in the experimental data indicating that the relative superiority of CSP was neutralized in the source space. This is likely because CSP is a supervised dimensionality reduction algorithm that finds spatial weights to maximize the ratio of variance from two known trial classes. More generally, CSP seeks to determine a projection that captures a form of the evoked response. By using the source space and prior neuroimaging to select RTPJ activity, we had effectively targeted the evoked response before applying dimensionality reduction. As we cannot exploit this information twice, applying the neuroimaging finding in a generalized source space (and then a more accurate individualized source space) likely diminished the relative superiority of the CSP accuracy causing comparable performance across the three dimensionality reduction techniques. Practically, source imaging may therefore reduce the dependence of classification accuracy on some of the experimenter's signal processing choices.

We also compared the efficacy of two standard approaches within source imaging: generic and individualized forward models. We found that the use of individualized MRI information (needed to construct a BEM and position current dipoles along the anatomically accurate gray/white matter boundary) conferred an increase in classification accuracy over a generic model (i.e., spherical conductor model with dipoles placed according to a template brain). In the experimentally recorded data, this increase was statistically significant. Collecting the required structural MRI scan carries its own practical concerns as discussed in the next section, but this work quantifies the advantage of anatomical MRI information for source imaging in a BCI-relevant context.

2.5.3 Practical considerations

Despite conferring a number of benefits, source imaging also comes with practical burdens. The distributed current dipole model used here requires solving an ill-posed problem as there are many more

variables being estimated (i.e., cortical sources) than known quantities (i.e., EEG potentials). This requires the choice of a regularization scheme and associated parameters to ensure a unique and stable solution. A number of reasonable regularization choices exist – each with strengths and weaknesses (for review, see Baillet, et al., 2001). As described in the Methods, we employed the prevalent minimum-norm regularization strategy, which has a computationally convenient linear solution though it often converges on diffuse current estimates (even for focal sources of activity). While we obtained benefits using default parameters with the MNE algorithm, further research is needed to quantify how BCI classification accuracy increases or decreases using other source imaging strategies. Previous theoretical work, however, has shown that many conventional source imaging techniques, which includes MNE, are theoretically equivalent with a much greater dependency on the sensor covariance than the particular regularization approach (Mosher, et al., 2003). Therefore, we do not expect our accuracy results to differ between conventional source imaging approaches, but more investigation is needed to quantitatively assess this hypothesis.

The source imaging approach used here is similar to most sensor-based BCIs in that it does not account for potential non-stationarities in the recorded EEG. Non-stationarities (i.e., time-varying changes in the signal statistics) can manifest in neural recordings for a number of reasons (e.g., fatigue, learning or adaptation to the task, artifacts, etc.; Woolrich, et al., 2013; Samek, et al., 2012) and are known to hamper both BCIs and neuroscience data analysis. Recently, researchers have begun to develop source imaging techniques (Gramfort, et al., 2013; Woolrich, et al., 2013) and to modify common dimensionality reduction techniques (Samek, et al., 2012; Kawanabe, et al., 2014) that can account for non-stationary signals. While we showed that the conventional MNE source imaging approach maintained its classification advantage even when the signal was contaminated with eye blinks – a nonstationary artifact – more studies are needed that explore the robustness of other sensor- and source-based BCIs to a variety of non-stationarities.

Source imaging can also add time-consuming procedures to a BCI protocol. Though not strictly required, source imaging approaches typically use 3D position of each electrode within the forward model. Manual

electrode localization adds about five minutes to the EEG cap fitting procedure, but photometric methods (using cameras) already exist to expedite this process (Cincotti, et al., 2008). When constructing the generic forward model, reasonable defaults exist for the spherical shell parameters and a template brain (like FreeSurfer's "fsaverage") is often used to position the current dipoles. Overall, this adds little time or cost especially with the advent of open source software packages developed for source space analysis. Individualized source imaging approaches, however, require more expertise and resources because of the structural MRI scan needed (in addition to the electrode localization). Still, an adult typically requires only a single scan session (lasting 20-30 min), and these costs are more tolerable in clinical BCI situations where even a meager increase in accuracy could have a meaningful impact on a patient's quality of life.

We found that the source space approaches gave a substantial gain over the conventional sensor space approach. With an absolute (offline) classification accuracy of about 65% for both the simulated and experimental studies, however, the source imaging paradigm used here is likely not ready for real-world implementation. Nevertheless, other systems aimed at analyzing the time-dependent dynamics of attention might benefit from the framework explored here. For example, a previous study has explored dynamically classifying a subject's attention when two people are speaking. The authors tracked the envelope of each talker and successfully used Hidden Markov Models (HMMs) to classify which talker the subject was attending to (at sub-second time intervals) by matching speech envelopes to the neural signals (Mirbagheri, et al., 2015). Detecting the reallocation of attention could augment these sorts of systems by regulating the transition probabilities within the HMM. A hybrid BCI of this form might be useful in the development of hearing aids that dynamically change their spatial tuning based on brain states (Lee, et al., 2013). The work here is also relevant for future BCIs where monitoring attention modulation is of interest (e.g., in classrooms, while driving, during conversation, etc.).

2.5.4 Future prospects linking neuroscience and BCIs

Source imaging has a number of other advantages that are worth exploring in future studies. For example, deep learning – a state-of-the-art machine learning technique – constructs representations directly from raw data and has made impressive advances in many classical machine learning problems.

However, this technique has not gained much traction in BCIs probably because it requires relatively large datasets (LeCun, et al., 2015). BCI systems rarely operate under these circumstances. Typically, there is only time to collect several dozen training trials, and further, those trials have considerable noise. Nevertheless, we showed here that source imaging allows systematic targeting of the most informative regions (thereby increasing the signal-to-noise ratio), and previous work has demonstrated that source imaging helps normalize intersubject differences to improve transfer learning (Wronkiewicz, et al., 2015). Thus, source imaging may help reframe EEG-based approaches to BCIs in a manner that is more amenable for advanced computational techniques.

There exists a collection of neuroscience findings relevant to other cortical regions that might be used in developing other novel BCI paradigms. We showed that simple binary weights (i.e., including or excluding dipoles) allowed us to leverage the known location of the RTPJ, but our framework can be appropriate for more advanced priors. Knowing *a priori* the most relevant cortical region could theoretically even help guide the construction of non-standard electrode montages that optimally target relevant ROIs. Further, since source imaging is capable of estimating activity across the entire cortex, it greatly facilitates capturing features that are based on distributed activity or network synchronization (Schoffelen & Gross, 2009) – a major focus in neuroimaging over the past decade. These neuroimaging findings, particularly those concerning areas comprising resting-state networks, may yield a new vein of neuroscience-inspired features to explore. Thus, source imaging provides a principled way to port new techniques and knowledge into BCIs thereby tightening the seam between neuroengineering and neuroscience.

2.6 Conclusions

In this work, we evaluated a BCI operating within a source space framework. The source space permits the analysis of EEG brain recordings at the cortical level, where knowledge and methods from modern neuroscience are readily applicable toward BCIs. Using both simulated and experimental data, we showed that focusing our analysis on a particular cortical region (based on previous neuroimaging literature)

substantially improved classification of when a subject switched auditory attention compared to the conventional sensor-based classifier – a gain that was consistent across three popular dimensionality reduction techniques. This demonstrates a promising strategy toward developing new BCI paradigms on the heels of recent neuroscience findings. Our results also indicate that the source imaging improves classification when using either generic or individualized forward models (i.e., regardless of whether structural MRI information was available or not), but the individualized approach provides additional improvement over a generic one. The collection of MRI information requires additional costs, so this result supplies a needed quantification of the expected benefit. As source imaging allows estimation of brain activity across the entire cortex, future work will extend to analyzing network activity and synchronization.

Acknowledgements

The authors thank Dan McCloy, Ross Maddox, Majid Mirbagheri, and the two anonymous reviewers for helpful comments and suggestions. This research was funded by the National Science Foundation Division of Graduate Education, Graduate Research Fellowship Program (DGE-1256082 to MW), The Department of Defense Air Force Office of Scientific Research Young Investigator Program (FA9550-12-1-0466 to AKCL), and The Department of Defense Office of Naval Research Young Investigator Program (N00014-15-1-2124 to AKCL).

2.7 References

Ahn, M., Hong, J. H. & Jun, S. C., 2012. Feasibility of approaches combining sensor and source features in brain-computer interface. *Journal of Neuroscience Methods,* 204(1), pp. 168-78.

Baillet, S., 2010. The Dowser in the Fields: Searching for MEG Sources. In: *MEG: An Introduction to Methods.* New York: Oxford University Press, Inc., pp. 83-123.

Baillet, S., Mosher, J. C. & Leahy, R. M., 2001. Electromagnetic Brain Mapping. *IEEE Signal Processing Magazine,* Issue Nov., pp. 14-30.

Ball, T. et al., 2008. Movement related activity in the high gamma range of the human EEG. *NeuroImage,* 41(2), pp. 302-10.

Bashashati, A., Fatourechi, M., Ward, R. K. & Birch, G. E., 2007. A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *Journal of Neural Engineering,* 4(2), pp. R32-R57.

Bentivoglio, A. R. et al., 1997. Analysis of Blink Rate Patterns in Normal Subjects. *Movement Disorders,* 12(6), pp. 1028-1034.

Besserve, M., Martinerie, J. & Garnero, L., 2011. Improving quantification of functional networks with EEG inverse problem: Evidence from a decoding point of view. *NeuroImage,* 55(4), pp. 1536-47.

Blankertz, B., Tangermann, M. & Müller, K.-R., 2012. BCI applications for the general population. In: J. R. Wolpaw & W. E. Winter, eds. *Brain-Computer Interfaces: Principles and Practice.* New York, New York: Oxford University Press, pp. 363-83.

Chapman, R. M. & Bragdon, H. R., 1964. Evoked Responses to Numerical and Non-Numerical Visual Stimuli while Problem Solving. *Nature,* Volume 203, pp. 1555-7.

Chen, Y., Wiesel, A., Eldar, Y. C. & Hero, A. O., 2010. Shrinkage Algorithms for MMSE Covariance Estimation. *IEEE Transactions on Signal Processing,* 58(10), pp. 5016-5029.

Cincotti, F. et al., 2008. High-resolution EEG techniques for brain-computer interface applications. *Journal of Neuroscience Methods,* 167(1), pp. 31-42.

Congedo, M., Lotte, F. & Lécuyer, A., 2006. Classification of movement intention by spatially filtered electromagnetic inverse solutions. *Physics in Medicine and Biology,* 51(8), pp. 1971-89.

Dale, A. M., Fischl, B. & Sereno, M. I., 1999. Cortical Surface-Based Analysis. *NeuroImage,* 9(2), pp. 179-94.

Dale, A. M. et al., 2000. Dynamic Statistical Parametric Mapping: Combining fMRI and MEG for High-Resolution Imaging of Cortical Activity. *Neuron,* Volume 26, pp. 55-67.

Darvas, F. et al., 2010. High gamma mapping using EEG. *NeuroImage,* 49(1), pp. 930-8.

Destrieux, C., Fischl, B., Dale, A. & Halgren, E., 2010. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage,* 53(1), pp. 1-15.

Edelman, B. J., Baxter, B. & He, B., 2015. EEG Source Imaging Enhances the Decoding of Complex Right-Hand Motor Imagery Tasks. *IEEE Transactions on Biomedical Engineering,* 63(1), pp. 4-14.

Farwell, L. A. & Donchin, E., 1988. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogray and Clinical Neurophysiology,* 70(6), pp. 510-23.

Friston, K. J. et al., 1995. Spatial Registration and Normalization of Images. *Human Brain Mapping,* 3(3), pp. 165-89.

Gramfort, A. et al., 2013. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience,* Volume 7, pp. 1-13.

Gramfort, A. et al., 2014. MNE software for processing MEG and EEG data. *NeuroImage,* Volume 86, pp. 446-60.

Gramfort, A. et al., 2013. Time-frequency mixed-norm estimates: Sparse M/EEG imaging with non-stationary source activations. *NeuroImage,* Volume 70, pp. 410-22.

Grosse-Wentrup, M., Liefhold, C., Gramann, K. & Buss, M., 2009. Beamforming in Noninvasive Brain-Computer Interfaces. *IEEE Transactions on Biomedical Engineering,* 56(4), pp. 1209-19.

Grosse-Wentrup, M. & Schölkopf, 2014. A brain-computer interface based on self-regulation of gamma-oscillations in the superior parietal cortex. *Journal of Neural Engineering,* 11(5), pp. 1-13.

Hämäläinen, M. et al., 1993. Magnetoencephalography - theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics,* 65(2), pp. 413-97.

Hämäläinen, M. & Ilmoniemi, R. J., 1984. *Interpreting Measured Magnetic Fields of the Brain: Estimates of Current Distribution,* Helsinki, Finland: Helsinki University of Technology.

Kamousi, B., Liu, Z. & He, B., 2005. Classification of Motor Imagery Tasks for Brain-Computer Interface Applications by Means of Two Equivalent Dipoles Analysis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* 13(2), pp. 166-71.

Kamousi, B., Nasiri, A. & He, B., 2007. Classification of motor imagery by means of cortical current density estimation and Von Neumann entropy. *Journal of Neural Engineering,* 4(2), pp. 17-25.

Kawanabe, M., Samek, W., Müller, K.-R. & Carmen, V., 2014. Robust Common Spatial Filters with a Maxmin Approach. *Neural Computation,* 26(2), pp. 349-76.

Krusienski, D. J., McFarland, D. J. & Principe, J. C., 2012. BCI signal processing: feature extraction. In: J. R. Wolpaw & E. W. Wolpaw, eds. *Brain-Computer Interfaces: Principles and Practice.* New York, New York: Oxford University Press, pp. 147-63.

Larson, E. & Lee, A. K. C., 2013. The cortical dynamics underlying effective switching of auditory spatial attention. *NeuroImage,* Volume 64, pp. 365-70.

Larson, E. & Lee, A. K. C., 2014. Switching auditory attention using spatial and non-spatial features recruits different cortical networks. *NeuroImage,* Volume 84, pp. 681-687.

LeCun, Y., Bengio, Y. & Hinton, G., 2015. Deep learning. *Nature,* Volume 521, pp. 436-44.

Ledoit, O. & Wolf, M., 2004. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis,* 88(2), pp. 356-411.

Lee, A. K. C., Drews, M. K., Maddox, R. K. & Larson, E., 2013. Brain Imaging, Neural Engineering Research, and Next-Generation Hearing Aid Design. *Audiology Today,* pp. 40-7.

Lotte, F., 2015. Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based Brain-Computer Interfaces. *Proceedings of the IEEE,* 103(6), pp. 871-90.

Lotte, F., Lécuyer, A. & Arnaldi, B., 2009. FuRIA: An Inverse Solution Based Feature Extraction Algorithm Using Fuzzy Set Theory for Brain-Computer Interfaces. *IEEE Transactions on Signal Processing,* 57(8), pp. 3253-63.

Makeig, S. et al., 2012. Evolving Signal Processing for Brain-Computer Interfaces. *Proceedings of the IEEE,* 100(Special Centennial Issue), pp. 1567-84.

Menendez, G. d. P. et al., 2005. Non-invasive estimation of local field potentials for neuroprosthesis control. *Cognitive Processing,* 6(1), pp. 59-64.

Mirbagheri, M., Ekin, B., Atlas, L. & Lee, A. K. C., 2015. *Flexible Tracking of Auditory Attention.* Dresden, s.n., pp. 2425-9.

Mitchell, J. P., 2008. Activity in Right Temporo-Parietal Junction is Not Selective for Theory-of-Mind. *Cerebral Cortex,* 18(2), pp. 262-71.

Mosher, J. C., Baillet, S. & Leahy, R. M., 2003. Equivalence of Linear Approaches in Bioelectromagnetic Inverse Solutions. *2003 IEEE Workshop on Statistical Signal Processing,* pp. 294-7.

Noirhomme, Q., Kitney, R. I. & Macq, B., 2008. Single-Trial EEG Source Reconstruction for Brain-Computer Interface. *IEEE Trans. Biomed. Eng.,* 55(5), pp. 1592-601.

Nunez, P. L. et al., 1997. EEG coherency I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. *Electroencephalography and Clinical Neurophysiology,* 103(5), pp. 499-515.

Parra, L. C., Spence, C. D., Gerson, A. D. & Sajda, P., 2005. Recipes for the linear analysis of EEG. *NeuroImage,* 28(2), pp. 326-41.

Penfield, W. & Boldrey, E., 1937. Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain,* 60(4), pp. 389-443.

Penfield, W. & Rasmussen, T., 1950. *The cerebral cortex of man.* New York, New York: The Macmillan Company.

Pfurtscheller, G. & Aranibar, A., 1979. Evaluation of event-related desynchronization (ERD) preceding and following voluntary self-paced movement. *Electroencephalography and Clinical Neurophysiology,* 46(2), pp. 138-46.

Qin, L., Ding, L. & He, B., 2004. Motor imagery classification by means of source analysis for brain-computer interface applications. *Journal of Neural Engineering,* 1(3), pp. 135-41.

Samek, W., Vidaurre, C., Müller, K.-R. & Kawanabe, M., 2012. Stationary common spatial patterns for brain-computer interfacing. *Journal of Neural Engineering,* 9(2), pp. 1-14.

Schoffelen, J.-M. & Gross, J., 2009. Source Connectivity Analysis With MEG and EEG. *Human Brain Mapping,* 30(6), pp. 1857-65.

Seabold, S. & Perktold, J., 2010. *Statsmodels: Econometric and statistical modeling with python.* Austin, TX, 9th Python in Science Conference, pp. 57-61.

Sutton, S., Braren, M. & Zubin, J., 1965. Evoked-Potential Correlates of Stimulus Uncertainty. *Science,* 150(3700), pp. 1187-8.

Uusitalo, M. A. & Ilmoniemi, R. J., 1997. Signal-space projection method for separating MEG or EEG into components. *Medical and Biological Engineering and Computing,* 35(2), pp. 135-40.

van den Broek, S. P., Reinders, F., Donderwinkel, M. & Peters, M. J., 1988. Volume conduction effects in EEG and MEG. *Electroencephalography and Clinical Neurophysiology,* 106(6), pp. 522-34.

Vidal, J. J., 1973. Toward direct brain-computer communication. *Annual Review Biophysics and Bioengineering,* Volume 2, pp. 157-80.

Vidal, J. J., 1977. Real-time detection of brain events in EEG. *Proceedings of the IEEE,* 65(5), pp. 633-41.

Witt, S. T., Laird, A. R. & Meyerand, M. E., 2008. Functional neuroimaging correlates of finger-tapping task variations: An ALE meta-analysis. *NeuroImage,* Volume 42, pp. 343-56.

Wolpaw, J. R. et al., 2002. Brain-computer interfaces for communication and control. *Clinical Neurophysiology,* 113(6), pp. 767-91.

Wolpaw, J. R., McFarland, D. J., Neat, G. W. & Forneris, C. A., 1991. An EEG-based brain-computer interface for cursor control. *Electroencephalography and Clinical Neurophysiology,* 78(3), pp. 252-9.

Wolpaw, J. & Wolpaw, E. W., 2012. *Brain-Computer Interfaces: Principles and Practice.* New York: Oxford University Press.

Woolrich, M. W. et al., 2013. Dynamic state allocation for MEG source reconstruction. *NeuroImage,* Volume 77, pp. 77-92.

Wronkiewicz, M., Larson, E. & Lee, A., 2015. Leveraging anatomical information to improve transfer learning in brain-computer interfaces. *Journal of Neural Engineering,* 12(4), pp. 1-12.

Yuan, H., Doud, A., Gururajan, A. & He, B., 2008. Cortical Imaging of Event-Related (de)Synchronization During Online Control of Brain-Computer Interface Using Minimum-Norm Estimates in Frequency Domain. *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* 16(5), pp. 425-31.

Chapter 3. Leveraging anatomical information to improve transfer learning in brain-computer interfaces

This chapter was published in the Journal of Neural Engineering by Mark Wronkiewicz, Eric Larson, and Adrian KC Lee with the title "Leveraging anatomical information to improve transfer learning in brain–computer interfaces" here: http://dx.doi.org/10.1088/1741-2560/12/4/046027.

3.1 Abstract

Objective: Brain-computer interfaces (BCIs) represent a technology with the potential to rehabilitate a range of traumatic and degenerative nervous system conditions but require a time-consuming training process to calibrate. An area of BCI research known as transfer learning is aimed at accelerating training by recycling previously recorded training data across sessions or subjects. Training data, however, is typically transferred from one electrode configuration to another without taking individual head anatomy or electrode positioning into account, which may underutilize the recycled data. *Approach:* We explore transfer learning with the use of source imaging, which estimates neural activity in the cortex. Transferring estimates of cortical activity, in contrast to scalp recordings, provides a way to compensate for variability in electrode positioning and head morphologies across subjects and sessions. *Main Results:* Based on simulated and measured EEG activity, we trained a classifier using data transferred exclusively from other subjects and achieved accuracies that were comparable to or surpassed a benchmark classifier (representative of a real-world BCI). Our results indicate that classification improvements depend on the number of trials transferred and the cortical region of interest. *Significance:* These findings suggest that cortical source-based transfer learning is a principled method to transfer data that improves BCI classification performance and provides a path to reduce BCI calibration time.

3.2 Introduction

Brain-computer interfaces (BCIs) are designed to provide alternate communication pathways for the brain and hold potential to rehabilitate people with traumatic or degenerative brain damage. Yet, this technology still faces some obstacles including the time required to calibrate these systems (Blankertz, et al., 2008; Wolpaw, et al., 2002). A typical non-invasive BCI session can be divided into three phases: electroencephalography (EEG) cap setup, a calibration period to train a machine-learning algorithm, and a feedback phase in which the user controls some output, such as a computer, robotic prosthetic, or spelling system. The first two phases are simply means to an end – setup and training prepare the BCI system to infer a user's intended action from recorded brain signals (Krauledat, et al., 2008). Given a fixed length BCI session, time spent in these two phases reduces time spent in the feedback phase, so techniques that shorten setup or training make BCIs more appealing for clinical and commercial applications by allocating additional time to the feedback phase. Recent improvements in recording hardware (e.g., rapidly deployable caps and dry electrodes) have reduced the setup time (Chi, et al., 2010), but less progress has been made in reducing the 20-30 minute training phase (Krauledat, et al., 2007; Lotte & Guan, 2010).

The purpose of the BCI training phase is to accumulate data to build a statistical model that can recognize spatial and temporal patterns in an individual's neural activity. These patterns typically vary across subjects (and even across sessions within individual subjects) due to a number of factors, so training is required to tune the classifier to the available features. A branch of BCI research known as transfer learning is focused on reducing training time by reusing previously recorded data to calibrate a new system (Lotte & Guan, 2010; Krauledat, et al., 2008; Kindermans, et al., 2014). Recycled training data can originate from the same subject's previous session(s) (i.e., subject-specific) or be transferred from other BCI users (i.e., subject-independent). Transfer learning thus provides a way to reduce or eliminate the need to collect training data for an ongoing BCI session by recycling preexisting training trials.

Past transfer learning strategies have employed a number of techniques to recycle BCI data. Many groups have developed methods to generate spatial filters and/or classifiers with the aim of creating BCI systems that are session- or subject- independent (Devlaminck, et al., 2011; Fazli, et al., 2009; Kang, et al., 2009; Kindermans, et al., 2014; Krauledat, et al., 2007; Krauledat, et al., 2008; Lotte & Guan, 2010; Tu & Sun, 2012; Wang, et al., 2012; Yang, et al., 2014, but see Samek, et. al., 2013 for an alternate approach). While the specific implementations vary, all of these methods transfer information based on EEG spatial patterns on the scalp, which neglect variation in individual cortical anatomy, head anatomy, and electrode locations.

It is preferable to classify across-subject data after explicitly accounting for deviations in head morphology and electrode positioning. Source imaging achieves these goals by mapping EEG sensor measurements recorded from the scalp (or sensor space) onto the underlying neural sources of the activity in the cortex (or source space; see Sections 1.1 and 1.2 as well as Figure 2.1 for details). Variation in individual anatomy and electrode placement is controlled as both are incorporated in the activity estimation process using structural MRI and electrode coregistration. Once brain activity is estimated in the source space, established neuroimaging tools allow for the morphing of brain activity between different subjects' brain spaces – a technique known as spatial normalization (Friston, et al., 1995). In this way, we can estimate how brain activity from one subject would manifest in any other subject's brain space. Therefore, the source space approach serves as a more robust reference frame for recycling non-invasive brain data across sessions and subjects.

This work explores the feasibility of transferring training trials through the source space to conduct subject-independent transfer learning in a single-trial BCI classification task. Compared to a subject-specific BCI classifier, we hypothesized that a system trained *solely* with data transferred from other subjects through the source space would yield equivalent or improved performance. To test this, we first simulated EEG data to examine classification gains when exploiting 3-dimensional anatomical information conferred by the source space. We then validated the simulation results by classifying the switching of attention in previously recorded EEG data. In the real EEG and most of the simulated EEG experiments,

we found that the transfer learning method surpassed the subject-specific classifier with a moderately sized pool of transferred data. This work illustrates the potential of source space to achieve anatomically normalized transfer learning as well as provide a formal avenue to include neuroscience priors in BCI work.

3.3 Methods

3.3.1 Anatomical information

Structural information was obtained for 21 subjects using three T1-weighted MRI scans consisting of one structural multi-echo magnetization prepared rapid gradient echo scan and two multi-echo multi-flip angle (5° and 30°) fast low-angle shot scans. From this structural data, a 3-layer boundary element model (BEM) was constructed for each subject using FreeSurfer (http://surfer.nmr.mgh.harvard.edu/; Dale, et al., 1999) to characterize the propagation of neural electromagnetic fields through the different layers of the head to the EEG sensors (Baillet, 2010). Positions of EEG electrodes, cardinal landmarks (nasion, left/right preauriculars), and other points on the scalp were recorded using a 3Space Fastrak (Polhemus) to coregister the electrodes with the BEM's coordinate frame.

3.3.2 Source vs. sensor space

This work used source imaging based on the algorithm implemented in MNE-Python. See Section 1.2 for an in depth discussion of these methods. In the real EEG data here, sensor noise covariance matrices were calculated from 200 ms baseline periods before each trial for every subject.

3.3.3 Simulated EEG data

The cortical dipoles in each of the N=21 subject's cortical models were partitioned into anatomically defined cortical patches, or regions of interest (ROIs), using FreeSurfer's automatic parcellation algorithm (Destrieux, et al., 2010). This technique uses cortical curvature to define a boundary between all gyri and sulci and then splits the cortical surface into distinct ROIs according to an anatomical atlas. We investigated 69 automatically defined non-overlapping sulcal and gyral ROIs, as well as 3 additional

custom hand motor cortex ROIs to further study the effects of ROI size. The center of these hand motor ROIs was defined from a hand motor functional localization meta-study (Witt, et al., 2008). Three increasingly larger hand motor ROIs were then constructed by finding the modeled dipoles that satisfied two conditions: 1) they were part of FreeSurfer's precentral gyrus ROI, and 2) were within 5, 10, and 15 mm of this center. To limit computational complexity, only ROIs located in the left hemisphere were used.

We conducted an EEG-based classification task using these 72 cortical ROIs to classify whether each had elevated or baseline levels of activity. To accomplish this, cortical activity $\mathbf{j}_i$ was simulated for the ith subject, where $i \in \{1 \ldots N\}$ at each ROI and normalized by the cortical area (as approximated by the number of included dipoles) to yield a constant current density. The simulated EEG profile $\mathbf{x}_i$ was then calculated by multiplying the activity $\mathbf{j}_i$ by the gain matrix $\mathbf{A}_i$ and adding Gaussian noise $\mathbf{n}_i$ (see Figure 3.1, dashed blue box):

$$\mathbf{x}_i = \mathbf{A}_i\mathbf{j}_i + \mathbf{n}_i \qquad (3.1)$$

Simulated EEG profiles consisted of a single time point of activity and each served as a trial. Sets of 10, 20, and 40 trials were simulated for each condition (elevated or baseline activity) for each subject. As there exists no ground truth as to which brain signals constitute "signal" and "noise" in real EEG recordings, it is difficult to estimate the SNR of experimentally recorded data; however, in simulations we had explicit control over the SNR. Thus, we repeated the simulation at SNRs of -15, -10, and -5 dB, as the resulting classification performances best spanned accuracies of actual BCI systems previously reported in the literature. As in the MNE toolbox (Gramfort, et al., 2013), SNR was defined as $10 * \log_{10}\left(\sigma^2_{\text{signal}}/\sigma^2_{\text{noise}}\right)$.

3.3.4 Real EEG data

We also used EEG data from a previously recorded attentional modulation task (Larson & Lee, 2014) to complement our simulation study. For details on the experimental paradigm, see the previous manuscript. All participants gave informed consent according to the procedures approved by the University of Washington. Briefly, N=10 subjects (all also a part of the previously mentioned 21 subject

pool) listened to a target and masker auditory stream with each stream comprised of spoken letters. In the experimental condition analyzed here, each stream was delayed by 300 μs either in the right or left channel to elicit a leftward or rightward spatial percept. Before each trial, listeners were cued to either maintain attention to the same auditory stream (left or right) throughout the trial or switch streams halfway through during a gap period. Using EEG data from this gap period, we classified whether or not the subject switched auditory spatial attention. The subject was required to report a behavioral metric for each trial, and only correct trials were included in our analysis. For this reason, the number of trials varied for each subject (mean=38). For simplicity, EEG data was temporally averaged over the 600 ms gap period to create a single spatial profile (i.e., one time point) of activity just as in the simulated data.

3.3.5. Traditional vs. transfer learning training data

Two different training formulations were tested in this BCI classification task for both the simulated and real experiments: a traditional approach and the transfer learning approach. We emulated a traditional BCI classifier (detailed below) by training and testing on simulated (and real) EEG data from each of the 21 (10) subjects individually. The majority of current BCI applications use this subject-specific strategy, so the classification accuracy here served as a benchmark for our transfer learning technique. The transfer learning approach recycled EEG trials based on a leave-one-*subject*-out method—when training a classifier for subject *k*, we used trials from all other subjects and transferred them through the source space to subject *k*'s brain space for classifier training. In other words, the transfer learning approach was subject-independent, as no data from the subject of interest was used to train that individual's BCI system.

For both the simulated and real EEG experiments, data transfer was achieved by first multiplying the training EEG activity for each subject *i* by the pseudoinverse of the gain matrix $\mathbf{M}_i = \mathbf{A}_i^{\dagger}$ to estimate the cortical currents $\hat{\mathbf{j}}_i$ (see Figure 3.1 and Section 1.2 for details). Next, brain activity was transferred by morphing these current activity profiles to subject *k*'s brain space in a manner analogous to the spatial normalization of brain activity across subjects in neuroimaging. This transfer step used an isotropic diffusion process described in Gramfort et al. (2014) and FreeSurfer's spherical brain morphing procedure

(Fischl, et al., 1999). Source estimates were first spatially smoothed onto a high-resolution version of the cortical mesh with the aforementioned diffusion process. Next, the mesh was inflated into a unit sphere where the entire set of sulcal/gyral boundaries, which delineate many functional areas, were optimally aligned with the target brain. A linear transformation from the original to the target sphere was then calculated and applied in FreeSurfer (using the *morph maps* function), and the spatially normalized data was subsampled to the target spatial resolution (of ~7000 dipoles). This entire set of linear algebra operations comprises the transformation matrix $\mathbf{T}_{i\to k}$. Therefore, the transformation of cortical activity from subject *i* to *k* was accomplished by multiplying each subject's source estimates by the appropriate transfer matrix (see Figure 3.1):

$$\hat{\mathbf{j}}_{i\to k} = \mathbf{T}_{i\to k}\hat{\mathbf{j}}_i \qquad (3.2)$$

After the training trials from each of the N-1 subjects were transferred to subject *k*'s source space, the data were multiplied by subject *k*'s gain matrix and optionally weighted to generate N-1 sets of EEG training data for subject *k*'s classifier (see Figure 3.1):

$$\hat{\mathbf{x}}^{\mathrm{W}}_{i\to k} = \mathbf{W}_k\mathbf{A}_k\hat{\mathbf{j}}_{i\to k} \qquad (3.3)$$

The weighting procedure will be discussed subsequently in section 3.2.7.

Using this leave-one-*subject*-out scheme, similar to previous transfer learning work (Fazli, et al., 2009), data from all *i* subjects trained the classifier while data from subject *k* was used to test the classifier. The simulations for both the traditional (subject-specific) and transfer learning (subject-independent) versions of the classification task were repeated 25 times per subject to ensure stability before comparison. All of the code for this analysis is online (https://github.com/LABSN-pubs/2015-JNE-Wronkiewicz).

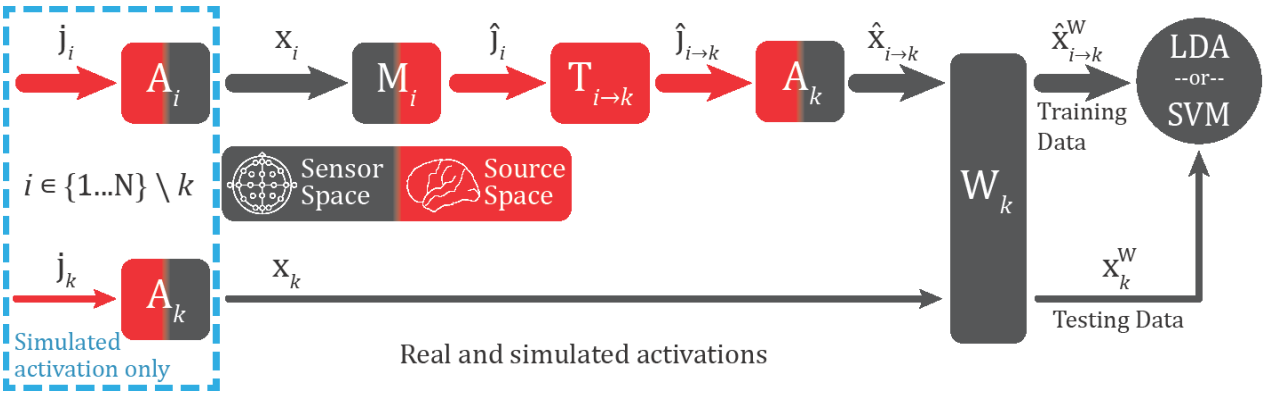

Figure 3.1 Schematic of source-based transfer learning method. Blue, dashed box: In the simulation experiments only, cortical activity was generated for all subjects and multiplied by the appropriate gain matrix to simulate EEG activity. Top row: In both the simulated and real experiments, EEG data for subjects 1 to N (excluding subject *k*) were multiplied by the pseudoinverse matrix to obtain source activity estimates. Source estimates were subsequently morphed to subject *k*'s brain using the appropriate transformation matrix (see text for details and citations), multiplied by subject *k*'s gain matrix, and (optionally) weighted yielding N-1 sets of classifier training data in subject *k*'s sensor space. Bottom row: In both the simulated and real experiments, EEG data from the subject of interest (subject *k*) were (optionally) weighted yielding data to test the classifier. Note that no data from the subject of interest were used to train the BCI classifier.

We repeated this classification task with different numbers of subjects in the training set to see how classification performance was affected by training pool size. In the simulation task, training sets ranged from 2 to 20 randomly selected subjects with 40 trials per condition and an SNR of -10 dB. As before, the classification was repeated 25 times with random subsets of $i \in \{1 \ldots N\} \setminus k$ subjects for each test subject *k* for stability. Here, classification accuracy was investigated in a set of ROIs relevant to common BCI paradigms, which included six cortical areas potentially associated with the P300 response, four areas associated with motor-imagery, three steady state visually evoked potential (SSVEP) areas, and one ROI in primary auditory cortex. In the real EEG data, training sets ranged from 2 to 9 randomly selected

subjects with an average of 38 trials per condition, and the classification was also repeated 25 times. The transfer of real attentional data used the same process as in the simulation experiment.

3.3.6 Classification methods

In the simulation experiment, we used regularized Linear Discriminant Analysis (LDA) (Friedman, 1989) to classify whether each ROI had elevated or baseline levels of activity. A conventional regularization parameter of 0.05 was used as it optimized the subject-specific (benchmark) classifier's performance. In the real EEG data, we used support vector machines (SVM) with a radial basis function kernel and an optimization strategy that was the same for both the benchmark and transfer-learning data. Leave-one-trial-out cross-validation was used in the simulated data, while 10-fold cross-validation was used in the real EEG data because of computational demands. LDA and SVMs were chosen for classification because of their efficiency and widespread use in the BCI field (for review, see Lotte, et al., 2007).

3.3.7 Simulation: weighting electrode importance

In the simulated experiment of this transfer learning approach, we also investigated the impact of weighting the EEG data according to the presumed location of active cortical sources. We hypothesized that the classifier training efficiency and accuracy could be improved by exploiting this spatial information. Weighting the EEG data warped the feature space by compressing the variance of signal dimensions that were deemed unimportant. As LDA attempts to simultaneously maximize the mean separation between classes and minimize within-class variance, this feature weighting made the discrimination boundary more sensitive to highly weighted EEG electrodes.

To achieve this weighting, we combined the gain matrix with each subject's presumed ROI location to define cortical areas important for classification. The weighting was tested using three different formulations. Consider a region like the precentral gyrus, for example (Figure 3.2a). The first and simplest method did not take location information into account and treated each dipole with equal importance (Figure 3.2b). The second, called centroid weighting, found the spatial center-of-mass of the anatomical region and weighted importance of dipoles according to a cortical distance-based Gaussian decay (Figure 3.2c). The last weighting strategy, Gaussian mixture weighting, used a Gaussian kernel

spatially convolved with each dipole to accumulate each dipole's contribution into a non-uniform weighting map (Figure 3.2d) inspired by previous fuzzy ROI work (Lotte, et al., 2009b). Here, dipole importance was determined according to the number of times that a particular dipole belonged to the ROI under evaluation across the entire subject set. The centroid and Gaussian mixture methods both made use of *a priori* knowledge about the presumed origin of certain cortical activity to weight those areas more highly during classification. In all weighting schemes, cortical dipole weights were multiplied by the gain matrix $\mathbf{A}$ and normalized by the maximum weight to obtain a weighting profile in the sensor space. EEG data was weighted according to each method and classified to explore any potential increases in classifier accuracy when including information about the ROI's spatial location.

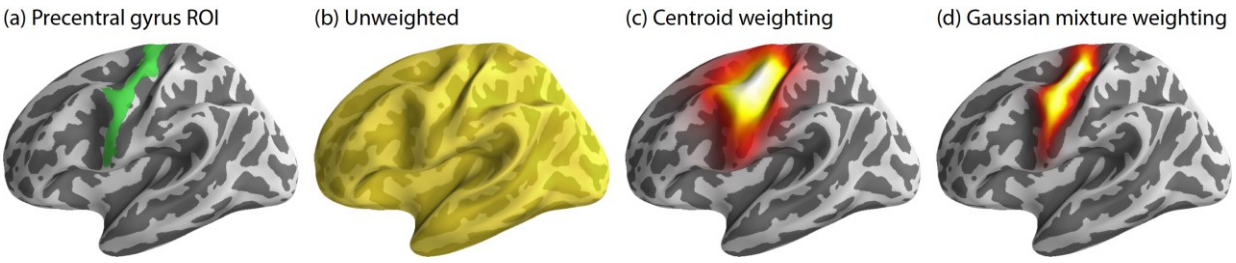

Figure 3.2 Three different weighting strategies for an exemplar ROI. (a) Precentral gyrus ROI (green). (b) Unweighted: dipoles are given equal weight when classifying. (c) Centroid weighting: dipoles are weighted by the cortical surface distance to the ROI's center. (d) Gaussian mixture: dipoles are weighted according to a spatial convolution of a Gaussian kernel with their importance to generate a non-uniform weighting profile (see text for details).

3.4 Results

3.4.1 Simulation: transfer learning vs. traditional classification performance

We compared classification accuracies between a traditional BCI approach and our source space transfer learning method using simulated data. The traditional (subject-specific) classifier was trained and tested on data from each subject individually, and served as our benchmark. The (subject-independent) transfer

learning method trained the classifier for the subject of interest using *only* transferred data from the other subjects.

In the traditional classifier, average accuracy across all 72 ROIs increased with both SNR and trial count and spanned a range of accuracies typical for real-world BCIs (Figure 3.3a). The transfer learning classifier, set at 40*(N-1) trials per subject, also showed accuracy increases with SNR and incremental gains with increasingly specific spatial weighting profiles (Figure 3.3b). The most accurate transfer learning classifier (Figure 3.3c; copy of bottom row of Figure 3.3b) performed comparably to the subject-specific classifier and achieved a small improvement in the two conditions with the lowest SNR. Again, zero training trials originated from the subject of interest (whose data was being classified) in the transfer learning classifier.

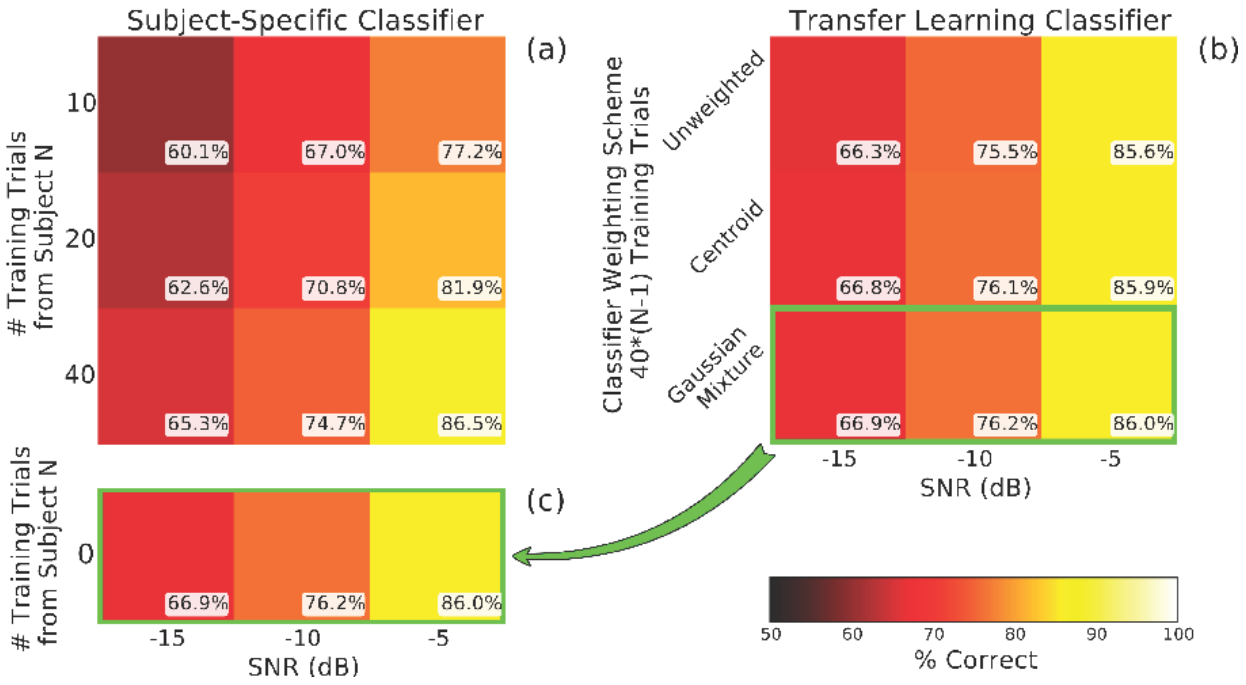

Figure 3.3 An LDA classifier trained with source-based transfer learning performed comparably to the traditional subject-specific classifier. Each score represents the average accuracy of 72 ROIs for 21 subjects repeated 25 times. (a) The traditional classifier was evaluated using 10, 20, and 40 training trials at SNRs of -15, -10, and -5 dB, and the classifier was trained and tested on each subject individually. (b) The transfer learning classifier was trained with three different weighting schemes. Each condition

represents the average as in (a), but the classifier was trained exclusively on data from all other (N-1) subjects in the pool and then tested using data from the excluded subject. (c) The Gaussian mixture accuracy is copied for comparison with the bottom row of (a), where the traditional classifier showed performance levels in the same range as many actual BCI systems. The transfer learning method surpassed the subject specific classifier in the two lowest SNR conditions. Theoretical chance accuracy is 50%.

The percent accuracy difference between the (Gaussian mixture weighted) transfer learning and the traditional classifier for each ROI is shown in Figure 3.4. Here, the SNR was set at -10 dB as the traditional classifier performance best represented real-world BCI systems when using a realistic training set size of 40 trials per class. Performance gains were not uniform across the cortical surface: the transfer learning approach better classified ROIs in many of the frontal and parietal dorsal areas, while the traditional classifier better classified ROIs in the occipital lobe and near the lateral sulcus (or Sylvian fissure).

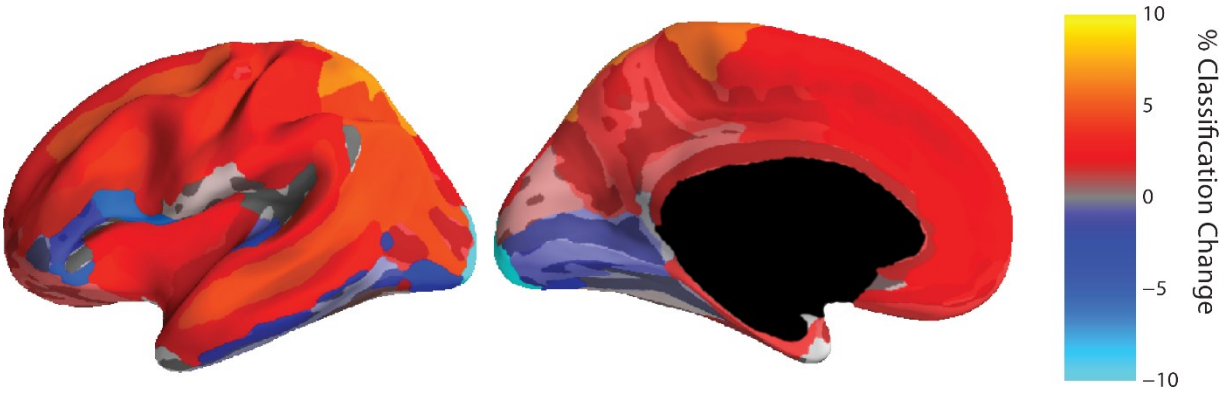

Figure 3.4 Difference in accuracy between the (Gaussian mixture weighted) transfer learning approach and the traditional BCI classifier at 72 ROIs distributed across the cortical surface. Lateral (left) and medial (right) views of cortical ROIs used in the classification task. Each ROI represents the average accuracy of 25 repetitions over 21 subjects. Here, the SNR was set at -10 dB as this gave a classification accuracy representative of real-world BCI systems in the traditional classifier.

3.4.2 Simulation: effect of training set size

We further explored how training set size affected transfer learning performance in ROIs related to common BCIs. To assess classification improvements in motor imagery BCIs, we tested three increasingly larger ROIs in the hand motor knob (Figure 3.5, increasingly lighter green shades) as well as the precentral sulcus (Figure 3.5, blue). In all ROIs, the transfer learning approach eventually achieved a classification accuracy greater than the traditionally trained classifier (Figure 3.5, circular markers). The pool size required to surpass the traditional classifier ranged from 2 to 6 subjects, and all ROIs reached an accuracy plateau after 10-12 subjects when using 40 trials per condition per subject.

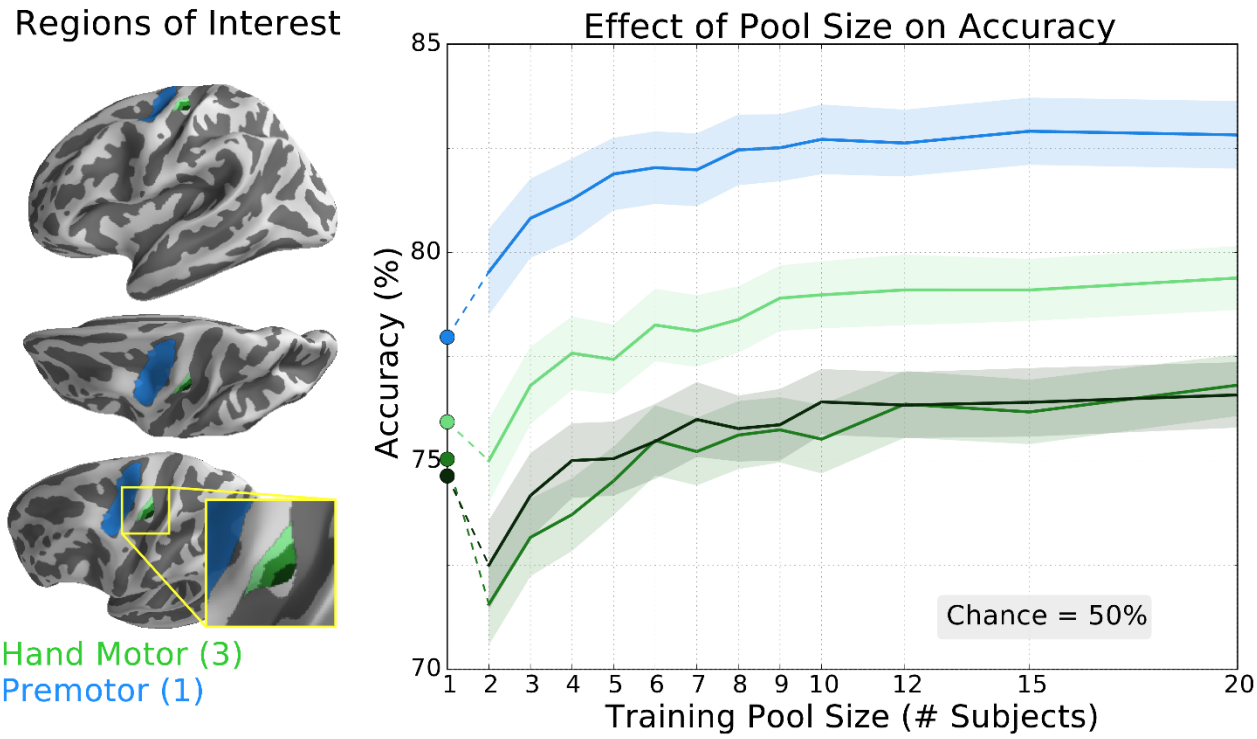

Figure 3.5 The classifier trained with transfer learning achieved higher classification accuracy than the traditional classifier for all tested ROIs related to motor imagery. Left: ROIs included small, medium, and large hand motor ROIs (increasingly lighter green shades; hand motor knob) and the premotor area (blue; pre-central sulcus). In the hand motor region, the three ROIs overlap so that each individual region contains the smaller ROI(s). Right: Accuracy of the transfer learning classifier ± SEM as a function

of pool size for these motor imagery ROIs averaged over 21 subjects repeated 25 times. Accuracy of the subject-specific classifier is indicated by the circular markers, left. Theoretical chance accuracy is 50%.

We repeated this analysis in six brain regions that are implicated in the P300 "oddball" event-related potential (ERP) in both early (Figure 3.6; P300a, green) and late (Figure 3.6; P300b, blue) components of this response. A large number of ROIs were tested here as the precise origin of the P300 ERP has remained elusive (Polich, 2007). For all ROIs, the transfer learning method surpassed the traditional classifier (Figure 3.6, circular markers) when the pool was comprised of 3 or more subjects and showed diminished accuracy improvements after 8-10 subjects.

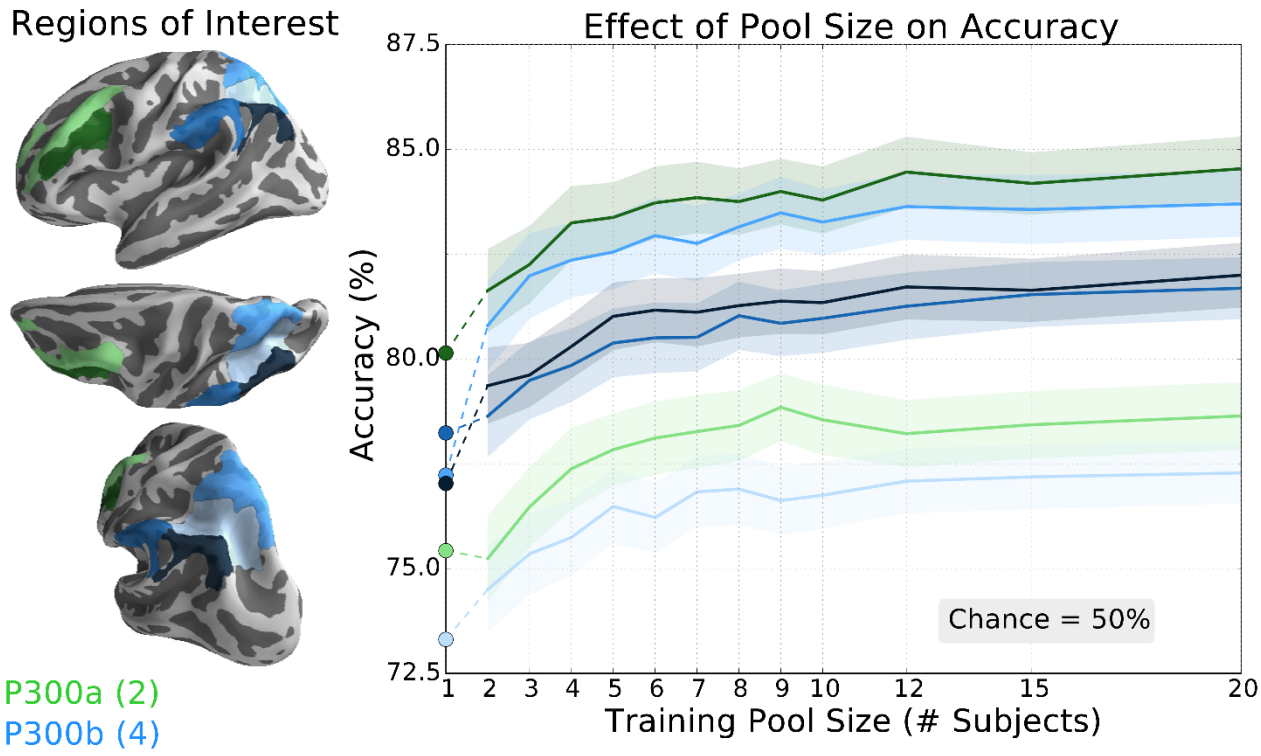

Figure 3.6 The transfer learning classifier achieved higher classification accuracy than the traditional classifier for all tested ROIs thought to be related to the P300 ERP. Left: P300a ROIs included middle frontal gyrus and inferior frontal sulcus (light and dark green, respectively). P300b ROIs include intraparietal sulcus and transverse parietal sulci, superior parietal lobule, supramarginal gyrus, and

angular gyrus (increasingly darker shades of blue, respectively). Right: Accuracy of the transfer learning classifier ± SEM as a function of pool size for these P300 ROIs averaged over 21 subjects repeated 25 times. Accuracy of the subject-specific classifier is indicated by the circular markers, left. Theoretical chance accuracy is 50%. Note scale difference from other BCI ROI figures.

Finally, we evaluated three SSVEP regions in primary visual (Figure 3.7, blue) and auditory regions (anterior transverse temporal gyrus; Figure 3.7, green). Here, only one of the three visual ROIs eventually reached accuracy levels matching that of the subject-specific accuracy (Figure 3.7, circular markers). The transfer learning method also failed to reach the subject-specific training accuracy in the primary auditory ROI. Even so, only the occipital pole region of primary visual cortex showed an accuracy decrease of more than 1.5% for the largest pool size.

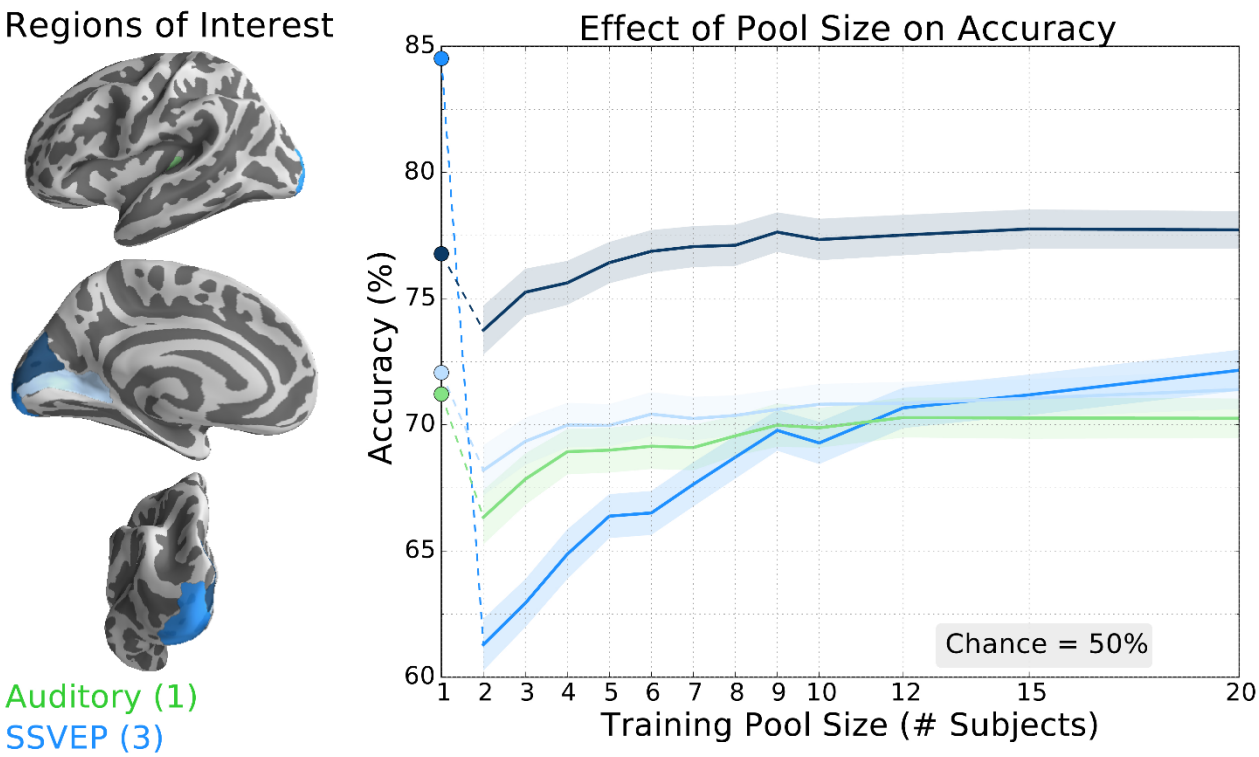

Figure 3.7 The classifier trained with transfer learning achieved higher classification accuracy than the traditional classifier in one of the primary auditory and visual ROIs tested. Left: ROIs include components

of primary auditory cortex (anterior transverse temporal gyrus; green) and primary visual cortex including calcarine sulcus, occipital pole, and cuneus (increasingly darker shades of blue, respectively). Right: Accuracy of the transfer learning classifier ± SEM as a function of pool size for the auditory and visual ROIs averaged over 21 subjects repeated 25 times. Accuracy of the subject-specific classifier is indicated by the circular markers, left. Theoretical chance accuracy is 50%. Note scale difference from other BCI ROI figures.

3.4.3 Real data: classification performance and effect of training set size

We tested the source-based transfer learning strategy on real EEG data from an attentional switching task. Subjects were instructed to maintain or switch attention between two auditory streams and data from the potential switching period was used to classify if spatial attention was reoriented. The transfer learning classifier increased with training pool size eventually surpassing the mean subject-specific accuracy (Figure 3.8) consistent with P300 and motor imagery simulation results. When using the maximum of nine subjects, the transfer learning method performed 3.7 percent better than the subject-specific approach with an average accuracy of 60.0 percent. The accuracy continued to increase with the number of subjects in the training pool, but, unlike in simulation, improvements did not appear to reach a plateau by the 9^{th} subject.

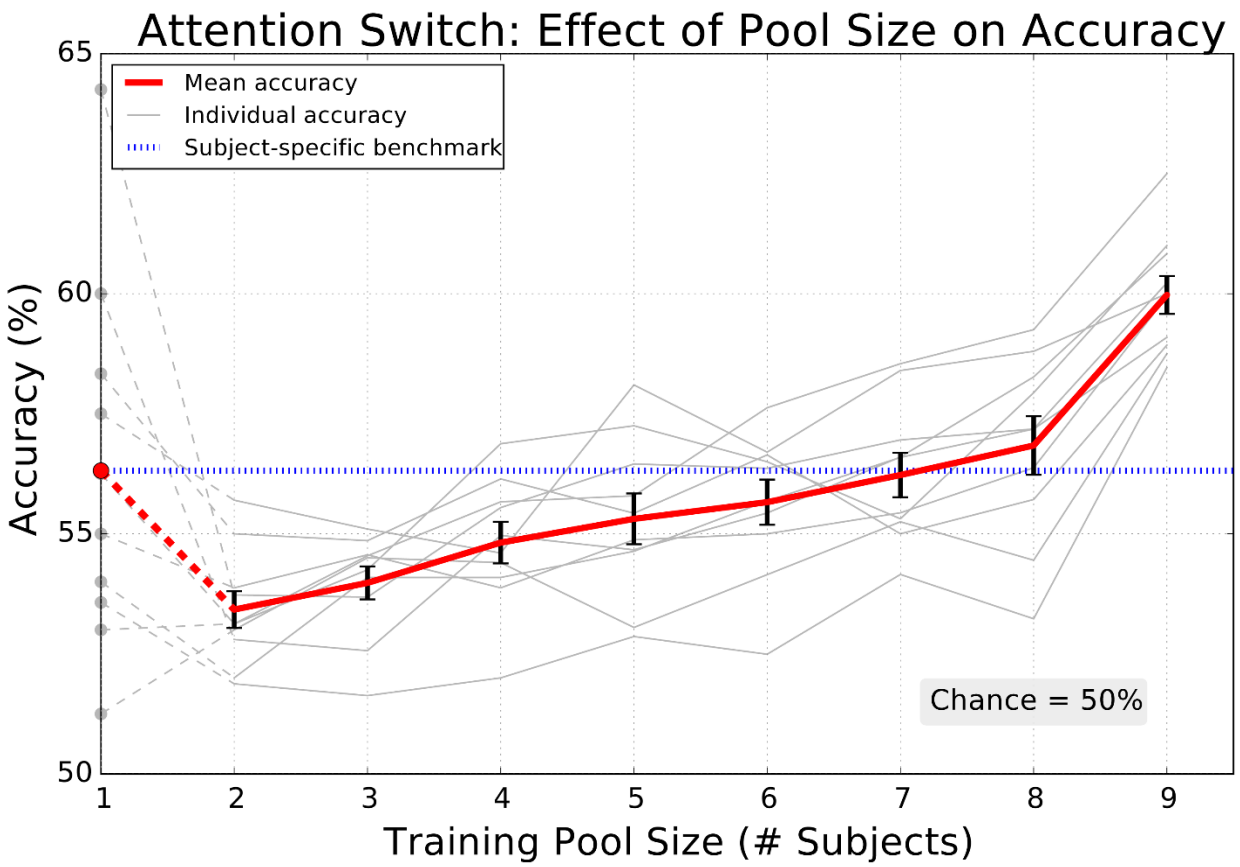

Figure 3.8 The transfer learning classifier achieved incremental performance gains in classifying real data from an attentional switching task and eventually surpassed the subject-specific benchmark. Mean ± SEM (red, black error bars) for 10 individual (gray) accuracies repeated 25 times when classifying attentional switching. Subject-specific classifier accuracy is indicated by the circular markers (left) and the mean is traced across all pool sizes for comparison (blue dotted line). Chance accuracy is 50% and the data consisted of 38 trials per condition per subject on average.

3.5 Discussion

This work demonstrates the feasibility of training a subject-independent BCI classifier (i.e., using data exclusively from other subjects) by transferring brain recordings from other subjects through the source space. The source space allows for the normalization of individual head anatomy and electrode locations,

in contrast to a sensor space approach. We evaluated this approach using simulated cortical activations, which provide ground truth regarding the spatial location and SNR of neural activity, as well as real EEG data from an attentional modulation task. A traditional subject-specific BCI training method, which was trained and tested on each individual's data, served as our benchmark.

Both the simulated and real experiments suggest that a classifier trained with data transferred from fewer than 10 subjects through the source space tends to match or outperform a subject-specific classifier. In simulation, we found that the average classification accuracy was comparable to the benchmark for all realistic SNRs and training set sizes (Figure 3.3) with up to 6.5% improvement (Figure 3.4) and surpassed this subject-specific classifier in some BCI specific regions (Figures 3.5-3.7). By more heavily weighting relevant cortical regions (essentially including functional localization priors), further incremental accuracy improvements were obtained (Figure 3.3b). In real data, the transfer learning method classified attentional reorientation 3.7% better than the subject-specific classifier (Figure 3.8).

This is encouraging for source-based transfer learning as we were unable to find a subject-independent transfer-learning study that reported an accuracy improvement. The majority of studies use *session*-independent data or a combination of subject-specific and subject-independent motor-imagery data and report small accuracy changes from their benchmark (almost all are within ±5%). The only exception that uses a true subject-independent approach (Fazli, et al., 2009) reported small decrease in accuracy, which was subsequently observed elsewhere (Samek, et al., 2013). This suggests that leveraging the tools and priors afforded by the source space can bring advantages to transfer learning (including classification of cortical activity not conventionally used in BCIs).

3.5.1 Source space transfer learning

The primary goal of transfer learning is to shorten or eliminate the time spent calibrating a BCI system by reusing training data. Theoretically, the most informative training data will always come from the subject in the ongoing BCI session, but compiling these training trials consumes a portion of that session's time. Therefore, transfer learning techniques can be advantageous when they merely match the classification accuracy of subject-specific BCIs, or even sustain minor accuracy decreases, because of the time saved.

This is especially relevant in clinical settings where the BCI user may have other health-related time obligations. Past transfer learning research, to the best of our knowledge, has been conducted exclusively in the sensor space, which makes assumptions about the data consistency across subjects and sessions.

A source space approach to transfer learning circumvents many of the assumptions associated with the sensor space. First, transferring spatial filters between subjects in sensor space does not explicitly account for individual variations in anatomical brain and head morphology. In source imaging, the use of individual head models obtained through MRI helps spatially normalize variation in cortical and head morphology and therefore, normalize effects of volume conduction as electromagnetic signals propagate from the cortex to the EEG sensors. Second, sensor space transfer learning assumes that differences in EEG electrode locations will be negligible between sessions. In practice, small shifts between the EEG and brain coordinate frames will result in a different composition of recorded brain activity, which may undermine algorithms based on spatial filtering (Ramoser, et al., 2000; Bashashati, et al., 2007). Coregistering electrodes to each individual's head model normalizes this spatial jitter in the electrode positioning relative to the brain. Third, the electrode montage must be conserved to transfer training data as different electrode arrangements will alter the sensitivity to particular brain areas, especially when the number of electrodes is low. However, conforming to a specific montage is undesirable as it limits the ability to choose other arrangements that are more practical or target pre-determined brain regions. Source imaging allows the computation of source estimates regardless of the montage choice or number of electrodes (within a reasonable limit). Additionally, the work here focuses on transferring the cortical activations themselves (when data are at an earlier step in the BCI processing chain). Therefore, predetermining the feature generation method or classifier is not necessary like in many sensor-based strategies.

3.5.2 Leveraging neuroscience priors in the source space

The source space provides a generalized pathway for connecting BCIs to other neuroscience research. For example, neuroimaging information can help define a hand motor ROI for BCI control as demonstrated here. This is similar to preselecting the C3/C4 electrodes as control signals for an imagined

hand motor task, but the source space allows the controlled targeting of specific ROIs independent of electrode positioning or montage choice. We showed that this targeting of a cortical region through electrode weighting gives incremental gains in transfer learning even with no training trials from the subject of interest. This targeting may aid in the development of BCIs driven by cortical regions beyond the small handful of paradigms that dominate the field (Vansteensel, et al., 2010). As an example, recent research has implicated the right temporoparietal junction (RTPJ) in switching auditory attention between sound sources at different spatial locations (Larson & Lee, 2014; Larson & Lee, 2013). While we used data from this attentional paradigm, a system that is further optimized by using the precise RTPJ location might be used in a hearing aid BCI with spatially selective sound amplification that mirrors a user's attention (Lee, et al., 2013). Even if a neuroscience prior is not available, past work has shown that non-standard target ROIs can be determined statistically (Lotte, et al., 2009a).

Operating in source space also opens the possibility to link findings from BCI research with other areas of neuroscience because of the shared spatial reference frame. Examples of these areas include functional and structural neuroimaging as well as invasive electrophysiological methods (acquiring signals via electrocorticography or microelectrode arrays, for example). Linking sensor-based BCI findings to the cortex is difficult with EEG scalp maps because they are only a proxy to the underlying brain activity. Thus, estimating cortical activity from non-invasive EEG data provides a standardized frame to both exploit knowledge from other areas of neuroscience as BCI priors and relate results with other existing bodies of literature.

3.5.3 The ideal training pool

To investigate the effects of training pool size in transfer learning, we explored a range of transferred training pool sizes. All simulated tests attained the highest proportion of total accuracy gain during first 8 to 10 subjects and diminishing improvements thereafter (with 40 training trials per class per subject). In the real EEG experiment, accuracy consistently increased as the training pool grew but did not appear reach a plateau by the 9th subject suggesting that more improvement might be possible with additional subjects. Additionally, we found that the maximum performance varied depending on the ROI, which may

be a consequence of morphological inconsistencies across subjects leading to different compositions of recorded cortical activity. Another possibility is that electrode coverage over certain ROIs – traditionally, the occipital lobe is a relatively difficult place to obtain reliable electrode contact. Future work focused on subselecting training data for personalized training data pools may yield further accuracy improvements.

3.5.4 Caveats

This work has demonstrated the potential of source-based BCIs, but some practical concerns must be addressed. Source imaging requires that each subject undergo a structural MRI scan to build a source space model. This adds a separate scan session of approximately 20 minutes and scanner costs prior to the first BCI use, but only one MRI scan will typically ever be needed for an adult BCI user as the source space model can be reused. An electrode coregistration procedure is also necessary to create a coordinate transform between the EEG electrodes and the subject's head model. Currently, the 3-dimensional mapping of each electrode adds several minutes to the cap setup, but there are groups actively developing accelerated coregistration techniques. This coregistration process is also likely to have some spatial error, but previous work has suggested that this has a minor impact on source localization (Larson, et al., 2014). Finally, an understanding of source imaging and associated software tools is necessary to generate the forward and inverse solutions. There already exists an active online community, however, where the tools (including all of those used in this work) are openly developed and shared with tutorials.

There are some limitations of the specific approach used here. This work focused on determining the feasibility of a source-based transfer learning for BCIs that spatially normalized head morphology across subjects. To have ground truth concerning the location and SNR of activity, the simulated data were comprised of single time point "snapshots" of cortical activity (in conjunction with actual head and brain models). This simplification allowed us to manipulate and evaluate spatial information in isolation but it did not explore temporal dynamics or functional variation across subjects. Past work has also detailed EEG nonstationarities arising from differences in mental strategy, fatigue, or attention (Krauledat, et al., 2007; Baillet, et al., 2001; Samek, et al., 2013; Lotte, et al., 2007) that were not considered in our

simulation, but stationarity issues remain open questions throughout many areas of BCIs and neuroimaging (see Tsai, et al., 2014 for some early work on addressing this issue). Even so, these time averaging and non-stationary issues were present in the real (attentional) EEG data analyzed, and we still achieved classification results consistent with our simulation experiment despite these concerns. We also took transferred data to the sensor space before classification for computational simplicity, but previous research has demonstrated that classification directly in source space is possible (Besserve, et al., 2011; Qin, et al., 2004; Cincotti, et al., 2008).

3.6 Conclusions

In this work, we developed and evaluated a method to train a BCI classifier exclusively with data transferred from other subjects through the source space. Within transfer learning, utilizing the source space obviates many assumptions made in sensor space regarding both head morphology and electrode positioning and provides the opportunity to more easily incorporate neuroscience priors. Using this approach, we obtained higher average (single-trial) classification accuracy than a traditional approach for both simulated and real EEG data even in cortical regions not classically used in BCIs. Future work will focus on applying this source-based transfer learning methodology in a real-time paradigm and further exploring the benefits conferred by source imaging in BCIs.

Acknowledgements

This research was funded by the National Science Foundation Graduate Research Fellowship Program (MW; DGE-1256082), the Department of Defense Air Force Office of Scientific Research Young Investigator Program (AKCL; FA9550-12-1-0466), and the National Institutes of Health Pathway to Independence award (AKCL; R00DC010196).

3.7 References

Baillet, S., 2010. The Dowser in the Fields: Searching for MEG Sources. In: *MEG: An Introduction to Methods.* New York: Oxford University Press, Inc., pp. 83-123.

Baillet, S., Mosher, J. C. & Leahy, R. M., 2001. Electromagnetic Brain Mapping. *IEEE Signal Processing Magazine,* Issue Nov., pp. 14-30.

Bashashati, A., Fatourechi, M., Ward, R. K. & Birch, G. E., 2007. A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *Journal of Neural Engineering,* 4(2), pp. 32-57.

Besserve, M., Martinerie, J. & Garnero, L., 2011. Improving quantification of functional networks with EEG inverse problem: Evidence from a decoding point of view. *NeuroImage,* 55(4), pp. 1536-47.

Blankertz, B. et al., 2008. The Berlin Brain-Computer Interface: Accurate performance from first-session in BCI-naive subjects. *IEEE Transactions on Biomedical Engineering,* 55(10), pp. 2452-62.

Chi, Y. M., Jung, T.-P. & Cauwenberghs, G., 2010. Dry-Contact and Noncontact Biopotential Electrodes: Methodological Review. *IEEE Reviews in Biomedical Engineering,* Volume 3, pp. 106-19.

Cincotti, F. et al., 2008. High-resolution EEG techniques for brain-computer interface applications. *Journal of Neuroscience Methods,* 167(1), pp. 31-42.

Dale, A. M., Fischl, B. & Sereno, M. I., 1999. Cortical Surface-Based Analysis: I. segmentation and surface reconstruction. *NeuroImage,* 9(2), pp. 179-94.

Dale, A. M. & Sereno, M. I., 1993. Improved Localization of Cortical Activity by Combining EEG and MEG with MRI Cortical Surface Reconstruction: A Linear Approach. *Journal of Cognitive Neuroscience,* 5(2), pp. 162-76.

Destrieux, C., Fischl, B., Dale, A. & Halgren, E., 2010. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage,* 53(1), pp. 1-15.

Devlaminck, D. et al., 2011. Multisubject Learning for Common Spatial Patterns in Motor-Imagery BCI. *Computational Intelligence and Neuroscience,* Volume 2011, pp. 1-9.

Fazli, S. et al., 2009. Subject-independent mental state classification in single trials. *Neural Networks,* 22(9), pp. 1305-12.

Fischl, B., Sereno, M. I., Tootell, R. B. & Dale, A. M., 1999. High-Resolution Intersubject Averaging and a Coordinate System for the Cortical Surface. *Human Brain Mapping,* 8(4), pp. 272-84.

Friedman, J. H., 1989. Regularized Discriminant Analysis. *Journal of the American Statistical Association,* 84(405), pp. 165-75.

Friston, K. J. et al., 1995. Spatial Registration and Normalization of Images. *Human Brain Mapping,* 3(3), pp. 165-89.

Gramfort, A. et al., 2013. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience,* Volume 7, pp. 1-13.

Gramfort, A. et al., 2014. MNE software for processing MEG and EEG data. *NeuroImage,* Volume 86, pp. 446-60.

Hämäläinen, M. et al., 1993. Magnetoencephalography - theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics,* 65(2), pp. 413-97.

Hämäläinen, M. S. & Ilmoniemi, R. J., 1984. *Interpreting Measured Magnetic Fields of the Brain: Estimates of Current Distribution,* Helsinki, Finland: Helsinki University of Technology.

Hämäläinen, M., Lin, F.-H. & Mosher, J., 2010. Anatomically and Functionally Constrained Minimum-Norm Estimates. In: *MEG: An Introduction to Methods.* New York: Oxford University Press, Inc., pp. 186-215.

Kang, H., Nam, Y. & Choi, S., 2009. Composite Common Spatial Pattern for Subject-to-Subject Transfer. *IEEE Signal Processing Letters,* 16(8), pp. 683-6.

Kindermans, P.-J., Tangermann, M., Müller, K.-R. & Schrauwen, B., 2014. Integrating dynamic stopping, transfer learning and language models in an adaptive zero-training ERP speller. *Journal of Neural Engineering,* 11(3), pp. 1-9.

Krauledat, M., Schröder, M., Blankertz, B. & Müller, K.-R., 2007. Reducing Calibration Time for Brain-Computer Interfaces: A Clustering Approach. *Adv. Neural Inf. Process. Syst.,* pp. 753-60.

Krauledat, M., Tangermann, M., Blankertz, B. & Müller, K.-R., 2008. Towards Zero Training for Brain-Computer Interfacing. *PLoS One,* 3(8), pp. 1-12.

Larson, E. & Lee, A. K. C., 2013. The cortical dynamics underlying effective switching of auditory spatial attention. *NeuroImage,* Volume 64, pp. 365-70.

Larson, E. & Lee, A. K. C., 2014. Switching auditory attention using spatial and non-spatial features recruits different cortical networks. *NeuroImage,* Volume 84, pp. 681-687.

Larson, E., Maddox, R. K. & Lee, A. K. C., 2014. Improving spatial localization in MEG inverse imaging by leveraging intersubject anatomical differences. *Frontiers in Neuroscience,* Volume 8, pp. 1-11.

Lee, A. K. C., Drews, M. K., Maddox, R. K. & Larson, E., 2013. Brain Imaging, Neural Engineering Research, and Next-Generation Hearing Aid Design. *Audiology Today,* pp. 40-7.

Lotte, F. et al., 2007. A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of Neural Engineering,* 4(2), pp. R1-R13.

Lotte, F. & Guan, C., 2010. Learning from other subjects helps reducing brain-computer interface calibration time. *2010 IEEE International Conference on Acoustics Speech, and Signal Processing,* Volume 1, pp. 614-7.

Lotte, F., Guan, C. & Ang, K. K., 2009. Comparison of Designs Towards a Subject-Independent Brain-Computer Interface based on Motor Imagery. *Conference Proceedings for the IEEE EMBS,* pp. 4543-6.

Lotte, F., Lécuyer, A. & Arnaldi, B., 2009. FuRIA: An Inverse Solution Based Feature Extraction Algorithm Using Fuzzy Set Theory for Brain-Computer Interfaces. *IEEE Transactions on Signal Processing,* 57(8), pp. 3253-63.

Polich, J., 2007. Updating P300: An Integrative Theory of P3a and P3b. *Clinical Neurophysiology,* 118(10), pp. 2128-48.

Qin, L., Ding, L. & He, B., 2004. Motor imagery classification by means of source analysis for brain-computer interface applications. *Journal of Neural Engineering,* 1(3), pp. 135-41.

Ramoser, H., Müller-Gerking, J. & Pfurtscheller, G., 2000. Optimal Spatial Filtering of Single Trial EEG During Imagined Hand Movement. *IEEE Transactions on Rehabilitation Engineering,* 8(4), pp. 441-446.

Samek, W., Meinecke, F. C. & Müller, K.-R., 2013. Transferring Subspaces Between Subjects in Brain-Computer Interfacing. *IEEE Transactions on Biomedical Engineering,* 60(8), pp. 1-10.

Tsai, A. C. et al., 2014. Cortical surface alignment in multi-subject spatiotemporal independent EEG source imaging. *NeuroImage,* Volume 87, pp. 297-310.

Tu, W. & Sun, S., 2012. A subject transfer framework for EEG classification. *Neurocomputing,* Volume 82, pp. 109-116.

Vansteensel, M. J. et al., 2010. Brain-Computer Interfacing Based on Cognitive Control. *Annals of Neurology,* 67(6), pp. 809-16.

Wang, Y., Wang, Y.-T. & Jung, T.-P., 2012. Translation of EEG Spatial Filters from Resting to Motor Imagery Using Independent Component Analysis. *PloS One,* 7(5), pp. 1-12.

Witt, S. T., Laird, A. R. & Meyerand, M. E., 2008. Functional neuroimaging correlates of finger-tapping task variations: An ALE meta-analysis. *NeuroImage,* Volume 42, pp. 343-56.

Wolpaw, J. R. et al., 2002. Brain-computer interfaces for communication and control. *Clinical Neurophysiology,* 113(6), pp. 767-91.

Yang, H. et al., 2014. Detection of motor imagery of swallow EEG signals based on the dual-tree complex wavelet transform and adaptive model selection. *Journal of Neural Engineering,* Volume 11, pp. 1-13.

Chapter 4. Incorporating modern neuroscience findings to improve brain-computer interfaces: detecting network activity

This chapter is a manuscript draft by Mark Wronkiewicz, Nick Foti, Eric Larson, Kurt Weaver and Adrian KC Lee with the working title "Incorporating modern neuroscience findings to improve brain-computer interfaces: interpreting resting state networks."

4.1 Introduction

Brain-computer interfaces (BCIs) are devices that interpret brain activity and use it to generate artificial output (Wolpaw et al., 2002). Because they accomplish this independent of standard biological output pathways (e.g., muscles or peripheral nerves), BCIs have a great deal of potential as a rehabilitation tool after neural damage (e.g., stroke, spinal cord injury, ALS, etc.) or augmentative applications for healthy users. This technology, however, is still limited to a few specific use cases due to a number of remaining obstacles including some related practicality and focus on a few cortical signals.

One obstacle lies in the fact that the large majority of modern BCIs are "synchronous." That is, the speed and timing of the BCI application is predetermined by the system and the user is forced to follow suit. As an example, motor-imagery BCIs often cue users to either imagine hand movement or rest during a series of system-defined time windows instead of allowing users to proceed or take breaks at their own pace. This rigid cadence limits the application of BCIs toward real world scenarios because it is unnatural for a number of tasks (Scherer et al., 2008; Mason et al., 2012) and may quicken fatigue (Fazli et al., 2009). Beyond these considerations, synchronous BCIs suffer from additional problems. For example, when a user is truly not engaged in the task, a BCI that attempts to interpret task-related activity may adapt to incorrect brain recordings or waste battery if the system is wireless. Further, if a user's attention lapses, synchronous BCIs are also likely to result in unintended actions, which are annoying at best and dangerous at worst (Mason et al., 2012). For these reasons, it is worth developing "asynchronous" BCIs supporting a self-paced operation where the user controls the system's progression.

The core of an asynchronous BCI is the ability to detect "intentional-control" (IC) and "no-control" (NC) states – periods of time when the user either does or does not intend to control the BCI system, respectively (Mason and Birch 2000). By classifying if the user is resting or in some other NC state, the BCI can pause its processing at the will of the user. Some of the first work in this domain was carried out by Mason and Birch (2000) on low frequency (1-4 Hz) signals to classify in a rolling window if subjects were resting or actively moving their hand. Other approaches have used two stage classifiers where the first acts as a gatekeeper (i.e., it allows or prohibits an action) and the second determined the intended movement (Bashashati et al., 2007; Scherer et al., 2008). Still other groups experimented with signal detection methodology, which quantifies the likelihood that the ongoing brain activity belongs to some reference distribution in the feature space. For example, signal detection methods were used to generate an output signal proportional to the distance from a distribution built with data while the subject was resting (Schalk et al., 2008) while others constructed a BCI where the system remained idle unless feature vector was relatively close to regions in the feature space associated with movement (Fazli et al., 2009). Note the difference of these approaches from conventional classification approaches that partition the entire feature space into a finite number of regions and always produce an action for each trial.

In BCIs, there is often limited a priori information to define appropriate brain features for classifying a user's brain state (Schalk et al., 2008) – a point demonstrated by the empirically developed features in previous asynchronous BCIs. Neuroimaging work, however, has uncovered patterns of network activity between certain sets of brain regions during periods of rest that may be relevant for classifying IC and NC states. There are approximately 7-10 of these resting state networks (RSNs), and each comprises a small set of brain regions that tend to vary their functional connectivity depending on task engagement (Damoiseaux et al., 2006). The default mode network (DMN) is one such RSN, which is composed primarily of medial parietal regions, temporal regions, and temporoparietal junctions of both hemispheres. Regions belonging to the DMN demonstrate higher functional connectivity while a subject unengaged with the outside world (e.g., wakeful rest, internally-directed thought, or daydreaming) and suppressed functional connectivity during task participation (Hellyer, et al., 2014). As a contrast, the dorsal attention network (DAN) is another RSN made up of the frontal eye field, inferior parietal sulcus,

and middle temporal area. Previous work strongly implicates it in voluntary attention as it shows increased functional connectivity while a subject is engaged in a task (Corbetta and Shulman, 2002). (For an in-depth review, see van den Heuvel and Pol, 2010 or Rosazza and Minati, 2011). The opposing properties of these RSNs makes them potentially relevant for separating IC and NC brain states in a BCI.

Initial RSN research was carried out with PET and fMRI (Biswal et al., 1995, Raichle et al., 2001, Damoiseaux et al., 2006), but more recent work has begun to be characterized with magnetoencephalography (MEG) making these findings useful in a BCI context. Using MEG, de Pasquale et al., 2010 found modulation of power between DAN regions in theta and beta bands as well as between DMN regions in theta and alpha bands within the resting state. Using a related approaches, Brookes et al., 2011 found the highest functional connectivity between 10 and 30 Hz and Hipp et al., 2012 found highest power envelope correlation in alpha and beta bands for these networks. Therefore, previous work suggests that the correlation in band-limited power (BLP) between these regions may help quantify brain states related to the DMN and DAN.

The spatial extent of RSN cortical regions is characterized in the source space (i.e., at the cortical level). MEG data, however, are recorded in the sensor space (i.e., at or above the surface of the scalp). We used a neuroimaging technique called "source imaging" to project noninvasive MEG data onto a 3D model of the brain thereby connecting the sensor and source spaces (see Sections 1.1 and 1.2 for an in depth explanation). By estimating cortical activity within areas belonging to the DMN and DAN, we were able to leverage the aforementioned neuroimaging research. Toward our goal of classifying IC and NC states, we used the functional connectivity between RSN regions as features to classify whether brain activity came from a time period when the subject was engaged in a task or resting. Note that in contrast to event-related potentials or changes in power of one cortical region, this approach relies on changes in functional connectivity between multiple anatomically separated regions.

4.2 Methods

Data used in this experiment originated from the motor task within the MEG component of the Human Connectome Project (Van Essen, et al., 2013) – a large, freely available dataset. To ease computational constraints, we analyzed data from the first 20 subjects (in terms of subject identification number) of the HCP MEG release. These MEG data were recorded using a MAGNES 3600 (4D Neuroimaging, San Diego, CA) system. Data from the motor task (and not the language or working memory tasks) were analyzed as this is most relevant to existing brain-computer interface research. In this motor task, subjects were instructed to make repeated overt movements that were visually cued by a pacing arrow at the center of the screen. Before each block, the system cued either hand or foot movement and indicated the left or right side of the body to use. The motor task consisted of 2 runs. In each run, there were 8 blocks per limb, and 10 overt movements per block. Each run also included 10 resting blocks where the screen was completely black for 15 seconds. During the recording, electrooculography (EOG), electrocardiography (ECG), and electromyography (EMG) of hand and foot muscle movements were also recorded. Here, we analyzed the unprocessed version of the MEG data rather than the preprocessed or fully processed versions.

During these tasks, M/EEG data was recorded at approximately 2 kHz. Offline, it was downsampled to 200 Hz and then band-pass filtered between 0.1 and 55Hz. EOG and ECG artifacts were removed with Signal Space Projection (SSP; Uusitalo and Ilmoniemi 1997) computed from channels recording eye and heart activity. Both the task and rest data were then divided into 1 second blocks yielding about 500 task trials and 600 rest trials for each subject.

To apply source imaging, we first computed a forward (or gain) matrix, which models how cortical activity affects each of the 248 MEG magnetometers (Section 1.2). To do this, the sources of cortical activity are modeled using several thousand simulated current dipoles distributed evenly across the entire cortex. Before and after each recording, the head position was recorded. The spatial location of fiducials and the head shape were also recorded using a FASTRAK-III system (Polhemus). This information served

as the basis for a transformation between the head and MEG device coordinate frames so that arbitrary modeled cortical activity could be uniquely computed in the MEG sensors.

Once the forward solution was computed, we generated an inverse solution according to Section 1.2 using the minimum-norm estimate (MNE) algorithm available in the MNE-Python open source software package (martinos.org/mne; Gramfort et al., 2013). Using this inverse solution, we projected MEG data recorded during this task to the source space. See Sections 1.1 and 1.2 for mathematical details. Importantly, the sensor noise estimate **C**, which is necessary for spatial normalization across electrodes in source imaging, was computed using data recorded during an empty room session (rather than activity during intertrial intervals with the subject in the scanner). This ensured that the resting state activity pattern was not inadvertently included in the noise model and partially removed during the source imaging step. In other words, we wished to include non-biological noise in this covariance estimate but not resting brain activity as is typically done in neuroimaging experiments searching for a change in brain activity from rest.

Relevant cortical regions of interest (ROIs) were defined according to Yeo et al., 2011. In that work, resting state data was recorded for 1,000 subjects using functional MRI and the boundaries of seven resting state networks were defined with each containing a handful of regions. We focused on two of these networks – the default mode network (DMN) and dorsal attention network (DAN). Data was analyzed at dipoles belonging to ROIs within these networks. For the DMN, these included the medial prefrontal cortex (MPFC), superior temporal sulcus (STS), posterior cingulate cortex (PCC), temporoparietal junction (TPJ). For the DAN, these included the frontal eye field (FEF), superior parietal lobule (SPL), and medial temporal visual area (MT).

For these ROIs, we first computed a time-frequency estimation at each dipole in the ROIs to obtain the BLP over the course of the trials. Here, we used 5-cycle Morlet wavelets at 10, 12, 16, 20, and 24 Hz. We then averaged over the dipoles in each ROI in each trial to create an average time-frequency estimates. Finally, we calculated the power correlation between ROI pairs using a 0.5 second sliding window. When

attempting to classify which condition (task or rest) the trial originated from, we used features based on BLP in each ROI as well as the correlation of BLP between pairs of ROIs.

To classify each trial, we used Scikit-learn (scikit-learn.org; Pedregosa et al., 2011) to test two standard machine learning techniques: linear support vector machines (SVMs) and random forests. Briefly, SVMs seek to map features into a feature space where the margin between the two classes is maximized. Here, we examined a grid of parameters with C, which determines the smoothness of the decision boundary, ranging in order of magnitude steps from 10^{-6} to 10^{4} and ⋅, which determines the radius of influence of each support vector, ranging in order of magnitude steps from 10^{-6} to 10^{3}. Random forest classifiers train a large number of simple estimators each trained on a subset of the data and aggregate their outputs to generate classification decisions. Here, we tested classifiers over a grid of parameters with 50, 100, 1000, 2000, 5000, and 10000 estimators and maximum number of features ranging in order of magnitude steps from 10^{1} to 10^{7}. In both the SVM and random forest classifiers, we conducted five-fold cross validation. Additionally, the entire classification process was repeated five times and then averaged to obtain a more stable accuracy estimate.

As a control condition, we computed the average BLP at each sensor over the course of each trial. Again, we classified this data using SVMs and random forests to determine their efficacy in signals not transformed with source imaging. Here, the same range of parameters was used to ensure a fair comparison.

4.3 Results

We investigated if a BCI could classify whether data was recorded during a resting or motor period in 1-second trials. Toward this goal, we classified features generated in three ways. First, sensor space data was classified after computing power averaged in time for each MEG channel. Second, source space data was classified after computing power averaged in time for each ROI. Finally, source space data was classified after computing correlation in BLP between relevant ROI pairs in the RSN networks. While we found similar performance in source and sensor space based on average power, the performance using

features based on functional connectivity was much lower and average accuracy was within two percent of chance. The data were classified using both SVMs (Figure 4.1) and random forests (Figure 4.2).

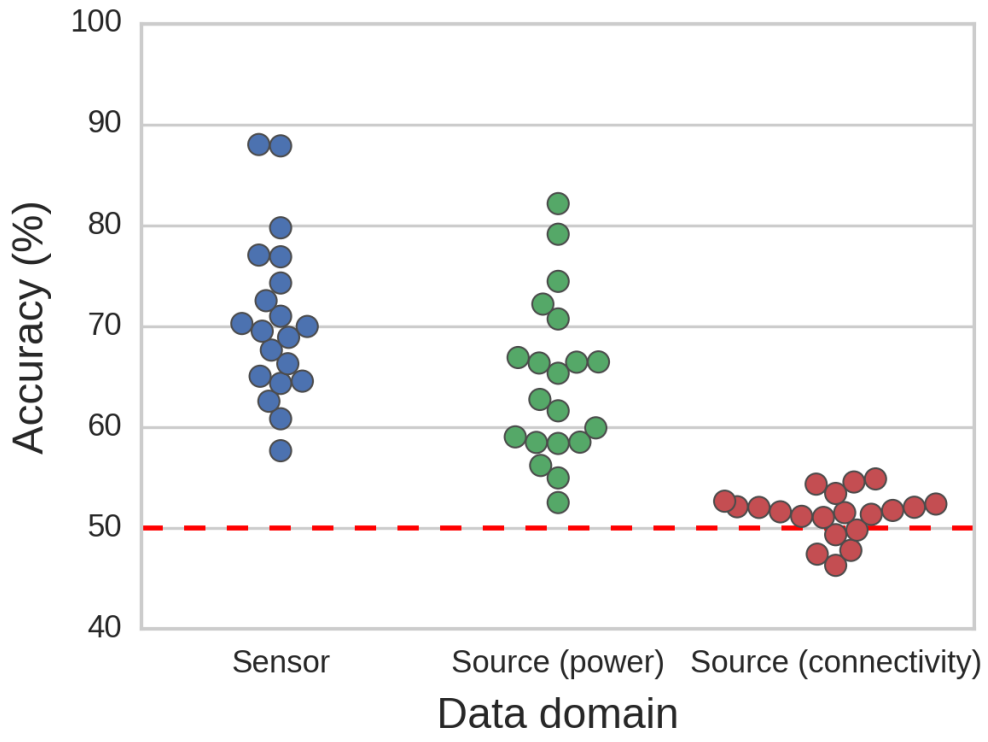

Figure 4.1 Accuracy for 20 subjects when classifying rest and task data from a motor task using support vector machines. Left: Sensor space data consists of average power at each channel. Middle: Source space data consisting of average power within each ROI tested. Right: Source space data consisting of average correlation in BLP between RSN regions. From left to right, the average percent classification accuracy was 70.8, 64.7, and 51.4.

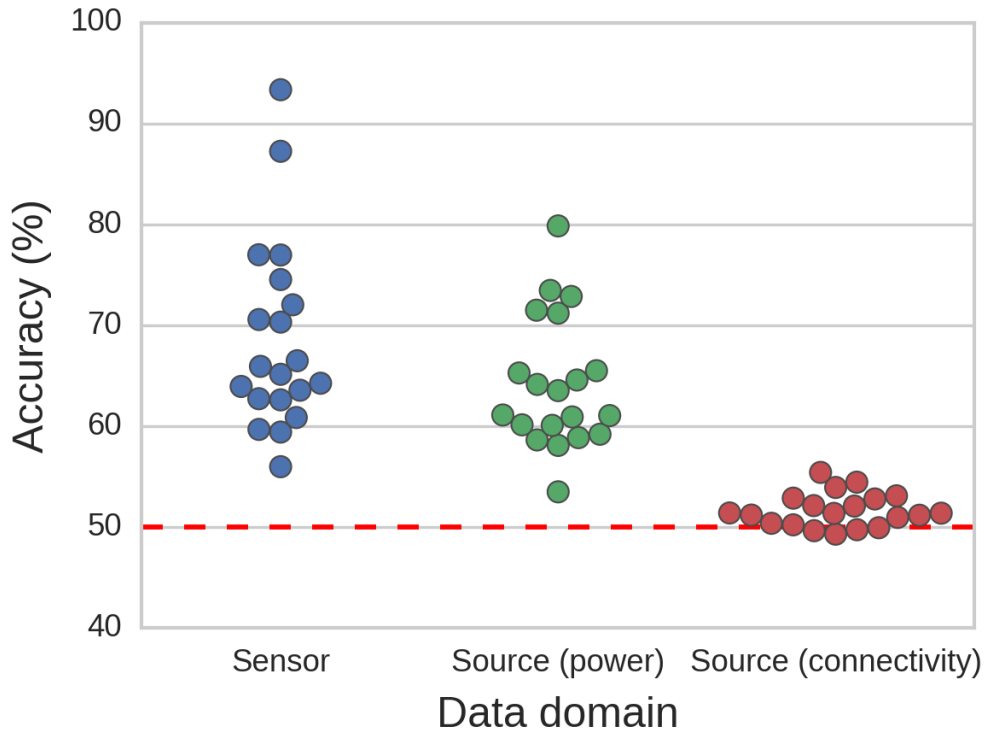

Figure 4.2 Accuracy for 20 subjects when classifying rest and task data from a motor task using random forests. Left: Sensor space data consists of average power at each channel. Middle: Source space data consisting of average power within each ROI tested. Right: Source space data consisting of average correlation in BLP between RSN regions. From left to right, the average percent classification accuracy was 68.6, 64.2, and 51.7.

4.4 Discussion

In this work, we found that, contrary to our hypothesis, correlation between BLP did not serve as an effective feature to distinguish rest from task when using data from an overt motor experiment. More straightforward features, like average BLP in sensors and in RSN ROIs gave substantially better performance (about 15 to 20 percent) than those based on functional connectivity feature when classifying with linear SVMs and random forests. However, there remain a number of future directions that should be explored related the original hypothesis.

Future work should first explore other functional connectivity metrics. Here, we tested correlation in BLP – a simple metric, which is known to suffer from field spread and other issues (van Diessen, et al., 2014). Other metrics like imaginary part of coherence (Nolte, et al., 2004) or Multivariate Interaction Measure (Ewald, et al., 2012) can theoretically account for the effects of field spread. However, there are also many metrics with other properties (e.g., ability to capture linear or nonlinear interactions, operating in the time or frequency domain, and measurement of amplitude or phase information; van Diessen, et al., 2014). Beyond functional connectivity metrics, alternative ROIs should also be investigated. Here, the spatial parcellation used to define relevant RSN ROIs originated from a large fMRI study (Yeo et al. 2011). Because of the methodological differences between signals recorded in fMRI and MEG (e.g., spatial resolution, temporal resolution, and type of signal measured), alternative ROI designations may better capture RSN connectivity. To the best of our knowledge, however, there is no standard electrophysiologically defined parcellation for RSN regions. Finally, alternative classifiers may provide another method of better distinguish brain-states using functional connectivity. As an example, deep learning is capable of discovering higher order patterns within data when given a large training set. Deep learning techniques use successive layers of simple features to build high-level abstractions capable of capturing complex features in data. Thus, these techniques may accomplish the engineering goal of detecting and leveraging patterns in functional connectivity to control a BCI if conventional metrics are not successful. Significantly more experimentation is needed to determine if different functional connectivity metrics, ROI definitions, or classifiers may offer useful improvements for BCIs based on network connectivity.

Despite the results presented here, the search for new types of features to control BCIs is an important endeavor for the BCI field. The majority of non-invasive BCI work relies on a handful of brain activations that were published many decades ago (Chapter 1). This has left a gap between the neuroscience that current BCIs rely on and the present state of the neuroscientific field. Future studies should further explore RSNs and other recently discovered brain activations to address challenges like better detecting the NC state. In the case of asynchronous BCIs, the field must also eventually show the ability to detect the NC state across multiple tasks rather than just one as has been done previously – a goal possible with

the HCP data, which has data from 3 separate tasks. We believe that newly published neuroscience research will continue to offer priors that will help in addressing current BCI challenges in a scientifically principled manner.

Acknowledgements

Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

4.5 References

Bashashati, A., Ward, R. K. & Birch, G. E., 2007. Towards Development of a 3-State Self-Paced Brain-Computer Interface. *Computational Intelligence and Neuroscience,* Volume 2007, pp. 1-8.

Biswal, B., Yetkin, F. Z., Haughton, V. M. & Hyde, J. S., 1995. Functional Connectivity in the Motor Cortex of Resting Human Briain Using Echo-Planar MRI. *Megnetic Resonance in Medicine,* 34(9), pp. 537-541.

Brookes, M. et al., 2011. Investigating the electrophysiological basis of resting state networks using magnetoencephalography. *PNAS,* 108(40), pp. 16783-8.

Corbetta, M. & GL, S., 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience,* Volume 3, pp. 201-15.

Damoiseaux, J. et al., 2006. Consistent resting-state networks across healthy subjects. *Proceedings of the National Academy of Sciences,* 103(37), pp. 13848-13853.

de Pasquale, F. et al., 2010. Temporal dynamics of spontaneous MEG activity in brain networks. *PNAS,* 107(13), pp. 6040-45.

Ewald, A. et al., 2012. Estimating true brain connectivity from EEG/MEG data invariant to linear and static transformations in sensor space. NeuroImage, 60(1), pp. 476-88.

Fazli, S. et al., 2009. Using Rest Class and Control Paradigms for Brain Computer Interfacing. In: J. Cabestany, F. Sondoval, A. Prieto & J. Corchado, eds. *Bio-Inspired Systems: Computational and Ambient Intelligene. IWANN 2009.* Berlin, Heidelberg: Springer, pp. 651-65.

Gramfort, A. et al., 2013. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience,* Volume 7, pp. 1-13.

Hellyer, P. J. et al., 2014. The Control of Global Brain Dynamics: Opposing Actions of Frontoparietal Control and Default Mode Networks on Attention. Journal of Neuroscience, 34(2), pp. 451-61.

Hipp, J. et al., 2012. Large-scale cortical correlation structure of spontaneous oscillatory activity. *Nature Neuroscience,* 15(6), pp. 884-90.

Mason, S. G., Allison, B. Z. & Wolpaw, J. R., 2012. BCI Operating Protocols. In: *Brain-Computer Intefaces: Principles and Practice.* New York: Oxford University Press, pp. 189-95.

Mason, S. G. & Birch, G. E., 2000. A Brain-Controlled Switch for Asynchronous Control Applications. *IEEE Transactions on Biomedical Engineering,* 47(10), pp. 1297-1306.

McCloy, D. R. et al., n.d. Pupillometry shows the effort of auditory attention switching. *Submitted to the Journal of the Acoustical Society of America.*

Nolte, G. et al., 2004. Identifying true brain interaction from EEG data using the imaginary part of coherency. Clinical Neurophysiology, 115(10), pp. 2292-307.Pedregosa, F. et al., 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research,* Volume 12, pp. 2825-30.

Raichle, M. E. et al., 2001. A default mode of brain function. *Proceedings of the National Academy of Sciences,* 98(2), pp. 676-82.

Rosazza, C. & Minati, L., 2011. Resting-state brain networks: literature review and clinical applications. *Neurological sciences,* 32(5), pp. 773-85.

Scherer, R. et al., 2008. Toward Self-Paced brain-Computer Communication: Navigation Through Virtual Worlds. *IEEE Transactions on Biomedical Engineering,* 55(2), pp. 675-82.

Uusitalo, M. A. & Ilmoniemi, R. J., 1997. Signal-space projection method for separating MEG or EEG into components. *Medical and Biological Engineering and Computing,* 35(2), pp. 135-40.

van den Heuvel, M. P. & Hulshoff Pol, H. E., 2010. Exploring the brain network: a review on resting-state fMRI functional connectivity. *Eur. Neuropsychopharmacol.,* 20(8), pp. 519-34.

van Diessen, E. et al., 2014. Opportunities and methodological challenges in EEG and MEG resting state functional brain network research. Clinical Neurophysiology.

Wolpaw, J. R. et al., 2002. Brain-computer interfaces for communication and control. *Clinical Neurophysiology,* 113(6), pp. 767-91.

Yeo, B. T. et al., 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology,* 106(3), pp. 1125-65.

Chapter 5. Outlook and future directions

This chapter contains a small amount of material by Mark Wronkiewicz, Eric Larson, and Adrian KC Lee as a book chapter for the book titled "Brain-Computer Interfaces Handbook: Technological and Theoretical Advances" with the title "Facilitating the integration of modern neuroscience into non-invasive BCIs" scheduled to be published in 2017.

5.1 Next steps in BCIs

Given that many of the ideal BCI applications have yet to be realized, the field still has a great deal of room to expand. However, we have many staunch problems to solve before BCIs technology will advance out of the laboratory in any real sense. As we have shown in Chapters 2-4, it is possible that the field's current strategy of empirically developing feature detection and classification algorithms is not the optimal approach. Our work instead suggests that basic neuroscience knowledge and techniques may quicken progress on these issues, and that source imaging is a useful bridge between neuroscience and neuroengineering. Fully adopting a source imaging pipeline may be untenable for some groups, but note that the barrier for entry is lowering in the face of open source toolboxes and a focus on developing tutorials. For many, a smaller but meaningful step may be to simply revisit the neuroscience papers describing BCI signal features – including the signals used in conventional (e.g. SSVEP, motor-imagery, and P300) BCIs. Neuroscience research, however, has grown well beyond the signals use in the canonical BCIs leaving a backlog of untapped insights that should be investigated for new potential BCI paradigms.

This dissertation should also make it clear that the quantitative relationship between brain activity and EEG electrodes changes with every subject and session. Although it is tempting to incorporate subjective choices into a processing pipeline to squeeze out every point of accuracy, these source of variability have the potential to corrupt any resulting findings. Instead, a general goal should always be to automate BCIs as much as possible to facilitate reproducibility and scalability of any discoveries; we have shown how neuroscience offers one path to accomplishing this, and with some moderate transitions in BCI

methodology, the field may again enjoy advancements at a pace not seen since its original expansion. Despite making some progress (as discussed in this dissertation), we have raised many more questions and future lines of BCI research than we actually explored and published. The rest of this section is dedicated to describing those research opportunities and other points relevant for future BCI work both inside and outside of our lab.

5.2 Alternative and individualized electrode montages

Typically in EEG, electrodes are positioned accord to the International 10-20 system. This system was defined by an international consortium and led by Herbert Jasper (Jasper, 1958). Along with other standards outlined in his report, Jasper and other researchers suggested the 10-20 electrode placement map to provide "adequate coverage of all parts of the head" and facilitate comparison across laboratories. The name arises from the 10 or 20 percent spatial increments between electrodes on the scalp, but it is worth pointing out that this electrode system may not be ideal for all situations – a point explicitly made by the original authors. This layout was initially designed to optimize spatial coverage of the scalp in a clinical setting. The fact that this original set of guidelines also discusses "pen deflections" and provides recommendations on suggested paper speed of the recording device both highlight its age. Modern neuroscience and BCI work, however, may benefit from different electrode placements to optimally capture relevant brain signals. New tools like the electromagnetic Polhemus FastTrak and BrainVision's stereoscopic camera Captrak system can also record electrode 3D location at sub-millimeter precision. This allows experimenters to deviate from the 10-20 system while still providing precise electrode configuration for comparing experiments between research groups. This opportunity to experiment with electrode positioning could lead to a number of interesting research hypotheses, two of which I think are worth exploring in the lab.

First, forward modeling can predict the suitability of a montage (standard or not) for capturing activity from any ROI and do so on an individual-by-individual basis – a goal some have just begun to pursue

(Zich, et al., 2015). Quantifying the expected SNR at each electrode will facilitate the BCI prototyping process while also reducing trial and error experimentation in selecting the best set of EEG electrodes or an optimal spatial filtering technique. For example, a few groups have developed mobile electrode montages that fit behind the ear (Do Valle, et al., 2014; Debener, et al., 2015). Forward modeling could be used to predict how appropriate these experimental electrode montages are for a given paradigm (e.g., motor-imagery or P300 systems). Evaluating the overall sensitivity of an electrode montage to a given cortical ROI in a particular subject (as we did in Figure 2.4) may also shed some light on why certain users are unable to control certain BCIs – a problem sometimes referred to as "BCI illiteracy."

Second, the forward modeling process could be used to construct optimal electrode montage; in other words, it could be used to calculate an electrode arrangement specifically engineered for a specific ROI rather than just quantify how receptive existing montages are. As before, this approach could take into account an individual subject's cortical geometry. A first pass, for example, might employ a greedy algorithm that successively positions 10 electrodes at scalp location where a predefined ROI's influence is highest subject to the constraint that it is at least 1cm from any other electrodes. A second more sophisticated version might position 10 electrode montage as a complete set (i.e., at once) while maximizing some information theoretic optimization function. The theoretical component here (i.e., calculating an optimal montage) would only require basic source imaging skills to complete making it appropriate for a student learning how use source imaging. Any positive results regarding this custom montage should then be tested against the International 10-20 system in an online BCI. Eventually, the lab could use 3D-printing or some other customizable cap construction techniques to personalize an optimal montage for each ROI-subject pair.

5.3 Real-time implementation: the proof is in the pudding

My biggest regret is not having the opportunity to validate our research results in an online BCI. We took some important theoretical steps in terms of incorporating neuroimaging into BCIs and showed their

promise with pre-recorded and simulated data, but the ultimate test remains real-time implementation. In order to evaluate our approach to an online setting, our lab will require a new PhD student or post-doc dedicated to this task (if not more than one). This section outlines some of my suggested goals once we have the personnel in place.

Our work targeting activity from a cortical region – published in our 2016 Journal of Neural Engineering paper (and Chapter 2) – is likely the best starting point for an online BCI. To briefly review this project, we used a previous neuroimaging publication (Larson, 2013) that found the RTPJ to be significantly more active during switching of auditory spatial attention. Focusing on activity from this region provided a significant classification boost when tracking subjects' attention. As a first step, this strategy should be repeated in a classical BCI paradigm. By avoiding a more experimental paradigm, I think we would have the best chance at gaining the attention of the BCI community and convincing them of the improvements to be had when using a source-based strategy. While better BCI tasks more likely exist (in terms of information that might be transferred, accuracy, reliability, etc.), it is important to show that a source-based approach is effective with one of the venerable BCI paradigms that the field is comfortable with.

Toward this goal, the P300 speller paradigm may be a good starting point for a few reasons. First, there is room for improvement; typically, the P300 speller will recorded average activity over 5-10 trials to increase the SNR before selecting a letter. Any significant reduction in the number of trials will translate to immediate gains in the information transfer rate. Second, the P300 ERP is relatively diffuse in the sensor space. However, some work has been done to localize it on the cortex (Polich 2007); because the ability to spatially target relevant activity in source space is the core of our work, we can expect more room for improvement given that the spatial signature is better known in the source space than sensor space. Thirdly, progress on the P300 BCI is an extremely relevant BCI for locked-in patients – perhaps the most appropriate end-user given today's state of rehabilitation-focused BCIs.

A second canonical BCI potentially worth testing is one based on motor-imagery. Unlike the P300 event-related potential, motor-related activity is well-researched in terms of spatial extent. Therefore, there may be leave less room for improvement because motor cortex is spatially consistent across subjects –

hence the pattern of previous work that use C3/C4 electrodes (over motor cortex) to assess motor intent. Regardless, this paradigm is very well-established, so any gains could be easily communicated to the BCI community.

As far as experimental details, early work should focus on testing the source-imaging technique on a conventional BCI paradigm with 10-12 subjects that we currently have head models for (to avoid additional structural MRI scans). This work should include two separate sessions per subject – one based on standard sensor space techniques and one using our experimental source space approach. There are several software options for displaying information to the user including BCPy2000 (developed by researchers at Max-Planck), Wyrm (developed by the Berlin-BCI group), and our own MNE-Python library. As the former two do not appear to be actively maintained, I recommend expanding MNE-Python's online BCI support. With the lab's extensive background in designing behavioral paradigms, I think we are capable of designing a better controlled task than is available in the first two software packages mentioned.

Assuming the source space generates some promising results, the data recorded P300 or motor-imagery experiment would immediately set the stage for validating our transfer learning work (with only minor additional data collection). Briefly, our 2015 transfer-learning paper used the source space to estimate how activity recorded from one subject would manifest for another subject's head model. By accounting for several sources of variability (e.g., head shape, brain morphology, and electrode positioning), we were able to train a classifier using only other subjects brain recordings (i.e., a subject-independent classifier). Data from the 10-12 subjects recorded as part of the aforementioned source-based BCI work would be useful in validating this approach. I recommend running the exact same BCI task for approximately 5 more subjects with a classifier that is trained only with data from the original 10-12 subjects. This should be done using a conventional transfer learning technique as a control (perhaps chosen from those outlined in Lotte et al., 2015) as well our experimental source-based transfer learning technique.

5.4 Human connectome project, network BCIs, and deep learning

Data from the Human Connectome Project (HCP) also presents an opportunity to continue the lab's source-based BCI line of research. It consists of a massive publically available neuroimaging dataset that includes experiments with EEG and MEG recordings for 100 subjects. Along with a number of fMRI recordings and subject metadata, the HCP organizers are obtaining structural MRI information (permitting source estimation). While the recordings are still ongoing, a portion of the planned experiments were complete for almost all of the planned subjects; for 60 of those subjects, both resting state and task data is already recorded, preprocessed, and available online. Therefore, this dataset is an excellent resource for exploring functional connectivity metrics as features for a BCI – the final aim of my thesis work. Analyzing the complex coordination between sets of brain regions, rather than just the activation of single ROIs, may provide a window into more abstract states of the user (e.g., if the user is resting or engaged in a task), which would represent an important step toward developing asynchronous BCIs (as discussed in Chapter 4). Assessing network activity will likely require a large dataset (like that available with the HCP) to better characterize nuanced functional connectivity features. Additionally, the HCP data set consists of multiple different tasks involving, for example, working memory, motor activity, and language processing. Using this data, resting state networks that the lab should consider investigating include the default mode network, dorsal attention network, motor network, and working memory network. Using the activity of these networks, potential grant project might be pursued to correctly identify rest and task trials, or, more ambitiously, the type of task being performed, purely from electrophysiological recordings of coordinated cortical areas.

Deep learning is another technology that should be further explored in BCIs (and potentially with the HCP data). It represents a machine learning technique that builds many layers of abstraction – typically from very large data sets – and can drastically outperformed standard machine learning under the right conditions (e.g., as has been shown on image and speech recognition problems). This represents a promising alternative to the typical approach of feeding human engineered features to LDA- or SVM-based classifiers. It is worth noting that perhaps the biggest advancement in BCIs came with the original

application of machine learning to BCIs – this provided a crucial step going from operant BCIs to those trained with machine learning. The former required the human to adapt to the system taking weeks of training to gain control while the latter took advantage of statistics to adapt to the user within tens of minutes – a major improvement in practicality. For this reason, major advances in machine learning should be explored in a BCI context as they could very well fuel a new round of advancement.

5.5 The peril of overpromising

A final thought I feel compelled to discuss is the dynamic between BCI research and the media. Recently, there have been a growing number of projects focused on making a media splash *before* publication (rather than after). While discussion sections in BCI papers have also always been full of grandiose dream applications that we are far from realizing, this is not a problem in and of itself; the peer-review process also pushes researchers to temper these dreams applications by noting the challenges that remain (if the writers do not acknowledge them themselves). The problem arises, however, when grand ideas are described in detail to the public before peer review or before any results are even obtained.

A particularly egregious example was Miguel Nicolelis's anticlimactic BCI demonstration at the start of the 2014 World Cup. While his project is certainly not the only neuroengineering example of press before publication, it is especially salient because the research he regularly promised in high-profile media releases never panned out. He supplied a crescendo of press including professional quality video simulations (NewsDirect, 2014) and excessive press explaining how a paralyzed patient would stand up, walk several meters, and kick a ball with a brain-controlled exoskeleton. Despite months of promotion, the demonstration failed to accomplish any of these goals. Rather than standing for a wheelchair, the user was carted onto the field with a golf cart before the ceremony (de Hond, 2014). Instead of walking up to the ball using a brain-controlled exoskeleton, two technicians physically supported the user and his backpack throughout the event as he stood still. Finally, all that was broadcast during the critical kick is a

3-second clip of the BCI user and the soccer ball rolling away from his foot – the footage never actually shows the foot and ball make contact. Nevertheless, Nicolelis claimed victory tweeting, "We did it!!!!" right after.

Of course, many would point out that this was supposed to be an absolutely unprecedented BCI demonstration on the world-stage. They might explain that is reasonable to expect some unforeseeable problems given these ambitious goals as outlined in a video produced by the NSF (National Science Foundation, 2014). Those people would probably also mention that his grand plan was meant to provide hope to those all over the world who suffer from neural damage, so perhaps visionaries should be given some leeway – a bit of creative license to accommodate their lofty goals.

This attitude, however, can be dangerous. A focus on press before publication is not new to science, but I think our field has some flagrant examples of the moral dangers with this tactic – the aforementioned World Cup example is one that would make an excellent case study in an undergraduate course on ethics. These videos have a number of disheartening comments on Youtube including disabled persons asking how much the device will cost or expressing hope that his work will soon allow a spouse to walk again soon. A Dutch paraplegic publically responded to the anticlimactic performance in a video and detailed how existing methods (not under brain control) already perform better than what was promised at the World Cup (de Hond, 2014), and are clinically more appropriate for the large majority of paralyzed individuals. From a scientific standpoint, Nicolelis's World Cup demonstration also does not to advance the state of the art; exoskeletons were already being tested for paralyzed individuals, and motor-imagery BCIs using EEG signals were first published by Wolpaw in 1991. In the end, this rush to simulate and promise the end results before peer reviewed publication cheapens the scientific process and comes off as an empty promise to the public. Our field's leaders need to show younger neuroengineers how to keep the horse before the cart – especially when their work has potential to significantly improve an injured person's quality of life. Given the number of real challenges that remain before BCIs become a mainstream technology, we must focus on the gritty but necessary work that actually advances the field toward a solution instead of succumbing to pomp and glamor.

5.6 References

de Hond, Marc, 13 June 2014. What went wrong when the paraplegic man in robotic suit kicked off the World Cup 2014? *Youtube.* Retrieved from https://www.youtube.com/watch?v=WaQcC8yJmMU.

Debener, S., Emkes, R., De Vos, M. & Bleichner, M., 2015. Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear. *Scientific Reports,* Volume 5, pp. 1-11.

Do Valle, B., Cash, S. S. & Sodini, C. G., 2014. *Wireless behind-the-ear EEG recording device with wireless interface to a mobile device (iPhone/iPod touch).* Chicago, IL, 2014 36th Annual International Conference on the IEEE EMBS.

Jasper, H. H., 1958. Report on the Committee on Methods of Clinical Examination in Electroencephalography: 1957. *Electroencephalography and Clinical Neurophysiology,* 10(2), pp. 370-5.

Larson, E. & Lee, A. K. C., 2013. The cortical dynamics underlying effective switching of auditory spatial attention. *NeuroImage,* Volume 64, pp. 365-70.

Lotte, F., 2015. Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based Brain-Computer Interfaces. *Proceedings of the IEEE,* 103(6), pp. 871-90.

National Science Foundation, 9 June 2014. World Cup exoskeleton allows paraplegic to walk again. *Youtube*. Retrieved from https://www.youtube.com/watch?v=6WO71e0XLqs.

News Direct, 6 February 2014. Brain-controlled exoskeleton will allow paralyzed teenager to kick off the World Cup 2014. *Youtube*. Retrieved from https://www.youtube.com/watch?v=MVrFkDGg_nI.

Polich, J., 2007. Updating P300: An Integrative Theory of P3a and P3b. *Clinical Neurophysiology,* 118(10), pp. 2128-48.

Wolpaw, J. R., McFarland, D. J., Neat, G. W. & Forneris, C. A., 1991. An EEG-based brain-computer interface for cursor control. *Electroencephalography and Clinical Neurophysiology,* 78(3), pp. 252-9.

Zich, C., De Vos, M., Kranczioch, C. & Debener, S., 2015. Wireless EEG with individualized channel layout enables efficient motor imagery training. *Clinical Neurophysiology,* 126(4), pp. 698-710.