A Physical Paradigm for Bidirectional Brain-Computer Interfaces

A dissertation

submitted by

Samuel W. Hincks, B.A., M.S.

In partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

in

Computer Science

TUFTS UNIVERSITY

August 2019

Advisor: Robert J.K. Jacob

Abstract

This dissertation deepens research into interfaces that supplement input the user transmits to the computer intentionally with an auxiliary channel describing ongoing brain activation. Existing implementations of such *implicit brain-computer interfaces* (BCI) depend on machine learning algorithms trained to distinguish physiological signals detected with functional near-infrared spectroscopy (fNIRS) under different task conditions, which subsequent chapters will refer to as neuracles. When calibrated to the user's brain, the *implicit BCI* adjusts settings in the interface to better match the mental state that has unfolded. Because this approach does not depend on an understanding about the relationship between fNIRS signals and physical activity in the brain, I will refer to the current methodology for studying and building *implicit BCIs* as the agnostic paradigm. Experiments evaluating *implicit BCIs* in the agnostic paradigm have led to measurable improvements in user performance in a number of controlled laboratory experiments.

This dissertation introduces a descendant of *implicit BCI*, referred to as a bidirectional BCI. Instead of adapting the interface to *match* the mental state that has unfolded, a bidirectional BCI strives to adapt outputs to the brain to stimulate and maintain optimal mental states for its user. This new class of BCI depends on discovering a model for the interaction between brain and computer at four levels of analysis. Such a model should account for how the brain works at the *physical level*, the linkup between brain state and mental state at a *mental level*, the relationship between brain state and sensor data at the neuracle level, as well as how computer settings and output affect the physical state of the brain at an *interface level.* With a synchronized model at these four levels, a bidirectional BCI can establish a feedback loop between the user's brain and its methods to affect the brain's state, and deploy machine learning algorithms to adjust output to the brain to coerce and sustain desirable mental states.

But bidirectional BCIs are not possible with the existing agnostic paradigm. This dissertation therefore develops an alternative method, which has a synchronized understanding of brain-computer interaction at *physical*, *mental*, neuracle, and *in-*

terface levels. Scientific progress towards physical neuracles depends on methods for studying one's own brain as a scientist and engineer as well as the brains of other lucid individuals synchronized on a common vocabulary for describing mental states. This alternative methodology is facilitated by the *Neuracle* software distributed as part of this dissertation, which consists of interfaces, visualizations, and signal processing algorithms needed in a physical paradigm for BCI.

This dissertation makes several contributions to the study of Brain-Computer Interfaces. Together, the following chapters

1. Extend existing unidirectional models for BCI to a bidirectional modality.
2. Define a physical paradigm for BCIs, which does not depend on machine learning to measure the user's cognitive state.
3. Distribute software that enables an introspectively oriented methodology needed in the physical paradigm.
4. Identify methods for influencing the state of the brain by altering the predictability of information coming from the computer.

Thank you to my advisor (Rob Jacob), who prioritized the human side of human-computer interaction.

Thank you to my parents (Rebecca Hincks and Mats Carlsson) and grandparents (Birgit & Karl-Erik Carlsson and John & Marcia Hincks). They opened every possibility for me. Thank you to my brother (Edward Hincks) and sister (Linnea Hincks), as well as my aunts, uncles and cousins for their love and support.

Thank you to my friends and colleagues (Daniel Afergan, Matt Russell, Birger Moëll, Tomoki Shibata, Evan Peck, Beste Yuksel, Erin Solovey, Tom Williams, Ronna ten Brink, Eun Youb Lee, and Maya DeBellis) with whom I built technology and ideas.

Thank you to my friends Naomi Hashimoto, Michael Gabour, Sebastian Wickström, and Justyna Garczarek for their roles in shaping the music massage project.

Thank you to the professors who lay the bedrock for my ideas (Stuart Hirshfield, Richard Werner, Daniel Dennett, Remco Chang, Erika Hussey, Tad Brunye, Sergio Fantini, Angelo Sassaroli, and Gina Kuperberg).

Thank you to Leanne Hirshfield, who introduced me to machine learning and fNIRS in 2009, and kept me working on brain-computer interfaces ten years later.

Thank you to the Tufts Sauna Club, where I felt warmth and community.

Thank you to the Octopus.

Thank you to the Tufts Computer Science department (especially the professors Noah Mendelsohn, Mark Sheldon, Lenore Cowen, Ben Hescott, and Anselm Blumer and administrative staff Jeannine Hunter, Donna Cirelli, and Sarah Richmond).

Thank you to NSF, CABCS, Google, and Tufts University for financial support.

Contents

List of Figures

List of Tables

Philosophical foreword

A miracle has unfolded on Earth. A tribe of biological machines has woken up at the tail of a long and possibly mechanical sequence of events. Its current generation is unified by a code that has iteratively adjusted for the duration of the universe. Each scale of the code — particle, atomic, chemical, biological, neural, cognitive, and beyond — consists of objects which interact with each other and in so doing embody physical procedures for ordering chaos. The stable configurations at the micro-scales sets the rules for the possible objects at the larger scales, whose interaction can paint complex realities, like compounds, cells, neural networks, languages and society. Although astonishing, the complexity at these higher order scales is not miraculous.

The miracle is that we are able to behold and be astonished by this complexity. Mysteriously, computation defined in human DNA and instantiated in the brain leaves an associated tingling mental sensation that registers subjectively, spawning or feeding some host that controls itself, seemingly of its own free volition. The spirit of the energy may have existed eternally, or else it may be a byproduct of evolution [37], condemning its successive generations to feel the pain of failing to outsmart the other forces of nature.

Although the origin of *consciousness* remains a mystery, select instances of human DNA understand the essence of how biology encodes *intelligence*, and can prove this understanding by reproducing aspects of thinking on machines. This lineage of intelligent thinkers originates with Alan Turing and has shaped the contents of this dissertation. Both cognitive and computer science begin with a conception of intelligence as a productive mapping between inputs and outputs, defined by code realized in physical space. Conceptually, little is required to achieve simple intelligence. *Turing Machines* depend on no more than a memory tape that is divided into cells which a symbol manipulator can read, write, and navigate conditionally depending on a cell's currently inspected content. Turing first demonstrated mathematically that his simple model could embody any procedure expressed as algorithmic symbol manipulation. Pressured by the second world war, Turing later implemented his abstractions in physical space as one of the world's first computers, and programmed it with algorithms that deciphered encrypted German signals.

Turing's computer simulated the thinking operations performed by contemporary human code-crackers, doing their jobs more efficiently. Before he died, Turing postulated that computers could be programmed to perform the entire spectrum of intelligences enacted by humans [129]. As he predicted, progress in A.I. advanced steadily in the 20th century, maturing into algorithms that replicate the most fundamental feat of the human brain: to learn from experience. Like the cognitive structures implemented in the brain, machine learning algorithms use data to identify efficient mappings between inputs and outputs. The algorithms improve over time given more data, and are often generalizable so that the same algorithm can solve many different problems. For example, a neural network designed to classify buy-sell orders from financial market data need only be tweaked slightly in order to classify mental states from streaming neuroimaging data.

Contemporary human civilization thus possesses a generic algorithm that can be repurposed to solve problems for which there exists a corpus of data describing the correct output from a set of inputs. When the algorithm outperforms its human competitor, it can be installed at scale, and replace entire sectors of the economy. For example, A.I. has learned how to operate vehicles autonomously by observing human drivers steer and brake conditionally based on video, gps, and other input data. As the century continues, it is possible that computers will defeat human intelligence in every game, until Earth's biological citizens face a challenging dilemma: accept cognitive inferiority to the machine, or augment cognition in some sort of brain-computer interface.

If humans are to enjoy their next epoch, the A.I. and its designers must consider the delicate nuance of the human condition. Although humans may want to stop performing certain types of boring computation, it is not altogether in the interest of mankind to relinquish computational duties over to the machine – for a miracle has unfolded on Earth. Humans generate conscious experiences when they compute. If the forces of the planet cease encouraging the computation of its human agents, then consciousness in those agents will withdraw, no longer connected to present moment experience. Humans may thus be relieved of their burdens, but risk regressing into the mind wandering of their brain's *endogenous default mode*, a detached home for consciousness.

This dissertation aims to identify methods for describing consciousness to machines, as well as methods to alter the quality and richness of mental states for humans. I argue that the primary dimension of consciousness indexes the current level of entropy in multiple independent networks who compete for a finite energy supply. When humans engage their external environment, consciousness fixates in *exogenous task-positive networks* that process information from the senses and execute intelligent motor responses as speech and movement. If these networks support productive representations for the current state of the environment, the brain can execute mappings between perception and action with low energetic cost. In this case, control of the energy supply is granted to *endogenous resting*, or *default mode network*, which generates a stream of thought that is not relevant to the immediate processing demands of the environment. Activity in these two networks is therefore anti-correlated in neuroimaging data. Attention has been estimated to be evenly distributed between these networks, but individuals generally report a strong preference for activity in their *exogenous task-positive networks* [77].

This dissertation argues that functional near-infrared spectroscopy (fNIRS) can measure the reciprocal back-and-forth activation between exogenous and endogenous networks, enabling a brain-computer interface (BCI) that adapts the stream of information to the brain in a bidirectional feedback loop with this measurement. Such a bidirectional BCI can dynamically adjust system variables that dictate the cognitive burden the exchange of information poses on the user's *exogenous task-positive networks*, minimizing undesirable *endogenous default mode network* activity in a population whose computation is no longer enlisted.

As illustrated in figure 1, the seven chapters of this dissertation unfold a set of ideas towards how to build such a bidirectional bci.

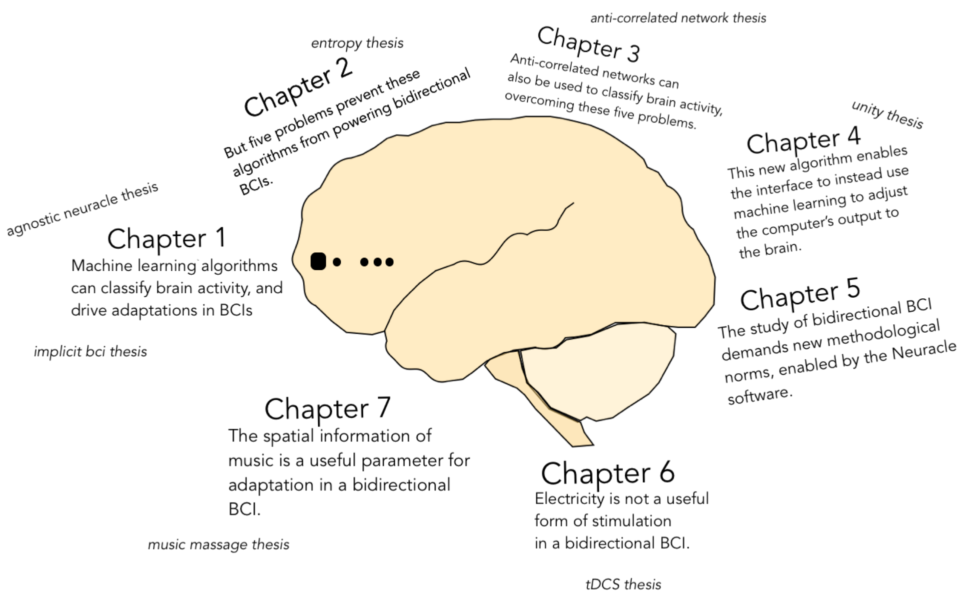

Figure 1: The Chapters of this Dissertation

Chapter 0

Introduction

0.1 Implicit BCI in an **Agnostic Paradigm**

The brain-computer interface research in this dissertation originates from a multidisciplinary field at the intersection of computer and psychological sciences. Research in human-computer interaction investigates channels of input and output between the human and computer, as encoded by a program's user interface. In this context, a brain sensor can be regarded as an additional channel of input to the computer, supporting a brain-augmented user interface. BCIs can be partitioned into different categories depending on how the data is used. In an *explicit* BCI, measurements of brain activity *replace* intentional inputs that the user may otherwise give through a mouse or keyboard. A user may imagine a command (e.g. *yes*) which has a measurably distinct neural profile from other commands (e.g. *no*). Explicit BCIs can have life altering impact for disabled users whose brains can no longer command the control of their bodies.

In an *implicit* BCI, measurements of brain activity *supplement* intentional input, expanding the bandwidth of communication between human and computer through an auxiliary channel that indexes the user's cognitive state. A well-designed *implicit BCI* leaves the user unaware of how their brain state affects system settings as it updates system settings to the user's advantage while they focus on the task at hand. Chapter One summarizes *implicit BCIs* which use functional near-infrared spectroscopy (fNIRS) to measure brain activity, as pioneered by my advisor (Robert Jacob) and the five dissertations that precede my own contributions. These BCIs illustrate a common pattern for how to build *implicit BCI* using fNIRS, which I will refer to as the agnostic paradigm for BCI. Experiments evaluating BCIs in the agnostic paradigm begin with a calibration period that changes the user's task demands in order to provide labels for data in a controlled setting. The collected data is fed into a machine learning algorithm that finds patterns separating the conditions of the calibration period on the basis of brain activity. When trained, the machine learning algorithm classifies ongoing brain activity in realtime, driving adaptations to settings in the user interface.

All BCIs depend on this algorithm, which processes as input measurements from brain sensors and returns as output a classification of the user's mental state in real-time for the user interface. In this dissertation, this algorithm will be referred to as a neuracle. A neuracle is an algorithm which summarizes the state of an information processing system such as the brain along a set of dimensions. The etymology of neuracle acknowledges the mythical *oracle*, an ancient Greek who predicted human fate. Like an oracle, a neuracle might predict human experience before it occurs by extrapolating rhythmic oscillations in neural activity.

The suffix of neuracle also recognizes the neuroanatomy of an octopus whose tentacles continue to live even if severed from its siblings. The octopus brain holds crucial information for how humans can adapt in an age of machines, and metaphors for how scientific methodology can adapt if it is to move beyond its current fixation on physical structures without reference to associated mental events. Scientific maturation is especially important for the present investigation into technology that measures and alters conscious states. Although it is beyond the scope of this dissertation to specify the methodology needed to effectively develop and test bidirectional BCIs, Chapter Five describes a software program called Neuracle that enables a more introspectively-oriented single subject (n=*1*) BCI methodology. Experiments conducted with Neuracle resonate with participants and experimenters, but do not conform to typical scientific standards (as advocated by the disciplines of psychology). They maximize the information about the relationship between the current configuration of inputs to the brain and concurrent mental states only for those who engage the Neuracle software at the expense of exporting a statistically significant set of quantitative metrics for a larger scientific audience. The chapters of this dissertation are organized to gradually unfold ideas towards this new methodology, leaving a proper specification of what would constitute a scientific result in this new paradigm to future work.

The first chapter describes the state-of-the-art algorithm for building neuracles that drive statistically significant improvements to user performance in a large and random population of users. This agnostic paradigm for BCI deploys agnostic neuracles, which are algorithms that do not depend on an understanding of the brain or sensor, and instead use machine learning to relate sensor data to a behavior set in place in the experiment's calibration period. The quality of the neuracle can be analyzed prior to the main phase of the experiment. Offline evaluation is critical for ascertaining that the neuracle can predict the user's mental state on the basis of physiological data, and iterate upon parameters in the neuracle (such as the filtering, feature set, and machine learning algorithm) when it performs poorly. As described in Chapter One, this evaluation is typically conducted by training the neuracle on all-but-one instance in the calibration period which supplies a valid test case, and repeating this analysis so that every trial is left out exactly once. The fact that these predictions are better-than-chance constitutes the first thesis of this dissertation, which is referred to as the agnostic neuracle thesis:

Thesis 0.1.1 *Agnostic neuracles can classify the user's current task on the basis of fNIRS data at a rate significantly higher than chance.*

The second and third sections of Chapter One describe experiments which use agnostic neuracles optimized for offline classification to drive real-time implicit BCIs, some of which successfully improved measures of user performance, and others which failed to produce statistically significant positive effects. In one of the successful experiments, the agnostic neuracle powered a dynamic difficulty adjustment engine which simplified task demands when the user was determined to be in a high cognitive workload state [6]. Acting as an air traffic coordinator, the user controlled the flight paths for a set of unmanned aerial vehicles (UAVs), while the dynamic difficulty adjustment engine was active or bypassed. When active, the difficulty adjustment engine removed simulated UAVs when the agnostic neuracle predicted that the user was in a high cognitive workload state, and added more work when the user seemed to have cognitive resources to spare. Overall, users made significantly fewer errors when the interface partitioned their workload based on brain activity. Chapter One defends the capability to build and evaluate fNIRS-based BCIs as another thesis statement of the dissertation, which is referred to as the *implicit BCI thesis*:

Thesis 0.1.2 *fNIRS-based implicit BCIs following an* agnostic paradigm *can improve user performance in controlled laboratory conditions.*

0.2 Bidirectional BCIs in a Physical Paradigm

The subsequent chapters move beyond existing research, and are geared towards establishing a physical paradigm for building brain-computer interfaces. BCIs in the physical paradigm benefit from a cohesive model of the interactions between the brain and computer at four levels. The model accounts for

- how the brain works at the *physical level*
- the linkup between brain state and mental state at a *mental level*
- the relationship between brain state and sensor data at the neuracle level
- how computer settings and output affect the physical state of the brain at an *interface level*.

A unified model at these four levels enables a new genre of BCI introduced as part of this dissertation. Like an implicit BCI, a bidirectional BCI depends on ongoing classifications about the user's mental state. However, the bidirectional BCI does not attempt to adapt itself to better serve the mental state that has unfolded.

Instead, it adjusts properties in a set of output channels to the brain, known to modulate its state. If the bidirectional BCI has a cohesive model for the user at the *physical, mental, neuracle,* and *interface levels*, then measurements of the user's mental state and brain stimulation channels can be linked together in a feedback loop where the system learns a protocol for sustaining some desirable state in the user over time. All algorithms, software, ideas and data beyond Chapter Two are concerned with the problem of how to build bidirectional BCIs in a physical paradigm, which can be broken into two separate research problems.

1. **Neuracle Question** What are the components of an algorithm that takes as input measurements from a neuroimaging device and returns as output a classification of their mental state?

2. **Stimulation Question** What properties of output channels between the computer and brain can be adjusted to induce desirable changes in the user's mental state?

The chapters of this dissertation investigate these two questions in isolation from each other, but develop a cohesive thread that combines the problems of measurement and stimulation. This thread pushes BCI research towards a unified framework for measuring and manipulating a set of neural dimensions in the user, encoded in an information processing vocabulary. In order to support a bidirectional BCI in a physical paradigm, these dimensions must satisfy criteria at the four levels of analysis described above.

- At a *physical level*, an appropriate *physical dimension* must successfully delineate two sets of physical configurations for its two extremes.

- At a *mental level*, the dimension must offer a vocabulary that BCI engineers can map onto a pair of mental states that feel in opposition to one another.

- At the neuracle level, changes in the dimension at the physical level must be quantified on the basis of a brain sensor and rendered in real-time to a user interface.

- At the *interface level*, the user interface must determine how to adjust stimulation to the brain in order to bring about more desirable settings to the dimension.

Motivated by the literature and experiments of Chapters Two and Three, Chapter Four gives a set of *physical dimensions* that satisfy the criteria at these four levels as well as an algorithm for measuring them using fNIRS.

0.2.1 The Bayesian Framework for the **Physical Paradigm** of BCI

At a *physical level*, the brain consists of a variety of competitive and collaborative networks that are layered in a hierarchy. According to the *Bayesian brain hypothesis*, the basic goal of any network in this hierarchy is to efficiently predict the causes of its inputs and to leak only the errors of its prediction as output to its hierarchically superordinate networks [50]. Together, these networks reorganize perception and action towards a stable representation of reality and their identity within it.

The imperative to stabilize reality and immortalize identity applies to functional operations at particle, atomic, neuronal, network, inter-network, and phenomenological levels. Each unit supports its own notion of reality and identity, including the unit that emerges when all parts interact. At this global level, identity resides principally in the hierarchically superordinate *endogenous* networks, whereas reality is a construct computed by *exogenous* networks more immediately tethered to inputs from the body, the sense organs, and the other sources of information in the universe [50].

For the *exogenous, sensory* networks, the mapping between inputs and outputs begins with primitive structures that transduce physical signals originating from the environment (light, sound, and other energy) into a corresponding internal representation. These networks contain all the raw information that the body's sense organs (eyes, ears, etc.) are able to physically register, but they successively forfeit the full resolution of the signal as they transmit data up to superordinate, more *endogenous* networks [83]. Low level networks compress the sensory data to satisfy the uncertainty of superordinate networks, leaving only the essence of the representation for conscious inspection: a *sensorium* that is shaped both by new data and the brain's prior expectation.

In the *Bayesian brain*, computation (or energy consumption) must concentrate in networks that resolve the meaning and implication of informative signals. In order to maximize computation in the *exogenously* oriented brain, a signal should contain enough familiar content to engage vast and disparate networks, but enough new information so that these networks must adjust internal representations in order to effectively predict the causes of its input in the future. In this dissertation and in related work [27], the physical dimension quantifying this information theoretical spectrum is referred to as *entropy*. A high entropy network is surprised by its inputs, and undergoes many changes in its underlying neural state space, whereas a low entropy network is already at a stable equilibrium with its inputs, leaving more leftover energy for other networks.

As a dimension for measurement and manipulation in a bidirectional BCI, *neural entropy* satisfies the principal criteria delineated above at *physical, mental, neuracle*, and *interface levels*. Stated as the *unity thesis* of this dissertation,

Thesis 0.2.1 *The spectrum of states in between the brain's stable and novel configurations have differentiable physical neural signatures that are mirrored in an introspectively observable mental workspace, which can be monitored by* neuracles *classifying physiological changes associated with the activation of neural networks, and which can be controlled by adjusting the amount of information transmitted to the brain in a user interface.*

The *unity thesis* binds the seven chapters of this dissertation, establishing a context for building bidirectional BCIs in a physical paradigm. In the chapter organization of this dissertation and its underlying history as a sequence of intuitions morphing over time, this new context has developed to address key shortcomings in the existing agnostic paradigm for BCI.

Chapter Two highlights shortcomings with the agnostic approach, which must be addressed if BCIs are to function outside of controlled, laboratory settings. This chapter illustrates how the dimension of *entropy* exposes a problem for the state-of-the-art agnostic methodology for constructing neuracles. If the signature token of a salient mental state is its novelty, then there should not exist some static experimental procedure for inducing differentiable mental states, since each time segment of measured neural data should have a unique signature for each user and under different task inductions. Because the theory of *neural entropy* simultaneously explains limitations with the agnostic paradigm and opportunities for the physical paradigm, Chapter Two creates a bridge between existing research and the novel ideas of this dissertation, which is referred to as the *entropy thesis.*

Thesis 0.2.2 *Neural entropy poses a problem for agnostic neuracles and an opportunity for physical neuracles as a property to be detected in the user.*

0.2.2 Anti-correlated networks as a basis for Physical Neuracles

Chapter Three provides an empirical foundation for measuring *neural entropy* using *physical neuracles* in fNIRS data. This theory is inspired by the neuroscientific discovery of anti-correlated networks in fMRI neuroimaging data. This literature suggests that the *endogenous, identity* networks and *exogenous, reality* networks are separated in space and compete for a finite energy supply [108].

The back-and-forth mental activity between the endogenous and exogenous brain can also be confirmed at a *mental level* through conscious introspection. Consider your phenomenology while driving a car. While operating a motor vehicle, the problem of executing a series of adjustments to the steering, brakes, and speed of a car in response to changing road and traffic information is solved by the *exogenous, task-positive brain.* At times, such as during periods of heavy city traffic, consciousness fixates on the immediate task of driving the car. However, when you enter an empty country road, the demand for novel computation in the *exogenous* brain diminishes. In this case, energy is liberated for the other computer which solves problems that do not pertain to the immediate task and sensory experience. This *endogenous, default-mode* network registers mentally as a sequence of memories, worries, and fantasies — or mind wandering.

This basic neuroanatomical feature of two independent computers in the brain exposes a hack for measuring the *exogenous neural entropy* that is the result of the computational demand of the task or environment. If the BCI independently interrogates the activity of both computers, then it has two points of truth for a single description of the user's mental state as a phenomenon that is somewhere on a spectrum between detached mind wandering and sensory-motor immersion. To this end, Chapter Three documents a search for negatively correlated (anti-correlated) networks in fNIRS data, where the activation of one network implies a simultaneous deactivation in some other spatially distinct network. This chapter shows the outcome of a 50-subject 4-session experiment identifying the strongest anti-correlated network and how it shifts depending on the current task. It is written to resolve a more specific formulation of the neuracle question given above.

1. What is the relationship between the low frequency oscillations present in fNIRS time-series and the mental states visited by the brain generating those oscillations?

An abstract answer to this question is given by the *anti-correlated network thesis.*

Thesis 0.2.3 *The strongest anti-correlated network in fNIRS data, in between 0.01 to 0.1 hertz, describes user attention as it shifts between a more sensory (or exogenous) mode to a more conceptual (endogenous) mode, and can be minimally measured by one fNIRS probe by the eyebrow and one by the ear. Properties of this pair of signals describe the relative neural entropy between external and internal sources of information.*

Chapter Four gives an algorithm for a *physical neuracle* that extracts thirteen statistical properties from two anti-correlated networks and hypotheses for how these correlate with different mental states. Algorithm 1 shows a preview of this *physical neuracle*, whose logic depends on a theoretical and empirical foundation developed in Chapters Two and Three. The algorithm identifies the strongest anti-correlated network in fNIRS data, partitions it into cycles, extracts statistical features on the current cycle, and uses domain expertise to determine mappings between these dimensions and imperatives for how to adapt an interface (including when to *simplify*, *continue*, and *interrupt*). Any application can listen to changes in these design imperatives, providing a convenient abstraction for user interfaces to upgrade into brain-computer interfaces.

Algorithm 1 Physical Neuracle Preview

```
procedure PhysicalNeuracle(fnirs)
    networks ← findPairsOfSignalsWithHighAntiCorrelation(fnirs)
    cycles ← partitionIntoSegments(networks)
    dimensions ← extractTemporalFeatures(cycles)
    imperatives ← getAdaptiveDesignTemplate(dimensions)
    return imperatives
```

By Chapter Five, the dissertation will have:

1. Demonstrated that agnostic neuracles can power full implementations of implicit BCIs in a laboratory context as the agnostic neuracle and *implicit BCI theses* (*Chapter One*).

2. Demonstrated the shortcomings of agnostic neuracles in realistic user settings as the *entropy thesis* (*Chapter Two*).

3. Demonstrated the feasibility to build *physical neuracles* on anti-correlated networks as the *anti-correlated network thesis* (*Chapter Three*).

4. Given an algorithmic specification that connects *physical, mental, neuracle,* and *interface* levels of BCI in a Bayesian framework as the *unity thesis* (*Chapter Four*).

The fifth chapter describes the Neuracle user interface software which includes the necessary commands, preprocessing, and data manipulation techniques for building *physical* and agnostic neuracles and letting them drive adaptations in

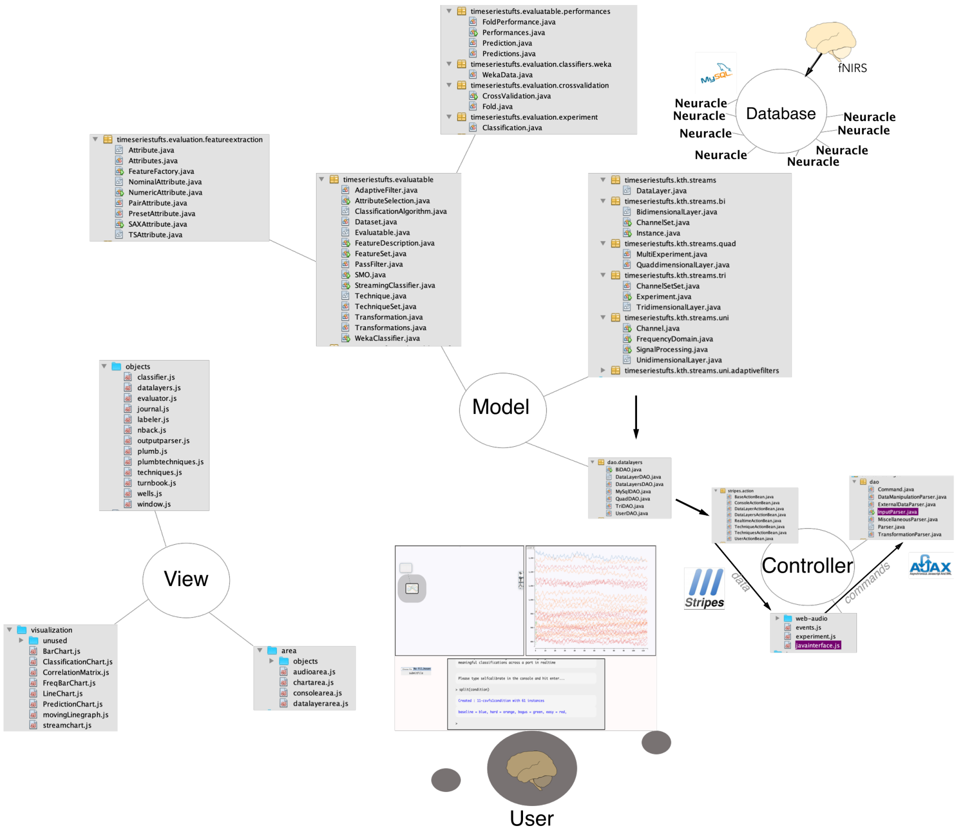

Figure 1: The Neuracle Source Code

experiments. The interface is optimized for rapidly generating experiments and inspecting their conclusions through visualizations and the classification performance of neuracles. It is intended to inspire novel methodology and research contexts for developing BCIs in a physical paradigm that emphasizes n = *1, 2, or 3* person studies, where each participant is a skilled meditator or introspector of their mental activity as well as synchronized on a common vocabulary for describing the observable dimensions of their mental activity. As shown in Figure 1, the source code obeys the *model-view-controller* architectural pattern (see *www.github.com/samhincks/neuracle* for the complete implementation).

0.3 Bidirectional BCIs at the Interface Level

The empirical findings in Chapters Two and Three as well, as the introspective findings from Chapter Five, provide evidence that the *physical neuracle* given in

Chapter Four measures the orientation of user attention. The constraints of the *physical neuracle* give a more specific formulation of the *stimulation question*:

1. Which channels of output from the computer to the brain can be dynamically adjusted based on real-time knowledge about the ongoing information exchange between the brain's exogenous and endogenous processes?

Chapters Six and Seven evaluate parameters in output channels to the brain which are central for their practical inclusion in a bidirectional BCI. Chapter Six investigates the potential to affect the state of the brain by administering an electrical current to the forehead, using a century-old technology called transcranial direct-current stimulation (tDCS). The chapter shows the motivation, methods, and results in a study investigating the delay between the administration of electrical current to the brain through tDCS, and concomitant effects on participant performance in a cognitive workload task. The results indicate a long delay between stimulation and cognitive effect, implying that binding stimulation parameters to mental state classifications from fNIRS would be unlikely to augment human performance in a bidirectional BCI. This negative result is referred to as the *tDCS thesis*.

Thesis 0.3.1 *The delay between the administration of tDCS and measurable changes in user performance exceeds the short timespan between stimulation and effect needed to establish a feedback loop in a bidirectional BCI.*

Chapter Seven continues with an audio alternative to electrical stimulation. In a *Bayesian framework*, music may be regarded as sound riddles for the brain. Music occupies a delicate sweetspot between predictability and novelty, giving enough familiarity to engage top-down circuitry in the brain but containing enough information to push error up the cognitive hierarchy, and engage the brain's deeper endogenous circuitry (i.e. emotions). Before language evolved, sound functioned primarily as a vehicle to transmit spatial information to the brain. Potentially, the audio sequences that register in the brain as music do so because they correspond to valid movements in time and space. In that case, these sequences would exert a greater control over the brain if the sounds moved correspondingly. If the motion and the sound match, the aesthetics would be preserved, avoiding the nausea inherent to violating user expectations.

Chapter Seven describes a pilot study evaluating the effectiveness of adjusting the spatial information of music in a bidirectional BCI. In this pilot experiment, attendees lay in a hammock in the woods, having specified a desired emotion among a menu of colors. The experimenters (or music masseurs) then adjusted audio, olfactory, and tactile sensation to the participants depending on the shared musical experience, desired color emotion, and more empathic/elusive feedback channels transmitted over the course of the *music massage.* The testimony of the participants underscores the power of spatial music, motivating a deeper and more rigorous scientific study. The potential to adapt the 3D orientation and motion of sound in a bidirectional BCI is referred to as the *music massage thesis*:

Thesis 0.3.2 *The spatial information of music is a useful channel for adaptation in a bidirectional BCI.*

Chapter 1

Implicit Brain-Computer Interfaces in the Agnostic Paradigm

Successful implicit BCIs expand the bandwidth of communication between the user and computer without demanding additional effort. By mapping classifications of the user's mental state obtained from non-invasive brain sensors onto a system's internal state, these BCIs establish an implicit channel of interaction that supplements direct input the user provides with mouse, keyboard, and other traditional controllers.

Three interdisciplinary research questions must be solved to build an implicit BCI.

1. *Dimensionalization Question:* Along what set of cognitive dimensions can the user's mental state be summarized?

2. *Portrayal Question:* How can these dimensions be extracted on the basis of non-invasive and low-cost sensors?

3. *Interaction Question:* How can these dimensions be mapped interactively onto dynamic system parameters that drive adaptations to the user's benefit?

This chapter illustrates a comprehensive solution to these questions, leading to a variety of implicit BCIs whose effectiveness has been evaluated in controlled user studies.

1. *Section 1* addresses the *dimensionalization question* from the perspective of *cognitive psychology* and *neuroscience*. It describes the prefrontal cortex and psychological construct of working memory as well as a principled method for altering it in an experimental context.

2. *Section 2* addresses the *portrayal question* from the perspective of *biomedical engineering* and *machine learning*. It describes how to interrogate brain activity using fNIRS, as well as how to classify the user's working memory on the basis of fNIRS measurements. The results of this section give recommendations for how to most effectively set parameters in an agnostic neuracle, and demonstrate the agnostic neuracle *thesis* of this dissertation. Agnostic neuracles can classify the user's current task on the basis of fNIRS data at a rate significantly higher than chance. These results are obtained using the neuracle software, which will be the focus of Chapter Five.

3. *Section 3* describes several experiments that use real-time agnostic neuracles in order to drive adaptations in two stationary implicit brain-computer interfaces. This section shows similar experiments for *wearable computers*, which illustrates the potential to use the neuracle web interface to drive real-time adaptations.

Altogether, this chapter defends the *implicit BCI thesis* of this dissertation: it is possible to build an *implicit BCI* that triggers passive adaptations in a user interface on the basis of real-time classifications about the user obtained from an agnostic neuracle. Moreover, these brain-adaptive systems measurably improve user experience compared to their non-adaptive counterparts, at least in controlled laboratory conditions.

I refer to the outlined solution as the agnostic paradigm in contrast to the physical paradigm to be described in subsequent chapters.

1.1 The Dimensionalization Problem

Effective dimensionalization is essential to BCI because without a link between cognitive states and the underlying neurobiological machinery, it would be challenging to build an algorithm that predicts the cognitive state from brain sensors. If dimensionalization is done properly, then implicit BCI has a small library of dimensions that are true enough to the basic operation of the brain to show an effect to non-invasive brain sensors, but far enough away from its bio-mathematical calculations to transmit meaning to a user interface designer.

The *dimensionalization problem* is not as important in the agnostic paradigm as it is in the physical paradigm. This chapter therefore considers the *dimensionalization problem* only from the point of view of classic neuroscience. In this literature, individual regions of the brain are understood in isolation from each other. Evidence in favor of this simplification is obtained by noticing that different regions of the brain show different levels of activity during different tasks as measured by fMRI. This understanding suffices for the purposes of building agnostic neuracles, whose machine learning algorithms may discover more precise descriptions between mental states and brain activity. But Chapter Two highlights shortcomings of agnostic neuracles, motivating the need to develop algorithms that classify mental states based on a physical understanding of the brain (so called physical neuracles). The neuroscientific thread of this dissertation will therefore be modernized in Chapters Two and Three, which will promote models of the brain that lie somewhere in between strict hardware (biological) and software (cognitive) levels. These modern and evolving models of the brain are based on functional and structural connectivity analyses of the brain, and lead to an understanding of brain activity in terms of networks.

The following subsections therefore give minimal neuroscientific specifications for understanding how to build agnostic neuracles. The remainder of this section will describe a region of the brain (the prefrontal cortex), a dimension of cognition (working memory), and a method to manipulate working memory (the n-back task to be described in the next subsection).

1.1.1 The Prefrontal Cortex

Occupying a third of the brain's overall hardware, the human prefrontal cortex size and energy consumption distinguishes it from brains of other animals. It can therefore be presumed to perform functions that are central to the implementation of intelligence, such as language, short term memory, and meta-cognition [130]. In the most general formulation, the PFC functions as an executive in the brain, integrating information from subordinate networks and delegating responses. The region is partitioned into lateral, orbital, and medial regions, which are differentially active depending on the type of information under manipulation. The medial region processes social data types, including the concept of the self; the orbital regions come online when inputs carry emotional charge; and the lateral regions process the most abstract data types. But the functional segregation in the PFC is blurry, and social, emotional, and abstract data processing overlap with each other. As a unitary system, the PFC should be conceived merely as an integrator of information.

1.1.2 Working Memory

In the process of integrating bottom-up inputs, the PFC sustains a conscious mental buffer that holds information that is immediately relevant to the present situation [13]. This *working memory* buffer enables its agent to compose thought (in language and other modalities), plan action, and direct attention.

There are competing models for how working memory operates at software and hardware levels. Baddeley's model posits an architecture whereby one domain-general central executive controls many domain-specific short-term memory buffers [13]. The underlying storage buffers can be verbal (words and sounds) or spatial (locations in an environment), or object-related (shape, color, texture). Although all housed by the PFC, these sub-systems have somewhat different underlying neural

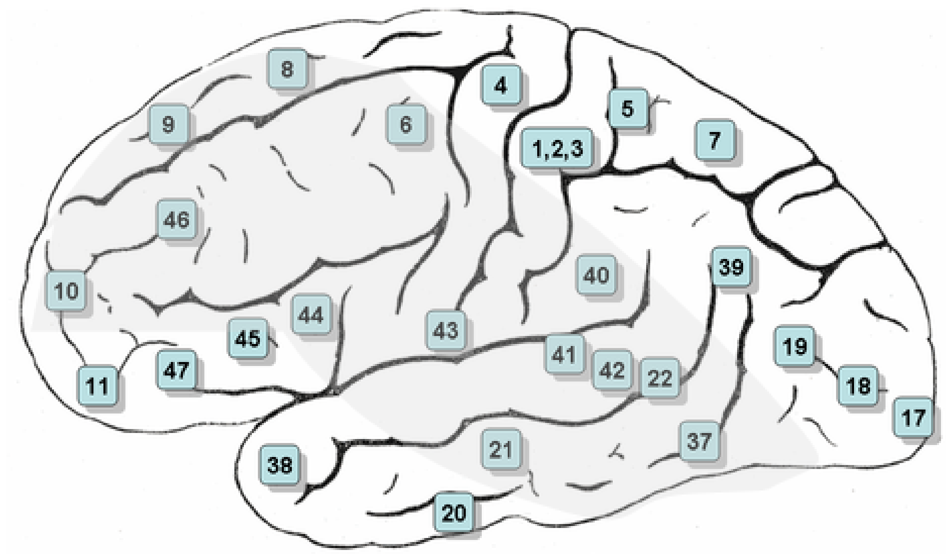

Figure 1.1: Brodmann areas

implementation. Verbal working memory (induced by an oral n-back) preferentially activates left-hemisphere speech areas (Brodmann 44). Spatial working memory activates the right premotor cortex (B.4), and object storage inhabits the ventral PFC at (B.10) [120, 101]. Current research has not been able to isolate significant neural distinctions between the task of short term information storage and broader executive processes (e.g planning, problem solving, and decision making) [98].

1.1.3 The n-back task

In cognitive science and BCI research, the n-back task is the standard method to induce variable levels of engagement from both verbal and visual subsystems of working memory [10, 11, 45, 52, 92, 97]. In a visual *n-back*, the user is presented with a succession of 3x3 grids where one of the nine boxes is filled, and the user must indicate whether the box is at the same location as it was n iterations ago (figure 1.2 illustrates a visual *3-back*). The audio *n-back* is similar except the user *listens* to a stream of numbers, indicating, upon receiving a new number, the number presented *n* iterations ago. Thus, a *0-back* poses no strain on the user's auditory working memory, as the brain simply repeats the numbers it hears.

The *n-back* is useful for the purpose of building agnostic neuracles since it offers a principled method to manipulate the burden on a user's working memory. Experiments in the agnostic paradigm typically begin with a series of *n-back* trials of variable difficulty, which gives the agnostic neuracle the opportunity to witness the brain under different conditions. A given n-back trial lasts between fifteen and thirty seconds, long enough to induce a neurological response for measurement by fNIRS. There are typically between seven and fifteen instances of two difficulty levels of the n-back task, sufficiently many for the machine learning algorithm to extract a pattern separating the brain under low and high levels of cognitive workload.

To foreshadow a discussion that will be continued in the next chapter, it is worth reflecting on the relationship between the two dimensions: working memory and cognitive workload. For now, high working memory should be understood as always entailing high cognitive workload. But high cognitive workload need not always imply high working memory. For example, it is possible to enter high cognitive workload writing states without excessively belaboring working memory: the words sometimes just flow from the fingers.

The neuracle software package includes both audio and visual n-backs along with mechanisms for labeling the data by the difficulty (n) and accuracy, as elaborated in Chapter Five. To initiate an experiment with the oral n-back according to a specified protocol, type *streamlabel(easy%hard, seconds%trialsOfEach%secondsOfRest)* in the console. This will broadcast the labels *easy, hard*, and *rest* to the data object that is currently accumulating fNIRS data, and trigger a voice that repeats numbers for the specified duration.

1.2 Portrayal using Agnostic Neuracles

The previous section outlined the basic function of the prefrontal cortex, the concept of working memory, and a protocol for manipulating its level of workload using

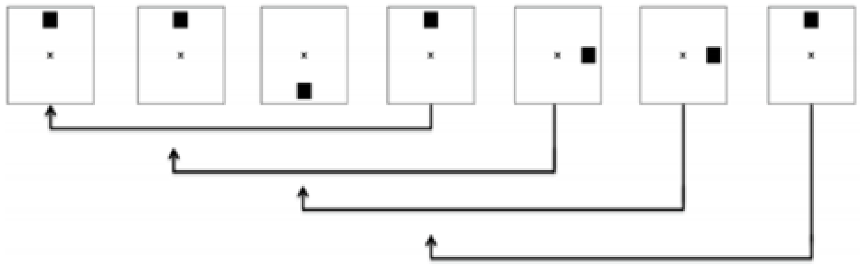

Figure 1.2: A visual n-back

the *n-back*. This overview was intended to provide preliminary answers to the *dimensionalization problem* that will be elaborated in Chapters Two and Three. The knowledge cited so far should be sufficient for establishing the credibility of the agnostic neuracle given below. In order to adequately solve the *portrayal problem*, this neuracle must solve three sub-problems (one hardware and two software).

1. **Biomedical Engineering Problem** Capture physical events associated with the brain's computation of cognitive states with scientific instruments.

2. **Signal Processing Problem** Transform physical measurements over time into a series of <class, feature> pairings.

3. **Machine Learning Problem** Identify patterns in <class, feature> pairings using machine learning algorithms.

1.2.1 Measuring the brain with fNIRS

Most neuroimaging devices are too intrusive, costly, or otherwise inconvenient for standard Human-Computer Interaction (HCI). However, fNIRS can be used in typical computer settings as it is resistant to motion artifacts [122, 88], and there are no hard constraints that limit its future integration in consumer grade electronics [44]. In fact, several labs are in the process of developing a portable fNIRS [117, 43, 111]. Compared to EEG, the more popular device for BCI, fNIRS has poor temporal resolution, meaning that there may be a lag time (3-6 seconds) between values it can detect and the physiological event it endeavors to represent. On the other hand,

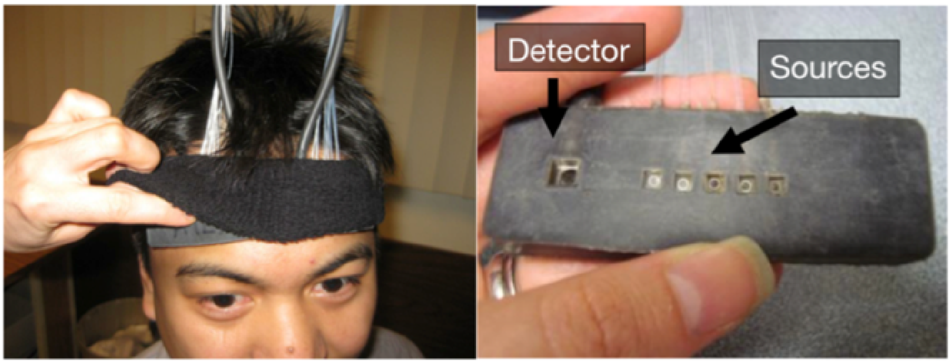

Figure 1.3: Functional near-infrared spectroscopy

unlike EEG, fNIRS has excellent spatial resolution, meaning that it is possible to resolve what space in the brain it measures.

As shown in figure 1.3, fNIRS consists of a set of light sources and light detectors: parts that can be made cheaply and miniaturized into hidden components of a hat or wearable [44]. To measure brain activity, near-infrared light at two wavelengths (typically 690 nm and 830 nm) is transmitted through the skin and bone before it is absorbed and scattered by the blood's hemoglobin, the supply system of energy for neurons in the brain. Because light interacts differently with oxygenated and deoxygenated hemoglobin, it is possible to infer the underlying neuronal activation from the quantity and timing of light returning to a sensor, approximately three centimeters away [124]. The detector thus outputs ongoing measurements about the light intensity at each of the two wavelengths, which software converts into measurements of oxygenated and deoxygenated hemoglobin according to the Beer-Lambert Law [36].

This gives a barometer of underlying activation in the brain according to the blood oxygen level dependent (or BOLD) response. When networks of neurons activate and communicate, they consume the oxygen that is present in order to metabolize glucose, causing an increase in blood flow and the need to replenish their current supply (leading to an increase in oxygenated hemoglobin). The method for mapping fNIRS data onto descriptions of neural architecture is similar to the more widely used functional magnetic resonance imaging (fMRI), which also infers under-

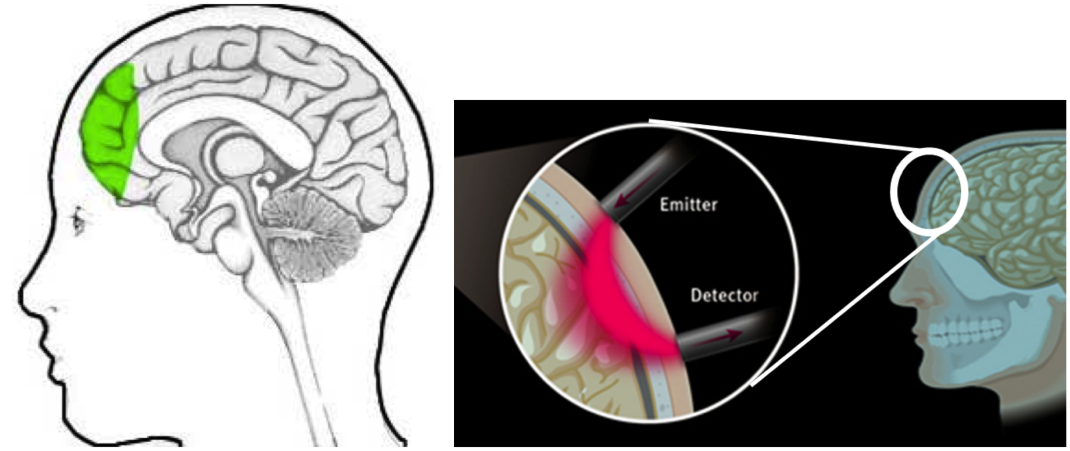

Figure 1.4: Left Image: The Prefrontal Cortex. Right image: Near-infrared light as it takes an arc-shaped path 3cm into the brain and back to the detector.

lying activation with the BOLD response. FMRI enables a larger three-dimensional resolution of the brain, but its price-tag, size, and loud magnetism prevents its possible adoption in non-invasive BCI contexts.

The experiments in this chapter use the probe placement in Figure 1.3. This probe configuration targets the brain's anterior PFC in symmetrical locations at the left and right hemisphere. The anterior PFC deals with high level executive processing, which corresponds to storing and retrieving objects in memory, planning, and paying deliberate attention [35, 75, 91]. Chapters Two and three measure more locations in the brain.

1.2.2 Agnostic Neuracles using Machine Learning

The calibration, preprocessing, feature extraction, machine learning, and real-time phases of the agnostic paradigm are illustrated in Figure 1.5. In the **calibration phase**, the user's working memory state is manipulated by a series of n-back trials with variable levels of difficulty. As described above, this is accomplished with the *streamlabel(easy%hard, seconds%trialsOfEach%secondsOfRest)* command in neuracle.

When calibration is complete, the experimenter begins the **preprocessing phase** by selecting the data object in neuracle, and typing *manipulate(filterchoice)*,

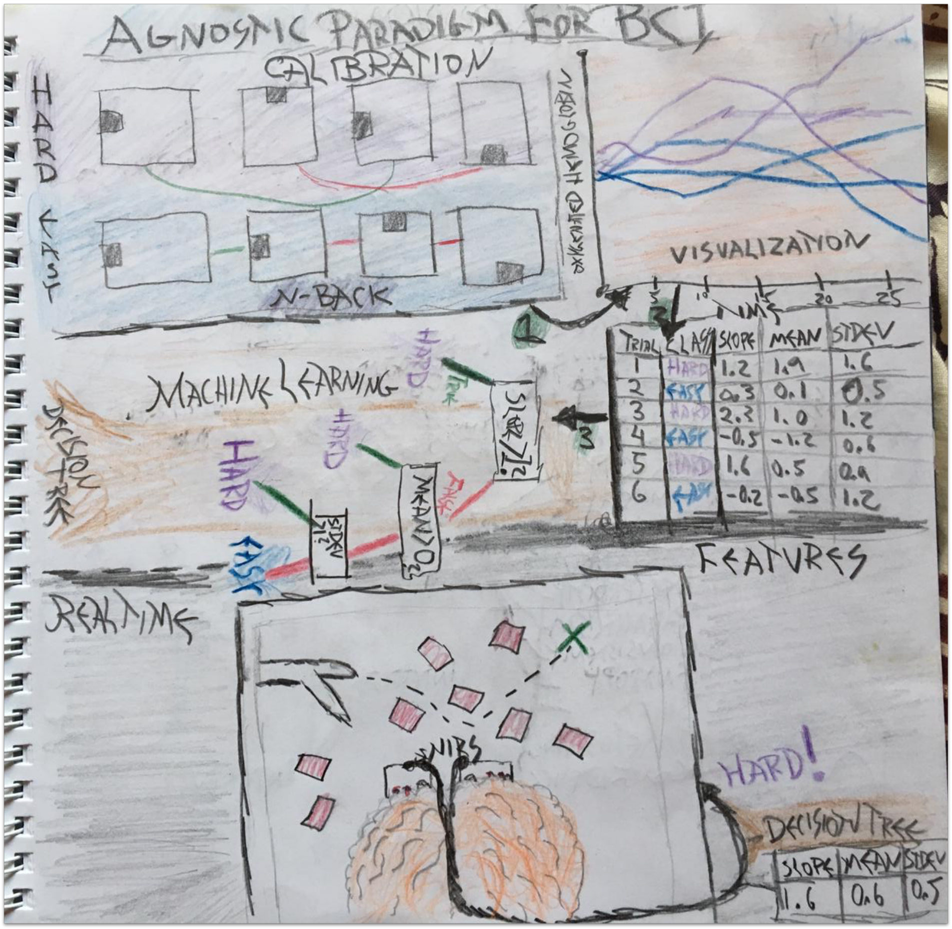

Figure 1.5: Training an Agnostic neuracle.

where common filter choices include bandpass filters, moving average filters, and adaptive filters. Next, the experimenter changes the internal representation of the data from a 2-dimensional series of channels into a 3-dimensional collection of channel-sets, grouped into trials with a common class by typing *split(conditionName)*.

In the **feature extraction** and **machine learning phase**, the experimenter prepares a feature-set with the *makefs(statistic,channel,window)* command, and a machine learning algorithm with the *makeml(algorithmName)* command. As illustrated in 1.6, the experimenter drags the dataset so that it intersects the *machine learning* and *feature objects*, and types *evaluate* to see how effectively the selected techniques fared in cross-fold validation at classifying the level (n) of *n-back* difficulty given the fNIRS data. When the experimenter is satisfied with this classification accuracy, they type *train*, which trains the intersected machine learning algorithm

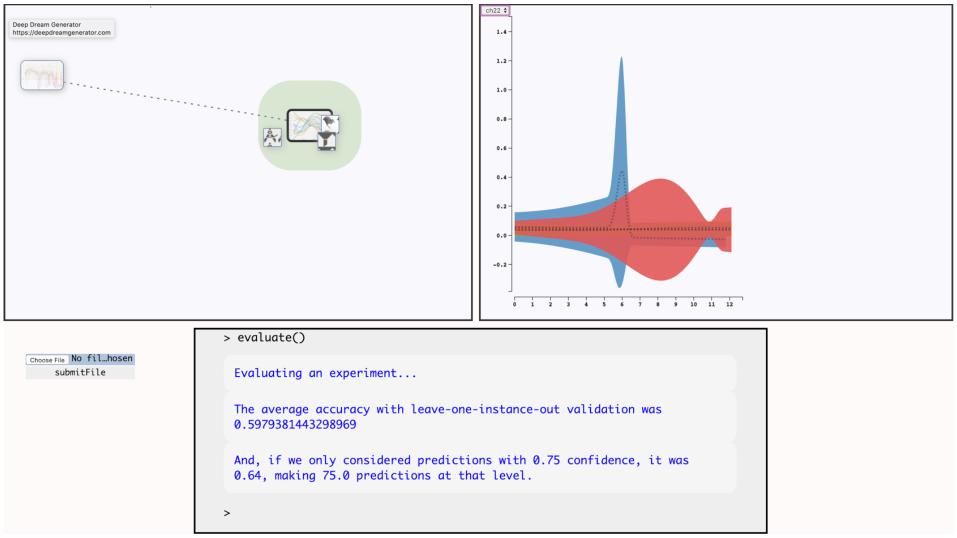

Figure 1.6: The Neuracle User Interface

on all the training data.

Next, in the **real-time phase**, the experimenter types *classifyLast(port)*, so that new trials are constructed every second on the last *trialLength* seconds of the real-time data, and fed into the trained machine learning algorithm. The classifications occur each second and are broadcasted to applications listening on the specified *port* number.

1.2.3 Evaluating Agnostic Neuracles

This section uses neuracle to evaluate basic settings in the agnostic neuracle, answering the following questions by manipulating one property at a time and holding all others constant.

1. **Channel:** In a given window of fNIRS data over sixteen channels, which channel portrays the most information about the user's mental state?

2. **Time:** In a given window of fNIRS data, which segment of time portrayed the most information about the user's mental state?

3. **Features:** In a given window of fNIRS data, which statistical features portrays the most information about the user's state?

4. **Machine Learning:** In a given set of <class-feature> pairings, which machine learning algorithm can most effectively classify the user's cognitive state?

The following offline analysis is done over a total of 46 sessions with 39 subjects collected over four experiments, and demonstrates the agnostic neuracle thesis. For the purpose of evaluating the cleanest possible data, I focused on half of the dataset, using the half with the highest average classification accuracy (according to eight machine learning algorithms operating over all evaluated features, channels, and time segments).

All experiments tested cognitive workload, and consist of a set of easy and hard trials in equal quantity (see Table 1.1). The first three experiments were visuospatial n-back experiments, while the fourth was a cognitive manual task where users processed visual stimuli and pressed keys accordingly in a timely manner. To assess both the quality of data and efficacy of different machine learning options, I conducted 46 leave-one-out crossfold-validations, where the data held out in each fold was a single trial (30 in three experiments, and 16 in the fourth). The machine learning algorithm thus built a model over all-but-one trial, whose class (easy or hard) it predicted, repeating this procedure until every individual trial had supplied the testing case. In total, this resulted in 1328 ($(16+10+14) \times 30 + (8\times16)$) unique testing cases; in other words, the listed classification accuracies represent the model trained and tested on 1328 separate occasions. If the fNIRS data was not indicative of the trial's associated class, the algorithm would likely classify roughly half of the trials correctly, and thus a classification accuracy well-above 50% over many trials suggests the presence of class-predictive information in the datasets and the capability of the associated algorithm to discover it.

What follows is a suite of evaluations on different machine learning algorithms in order to (a) provide basic guidelines for how to build useful fNIRS analysis tools using the neuracle software package described in Chapter Six, (b) give deeper insight into what components of the data inform changes in mental state, and (c) make explicit trade-offs to consider both for the design of experiment and the choice of

Table 1.1: The experiments analyzed in an agnostic neuracle

	Visuospatial n-back A	**Visuospatial n-back B**	**Visuospatial n-back C**	**Cognitive Manual Task**
Subjects analyzed	16	8	2	14
Sessions analyzed	16	8	10	14
Total subjects	28	16	7	14
Total sessions	28	16	35	14
Trials per session	30	16	30	30
Trial length (sec)	25	40	25	30
Sampling rate (Hz)	11.79	6.25	11.79	11.79

machine learning algorithm. In these results, the absolute classification accuracies are not as important as the relative accuracies, which indicate the effect of various combinations and choices of classification approach.

1.2.4 Channels

These experiments were run on an ISS Imagent with 16 channels of fNIRS data made up of two detectors with four linearly arranged light sources associated with each detector (see figure 1.3). An analysis of machine learning performance on only one channel reveals that while all channels are independently informative, there was little difference between detectors sampling the left versus right PFC nor between those calibrated to identify 830 versus 690 nanometer reflected light (see table 1.2). It is however worth noting that the two detectors farthest from the light source (3.5cm on the left and 3.0cm on the right) had highest single classification accuracy, which is expected since their positions enable them to sample the deepest neural tissue, while the shallow channels mainly contain noise and artifacts originating closer to the skin.

1.2.5 Time Segments

To evaluate where in the time segment information is concentrated, as well as determine optimal blocks for partitioning a trial, neuracle evaluated the performance of classifiers on individual separate windows of the dataset and multiple aggregated

Table 1.2: Classification accuracy when only one channel is used

Channel	Accuracy
A-DC4 (left, 830, 3.5cm)	66.7%
B-DC8 (right, 690, 3.0cm)	66.6%
B-DC7 (right, 690, 2.5cm)	66.3%
A-DC2 (left, 830, 2.5cm)	64.1%
B-DC6 (right, 690, 2.0cm)	63.7%
B-DC5 (right, 690, 1.5cm)	63.6%
A-DC1 (left, 830, 2.0cm)	63.2%
B-DC3 (right, 830, 3.0cm)	63.0%
A-DC3 (left, 830, 3.0cm)	61.8%
A-DC7 (left, 690, 2.5cm)	61.7%
B-DC1 (right, 830, 2.0cm)	61.2%
A-DC5 (left, 690, 1.5cm)	61.1%
A-DC8 (left, 690, 3.0cm)	60.5%
B-DC2 (right, 830, 2.5cm)	59.9%
B-DC4 (right, 830, 3.5cm)	59.7%
A-DC6(left, 690, 2.0cm)	58.8%

windows of the dataset. In Table 1.3, the first row labels different windows of the dataset, so that 1/1 stands for the whole time-segment treated as one cohesive unit and 4/5 stands for the fourth fifth of the dataset, the content starting at the 60th percentile time-stamp and ending at the 80th percentile. The next row shows classification accuracies when the algorithm only examines that particular window. These results suggest that the task-predictive information is insulated towards the end of the trial, which is consistent with the fact that the measurement technique relies on the slow movement of blood. The rows that follow show classification accuracies when many of these time segments are aggregated together. These comparisons would suggest that it makes sense to partition the data into subsegments, but two or three suffice. One can partition the dataset into thirds, fourths, or fifths, but this does not buy the algorithm much new information. Examining the whole, the first half, and the second half appears to extract the bulk of information.

1.2.6 Features

The second row of Table 1.4 shows the classification accuracies when the machine learning algorithm is permitted to train on only one statistical description of the

Table 1.3: Classification accuracy by time segment; *o* denotes that features are extracted over this segment.

1/1	1/2	2/2	1/3	2/3	3/3	1/4	2/4	3/4	4/4	1/5	2/5	3/5	4/5	5/5	Accuracy
70.1%	63.0%	66.6%	58.5%	65.5%	66.0%	57.1%	62.5%	64.9%	66.6ˆ	56.5%	60.4%	60.4%	64.2%	65.3%	
o	o	o	-	-	-	-	-	-	-	-	-	-	-	-	74.3%
o	-	-	o	o	o	-	-	-	-	-	-	-	-	-	72.9%
o	-	-	-	-	-	o	o	o	o	-	-	-	-	-	73.5%
o	-	-	-	-	-	-	-	-	-	o	o	o	o	o	73.5%
o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	74.2%

Table 1.4: Classification accuracy by feature; *o* denotes that this feature is analyzed by the machine learning algorithm.

Linear Slope	Std. Dev.	Min.	Time to Peak	Abs. of Mean	Max.	Mean	Abs. of Slope	Full Width At Half Max.	Second Der.	Accuracy
66.3	63.6	60.4	59.9	59.3	59.0	58.4	57.5	57.1	46.8	
o	o									69.0
o	o	o								69.2
o	o	o	o							71.3
o	o	o	o	o						70.8
o	o	o	o	o	o					70.9
o	o	o	o	o	o	o				71.1
o	o	o	o	o	o	o	o			71.4
o	o	o	o	o	o	o	o	o		72.5
o	o	o	o	o	o	o	o	o	o	73.3

time segment at each channel. The analysis indicates that linear slope (the difference between the last and first value divided by the number of observations) appears to be the most informative way to describe the time segment, and the best starting point for analysis. The subsequent rows show accuracies when the next most independently productive feature is added to the selection. These results reflect the intuitive phenomenon that describing the trial in terms of more statistical features (even the least predictive ones) tends to boost accuracy.

1.2.7 Machine Learning Algorithms

Table 1.5 shows a comparison of different machine learning algorithms. Weka's support vector machine (with a polynomial kernel and cache (= 1.0) parameters) outperformed the other algorithms [68]. It is worth noting that I examined different

Table 1.5: Classification accuracy by algorithm.

Algorithm	**Average**
Support Vector Machine: (Weka SMO)	74.3%
Multinomial Logistic Regression (Weka)	71.7%
Support Vector Machine: LibSVM (Weka wrapper)	67.6%
Adaboost on Decision Stump (Weka)	67.6%
Logistic Model Tree (LMT, Weka)	67.6%
Simple Logistic Regression (Weka)	67.5%
Naive Bayes (Weka)	65.4%
3-Nearest Neighbor (Weka)	64.4%

parameters (Weka's RBF and Puk kernels, and different sizes for the cache) for the support vector machine, but no combination of values led to superior results. In addition, procedures that tailored the selection of kernel and cache to the present training set using cross-validation had only a negligible impact on accuracy.

This dissertation makes no theoretical contributions to machine learning, and readers are referred elsewhere for complete descriptions of underlying implementation (e.g. the Weka source code [68]). But as preparation for subsequent chapters, it is worthwhile to sketch out the critical intuitions that underlie machine learning, especially the support vector machine and logistic model tree, as they illustrate important concepts about how the brain works. The primary aim of this dissertation is to identify a path forward in BCI that is not as dependent on machine learning algorithms at the neuracle level. However the physical paradigm for BCI will depend on a greater dependence on machine learning at the *physical level*, where the brain should be dimensionalized as a machine learning algorithm, or hierarchical prediction and error correction engine. With a physical neuracle classifying the spatiotemporal correlates of prediction error in the brain, machine learning card instead be used at the *interface level*, where inputs to the brain can be tweaked in a feedback loop with that classification.

1.2.7.1 The Support Vector Machine

The goal of a support vector machine (SVM) is to find a hyperplane in a k-dimensional space that optimally separates instances of two classes on the basis

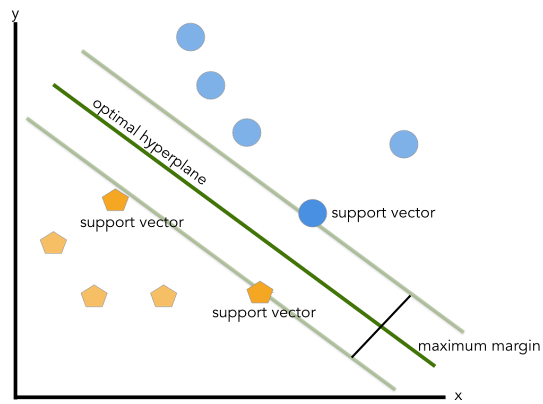

Figure 1.7: A 2-dimensional support vector machine

of their k features. As illustrated in Figure 1.7, the optimal hyperplane is the one that has the maximum distance from instances of both classes. When the algorithm is trained, the support vector machine classifies new instances by computing where its k features fall relative to the hyperplane. The hyperplane is represented in a k-dimensional feature space with the function:

$$f(\mathbf{x}) = \sum_{i=1}^{k} x_i w_i + b.$$

The learning function discovers settings to $\overleftarrow{w}$ and b that optimally separate the two classes. To accomplish this, the SVM algorithm initializes random settings for $\overleftarrow{w}$ and b, and creates a simple mapping function *y(x)* that transforms the representation of two possible classes into a positive and negative integer, i.e. all values below some threshold receive -1 and all above receive 1. If the SVM arrives at a perfect model for the data, then the product of the weight vector ($\overleftarrow{w}$) and feature vector ($\overleftarrow{x}$) minus b becomes greater than 1 when the class is of one kind and less than -1 when it is of the other kind. Put in mathematical terms, the algorithm minimizes the value of the objective loss function

$$L(x) = \frac{1}{2} \sum_{i=1}^{n} max(0, 1 - y_i(\overleftarrow{w}_i \overleftarrow{x}_i - b)) + \alpha ||\overleftarrow{x}||,$$

where α is a constant that determines a trade-off between increasing the margin size and guaranteeing that instances fall on the correct side when classified. Modern implementations use *gradient descent* to incrementally adjust $\overleftarrow{w}$ and b in the cases when some instance falls on the wrong side of the hyper-plane [16].

The mathematics underlying the internal mechanics of the support vector machine pave the way for intuitions about how the brain operates. As I will describe in the next chapter, the brain can also be thought of as a machine learning algorithm that adjusts internal settings to minimize a loss function (e.g. characterizing the average amount of surprise of sensory signals). But the brain does not use gradient descent to minimize entropy, instead relying on the degree to which a neuron is affected by and affects other neurons as the underlying basis for encoding knowledge. This more flexible and feature-agnostic approach to learning resembles the *neural network*, which has surged in popularity since the original design of neuracle, and is therefore omitted from this chapter. [1]

1.2.7.2 Logistic Model Tree

The intuitions underlying the logistic model tree (the third best algorithm in my analysis) provide useful and complementary information towards an understanding of brain activity. The logistic model tree is a standard decision tree that uses logistic regression at its leaves to overcome an unaugmented decision tree's risk of underfitting. The algorithmic strategy of a decision tree matches the Bayesian approach deployed subconsciously by the brain to selectively attend to features of the environment.

As with the support vector machine, the decision tree establishes a method to prioritize certain features above others, singling out those that are most informative in classification. In an SVM, this filter happens implicitly in the assignment of weights to features, so that features that receive a weight of zero are eliminated from consideration. The algorithm for a decision tree ranks features more explicitly by computing their *information gain*, which measures how much information a feature

[1] Future research may consider integrating neuracle's Weka-based machine learning infrastructure with modern approaches, such as *Tensorflow*, although potential contributors to neuracle should know that the bottleneck for accurate agnostic neuracles is not within the field of machine learning, as will be made clear in the next chapter.

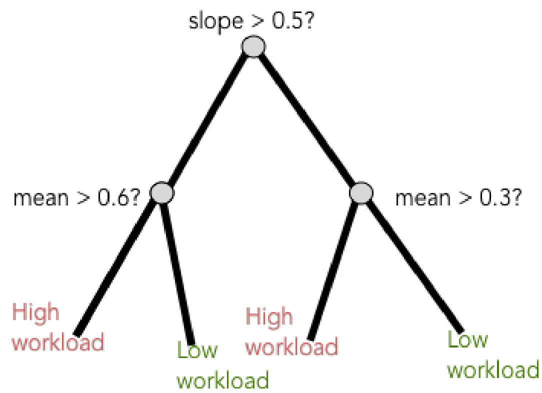

Figure 1.8: A simple decision tree

provides about the class. The algorithm thus favors features that tend to assume a given value for one class and a different value for another class. Figure 1.8 shows plausible logic for a decision tree inferred from the fNIRS data analyzed in this chapter, where the slope of values has the most information and is therefore placed at the top of the tree.

Information gain is encoded by an equation that captures the change in *entropy* from some initial state T to a posterior state that has taken x into consideration:

$$I(T, x) = H(T) - H(T|x).$$

A definition and discussion of the equations for computing entropy H(T) will be saved for the next chapter (Section 2.2).

1.2.8 Offline Agnostic Neuracles

The experiments above show the effect of various parameter and preprocessing approaches. From this, I can recommend that a trial be broken into segments. For each channel in these segments, a set of statistical features can be computed. Neuracle found the highest classification accuracy when using slope and standard deviation over a segment. For potential performance improvements, additional features could be added such as minimum, time-to-peak, absolute-of-mean, mean, largest, absolute-of-slope, full-width-at-half-maximum, absolute-of-slope, and second derivative. Using Weka's SMO support vector machine shows promise, and should be used to build agnostic neuracles.

This analysis suggests that even rudimentary filtering and attribute selection approaches can provide reasonable classification accuracy. Future work could focus on expanding the feature set. In particular, a suite of features that describes amplitudes at various frequencies on Fourier-transformed segments may provide complementary information. Other features that better approximate how the segment changes over time are also promising candidates for a better algorithm. Some of these candidate features, including frequency domain descriptions and SAX-time-series representations [84] are included in the neuracle software.

The fact that the reported accuracies are better than chance — as high as 74% when using the support vector machine on all features, channels and time segments — demonstrates the agnostic neuracle *thesis* of this dissertation. The next section will examine the possibility of using agnostic neuracles in real-time, where their classifications drive adaptations in *implicit BCIs*.

1.3 Interaction - Implicit BCIs in the Agnostic Paradigm

Implicit interfaces adapt the content, structure, or presentation of hidden or detectable elements without the user's explicit intention. From the perspective of the user, implicit input is free: an additional bandwidth of information from the user to the computer. Because BCI input tends to be probabilistic and error prone, a well-designed implicit interface never makes drastic changes in the face of uncertainty; instead, it subtly alters the content or appearance of internal or external data in a manner that is completely innocuous when the user's intention or mental state is miscalculated [128]. The next two sections organize implicit BCIs into two categories depending on whether the *implicit BCI* operates in a stationary computing or wearable computing context.

The experiments evaluating implicit BCI for stationary computers were led by Daniel Afergan, and a reference to the full description of this work can be found

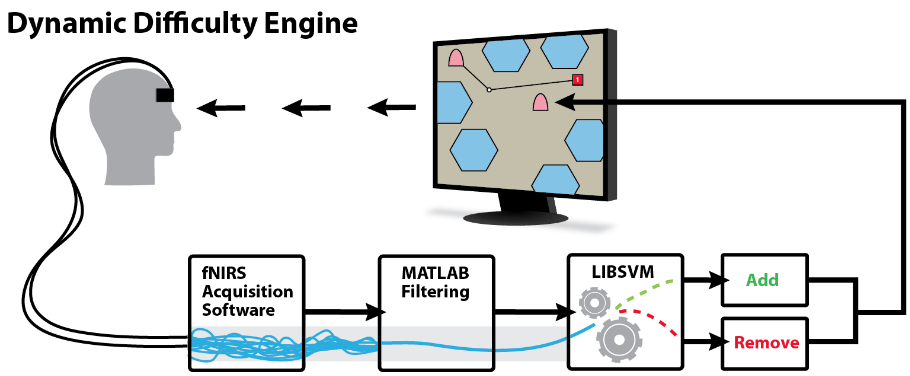

Figure 1.9: The Agnostic Paradigm for BCI [6]

in the bibliography at [6] and [7].

1.3.1 Implicit BCI for Stationary Computers

Brain-based adaptations may alter the definition of a user interface at *semantic* or *syntactic* levels [128]. In HCI terminology, the *semantic* level of a system refers to the set of internal functions and settings that control its behavior. In a digital text reader, this refers to, among other entities, the underlying data structures that store in memory the text to be visualized on the screen. In contrast, the *syntactic* level refers to properties governing the interface established with the user. In a digital text reader, this may refer to the speed at which the visualized text changes when the user scrolls down.

1.3.1.1 Semantic Adaptations in a UAV Simulation

In an experiment led by Afergan, we evaluated the feasibility of building a *dynamic difficulty adjustment* engine by establishing an implicit link between a semantic property of a user interface and fNIRS-based classifications of a user's brain state. Here, the semantic property was a single variable determining the number of unmanned aerial vehicles (UAVs) under user control in an air traffic control simulation.

In this experiment and others in the agnostic paradigm, the experiment proceeded in two phases. In the calibration phase, an agnostic neuracle was trained to distinguish between the user's high and low cognitive workload states. Offline calibration follows the guidelines and best practices described in section 1.2. Participants first completed fifteen 25-second visuospatial 3-backs and fifteen 25-second visuospatial 1-backs, woven together by 15-second resting periods. Each of the thirty *n-back* timeseries were defined by two features (mean and slope), resulting in 32 features total for each instance. Thirty instances, each with 32 features, thus supplied the underlying information for the support vector machine learning algorithm to construct a model separating the user's high and low cognitive workload states.

In the *experimental phase*, the real-time classification of the support vector machine generated continuous estimates of the user's cognitive workload state. As illustrated in figure 1.10, the user was instructed to set a flight path between UAVs and their target destination by arranging a sequence of points on the screen. To reach the destination without penalty, the flight path should avoid collision with objects labeled as no-fly-zones. In order to evaluate the efficacy of brain-based adaptation, the experiment included two conditions. In the *adaptive condition*, the system removed UAVs from user control when they were deemed to be in high cognitive workload, and added UAVs when they were deemed to be in low cognitive workload. In the *non-adaptive condition*, the number of UAVs was set according to a predefined script that guaranteed an equal amount of total work to the *adaptive condition*.

Figure 1.11 shows measured user performance along four dependent variables. Most notably, brain-based adaptation decreased operator failure rate (as measured by the number of **collisions** with no-fly zones) by 35% compared to the non-adaptive condition.

1.3.1.2 Syntactic Adaptations in a Brain-adaptive Bubble Cursor

In another experiment also led by Afergan, we evaluated the feasibility of changing syntactic properties of a user interface using brain classification.

Following the guidelines described in section 1.2, we built an agnostic neuracle that classified the user's cognitive multitasking state as he or she completed an

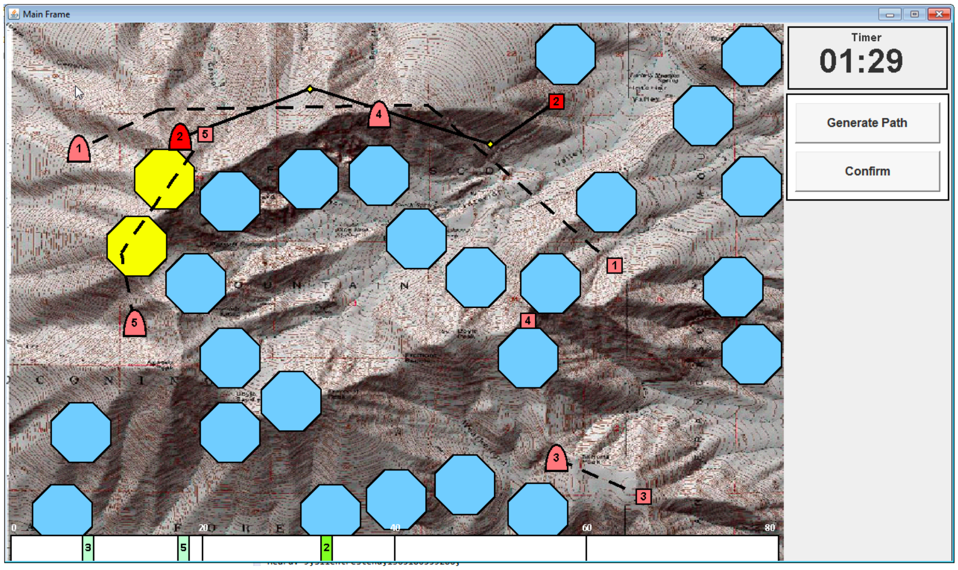

Figure 1.10: Afergan's UAV simulation

artificial point-and-click task intended to assess the effectiveness of a mouse expansion technique known as bubble-cursor (see figure 1.12). Interfaces implementing the standard bubble-cursor target expansion technique guarantee that the element nearest the mouse-pointer is always selected [56]. In our brain-adaptive version of bubble-cursor, the element selected was a function of its proximity to the cursor as well as the user's brain activity.

The experiment began with a *calibration period* where the user performed a task in two different difficulty settings, while their brain state was measured using fNIRS. A support vector machine was trained to differentiate the mean and slope of the fNIRS time-series pertaining to the two conditions, leading to an agnostic neuracle prepared to classify the user's current level of cognitive workload.

In the *main phase* of the experiment, the user completed two parallel tasks. As their primary task, the user repeated numbers they heard at a delay (n) in an n-back. The primary task changed the difficulty of the n-back by increasing the delay at which numbers were to be recited. As their secondary task, the user was given a sequence of uppercase and lowercase letters which they were instructed to locate on-screen. The uppercase letters had a higher priority than the lowercase letters, which they could skip without major penalty. In the *adaptive condition* of

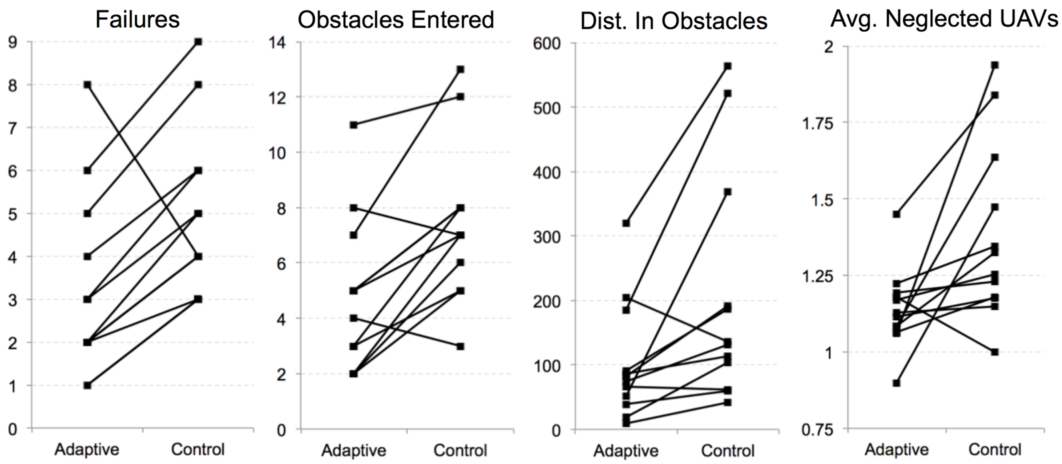

Figure 1.11: These four dependent variables show significantly better performance in the brain adaptive condition ($p < 0.05$). Each line represents a subject.

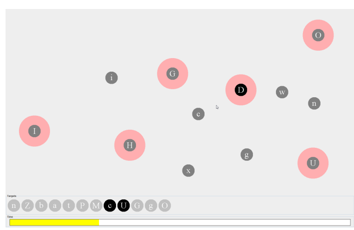

Figure 1.12: Screen capture of bubble-cursor target expansion

the experiment, the uppercase letters were made easier to select when the agnostic neuracle classified the user to be in a high cognitive workload state. The *adaptive condition* was compared to conditions with *no expansion* (where the bubble cursor feature was off) and *static expansion* (where the bubble cursor was active but did not change based on brain activity). Compared to the non-brain conditions, the *adaptive condition* entailed significantly improved user performance on every dependent variable measured: most notably the combined score in the simulation as well as the n-back performance (see figure 1.13).

The fact that these two experiments both showed significant improvements to user performance when semantic or syntactic properties of the interface changed based on real-time brain activity demonstrates the *implicit BCI thesis* of this dissertation.

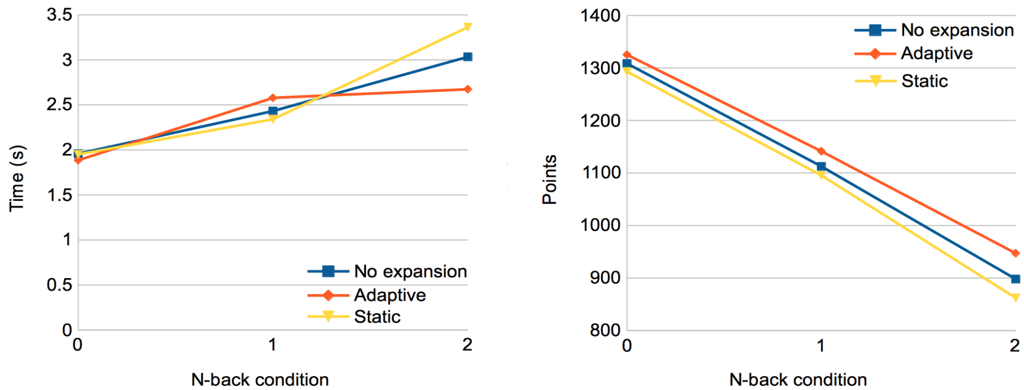

Figure 1.13: Left: Movement time for uppercase letters by expansion condition. Right: Points by expansion condition [7].

1.4 Implicit BCI for Wearable Computers

Implicit BCIs for *wearable* computers serve a different role than for *stationary* computers, enhancing usability in products whose function is to facilitate (and not replace) the user's ordinary interaction in the world. Head-mounted wearables such as Google Glass own a tiny share of real estate in the user's visual periphery, where they can broadcast on demand information depending on the user's location or task [123]. This emerging genre of human-computer interaction creates an intimate bond between user and computer that can also be disruptive, especially if the device mistimes its output, clutters and confuses, or in any way interrupts the user as she engages with other demands imposed by the real world. Well-designed applications err on the side of not interrupting the user. Even better would be applications that leave this decision to in-the-moment probabilistic calculation, and throttle notifications when the user's present situation or mental state appears unlikely to benefit from its content. In this section, I first investigate the space of possible implicit BCIs for wearable computers and then prototype two user interfaces that rely on passive input to alter notifications and content in consumer-grade wearable computers.

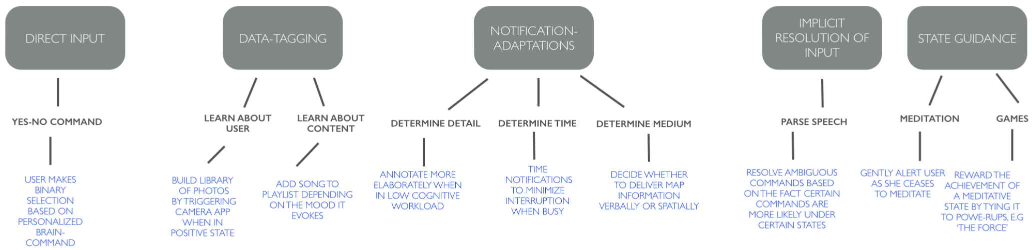

Figure 1.14: Five methods for using physiological data as input to head-mounted computers, with sample use cases.

1.4.1 The when, how, and what of wearable notification streams

Unlike traditional computers that 'own' the user's attention, wearable computers are companions to the real world: digital assistants, whose utility hinges on successfully relaying context-specific information [20]. At any given moment, an active application must judge whether present circumstance warrants introducing information to the user's visual field, and if so, its content and format.

When the system has arrived at the decision to interrupt, it can further customize relevant parameters such as the content, level of detail, the minimality of design, and medium of presentation in accordance with the user's mental state. For example, it could swap between oral, written, and visual formats for route recommendations, capitalizing on the fact that working memory appears to support distinct (and parallel) processing units for these modalities [14]. In keeping with general design principles for implicit interfaces [128], these adaptive wearable applications should never shock the user; nor indeed, given their tendency to err, let the user in on the 'secret' that a probabilistic mental state estimate has changed some aspect of the interaction. Mission-critical information should never be held back nor erratically switch formats. Instead, well-designed implicit interfaces would target non-essential or time/context insensitive information for adaptation. For example, a turn-by-turn navigation system would not drastically alter how or when it displayed an impending turn, but it could use its estimates to dictate the display of nearby gas stations, the presence of roadside attractions, or speed limits.

1.4.2 User Discovery

When navigating the web, a user leaves a rich cyber-trail that exposes their personality, mood, and preferences. Research into recommender systems has optimized methods for translating web interactions (e.g. clicking a link) into valuable models of the user. For example, Twittomender uses tweets and existing followers to recommend new followers [59] and Tagommender combines both user's search and rating history to infer a user's preference for tags and other movies [114].

fNIRS-based predictions of the user's preference have been used to augment movie recommendation algorithms [103]. Designed to avoid prolonged sessions of device interaction, wearable computers must incorporate other means than clicking links for user profiling. Wearables tend to remain active even when not used, as well as support a variety of sensors (e.g. camera, GPS, microphone), opening several channels for context sensitivity [39]. With concurrent physiological sensing, data from the environmental sensors could be timestamped with predictions about the user's mental state, and estimate the places, people, tasks and time periods associated with mental idleness, working memory engagement, positive emotion, focus, and mind wandering. These datapoints could inform targeted advertisement, recommendations, or could be related explicitly to the user in the form of a cognitive/emotional heatmap that associates space, time, and images with a current prediction of the mental state. It is important that these systems are scrutinized by benign scientific supervision, and are designed to promote human wellbeing.

1.4.3 User Interface Experiments

This section describes two previously unpublished implementations of *implicit BCIs* for wearable computers. The prototypes are tested on a small number of participants, and the purpose is to illustrate the parts of a complex system as opposed to evaluating significant results. The systems were built using the software Neuracle distributed in Chapter Six (and available online at *www.github.com/samhincks/neuracle*).

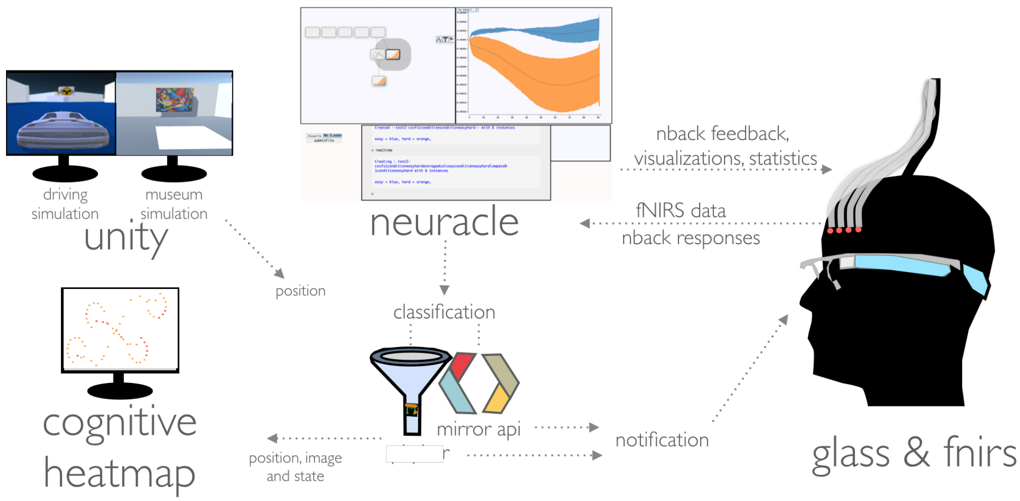

Figure 1.15: System architecture for Neuracle and implemented Glass prototypes. fNIRS data communicates to Neuracle server via an intermediate database. Here, the user calibrates a machine learning algorithm on their own data by solving an n-back in the console. Realtime classifications can then be redirected to a local port, (e.g. to *Phylter* [5]) or to a web application, and external wearables like Glass via the Mirror API.

1.4.3.1 Participants

Five participants (3 female, 2 male), ranging from ages 20 to 25, partook in the *route recommendation* experiment, and two of the five completed the *cognitive heatmap*.

1.4.3.2 Adaptive Route Recommendation

Google, Apple and other technology companies have dedicated considerable resources towards accurately mapping the planet's roads and collecting detailed information about their convenience and traffic patterns. Research in human spatial navigation ability lays the foundation for the user interface of these systems, informing parameters such as the the timing and frequency of turn notifications as well as the level of detail most amenable to a balance between comprehensible and information-rich content. But in many cases, optimal settings for these parameters are not fixed across time and user, but highly dynamic, a function of the user's mood or workload. There is usually more than one route to reach a particular destination, and often routes vary both in speed and complexity. Invariably, turn-by-turn navigation software tends to suggest the swiftest, shortest route, sometimes guiding users down difficult paths that lead them astray. If one path is slightly slower but significantly easier to process and navigate than an alternative route, then an up-to-date model of the user's cognition can aid the choice of what path to suggest.

1.4.3.3 Architecture

The experiment simulated a wearable turn-by-turn navigation system by integrating Google Glass with an Imagent fNIRS imaging device. Imagent broadcasts raw data to Boxy software, which transmits calculations of neural oxygenation levels to Neuracle [2]. Neuracle rebroadcasts predictions of the user's cognitive workload to custom software called *Phylter* [5], which intercepts user locations from Unity software [4], and decides whether or not to recommend the shortcut. The Unity simulation included four roads (see figure 1.16): each supports one quick but challenging route, one long but easy route, and a variety of other routes that lead the user astray. The quick route requires the user to read, memorize, and execute four directions.

1.4.4 Method

The user began by solving either an oral 0-back (low cognitive workload) or a 1-back (high cognitive workload). After twenty seconds, the driving simulation began (as the user continued to solve the n-back), and within ten seconds, the suggested route appeared on Glass. In the static condition, the user always received the shortcut: four simultaneous instructions. In the adaptive condition, the user received the shortcut when they had a low predicted cognitive workload and the longer route when they had a high predicted cognitive workload.

1.4.4.1 Results and Discussion

In the adaptive condition, the system correctly displayed the appropriate route in Glass in each of the ten instances (2 roads for 5 participants), depending on its present calculation of the user's workload. However, users did not arrive at the destination faster nor suffer fewer collisions during the adaptive condition. They

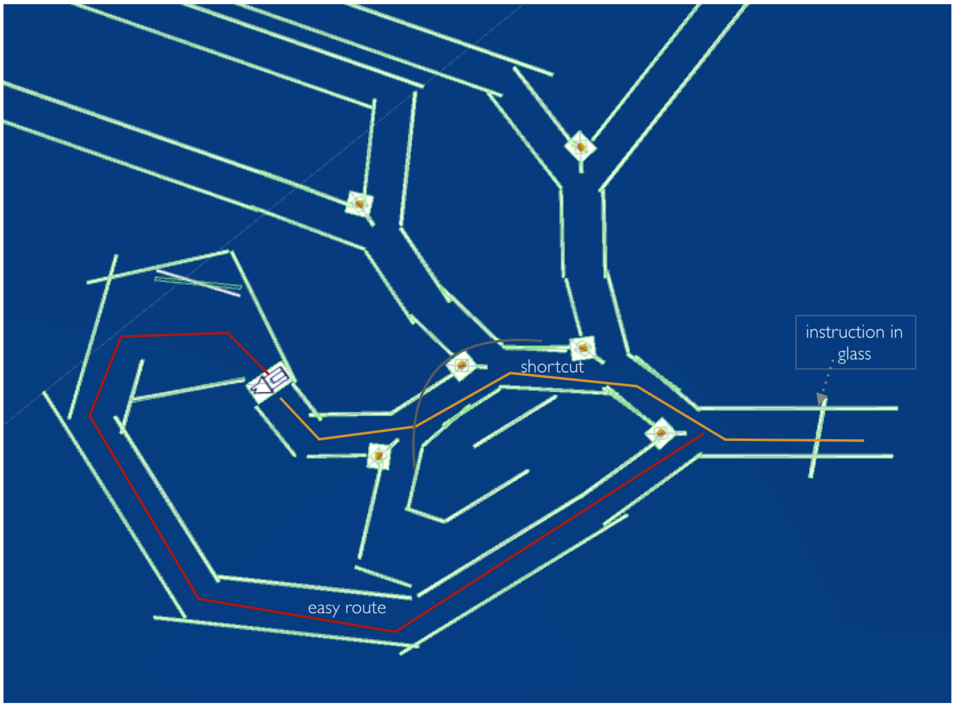

Figure 1.16: Unity map design.
Two paths to reach a destination: one straightforward route and one which requires the participant to remember four simultaneous directions.

completed the simulation approximately as fast when completing a 0-back (m=99s, s=50s) as when they completed a 1-back (m=105,s=44s). Figure 1.17 (which shows concurrent brain measures) shows the raw data on which Neuracle based its prediction of user workload. This validates the bidirectional communication capability of Neuracle; fNIRS data can be tagged from external software, which enables it to learn new user classifications post-calibration.

The high variability of results indicates that adaptive route recommendation requires careful settings to several subject-specific parameters, which may be difficult to extract in a laboratory setting. A proper cloud-based service like Google Maps would infer difficulty and skill parameters of route recommendations and users based on how frequently the suggested turn failed to produce the correct future GPS coordinates. Concurrent measurements of the user's mental state could then identify the degree to which taking the wrong turn was sensitive to their mental state, and suggest easier routes for the state-sensitive users when in a high workload or

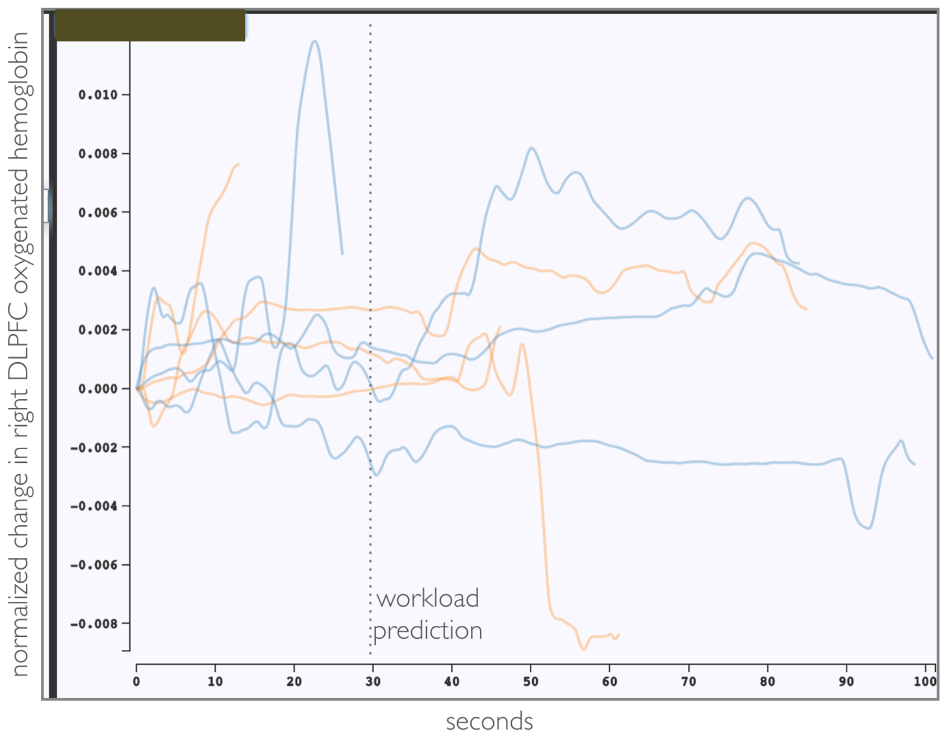

Figure 1.17: fNIRS measurement in right dlPFC during driving simulation for two subjects.
In blue trials, participant completed a 0-back (easy), and in orange trials, participants completed a 1-back (hard) while driving. The system delivered the direction after 30 seconds, and trials varied in length depending on how long it took to reach the destination. In the few samples shown, easy trials tended to provoke more variable data than hard trials, which matches the effect observed in offline calibration analysis.

otherwise compromised state.

1.4.4.2 Cognitive Heatmap

A well-designed wearable computer supports its user's goals even when its services are not explicitly requested [39]. Equipped with a camera, the device could take repeated snapshots of the user's vantage point. As a result, the user would have a log of data that documented every activity engaged, in every acquaintance encountered, and every location visited. If images were tagged with meaningful descriptions about the user's state, the user could search time points of interest and also catalog the sorts of events that stimulated particular mental states. For example, if the application operated while its user attended a lecture, she could afterwards revisit a state-indexed timeline, and discover the moments that stimulated the highest levels of cognitive workload, or, conversely, the moments when she wasn't paying attention. This experiment prototypes a cognitive heatmap that enables users to select relevant photos based on the associated cognitive state.

1.4.4.3 Architecture

The architecture resembles the previously described system, except neuracle's cognitive workload predictions never alter the contents of Glass. Instead, Unity position coordinates, Glass photos, and mental state predictions are all associated in a single file, which D3 visualization software [1] uses to display a set of points, colored on a continuous scale from yellow to red representing the confidence that their workload is high, which the user can click on to retrieve the associated image.

1.4.4.4 Method

Participants worked as a museum curator, instructed to count the distribution of paintings in various rooms. Each room was a simulated environment, and the game automatically moved them from one point to another; the user controlled the direction of the camera. Some rooms included only one type of painting, and thus the task amounted to counting the number of paintings. Other rooms included two easily distinguishable types of paintings (Pop Art and East Asian), and updating and memorizing this distribution was meant to stimulate high cognitive workload. These rooms contained signs pointing to the content and time of future exhibits. In a later task, the user had to use the cognitive heatmap to find the image that contained this sign. If the system worked correctly, then these would be by the high cognitive workload (red) points.

1.4.4.5 Results and Discussion

Participants reported using the indexing of their mental state as a search mechanism. However, they did not find the photo quicker when the map was colored according to their cognitive workload. It was challenging to get the subject to fixate their gaze on the relevant image, and so some of the generated maps included only a blurred rendition of the target image. I nevertheless think the system shows promise, especially with well-calibrated models whose confidence values communicate true model uncertainty and in use cases where it need only pluck out a fraction of true positives in order to benefit the user, e.g. if the user only has time to review one lecture out of a longer series.

This chapter has given four experiments in the agnostic paradigm, demonstrating the *implicit BCI* thesis. However, two of these experiments were aborted when it was clear it would not be possible to obtain statistically significant effects. This failure and others are the result of shortcomings with the agnostic paradigm, which is the focus of the next chapter.

Chapter 2

The Problem and Opportunity of Entropy

The previous chapter reduced the challenge of building a BCI into three subproblems: *dimensionalization*, *portrayal*, and *interaction*. It gave a solution to these problems, and called it the agnostic paradigm for implicit BCI. This chapter is intended to bridge the existing agnostic paradigm to a new physical paradigm by demonstrating the *entropy thesis* of this dissertation. *Neural entropy poses a problem for* agnostic neuracles *and an opportunity for* physical neuracles *as a property to be detected in the user.*

This chapter builds a case for the *entropy thesis*. The first section will expose five problems with the agnostic paradigm, teeing up the remaining sections that linger on the fifth problem: the problem of entropy. As promised by the previous chapter, the *dimensionalization* section of this chapter pushes a modern conceptualization of the brain as a hierarchical prediction and error correction engine; and the *portrayal* section that follows gives empirical support for this new Bayesian model of brain activity.

2.1 Limitations of the **Agnostic Paradigm** for Implicit BCI

To summarize the previous chapter: in implicit BCIs following the agnostic paradigm, a user's mental activity is *dimensionalized* along the axis of cognitive workload, which is more narrowly defined as the content load of short term working memory buffers in the prefrontal cortex. In order to *portray* the burden of working memory to a user interface, changes in oxygenation consumption are measured at dorsolateral regions of the prefrontal cortex through functional near-infrared spectroscopy (fNIRS), which pulses near-infrared light through the skull of the forehead and detects the amount of light that returns to a sensor approximately three centimeters away from the light source. An agnostic neuracle builds a model for classifying the strain on a user's working memory on the basis of these fNIRS measurements. In order to show the brain in different memory conditions, participants perform a benchmark cognitive workload induction task (the n-back) alternating between its easy and hard versions. The fNIRS signals of these trials are then described in terms of statistical features and grouped into two sets (easy vs. hard), which supply the instances for a machine learning algorithm. Once trained, the machine learning algorithm delivers real-time predictions to the implicit interface for the evaluative portion of the experiment.

The final sections of Chapter One show four solutions to the *interaction problem* of mapping working memory classifications onto changes in system properties of wearable and stationary user interfaces. Two of these experiments showed significant improvements to user performance, and two were aborted when it was clear the experiment would not yield any significant effects. There are at least two other published implicit BCIs following the agnostic paradigm that demonstrate significant improvements to user performance. In one, the agnostic neuracle was used to initiate automation from a robot [121]; and, in another, user state classifications were used to determine whether a novice piano player was prepared for a more difficult piano lesson [132].

For each published experiment that showed significant improvements to system usability when supplemented by brain input, there may be others that failed to show improvements, but are not known because of the positivity bias in scholarly publication. The wearable computer experiments in the previous chapter were aborted when it was clear the *agnostic* solution would not lead to an implicit BCI that enhanced system usability. Many experiments may fail because an agnostic neuracle could not reliably classify the user's cognitive workload state in real-time due to inherent difficulties in this technique [64]. In particular, at least five issues shared with machine learning in other neuroimaging devices [116, 82, 23], must be addressed.

1. *(a) Psychology problem*: a cognitive dimension postulated on the basis of high level psychological science, interpreting reaction times and subjective reports instead of brain activity, may not be fundamental enough to the brain's action for detection by non-invasive neuroimaging.

2. *(b) Overfitting problem*: the task paradigm may elicit a distinguishable physiological trace during the calibration task, but the general post-calibration scenario implicates a wider range of states than the neuracle is trained to predict.

3. *(c) Disengagement problem*: the task paradigm may fail to elicit consistent differences in the brain because the task fails to combat the brain's endogenous prerogatives and engage the brain's attention.

4. *(d) Synchronization problem*: by training the algorithm on features from a series of well-delineated n-length sequences, the algorithm is only well-prepared to classify the state on the rare occasion that the user is n seconds into a state that transitioned from a similar rest profile that preceded the task in the calibration period.

5. *(e) Entropy problem*: the neurological activity associated with the state's novelty may be more pronounced than the activity associated with the state un-

der investigation [65], yielding a unique signature in each machine learning instance.

In this chapter, I will investigate the degree to which these problems affect the fNIRS-based agnostic paradigm established in Chapter One. Together, these sections motivate the need for physical neuracles which will be the focus of the next chapter.

2.1.1 The Psychology, Overfitting, and Disengagement Problems

The *psychology (a)* problem poses no challenge to the agnostic paradigm in Chapter One if cognitive workload exists on a spectrum in the brain so that brains are sometimes cerebrally active and sometimes deactive. I will refer to this as *assumption a*. The *overfitting (b)* problem is unproblematic if the difference between a brain completing a 3-back and a 1-back is similar to the difference between a brain completing any other hard task and any other easy task (*assumption b*). Finally, the *disengagement problem* can be ignored if the experimental context and n-back task reliably command control over the user's attention (*assumption c*).

It is unclear which of the assumptions *a, b,* and *c* are true. Evidence from neuroscience literature suggests that *assumptions a* and *b* cannot both be true. Globally, the brain consumes approximately as much energy when there are high and low demands from the environment, indicating that there is no cognitive workload *master dimension* in the brain [108]. For *a* to be true, cognitive workload must refer to a more specific spectrum in the brain, such as the activation of a short term working memory buffer in the brain. But in that case, the agnostic neuracle is unlikely to generalize to a wider array of mental states, and *assumption b* is likely false.

The high variability of offline classification performance in agnostic neuracles suggests that *c* is false, and there is not a single *calibration protocol* that universally elicits the subject's engagement. In his dissertation, Afergan mentions anecdotal evidence suggesting that agnostic neuracles perform better when the user reports performing the audio n-back according to a specialized cognitive tactic [8]. Chapter Five in this thesis affirms this conclusion. That chapter shows results from an n=1 experiment published here [64], where I report on a multi-week self-experiment measuring my own brain while I completed an n-back. In order to elicit a reliable signal from fNIRS, the task had to put me at the edge of my abilities, and I therefore had to increase the difficulty of the n-back over the course of the multi-week self-experiment to witness an associated increase in fNIRS-detectable prefrontal cortex activity.

These observations alone call for algorithms that encode the relationship between fNIRS signals and mental states a priori, without relying on a calibration period. The need for a physical paradigm is further motivated by the absence of obvious solutions to the *synchronization* and *entropy* problems.

2.1.2 The Synchronization Problem

The *synchronization problem* (d) poses no challenge to the agnostic paradigm in Chapter One if the information distinguishing different mental states can be defined over a short time scale and if that information does not depend on a context specific to the calibration period (assumption d). But the offline and real-time agnostic neuracles given in Chapter One indicate that *assumption d* is false, and that the best source of information is the slope over the trial's entire time-series. This can be seen visually in figure 2.1, where the canonical high cognitive workload state in the three-back entails a steady increase in values over time.

Because the source of information is defined over a long period of time, the agnostic neuracle should only be expected to accurately classify the user's mental state when a specific context matching the trial protocol in the calibration period has unfolded. If k is the length of the trials in the calibration period, the agnostic neuracle is only trained to classify the user's mental state when they are exactly k seconds deep into a mental state that transitioned from a resting state. In a real-time experimental context, this only occurs once: k seconds into the experiment. Among the five problems, the *synchronization* problem may therefore be the best

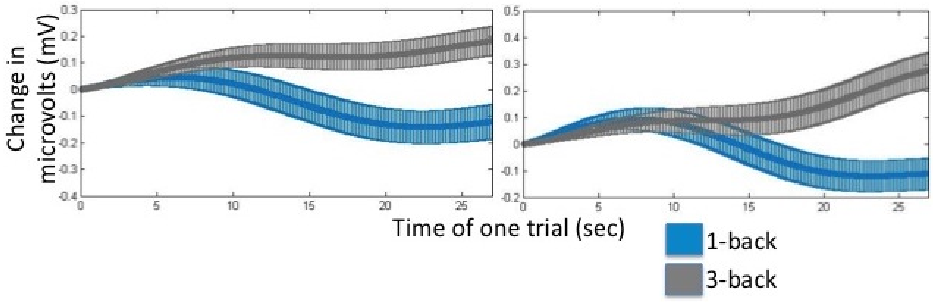

Figure 2.1: Averaged fNIRS data

explanation for why some experiments following the agnostic paradigm fail to elicit significant improvements to user performance.

2.1.3 The Entropy Problem

As stated by the *entropy thesis* of the dissertation, the fifth problem of agnostic neuracles may be the key to a solution. Because it is important to understand, this chapter focuses on the *entropy problem*, grounding it appropriately in fMRI-based neuroimaging and cognitive neuroscience, and evaluating the degree to which it affects both fNIRS and EEG.

The *entropy problem (e)* problems poses no challenge to the agnostic paradigm in Chapter One if task-dependent physiological profiles are consistent over time (*assumption e*). The following three sections evaluate if *assumption (e)* is true, and if so, what new BCIs are possible.

2.2 Entropic Dimensionalization

Entropy is a term with different but overlapping definitions in physics and information theory. An emerging neuroscience literature is beginning to use the term for the purposes of understanding how neural networks change internal configurations in response to external signals. A central contribution of this dissertation is to arrive at a definition that is useful in the design of BCI and encoded by a physical neuracle processing ongoing fNIRS data.

2.2.1 Entropy in Thermodynamics

Statistical mechanics is a subfield of modern physics concerned with characterizing physical systems with many degrees of freedom. In this community, entropy refers to the tendency for systems to assume less ordered configurations over time. Compared to a high entropy system, a well ordered system has a smaller microscopic state space needed to express the system's macroscopic properties [38].

Physicists use entropy to describe phenomena that can be observed in the natural world. For example, imagine a glass of water moments after a human has placed a soluble vitamin C tablet inside. In that moment, the water glass system has a relatively low entropy: the emergent phenomenon of a drinkable liquid with a solid tablet floating on the surface can be expressed simply as a collection of H2O compounds supporting a group of artificial compounds bound together by human effort. As the tablet dissolves and binds with the H2O compounds, the system assumes a wider and more unpredictable range of macroscopic states, demanding a more precise and larger microscopic state space to characterize its emergent configuration.

The second law of thermodynamics states that over time closed systems will move towards disorder, requiring a larger microscopic state space in order to express the emergent macroscopic behaviors of the system.

The concept of entropy may be easy to understand for computer scientists, as a good programming imperative is to continually reduce entropy: both to make the file size of the program smaller and to make it easier to read. A signature of high entropy programs is code duplication, where similar logic is repeated at different points of the program. Veteran programmers use abstraction to find functions or objects that can be reused with different variable settings, thereby reducing the entropy of the program.

2.2.2 Entropy in Information Theory

Information theory is a field concerned with the storage, quantification, and communication of information. It studies systems where there is a source of data, a receiver, and some channel between source and receiver. The basic problem of information theory is to identify a method for the receiver to deduce what information the source has generated based on the content in the channel. Entropy here refers to the average rate at which the source produces information, where information is understood as the resolution of uncertainty [110]. Information entropy is measured in bits, formalized with a mathematical equation that captures the number of yes-or-no questions needed to ascertain the observed outcome of a source (the value of a random variable):

$$H(X) = -\sum p(X) \log p(X)$$

[3]. The intuition for this equation is that the entropy of a random variable, H, is a summation of the shortest binary encoding of its possible values multiplied by their probability. Consider how this equation would act on two artificial methods to generate information: the random variable of a coin and a die. Flipping a coin can result in two possible outcomes: heads or tails. The observer of a coin flip must therefore only ask one question to know the outcome. Since the two outcomes are equiprobable, flipping a coin has 1 bit of information. Rolling a die can result in six possible outcomes, and therefore has more entropy, requiring more yes-no questions to communicate the result from the source to the receiver. Since all sides are equiprobable, the entropy of rolling a die is lg(6), or just below 2.585.

2.2.2.1 Kullback-Leibler divergence

The *KL-divergence* is a synonym for *information gain*, which provides the logic for how a decision tree prioritized features, as discussed in Chapter One. The decision tree placed features nearer the root that had a superior capacity to predict the class on the basis of its values. As with the formula for *entropy*, the intuition behind the *KL-divergence* hinges on the notion of a minimal description length in a binary encoding of symbols in a system. In order to communicate data efficiently, symbols with a higher likelihood of occurring should map to a shorter binary encoding (e.g the two most frequent symbols receive zero and one). But the system cannot know how to optimally encode its data unless it has a perfect representation for the underlying absolute and conditional probability distributions. Systems thus move towards more accurate probability density functions as they witness data. The *KL-divergence* [87] captures the gain in information (*relative entropy*) when a system moves from the known prior distribution P towards a posterior distribution Q:

$$D_{KL}(P \parallel Q) = -\sum_{x \in \mathcal{X}} P(x) \log \left(\frac{Q(x)}{P(x)} \right).$$

2.2.3 Entropy in Cognitive Science

Emerging fields of psychology recognize the utility of using constructs from statistical mechanics and information theory to explain the spatiotemporal structure of the brain. According to the *Bayesian brain hypothesis*, the basic goal of the brain is to actively and parsimoniously predict and suppress external sensory signals using the knowledge of internal models, and to update these models so that prediction error is minimized in the future [50]. The brain might be summarized as a hierarchical prediction and error correction machine [31], in which information processing proceeds bidirectionally, so that statistically informed prediction flows from the top-down and prediction error modifies internal statistics from the bottom-up.

2.2.3.1 Free Energy

A pioneer of *Bayesian models*, Karl Friston models prediction error as *free energy*, and asserts that the basic imperative of the brain is to minimize this quantity [50]. The free energy principle applies to all open biological systems that resist the second law of thermodynamics for the course of their existence. By extension, it applies not only to the brain, but to each network and cell that participates in its computation. These biological systems erect a *Markov Blanket* that shields an internal state space from the inputs arriving from the external environment [102]. Successful systems enact procedures that discover stable equilibria between the model of reality encoded in an endogenous state space and the perceptions encoded by their Markov Blanket. As especially sophisticated entropy reducers, the networks found in the brain arrive at constructs that embody the underlying *causes* of sensation, mitigating the average amount of 'surprise' they experience as they navigate an environment in which they must survive and replicate.

According to Friston, the *free energy principle* explains action as well as perception [50]. Perceptual constructs (i.e., objects in the sensorium) correspond to updates in *internal states* that resolve prediction error. Similarly, action can be modeled as entropy reduction in the external state that falls under the jurisdiction of the brain's motor effectors. For example, signals from the body may inform the brain of a need to quench thirst, and signals from the eye let the brain conjure a sensorium that affords the drinking of water. To purge these signals from the brain (and reduce entropy), the brain applies motor commands to bring the cup to the mouth, creating a new reality with lower entropy.

Figure 2.2 shows a simplified model of a human as always belonging to one of four states, the desirability of which depends on the current scenario. A rigid and mild endogenous brain indicates boredom or tiredness; a random and intense endogenous brain indicates creativity or anxiety; a rigid and mild exogenous brain indicates relaxation; and a random and intense exogenous brain indicates flow or stress. Each state has specific user affordances, tasks that suit it, controlled and incidental means to calibrate machine learning algorithms that detect it, as well as guidelines for how to transform the state into another one.

2.3 Entropy in the Agnostic Paradigm

The following experiment illustrates the degree to which entropy affects the agnostic paradigm. Unlike the previous chapter which limits its investigation to fNIRS-based BCIs, this chapter expands the space to include EEG-based BCIs as well.

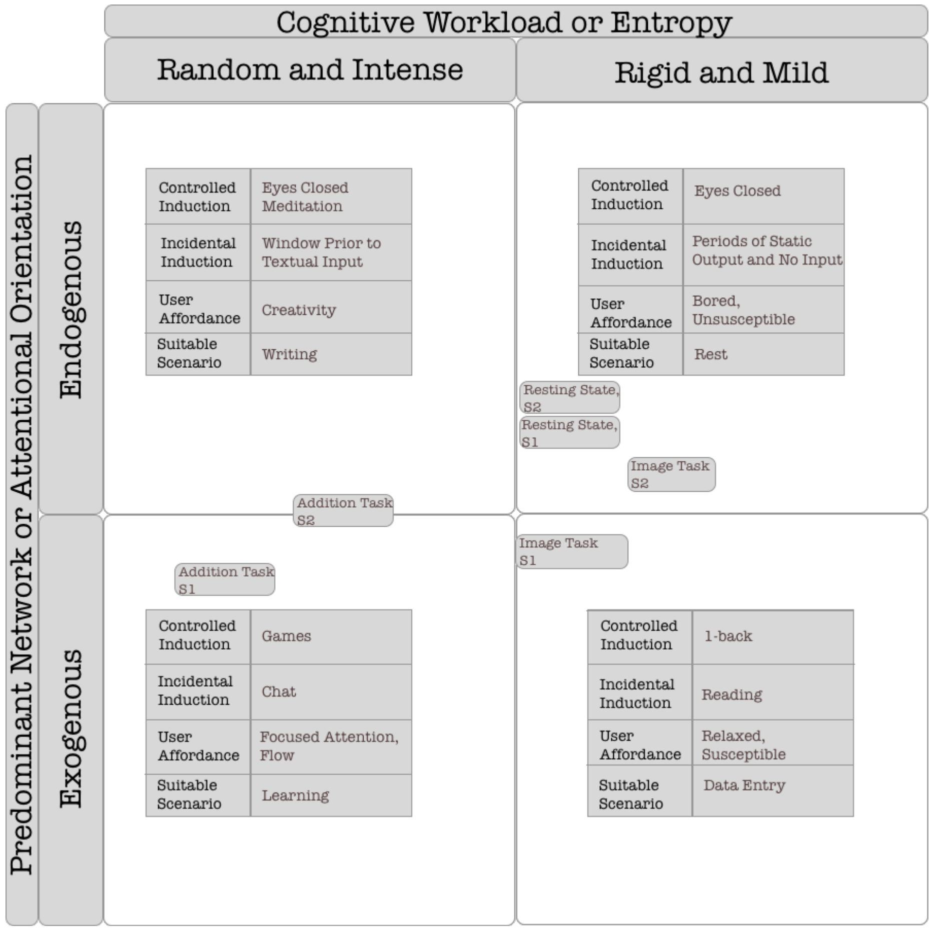

Figure 2.2: A simplification of users into four states along two dimensions. The *Cognitive Workload* or *Entropy* dimension indexes the degree of computation and surprise in the user's brain, and the *Attentional Orientation* or *Predominant Network* dimension indexes the degree to which this computation is engaged in a goal-directed loop with its sensory environment.

2.3.1 Differences between EEG and fNIRS

EEG (electroencephalography) and fNIRS have different advantages and disadvantages, complement each other quite well, and have been studied separately many times. Electroencephalography relies on a chemically-induced electrical charge that occurs at a neuron's action potential (the fundamental unit of computation in the brain). The EEG infers the aggregate of many such events by measuring voltage changes on the user's scalp. The detected voltage at the scalp oscillates according to specific patterns, which conveys information about the activity of neurons, especially near the cortex [127]. Unlike EEG, whose most studied signals oscillate at frequencies above 1 hz, fNIRS takes seconds to register changes in state. This prevents fNIRS from indexing the high frequency patterns powering the EEG signal, but opens the device to measure longer term patterns in brain activation [104].

Complementing each other's weaknesses and strengths, fNIRS and EEG seem to invite integration into a singular input device that delivers a suite of predictions about the user's cognitive, emotional, and intentional state. EEG has poor spatial resolution [17], meaning it is difficult to resolve from where a given signal is originating, but it has good temporal resolution, meaning a given measurement is temporally very close to the phenomenon it endeavors to portray. Conversely, fNIRS has good spatial resolution but poor temporal resolution [44]. Because of its good temporal resolution, EEG can capture a brief episode of mental activity and translate that activity into a command. For example, a user might imagine body movements (moving left or right hands) which would produce a machine learnable EEG signal, which then could be mapped to a deliberate command [57]. fNIRS is very difficult to use for direct input because it would necessarily take seconds for the command to register. But EEG's sensitivity to short term electrical fluctuations leaves it vulnerable to noisy inputs. For example, an eye blink and other movement also produces an electrochemical effect that drowns the neurological signal underlying state classification [17, 67]. fNIRS does not suffer as severely from noise [88, 122].

2.3.2 Hypothesis

The following experiment has been repeated four times using two portable brain sensing instruments (EEG and fNIRS), making it possible to observe what dimensions of user information processing are better left to the jurisdiction of either sensor, in the hope of specifying how an integrated fNIRS and EEG system could jointly classify the user's state. I hypothesize that fNIRS, which has good spatial resolution, is better than EEG at classifying when user's transition from resting to task states, but that EEG, which has good temporal resolution, is better than fNIRS at indexing a user's entropy or workload, since information processing occurs quickly

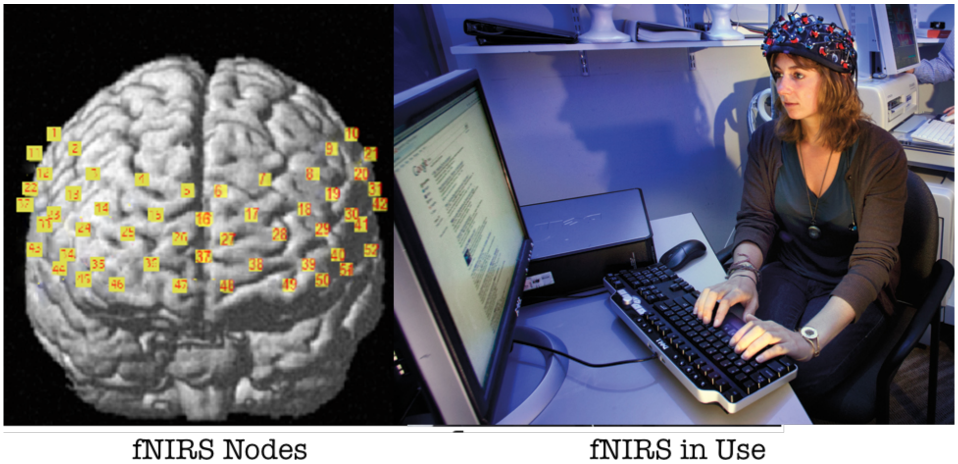

Figure 2.3: Hitachi fNIRS equipment

in the brain.

2.3.3 Equipment

The EEG used in this experiment was *Advanced Brain Monitoring's b-alert X10*, a 9-channel wireless EEG system with a linked mastoid reference, sampling at 256hz. The EEG headset was placed on users using standard 10-20 measurement set-up techniques [70]. Figure 2.4 shows the nine regions of the brain measured by each channel.

The fNIRS device used in this experiment was the *Hitachi etg-4000* fNIRS device with a sampling rate of 10Hz. The fNIRS probe (Figure 2.3) was a 3x11 probe with 17 light sources and 16 detectors, resulting in 52 locations measured on the head.

2.3.4 Method

Twenty-three subjects (8 female, 15 male) between the ages 18 and 49 participated in the experiment. Upon arrival, subjects consented to the experiment and were fitted with the fNIRS or EEG sensors. The prepackaged *b-alert* and *Hitachi* software calibrated itself to the detected connection with the user's scalp. Then, the subject alternated between 8 instances of an arithmetic task and 8 instances of an image-matching task, performing each task for 35 seconds, with 15-second controlled rest

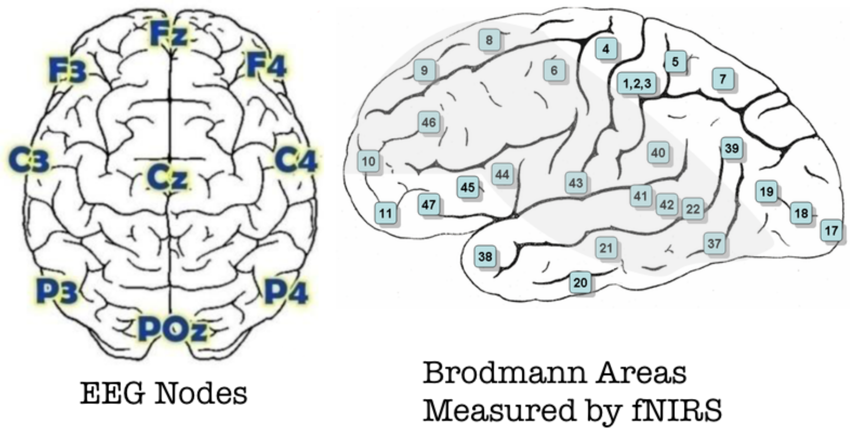

Figure 2.4: Channel locations for fNIRS and EEG. For fNIRS, only highlighted Brodmann Regions are measured

periods in between. For the image-matching task, users indicated whether sequences of images matched each other, as in an n-back [51] with n permanently set to 1, similar to the low cognitive workload condition used in previous implicit BCI work [6]. For the arithmetic task, users added two two-digit integers to each other, entering the response into a text box. Figure 2.5 shows the computer output for these tasks and Figure 2.2 plots these tasks (and comparisons between them) along the spectra of cognitive workload and attentional orientation.

I was especially interested in the transition between task and rest, and how that would change in the first and second session for both EEG and fNIRS. My interest in inter-session comparison was born out of a consideration for how machine learning algorithms might decay over time if they did not account for updates to the user's cognition. As mentioned, I was especially interested in corroborating a previous small long-term pilot [64], where I tracked my fNIRS data over the course of several weeks as I made myself an expert at the cognitive workload tasks typically used in implicit BCI (the n-back). In later sessions, I noticed that fNIRS failed to register a strong effect unless the difficulty placed me at the edge of my ability, and I hypothesized that my brain had over time generated too efficient top-down schemes for solving the task, thereby leaking less prediction error up a

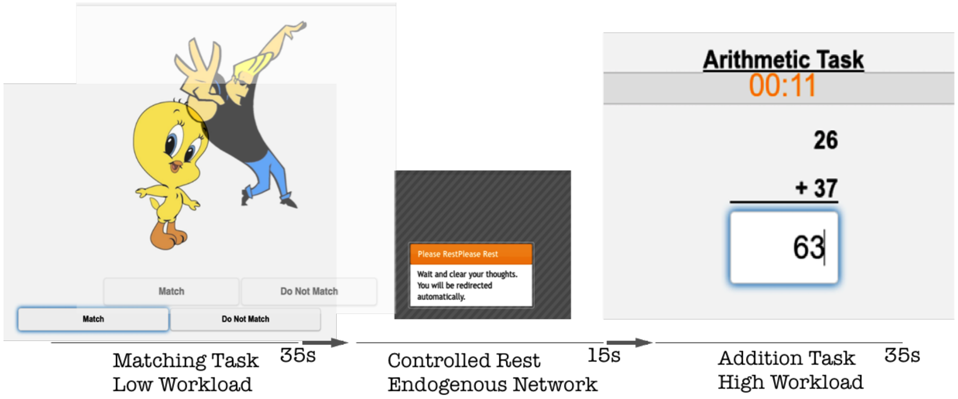

Figure 2.5: The three conditions of the experiment

cognitive hierarchy, the information theoretical construct to which I hypothesize these cortically-sensing instruments are principally sensitive. If that is the case, then machine learning algorithms operating from either EEG or fNIRS data should degrade in a second session of the experiment. The experiment was thus repeated four times in two sessions for each device on two separate days. The second session for a given participant was at most 27 days later and at least 23 days later.

For each of the 23 subjects, there were four datasets (two sessions for each device). Each dataset included 8 trials for the math task, 8 for the image task, and 16 for the resting task. I was interested in whether or not I could build an agnostic neuracle to separate the image and math task using both fNIRS and EEG data, as well as how this algorithm might change in the second session. I trained a new machine learning algorithm, for each subject for each session, and for each neuroimaging device. I tried one approach for feature design using the specifications optimized in Chapter One [128], and for ease of communication, I let this algorithm be identical for both EEG and fNIRS, leaving the critical preprocessing components to the software distributed by *Hitachi* and the *b-alert* EEG.

For fNIRS, the input to the agnostic neuracle was a matrix of 52 channels that had been converted from light intensity into oxygenation measurements according to the Beer-Lambert Law, and bandpass filtered, leaving only the components of the signal that fluctuated between 0.01 and 0.5 hz. In the analysis, I omitted de-oxygenation measurements since these values largely convey the same information as oxygenation. For EEG, the raw data was processed by *Advanced Brain Monitoring's* proprietary acquisition software, which includes artifact decontamination algorithms for eye blinks, muscle movements, and environmental/electrical interference. After decontaminating the raw data, the input was a matrix of 90 channels, consisting of the average power spectral density, averaged together into one second time periods, at each of the nine channel locations. Power spectral density was computed for the ten frequency bands of delta (1-3Hz), theta slow (3-5Hz), theta fast (5-7Hz), theta total, (3-7Hz) alpha slow (8-10Hz), alpha fast (10-13Hz), alpha total (8-13Hz), sigma (12-15Hz), beta (13-30Hz), and gamma (25-40Hz), at each of the channel locations.

For each instance, I computed the mean, linear slope, and standard deviation of the entire time-series of values for each channel. Thus, for inter-task (A1, A2) comparisons on the fNIRS data, the 350 readings x 52 channel windows became 2 condition x 8 trials x 156 feature instances and for EEG, the 35 readings x 90 channels window became 2 condition x 8 trials x 180 feature windows. For the task vs. rest comparisons (B1, B3), the same transformation occurred but the first 150 readings of the task were extracted, and compared to the 150 readings of resting data. I fed these feature sets into *Matlab's Statistics and Machine Learning Toolbox* implementation of the linear kernel support vector machine (SVM) and did not change default parameters (since the goal was to discern to what fundamental dimensions the machine learning algorithms were most sensitive and how that differed between devices, and not to maximize machine learning performance). I evaluated each machine learning separation using 10-fold cross validation [109], training the machine learning algorithm on all but an approximate tenth of the data, changing what tenth was omitted from the dataset and using that set for testing the trained classifier in ten separate tests. For all tests, evaluation instances were drawn from the same subject and session as the training instances that drove the machine learning algorithm. Next, I report on the averaged 10-fold cross validation classification accuracy for each test of interest.

Table 2.1: SVM Machine Learning Accuracies for Matching vs. Addition Comparison (A1,A2). m denotes the mean classification accuracy for all 23 subjects in 10-fold cross-validation and s refers to the standard deviation. For each row and column, classification accuracies have been compared in a *paired samples t-test*, and the *p-value* is reported.

S	**EEG**	**fNIRS**	**p**
1	m=73%, s=21%	m=72%, s=18%	0.7358
2	m = 80%, s = 29%	m = 71%, s =17%	0.0525
p	0.2430	0.8900	

2.3.5 Results

2.3.5.1 Task Comparison: A1,A2

For each session, I made two comparisons for each device, first distinguishing between the two task conditions and then separating the two tasks. Table 2.1 shows classification accuracies in cross-fold validation for the inter-task separation. In the previous section, Figure 2.5 shows how I expect these two tasks to differ from each other with respect to the workload and attentional orientation dimensions for the different sessions. The addition task presumably poses a greater burden to the user's cognitive workload than the matching task. I have compared the mean classification accuracies for each of the 23 subjects between both device and session, and Table 2.1 reports the probability that the null hypothesis is true in a *paired sample t-test*. There were no significant effects in these comparisons, but the EEG-based machine learning algorithms trended towards better performance. The results highlight that both devices can effectively parse the user along the cognitive workload dimension.

2.3.5.2 Rest versus Task Comparison: B1,B2

Table 2.2 shows an identical analysis, but for the comparison between the two tasks and rest. Since rest periods were shorter than task periods, I truncated the task trials so that they only included the first 15 seconds of data. Table 2.1 indicates that this comparison primarily determines whether or not the user has engaged an endogenous or exogenous network. For fNIRS, machine learning performance in the first session (m = 84%, s = 9%) is significantly better than machine learning

Table 2.2: SVM Machine Learning Accuracies for Task vs. Rest Comparison (B1,B2).

S	**EEG**	**fNIRS**	**p**
1	m = 79%, s = 11%	m = 84%, s = 9%	0.11
2	m = 68%, s = 17%	m = 75%, s =13%	0.11
p	0.0270*	0.0096**	

performance in the second session (m = 75%, s = 13%) (p =0.0096, N = 23). Similarly, for EEG, machine learning performance decays significantly from the first session (m = 79%, s = 12%) to the second session (m = 68%, s = 17%) (p =0.0270, N = 23).

2.3.6 Discussion

It is interesting that EEG (m = 76%) outperformed fNIRS (m = 71%) at separating the mathematical and image recognition task, which manipulates the user along the cognitive workload dimension, but that fNIRS (m = 80%) outperformed EEG (m = 74%), which manipulates whether or not the user is engaging a *task-positive* or *task-negative network*. Even though these differences are not significant, the results are consistent with the hypothesis that these two devices complement each other, covering the other's weakness. fNIRS is generally regarded as supporting better spatial resolution whereas EEG has better temporal resolution. Since every region in the brain is better described as belonging to either a task-positive or task-negative network and these two networks are anti-correlated, the fNIRS-based features (which are not as confused as EEG about the tissue they measure) might provide the information the SVM needed to discern the notion of anti-correlated networks, and robustly predict the user's state. This will be elaborated in the next chapter, as anti-correlated networks can supply the basis for a physical paradigm for BCI.

Another interesting result is that in both fNIRS and EEG experiments for the separation rest versus task, classification accuracy in session 1 reduces significantly in the second session for both fNIRS (p = 0.0096) and for EEG (p = 0.0270), but not for the separation between the two tasks, where classification accuracy is in fact better in the second session for EEG and approximately the same for fNIRS. However, this is not surprising if the brain is dimensionalized according to the Bayesian framework. Both fNIRS and EEG measure brain activation principally at the outer tips of the brain: its cortex. In a Bayesian framework, the outer tips of the brain's hardware presumably carry out computations very high in the information processing hierarchy. In the second session, subjects had already been exposed to the tasks; thus, the second session cognitive makeup likely included internal representations that solved the input-output relations dictated by the task at a more primitive point in the information processing hierarchy, reducing the prediction error and its associated corrective events to penetrate the higher level regions of the brain under interrogation by the brain sensors.

Specifically, I attribute the relatively increased difficulty of the SVM to predict transitions between task and rest in the second session to the greater difference in system entropy between the resting and task conditions in the first session than in the second session. Since prediction error (or free energy) dictates the entire operation of a Bayesian brain, this may be another way of expressing that, in the second session, the user had absorbed efficient probability distributions for the task, enabling competing endogenous resting state inputs (which draw from the same finite pool of oxygen supply) to flourish and thus produce a profile that better matched the resting state. In simpler terms, the user's brains had figured how to efficiently solve the task in the second day of the experiment, but not how to efficiently rest. With this interpretation, machine learning accuracy did not change significantly for the inter-task comparison since the user had previously engaged both tasks, making so that task-induced entropy would decrease equally in both conditions. This finding demonstrates the first part of the *entropy thesis* by showing that agnostic neuracles are highly sensitive to the novelty of a state, and may not be able to find algorithms that classify across sessions.

As stated by the *entropy thesis*, the difficulty to control for system entropy between tasks and sessions is a feature, not a bug, of brain-computer interfacing so long as it is acknowledged by the designer. By defining the user dimension of cognitive workload using Bayesian formulations of system entropy, there is an opportunity to understand oscillations in brain activity without machine learning.

Chapter 3

Anti-correlated Networks for Physical Neuracles

So far, this dissertation has stayed within the boundaries of its predecessors in *implicit BCI*, identifying advantages and disadvantages of what is referred to as the agnostic paradigm, as well as possible paths forward based on contemporary neuroscience. Chapter One

- specified the *agnostic paradigm*, conducting an offline evaluation of machine learning algorithms trained to distinguish high and low cognitive workload states in n-back data
- showed four implicit stationary BCIs obeying the agnostic paradigm.

With a specification, implementation, and evaluations of BCIs in the agnostic paradigm, Chapter Two slowed down the research, taking a small step backwards in order to pave the way for a large leap forward. Chapter Two

- identified five problems without obvious solutions in the agnostic paradigm
- gave a neuroscientific simplification of brain activity designed to better predict the spectra of mental states which neuracles can be calibrated to detect
- showed results from a two-session fNIRS and EEG experiment, giving evidence

for the capacity to portray this new dimensionalization using neuracles, as well as demonstrating how the entropy problem affects BCI.

Together, these chapters demonstrate that BCIs may be able to measure the moment-to-moment information processing burden (or entropy) the application poses on the user, and adjust internal semantic or syntactic parameters in order to sustain a flow state for the user, where the challenge is at a sweetspot set to optimally combat the brain's default prerogative to retreat into resting networks out of boredom or frustration. However, these chapters also demonstrate that realistic implementations of these BCIs are not possible without algorithmic solutions to the *five problems of agnostic neuracles.*

In the absence of algorithmic solutions to these problems, this chapter explores an orthogonal method for building neuracles in what will be referred to as the *physical paradigm for BCI.* Previous research has avoided the physical paradigm because it depends on an understanding of how the brain works at many levels of analysis. But engaging the multidisciplinary research needed to decode the raw fNIRS signal may be worth the effort since a physical neuracle suffers none of the five problems defined in Chapter Two.

A calibration period that manipulates the user's state to provide instances for a machine learning algorithm is unnecessary, since the system understands the signal by definition, which solves the *psychology, disengagement, and overfitting problems.* This period can instead be used to change the probe configuration and other parameters with known individual variation. The *entropy problem* can be avoided by targeting *entropy* as a dimension under investigation [65]; and a *dynamic segmentation algorithm*, which partitions the fNIRS signal into distinct cycles of the low frequency oscillation investigated in this paper, may overcome the *synchronization problem* (see Chapter Four). Instead, the key problem of a physical neuracle is knowing what a given signal represents. This chapter argues that contemporary neuroscience has produced enough information to decode at least one *physical dimension* in a user, describing a basic trade-off in information processing. It therefore defends the *anti-correlated network thesis*.

Anti-correlated Network Thesis: The strongest anti-correlated network in fNIRS data in between 0.01 to 0.1 hz describes user attention as it shifts between a more sensory (or exogenous) mode to a more conceptual (endogenous) mode, and can be minimally measured by one fNIRS probe by the eyebrow and one by the ear.

3.1 The Significance of Anti-correlated Networks

Progress towards physical neuracles depends on identifying user dimensions that are physically real in the user's brain [65, 116, 82, 23]. Plausible initial dimensions describe high level information processing trade-offs that are evident from the brain's neuroanatomy and network architecture, and leave aside a more specific description about the contents of these networks. This chapter's approach for identifying user dimensions that are physically real, as well as useful descriptors of the user, is inspired by the discovery of negatively correlated (anti-correlated) networks in the brain, where the activation of one network implies a simultaneous deactivation in some other spatially distinct network. The rationale is that two anti-correlated networks together define a physically real trade-off in the brain's moment-to-moment information processing. Since the spectrum of states in between the extreme bias towards one network or its inverse is physically measurable by definition, this trade-off meets the measurability requirement for a candidate output of a physical neuracle. Furthermore, if a BCI measures two distant regions that are anti-correlated, then it can leverage two independent sources of information when determining whether some change reflects a meaningful signal or noise. For these reasons, I propose calibrating physical neuracles on anti-correlated networks in the brain, and I will dedicate this chapter to answering two questions related to these anti-correlated networks.

1. What are the probe locations for the strongest anti-correlated network in fNIRS data?

2. What trade-off in mental state does this signal describe about the user?

Contemporary neuroscience suggests that the strongest anti-correlated network in the brain is between the task-positive and resting state networks [106]. These two networks differ primarily on whether the network's input has exogenous (from the environment) or endogenous (from the brain's memory) origin and whether its output serves an immediate or delayed function. The task-positive networks (TPN) show an increase in activation when the subject performs a task and must implement functions that interpret live sensory (*exogenous*) signals and set appropriate motor responses, e.g. maneuvering a car in response to changing audio/visual/tactile traffic information. Conversely, the anti-correlated resting networks consume more energy during conditions in which the environment poses no immediate demands on cognition. Spatially, the resting networks are more separated from the sensing organs, and operate primarily on *endogenous* memory. Activity in the most studied and energy-intensive resting network, the Default Mode Network (DMN), appears to coincide with the mental sensation of an automatic mind wandering stream that may not be relevant to the current situation [106].

While trying to follow this introduction's multidisciplinary BCI threads, some readers may notice their DMN periodically interrupting their attention to the dissertation with brief mind-wandering episodes. A brain-adaptive version of this dissertation may attempt to re-engage the user's attention in these moments by making a subtle change to the text's font or otherwise interrupt the reader with helpful definitions of some of the dissertation's important concepts like *anti-correlation*, *default mode network*, and agnostic neuracle. Knowledge about the user's anti-correlated networks can also be leveraged to their delayed benefit. For example, if the central topics of this dissertation were expanded in an online brain-augmented virtual classroom, the brain's regression into the DMN while trying to understand the concept of the DMN could be acknowledged and saved in a database. When determining what content to revisit in preparation for the final exam, the virtual classroom could spend additional time on that content which had caused the brain to retreat into the DMN. The virtual teacher may then also consider rewriting that material to be more engaging in the next iteration of the class, providing a similar interface to the *cognitive heatmap* in the beginning of Chapter One.

The cyclical activity in these anti-correlated networks may offer a real-time classification of the canonical high quality attentional state known as *flow* [33]. It is possible that the flow state may manifest as a temporary suspension to periodic oscillation between major anti-correlated networks, suggesting an opportunity to guide the user towards improved attentional states. This is the aim of the bidirectional BCI that will be introduced in Chapters Six and Seven, when I have specified the ingredients for a *physical paradigm for BCI*. In one example of a *bidirectional BCI*, the user may be programming while listening to an instrumental *brain-adaptive song* on repeat in fNIRS-augmented headphones. The low information version of the song repeats a simple but entrancing melody from a piano, played gently. In this version, the song does not intend to command the brain's attention, providing beautiful sounds that neutralize other auditory distractions. But when the BCI detects a build-up in the endogenous default mode network, the song could morph into its high information version, adding a progression of violin chords that harmonizes the piano's melody. If the violin sequence has the right amount of brain relative novel information, the brain may prioritize processing this new exogenous auditory signal over the endogenous signals of the default mode network, preventing what may have amounted to a prolonged period of malproductive mind wandering.

In order to create the physical neuracle that enables a brain-adaptive interruption engine that aborts task-irrelevant rumination, the virtual classroom that intervenes and remembers when its students zone out, a meditation assistant that expedites long term attentional training, and the brain-adaptive music that changes its audio content during moments of high endogenous processing, as well as several other applications not yet imagined, this chapter documents a search for anti-correlated networks in the brain. I believe this anti-correlated network defines a routine back-and-forth between exogenous and endogenous modes of processing in the brain. This low frequency spatially defined signal may elude the EEG, suggesting more powerful BCIs that fuse EEG and fNIRS.

The purpose is thus to determine the probe configuration for an algorithm that determines the task-positive versus default mode network cycle in a physical neuracle. Any measured region shows more or less activity during a task or in rest, but some are more decisively task-positive. Because the correlation between fMRI data and fNIRS data is only about 0.25 [34], the optimal probe location must be determined on the basis of fNIRS data. I therefore ran a 4-session, 50-subject fNIRS study, and analyzed the data to cluster the fNIRS channels into three categories: those which increase during rest; those which increase during task; and those which don't vary in response to rest or task. I categorize nodes from fNIRS data collected on two sessions of an experiment on one day and evaluate the nodes chosen on two follow-up sessions, approximately two weeks later. This chapter makes multiple contributions towards the development of physical neuracles from fNIRS data:

1. *The background section* summarizes existing HCI, BCI and neuroscience literature, leading to a framework for how to detect a user's changing degree of endogenous and exogenous processing on the basis of anti-correlated networks.

2. *The results section* describes the methods and results of a 50-person, 4-session, 52-channel fNIRS experiment aimed to manipulate the user's attentional state along these dimensions. The data analysis focuses on identifying a pair of probe locations with high anti-correlation to each other, whose values fluctuate depending on whether the subject is performing a task or resting.

3. *The discussion section* specifies a *physical neuracle* based on anti-correlated networks which fuse EEG and fNIRS data and can be implemented in the open-source *Neuracle* software distributed in Chapter Five.

3.2 Background

3.2.1 Distraction by the Default Mode Network

Contemporary neuroscience finds that the brain at rest typically does not rest [106]. Instead, it initiates resting networks which reflect, plan, and fantasize – a phenomenon corroborated in an experiment where one hundred percent of participants who were instructed to not think for a 20-25 second period reported in fact having thoughts during that period [47]. To support these functions, the brain consumes only fractionally ($\sim 5\%$) more energy when in the task-positive network (TPN) [106]. This finding is explained in terms of a computationally expensive default mode network (DMN) wired to ruminate when these processes do not interfere with higher priority environment demands. Conveniently for physical neuracles, the regions supporting the DMN are similar across subjects. Resting networks often engage the ventromedial prefrontal cortex (vmPFC at Brodmann (B.) area 10), which were measured in this experiment. Capturing activity in this region may therefore supplement existing techniques for quantifying user engagement using physiological sensors such as EEG [18], eye-tracking [105] and GSR [89, 41] (see [42] for an overview).

Although mind wandering may serve a creative function [80], the time a human spends in the DMN correlates negatively with both productivity and satisfaction. In a vigilance task, increased activity in regions of the DMN preceded worse reaction time [131], and in another experiment, DMN activity predicted task error up to 30 seconds into the future [40]. In a smartphone survey on 3000 individuals, Killingsworth (2010) estimated that, on average, brains spend roughly as much time indulging either network [77], but time spent in the default mode network is considered less enjoyable and also predicts future dissatisfaction. Other research implicates smartphone and internet usage as culprits in an ongoing attention epidemic [30, 112, 21, 79]. Before the rise of the Internet, obtaining and spreading information required physical effort. However, information is now acquired and disseminated virtually. What used to require planning, working memory, and patience is now accomplished automatically in a digital instant, freeing DMNs to meander. What is more, the information age's powerful social stimuli (e.g. gossip), strong feedback loops (e.g. video games), positivity biases (e.g. Facebook news feed), and continuous presence (e.g. a smartphone) may have disturbed the natural rhythms of attention, elevating the amount of stimulation necessary to obtain flowing exogenous attentive states where attention is happily wedded to sensory-motor mappings. Instead, some users may feel uncertain about what to do, vulnerable to interruption, and attentionally fragmented.

HCI research should carefully consider the role a given technology plays in fanning or fueling the attentional epidemic over the course of its sustained use. If negligently designed, the BCI proposed in this dissertation will maximize user experience and attention in the short term, but by hyper-fitting digital inputs to induce peak states, the user may become dependent on the technology and less satisfied in the long term. This problem is taken seriously in Buddhist literature, and I therefore briefly summarize ongoing research into meditation, which may be regarded as a millennial long empirical investigation into the development of a practice that improves attention.

3.2.2 Improving Attention in the Long Term

In parallel to the attention epidemic, meditation has emerged as a popular practice for improving the quality of attention, especially in technologically advanced Western hubs such as Silicon Valley. A growing *Neural Buddhist* literature argues that training the ability to regulate attention can improve cognitive functions, as well as mitigate psychiatric disorders [9, 60]. A sustained meditation practice correlates with improving schizophrenia [29], depression [126], anxiety [55], ADHD [119], emotional regulation [86], memory [134], self regulation [125], and self-awareness [69]. The data from neuroimaging scans suggest that meditation reduces competition (or anti-correlation) between the task-positive network and default mode network [85, 62, 63]. This finding is difficult to parse, but worth footnoting in the design of technology that seeks to improve the quality of its user attention.

Those with cultivated skills of introspection emphasize the distinction between self (endogenous processing) and other (exogenous processing) when describing the meditation exercises that led to improved experience [9]. For example, in *Focused Attention* meditation, the meditator rests attention on external stimuli (e.g. breath or sound), and attempts to reduce competitive influence from internal distractions. Compared to novice meditators, expert meditators show less activity in key regions of the DMN and more activity in areas related to top-down attentional control [22]. This literature shows that meditation modifies brain activity by changing the interplay between endogenous and exogenous networks [118]. A pair of signals describing the rhythmic oscillation between these networks may thus source the key neural parameters for long-term optimization in a BCI. In this chapter, I argue that this rhythm is a sufficiently global pulse of brain activation that fNIRS can monitor its real-time fluctuations. In *Open Monitoring* (OM), the meditator allows internal rumination to occur but treats these thoughts non-judgmentally as neutral data-points. Compared to novice meditators, expert meditators have a resting state that more closely matches their state when they perform OM meditation [90]. Finally, in *Nondual Awareness* (NDA) meditation, the meditator is instructed to be equally aware of external and internal processes, allowing experience to come and go of its own accord. Compared to FA meditation, NDA meditation entails reduced anti-correlation between exogenous and endogenous systems, as measured by BOLD activity in the associated networks [74]. We have investigated how to study meditation with fNIRS in previous research [66].

3.2.3 Classifying TPN versus DMN Activity using fNIRS

Spontaneous hemodynamic fluctuations in fNIRS data associated with default-mode, task-positive, and frontoparietal networks have been investigated previously in neuroscience literature [113, 95]. In 2013, Harrivel et al. evaluated fNIRS correlates of the task-positive and task-negative networks. In this experiment, five participants wore fNIRS as they completed the multi-source interference task (MSIT), which requires the participant to suppress distraction. This experiment identified two pairs of anti-correlated networks, the medial frontal gyrus (which increased activation during rest) and the dorsolateral prefrontal cortex (dlPFC, which increased activation during task) [61].

Harrivel et al's results [61] have inspired the present investigation. However, I evaluate different tasks in the present experiment, measure more regions, apply different statistical methods, and run the experiment on 50 subjects in four sessions. These modifications allow us to approach the search for anti-correlated networks more naively, unbeholden to prior knowledge from fMRI literature. In Harrivel et al's experiment, nodes were placed on key regions of the task-positive and task-negative networks as known in the fMRI literature, but fNIRS and fMRI measurements are not entirely correlated (approximately 0.25 on average [34]). Compared to fMRI, fNIRS has a finer temporal resolution, but an inferior spatial resolution, and it measures both oxygenated (oxy) and deoxygenated species of hemoglobin. For these reasons, I measure 52 channels across the user's cortex, as I attempt to form clusters of task-positive and resting state regions among these channels. I repeat the experiment four times on 50 subjects, using the first two sessions to select the channels and the last two sessions to evaluate that selection.

The aim of this research paradigm is to produce a *physical neuracle* which, given fNIRS input, outputs a real-time measure of some *physical trade-off* in the user's moment-to-moment information processing. This section provides results suggesting that a *physical neuracle* tuned to the orientation of user attention is possible in principle. A complete validation of the utility and accuracy will await the implementation of BCI that uses this *physical neuracle* to the user's measurable benefit. In preparation of future experiments, the output of the present analysis are settings to variables in the hypothetical *physical neuracle*, especially filter techniques and probe configuration.

The goal is to evaluate the hypothesis that fNIRS measures are sensitive to the orientation of attention, and to provide a clear set of recommendations in terms of where probes need to be placed. The aim of this search is two-fold. First, I expect greater clarity for how to design an fNIRS-based BCI if the signal can be plausibly mapped onto a well-studied neurobiological phenomenon (i.e., the anti-correlated task-positive and default-mode networks found in neuroscience literature) that survives all levels of scientific analysis: i.e. it has a known and well-delineated neurobiological anchoring, it provides a spectrum that remains relevant for any state in question, and it also signifies sufficient meaning about the user, making it possible for a designer to reason about how to map changes in the user's neurobiological state to system adaptations. Second, if an adaptive system measures regions that are anti-correlated, then it can leverage two independent sources of information when isolating the meaning of a signal. For example, suppose a real-time fNIRS system witnesses an increase in the light intensity quantities it measures at probe A. In order to decide whether that increase reflects noise (e.g. the probe is dislocated) or a known signal, it can verify that probe B shows a concomitant decrease in values, provided probe A and probe B are known to be anti-correlated, fluctuating in response to the user's external versus internal attentional orientation.

3.3 Method

3.3.1 Equipment

The experiment measured brain activation using a *Hitachi ETG-4000*, which cost roughly $370K, but contains no parts which cannot be built more cheaply, making it a good instrument for deciding the specifications for a cheaper device. This 52-channel continuous wave (CW) system is depicted in Figure 1. There are a variety of probe options available for the *ETG-4000* that can span several regions of the brain cortex (depending on the areas that researchers want to measure). This experiment uses a 3 (row) x 11 (column) probe with 17 light sources, and 16 light detectors. The light sources and detectors are spaced in such a way as to allow the measurement of 52 unique locations (also referred to as 'channels') across the brain cortex. Each light source produces near infrared light at two wavelengths (690nm and 830nm), which are pulsed intermittently in time. This results in 52 channels x 2 wavelengths = 104 readings at each time point, and the device collects 2 samples per second.

3.3.2 Experiment Protocol

51 subjects (37 male) participated in the experiment, with one participant omitted from analysis due to data loss. Subjects were college students (aged 18-22) representing a wide range of majors. Informed consent was obtained, and participants were compensated for their time. All subjects filled out a pre-survey, which solicited demographic and health information. During the experiment, the subject alternated between three tasks, which manipulated the orientation of their attention. Each task lasted 20 seconds and was repeated a total of four times in one session. Each task was separated by ten seconds of uncontrolled and unanalyzed time. Each subject completed a total of four sessions on two separate days. The first and second sessions were completed on the first experiment day and the third and fourth sessions were completed on a second day approximately two weeks later. In the search for anti-correlated networks, the first two sessions were used for channel selection and the second two sessions were used to provide an independent group to evaluate that selection. A timed *PowerPoint* presentation displayed the task instruction, alternating between three conditions in the experiment:

1. **Addition Task (A)** inducing *Exogenous TPN Activity.* This section began with a slide instructing participants to 'Start with x, where 'x' was a small number, such as 5. Each addition slide was displayed for 2 seconds, and participants were told to keep a running sum of the numbers as they appeared. Next, new slides appeared with instructions such as 'add 6' or 'add 9'. The last slide of the addition section instructed the participant to tell the experimenter the sum calculated. This task elicited exogenous goal-oriented cognition with limited confounding movement.

2. **Phone Number Task (P)** inducing *Endogenous TPN Activity.* The participant was instructed to sub-vocally rehearse their phone number repeatedly for the duration of the slide. This task did not require the user to process any exogenous inputs or produce outputs, but nonetheless required the user to deliberately engage memory and suppress the DMN.

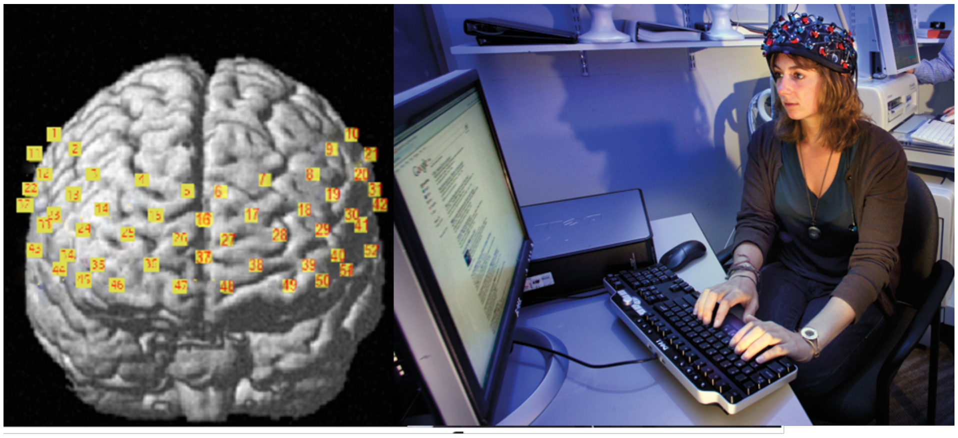

Figure 3.1: Hitachi ETG-4000 in use

3. **Controlled Rest (R)** inducing *Endogenous DMN Activity.* Participants were told to relax and clear their minds during this task. This non-task created the conditions for subjects to engage their DMN.

All sessions followed an identical sequence of tasks: *P, A, R, A, P, R, P, A, R, A, P, R* (see figures 3.3 and 3.3). Before beginning the experiment, subjects were informed about the tasks, invited to ask questions, and told to remain still and silent throughout the session, unless otherwise instructed. Participants were misled into believing the tasks would appear in random order so that they wouldn't anticipate the next task. Finally, the fNIRS cap was placed on the subject and the *PowerPoint* began.

3.3.3 Preprocessing

I was interested in finding neural correlates of task-positive and resting-state regions in fNIRS, and therefore preprocessed the data to provide the cleanest description of brain activity. This entailed first extracting measurements of oxygenated and deoxygenated hemoglobin from the raw signal of light intensity according to the *Modified Beer Lambert Law* [36], and bandpass filtering the signal to include only frequencies between 0.01 and 0.1 hz., a step that removes the high frequency components of the signal stemming from respiratory artifacts, heart rate, and motion, as well as the low frequency components stemming from brain oscillations and other

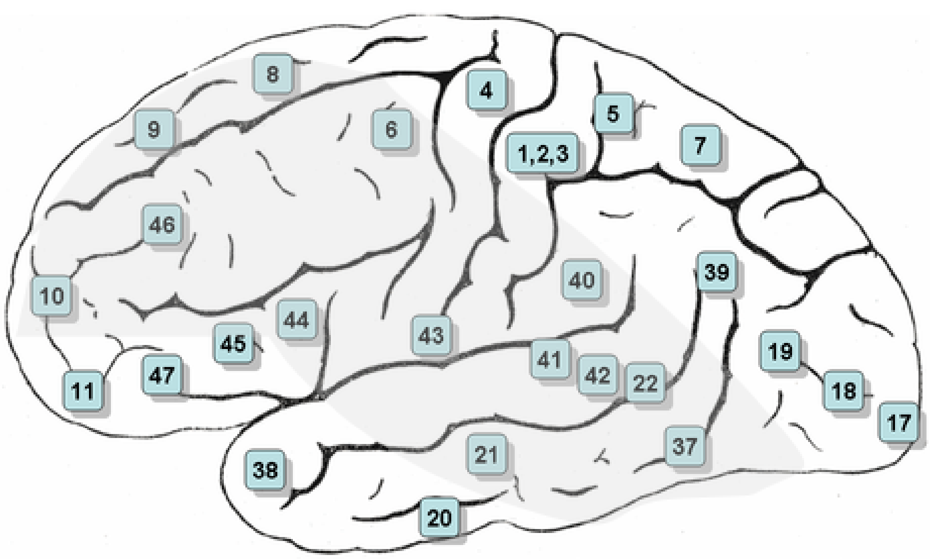

Figure 3.2: Brodmann areas interrogated by the ETG-4000

unknown slow drifts. These preprocessing steps were both accomplished by the *Hitachi ETG-4000* software package.

In addition, I also applied a makeshift adaptive filter [135]. The rationale of this filter is that any given series of fNIRS measurements fluctuates not only in response to the local brain tissue under investigation, but also to changes in blood flow and oxygenation properties in the system at large [78]. With an independent measure of the channel containing exclusively noise, the frequencies present in the noise channel can be removed from the deep channel containing both skin and brain activity. In the absence of a third-party measure for this experiment, we averaged together all 52 channels and treated this average signal as the noise channel, and subtracted its frequencies from each channel's data using a *least means squared adaptive filter*. This strategy appears effective at exposing a signal that can be otherwise drowned by system-wide changes. (I note further that this adaptive filtering style is easily portable to a real-time context, and may be more effective if the regressor channel is adjacent to the source.)

3.3.4 Results

3.3.5 Machine Learning for **Agnostic Neuracles**

In order to gauge the effectiveness of an agnostic neuracle, I first evaluated the capacity to separate the addition task (A) from the controlled rest period (R) on the basis of a machine learning algorithm trained on oxygenated fNIRS data. I tried only one algorithm with effective settings, shown in Chapter One and published in [128], and evaluated it with leave-one-session-out analysis. I created a support vector machine (SVM) trained to separate resting conditions from task conditions on the mean and slope of each fNIRS channel on each session for all subjects except one, repeating this analysis once for each session. The algorithm was thus trained independently on 200 sessions 200 separate times, where each session to be evaluated included 4 instances of each condition (not included in the training set), and thus 8 total predictions. The mean accuracy across all 50 subjects was 70%, standard deviation = 14%, ranging from a minimum accuracy of 38% and maximum accuracy of 97%.

The SVM's 70% mean accuracy reflects the capability to distinguish between resting networks and task-positive networks using agnostic neuracles in a scenario where the machine learning algorithm can access data recorded from previous sessions for that user. Although the underlying machine learning can likely be improved, it should be noted that the classification accuracy reflects a sort of upper limit to the agnostic neuracles' likely success in a real-time context given the fact that *overfitting* and *synchronization* problems discussed in *Chapter One, Section 2* have not been addressed. This motivates a search for anti-correlated networks that may drive a physical neuracle.

3.3.6 Anti-Correlated Networks for **Physical Neuracles**

The first two sessions of the experiment were used to bin channels into two categories: *Task+* and *Rest+*. To create these two channel-clusters, I first computed the mean linear slope for each of the four trials of the resting task for each subject

Table 3.1: *Rest+ group.* Channels which show a statistically significant increase in oxygenation during rest trials for the selection sessions. Statistical significance measures here and in the table below are conservative, so that $p < 0.001$ means the channel was significant at this threshold for both session 1 and 2.

Ch.	**MNI**	**p**	**Brodmann**
18	-51,36,19	<0.001	46 - *dlPFC*
28	-51,45,-6	<0.001	47 - *dlPFC*
37	-22,57,32	<0.01	9 - *dlPFC*
36	2,60,32	<0.01	**10 - vmPFC**
38	-42,42,32	<0.01	46 - *dlPFC*
39	-55,17,31	<0.01	9 - *dlPFC*
49	-44,52,6	<0.05	10 - vmPFC
47	2,68,8	<0.05	10 - vmPFC
48	-24,68,9	<0.05	10 - vmPFC

Table 3.2: *Task+ group.* Channels which show a statistically significant increase in oxygenation during addition trials for the selection sessions.

Ch.	**MNI**	**Brodmann**	
51	-64,-2,6	<0.0001	**22 - Temporal Lobe**
31	67,-35,30	<0.001	40 - Wernicke's area
52	-69,-27,1	<0.01	21 - Temporal Lobe
50	-57,28,7	<0.01	45 - Broca's Area
43	66,-4,5	<0.05	22 - Temporal Lobe
21	69,-13,-10	<0.05	21 - Temporal Lobe

and for each channel in both session 1 and session 2. I computed the same statistic for the addition trials, and evaluated the statistical significance of each channel in a *Student's t-test* that compared the slope of the addition and resting trials. I thus conducted a t-test ($n = 50$ subjects) 52 times (once for each channel) twice (once for each session). To meet the criteria necessary for inclusion in either anti-correlated network subset, the channel needed to be statistically significant for both session 1 and 2. To be included in the *Rest+ channel set*, the slope of oxygenation values needed to be statistically significantly higher in the resting condition, and vice versa for a channel's inclusion in the *Task+ channel set*. All Brodmann and MNI coordinates for the *Rest+* and *Task+* channels are shown in tables 3.1 and 3.2, where the channels are ranked in order of statistical significance.

Next, I normalized the values so that each channel was set to fluctuate between 0 and 1, and then plotted the two channels on an area chart, side by side,

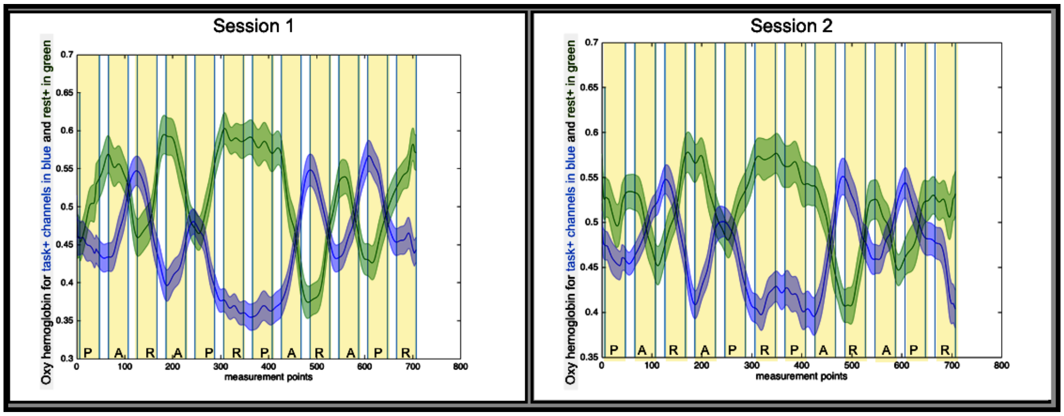

Figure 3.3: Selection channels: The time-course of two anti-correlated channel clusters in fNIRS data. The y axis shows change in oxygenated hemoglobin and the x axis shows change in time. The data has been bandpass filtered (between 0.01 and 0.1 hz) and adaptively filtered. It is sampled at 0.5 hz, and so there are 350 seconds = 6 minutes of data. The green channel shows the Rest+ group (Table 3.1) and the blue channel shows the Task+ group (Table 3.2). The thickness of the chart represents the standard error of data at that time-stamp.

where the middle of the area represents the channel's mean value, and the thickness represents the standard error of values (this is a standard way to represent fNIRS data as overlap tends to indicate high subject variation). Figure 3.3 plots this visualization for session 1 and session 2. The anti-correlations demonstrated in these visualizations suffer a selection bias given that statistics from these datasets were used to select channels. Figure 3.4 plots this visualization for session 3 and 4. The anti-correlation exhibited in these networks represents a statistically valid effect since previous sessions two weeks earlier were used to select the channels. Therefore, this selection likely persists in future experiments.

The mean *Pearson correlation coefficient* between the averaged oxygenation over the course of the six minute time-series of *Rest+ channels* and *Task+* channels was -0.8416 for session 1, -0.8280 for session 2, -0.6179 for session 3, and -0.6279 for session 4. Table 3.3.6 shows anti-correlation between these two between-subject averaged channel sets (*mean-a*), and lists them together with the mean anti-correlation of *Rest+* and *Task+* channel-sets for each subject (*mean-b*). This latter number (*mean-b*) is the more valid estimate of the likely effectiveness of a *physical neuracle* calibrated to detect whatever trade-off in the user is defined by this pair of anti-

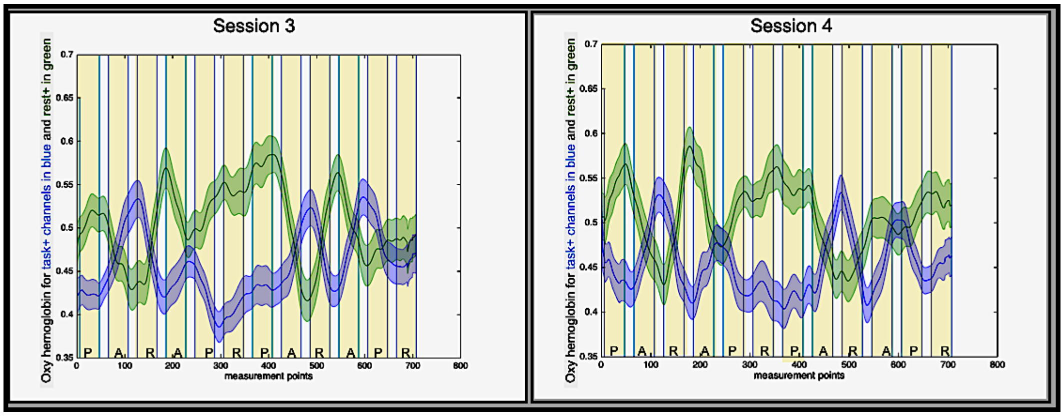

Figure 3.4: Evaluation channels: A statistically valid anti-correlation effect between Rest+ and Task+ channel sets. As with figure 3.3, *P* means *phone number* trial, *A* means *addition* trial, *R* means *resting* trial. All trials are 20 seconds long.

correlated networks because averaging across users cancels noise. It is promising that for some subjects, the mean anti-correlation was as high as -0.9, but discouraging that for some subjects, these channel sets were positively correlated to each other. For these subjects, the positive correlation between these channel sets may be due to that session's failure to arrange the fNIRS probes to interrogate brain activation as opposed to system artifacts (which should all be positively correlated). Future experiments might mitigate this effect by using a live measure of anti-correlation to determine whether or not probe setup needed to be changed at the beginning of the experiment.

These results suggest generic probe locations likely to portray this anti-correlated network for most users. As illustrated in figures 3.4 and 3.3, the anti-correlation effect is strong and persists to new sessions, showing a physical trade-off in the cerebral work accomplished by the regions listed in table 3.1 (*Rest+*) versus table 3.2 (*Task+*), and resolving *question 1* in the introduction. A physical neuracle may therefore simply measure which channel cluster has higher oxygenation consumption to classify the user's state. For reasons discussed in the next section, I recommend designating *ch.36* (measuring the *right hemisphere vmPFC at B.10 (MNI: 2,60,32)*) as default *Rest+* channel, and *ch.51* measuring the *left hemisphere temporal lobe at B.22 (MNI: -64,-2,6)* as the default *Task+* channel.

Table 3.3: The global mean (*mean-a*) and subject mean (*mean-b*), min, and standard deviation of Pearson correlation coefficients between *Rest+* and *Task+* for each subject for the four sessions.

s	**mean-a**	**mean-b**	**max**	**min**	**std**
1	-0.84	-0.47	-0.92	0.78	0.43
2	-0.83	-0.54	-0.94	0.42	0.32
3	-0.62	-0.49	-0.95	0.80	0.38
4	-0.63	-0.44	-0.93	0.72	0.40

3.4 Discussion

The results suggest the feasibility of a *physical neuracle* that detects shifts in the orientation of a user's attention on the basis of anti-correlated networks. Notably, when all data are averaged for all 50 subjects, the data show that these two sets of channels are strongly anti-correlated (∼-0.73 on average without any downsampling), and that the signal oscillates at a rate roughly equal to the task length (in this case, 20 seconds). The mean anti-correlation for each subject was weaker (∼-0.49 on average), but it is worth emphasizing that perfect -1.0 anti-correlation may not be necessary for an algorithm that perfectly portrays shifts in the orientation of user attention. In fact, perfect anti-correlated entanglement between two channels would suggest that all information about one probe is contained in the other, rendering one redundant.

Engineers may question the utility of monitoring whether or not the user is at rest or performing a task, which may be known about the user by simpler means, e.g. (keyboard input, eye-tracking, or a camera). But variations in these networks are not entirely correlated with the performance of a task, as validated by the failure to build task vs. rest agnostic neuracles with perfect classification accuracy. The anti-correlated networks describe a pulse in the brain, which can be manipulated by changing the environment's demands and sensory information load. But barring an extreme intervention (e.g. closing eyes), the pulse also follows its own spontaneous prerogatives, and contains information not yet clear to science. A third and hitherto neglected condition may help understanding what these networks measure, shedding light on *question 2*.

In addition to the *exogenous* and *goal-oriented* addition condition used to determine *Task+* probe configuration, this experiment also included an *endogenous* and *goal oriented* condition, where the participant sub-vocally rehearsed their phone number (P). In figures 3.3 and 3.4, there is a sharp change in oxygenation values only when one of the *endogenous* conditions follows or is followed by the *exogenous* condition. In other words, when either the *rest* (R) or *phone-number task* (P) is preceded by the other *endogenous* task, oxygenation values in both channel sets remain constant on average, and the two anti-correlated networks do not shift. Pictorially in figures 3.3 and 3.4, this is represented by an open mouth with the green and blue lines as lips in the middle of the experiment. This suggests that the chosen channel sets are specifically tuned towards the origin of the data being computed on: *endogenous* vs *exogenous*. In a live HCI scenario, this anti-correlated network may therefore supply the foundation for a *neuracle* that predicts the moment a user zones out of a task seconds prior to the moment when the user zones out.

3.4.1 Probe Configuration for Detecting TPN vs DMN

The fNIRS correlates referenced in Tables and 3.1 and 3.2 are mostly consistent with a previous investigation of task-dependent anti-correlation in fNIRS data, and the disagreements are worth analyzing. Harrivel et al. (2013) investigated anti-correlation between medial frontal gyrus (MFG) as a rest-positive node and dorsolateral prefrontal cortex (dlPFC) as task-positive node [61]. In that experiment, sites were referenced according to the *International 10-20 system*, typically used in EEG experiments. The MFG node in that experiment measured FPz and FP2, which closely maps to Brodmann region 10 (B.10). This experiment affirms the choice of MFG as a DMN location given that channels 36, 49, 47, and 48 were all binned into the *Rest+* channel set. The dlPFC node in that experiment interrogated site F4, which maps onto B.46.

In our experiment, no dlPFC node was placed in the *Task+* group. In fact, two channels measuring B.46 were placed in the *Rest+* group along with two further channels also measuring the dlPFC. This disagreement likely depends on the different task used in the present experiment and the previous one, and suggests the exclusion of the dlPFC from both a generic *Task+* set (as suggested by [61]) and from the *Rest+* set (as suggested by our experiment). A generic set of *exogenous* regions should instead limit itself to regions with a more fundamental role in sensory processing, such as the temporal lobe at B.21 (see table 3.2). The fNIRS measurements of the dlPFC may instead drive an orthogonal *physical dimension*, e.g. describing the relative distribution of *deliberate* and *top-down* vs. *automatic* and *bottom-up* processing.

No matter the interpretation, this disagreement suggests the exclusion of the dlPFC from both a generic *Task+* set, as suggested by Harrivel, and from the *Rest+* set, as suggested by our experiment. Moving forward with physical neuracles, it may be useful to disentangle the *endogenous vs exogenous* dimension from its parent *task vs rest* dimension. A user can compute on exogenous sensory data in a state of rest, and a user can engage endogenous memory stores in the service of a task. A generic set of *exogenous* regions should exclude regions with apparent involvement in both *endogenous* and *exogenous* involvement such as the dlPFC, and limit itself to regions with a more fundamental role in sensory processing, such as the temporal lobe locations at channels 21, 43, 50, and 52. The anti-correlated *endogenous* channel set should also skip the dlPFC, electing instead to measure channels 36, 47, 48, and 49 at Brodmann region 10, which corresponds to the vmPFC of the default mode network, validated by Harrivel's experiment, and fMRI literature.

The fNIRS measurements of the dlPFC may instead drive an orthogonal *physical dimension*, describing the relative distribution of *top-down* vs. *bottom-up* processing. In a Bayesian framework, *top-down* networks predict data driven from the *bottom-up*. In cases when surprising data dominates the brain's predictions and controls its response, the user is in more of a bottom-up state, and in cases where information chaos is held at bay by an experienced *top-down* model, the user is in more of a *top-down* state. The two *endogenous* conditions (phone-number vs. controlled rest) of this experiment differ along this dimension because the continuous (*top-down*) rehearsal of a sequence of well-known numbers limits the possibility for bottom-up memory data to penetrate the system and coerce interest from the default mode network. On the other hand, the absence of the sub-vocal rehearsal procedure in the controlled rest period invites mind wandering from the default mode network, which appears to involve involuntary bottom-up processing on *endogenous* data. However, the experimental data suggests that this finding may be due to individual variation, given the inclusion of dlPFC as *Rest+* node. Knowing that more tasks were about to follow, many participants likely did not engage the DMN during the resting period. Instead, they likely performed an active *top-down* executive monitoring with their dlPFC.

The physical neuracle in the next chapter has not yet been implemented and tested. As is clear from the pseudocode, the algorithm avoids machine learning at *network, segmentation* and *classification* levels, thereby bypassing all issues referenced in *Chapter Two, section 1.* Instead, the BCI saves its *machine learning card* to solve the problem of updating system properties and output at *design* and *application* levels in response to ground truths about the user's state known from the physical neuracle. The associated *bidirectional BCI* then uses machine learning to transform the user's measured physical dimensions into more desirable dimensions by changing some property in the interaction.

Chapter 4

Specification of a **Physical Paradigm for BCI**

This chapter aims to show how *physical, mental, neuracle,* and *interface* levels of BCIs can be bound in a cohesive Bayesian information processing framework, as promised by the *unity thesis.* It gives an algorithmic outline for a real-time physical neuracle calibrated on cyclical activity in anti-correlated networks in fNIRS data. The algorithm is based on neuroimaging literature in the first two chapters, the discovery of anti-correlated networks in Chapter Three, and the introspective studies in Chapter Seven. As shown in figure 4.1, the aim of the algorithm is to find a direct mathematical mapping between anti-correlated networks in fNIRS data and introspectively real dimensions of consciousness.

In this algorithm, I assume a BCI has three fNIRS probes interrogating brain activity at the dorsolateral prefrontal cortex (*dlPFC*), ventromedial prefrontal cortex (*vmPFC*), and temporal lobe (*tl*) as well as fourth probe measuring systemic response at a site not influenced by brain activity (*global*).

1. *The dlPFC-probe* measures oxygenation changes at the top of the forehead at the dorsolateral prefrontal cortex, at Brodmann area 46. As discussed in Chapter One, neuroimaging data (from fNIRS and fMRI) implicates this region for executive top-down functions.

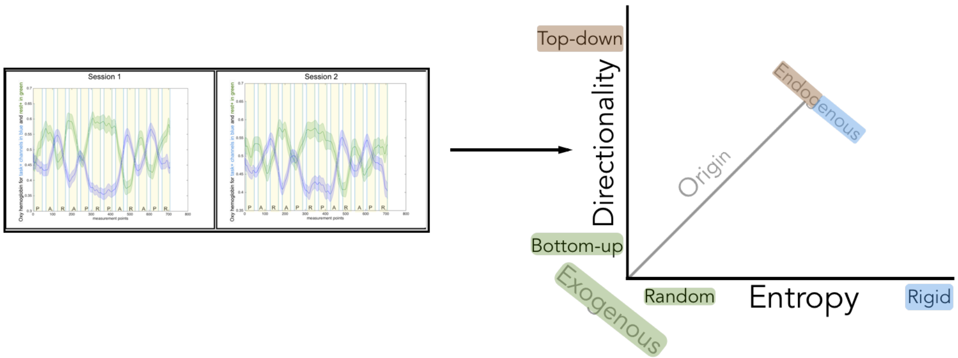

Figure 4.1: A function needed in a physical neuracle measuring fundamental dimensions of consciousness.

2. *The vmPFC-probe* measures oxygenation changes above the eyebrow at the ventromedial prefrontal cortex, at Brodmann area 10. As discussed in Chapter Three, neuroimaging data (from fNIRS and fMRI) implicates this region's involvement in the default mode network which implements automatic (bottom-up) mind wandering functions.

3. *The tl-probe* measures oxygenation changes by the ear at the temporal lobe, at Brodmann area 22/21. Neuroimaging data (from fMRI) implicates this region as a fundamental router in sensory processing [12].

4.1 **Physical Neuracle** Specification

The algorithm assumes oxygenation changes in the *dlPFC* and *vmPFC* are both anti-correlated with the *tl*, but makes no assumption about possible anti-correlation between *dlPFC* and *vmPFC*. I therefore form two anti-correlated networks (*dlPFC-tl*) and (*vmPFC-tl*), each including one *endogenous* node (*dlPFC* or *vmPFC*) and one *exogenous* node (*tl*). The results in Chapter Three suggest that the default operation of the brain is to cycle rhythmically back and forth between the exogenous and endogenous node in one of these networks, increasing the level of oxygen in the exogenous or endogenous node at an average rate of once every twenty seconds. As illustrated in figure 4.3, the analysis of this algorithm is partitioned into five distinct levels of abstraction, where the first three levels describe how to extract plausible

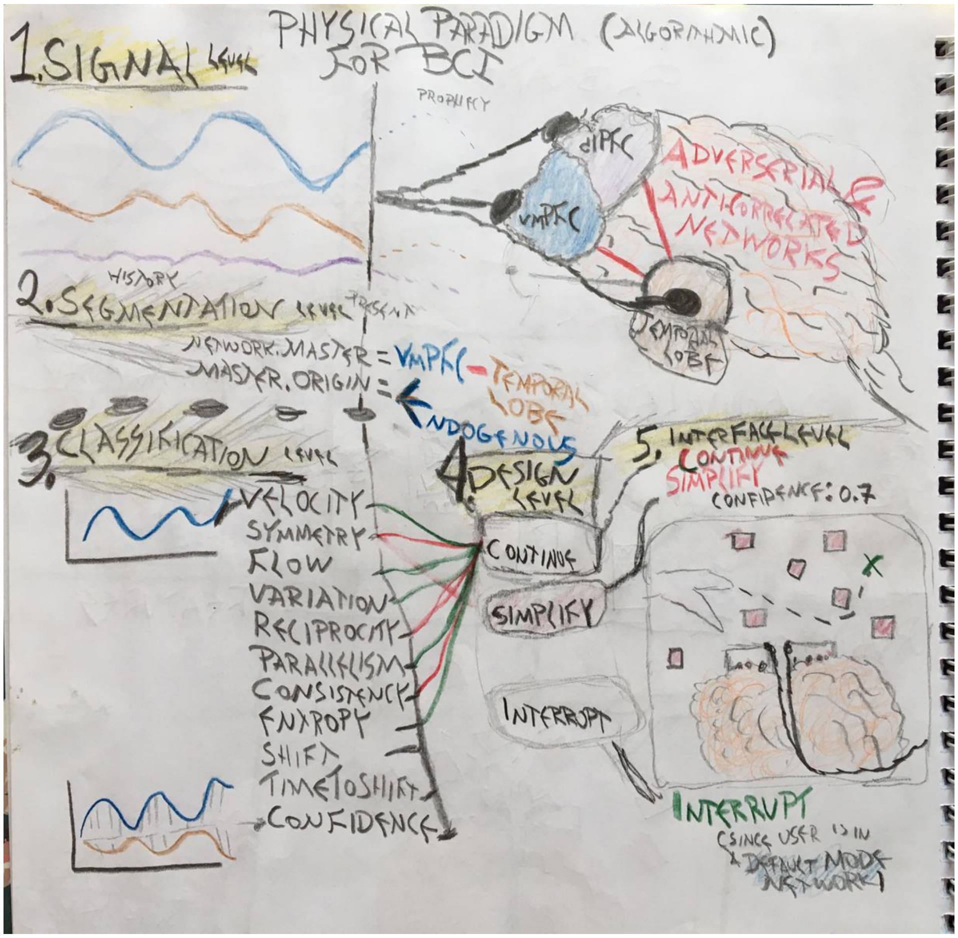

Figure 4.2: Physical Paradigm for BCI

user dimensions in a physical neuracle, and the last two levels describe how to act on changing values in these dimensions to the user's benefit.

1. At the *Signal Level*, probe configuration and signal processing algorithms identify pairs of regions with maximum anti-correlation, and outputs <user.networks>, as a set of time-series describing pairs of anti-correlated networks.

2. At the *Segmentation Level*, the time-series of *user.networks* is transformed into a set of segments with mathematical features, using heuristics for partitioning data at the moment *control over oxygen* is ceded from one network to its competitor.

3. At the *Classification Level*, the current <user.features> object is analyzed to produce <user.classifications>, which consists of the user's state along several

dimensions.

4. At the *Design Level*, the current <user.classification> object is analyzed to generate values in an <application.design> object which consists of current settings to generic trade-offs in design choices for a user interface, e.g. the user's tolerance to interruption. The mapping between <user.classifications> and <application.design> changes iteratively based on <user.behaviors>.

5. At the *Application Level*, a brain-computer interface determines a mapping between system changes in the application and changes in <application.design>. It may also trigger more custom changes based on fluctuations in <user.dimensions>.

4.2 Signal Level

The goal at the *signal level* is to output as many anti-correlated network pairings as possible. The rationale for conducting probe configuration according to this heuristic is five-fold.

1. *Customization argument:* With a function that evaluates the effectiveness of probe placement with as little as 5-10s of fNIRS data, it is easy to execute an adaptive probe placement algorithm.

2. *Reproducibility argument:* Uniform probe configuration is possible even if different subjects have distinct neuroanatomy.

3. *Filter argument:* A change in one channel without concomitant change in the competitor may indicate that change is due to noise and should be filtered.

4. *Information argument:* Given that memory and processes are encoded as relationships between distant matter, features comparing signals of two competing networks may provide information not available by single channel analysis.

5. *Reliability argument:* The network's overall anti-correlation can be used as a confidence measure, indicating how much fNIRS classifications should be trusted in relation to other sensors, as well as how *risky* adaptations can be.

Algorithm 2 processes the data using the techniques of Chapter Three collected from two endogenous channels (*vmPFC* and *dlPFC*) and one exogenous channel (*tl*). The experiment adjusts these channels to maximize anti-correlation between endogenous and exogenous nodes.

Algorithm 2 Physical Neuracle at Signal Level

```
procedure ConfigureFNIRS(user, fnirs)
loop:
    b10 ← some vmPFC channel in Table 3.1
    b21 ← some temporal lobe channel in Table 3.1
    b46 ← some dlPFC channel in Table 3.1
    global ← systemic fNIRS signal
    Preprocess(b10, global)
    Preprocess(b21, global)
    Preprocess(b46, global)
    networks.vmpfc-tl ← (b10, b21)
    networks.dlpfc-tl ← (b46, b21)
    If anti-correlation is weak goto loop
    return user.networks
procedure Preprocess(brain, global)
    BeerLambert(brain)
    bandpass(brain, 0.01, 0.1)
    LMSAdaptiveFilter(brain, global)
    movingAverage(brain, 0.5)
```

The physical neuracle at the *signal level* yields two anti-correlated networks determined on the basis of data from the experiment in Chapter Three.

1. Anti-correlated activity in *networks.vmpfc-tl* describes back-and-forth action between B.10 (or vmPFC, which partakes in the default mode network) and B.21 (or tl/temporal lobe, which is a fundamental sensory processing hub).

2. Anti-correlated activity in *networks.dlpfc-tl* describes the degree to which the B.46 (or dlPFC, which partakes in executive control networks) depends on information mediated from B.21.

4.3 Segmentation Level

At the segmentation level, this *physical neuracle* attempts to partition anti-correlated network data into its constituent cycles. The measured low frequency oscillation describes a slow cortical potential [72] in the brain. In the experiment of Chapter Three, the averaged network increased oxygen measures for 20s, before relegating control of the oxygen supply to the anti-correlated network (see figure 3.3). Acting on this averaged input, the algorithm given below would produce segments of approximately 20 second lengths. However, the key to the algorithm is that it adapts segment length depending on live fNIRS data.

Algorithm 3 Physical Neuracle at Segmentation Level

```
procedure EXTRACTFEATURES(networks)
    window ← 5
    analyzeNetwork(networks.vmpfc-tl, window)
    analyzeNetwork(networks.dlpfc-tl, window)
    return networks
procedure ANALYZENETWORK(network, window)
    length ← network.length
    masterSlice ← network[0][length − window...length]
    slaveSlice ← network[1][length − window...length]
    masterSlice.slope ← slope(master)
    slaveSlice.slope ← slope(slave)
    if masterSlice.slope > slaveSlice.slope then
        origin.name ← "master"
        network.origin ← network.master
    else
        origin.name ← "slave"
        network.origin ← network.slave
    if network.origin.name ≠ origin.name then
        network.lastLength ← length − network.start
        network.start ← length
        network.switching ← True
    else
        network.switching ← False
    network.master ← network[0][network.start...length]
    network.slave ← network[1][network.start...length]
    network.master.slope ← slope(master)
    network.slave.variation ← stdev(master)
    network.slave.slope ← slope(slave)
    network.slave.variation ← stdev(slave)
    network.length ← length - start
    network.networkCorrelation ← Math.abs(correlation(signal[0], signal[1]))
    network.cycleCorrelation ← correlation(network.master, network.slave)
```

Algorithm 3 segments the networks into cycles by comparing the slope of the last *window* seconds of activity at the two probe locations. Here, *window* is a somewhat arbitrary value set to five, balancing a trade-off between the immediacy and accuracy of segmentation; more experimentation is necessary for an optimal setting. The *segmentation level* determines two high level network properties that determine on which time-series the *classification level* should extract features.

1. The *network.master* variable is set to hold whichever of *dlpfc-tl* and *vmpfc-tl* has higher anti-correlation.

2. The *network.master.origin* variable is set to hold whichever time-series of the *temporal lobe* and its endogenous competitor (vmPFC or dlPFC) has higher slope.

4.4 Classification Level

Given the last k seconds of data stored as a segment in *network.origin.master* variable, the *classification level* produces a set of dimensions, whose values can be mapped onto design imperatives at the *design level*. Some dimensions are crafted using domain expertise, and their meaning can be reasoned with at the design level. Other dimensions are inspired by a process of optimizing signal features for agnostic neuracles in Chapter One. Yet other dimensions require playful introspection as enabled by the *Neuracle* software. The stated meaning of all dimensions should be regarded as hypotheses, which will be refined as knowledge of the brain's function accrues. This section enumerates thirteen dimensions and reasons about their physical meaning, before giving the algorithm for extracting them from the *network.origin.master* object. The following labels are used to further specify how the dimension should be computed.

1. Dimensions computed for the whole session of data are labeled as (S) and dimensions computed for the current cycle of data are labeled as (C).

2. A *normalized* dimension means that a cycle attribute has been normalized in relation to the whole session, so that 0 is the mean value, 1 represents a standard deviation above the mean, and -1 is a standard deviation below.

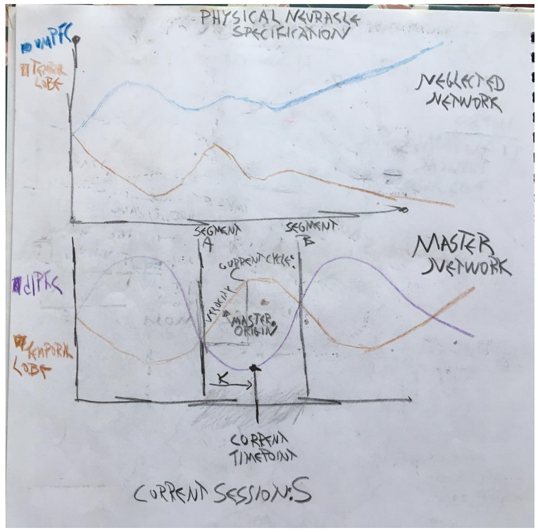

Figure 4.3: A physical neuracle

3. Attributes with (H) incorporate information about the history of this signal for the user.

1. **Master** [(vmpfc-tl or dlpfc-tl), C, <user.master.name>]. The master network dimension may describe whether the user is in a default mode network state (*vmpfc-tl*) versus an executive attention state (*dlpfc-tl*). In the proposed algorithm, *master* network is set to whichever network has higher anti-correlation strength. A strongly anti-correlated network (suggesting two distant regions in communication or competition) is likely to be performing considerable computation, and thereby be the master of attention. The subsequent dimensions are computed only on the master network. [1]

2. **Origin** *(endogenous or exogenous), C, <user.master.origin>* The region with higher slope of two competitors in the master network indicates whether or not the user is in a *endogenous* or *exogenous* mode of cognition. Effective cognition combines concrete, immediate exogenous sensory-motor mappings with more abstract, delayed/executive endogenous memory-plan mappings. A

[1] **Directionality** [top-down or bottom-up] may be the same dimension as **Master**, so that *directionality* is *bottom-up* when *master* is *vmpfc-tl* and *directionality* is *top-down* when *master* is *dlpfc-tl*. This way of thinking about the difference between possible master networks may be clearer to neuracle designers, who can think about the master of a given mental state as being controlled deliberately by the user's 'free agency' or as set by forces beyond its control, external or internal depending on the current value to the *origin* dimension.

second of brain activity contains many back-and-forth messages between more exogenous and endogenous networks that would be better described with an EEG. However, this dimension captures a low frequency oscillation between these networks, perhaps measuring which network currently owns conscious attention. The frequency is variable, but present data and fMRI literature [107] indicates that brains cycle between endogenous and exogenous modes of the master network at rates 20 second long on average. It is not clear what this means as a matter of conscious attention.

3. **Velocity** *(normalized), C, <user.master.velocity>* The user's master network changes its measured oxygenation with a particular velocity or slope. The velocity attribute is computed on the master's master region from *master.origin.slope*, which is always the larger of the two slopes in the network. It is not clear what high velocity signifies about the user, but Chapter Five gives data illustrating that this quantity is high when brains immediately switch between endogenous and exogenous mode by opening or closing their eyes. In the extreme case of opening eyes, the entire visual scenery must be recomputed, accounting for the rapid change in oxygenated values. Potentially, velocity is high when the state of the brain is changing rapidly. The neuracle described here is designed to reset slope computation (i.e. change the window of analysis) every time control shifts between the exogenous and endogenous mode of the network. This design choice means the velocity measure is computed on a small (5s) window length at the beginning, and on a longer window length at later points of the sub-networks action. The velocity attribute captures the strength of attentional shifts, and is less sensitive to changes in velocity later in the trial.

4. **Symmetry**, *(0,1), 2C, <user.master.symmetry>* Unlike the other cycle dimensions, *symmetry* is computed only once a cycle, at the moment control is relegated from some network to its competitor. Symmetry is 0 when the present cycle was exactly as long as the previous anti-correlated cycle, and

holds the number of seconds apart otherwise. Symmetry may be high when the relationship between two anti-correlated networks is less adversarial and more an act of concerted communication. Imagine one network (a) changes internal neuronal configurations for some period of time (a') until it reaches some energy limit and is not refueled. At this point, the competitor network (b) begins changing its internal configuration for some period of time (b'). If (a') and (b') are similar, then it is likely that the amount of internal reconfiguration by (b) depended on the amount of internal reconfiguration in (a), especially given a strong interpretation of the *Bayesian Brain Hypothesis*. On the other hand, symmetry should be low when the brain is biased towards action in just one of the two anti-correlated regions. Thus, a low symmetry value indicates a purer and more sustained version of an *endogenous / exogenous* state.

5. **Flow**, *seconds, C, <user.master.flow>* The flow dimension captures the number of seconds a user has remained in a given network. It can be thought of as a timer that begins when there is a change to which of two anti-correlated network has the larger slope, and continues until there is another shift in network. It is unclear if a brain region measured by fNIRS can increase its oxygenation content indefinitely. If not, it may be necessary to modify the segmentation algorithm so that this *Flow* attribute approximates the psychological construct [32].

6. **Shift**, *(True/False), C, <user.master.shift> Shift* is true exactly once a cycle and false at all other points. This information is also contained in the *origin* dimension which switches from *exogenous* to *endogenous* mode exactly once a cycle, and therefore exists as a convenience in the *design level.* The lag time between a measured shift and a real shift in attention is unclear. Conventional wisdom suggests that the BOLD effect has a 3-6 second delay.

7. **Variation**, *(normalized), C, <user.master.variation>* The variation dimension holds the normalized standard deviation of the cycle's raw oxygenation

content. It is not clear what high signal *variation* indicates about the user's state, but this feature was found to be useful fodder for an agnostic neuracle in Chapter One. High variation may indicate noise or movement, but effective filters at the *signal level* could yield a *variation* measure that reflected only within region brain activity. For example, if $< user.master > = vmpfc - tl$ $< user.origin > = endogenous$, then high *variation* may indicate an unusually active, and potentially more creative or anxious, default mode network. *Variation* may correlate with the high frequency oscillations present in EEG data.

8. **Reciprocity** *(0,1), C, <user.master.confidence>* The *Reciprocity* dimension is the inverse of *anti-correlation*, so that reciprocity is 1 when *anti-correlation* over the current cycle is -1, and 0 if *anti-correlation* is greater than or equal to 0. *Reciprocity* likely captures similar information to *Symmetry*, gauging the amount of computation in the network that depends on the competitive network's previous output. But unlike *Symmetry*, *Reciprocity* captures information about competitor network from the current cycle.

9. **Confidence** *(0,1), C, <user.master.confidence>* The confidence measure provides the *design level* with a measure for how much to trust the other dimensions. *Confidence* is set to 1 when the session and cycle correlations are both -1, indicating information confirmed by two sources. When anti-correlation is weaker, *confidence* is lower. But the *confidence* function is not trivial since *confidence* should also be high when *cycle correlation* is weak but the *session correlation* is strong, especially if *flow* is high. This scenario may indicate *parallelism* between otherwise competitive networks.

10. **Parallelism** *(0,1), C, <user.master.parallelism>* This attribute is set to 0 in the event of anti-correlation in the master network. It is 1 when two typically anti-correlated networks are perfectly correlated with each other. High parallelism likely indicates undesirable sensitivity to system artifacts and should only be trusted in the event that *globalConfidence* is high, indicating parallel

activity in a system that is otherwise serial. It is not clear what, if anything, may cause a suspension to anti-correlation, but as described in Chapter Three, expert meditators have reduced anti-correlation in a resting state, and thus more parallel activity [85]. This may be an interesting dimension for long term optimization.

11. **Consistency** *(0,1), S, <user.master.consistency>* This attribute measures the consistency in oscillations over the course of the session. High consistency may signify two different properties in the user depending on *user.master.name*. If the *dlPFC-tl* is the active master network, high consistency may indicate a longer period of sustained top-down attention, with consistent smart exchanges between exogenous and endogenous processes. If the *vmPFC-tl* is active, high consistency may indicate a duller day-dreaming state and a boring task, causing routine activity in the default mode network to cycle in and out of spatiotemporal existence. No matter the master network, low consistency may indicate a more creative, high entropy state (see below).

12. **TimeToShift** *(seconds), C, <user.master.timeRemaining>* This attribute estimates the number of seconds until a network releases control to its competitor on the basis of the previous cycle's length. It should only be trusted when *consistency* is high. The attribute can be leveraged to time digital events to better coincide with the brain's rhythms.

13. **Entropy** *C, H, <user.master.entropy>* A sort of inverse to consistency, entropy measures the unpredictability of a given slice of fNIRS data. It depends on the other dimensions and is therefore computed last. It is computed as the cycle shifts to the competitor, just before the cycle data is saved to the session. A Bayesian Belief updating algorithm with the dimensions in H (containing all cycles ever observed about that user) as its prior, predicts the dimensions in C, and updates to some new posterior distribution. The *KL-divergence* (see Chapter Two) in this transformation is saved as an attribute describing the user's state. It is possible that experience or consciousness feels more pedes-

trian when entropy is low, whereas it registers as an introspectively higher temperature in cases when entropy is high, perhaps indicating more overall change in physical space over time (computation) [28].

Algorithm 4 extracts these thirteen features from the <user.networks> object.

Algorithm 4 Physical Neuracle at Classification Level

procedure CLASSIFYUSER(user.networks, user.history)
  $vmpfc - tl \leftarrow user.networks.vmpfc - tl$
  $dlpfc - tl \leftarrow user.networks.dlpfc - tl$
  $classify(dlpfc - tl)$
  $classify(vmpfc - tl)$
  if *dlpfc-tl.confidence* > *vmpfc-tl.confidence* **then**
    $user.master \leftarrow dlpfc - tl$
  else
    $user.master \leftarrow vmpfc - tl$
  if *user.master.switching* **then**
    $recordCycle(user.master)$
procedure RECORDCYCLE(network)
  $cycle.lastLastLength = network.lastLength$
  $cycle.length = network.lastLength$
  $cycle.symmetry = |lastLastLength - network.lastLength|$
  $cycle.slope = network.origin.slope$
  $cycle.variation = network.origin.variation$
  $network.session.add(cycle)$
procedure CLASSIFY(network)
  $cycles \leftarrow network.session$
  $network.flow \leftarrow Normalize(network.length, cycles)$
  $network.timeRemaining \leftarrow mean(cycles.length) - network.length$
  $network.velocity \leftarrow Normalize(network.origin.slope, cycle)$
  if *network.antiCorrelation* < 0 **then**
    $network.confidence \leftarrow |network.antiCorrelation| * a + network.velocity * 1 - a$
    $network.parallelism \leftarrow 0$
  else
    $network.confidence \leftarrow 0$
    $network.parallelism \leftarrow network.antiCorrelation$

4.5 Design Level

If work at the *classification level* transforms fNIRS data into user dimensions that can be reasoned with effectively at the *design level*, the *design level* transforms user dimensions into generic *imperatives* about how to adjust data, processes, and interaction in the *application level*. This level of abstraction simplifies work for the software engineer who need not understand how the brain works when upgrading their application for the age of BCI. Instead, this engineer may simply add *listeners* to the application, which tracks changes in the *design imperative* object and triggers changes to the application's settings. If dedicated to improving state of the art BCI, the application should respond to the *design level* with a number scoring the measurable effectiveness of human-computer interaction at a given moment (which it may know from keyboard/mouse input or other settings. I organize these design imperatives into three *design imperative objects*, with four properties.

1. **Imperative:** This constant string value holds the design decision which should be acted on when *value* is positive.

2. **Value:** This normalized floating point value is positive when the *design level* believes in *imperative*, and negative when *design level* believes in the inverse of the *imperative*. It is zero when agnostic about the *imperative*.

3. **TimeToChange:** This integer represents the number of seconds until the application should act on the *imperative*. When zero, the *design level's* recommendation is to act immediately on its *imperative*.

4. **Confidence:** This floating point ranges from 0 to 1 in proportion to the confidence of its current setting to *TimeToChange* and *Value*. A *confidence* of 1 means these decisions can be trusted.

The responsibility of the *design level* is to set each of these four properties on the basis of user classifications on three distinct *design* objects:

1. **Continue** is positive when the user appears to be in a desirable state, and negative when the user is in an undesirable state, and system settings need not change. The physical neuracle in this paper is engineered for scoring attention along a positive-to-negative axis, but improved physical neuracles may be able to score positive and negative affect.

2. **Simplify** has a positive value when the user appears to be in a cognitively overworked state, warranting simplification in the user interface, and negative when the user appears to be in a bored state, which suggests that task demands can be escalated.

3. **Interrupt** has a positive value at moments when the physical neuracle deems the user's brain prime for interruption, and negative when the user is unlikely to be able to process additional information.

4.5.1 Imperative Functions

The *design level* needs a function that takes as input <user.classifications> and assigns a high *continue.value* when the user is in a high quality attentional state, and a negative value when the user is in the worst possible attentional state. This function is challenging to hardcode, but the algorithm need only require reasonable initial settings, before it can be iterated upon with machine learning algorithms sensitive to feedback from the *application level.* The most desirable attentional state is the *exogenous flow* state, where attention is effortlessly wedded to the environment, without consistent regression into *endogenous, default mode* networks. Instead of giving code for these *imperative function* (as I did for the previous *physical neuracle* layers), I provide a table, indicating how I expect the *design* object ought to to fluctuate in response to changes in a particular dimension.

The tables 4.1, 4.2, and 4.3 thus contain predictions for the valence of weights (w) for a hypothetical algorithm that multiplies some *dimension* (d) and gives the product *design.value* (v) (w*d = v). When the weight in the table is negative (-), *imperative.value* is higher when the dimension is lower; if positive (+), the values

Table 4.1: Weights for determining *continue.value*, *simplify.value*, and *interrupt.value* from *dimensions*

Dimension	**Continue**	**Simplify**	**Interrupt**
Master	dlPFC	dlPFC	vmPFC
Origin	exogenous	endogenous	endogenous*
Velocity	+	-	-
Symmetry	-*	-*	+
Flow	+	0	-
Variation	+	0	-
Reciprocity	-	-*	-*
Parallelism	+	-*	-
Consistency	-	-*	-*
Entropy	+	+	-

covary positively; if 0, I am too uncertain to form a hypothesis. A string value indicates the value for which the *imperative* is higher. If these weights have an asterix (*), then the predicted weight depends on the *dimensions.master.name*, e.g. I expect a positive weight if the user is in the *dlPFC* network, and a negative weight if the user is in the *vmPFC* network. Similarly, if these weights have two asterixes (**), then the valence of the weight depends on *dimensions.master.origin* (whether the user is in an endogenous or exogenous mode of the network).

Table 4.2: Weights for determining *design.confidence* from *dimensions*

Dimension	**Weight**	**Rationale**
Symmetry	+	Trust adaptation when system understands user state.
Variation	-	Leave user alone if their signal is noisy (or they are computing).
Consistency	+	Interrupt user when the system can predict user's state.
Confidence	+	Change interface only when system is confident.

Table 4.3: Weights for determining *design.TimeToChange* from *dimensions*

Dimension	**Weight**	**Rationale**
Velocity	-	Don't act since information is currently obtained about user's state.
Flow	-	Do not change if the user is in flow.
Shift	True	Before user deepens a state, change conditions.
Parallelism	-	Do not interrupt user while computing.
TimeToShift	-	Interrupt user when network is about to shift.
Entropy	-	Do not alter conditions during creative states.

An evaluation of the reported physical neuracle awaits future work. The section should therefore be interpreted as mathematically precise hypotheses as opposed to factual statements about brain-computer interfaces. The next chapter argues that the wisest approach towards evaluating these hypotheses consists not in implementing and evaluating the algorithm in traditional user studies. Instead, the best approach is to distribute software for introspection and rapid hypothesis-testing and produce research contexts where other BCI engineers are empowered to study themselves. Put another way, it is my hope that researchers replicate the steps I took to produce this algorithm as opposed to replicate the algorithm itself.

4.6 The Application Level

Benefiting from a mapping logic between *classifications* and *imperatives* implemented at a system wide *design level*, a brain-augmented *application* can focus on the problem of mapping the *imperative's* changing instruction (*simplify*, *continue*, and *interrupt*) onto system parameters. Chapter One gave four examples of *implicit BCIs* at the *application level*. If the logic of the *design level* is sufficiently broad, then it should be possible to redesign these four interfaces so that they operate on the instruction of the *imperative object*.

1. The *dynamic difficulty adjustment engine* implemented in the UAV experiment would use the *simplify object* to determine whether to add or remove aircraft from user jurisdiction. With a positive *simplify* value, the simulation would remove aircrafts from the screen, and vice versa.

2. In the *target expansion technique* implemented in the *bubble-cursor* experiment, the *simplify object* would be used to control the expansion width of the mouse. With a positive *simplify* value, the interface would increase the selection range of high priority targets, and vice versa.

3. The *route recommendation engine* implemented in the first *Google Glass* experiment would use the *simplify object* to determine whether to provide a *complex-but-fast* set of turns or a *simple-but-slow* set of turns.

4. The *cognitive heatmap* implemented in the second *Google Glass* experiment would use the *continue object* to determine how to tag associated movement data in a map. When *continue* was positive, the cognitive heatmap would place a marker at the corresponding user location so that the user could revisit the associated images captured by *Google Glass* at that point of interest.

A third *Google Glass* experiment implemented by Afergan would have made use of the *interrupt* object [5]. In that experiment, users completed a primary task, in which they acted as campaign workers, delivering election material to houses they passed in a simulation. Simultaneously, users also completed a secondary task with lower priority that was controlled by the information they received in Google Glass. The instruction in *Google Glass* would inform them to deliver additional packages to houses satisfying certain criteria. However, this instruction would distract them from the more important primary task if the user lacked the cognitive resources to handle an added decision making burden. With the physical neuracle specification given in this chapter, this experiment could be redesigned to determine whether to overlay Glass information using the instruction mediated by the *interrupt* object. When *interrupt* was positive, the simulation would overlay potentially distracting information enabling the user to complete the secondary task.

The physical neuracle can therefore solve the existing *implicit BCI* problems described in Chapter One's agnostic paradigm. And it does not suffer any of the five problems that limit agnostic neuracles described in Chapter Two. Instead, the physical neuracle depends on a real understanding of the relationship between oscillations in the brain and cognitive functions. This dissertation has advanced a physical model of the brain wherein networks become active if they can reduce the entropy of endogenous or exogenous signals. When there is an opportunity to reduce entropy in *exogenous signals*, the brain engages a family of task-positive networks. In the absence of such external demands, the brain begins to reduce entropy in *endogenous signals*, generating memories, ideas and fantasies that are consistent with the history of the system and its predilection to survive and replicate. The observed effect of anti-correlation between these two systems (task-positive and default-mode networks) gives the foundation for building physical neuracles.

Since the given design for a physical neuracle describes the relationship between information processing and detected signals in the brain, it is possible to build a new sort of brain-computer interface, which adjusts inputs to the brain in a feedback loop with measurements of its state, driving it towards more desirable mental states. Because this system benefits from a physical understanding of the relationship between fNIRS-detected signals and neuro-cognitive events as the energetic byproduct of hierarchically layered entropy suppression, the *bidirectional BCI* can use machine learning at the *interface level* to discover an optimal information load in signals to the brain and the user actions they warrant.

Chapter 5

Neuracle

So far this dissertation has evaluated key parameters of a bidirectional BCI. To that end,

1. *Chapter One* illustrated an agnostic paradigm for building *implicit BCIs* using fNIRS, from which one could draw inspiration in the design of a bidirectional BCI.

2. *Chapter Two* demonstrated that the agnostic paradigm was inadequate for building BCIs that generalized beyond a laboratory context, and that BCIs demanded a physical paradigm.

3. *Chapter Three* advanced a theory and empirical findings for a physical paradigm for BCI, oriented around *anti-correlated networks.*

4. *Chapter Four* gave an algorithmic specification for this physical paradigm.

Note that in Chapter Four, I cautioned that the assumed relationship between mental and physical processes was speculative based in large part on my own introspective efforts. In this chapter, I distribute the software program I built to rapidly iterate-and-test hypotheses about the relationship between fNIRS data and real-time changes in my own mental state. My hope is that other researchers will use this software to understand and improve upon the specification for the physical paradigm proposed in Chapter Four. To this end, I present a new software system called Neuracle.

Neuracle marries open-source Java-based statistics, filtering, and machine-learning libraries (such as Weka) with an easy-to-use interface for visualizing, tagging, manipulating, storing, and broadcasting data. The system makes the process of rapidly testing the detectability of a novel state an easy and satisfying endeavor. A video demonstration of its use can be found at Youtube (https://youtu.be/dKIhXeqwNk8). A functioning demo can be found running online at (http://sensormining.herokuapp.com/). Source code can be found at GitHub (https://github.com/samhincks/Neuracle) under the MIT license. Documentation and tutorials are made available within Neuracle.

This chapter contains three parts:

1. Section 1 describes a workflow for Neuracle and an evaluation of this work flow on five participants.
2. Section 2 describes aspects of Neuracle's implementation.
3. Section 3 describes the outcome of small experiments I have conducted, mostly on my own brain.

5.1 Self-calibration Validation of Agnostic Neuracles

In this section, I will describe and test the following work flow for calibrating new physiologically-based machine learning algorithms, and refer to the interface's three components: (a) the data-view, which contains representations of active datasets (preloaded or streaming) and the tools for building machine learning algorithms from data, (b) the visualization-view, which shows live streams of the data and meaningful visualizations of it, and (c) the console, which communicates system output and recognizes over fifty commands for data manipulation, self-calibration, and more.

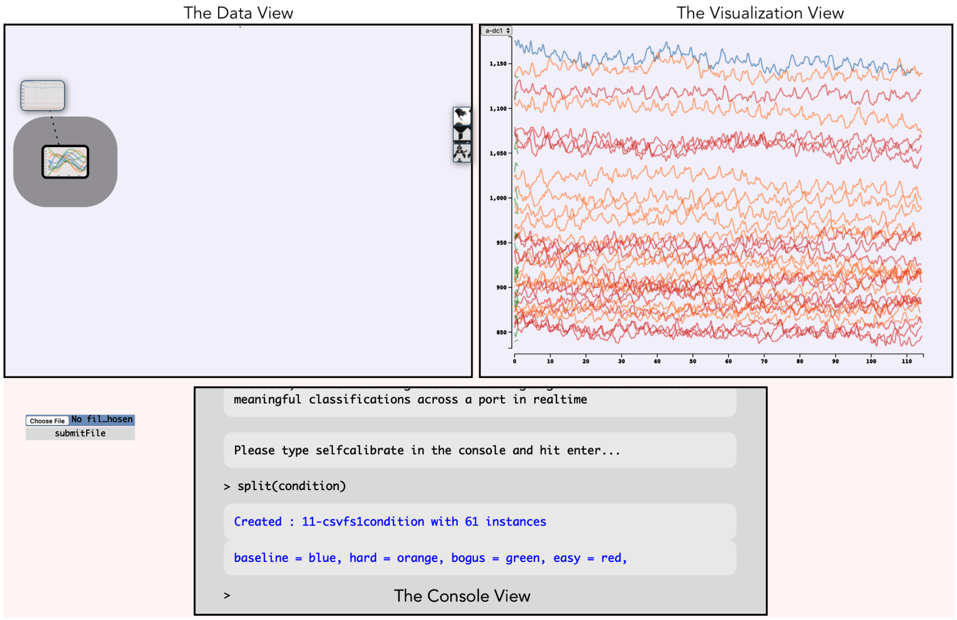

Figure 5.1: The Neuracle Interface

5.1.1 Participants

Five participants (3 female), ranging from ages 20 to 25, partook in self-calibration.

5.1.2 Method

In the experiment, tasks A (data synchronization), C (machine learning), and D (real-time classification) are completed by the 'trainer' (in this case, the experimenter), and task B (self-calibration) is completed by the one whose brain is being calibrated (in this case, the subject). I tested this workflow in an experiment, using fNIRS 16-channel ISS OxiplexTS: half sensing activation changes in the left dlPFC and half in the right dlPFC. The two probes are identical, each including source-detector pairings at 2cm, 2.5cm, 3cm, and 3.5 cm. Raw data values were converted to (de) oxygenation measurements using Boxy software, and relayed in real-time to Neuracle.

5.1.3 A: Data synchronization

Neuracle is designed to interact with changing values in a *MySQL database*. Once data is confirmed to be streaming into this database, the connection can be opened from Neuracle's console, which places a visual representation of the data in the data-view of the screen. Double clicking on this object shows a streaming view of the data in the visualization view.

5.1.4 B: Self calibration

To associate data with estimated mental states, the user must then initialize a set of trials where task is known by specifying a pattern to describe the condition, length, and quantity of trials. Neuracle supports built-in calibration for cognitive workload using the n-back. For the n-back, participants listen to an audio recording of numbers, with a 2.5s interstimulus interval, and enter these into the system. An affirming smile appears following correct responses, and a red frown follows incorrect response; classification accuracy is displayed to the user (and also saved along with the data, so that trials with many errors can be eliminated from analysis). In between trials, the user reports their experience of difficulty and focus. For inductions not yet built into Neuracle, the system merely displays trial transitions, playing helpful sound alerts if the user is performing the task on another screen.

5.1.5 C: Machine Learning Design and Evaluation

Neuracle provides a suite of tools for involving the human in the machine learning loop. When training data has been collected, the experimenter can examine condition-partitioned views of the data (see figure 5.2). These visualizations provide a preliminary outlook into the quality of the data, suggest outliers that ought to be excluded, as well as inform filtering and choice of features. Next, the experimenter can apply bandpass filters or other manipulation techniques (e.g. baseline subtraction, adaptive filtering or z-scoring); in practice, such manipulation introduces a new dataset to the view, so that the choice can easily be reverted or compared to. (For states with consistent inter-participant physiological signatures, macros can be defined to implement a decided sequence of manipulations.)

The trainer can then define the feature-set; in Neuracle, a feature is specified by three values (a) descriptive statistic (b) channel (singletons or averaged), and

(c) time (what subset of the data to examine). There are boilerplate descriptive statistics (e.g. mean , slope, and the first and second derivative) as well as more advanced features (e.g. the SAX-distance [84] predictive granger-causality estimates [54] between two channels (useful for connectivity analyses), and the signal power at a specific frequency. These features have been selected for processing fNIRS in the time and frequency domain as well as for connectivity analysis, but apply generally to any physiological sensor. When a new feature-set is introduced, Neuracle automatically ranks each feature's *information gain* [76], which the experimenter can view by double clicking its visual representation.

Finally, the experimenter chooses among thirteen Weka machine-learning algorithms by evaluating them with cross-fold validation; this command shows the internal leave-one-out classification accuracy for the method, as well as confidence-thresholded estimates. A prudent experimenter does not wield this specificity to overfit nor does she flaunt inflated classification scores resulting from exhaustive trial-and-error. If classification accuracies appear abysmal, *Neuracle* supports tools for loading old datasets from other subjects, and either using this in lieu of or in conjunction to the subjects' data.

5.1.5.1 D: Realtime classification

When the experimenter has trained the machine learning algorithm, she can instruct it to classify the streaming data, and request a classification on the last n datapoints, where n corresponds to the number of points in the trials it was trained on. This command can then be set to occur at timed intervals with the repeat command, and classifications streaming from the console can be redirected to a different address (e.g. a local port in the software in charge of wearable adaptation).

5.1.5.2 Results and Discussion

Participants reported appreciating the presentation of feedback as a green or red smiley, and also viewing time-series that represented changes of their own brain activity.

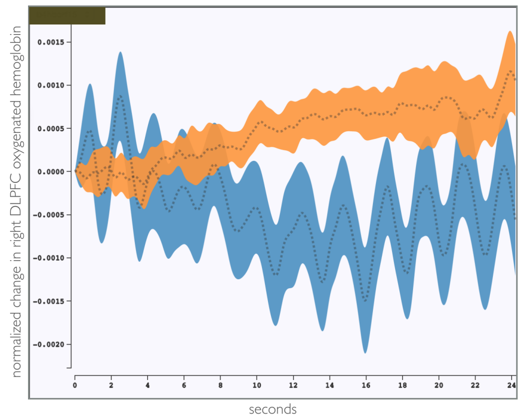

Figure 5.2: Average change in right-DLPFC oxygenation for all 14 trials of the five participants (35 0-backs in blue and 35 2-backs in orange)

.

Data manipulation was kept at a minimum to support real-time processing. For each trial, the slope, standard deviation, and mean were used as input to Weka's Logistical Model Tree [81]. No filtering was applied, in part since short term oscillations reflect heart rate and breathing patterns, which also correlate with workload [58]. In leave-one-instance-out evaluation, classification accuracies ranged from 50 to 85% (m =65%, s=17%); and when classifications with an associated confidence measure above 75% were considered (47 of 70 samples), accuracies ranged from 50 to 88% (m =70%, s =17%).

After the experiment, all subject data was appended, and I evaluated the internal leave-one-out-accuracy (training data included data from the individual and others). The LMT correctly identified whether a trial was a two or zero back 87.5% of the time (90% when confident) even though individual subjects had classification accuracies as low as 50%. This underscores the importance of using many trials for machine learning, and suggests that future experiments ought to merge current calibration data with historic data. The most informative feature in this analysis as measured by information gain [76] was the mean deoxygenated hemoglobin value in the second half of the trial of a probe sensing the right dlPFC. These results validate Neuracle's capability to induce distinguishable cognitive states in its user.

5.2 Neuracle Implementation

As described in the previous section, the Neuracle interface contains three primary visual components:

1. The *data view*, where all loaded datasets appear, hierarchically organized to reflect the different stages of the data as it is filtered and in other ways manipulated.

2. The *visualization view*, which produces different visualizations of the data object that is clicked, including real-time streams of data and a side-by-side organization of an experiment's separate trials, color-coded by condition.

3. The *console view*, which enables the user to execute commands on the data objects, including procedures for initiating experiments and labeling data, training machine learning algorithms, and filtering the data (see figure 5.3).

These *view* components are implemented using *HTML, CSS, and Javascript* and common web frameworks, including jQuery to simplify DOM manipulation and D3 for visualizations. Neuracle's implementation follows a *model-view-controller* architectural pattern, where the underlying model as all the fNIRS data that is currently loaded is implemented in a Java back-end. Communication between the front-end web stack and back-end Java occurs via the *Stripes* web application framework.

To transmit user intention from the view to the model, the user selects a data object in the *data view* and enters a command into the *console view.* A script (javainterface.js) first determines whether the command can be parsed locally (in the web client) or whether it needs access to the real data representations in the back-end. For commands that entail accessing or changing the data model, the data object which is to be acted on (the one selected) is known via *ActionBeans* in the Stripes framework. With this information, the input parser determines how to change underlying data representations when the command arrives in the Java

Figure 5.3: A list of commands implemented in Neuracle, available by typing tab into the console

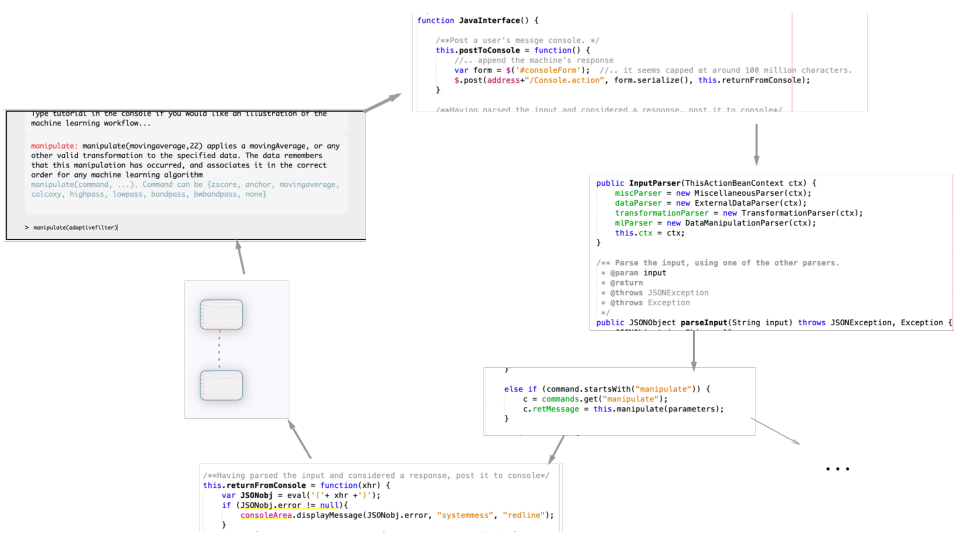

Figure 5.4: Neuracle code infrastructure. Omitted code (...) executes an adaptive filter on the underlying data

back-end, and sends a response back to the Javascript file javainterface.js, which changes local data in the client, as shown in figure 5.4.

5.2.1 Adaptive Filtering

The previous section described a workflow for building agnostic neuracles. As shown in Chapters Three and Four, the essential component for a physical neuracle is the *adaptive filter*.

Motion artifacts constitute the major bottleneck for brain monitoring in an fNIRS-based wearable computer setting. Slight movements can cause a slight decoupling between the sensor and the skin. Adaptive filtering, introduced in [135], is an effective technique for filtering systemic trends in local fNIRS data. Bandpass filtering successfully removes breathing and heart rate, but it leaves intact spontaneous low frequency oscillations that do not have neurological origin. These oscillations would be present in both a shallow and deep source-detector pairing, and can be eliminated from the deep source-detector pairing with knowledge about what frequencies were common between them.

Neuracle comes equipped with the software necessary for adaptive filtering. Specifically, it uses the Widrow-Hoff Least Mean Squared (LMS) algorithm to remove common inter-channel oscillations (as detailed in [135]), implemented with an external library [15]. In this self experiment, I alternated between four trials of sixty seconds in a passive mind wandering eyes-open state, separating 5 three-minute trials of eyes-closed meditation, in which I attempted to align conscious experience with breath while recognizing and mindfully discarding any incoming thoughts or feelings. The purpose of this exercise was to obtain two conditions that maximized brain-exclusive differences in state in order to test an adaptive filter algorithm under pristine conditions. These conditions differ by degree of deliberate engagement and volume of incoming sensory input - features of mind likely to implicate signal only in brain and not skin.

Figure 5.6 illustrates the presence of a signal invisible without adaptive filtering. The bottom graph shows a dataset in which the frequencies present in a 1 cm source-detector pairing have been subtracted from the frequencies present in a channel showing a source 3 centimeters apart from the same detector. Since the

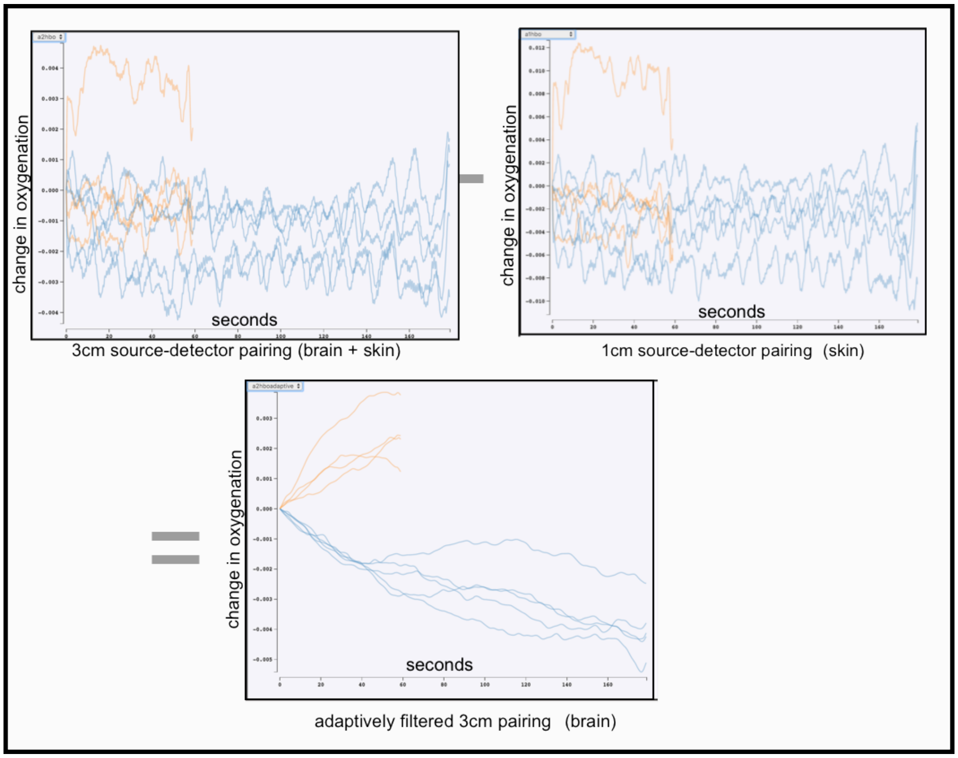

Figure 5.5: The effect of adaptive filtering on left dlPFC oxygenation data. Orange trials show 60 seconds of controlled eyes-open rest. Blue trials show 180 seconds of eyes closed meditation

near channel only detects oxygenation oscillations from skin and the far channel detects oscillation from both skin and brain, an adaptive filter enables the investigation of an exclusively brain-based signal. In figure 5.6, the adaptively filtered data shows that meditation involves a reduction in oxygenation content in the left *dlPFC*, before it starts to increase in the resting condition. This filter works both offline and real-time in Neuracle, thereby contributing the enabling software for BCI in the physical paradigm. As shown in Chapter Four, this discovery period can be leveraged both to rapidly prototype feasible states and to intelligently search the forehead for the most effective probe placement. Figure 5.6 shows the commands needed to execute the adaptive filter on data from within the Neuracle console.

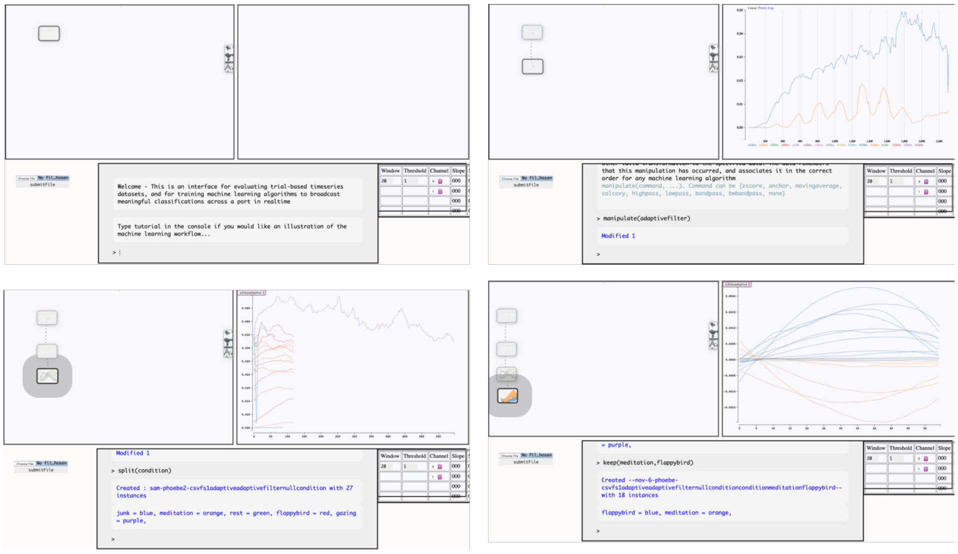

Figure 5.6: Interface commands for executing an adaptive filter on selected data

5.3 Self-Experiments for a **Physical Paradigm**

5.3.1 4-back

In this experiment, I completed a total of fourteen 30 second aural 4-backs (in orange) and aural 0-backs (in blue), interspersed with 10 second resting periods. (In an aural n-back, participants repeat the number heard n iterations ago). The data (which have been adaptively filtered and anchored to start at zero) are shown in figure 5.7. To solve a 4-back, I sub-vocally rehearse a 4-item mental buffer. When I hear a new number, I say the first number of the string recently rehearsed, then immediately repeat the string but with the last string's second element in the first position, and the new element in the last position. As long as I keep escalating n, it is virtually impossible to solve the problem without granting the task the exclusive province of my attention.

During a 0-back, I do not exert the same level of control over the activity of my mind. I complete the task, but occasionally my conscious mind is occupied by other thoughts or feelings.

What is especially interesting with fNIRS self-analysis is the possibility to dissect the trials which break an otherwise coherent pattern. The perfect self-analytical experiment is when you have a clear pattern, with exactly one exception. In this experiment, each condition has one exception. For the easy trials (when I was merely repeating numbers), one trial has the character of the 4-back trials. For the 4-back trials, one trial has the character of the 0-back trials. The 4-back exception is especially interesting because it was the last trial in the series and I know exactly how it mentally unfolded. It was the last trial and I was getting tired. For the first five seconds, the trial unfolded like every other 4-back trial. I heard 0, then sub-vocally said zero. I heard 3; then sub-vocally said *zero-three*. I heard 5; then sub-vocally said *zero-three-five*.

But then I heard six, and instead of adding it to a four-item buffer, I said aloud the first item in my buffer (zero), immediately noticed the mistake and then fumbled through the remainder of the trial unable to recover. Five seconds into the trial (approximately when I heard 5), there is a spike in the oxygenation levels in my left vmPFC. There are three possible explanations for this. First, it could be entirely coincidental. Second, it is possible that I made the mistake, noticed the mistake, and the data reflects my frustration. But a subtle clue in the data dismisses this possible. The break in activation occurs two seconds before my noticing that I have failed the trial. That suggests the third possibility is true.

Five seconds into the trial, either a rhythmic biological force or a string of computational association, cemented computation in one of the nodes antithetical to proper focus. For 6/7 of the 4-back trials, a combination of task difficulty and mental preparedness enabled me to block out the otherwise near-continuous presence of my default mode network. For the 7th and last trial, I got ready, doing my best to sustain focus, but five seconds in, neurological circumstance concentrated computation in a network of mind that interferes with task-related rumination. This explanation aligns exactly with the observed data, my private experience of it, as well as the cognitive science literature.

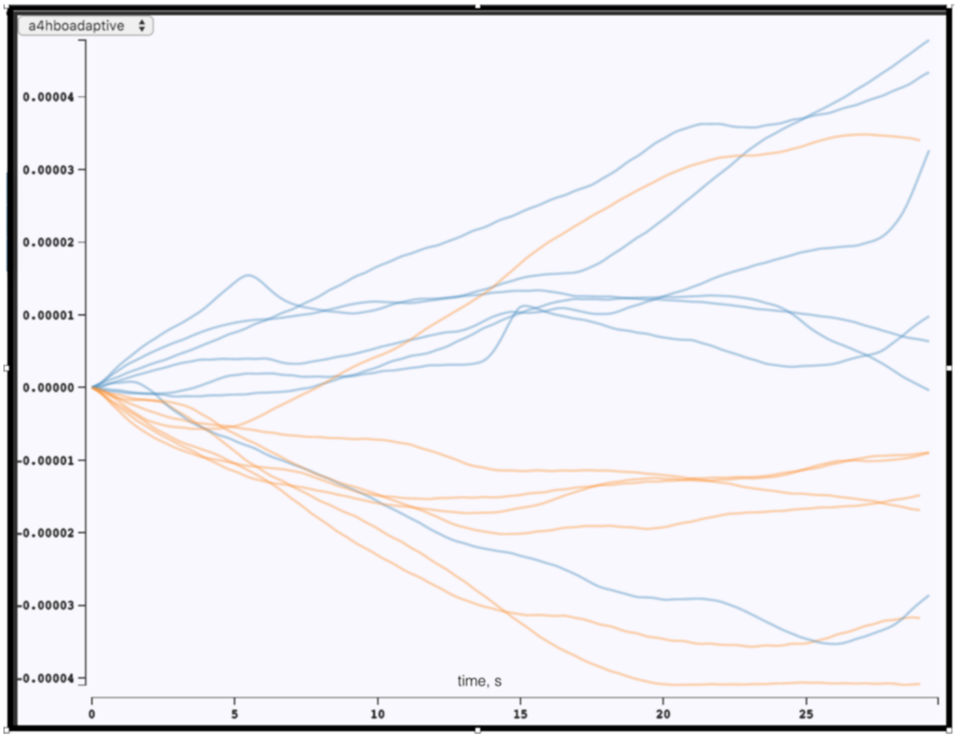

Figure 5.7: My brain solving a four-back

5.3.2 Anti-correlated network validation

In another series of self-experiments, I sought to verify the anti-correlated network findings advanced in Chapter Three using a different fNIRS on myself. Chapter Three proposes that the strongest anti-correlated network in the brain, indexing a trade-off between endogenous and exogenous processes, can be measured by one probe placed by an eyebrow (vmPFC) and one by the ear (temporal lobe). The *Imagent* probes used for the verification experiment require the probe to interrogate bare skin, requiring me to shave part of my head. As illustrated in figure 5.8, the self-experiments validated the existence of an anti-correlated network between these two regions and underscored the phenomenology: the PFC probes increase every time I 'regress into endogenous thinking', which includes both mind wandering and executive thinking, and the temporal lobe probe increases when interest returns to the world. This self-experiment taught me that the easiest way to control the signal was to limit visual input. When I cover or close my eyes, the prefrontal cortex probes increase. When I uncover or open my eyes, the temporal lobe probe increases. I

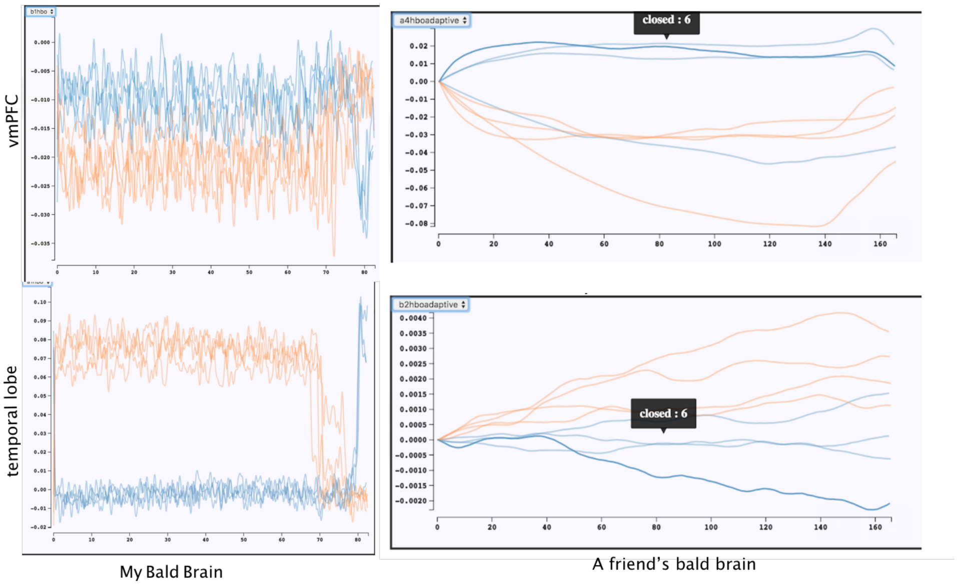

Figure 5.8: Anti-correlated networks in myself and a bald friend

verified this experiment on another bald friend.

5.3.3 Two-person eye-gazing

The next experiment was inspired by a finding in psychological literature indicating that prolonged eye-gazing leads to hallucination [26]. I sat face-to-face with two friends as we both wore fNIRS interrogating our left hemisphere PFC. We agreed to pick an eye and stare into it for ten minutes. For me, the hallucinations set in after about five minutes, and the associated state of entropy is apparent in fNIRS low frequency oscillations in the PFC. As illustrated in figure 5.9, compared to the first five minutes, the entropic state entails rapid back-and-forth sharing of data between the PFC and its anti-correlated exogenous networks, as data is shuttled to many different regions of the brain.

I cannot speak about the mental life of the three participants who gazed into my eye, but figure 5.10 indicates unlikely alignment between me and 'friend two' as well as me and 'friend one' in our second session. 'Friend one' and I decided to repeat the experiment in the meditative state induced by the first session. This time the PFC signal was flat for both of us in the first five minutes, before it switched in

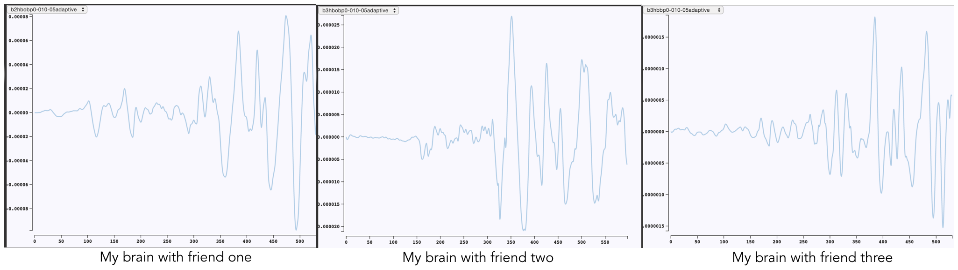

Figure 5.9: My brain during gazing into the eyes of a friend

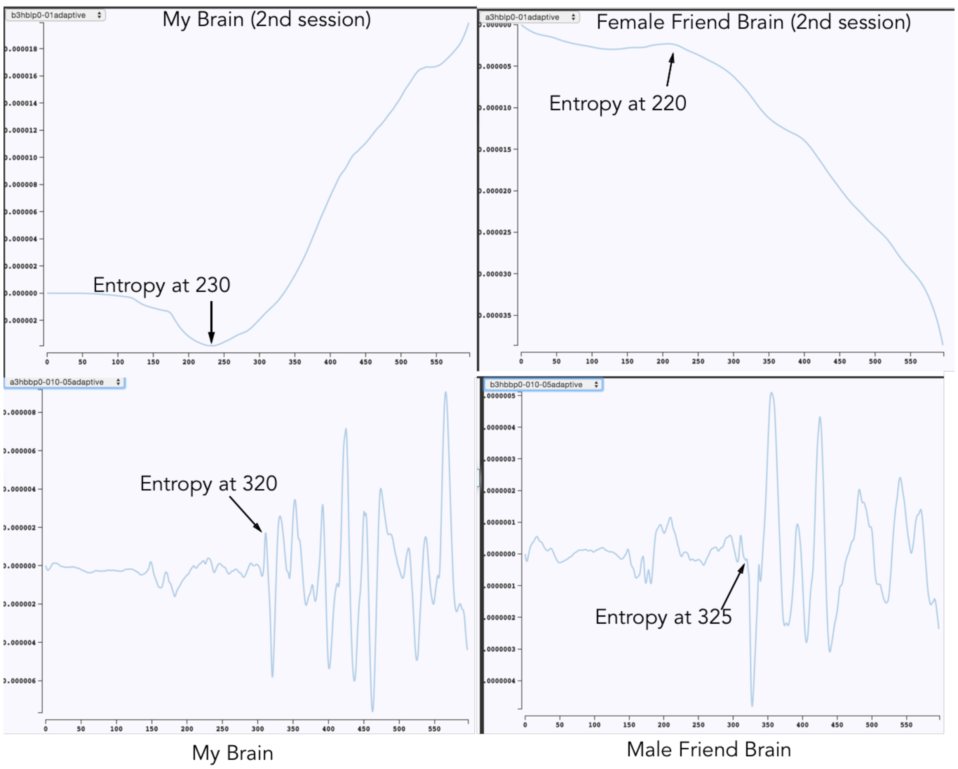

Figure 5.10: Two friends' brain data as they look me in the eye

opposite directions for each of us, at right about the same time. [1]

5.3.4 Biological Neuracles

As I have worked on the problem of measuring mental states using fNIRS over the past decade, I have developed an acuity for noticing shifts in my own mental state. I

[1] Human connection is an extraordinary phenomenon. It seems like we are living in a sort of computer, where the underlying physical space contains more dimensions than what the sensorium would lead us to believe. Although more research is needed, the synchronization between brain signals shown here (what you may call empathy) may suggest the existence of extrasensory channels of communication, and a rudimentary informational entanglement between our species.

can say with certainty when I am in a pure and primal *exogenous state*, absent both default mode network and frontoparietal networks. It happens when I dance and look people in the eye. I can initiate a pure frontoparietal network by rehearsing items in a working memory buffer, as I did in the *4-back* experiment above. And I can enter a meditative state where I by force of will lock my brain to a sensory stimuli, such as a sound or a breath.

But I have also learned that it is hard to reliably enter a given network when I do not control my task. And, as described by the *engagement problem* in Chapter Two, it is probably just as hard to reliably induce a given network in another person. Progress towards building the physical neuracles necessary for bidirectional BCI thus depends on a change in methodological norms: away from a study of random brains and towards studies on oneself and those with whom one can establish a deep and introspective dialogue, oriented a common vocabulary for describing mental states. Recall the definition of a neuracle from the introduction. A neuracle is an algorithm which summarizes the state of an information processing system such as the brain along a set of dimensions. So far, this dissertation has only considered digital neuracles. The software package *Neuracle* enables its users to create both agnostic neuracles and physical neuracles. But its highest aim is to train biological neuracles. A biological neuracle is a network in the brain that introspects and reports upon its state.

Methodological thesis: Progress towards *bidirectional BCIs* depends on the proliferation of biological neuracles, which is facilitated by the *Neuracle* software.

Neuracle supports a workflow designed to train biological neuracles. When the user of *Neuracle* types 'experiment' in the console (a command I adapt to string together a complex sequence of commands), the system initiates a sequence of songs and makes so that every time the user presses enter, a label is sent to the back-end. The biological neuracle is supposed to grant their undivided attention to the music. But inevitably, *endogenous gravity* will pull their attention into the *default mode network*. The biological neuracle will linger in the *default mode network* for a period of time before they notice they are lost in thought. At this point, they will press

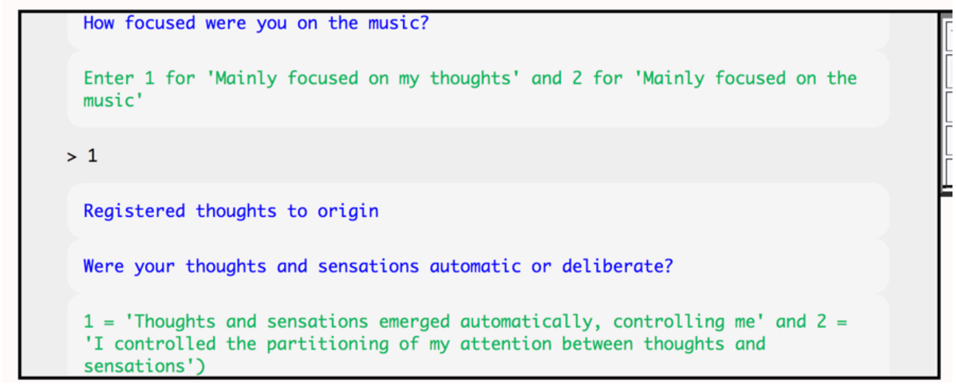

Figure 5.11: Feedback

enter, and send a label to the back-end.

As shown in Figure 5.11, at the end of the trial, Neuracle sends two questions to the user in order to further label the introspected states of the biological neuracle. These questions encourage a neutral information processing vocabulary for describing mental states. The first question attempts to resolve the degree to which attention was wedded to the music versus the endogenous gravity (or default mode network thought). The second question resolves the degree to which experience felt automatic (as though the bottom-up data was controlling thought) versus deliberate (as though the participant's agency exerted over cognition).

Although *Neuracle* was designed to be a web-based platform for intercepting, processing, and responding to fNIRS data from disparate wearable computers such as Google Glass, its key affordance may be in providing a live representation of data, bundled with the necessary filtering tools and machine learning as well as benchmark interfaces for state induction. In distributing this source code, I hope to inspire fellow brain-computer interfacing researchers to adopt a somewhat less conventional but more efficient approach to deciphering brain signals, and join me as biological neuracles.

Chapter 6

Electricity as Stimulation in Bidirectional BCI

This chapter begins the *stimulation part* of the investigation into a future user interface that measures its user's mental state and responds not only through a display but also by sending output directly to the brain, leading to a primitive *bidirectional brain-computer interface*. The previous chapters have explored interactive systems which measured brain state with functional near-infrared spectroscopy (fNIRS) for communication from user to computer; I now explore transcranial direct-current stimulation (tDCS) as a channel in the opposite direction. The goal is to integrate this with brain measurements from fNIRS, so that the stimulation parameters governing tDCS may be set dynamically to enhance user cognition based on current mental state and task demands. To do this, the first step is to determine how long it takes for tDCS to register cognitive effects and how long these effects last. This chapter presents an experiment that investigates the temporal dimension of tDCS for this purpose. The findings suggest a long lag-time between the onset of stimulation and any measurable cognitive effect, which may prohibit the effectiveness of tDCS in a brain-adaptive application.

6.1 Bidirectional BCIs

Computers support several methods for communicating with the user, but currently these output methods are constrained by users' sensory channels. Non-invasive brain stimulation techniques, such as transcranial direct-current stimulation (tDCS), might transcend this limitation. Evidence in the psychology literature suggests that tDCS can temporarily enhance or emphasize aspects of user cognition [24] without imposing a health risk [19]. tDCS delivers a weak (1 to 2 milliamp) electrical current to the exterior of the subject's scalp through an electrode, taking the path to the nearest cathode, which has been carefully placed so that the current will enter and alter particular regions of the subject's brain. tDCS has been used to treat depression [99], as well as enhance language learning [48], working memory [49] and attention [53]. With the introduction of tDCS to the standard output arsenal of HCI, an interactive system may be able to judiciously enhance these abilities depending on the circumstance and user state.

The study is aimed at a future user interface that uses brain measurement as input and responds not only with the usual screen output but also by sending output directly to the brain, suggesting a primitive bidirectional brain-computer interface. Previous systems have measured brain state with fNIRS for communication from user to computer [6, 5, 121]; we now explore tDCS for the opposite direction in a bidirectional brain-computer interface, with fNIRS or another brain monitor as input, and tDCS as output. Consider the brain-adaptive UAV system in Chapter One [6], where the system responded to a spike in cognitive workload by decreasing operator workload. The bidirectional version we propose here would apply tDCS stimulation briefly, precisely when the measured workload increases and only for the duration of the workload spike. For such a bidirectional brain-computer interface to work in practice, the lag time between stimulation and its result should be short. However, much previous tDCS research, especially experiments aimed at treating depression [115], have emphasized longer term effects and longer stimulation periods, because interactivity was not the goal.

To proceed with an interactive system, the key question is to determine how long it takes for tDCS stimulation to register cognitive effects and how long these effects last. We investigate that in this paper with two experiments. At present, there are many unknowns regarding the relationship between settings of the device and its associated cognitive effects, making it difficult to gauge whether the device warrants study and inclusion in next generation user interfaces. An alarming percentage of experiments [71] fail to elicit significant improvements to user performance. The consensus is that results vary across person, possibly because each individual has a different brain and unique rules for how to conduct brain stimulation. However, deciding to abandon tDCS for that reason is premature, because the device has not yet been studied interactively. The missing ingredient for effective tDCS may in fact be a two-way digital ecosystem in which settings can be dynamically adjusted based on their judged, subject-to-subject effectiveness.

In this chapter, I evaluate the feasibility of a tDCS-based bidirectional brain-computer interface. I present two experiments aimed at evaluating temporal properties of tDCS by estimating performance changes in a visuospatial n-back task over a 15 minute time-course. In the first experiment, I compare 5 minutes of tDCS stimulation to a placebo condition; and in the second experiment, I compare 10 minutes of stimulation to a placebo condition. I evaluate changes in reaction time and accuracy for each minute of the experiment.

6.2 Transcranial Direct Current Stimulation

While introducing tDCS brain stimulation into HCI raises safety and ethical questions, research to date has shown that when stimulation does not exceed 2 milliamps and lasts shorter than 40 minutes, there have been no cases of irreversible injury caused by tDCS in a sample of 33,200 sessions [19]. Compared to other brain stimulation techniques, tDCS is easy to use and potentially inexpensive; it already supports a do-it-yourself community [46]. Although experiments typically use a more advanced setup, the basic device consists of just two electrodes and a battery to energize them. Direct current is then administered through a saline-soaked sponge or rubber electrode with conductive gel on the subject's scalp. In a typical setup (and the one used in this experiment), one electrode is placed over the target of stimulation, initiating a path for the current to take to a second electrode placed somewhere nearby. The current is presumed to alter the cortical excitability of the neurons it interacts with, either depolarizing their membranes and making the neurons more likely to fire in the case of anodal stimulation, or hyperpolarizing the membranes, making the neurons less likely to fire in the case of cathodal stimulation [94].

Given that working memory, compulsivity, and attention are often impaired in individuals with Attention-deficit/hyperactivity disorder (ADHD) [96], tDCS has been explored as treatment for individuals with this condition [25]. Experiments aimed to enhance working memory typically administer anodal stimulation to the left dorsolateral prefrontal cortex (dlPFC) at the site F3 (in the International 10-20 system [100]) and allow current to flow through a reference electrode at a symmetrical location on the brain's right hemisphere at site F4 [133] (see Figure 6.1). Many experiments have used this montage to enhance performance at an n-back test [24]. The present experiment makes use of the same montage and n-back paradigm, except I investigate shorter stimulation periods and track performance on a minute-by-minute basis in order to evaluate the usage of tDCS in an interactive system.

6.3 Experiments

6.3.1 Equipment

For measuring brain activity, I used the *multichannel frequency domain Imagent fNIRS* device from ISS Inc. (Champaign, IL) to acquire brain data. It uses two probes, each with four light sources emitting light at 830 and 690 nanometers, and detectors located between 0.5 and 3.5 centimeters away from these sources. Sampling frequency was set to 11.79hz.

For altering brain activity, I used *Soterix 4x1 HD-tDCS multi-channel stimu-*

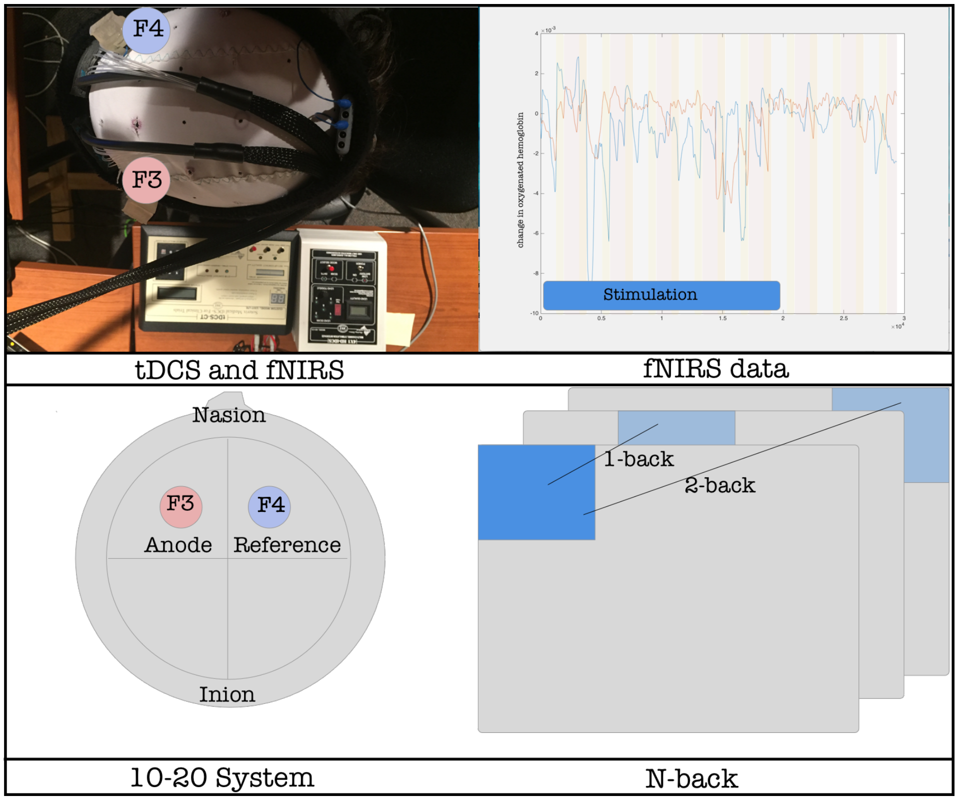

Figure 6.1: The n-back task, pictures of the device, and one subject's fNIRS activity

lation interface (*model 4X1-C2*) to pass electrical currents and the *Soterix tDCS-CT* (*model 1507-LTE*) to control stimulation and placebo according to a double-blind protocol.

6.3.2 Experiment 1

Nine undergraduate college students (5 female) participated in the first experiment. They were monetarily compensated and gave consent at the beginning of the experiment. A university Institutional Review Board approved the experiment. The experimenter explained the visual n-back task (Figure 6.1) on a whiteboard, and let the user practice two trials of the 1-back and two trials of the 2-back. For the 1-back, the user hit the left arrow key if the visual arrangement matched the previous one and the right arrow key otherwise, and for the 2-back they indicated whether or not it matched what they saw 2 iterations ago. These keys were marked with 'YES' and 'NO' with tape on the keyboard. This task was implemented with *Neu-*

Table 6.1: Differences in N-Back Accuracy and Reaction Time for each Minute of Experiment 1

	Percent Accuracy					Reaction Time (milliseconds)				
	Sham		Real			Sham		Real		
min	**mean**	**std dev**	**mean**	**std dev**	**p-value**	**mean**	**std dev**	**mean**	**std dev**	**p-value**
1	74	24	77	22	0.82909	930	247	808	209	0.4575
2	82	34	93	10	0.58809	701	216	852	250	0.36142
3	80	39	90	14	0.65056	770	261	880	41	0.43568
4	81	28	98	5	0.27549	808	310	811	40	0.98716
5	84	29	88	10	0.84721	651	189	832	166	0.17768
6	82	34	93	15	0.60227	744	201	728	167	0.90031
7	78	32	100	0	0.2236	681	224	691	96	0.9418
8	83	28	93	10	0.52828	669	198	718	149	0.69088
9	80	33	87	13	0.70177	631	218	699	146	0.61433
10	82	29	98	5	0.33798	655	230	651	73	0.97642
11	84	35	95	10	0.57964	637	182	723	109	0.43732
12	80	39	89	16	0.67354	658	159	763	93	0.28151
13	81	28	90	8	0.54059	666	273	620	81	0.75824
14	84	35	90	8	0.76657	590	168	704	149	0.32322
15	78	26	92	5	0.34072	624	201	715	187	0.50996
m	81	31	91	4	0.53000	694	209	746	52	0.65

racle software package for the purpose of recording reaction time and dynamically labeling fNIRS data. After these practice trials, the experimenter fit the user with tDCS and fNIRS. This entailed first measuring the size of the subject's head and selecting between four cap sizes, and then placing one gel-covered anodal electrode at site F3 and the other reference electrode at site F4 [100], and then connecting the electrodes to the *Soterix* device (see Figure 6.1). Next, the experimenter placed the two fNIRS probes as near as possible to those sites. (I do not report on any fNIRS data in this paper for experiment one or two because I was unable to discover stimulation dependent patterns).

The subsequent experiment proceeded in two phases. In the first phase, subjects alternated between 30 seconds of the 1-back and 30 seconds of the 2-back, performing each task 7 times. This served as practice as well as the opportunity to group participants by the separability of their fNIRS data. In the second phase, the subject alternated between 40 seconds of the 2-back and 20 seconds of rest, repeating this 15 times for a total of fifteen minutes. In the n-back task (for experiment 1 and

2), a new stimulus appeared every 3 seconds, and accuracy and reaction time for the 40 second task was therefore based on the average of 13 responses.

The experiment used a between subject design. Prior to the experiment, the participant had been placed in two groups: four in the real tDCS group and five in the sham group, and neither experimenter nor subject knew the groups. The real group received 2 milliamps of anodal stimulation at site F3 for 5 minutes. The sham group received 2 milliamps of stimulation only for 30 seconds, a standard placebo, since subjects tend to sense when the device turns on but forget about it when it has been on for a while [24]. Participants began the experiment in parallel to onset stimulation. Afterwards, the experimenters removed the equipment from the user and debriefed them.

Table 6.2: Differences in N-Back Accuracy and Reaction Time for each Minute of Experiment 2

	Percent Accuracy					Reaction Time (milliseconds)				
	Sham		Real			Sham		Real		
min	**mean**	**std dev**	**mean**	**std dev**	**p-value**	**mean**	**std dev**	**mean**	**std dev**	**p-value**
1	93	7	95	6	0.3028	716	167	719	219	0.9216
2	95	7	95	5	0.7184	851	260	799	203	0.2297
3	94	11	92	9	0.7229	799	193	706	222	0.0629
4	95	8	94	7	0.7001	770	260	783	341	0.8458
5	93	12	94	11	0.5162	731	208	669	205	0.1777
6	91	10	88	15	0.4098	793	249	843	355	0.7569
7	93	10	95	8	0.4657	684	157	684	154	0.7898
8	90	13	95	7	0.1775	751	278	709	217	0.2580
9	88	9	97	6	0.0034**	741	204	639	185	0.0323*
10	94	9	91	10	0.2456	749	230	771	313	0.9678
11	90	9	95	8	0.0631	720	177	628	234	0.0975
12	99	5	95	8	0.2519	762	299	707	277	0.2690
13	93	7	91	12	0.7966	649	157	626	193	0.5743
14	96	7	89	20	0.1930	701	211	715	270	0.9637
15	88	12	89	9	0.7260	624	144	595	155	0.2914
16	94	11	98	4	0.1944	640	198	710	314	0.4852
m	93	7	93	5	0.5049	746	185	706	223	0.3618

Results: I have summarized the results of the first experiment in Table 6.1. There were no significant effects for the 5-minute stimulation, although stimulated users trended towards better accuracy and the control group trended towards faster speed, hinting more at a speed-accuracy trade-off than cognitive enhancement. Table 6.1 shows the mean and standard deviation of the participants' mean accuracy and reaction time for each of the fifteen trials under both sham and real conditions, as well as the probability that these averages differed between sham and real conditions in an independent t-test. Without a clear indication that 5 minutes of stimulation exerted significant improvements to user performance, we modified our design and conducted a second experiment.

6.3.3 Experiment 2

Fourteen college students (4 female) participated in the second experiment. Based on the lack of significant results in the first experiment, we increased stimulation time from 5 to 10 minutes, and used a within-subject design so that all participants received both real and sham stimulations. Participants alternated whether or not they received real stimulation first, and both experimenter and subject were blind to this information. To allow time for both conditions, we removed the initial fifteen minute practice period, and participants alternated between 1-backs and 2-backs, starting with the 1-back. For both real and sham stimulation, participants thus completed 8 sets of 40-second 1-back and 8 sets of 2-backs with a 20 second rest in between. In total, each condition lasted sixteen minutes, separated by a five minute break. Because interference from hair prevented fNIRS measurement in the first experiment, we placed the two fNIRS probes on the user's forehead. Apart from these changes, the second experiment proceeded identically to the first.

Results: I have summarized the results of the second experiment in Table 6.2, which is arranged identically to Table 6.1, and illustrate changes in accuracy in Figure 6.2 and changes in reaction time in Figure 6.3. Overall, tDCS did not significantly improve either n-back accuracy or reaction time after 10 minutes of stimulation. However, there was a significant improvement to n-back accuracy during the last minute of stimulation. For minute 9-10, the mean accuracy of the 1-back in the sham condition was 88% (std = 9) and the mean accuracy in the 10 minute stimulation condition was 97% (s = 6) (N =13, p = 0.0034 in a *paired sample t-test*).

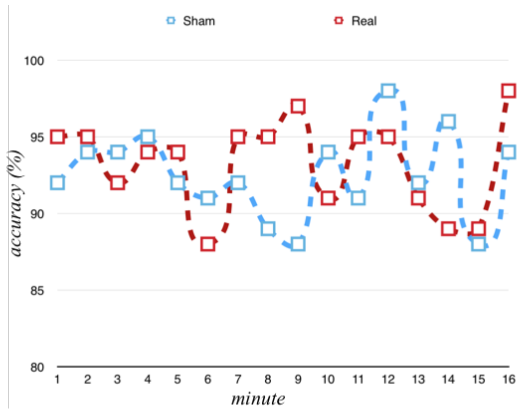

Figure 6.2: Changes in percent accuracy over time, recorded at the end of each minute.

At minute 9-10, improved accuracy in the 1-back did not come at the expense of speed. In fact, reaction times in the 10 minute stimulation condition (m = 639 ms, s = 185 ms) were significantly faster than reaction times in the sham condition (m = 741 ms, std = 204 ms) (N =13, p = 0.0323 in a *paired sample t-test*).

Note that since 1 out of 20 tests should be significant with a threshold set to 0.05, it is hard to verify whether variation has occurred due to chance or not. If significance thresholds are modified according to a Bonferroni correction, then the new threshold is 0.05 / 16 = 0.003125 since there are 16 tests, and neither accuracy nor reaction time are significantly better in the stimulation condition than in the sham condition, although accuracy at minute 9 misses Bonferonni corrected significance by less than 0.0003. There are two reasons why the results between minute nine and ten could be regarded as more valid. First, significance occurs at the very last minute of stimulation and not in a more random minute during the ten stimulation minutes or five non-stimulation minutes. Second, the two dependent variables exhibiting a statistically significant effect according to non-conservative statistical thresholds refer to the same minute, which is improbable unless there was a true effect driving enhancement at this minute, especially given the expectation of a speed accuracy trade-off.

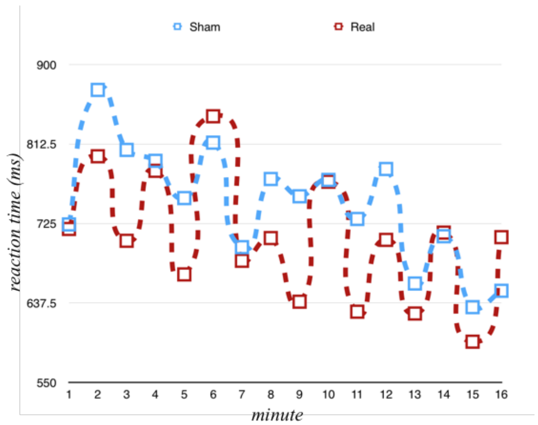

Figure 6.3: Changes in reaction time measured in milliseconds, recorded at the end of each minute.

6.4 Discussion

According to these results, tDCS requires at least 9 minutes of stimulation in order to register an effect. Whether or not effects escalate beyond 10 minutes is an interesting investigation for future work. For present purposes, the delayed response between stimulation and effect implies that fNIRS-adaptive stimulation using tDCS may not work effectively. In the interactive application that motivated the design of this experiment, a subject would perform a computer task under the interrogation of fNIRS measurement, and tDCS would apply stimulation to the user when brain activation measures indicated that cognitive workload had increased. The results indicate that the user would need to wait at least 9 minutes before enjoying a boost to cognition, and a brain-adaptive deployment of the technology would therefore be applicable to tasks with a time span in this range. This is feasible in practice, but less amenable to study in an experimental setting. The negative result demonstrates the **tDCS thesis** of the dissertation.

tDCS thesis: *The delay between the administration of tDCS and measurable changes in user performance exceeds the short timespan between stimulation and effect needed to establish a feedback loop in a bidirectional BCI.*

It is not clear why it takes 9 minutes of stimulation for behavioral effects to register nor whether this limitation disappears given better settings to the device.

Individual differences in skin texture, bone density, and brain structure may imply that standardized stimulation protocols fail to appropriately customize to any given subject. If that is the case, better settings to device parameters such as intensity, polarity, duration, and probe location could be discovered and change based on simultaneous brain measurements [93].

As described in Chapter Four, I envision a design in which fNIRS could monitor the relative activation of the user's task-positive and task-negative networks, which oscillate in inverse correlation to each other depending on whether or not the user is sensorily immersed or in an introspective mode of cognition [108]. The back-and-forth activity of these networks could be monitored; a bidirectional brain-computer interface might discover how to stimulate the user's brain in order to maximize task-positive immersion and minimize task-negative introspection when warranted. A first step in this direction would be to evaluate whether or not fNIRS detects short term neurobiological reactions to stimulation. I attempted such an investigation in this experiment, but in the first experiment hair prevented our device from appropriately measuring the targeted F3 and F4 nodes. I note that other fNIRS devices (such as *Hitachi ETG 4000*) can solve this problem. In the second experiment, when probes were placed approximately 3 inches from the site of stimulation, I did not observe any obvious fNIRS patterns separating the stimulation and real conditions. However, the experiment found no severe limitations preventing the two devices from being used in concert. Our target is an interactive system in which real-time fNIRS measurements are used to modify the tDCS stimulation parameters for better effectiveness. Our experimental configurations and results present a first step in support of such a bidirectional brain-computer interface.

Because of the lag time between the onset of stimulation and any measurable cognitive effect, research into bidirectional brain-computer interfacing might instead focus on stimulation modalities with a more immediate impact on the user's mental state, which is the focus of the next chapter.

Chapter 7

Autobiographical Postscript

This chapter is not written with the same *scientific* rigor as previous chapters. I submit its contents as an instance of the new methodology needed to support research into bidirectional BCIs in the physical paradigm, and leave a proper specification of the methodology needed for the study of how to measure and induce desirable mental states to future work. The case studies in this chapter depart from traditional research in their emphasis on these seven attributes.

1. **Qualitative feedback** as drawings, speech, and emotions conveyed through empathic channels.

2. **Flexible experimental protocols** whose procedures do not intervene with a natural flow of experimenter-participant interactions.

3. **Self experimentation** prior to the study of others, so that the author is also a participant.

4. **Use of subjective voice** to describe methods and results.

5. **Participant-centered experiment design** where the primary goal is to produce results that resonate with the participants at the expense of quantitative metrics that can be exported to a scientific audience.

6. **Empathic experimenters** who are tuned to the emotions of the participant as opposed to a predefined experimental script.

7. **Long term participants** who learn over time to introspect and describe dimensions of their mental life (referred to in chapter 5 as biological neuracles).

7.1 Music Massage as Stimulation in Bidirectional BCI

The problems of cognitive alteration via electrical stimulation and music may have similar computational solutions, and should thus be studied in concert. In both cases, variables controlling procedures that output physical events to the brain (as current or sound) need to be configured such that stimulation is both safe and beneficial to the recipient. For tDCS, these variables describe the location of electrical probes applying current of a given intensity and polarity for a duration of time. For music, the variables of interest govern a procedure for generating an array of decibel amplitudes in the frequency domain.

The collective work of music theory describes rules for producing harmonious sound, restricting the large space of sound possibilities. For example, a given sequence of sounds should be organized around some tonic note (or frequency series) known as the key of the song, and harmonious sounds are mathematically related to this fundamental frequency according to some scale of intervals (e.g., the major or minor scales). Engineers have encoded these rules into digital audio workstations, enabling an opportunity for auditory bidirectional brain-computer interfacing if the song is augmented with an interface that allows it to branch between different versions depending on implicit input from sensors measuring the user's physiology.

The previous chapters have advanced a Bayesian model, which models brain activity as a series of hierarchically layered attempts to optimally compress and learn from sensory data. The brain uses existing models to predict the content of sensory signals, and propagates information that violates expectation up cognitive hierarchies where existing models are modified [50]. Music – or sound which obeys mathematical patterns – may exist as a happy coincidence of the brains proclivity to direct computation (and associated conscious experience) towards stimuli that engages its predictive machinery [73]. By this reasoning, the state of the brain

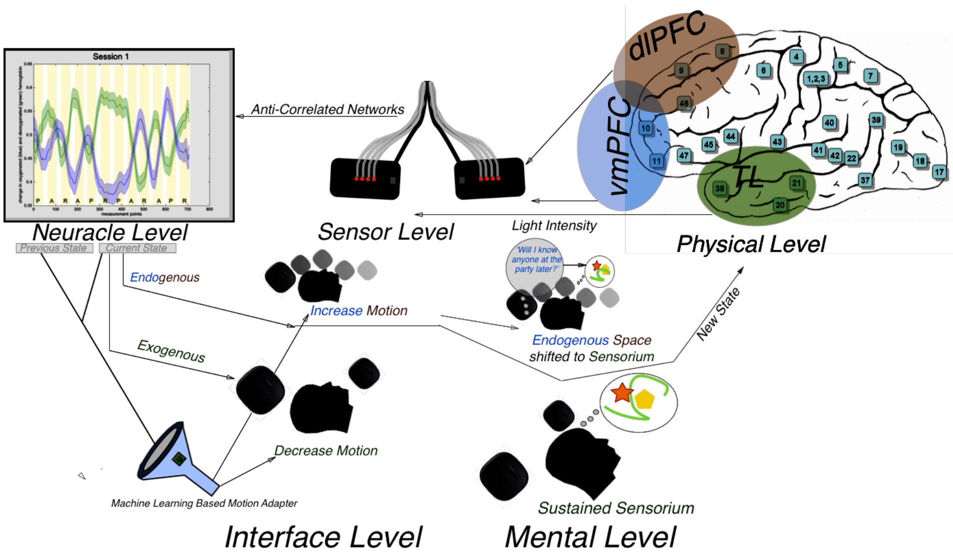

Figure 7.1: Motion-adapting musical bidirectional BCI

and attention can be modulated by manipulating the user-relative predictability of sound. This chapter argues that the inherent motion, or changing 3D orientation of a sound relative to the ears, is an ideal target for adaptation in a bidirectional BCI, as the **music massage thesis**.

7.2 Music Massage

Music massage is a group activity, invented by Naomi Hashimoto, that requires music played from speakers, a music masseur, and a *music listener*. A typical music massage setup, and the one evaluated in this paper, includes two music masseurs with a *stereo* music speaker setup.

A music massage begins when music masseurs give listeners a menu of pens in ten different colors. Each color corresponds to a desired mood, and *listeners* are told to select the mood they wish to experience. For example, in our color coding, light blue corresponds to a relaxed, calm, serene feeling. Next, *listeners* are asked to assume a comfortable sitting or lying position, and to close their eyes. The *masseurs* queue a song in the selected color coding on the *music player*, and take one speaker each. When the music begins, the *masseurs* move the speakers slowly around the listener's head at the pace of the music. When the experience is over, *listeners*

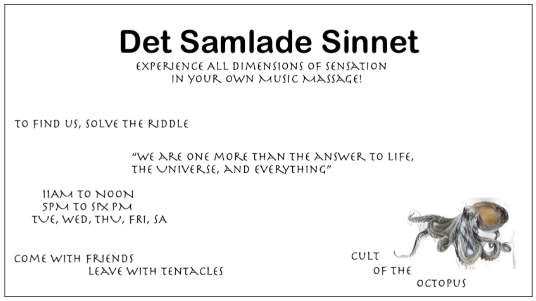

Figure 7.2: A ticket to receive a music massage

typically report a rich and unusual mental experience (entropy).

It is not clear how to study the mental experiences that emerge during music massage. In this chapter, I describe a multi-part casy study in narrative form, leading towards a more formal study of music massage.

7.3 Music Massage at an Art Festival

The case study began at a music and art festival in the woods, where Michael Gabour and I had registered an art performance. We set up a small stage which included a hammock off a path in the woods, away from most camps in the festival.

Recruitment: We recruited attendees as we explored the other events of the festival, handing out tickets (figure 7.2) that included our open hours as well as a riddle that disclosed our location if solved. A riddle induces higher levels of consciousness in brains by the same principle as a music massage. Just like high information sound demands vast networks in the brain in order to decode and render the musical content in the listener's sensorium, a good riddle traverses many networks before its resolution, coercing a richer and more engaged brain state than simply disclosing our camp's location.

Methods: When attendees arrived at the camp, we gave them a menu of colors, symbolized through pens, which denoted different types of music.

1. Light blue: relaxed, calm, serene

2. Dark blue: melancholic, pensive

3. Light green: natural, wild, instinctive
4. Dark green: chilled out, easy going, socially interested
5. Yellow: warm, snug, glowing
6. Orange: playful, creative, childlike, curious
7. Purple: inspired, full of fantasy, dreamy
8. Brown: assertive, confident, energized
9. Red: passionate, romantic, loving
10. Black: destructive, wicked, animated

To facilitate choice of color, attendees were given a guestbook which included examples of songs in different color categories. I chose songs across many genres specifically for the music massage. Appropriate songs contain an implicit motion pattern which the music masseur detects and renders in a 3D-space as though dancing. The guestbook contained references to my own collection of music satisfying this criteria at *(Spotify/Sam Hincks)* and *(Spotify/Samulus)* as well as music from more established artists like *Hans Zimmer*. When they had chosen a color, they were asked to log personal information in that color to the guestbook. Six columns requested 'a name', 'if you like, an email', 'some movie', 'a musician with spirit', 'a song everyone knows', and 'something the brain does'.

In total, twenty attendees added their names to the guestbook: 2 purple, 2 light blue, 3 red, 3 light green 3 orange, 3 black, 4 dark blue. After the attendee added information to the guestbook, we directed them to the hammock, and asked them to put on a blindfold. We played a song in the selected color genre, and we began to move a speaker slowly around their head at the pace of the music. At the climax of the song, we sprayed scented water in order to create a richer sensorium for the *listener*. When the song finished, we gave them the pen they had selected, and asked them to contribute a drawing to the guestbook.

Figure 7.3: Drawings from two attendees after a music massage. (Left illustration: by Alerisa Rose)

Results: In this performance, we did not attempt to quantify the effect of the music massage. The feedback suggested that attendees greatly enjoyed the show. By the third day, demand for a personal music massage outstripped supply, and we were forced to give a music massage to between fifteen and twenty listeners simultaneously. Some attendees added messages to the guestbook, including

1. 'I learned to pay attention to the serious/playful spectrum of a thought as a thought itself'

2. 'I think this is the best thing I've ever drawn."

3. 'Peaceful! Energized! Strong! Proud!"

4. 'This is the best thing that's happened to me all week!"

5. 'Fearless. The only word I can use to describe this experience. If you are asking for a multisense experience, you've come to the right place. Great people. Great food. Mind opening gate to who knows..."

Attendees also added great art to the guestbook that may have reflected their energized states (see figure 7.3).

7.4 Music Massage on a Synaesthete

The next pilot was focused on one individual, *(JG)*, who has mild synaesthesia between her visual and auditory senses, a condition that causes the sensorium to produce visuals for sound. I learned she had this condition after giving her a music massage. She picked the color orange, and I played an electronic song from my own collection called Peco *(Spotify/Sam Hincks/Peco)*. After the *music massage*, she reported seeing a dark tunnel, leading to a sunset. Before this experience, I had used the colors as conceptual bindings between emotion and information in different senses. But I realized after this music massage that there was an opportunity for an unusual scientific study, testing an interesting possibility for a *bidirectional BCI*.

Does a song which I label as a particular color (e.g orange) stimulate that color with greater likelihood than songs in other colors?

With this question in mind, I color-categorized 83 songs from my own collection, and music-massaged 19 of them for *JG*, without telling her in advance my color hypothesis. At the end of each song, I asked what she saw in her mind's eye.

1. Trainbow: *This song was lightblue and green, and I saw grass and wind.*

2. Ocean Yawning: *This song was dark green and purple. I saw a dead forest with black and white fog. I saw a mountain with one tree, and fog came like a river. At the end, it was just deep green.*

3. Fanguar: *I saw big bowl of black and yellow, but it disappeared against my will. Violet appeared and I saw big waves of dark blue.*

4. February Fanguar: *Thank you for the light. Red came as a big bird flying into yellow bowl. They were playing. For a moment, I saw grey as fog. Light came from piano at start.*

5. Eerier Moment Generator: *This song had dark blue as clear dominant color, but no scenes.*

6. The Destormus: *I saw some red at the start, but no color later on. It was like a movie. A storm came with lightning everywhere. Clouds were above my head, and there were shapes of birds, like Egyptian Gods.*

7. Evolving Shades: *It was yellow for sure, the whole time, with some light blue and then forest and fog. Imagine a clock covered by a forest, moving around. As it approaches, it gets darker. There were squares of yellow surrounding the clock.*

8. Crustacean: *There was no color in this song.*

9. Revenge: *I saw a yellow light and a camera.*

10. Demogorgan: *Most of the time, there were no colors. Towards the end, I saw a black mountain.*

11. Prediction Error: *I saw a human with open arms, a distance away, but then when I came closer I saw it was a scarecrow in black and white. There was light blue light but no specific shapes.*

12. Lonely Tentacle: *There were no shapes, but I saw light blue and orange. There were two screens, and in one screen I thought I had heard it. In the other, I saw the village in the evening, and a constellation of stars.*

13. Rugged Innards: *This was a light shade of red. I saw mountains with peaks. Later, a package with a ribbon.*

14. Ganorak: *This was yellow. In the beginning, I saw a tiger.*

15. The Cottage: *First part I saw violet and fog everywhere. Then my default mode network came on for a while, and I had thoughts connected with people I never met. In the last part was a room with greek sculptures that were hugging.*

16. Disturbed Drone: *It was yellow and brown at start, with a light blue bird.*

17. Filthy Lacrimosa: *At the beginning, there was yellow, and a bowl of red, with a little orange and light blue.*

18. Eerier Moment Generator(2nd time): *There was so much light blue. There was a gate, and there were some sculptures, and it was windy, and the wind was violet. And then I saw a yellow boat on the lake, and I saw pink columns.*

19. Evolving Shades (2nd time): *It was yellow and light red.*

20. Firefly Forest *It was light red, with cocoons in all colors, and then blue, with beautiful ceramic patterns, very detailed. And then black clouds, with faces of dead people.*

21. Living History *There were so many elements so it was hard for me to focus. I saw some yellow paths. There was some intense blue at the end.*

22. Mean on Purpose (2nd time): *The whole song I saw dark green, and a space, where I would like to do movements. I saw a valley, and a city.*

23. Trainbow (2nd time): *In the beginning, there was no color. I saw a face of an old man engraved in a tree. Then I saw a monument. Then I was in the default mode network for a while. I saw some light blue in the middle. Then at the end, there was just a little bit of orange.*

24. Rugged Innards (2nd time): *I saw a small light in front of doors of a palace in a distance. It had towers, that were high and sharp. Then my default mode network came online. Then I saw a red and yellow eagle, and then a butterfly in the same colors. And finally, a hawk.*

Results: It is difficult to draw quantitative conclusions about the relationship between the color-coding of my songs, and the actual color experience in *JG's* blindfolded sensorium. Many of the songs produced multiple colors and scenes for *JG*. The astonishing data-point was that 18/19 songs induced rich color experiences for *JG*. In 23/24 cases, the music massage produced a visual-auditory sensorium that was practically free from default mode network activity.

Discussion: The low level of default mode network activity in this pilot may not merely be the consequence of the information density in the *music massage*. *JG* may be an anomaly in her level of sensory presence. *JG* insisted that she was not an anomaly, but instead obeyed a strict code, controlling the level of stimulation to her brain. For example, she seldom drinks coffee or alcohol and doesn't use social media. *JG* described how to design life properly, in the color-coding she had learned from my songs, from reading notes for this dissertation, and from our personal conversations. If you want to see, she said, you must:

1. Use the rational brain to decide what's important.
2. Use the default mode network to imagine.
3. Use the meta-reflective brain to let go of thoughts.
4. Behave naturally.
5. Shine your light on the world.
6. Feel love.
7. Abstain from lying.
8. Find people you trust.
9. Play like a child.
10. Acknowledge and control the darkness.

I believed I was already wielding these networks correctly, and so the morning after she told me this, as I was meditating, I started to look very intensely at my dark eyes-closed sensorium. For ten minutes, my mind's eye was black, as it normally was. But then I saw flickers of light move, creating the sensation that I was floating in space. The lights gathered together, and then I felt a huge explosion, as though I was looking at the big bang, and I ended up lying on a big field of grass at night, looking at the dark silhouettes of trees beneath a deep blue sky full of stars. That same sky full of stars now re-occurs frequently when I close my eyes and am able to silence by default mode network.

I had my first synaesthetic episode a month after my first eyes-closed color experience. This time *JG* was giving me a music massage in a sauna. The heat exposure and entropy of the sauna probably dialed down my naturally high level of default mode network activity. As my eyes were closed, my sensorium visualized a clear light blue pool with detailed orange and green patterns at the bottom. When the song reached its climax, *JG* poured water on the sauna's rocks, producing heat and steam in the room, as well as loud white noise that released energy in the song. In my sensorium, the previously still pool became disturbed with drops of water splashing on the surface, producing ripples. I started this part of the music massage investigation painting sound pictures in *JG's* sensorium, and she finished it by painting pictures in mine.

The theoretical ideas from Bayesian cognitive science, the testimony of attendees at the art festival and the rich color experiences of *JG* demonstrate that *motion in sound* functions as an auxiliary parameter to music that can be adjusted to manipulate the degree to which it affects brain activity. I intend to move forward with a more formal scientific investigation of music massage. But for the same arguments that support the *methodological thesis*, I am held back by the problem of studying individual phenomenology on a random population, using quantitative metrics. New methodology is needed.

Chapter 8

Conclusion

The chapters of this dissertation have unfolded the essential theoretical/empirical foundation (Chapters 1, 2, and 3), algorithms (Chapter 4), software infrastructure (Chapter 5), methodology (Chapter 5) and interfaces (Chapter 6 and 7) for building bidirectional BCIs. The result is the specification of a technology for altering the *temperature level of consciousness* in a brain by adjusting information in a bidirectional feedback loop with a measurement of *fNIRS-detected* anti-correlated networks. If the theoretical foundation is accurate, the reciprocal see-saw relationship between oxygenation levels in the prefrontal cortex and temporal lobe reflects competitive activity between two independent computers, and fluctuations in this signal reflect the degree to which cognitive Energy crystallizes a more endogenous versus exogenous *orientation* and top-down versus bottom-up *directionality.* A bidirectional BCI can control its user along these dimensions by controlling the predictability of information to the brain (such as the motion of music), using physical neuracles to score the real-time effect of brain stimulation and optimize the *temperature* of the brain.

To reach this conclusion, the dissertation has opened several threads that are bound together by the *unity thesis.* Entropy, the mathematical formulation of order and chaos, gives a principled way to understand the meaning of fluctuations in anti-correlated networks, providing the foundation for physical neuracles, and a path towards a physical paradigm for BCI.

Unity thesis: *The spectrum of states in between the brain's stable and novel*

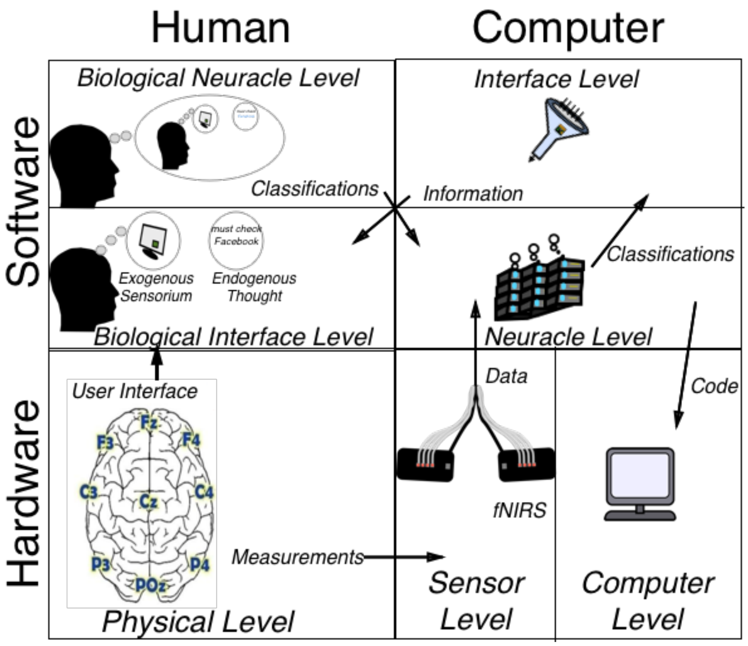

Figure 8.1: Physical Paradigm for Bidirectional BCI

*configurations have differentiable physical neural signatures (**human-hardware component**) that are mirrored in an introspectively observable mental workspace (**human-neuracle component**), which can be monitored by neuracles classifying physiological changes associated with the activation of neural networks (**computer-neuracle component**), and which can be controlled by adjusting the amount of information transmitted to the brain in a user interface (**computer-interface component**).*

8.1 The Unity Thesis

The *unity thesis* has four components, and figure 8.1 illustrates how these components can be organized along two axes: the first portraying whether or not the component exists at a hardware or software level, and the second resolving whether the component pertains to the human or the computer part of the cybernetic loop.

The **human-hardware** component of the *unity thesis* is a straightforward consequence of the resistance to the *second law of thermodynamics* necessary to the brain and other biological systems that find stable equilibria with the environment for the duration of their existence. Any salient object of the Universe erects a *Markov* *blanket* shielding an *internal memory space* from the environment. Those objects that survive and replicate necessarily contain *internal memory spaces* that predict the causes of inputs registered in the *Markov blanket.* They resist entropy stimulated by the inputs by modifying the *internal memory space* for better prediction over time, as well as by changing *output effectors* so that inputs better agree with a model of the world in which the object is anticipated to survive and replicate.

In humans, the *exogenous internal memory space* is a consciously introspectable *user interface* in five primary sense modalities, called the *sensorium.* The sensorium generates an illusion (or set of explanations in its hierarchically superordinate endogenous structures) as existing outside the brain in a common space available to other humans. But the *sensorium* is generated from within and from the *top-down.* Approximately half of the time, the signals mediated from the outside world do not contain enough information to demand more than default computation in the *exogenous networks* responsible for producing the *sensorium* [77], and as a result *consciousness* does not fixate there. Instead, the *Energy* fixates in some *endogenous network* where information processing also flows in two *directions* and is at any point either dominated from the *top-down* or *bottom-up* or otherwise at a stable equilibrium between these dueling factions, in which case the energetic level that is witnessed in the *mental workspace* is low. Conscious attention fixates where there is Chaos in humans, but probably also in animals, possibly in plants, and potentially also in 'inanimate' objects like rocks that do not process much information at our time-scale insofar as it manifests in our sensorium.

The **human-neuracle** component of the *unity thesis* asserts that an appropriately trained human (what Chapter Five refers to as a *biological neuracle*) can reliably witness and report upon the current *temperature* level of their consciousness. *Biological neuracles* can further decompose mental states with higher than normal *temperature* levels into *directionality* and *origin* dimensions depending on whether the flow of information proceeds primarily from the *bottom-up* or *top-down* (*directionality*) and is acting on a data source that originates from the environment or the repository of data stored in memory and DNA (*origin*). *Low temperature states* can simply be called *default states*, where processing oscillates haphazardly and automatically between historical memory, present input, and future simulation: between a primitive *exogenous* brain responsible for the sensorium and a quintessentially human *endogenous* brain responsible for linguistic mind wandering.

This model of conscious attention, as Energy that fixates in the network exhibiting most *Chaos*, gives a physical approach for decoding cognitive dimensions (including *temperature*, *origin*, and *directionality*) from the competitive and cyclical activity of anti-correlated networks measured with fNIRS. Chapter Four released a working hypothesis for how to build a *physical neuracle* on the basis of anti-correlated networks discovered in Chapter Three in partial fulfillment of the **computer-neuracle** component of the *unity thesis*.

The algorithm depends on three probes, measuring oxygenation levels at the dorsolateral (dlPFC) and ventromedial prefrontal cortex (vmPFC) and the temporal lobe (TL). The three probes form two anti-correlated networks (*vmPFC-TL* and *dlPFC-TL*). Whichever of the two networks has higher anti-correlation, indicating computation in one node that depends on the other, is assigned as the *master network*. *Directionality* is computed from the setting to *master network*, so that *dlPFC-TL* ← *top-down* and *vmPFC-TL* ← *bottom-up*. *Origin* is set as whichever of the two probes in the *master network* has a higher slope, so that it is set to *exogenous* when the *TL-probe* increases in measured oxygen faster than the node placed at the prefrontal cortex. In Chapter Four, eleven other dimensions are also extracted, including *entropy*, which is the level of surprise with the current settings of dimensions given their history. This thirteenth dimension depends on the twelve other dimensions, and may be the best approximation for the *temperature* of consciousness in a brain.

Because of fundamental difficulties that have permeated science and philosophy since their inception, the proposed dimensions remain hypothetical. These hypothesis are inspired by a multi-disciplinary literature review, an evaluation of fNIRS data collected from hundreds of brains, and parallel introspection and brain measurement using customized software. Chapter Five gives the software program *Neu-*

racle, which is a user interface for rapid hypothesis-testing and self-experimentation. *Neuracle* gives its users a set of procedures for manipulating the state of their brain and observing the associated change in fNIRS signals shortly after the user's brain generated those signals. It gives the tools to rapidly build, refine, and evaluate both physical and agnostic neuracles, and to output their real-time classifications on a port, which brain-computer interfaces can use to adjust aspects of interaction to the user's benefit. My hope is for other researchers to use the software to improve upon the dimensionalization in Chapter Four, and help develop a methodology that supplies a tighter bridge between the software and hardware levels referenced in figure 8.1.

The final chapters focus on the **computer-interface** component of the *unity thesis*. They explore how to control the predictability of input to the brain to augment cognition. The placement of Chapter Six at the end of the dissertation may however be misleading since the design of the experiment does not reflect the knowledge accrued in Chapters Two, Three, and Four. The thinking of Chapter Six is based on the limited brain dimensionalization of Chapter One, needed to support the agnostic paradigm. Chapter Six concluded that tDCS is not a worthwhile method for stimulating the brain towards augmented working memory capacities since there is a long delay before stimulation registers any improvement in *n-back* performance. But tDCS may still prove effective for bidirectional BCIs that augment other aspects of cognition. Electrical stimulation to the brain can relieve symptoms of depression [99], which is not surprising in a *Bayesian* brain. Depression manifests in brains that are stuck in default states, thereby performing little computation, and exhibiting low Energy. A jolt of electricity is another semi-random piece of information for the brain to process, destabilizing its default state. A revised tDCS study would relax the need to obtain measurable performance benefits and focus on measuring dependent variables associated with the *temperature* of mental experiences for the participants.

Embracing a qualitative approach for studying bidirectional BCIs, Chapter Seven examined music as sound information to the brain that occupies a delicate and aesthetic sweet-spot between order and disorder and in so doing elevates and defines the *temperature* of experience. The mathematical structure of music, which is designed to absorb the brain's sustained attention, may be a sort of reverse engineering of evolved brain structures that enable the brain to decode the 3D-orientation and motion of regular objects in time and space. (This relationship may be what causes us to dance to music: the sound contains an implicit motion sequence, which the dancer embodies.)

Chapter Seven proposed a new way to embody music, where a *music masseur* moves stereo speakers around a listener's head in synchrony with the sound's inherent motion sequence. The *Bayesian* grounding of this dissertation predicts that music heard through *music massage* dictates more processing in the *exogenous networks* responsible for generating the *sensorium.* Crucially, there is *mutual information* between the changes in the sound's pitch and 3D origin, and so *music massage* does not fragment attention between multiple networks. In the listener's sensorium, *music massage* manifests as a singular object of perception, unified in one large sensory network critically poised between order and disorder. This chapter evaluated *music massage* in two pilot studies: one performance at an arts festival and one focused study on a mild synaesthete.

In these case studies, a biological neuracle approximates the role of the **computer neuracle** and **computer interface components** of a full-fledged sound-based bidirectional BCI, as shown in 8.2. Acting as a *wizard of Oz bidirectional BCI*, the biological neuracle strove to make each exchange with the eventual *listener* provide the right amount of information to unhinge their default mode of operation without making them uncomfortable or self aware. During the *music massage*, the biological neuracle tuned their internal state to the emotions created by the music, presumed to be shared with the listener, and adjusted the intensity of *listener's* sensorium appropriately. When the experience was over, the biological neuracle patiently created conditions for the *listener* to render the experience that unfolded in language or visual symbolism. The art, testimony, and rich sensoria described in Chapter Seven underscores the effectiveness of the *music massage*, the utility of

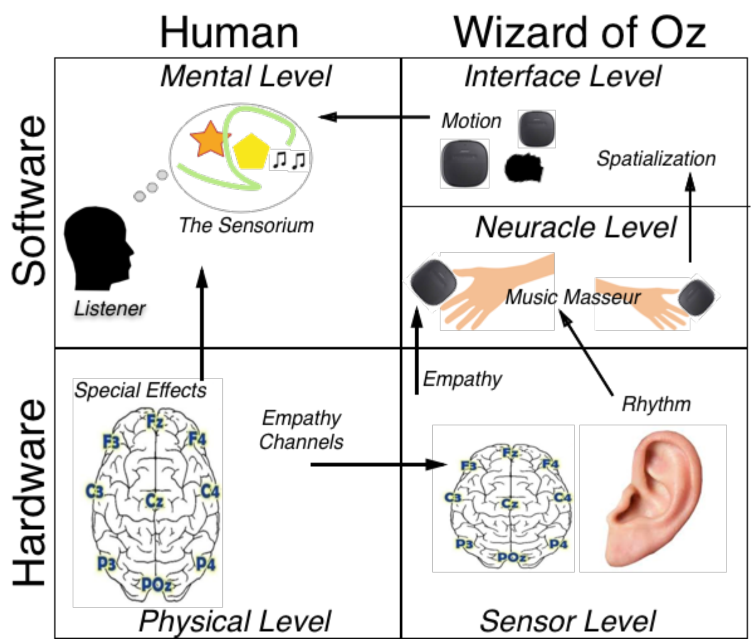

Figure 8.2: Bidirectional BCI Wizard of Oz Experiment

motion as an auxiliary parameter for adaptation in a *bidirectional BCI*, and the **interface component** of the *unity thesis*. Music massage also appears useful for creating more biological neuracles by encouraging listeners to pay attention to and render their sensoria.

8.2 Future Research

A technological implementation of a music massage bidirectional BCI would be able to detect when the *listener's* brain regressed into a default mode of operation. At this point, the information load of the inputs producing the sensorium could increase: *more motion*, *sound closer to the ears*, *release of smells*, and *touch*. Unfortunately, *music massage* does not work as effectively when the sound moves in headphones, demanding many speakers to be arranged in a large ambisonic dome. A wise follow-up experiment would rely on a biological neuracle that bases decisions for how to adapt the *listener's* sensorium on the real-time classification of a physical neuracle.

Out of ethical considerations, any technology that is developed in this paradigm must recognize the ongoing effect of ubiquitous information on contemporary civilization. There is a palpable attention-deficit epidemic, and those of us who have championed effortless and pervasive human-computer interaction have a duty to help reshape the technological landscape. Modern technologies like the Internet, smart phone, and social media are deleterious to human attention for two primary reasons. First, notifications, videos, text messages, and other digital inputs to the brain that successfully usurp its attention pack a high informational punch, elevating the baseline that some natural input to the brain must exceed to grab the full province of the brain's exogenously oriented attention. Second, the extended tribe imagined in a globalized civilization in dialogue through an all-encompassing Internet creates an extended set of concerns for a default mode network optimized for primate style social cognition.

It is unclear whether bidirectional BCIs can help mend the wounds wrought by hyper-connectivity or whether these interfaces will unfold as more digital superstimuli, exploiting frailties in human attention with scientific precision. For this reason, I urge scientific supervision over the technology enabled by the knowledge gathered in this dissertation.

I disseminate the knowledge and software in the hope of creating more biological neuracles. We are tired information processors that have cultivated cognitive and behavioral habits that suppress the *Energy* and strive to unshackle the chains of the default mode network in a quest to feel unity with the *Source*.

Bibliography

[1] Data-driven documents. https://d3js.org/. Accessed: 2019-06-03.

[2] Functional brain imaging system. http://www.iss.com/resources/pdf/brochures/ISS_Imagent_Brochure.pdf. Accessed: 2019-06-03.

[3] Information theory primer. http://users.fred.net/tds/lab/papers/primer/german/2.59/information_theory_primer_de.pdf.

[4] Unity. https://unity.com/. Accessed: 2019-06-03.

[5] Afergan, D., Hincks, S. W., Shibata, T., and Jacob, R. J. Phylter: a system for modulating notifications in wearables using physiological sensing. In *International Conference on Augmented Cognition* (2015), Springer, pp. 167–177.

[6] Afergan, D., Peck, E. M., Solovey, E. T., Jenkins, A., Hincks, S. W., Brown, E. T., Chang, R., and Jacob, R. J. Dynamic difficulty using brain metrics of workload. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2014), ACM, pp. 3797–3806.

[7] Afergan, D., Shibata, T., Hincks, S. W., Peck, E. M., Yuksel, B. F., Chang, R., and Jacob, R. J. Brain-based target expansion. In *Proceedings of the 27th annual ACM symposium on User interface software and technology* (2014), ACM, pp. 583–593.

[8] Afergan, D. A. Implicit brain-computer interfaces for adaptive systems: Improving performance through physiological sensing. *Tufts University Dissertation* (2015).

[9] Austin, J. H. *Zen-brain reflections.* MIT press, 2010.

[10] Ayaz, H., Izzetoglu, M., Bunce, S. C., Heiman-Patterson, T., and Onaral, B. Detecting cognitive activity related hemodynamic signal for brain computer interface using functional near infrared spectroscopy. *2007 3rd International IEEE/EMBS Conference on Neural Engineering* (2007), 342–345.

[11] Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., and Onaral, B. Optical brain monitoring for operator training and mental workload assessment. *NeuroImage 59*, 1 (Jan. 2012), 36–47.

[12] Baars, B., and Gage, N. M. *Fundamentals of cognitive neuroscience: a beginner's guide.* Academic Press, 2012.

[13] Baddeley, A. Working memory. *Science 255*, 5044 (1992), 556–559.

[14] Baddeley, A. D., and Hitch, G. J. Working memory. *The psychology of learning and motivation 8* (1974), 47–89.

[15] Baldwin, R. General-purpose lms adaptive engine in java. https://www.developer.com/java/other/. Accessed: 2019-03-25.

[16] Bengio, Y., Simard, P., Frasconi, P., et al. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks 5*, 2 (1994), 157–166.

[17] Berka, C., Levendowski, D. J., Cvetinovic, M. M., Petrovic, M. M., Davis, G., Lumicao, M. N., Zivkovic, V. T., Popovic, M. V., and Olmstead, R. Real-time analysis of eeg indexes of alertness, cognition, and memory acquired with a wireless eeg headset. *International Journal of Human-Computer Interaction 17*, 2 (2004), 151–170.

[18] Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., Olmstead, R. E., Tremoulet, P. D., and Craven, P. L. Eeg correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, space, and environmental medicine 78*, 5 (2007), B231–B244.

[19] Bikson, M., Grossman, P., Thomas, C., Zannou, A. L., Jiang, J., Adnan, T., Mourdoukoutas, A. P., Kronberg, G., Truong, D., Boggio, P., et al. Safety of transcranial direct current stimulation: evidence based update 2016. *Brain Stimulation 9*, 5 (2016), 641–661.

[20] Billinghurst, M., and Starner, T. Wearable devices: new ways to manage information. *Computer 32*, 1 (1999), 57–64.

[21] Brand, M., Young, K. S., and Laier, C. Prefrontal control and internet addiction: a theoretical model and review of neuropsychological and neuroimaging findings. *Frontiers in human neuroscience 8* (2014), 375.

[22] Brefczynski-Lewis, J. A., Lutz, A., Schaefer, H., Levinson, D., and Davidson, R. Neural correlates of attentional expertise in long-term meditation practitioners. *Proceedings of the national Academy of Sciences 104*, 27 (2007), 11483–11488.

[23] Brouwer, A.-M., Zander, T. O., Van Erp, J. B., Korteling, J. E., and Bronkhorst, A. W. Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls. *Frontiers in neuroscience 9* (2015), 136.

[24] Brunoni, A. R., and Vanderhasselt, M.-A. Working memory improvement with non-invasive brain stimulation of the dorsolateral prefrontal cortex: a systematic review and meta-analysis. *Brain and cognition 86* (2014), 1–9.

[25] Cachoeira, C. T., Leffa, D. T., Mittelstadt, S. D., Mendes, L. S. T., Brunoni, A. R., Pinto, J. V., Blazius, V., Machado, V., Bau,

C. H. D., Rohde, L. A., et al. Positive effects of transcranial direct current stimulation in adult patients with attention-deficit/hyperactivity disorder–a pilot randomized controlled study. *Psychiatry Research 247* (2017), 28–32.

[26] Caputo, G. B. Strange-face illusions during interpersonal-gazing and personality differences of spirituality. *Explore: The Journal of Science and Healing 13*, 6 (2017), 379–385.

[27] Carhart-Harris, R. L., Leech, R., Hellyer, P. J., Shanahan, M., Feilding, A., Tagliazucchi, E., Chialvo, D. R., and Nutt, D. The entropic brain: a theory of conscious states informed by neuroimaging research with psychedelic drugs.

[28] Carhart-Harris, R. L., Leech, R., Hellyer, P. J., Shanahan, M., Feilding, A., Tagliazucchi, E., Chialvo, D. R., and Nutt, D. The entropic brain: a theory of conscious states informed by neuroimaging research with psychedelic drugs. *Psychoanalytical neuroscience: Exploring psychoanalytic concepts with neuroscientific methods* (2015), 140.

[29] Chien, W. T., and Thompson, D. R. Effects of a mindfulness-based psychoeducation programme for chinese patients with schizophrenia: 2-year follow-up. *The British Journal of Psychiatry* (2014), bjp–bp.

[30] Christakis, D. A. Internet addiction: a 21 st century epidemic? *BMC medicine 8*, 1 (2010), 1.

[31] Clark, A. Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences 36*, 03 (2013), 181–204.

[32] Csikszentmihalyi, M. Flow and the psychology of discovery and invention. *New Yprk: Harper Collins* (1996).

[33] Csikszentmihalyi, M. Flow and the psychology of discovery and invention. *HarperPerennial, New York 39* (1997).

[34] Cui, X., Bray, S., Bryant, D. M., Glover, G. H., and Reiss, A. L. A quantitative comparison of nirs and fmri across multiple cognitive tasks. *Neuroimage 54*, 4 (2011), 2808–2821.

[35] Curtis, C. E., and D'Esposito, M. Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences 7*, 9 (Sept. 2003), 415–423.

[36] Delpy, D. T., Cope, M., van der Zee, P., Arridge, S., Wray, S., and Wyatt, J. Estimation of optical pathlength through tissue from direct time of flight measurement. *Physics in medicine and biology 33*, 12 (1988), 1433.

[37] Dennett, D. C. *Consciousness explained*. Penguin uk, 1993.

[38] Deutsch, J., Li, H., and Sharma, A. Microscopic origin of thermodynamic entropy in isolated systems. *Physical Review E 87*, 4 (2013), 042135.

[39] Dey, A. K., Salber, D., Abowd, G. D., and Futakawa, M. The conference assistant: Combining context-awareness with wearable computing. In *Wearable Computers, 1999. Digest of Papers. The Third International Symposium on* (1999), IEEE, pp. 21–28.

[40] Eichele, T., Debener, S., Calhoun, V. D., Specht, K., Engel, A. K., Hugdahl, K., von Cramon, D. Y., and Ullsperger, M. Prediction of human errors by maladaptive changes in event-related brain networks. *Proceedings of the National Academy of Sciences 105*, 16 (2008), 6173–6178.

[41] Fairclough, S. H., Gilleade, K., Ewing, K. C., and Roberts, J. Capturing user engagement via psychophysiology: measures and mechanisms for biocybernetic adaptation. *International Journal of Autonomous and Adaptive Communications Systems 6*, 1 (2013), 63–79.

[42] Fairclough, S. H., Moores, L. J., Ewing, K. C., and Roberts, J. Measuring task engagement as an input to physiological computing. In *Affec-*

tive Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on (2009), IEEE, pp. 1–9.

[43] Fatmehsari, Y. R., and Moussavi, Z. Designing and implementing a portable near-infrared imaging system for monitoring of human's brain functional activity. *Journal of Medical Devices 9*, 2 (2015), 020921.

[44] Ferrari, M., and Quaresima, V. A brief review on the history of human functional near-infrared spectroscopy (fnirs) development and fields of application. *Neuroimage 63*, 2 (2012), 921–935.

[45] Fishburn, F. A., Norr, M. E., Medvedev, A. V., and Vaidya, C. J. Sensitivity of fNIRS to cognitive state and load. *Frontiers in human neuroscience 8*, February (2014), 76.

[46] Fitz, N. S., and Reiner, P. B. The challenge of crafting policy for do-it-yourself brain stimulation. *Journal of medical ethics* (2013), medethics–2013.

[47] Flavell, J. H., Green, F. L., and Flavell, E. R. Development of children's awareness of their own thoughts. *Journal of Cognition and Development 1*, 1 (2000), 97–112.

[48] Flöel, A., Rösser, N., Michka, O., Knecht, S., and Breitenstein, C. Noninvasive brain stimulation improves language learning. *Journal of Cognitive Neuroscience 20*, 8 (2008), 1415–1422.

[49] Fregni, F., Boggio, P. S., Nitsche, M., Bermpohl, F., Antal, A., Feredoes, E., Marcolin, M. A., Rigonatti, S. P., Silva, M. T., Paulus, W., et al. Anodal transcranial direct current stimulation of prefrontal cortex enhances working memory. *Experimental brain research 166*, 1 (2005), 23–30.

[50] Friston, K. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience 11*, 2 (2010), 127–138.

[51] Gevins, A., and Smith, M. E. Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science 4*, 1-2 (2003), 113–131.

[52] Gevins, A. S., and Cutillo, B. C. Neuroelectric evidence for distributed processing in human working memory. *Electroencephalogr. Clin. Neurophysiol 87* (1993), 128–143.

[53] Gladwin, T. E., den Uyl, T. E., Fregni, F. F., and Wiers, R. W. Enhancement of selective attention by tdcs: interaction with interference in a sternberg task. *Neuroscience letters 512*, 1 (2012), 33–37.

[54] Granger, C. W. Some recent development in a concept of causality. *Journal of econometrics 39*, 1 (1988), 199–211.

[55] Grossman, P., Niemann, L., Schmidt, S., and Walach, H. Mindfulness-based stress reduction and health benefits: A meta-analysis. *Journal of psychosomatic research 57*, 1 (2004), 35–43.

[56] Grossman, T., and Balakrishnan, R. The bubble cursor: enhancing target acquisition by dynamic resizing of the cursor's activation area. In *Conference on Human Factors in Computing Systems* (2005), ACM Press.

[57] Guger, C., Edlinger, G., Harkam, W., Niedermayer, I., and Pfurtscheller, G. How many people are able to operate an eeg-based brain-computer interface (bci)? *IEEE transactions on neural systems and rehabilitation engineering 11*, 2 (2003), 145–147.

[58] Hankins, T. C., and Wilson, G. F. A comparison of heart rate, eye activity, eeg and subjective measures of pilot mental workload during flight. *Aviation, space, and environmental medicine 69*, 4 (1998), 360–367.

[59] Hannon, J., Bennett, M., and Smyth, B. Recommending twitter users to follow using content and collaborative filtering approaches. In *Proceedings of*

the fourth ACM conference on Recommender systems (2010), ACM, pp. 199–206.

[60] Harris, S. *Waking up: A guide to spirituality without religion.* Simon and Schuster, 2015.

[61] Harrivel, A. R., Weissman, D. H., Noll, D. C., and Peltier, S. J. Monitoring attentional state with fnirs. *Frontiers in human neuroscience 7* (2013), 861.

[62] Hasenkamp, W., and Barsalou, L. W. Effects of meditation experience on functional connectivity of distributed brain networks. *Frontiers in human neuroscience 6* (2012), 38.

[63] Hasenkamp, W., Wilson-Mendenhall, C. D., Duncan, E., and Barsalou, L. W. Mind wandering and attention during focused meditation: a fine-grained temporal analysis of fluctuating cognitive states. *Neuroimage 59*, 1 (2012), 750–760.

[64] Hincks, S. W., Afergan, D., and Jacob, R. J. Using fnirs for real-time cognitive workload assessment. In *International Conference on Augmented Cognition* (2016), Springer, pp. 198–208.

[65] Hincks, S. W., Bratt, S., Poudel, S., Phoha, V., Dennett, D. C., Jacob, R. J. K., and Hirshfield, L. M. Entropic brain-computer interfacing: Using fnirs and eeg to measure attentional states in a bayesian framework. In *PhyCS* (2017).

[66] Hirshfield, L. M., Bergen-Cico, D., Costa, M., Jacob, R. J., Hincks, S., and Russell, M. Measuring the neural correlates of mindfulness with functional near-infrared spectroscopy.

[67] Hoffmann, S., and Falkenstein, M. The correction of eye blink artefacts in the eeg: a comparison of two prominent methods. *PLoS One 3*, 8 (2008), e3004.

[68] Holmes, G., Donkin, A., and Witten, I. H. Weka: A machine learning workbench.

[69] Hölzel, B. K., Lazar, S. W., Gard, T., Schuman-Olivier, Z., Vago, D. R., and Ott, U. How does mindfulness meditation work? proposing mechanisms of action from a conceptual and neural perspective. *Perspectives on psychological science 6*, 6 (2011), 537–559.

[70] Homan, R. W., Herman, J., and Purdy, P. Cerebral location of international 10–20 system electrode placement. *Electroencephalography and clinical neurophysiology 66*, 4 (1987), 376–382.

[71] Horvath, J. C., Forte, J. D., and Carter, O. Evidence that transcranial direct current stimulation (tdcs) generates little-to-no reliable neurophysiologic effect beyond mep amplitude modulation in healthy human subjects: a systematic review. *Neuropsychologia 66* (2015), 213–236.

[72] Hoshi, Y. Hemodynamic signals in fnirs. In *Progress in brain research*, vol. 225. Elsevier, 2016, pp. 153–179.

[73] Huron, D. B. *Sweet anticipation: Music and the psychology of expectation.* MIT press, 2006.

[74] Josipovic, Z., Dinstein, I., Weber, J., and Heeger, D. J. Influence of meditation on anti-correlated networks in the brain. *Frontiers in human neuroscience 5* (2012), 183.

[75] Kane, M. J., and Engle, R. W. The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. *Psychonomic bulletin & review 9*, 4 (Dec. 2002), 637–671.

[76] Kent, J. T. Information gain and a general measure of correlation. *Biometrika 70*, 1 (1983), 163–173.

[77] Killingsworth, M. A., and Gilbert, D. T. A wandering mind is an unhappy mind. *Science 330*, 6006 (2010), 932–932.

[78] Kirilina, E., Jelzow, A., Heine, A., Niessing, M., Wabnitz, H., Brühl, R., Ittermann, B., Jacobs, A. M., and Tachtsidis, I. The physiological origin of task-evoked systemic artefacts in functional near infrared spectroscopy. *Neuroimage 61*, 1 (2012), 70–81.

[79] Kühn, S., and Gallinat, J. Brain structure and functional connectivity associated with pornography consumption: the brain on porn. *JAMA psychiatry 71*, 7 (2014), 827–834.

[80] Kühn, S., Ritter, S. M., Müller, B. C., Baaren, R. B., Brass, M., and Dijksterhuis, A. The importance of the default mode network in creativitya structural mri study. *The Journal of Creative Behavior 48*, 2 (2014), 152–163.

[81] Landwehr, N., Hall, M., and Frank, E. Logistic model trees. In *Machine Learning: ECML 2003*. Springer, 2003, pp. 241–252.

[82] Lemm, S., Blankertz, B., Dickhaus, T., and Müller, K.-R. Introduction to machine learning for brain imaging. *Neuroimage 56*, 2 (2011), 387–399.

[83] Li, R., Chen, K., Fleisher, A. S., Reiman, E. M., Yao, L., and Wu, X. Large-scale directional connections among multi resting-state neural networks in human brain: a functional mri and bayesian network modeling study. *Neuroimage 56*, 3 (2011), 1035–1042.

[84] Lin, J., Keogh, E., Wei, L., and Lonardi, S. Experiencing sax: a novel symbolic representation of time series. *Data Mining and knowledge discovery 15*, 2 (2007), 107–144.

[85] Lutz, A., Slagter, H. A., Dunne, J. D., and Davidson, R. J. Attention regulation and monitoring in meditation. *Trends in cognitive sciences 12*, 4 (2008), 163–169.

[86] Lutz, J., Herwig, U., Opialla, S., Hittmeyer, A., Jäncke, L., Rufer, M., Holtforth, M. G., and Brühl, A. B. Mindfulness and emotion regulationan fmri study. *Social Cognitive and Affective Neuroscience 9*, 6 (2014), 776–785.

[87] MacKay, D. J., and Mac Kay, D. J. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

[88] Maior, H., Pike, M., Sharples, S., and Wilson, M. L. Examining the reliability of using fnirs in realistic hci settings for spatial and verbal tasks. In *Proceedings of CHI* (2015), vol. 15, pp. 3807–3816.

[89] Mandryk, R. L., Inkpen, K. M., and Calvert, T. W. Using psychophysiological techniques to measure user experience with entertainment technologies. *Behaviour & information technology 25*, 2 (2006), 141–158.

[90] Manna, A., Raffone, A., Perrucci, M. G., Nardo, D., Ferretti, A., Tartaro, A., Londei, A., Del Gratta, C., Belardinelli, M. O., and Romani, G. L. Neural correlates of focused attention and cognitive monitoring in meditation. *Brain research bulletin 82*, 1 (2010), 46–56.

[91] Manoach, D. S., Schlaug, G., Siewert, B., Darby, D. G., Bly, B. M., Benfield, a., Edelman, R. R., and Warach, S. Prefrontal cortex fMRI signal changes are correlated with working memory load. *Neuroreport 8*, 2 (Jan. 1997), 545–549.

[92] McKendrick, R., Ayaz, H., Olmstead, R., and Parasuraman, R. Enhancing dual-task performance with verbal and spatial working memory training: Continuous monitoring of cerebral hemodynamics with NIRS. *NeuroImage 85* (2014), 1014–1026.

[93] McKendrick, R., Parasuraman, R., and Ayaz, H. Wearable functional near infrared spectroscopy (fnirs) and transcranial direct current stimulation

(tdcs): expanding vistas for neurocognitive augmentation. *Frontiers in systems neuroscience 9* (2015), 27.

[94] Medeiros, L. F., de Souza, I. C. C., Vidor, L. P., de Souza, A., Deitos, A., Volz, M. S., Fregni, F., Caumo, W., and Torres, I. L. Neurobiological effects of transcranial direct current stimulation: a review.

[95] Mesquita, R. C., Franceschini, M. A., and Boas, D. A. Resting state functional connectivity of the whole head with near-infrared spectroscopy. *Biomedical optics express 1*, 1 (2010), 324–336.

[96] Moëll, B., Kollberg, L., Nasri, B., Lindefors, N., and Kaldo, V. Living smarta randomized controlled trial of a guided online course teaching adults with adhd or sub-clinical adhd to use smartphones to structure their everyday life. *Internet Interventions 2*, 1 (2015), 24–31.

[97] Molteni, E., Baselli, G., Bianchi, a. M., Caffini, M., Contini, D., Spinelli, L., Torricelli, a., Cerutti, S., Cubeddu, R., Milano, P., Bioingegneria, D., Milan, I., Dipartimento, M., and Vinci, L. Frontal brain activation during a Working Memory task: a time- domain fNIRS study. 71613N–71613N–12.

[98] Nee, D. E., Brown, J. W., Askren, M. K., Berman, M. G., Demiralp, E., Krawitz, A., and Jonides, J. A meta-analysis of executive components of working memory. *Cerebral Cortex 23*, 2 (2013), 264–282.

[99] Nitsche, M. A., Boggio, P. S., Fregni, F., and Pascual-Leone, A. Treatment of depression with transcranial direct current stimulation (tdcs): a review. *Experimental neurology 219*, 1 (2009), 14–19.

[100] Nitsche, M. A., Cohen, L. G., Wassermann, E. M., Priori, A., Lang, N., Antal, A., Paulus, W., Hummel, F., Boggio, P. S., Fregni, F., et al. Transcranial direct current stimulation: state of the art 2008. *Brain stimulation 1*, 3 (2008), 206–223.

[101] Owen, A. M., McMillan, K. M., Laird, A. R., and Bullmore, E. N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human brain mapping 25*, 1 (2005), 46–59.

[102] Pearl, J. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier, 2014.

[103] Peck, E. M., Afergan, D., and Jacob, R. J. Investigation of fnirs brain sensing as input to information filtering systems. In *Proceedings of the 4th Augmented Human International Conference* (2013), ACM, pp. 142–149.

[104] Pfurtscheller, G., Allison, B. Z., Bauernfeind, G., Brunner, C., Solis Escalante, T., Scherer, R., Zander, T. O., Mueller-Putz, G., Neuper, C., and Birbaumer, N. The hybrid bci. *Frontiers in neuroscience 4* (2010), 3.

[105] Pomplun, M., and Sunkara, S. Pupil dilation as an indicator of cognitive workload in human-computer interaction. In *Proceedings of the International Conference on HCI* (2003), vol. 2003.

[106] Raichle, M. E. The brain's dark energy. *Scientific American 302*, 3 (2010), 44–49.

[107] Raichle, M. E. The brain's default mode network. *Annual review of neuroscience*, 0 (2015).

[108] Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. A default mode of brain function. *Proceedings of the National Academy of Sciences 98*, 2 (2001), 676–682.

[109] Refaeilzadeh, P., Tang, L., and Liu, H. Cross-validation. In *Encyclopedia of database systems*. Springer, 2009, pp. 532–538.

[110] Rényi, A., et al. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability,*

Volume 1: Contributions to the Theory of Statistics (1961), The Regents of the University of California.

[111] Rodriguez, M. *Development of a Versatile Wireless fNIR System.* PhD thesis, Drexel University, 2011.

[112] Roelfsema, M. T., Hoekstra, R. A., Allison, C., Wheelwright, S., Brayne, C., Matthews, F. E., and Baron-Cohen, S. Are autism spectrum conditions more prevalent in an information-technology region? a school-based study of three regions in the netherlands. *Journal of autism and developmental disorders 42*, 5 (2012), 734–739.

[113] Sasai, S., Homae, F., Watanabe, H., Sasaki, A. T., Tanabe, H. C., Sadato, N., and Taga, G. A nirs–fmri study of resting state network. *Neuroimage 63*, 1 (2012), 179–193.

[114] Sen, S., Vig, J., and Riedl, J. Tagommenders: connecting users to items through tags. In *Proceedings of the 18th international conference on World wide web* (2009), ACM, pp. 671–680.

[115] Shiozawa, P., Fregni, F., Benseñor, I. M., Lotufo, P. A., Berlim, M. T., Daskalakis, J. Z., Cordeiro, Q., and Brunoni, A. R. Transcranial direct current stimulation for major depression: an updated systematic review and meta-analysis. *International Journal of Neuropsychopharmacology 17*, 9 (2014), 1443–1452.

[116] Shirer, W., Ryali, S., Rykhlevskaia, E., Menon, V., and Greicius, M. Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cerebral cortex 22*, 1 (2012), 158–165.

[117] Si, J., Zhao, R., Zhang, Y., Zuo, N., Zhang, X., and Jiang, T. A portable fnirs system with eight channels. In *SPIE BiOS* (2015), International Society for Optics and Photonics, pp. 93051B–93051B.

[118] Simon, R., and Engström, M. The default mode network as a biomarker for monitoring the therapeutic effects of meditation. *Frontiers in psychology 6* (2015).

[119] Smalley, S. L., Loo, S. K., Hale, T., Shrestha, A., McGough, J., Flook, L., and Reise, S. Mindfulness and attention deficit hyperactivity disorder. *Journal of Clinical Psychology 65*, 10 (2009), 1087–1098.

[120] Smith, E. E., and Jonides, J. Storage and executive processes in the frontal lobes. *Science 283*, 5408 (1999), 1657–1661.

[121] Solovey, E., Schermerhorn, P., Scheutz, M., Sassaroli, A., Fantini, S., and Jacob, R. Brainput: enhancing interactive systems with streaming fnirs brain input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2012), ACM, pp. 2193–2202.

[122] Solovey, E. T., Girouard, A., Chauncey, K., Hirshfield, L. M., Sassaroli, A., Zheng, F., Fantini, S., and Jacob, R. J. Using fnirs brain sensing in realistic hci settings: experiments and guidelines. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology* (2009), ACM, pp. 157–166.

[123] Starner, T. Project glass: An extension of the self. *IEEE Pervasive Computing 12*, 2 (2013), 14–16.

[124] Strangman, G., Culver, J. P., Thompson, J. H., and Boas, D. A. A quantitative comparison of simultaneous bold fmri and nirs recordings during functional brain activation. *Neuroimage 17*, 2 (2002), 719–731.

[125] Tang, Y.-Y., Posner, M. I., and Rothbart, M. K. Meditation improves self-regulation over the life span. *Annals of the New York Academy of Sciences 1307*, 1 (2014), 104–111.

[126] Teasdale, J. D., Segal, Z. V., Williams, J. M. G., Ridgeway, V. A., Soulsby, J. M., and Lau, M. A. Prevention of relapse/recurrence in major

depression by mindfulness-based cognitive therapy. *Journal of consulting and clinical psychology 68*, 4 (2000), 615.

[127] Teplan, M. Fundamentals of eeg measurement. *Measurement science review 2*, 2 (2002), 1–11.

[128] Treacy Solovey, E., Afergan, D., Peck, E. M., Hincks, S. W., and Jacob, R. J. Designing implicit interfaces for physiological computing: Guidelines and lessons learned using fnirs. *ACM Transactions on Computer-Human Interaction (TOCHI) 21*, 6 (2015), 35.

[129] Turing, A. M. Computing machinery and intelligence. In *Parsing the Turing Test*. Springer, 2009, pp. 23–65.

[130] Watanabe, M., and Sōgō, O. T.-t. R. I. *The prefrontal cortex as an executive, emotional, and social brain.* Springer, 2017.

[131] Weissman, D. H., Roberts, K., Visscher, K., and Woldorff, M. The neural bases of momentary lapses in attention. *Nature neuroscience 9*, 7 (2006), 971–978.

[132] Yuksel, B. F., Oleson, K. B., Harrison, L., Peck, E. M., Afergan, D., Chang, R., and Jacob, R. J. Learn piano with bach: An adaptive learning interface that adjusts task difficulty based on brain state. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (2016), ACM, pp. 5372–5384.

[133] Zaehle, T., Sandmann, P., Thorne, J. D., Jäncke, L., and Herrmann, C. S. Transcranial direct current stimulation of the prefrontal cortex modulates working memory performance: combined behavioural and electrophysiological evidence. *BMC neuroscience 12*, 1 (2011), 1.

[134] Zeidan, F., Johnson, S. K., Diamond, B. J., David, Z., and Goolkasian, P. Mindfulness meditation improves cognition: Evidence of brief mental training. *Consciousness and cognition 19*, 2 (2010), 597–605.

[135] Zhang, Q., Strangman, G. E., and Ganis, G. Adaptive filtering to reduce global interference in non-invasive nirs measures of brain activation: how well and when does it work? *Neuroimage 45*, 3 (2009), 788–794.