Field Recordings, Mind, and the Eternal Hum or Whisper: Observing neural activity in the wild and at scale

Nick Merrill[1], Thomas Maillart[1] and John Chuang[1]

[1]*School of Information, UC Berkeley, California, USA*
ffff@berkeley.edu

Abstract: whysoever why and when, and why not

1   Introduction

Our ability to build effective, useful brain-computer interface (BCI) is limited by the capabilities of our scanning devices and by our inchoate knowledge of neural activity, especially about how it looks outside of lab settings. While we expect mobile brain scanners to improve dramatically over the next few years, our ability to understand brains "in the wild" is hampered primarily by a paucity of relevant data. Almost all neurological studies, BCI research included, occurs under controlled conditions rather than in the normal human habitat, a perceptually rich environment in which people are moving freely and interacting with space, objects and people.

Presently, there exists no large-scale repository for "field recordings" (i.e., long-timescale recordings taken in uncontrolled, non-lab environments) of neural signals, making it difficult for researchers to draw the sorts of observations that might improve BCIs, and our understanding of mind generally.

The difficulty of dealing with subject-to-subject and trial-to-trial changes in neural signals provides a salient example of our knowledge gap: a BCI that works with one person will most likely not work with another, and may even stop working on the original subject over time, due simply to the fact that minds shift and change and grow. It is tempting to believe that such changes, especially those observed within subjects over time, can be predicted statistically, at least to some extent, given enough data. However, without a large-scale source of neural field recordings, it is difficult or impossible for researchers to make the sorts of observations necessary to investigate such patterns.

This paper imagines a platform for neural field recordings, approaching the problem from an infor-mation architecture standpoint and from a user interface one. How do we scale a data collection platform to hundreds of thousands of users, all of whom are producing megabytes of data every minute? How do we incentivize users to donate their neural data, and how do we assure their privacy? How might we manage access and ownership in a publically available research repository while maintaining the confidentiality of individuals involved?

maybe we talk about what this paper covers here, in summary, when we figure that out

2   Background & Related Work

2.1   Brain-computer interface

Brain-computer interface aims to establish a direct connection between the human nervous system and a digital computer, without intermediate muscular intervention. Obvious applications include devices and tools for the severely disabled, for whom muscular activity is difficult or impossible, but BCI has a potentially very wide audience if it is convenient enough to use; it promises to increase our overall communicative capacity with digital machines, and to passively monitor our cognitive load and emotional states, helping us better adapt to our environment and to ourselves.

Attempts at BCI generally employ adaptive, machine-learning algorithms, as neural signals vary from person to person, and change within individuals over time (this is refferred to as "nonstationarity" - that is, that there exists no static model of neural functioning that will work reliably with all people at all times). Some *a priori* neurological knowledge is

generally employed, usually around a central strategy (e.g., P300, SSVP, motor imagery, etc; see [] for a review of BCI strategies). Thereafter, an adaptive algorithm, usually a linear classifier, is trained on sample data, and this calibration closes the gap between the theoretical basis behind the system and the observed data in reality.

2.2 BCI in the wild

There have been numerous attempts to deploy brain-computer interfaces in outside-of-lab settings. Many wireless EEG headsets exist - there are at least four on the market presently, and three new devices slated for release in the immediate future. From an application standpoint, some commercial ventures into "direct-control" BCI applications have recieved wide media attention as of late (e.g. an application for Google Glass that takes a picture when the user focuses). Historically, most consumer BCI solutions have centered more around affective and emotional measurement such as stress detectors, meditation coaches, applications that claim to monitor mental workload, and so on. The claims behind these devices are difficult to rigorously validate, as the developers of these applications understandably do not publish detailed information about their use.

From the academic side of things, a few applications have surfaced employing consumer-grade wireless headsets. One application authenticates users using their brainwaves alone, and various others have replicated the levels of control achieved with in-lab direct-control BCI systems, with varying levels of success. A few of these projects have cognitive load measurements in real-life work situations, and a few have looked at in-the-wild emotional response. Overwhelmingly, however, academic approaches to BCI have approached their subjects and analysis from inside the lab, while free-market ventures into BCI have not rigorously collected or disseminated data on their applications' use.

2.3 Large-scale data repositories for scientific research

3 Specifications

The BCI community, as well as the neuroscience and wearable computing *(have i really established this, does this paper really establish its relevance to the larger wearables community)* communities at large, could benefit from a large-scale repository of neurological data recorded in real-world settings. In this section, we describe the specifications for an open collection platform aimed at field recordings of neural data. We focus specifically on the use of electroencephalography (EEG) *(mention eeg stuff earlier?)* due to its wide adoption in BCI and as a format for consumer brainscanners.

3.1 Scalability

A system for storing neural recordings is unlike traditional web applications (Twitter, Facebook, Youtube) in that we expect each user to produce a much larger volume of (raw) data, and to submit that data more frequently. The Neurosky MindWave, a current-generation wireless EEG with a single electrode, produces approximately a megabyte of raw data every minute. This bitrate will increase linearly with the number of electrodes on the device, so even devices on the immediate horizon may produce up to five times this volume. With each user producing gigabytes of data every day in chronic recording scenarios, the system will not scale well to large number of users.

However, raw recordings may not be necessary to produce useful insights about neural data, especially at high volume. A lossy file format may be sufficient, the proverbial mp3 of neural data. What might this file format look like?

3.2 Redundancy & Control

3.3 End user concerns

4 Design recommendations

4.1 A lossy fileformat for EEG data

4.2 A decentralized network

4.3 A decoupled application layer

5 Discussion