CSEC-472 Authentication Paper

# Proposed Study in Mapping Optimal Probe Sites with Regard to Brainjacking

Jarryd Brits[1], Nicholas Graca[1], Gia Mule[1], Chris Vantine[1], and Adrianna Visca[1]

[1]Computing Security, Rochester Institute of Technology

May 6, 2020

## Abstract

Enhancing the human body has always been a crazy idea that has only been teased about in Sci-Fi films, but it is finally becoming commercially available with the use of Brain-Computer Interfaces (BCIs). Common use cases include individuals with an advanced prosthetic limbs, such as a hand that can clench and release by using the host's brain. Since this technology is still in its infancy, there is a lot of room for improvement. Our study looks at a vast amount of user data to find what locations on the human head are the optimal locations for Transcranial Direct Current Stimulation (TDCS). TDCS is when a current is sent through a specific part of the brain in order to stimulate it. This can be used to help treat chronic diseases, for example Parkinson's, which can cause uncontrollable shaking in the hands. In this study, we are only interested in the actions of rest, right hand squeeze and left hand squeeze. Once the optimal locations for classification of these movements are found, we can propose the use of TDCS at those locations to control the squeezing of a person's hands. With controlling a person's body comes great responsibility, and that is why we will be touching on how dangerous brainjacking, the name of this type of attack, can be if the BCI is accessed and used by an inappropriate party. By providing a thorough description of how we propose to conduct this experiment, we hope that it can be applied to more than just the opening and closing of hands and to evolve to look at more complex actions such as walking or throwing objects. An expansion of the scope of this project might even include the induction of non-motor thought, such as decision-making and memory alterations.

## Keywords

Brain-Computer Interface, BCI, Electroencephalogram, EEG, Brainjacking, Transcranial Direct Direct Current Stimulation, TDCS, Transcranial Magnetic Stimulation, TMS.

## 1 Introduction

The first brain-computer interface (BCI) implanted in a Homo Sapiens was in 1998 by researcher Philip Kennedy[1]. Since then, brain mapping technology has been a topic of interest in the scientific and medical communities. The applications for BCIs are vast - prostheses[2], cardiac research[3], and even non-medical use cases such as gaming[4]. As brain-computer interfaces become more widely used, particularly in medicine, a re-evaluation of the authentication and security measures in place must occur. The new technology introduces new challenges that may require development of novel conceptual frameworks, device ontologies, typologies, protocols, and operational practices[5]. The widespread use of BCIs also introduces a novel attack known as "brainjacking". Within the scope of brainjacking, this study aims to discuss induction of past thought. Inductino of elasticity in the human brain has already been achieved through the application of transcranial magnetic stimulation[6]. We aim to expand the potential for induction to include past motor-thought and transcranial direct current stimulation.

This study is built on the idea of optimal electrode site selection via machine learning. One study[7] makes use of a naive Bayes classifier to deduce the least number of necessary electrodes for clinical implementation of P300 Speller systems. On the same principle, we aim to deduce

the optimal electrode site(s) for the induction of past thought based on previously-collected EEG data for Physio[8] [9] [10]. Ideally, the study has two phases, grounded in the theory of generalization. The first phase is the deduction of optimal EEG electrode sites, specifically for opening and closing the right and left fists, from the previously collected data and the use of a convolutional neural network. The second phase will be the validation of site selection by means of transcranial direct current stimulation in a clinical environment. Future research may include the stimulation of the targeted sites from BCI feedback that has been forcibly generated, and stronger authentication procedures to ensure BCI feedback to the brain is not generated by an unauthorized entity.

In this paper, we will provide a deeper discussion on the background of brain-computer interfaces and authentication. A special regard for brainjacking and its significance to this research will be considered. We will describe the methods and procedures necessary to carry out both phases. Due to limitations in processing power and storage space, we cannot at this time perform the described experiment. In lieu of preliminary results, a complete framework will be provided and preliminary suppositions will be discussed. Finally, we will explore the importance and impact of this research with regard to medical, technological, and ethical concerns.

# 2 Background & Significance

## 2.1 Brain Computer Interface

In 1924, a German neuroscientist discovered the electrical activity of the human brain with EEG. In 1970, the Defense Advanced Research Projects Agency (DARPA) of the United States initiated a program to explore brain communications using EEG. In 1998, the first invasive, non-EEG, implant was implanted in a human brain and produced high quality signals. The next year, a BCI was used to air a quadriplegic for limited hand movement[11]. Since then, BCIs have been rapidly improving the quality of life of the disabled community and allowing them to interact with their environment. BCIs are typically centered around control applications, such as paralyzed body parts, prosthesis, cursors, etc. Most of these applications are with the purpose of aiding the needs of the disabled community by helping with communication or with quality of life.

The purpose of a BCI is to detect and quantify features of brain signals that indicate the user's intentions and to translate these features in real time into device commands that accomplish the user's intent. When looking at BCI systems from a high-level, there are four main components: signal acquisition, feature extraction, feature translation, and device output. During the signal acquisition phase, the brain signals are amplified, digitized, and sent to the computer for feature extraction and translation. Feature extraction is the process of analyzing the digital signals to distinguish pertinent signal characteristics relating to the user's intent. The resulting signal features are then passed to the feature translation algorithm, which converts the features into the appropriate commands for the output device, to accomplish the user's intent[12].

## 2.2 Transcranial Magnetic Stimulation

Transcranial magnetic stimulation (TMS) is a noninvasive procedure that uses magnetic fields to cause electric current at a specific area of the brain through electromagnetic induction. The first transcranial magnetic stimulation device was introduced in 1985 by Anthoney Barker and his colleagues as a non-invasive pain-free method to stimulate the human cortex. The device used physics principles discovered in 1881 by Michael Faraday, such as creating a magnetic field by running electricity through metal coils[13].

TMS has since shown diagnostic and therapeutic potential in the central nervous system with a wide variety of disease states in neurology and mental health. Since the mid-1900s, TMS has been used as a treatment for depression. During TMS treatment, an electromagnetic coil is placed against the subject's scalp near their forehead. The electromagnet painlessly delivers a magnetic pulse that stimulates nerve cells in the region of their brain involved in mood control and depression. It's thought to activate regions of the brain that have decreased activity in depression[14]. In general, TMS treatments have no severe risks since the process is non-invasive. However, there are some slight side effects, such as headaches, electrode site discomfort, and lightheadedness.

TMS also has the ability for physical stimulation. The body's movements are entirely controlled by electricity running through the nervous system. That electrical current tells the muscles in the body to either relax or contract. The electrical currents all originate from your brain. Therefore, the electromagnetic coil could be placed on a specific part of the brain to induce a current that tells the body to relax or contract. This method has been used effectively

to treat hand tremors which is a symptom of Parkinson's[15].

### 2.3 Research Significance

The purpose of this study is to provide a method for selecting the optimal electrode sites to induce past thought with transcranial direct current stimulation. Once optimal electrodes are found, stimulation can be created in a laboratory environment. Inducing thought with targeted current in a laboratory setting is a proof of concept of a potential attack during a brainjacking. Without proper authentication in place, feedback from a BCI may be sent unbeknownst to the system's host. This study aims to provide a stronger incentive to implement more secure authentication protocols within BCIs. In particular, this research addresses optimal electrode site selection via machine learning. The designed experiment was broken into two main methods of deducing the optimal location for electrodes. The first method proposes an algorithm to find the optimal electrode site when the subject is opening and closing their right and left fists. The second research method involves using this optimal location as the center of the anode/cathode configuration for transcranial direct current stimulation in a clinical environment. Unfortunately, due to time and resource limitations we were not able to perform definitive research using these methods, but rather, provide a detailed framework for future research in this field.

## 3 Related Work

### 3.1 Industry Standards

The BCI industry lacks a unified standardization of its many devices and variants, making it not only difficult to compare the effectiveness of devices of similar approaches, but also limiting application usage to the device the application was developed on[16]. Although there are standards in the way of sensor placement and measurement, such those set by the American Electroencephalographic Society for EEG electrode placement[17], these standardizations are limited in scope, and do not extend to hardware and software standards that would make research and development tasks consistent across the multitude of similar commercially-available BCIs on the market today.

### 3.2 Brainjacking

As of the writing of this paper, there is no official definition for brainjacking. This is due to the brain computer interface technology being so new. It is just starting to be commercially available. That is why security flaws are being talked about but since these devices are not popular, it is something that does not get a lot of recognition. The best way to describe brainjacking, based on current research, is the unauthorized access and/or control over a brain computer interface.

A practical example of what attackers are able to do is based on an advanced prosthetic hand. The BCI in that device reads the user's brain waves so that it knows when to close the prosthetic hand and when to open the hand. An experiment was conducted[18], and researchers were able to determine a subjects PIN when they read the subjects brain waves while they were flashing images at them. Of course an attacker would need to go through a lot of trouble in order to be able to do this. However, just like how BCIs are in their infancy, so are these attacks. Doctors are not able to efficiently read brainwaves and instantly know what the person is trying to do but, as BCI technology evolves and becomes more efficient, so will the attacks on them. With a better understanding of how to read brain waves, the attacker would not need to flash images of pictures of numbers to determine the victims PIN. They could wait for the victim to use the ATM and read their brain then.

Researchers have also had the ability to reduce the symptoms of Parkinson's, like hand tremors, by using a BCI to induce a current on the brain to counteract the tremors[15]. This better aligns with the main goal of our experiment. We wanted to isolate the probe sites on the brain that control the left and right clenching of the hand. If an attacker had access to the system that was responsible for controlling a users body, there is no telling what they could do. As previously mentioned, this technology is still up and coming. As it gets better with time, the actions that BCIs will be able to control or even facilitate will increase in complexity. It is not unreasonable to think that one day, a BCI will be able to control every body movement without the user having to do anything themselves. Thinking that an unauthorized user could have full control over another human is truly terrifying and an attack on human autonomy.

In today's world, with massive data breaches such as those at Target or Equifax, the average person is becoming much more security conscious than they used to be. This has allowed for more security-focused research to be conducted

on BCI technology, however, before it can be made more commercially available, BCIs must be redesigned with security in mind. Every computer has the ability to be compromised, but not every computer is directly connected to a person and has the ability to read brain waves and control bodily movements.

# 4 Research Design & Methods

## 4.1 Procedures

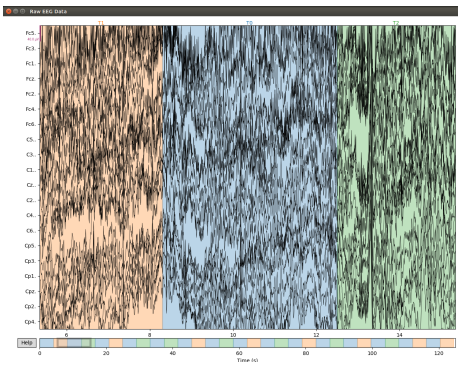
Using a dataset from Physio [8] [9] [10], a Convolutional Neural Network (referred to as CNN from now on) can be trained to classify 64-channel EEG data as either movement of the left fist, right fist, or resting state. Various other Neural Network models were considered alongside the CNN, such as the Long-Short Term Memory, the Recurrent Neural Network and the Deep Neural Network based primarily on backpropagation utilising sigmoid and relu activation functions. The CNN was chosen as the others posed various technical issues with our data set. For instance, both the LSTM and the RNN use the input of previous data for future predictions. Since the proposed dataset is composed of EEG readings from multiple different hosts, a feedback loop may cause inconsistent classification. With regard to the Deep Neural Network, this method seems viable, however, various other studies suggest the CNN model over the others[19][20][21]. Although EEG data, by nature, is continuous, once recorded it becomes discrete. After chunking based on the time period between the start and end point of the actual movement - from now on referred to as an epoch - the resulting data matrix is ideal for the CNN model after image conversion.

The data should be filtered to only include trials which focus on real movement, as the initial dataset includes imagined movement. As previously mentioned, only left fist and right fist movement is of interest for this study. After some initial observation of the data, the trials from subjects 88, 92, and 100 were sampled at a rate less than 160 Hz, which was the sampling rate used for the other 103 subjects. For consistency, these subjects' trials should be dropped. See the appendix for more information. From the remaining subjects, trials of interest are trials 3, 7, and 11. The new dataset includes three trials from 103 different subjects, resulting in 309 different trials. Each trial contains 29 epochs in total, resulting in 8,961 distinct events, each with 657 data points. We recommend a train/test split of approximately 80/20. Of the 309 trials, 247 can be used for training, and the remaining 62 will be split amongst two rounds of testing, as described below. All 309 trials should be prepared with the same pre-processing procedures. Test data will undergo sampling and further pre-processing before being passed through the CNN.

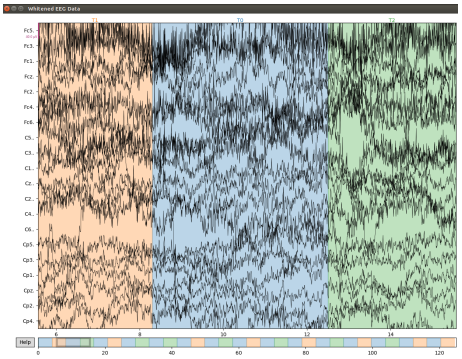
For data observation, we used MNE-python[22] to read, whiten, break into epochs, and display the raw EEG data. The epochs were then written to NumPy arrays[23] array for further pre-processing. Each array is of shape (64, 657) before processing; 64 corresponds to each of the channels, 657 corresponds to each datapoint. Each array itself is a single epoch. Figure 1 shows the original and whitened data, complete with colored blocks in the images representing annotations, where orange is a T1 event, which corresponds to movement of the left fist, green is a T2 event, which corresponds to movement of the right fist, and blue is a T0 event, which corresponds to rest, as well as a NumPy representation of the data after epoching, transformation, and reshaping.

We propose the use of continuous wavelet transformation (CWT) for data pre-processing. CWT can be implemented with PyWavelets[24]. A CWT was chosen over previously used methods, such as a fast fourier transform (FFT), due to the decision to use a CNN. Using CWT, the epoch arrays post transformation have a shape of (64, 657, 657), assuming a range of 1 to 657 is used as the frequency scales. The decision to use 657 for the frequency scales stems from the convenience of CNN processing with a square image. It should be noted that using scales from 1 to 657 violate Nyquist principles, however, for the sake of classification, this is not detrimental. The resulting arrays are rather large, and we suggest utilizing Python's h5py[25] library to save them in HDF datasets. Figure 1 contains a sample entry of such a dataset. Note that the data is reshaped prior to being written to the file. The new shape is (657, 657, 64), to reflect the square image with 64 different input channels as required by the TensorFlow module and is the default format for temporal data when using the Keras API.

The newly shaped dataset is what will be fed into the CNN for training. For training purposes, all 64 channels should be used as is. The Convolutional Neural Network has various parameters that, in general, are flexible, based on the input data and its various output expectations. Considering this research paper could not accomplish a fully functional model due to the extended period of time needed for data processing, the actual effectiveness of the parame-



(a) Raw EEG data before whitening



(b) Raw EEG data after whitening

```
Data shape:  (657, 657, 64)
Data Excerpt:
[[[6.89862273e-310 6.89862274e-310 6.89862274e-310 ... 6.89868457e-310
   6.89862274e-310 6.89862330e-310]
  [6.89862273e-310 6.89862273e-310 9.92609724e-315 ... 6.89862274e-310
   6.89862273e-310 6.89862960e-310]
  [6.89868480e-310 6.89862273e-310 6.89862330e-310 ... 9.92665882e-315
   6.89868457e-310 6.89862274e-310]
  ...
  [6.89871478e-310 6.89870175e-310 6.89862273e-310 ... 6.89862273e-310
   6.89862274e-310 6.89862274e-310]
  [6.89868480e-310 6.89862273e-310 6.89862274e-310 ... 6.89862274e-310
   6.89862274e-310 6.89862274e-310]
  [6.89862274e-310 6.89862274e-310 6.89862960e-310 ... 6.89868480e-310
   6.89868480e-310 6.89862274e-310]]

 [[1.01075466e-314 6.89862274e-310 6.89862273e-310 ... 6.89862274e-310
   6.89862274e-310 6.89871478e-310]
  [6.89871478e-310 6.89862273e-310 6.89862273e-310 ... 6.89862273e-310
   6.89862273e-310 6.89862274e-310]
  [6.89862274e-310 6.89868480e-310 6.89868480e-310 ... 6.89870175e-310
   6.89869302e-310 1.01081992e-314]
  ...
  [6.89862274e-310 6.89871699e-310 6.89862274e-310 ... 1.03117777e-314
   6.89868457e-310 6.89871699e-310]
  [6.89862274e-310 6.89862273e-310 6.89862960e-310 ... 6.89871478e-310
   6.89862274e-310 6.89862274e-310]
  [6.89862274e-310 6.89862273e-310 6.89862274e-310 ... 6.89862274e-310
   6.89862330e-310 6.89862274e-310]]

 [[6.89862273e-310 6.89862273e-310 6.89862274e-310 ... 6.89862274e-310
   6.89862274e-310 6.89862274e-310]
  [6.89862273e-310 6.89862273e-310 6.89871164e-310 ... 6.89868480e-310
   6.89862274e-310 6.89862273e-310]
  [6.89862273e-310 6.89862960e-310 6.89862274e-310 ... 1.03134422e-314
   6.89862273e-310 6.89862330e-310]
  ...
  [1.04923366e-314 6.89862330e-310 6.89862273e-310 ... 6.89862274e-310
   6.89862274e-310 1.04926958e-314]
  [6.89862274e-310 6.89862274e-310 6.89862273e-310 ... 6.89862273e-310
   6.89862274e-310 6.89871478e-310]
  [1.04928982e-314 6.89862273e-310 6.89862330e-310 ... 6.89862274e-310
   6.89862273e-310 6.89862274e-310]]

 ...
```

(c) A NumPy array representation of a single epoch after CWT and reshaping

Figure 1: EEG Data and Array Representation

ters are uncertain and thus the parameters and minute details that define the model's hidden layers and its activation functions are purely speculation based on previous research. To start, the number of classes was set to three considering our output possibilities only included the three actions: relaxed hands, tense left hand, and tense right hand. In terms of batch size, the model specified a value of 256 which is eight times the default value. We decided to use that specific value to broaden the number of samples per gradient update in order to increase the samples analysed by the model before updating the internal model parameters, and thus we assume it will provide the ideal set of output values that the model can compare with before attempting the backpropagation method.

The following CNN model structure was taken and modified slightly from a different internal project that was not published, but was intended for classifying network data with a similar data structure to the EEG data gathered. Figure 2 depicts the current structure of the neural network along with its input shapes at each hidden layer. There are only two Convolutional layers used in the model along with two Dense layers, a Max Pooling layer, a dropout layer as well as a flatten layer. Due to the immense number of data points present for input to the model, the Max pooling layer attempts to pool together and 'down sample' the features of the output of the previous layer thereby encapsulating the existing data into a smaller more compact dataset. For example, the initial output shape was (646, 646, 100) which, by utilising the Max Pooling function, was converted into (215, 215, 100) through the pool size of (3, 3). For instance, 646 divided by 3 is 215 which follows as the resulting tuple shape depicts this. The value of 100 is conserved between the two tuples as it represents the number of samples or input streams given to the model. By utilising the Dropout layer, we attempt to avoid overfitting the data given constant input to output methods, instead the dropout layer with a value of naught point five allows for a rate of fifty percent of units to effectively stop functioning. This method of preventing overfitting was initially mentioned by this study[26]. The Flattening layer allows the model to convert a multi-dimensional data set to a single dimension, which constitutes the passing on of data from the Convolutional and Dropout layers to the Dense layer, which only takes a single dimensional dataset as input.

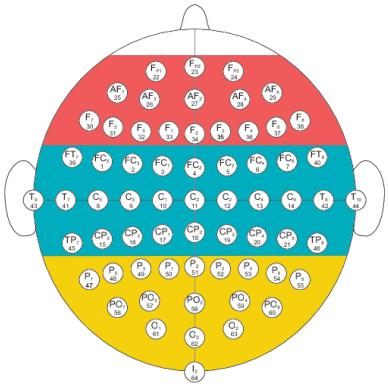
The CNN model utilises three of the thirty one different activation functions: ReLu, Sigmoid and Softmax. ReLu is the most frequently used activation function with regards to CNNs

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_1 (Conv2D)            (None, 646, 646, 100)     921700
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 215, 215, 100)     0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 213, 213, 100)     90100
_________________________________________________________________
dropout_1 (Dropout)          (None, 213, 213, 100)     0
_________________________________________________________________
flatten_1 (Flatten)          (None, 4536900)           0
_________________________________________________________________
dense_1 (Dense)              (None, 20)                90738020
_________________________________________________________________
dense_2 (Dense)              (None, 3)                 63
=================================================================
Total params: 91,749,883
Trainable params: 91,749,883
Non-trainable params: 0
```

Figure 2: Output from Keras's model summary function for the proposed CNN

as they have proven to produce positive results when working with images, however they are subject to failure and loss of information when dealing with negative values. The image created for CNN processing can be normalized to avoid negative values, if necessary. The Sigmoid function is a more generalized version of the Softmax function and usually acts as a measure of probability as it ranges between naught and one. In this instance we used the Sigmoid function on the second Convolutional layer in an attempt to utilise its probabilistic properties on the resulting dataset from both the ReLu function and the Max pooling layer. Softmax is generally a useful activation function for datasets that result in having more than two classes as output such is the case in this instance. The Softmax function is also primarily used, such as in this instance, as the last activation function the data is run through in order to permute the data into vector form which is supposed to represent the class vector. Ideally this would make it more human readable or rather, easier to interpret without the use of complex aggregate models. Granted, in this case, the softmax function was primarily used to facilitate the vector representation as an accuracy score was calculated from its final output.

We propose two "rounds" of testing to determine optimal electrode site selection. The first round requires the grouping of electrodes based on cranial location to narrow our sample of electrodes. The second round groups the remaining electrodes in the narrowed sample to determine the optimal location or locations. The sampling of the first round can be manual. By comparing the montage for the original dataset (see Appendix, figure 8) and relative locations of each electrode over the human brain, it can be concluded that this round of testing will focus on three groups of electrodes. Figure 3 gives two depictions, graphic and tabular, of the suggested grouping of electrodes. The decisions that lead to this sampling structure will be further explored in section 4.3 Novel Techniques.



(a) Color-coded visual representation of electrode grouping

**Electrode Grouping and Corresponding Brain Region**

| Group Number | Electrodes | Brain Region |
|---|---|---|
| 1 | $F_{p1}$, $F_{p2}$, $AF_7$, $AF_3$, $AF_z$, $AF_4$, $AF_8$, $F_7$, $F_5$, $F_3$, $F_1$, $F_z$, $F_2$, $F_4$, $F_6$, $F_8$ | Frontal lobe |
| 2 | $FT_7$, $FC_5$, $FC_3$, $FC_1$, $FC_z$, $FC_2$, $FC_4$, $FC_6$, $FT_8$, $T_9$, $T_7$, $C_5$, $C_3$, $C_1$, $C_z$, $C_2$, $C_4$, $C_6$, $T_{10}$, $TP_7$, $CP_5$, $CP_3$, $CP_1$, $CP_z$, $CP_2$, $CP_4$, $CP_6$, $TP_8$ | Frontal lobe, Parietal lobe, Temporal lobe |
| 3 | $P_7$, $P_5$, $P_3$, $P_1$, $P_z$, $P_2$, $P_4$, $P_6$, $P_8$, $PO_7$, $PO_3$, $PO_z$, $PO_4$, $PO_8$, $O_1$, $O_z$, $O_2$, $I_z$ | Parietal lobe, Occipital lobe |

(b) Tabular representation of electrode grouping

Figure 3: Grouping of electrodes for CNN testing round 1

The preparation of data for the first round of testing involves three additional steps in preprocessing. To reduce processing time, it's recommended that the sampling of round 1 test data occur prior to application of the CWT function. Figure 4 shows the mapping of electrode name to channel number.

For any group, only the electrodes in that group should be transformed with the CWT function. Remaining channels should be set to all zeroes. This will, effectively, only test classification with the transformed subset of chan-

| | | | | | | |
|---|---|---|---|---|---|---|
| 00: Fc5 | 08: C3 | 16: Cp1 | 24: Af7 | 32: C3 | 48: Cp1 | 56: Cp1 |
| 01: Fc3 | 09: C1 | 17: Cpz | 25: Af3 | 33: C1 | 49: Cpz | 57: Cpz |
| 02: Fc1 | 10: Cz | 18: Cp2 | 26: Afz | 34: Cz | 50: Cp2 | 58: Cp2 |
| 03: Fcz | 11: C2 | 19: Cp4 | 27: Af4 | 35: C2 | 51: Cp4 | 59: Cp4 |
| 04: Fc2 | 12: C4 | 20: Cp6 | 28: Af8 | 36: C4 | 52: Cp6 | 60: Cp6 |
| 05: Fc4 | 13: C6 | 21: Fp1 | 29: F7 | 37: C6 | 53: Fp1 | 61: Fp1 |
| 06: Fc6 | 14: Cp5 | 22: Fpz | 30: F5 | 38: Cp5 | 54: Fpz | 62: Fpz |
| 07: C5 | 15: Cp3 | 23: Fp2 | 31: F3 | 39: Cp3 | 55: Fp2 | 63: Fp2 |

Figure 4: PyCharm output showing the index of the channel in NumPy data arrays with its corresponding electrode name

nels. We recommend using 31 of the remaining trials for this round of testing. A pool of 31 trials allows for approximately 299 epochs to be pre-processed for each electrode grouping. Each batch of data should be passed through the trained CNN and results should be interpreted with a confusion matrix separately. This will allow the group with the most accurate classifications to be identified and focused on during the second round of testing.

The electrode group determined as optimal denotes the only electrode channels of interest during the final round of testing. All other electrode channels will be set to 0 as an initial step in pre-processing. We recommend splitting the remaining 899 epochs from the final 31 trials evenly amongst each electrode in the group. The same testing procedure as the previous round will be used. During data pre-processing, it is important to take note of which epoch is being processed, and which electrode is in use. Electrodes placed on the right side of the scalp (electrodes with an even number subscript) will control left-body movement. This means that even numbered electrodes should not be paired with T2 epochs, and odd numbered electrodes should not be paired with T1 epochs. Because epochs labeled as T0 are rest, the lateral placement of the electrode in question has little consequence. It is possible that no single electrode as an optimal selection site exists. The electrode(s) that prove most accurate in classification are what we will consider the optimal electrode site for past thought induction via transcranial direct current stimulation. Most probably, a minimum of two electrodes will be found as optimal, one for the right half of the brain (left body movement) and one for the left half of the brain (right body movement). This concludes phase 1 of research. Phase 2 will be described under the assumption that two optimal electrodes are found as described above.

For phase 2, we recommend a clinical study in which to validate the results of phase 1. Healthy subjects with full motor control of both hands and arms should be recruited for this phase. Two simple electrodes will be placed such that current flows through the brain region that corresponds to the electrode determined to be optimal. Previous studies that include anodal stimulation of the M1 primary motor cortex place the cathode at the contralateral supraorbital ridge (across the forehead)[27], the superior aspect of the trapezius[28], and the shoulder[29]. When choosing right or left sides of the body, both the anode and cathode were placed on the same side (on the trapezius or shoulder), or centrally (on the supraorbital ridge). There are no governing standards or even conclusive results for anode and cathode placement for stimulation of a single brain region[30]. Due to the lack of standards, as well as the dependency on data from the previous phase, we have created a guide to how we suggest placement of the electrodes, where the anode and cathode are on either side of the region of interest, orthogonal to the central succulus. An orthogonal electrode configuration has been found to produce more consistent current flow[31]. Figure 5 describes the placement recommendation using two 5 cm by 5 cm simple electrodes. It is important to note that the images in figures 5, 6, and 7 were created with SimNIBS[32], which is not intended for clinical usage, but only as a research tool. It should also be noted that each individual will display varying current fields due to various anatomical differences, including skull thickness, cortical folding, and skull conductivity[33] [34]. The simulations in figures 6 and 7 are based on a pre-loaded sample head model named Ernie.

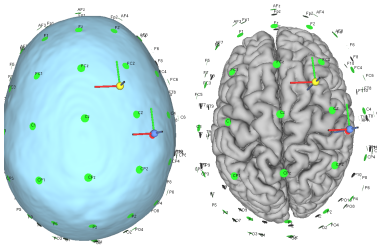


Figure 5: Anode (yellow) and cathode (blue) placement on simulation model

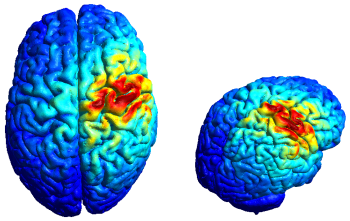


Figure 6: Current flow estimation on gray matter surface

While we do not have recommendations for the amount of study participants, we do recommend the use of two experimental groups and one control group. All three groups will be blind. All participants will be fitted with electrodes in the same orientation. Participants will sit, comfortably, and relax both arms on arm rests. A technician will either deliver a TDCS dose, or pretend to. Doses will be 1 mA for experimental group 1, and 2 mA for experimental group 2. The use of distinct doses is rooted in the

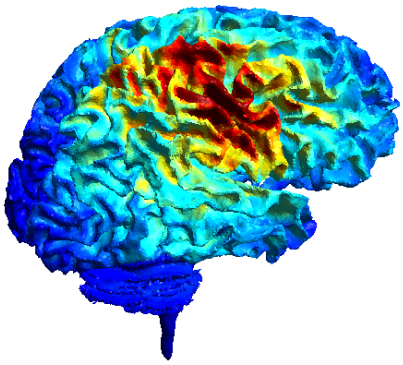


Figure 7: Current flow estimation on white matter

conclusions found in this study[35], citing that individualized dose-control may eliminate variance in electric field intensities at a cortical target site, in this case M1. All three groups will undergo at least two tests, one targeting the optimal electrode in the left brain, and one targeting the optimal electrode in the right brain. It is recommended each test is administered thrice to each participant. The re-testing of an electrode should take place on a different day for each of the three times. The groups will be fitted with EEG electrodes in addition to the anode and cathode used in TDCS to monitor current in the region of interest. This will help validate the placement of electrodes.

All three groups will be asked to identify when they feel a sensation either at the site of the electrodes, or in either forearm or hand. No participant should know when they are receiving a dose or how many doses they should expect in a trial. An important metric from this study will be perception of current in the body. Due to the use of simple electrodes rather than electrodes with saline soaked sponges, participants may feel a warming or tingling at the electrode site. The main interest will be what the participants feel in their arms and hands. Additionally, researchers should be monitoring for motor-evoked potentials (MEPs), as seen in this study[36], as well as in studies concerning TMS[37][38]. This means that the use of an EMG to measure potential in the muscle of interest (in this case, the right and left hands, ideally placed on the lower forearm, just above the wrists)[39]. A flexible electrode grid, as described in this study[40] and used in this study[39], is recommended for EMG measurement. Any muscular contraction that is observed should be tracked.

Upon the conclusion of the trials, researchers should have three forms of results: EEG current measurements in the area of the interest, perceived bodily current from participants, and monitored MEPs/visible contractions. EEG measurements from the control group can be compared to the EEG measurements from the experimental groups. The control group serves to show the baseline of current in the monitored area, whereas the experimental groups should provide a sort of timeline for when doses were administered. Reading amplitudes may vary due to the varied controlled dose used for each group. Finally, a record of both visually observed muscular contraction and EMG recorded MEPs should be available for analysis. A comparison between visible contraction and recorded MEPs can be made to determine the threshold TDCS does for induction.

## 4.2 Timeline

The anticipated timeline for a study of this caliber is a matter of weeks. Because the selected dataset has already been collected, minimal preparation is needed for phase 1. Due to the vast amount of data to pre-process, data preparation could take anywhere from two days to a week, depending on available processing power. The training and initial testing data can be prepared prior to obtaining results, however, the third set cannot be prepared until the first round of testing has been completed. These time estimates include preparation of all three sets of data. A feasible timeline for the training of the CNN as described is within a single day. The initial round of testing can also be done well within the confines of a day, most probably the same day as training. Finally, after preparing the final testing set, the testing can be completed in less than a day. After finding a suitable pool of volunteers, participant screening can likely be completed in a week or two, depending on the pool size and desired sample size of subjects. For the projection of a timeline, we assume a sample size of 100 subjects will be used. Screening for this study is minimal. Pertinent considerations include dominant handedness, age, and health. Finally, a single iteration of the physical portion of the experiment can be completed within a week. Ideally, subjects will undergo the same trial procedure on at least three separate days each, to speak to reproducibility of the results. If this is the case, likely three weeks will be needed for the experiment to be carried out. Analysis of the results can be completed likely within a week, or less. Overall, a projected timeline for this study is five to six weeks.

### 4.3 Novel Techniques

As described in section 4.1 Procedures, the first round of testing data is clustered into three groups of electrodes. These groups were determined by comparing the electrode location on the human scalp to the relative brain region corresponding to it. As depicted in 3, the electrodes are divided laterally. The number of electrodes per group is not even among the groups. Also in the figure, the table shows that the groupings do not directly correspond to lobe location. Rather, grouping was centered around the control function for various regions of the brain. Group 1 consists of electrodes over brain regions primarily concerned with attention and planning; group 2 consists of electrodes over brain regions primarily concerned with movement and touch; and group 3 is over brain regions primarily concerned with vision. As this experiment is focused on induction of past thought in the form of motor thought, the regions of the brain that are targets of interest are surface regions; most of the temporal lobe, and the cerebellum, are not locations of interest. We hypothesize that the grouping that will provide the most accurate classifications during this round of testing is group 2. We further hypothesize that of group 2, the optimal electrode sites for thought induction will lie on or above the medial line[41].

## 5 Preliminary Suppositions & Implications

Since the study is purely theoretical, due to current circumstances, the results gathered are not fit for an accurate representation of the project. Thus, the results cannot be properly applied to both our theoretical framework and the underlying assumptions that support our study. With respect to the future direction of this study and further research on the topic, the results of this study have two possible avenues with a more generalised scope. The first possibility is that there are no optimal points that can be clearly distinguished from each other. Therefore, the implications of past thought induction as a result of brainjacking may not be as severe as speculated. The second is that there are a variety of optimal points that can be focused on in order to optimize both the learning rate and the accuracy of the model. Thereby, potentially giving attackers the ability to isolate and consolidate their attacks in a more meaningful manner, which consequently, means that their efforts have been minimised with a maximised threat level on their behalf.

Subsequent research may also include further refinement of the Convolutional Neural Network machine learning model along with its various parameters that should be adjusted based on input data and variation of hosts. It may be beneficial that any further research relies not only on the CNN but rather a hybrid method. Specifically, a CNN model combined with a previous iteration feedback loop model such as the Recurrent and/or the Long-Short Term Neural Network. Using this hybrid method would rely on the assumption that the input dataset being used is from the same host as opposed to a larger dataset such as the dataset proposed for use in this study. A next step in research would include attempting targeted stimulation of the brain from a sort of "forced feedback" from an invasive/implanted device, such as a BCI. The primary research question for a study in this regard would be, if an attacker gains unauthorized access to a host's BCI, can they force the BCI to send current feedback to a specific target in the brain and thereby induce the same motor function as was induced in the clinical trial? As an expansion of the research, the techniques used to induce motor thought and thereby MEPs may prove useful as tools to induce past memories or influence decision making. A study of this nature might involve using the same optimal site selection techniques for classification of a host making a decision between two variables, for example.

The results of this study are meant to persuade software engineers, security engineers, and even mechanical and electrical engineers working on prosthesis and implantable medical devices that integrate with BCIs to focus on the development of novel conceptual frameworks, device ontologies, typologies, protocols, and operational practices with strong regard for authentication protocols[5]. Without proper foresight of potential attacks, a secure environment in a body area network where invasive brain connection is involved is not achievable. The results of the proposed study may cause a spike in hesitancy and social distrust in BCI technology, however, that is not to say that some wariness of a new frontier in technology is not unfounded. Identifying a set of optimal selection sites for attacks will not prevent the attacks, but the identification may provide insight into the prevention of the attack, and lay a foundation for future research in the authentication and security features and protocols built into BCIs.

## 6 Conclusions

This study in itself is instrumental in understanding the various ways in which the correctly

selected optimal probe points influence a more efficient insight into the authentication processes against external/remote attackers. Given current circumstances, the reach of the preliminary phases of this study are severely limited and thus further research into this area of study is required in both the extent of viable security of the BCI with a fewer number of probes as well as the accuracy and precision of using the selected electrode points with other cognitive motor functions. The proposed study serves as a proof of concept of such an attack in the hopes that future research will be steered towards the development of industry standards and practices that allow for mitigation and eventually prevention of thought induction via direct current stimulation.

The use of a CNN is aimed at reduction of unnecessary resource usage until a basis can be formed on which electrode locations are optimal for past thought induction. Once a basis is formed for electrode site selection, a direct laboratory study that includes transcranial direct current stimulation may be more beneficial than one that leverages transcranial magnetic stimulation due to the nature of feedback from BCIs in the "real world." As previously mentioned in section 4 Research Design & Methods section of this paper, the CNN model was chosen over the various other Neural Networks due to its compatibility with the data set, along with the model's specialisation with images as input data. Various other Neural Network models were initially considered alongside the CNN such as the Long-Short Term Memory, the Recurrent Neural Network and a Deep Neural Network based primarily on backpropagation utilising sigmoid and relu activation functions. Granted, the CNN was chosen as the others posed various technical issues with our data set.

With a deeper understanding of the various probe points on the scalp needed to both induce motor responses as well as its further application into 'reading minds' proposed by these studies [42] [15], comes the implication of how this information is stored and used. Ideally, a machine learning module such as the one described in this research paper is used to determine potential outcomes based on the input from a singular host, as opposed to the multitude of data mentioned above. Considering this, another neural network model may be preferable as previous iterations of similar actions or thoughts could be utilised in the model to more accurately determine output. This would mean the use of either a RNN or a LSTM model. It therefore follows that the information gathered from the EEG needs to be stored securely and locally creating a new avenue of attack for hackers. If the information stored on the device along with its proposed meaning becomes 'public' knowledge there are a number of privacy and ethical concerns since attackers could potentially 'read' the hosts thoughts.

The scope of brainjacking is an attack on human autonomy itself. There are various ethical, social, and technical barriers to overcome at the use of BCIs, particularly in the medical field, becomes prominent. With the goal of a secure integration and strong authentication at the forefront of development, we hope that a collaborative effort amongst engineers of all disciplines brings into existence a more secure brain-computer interface that allows any user a more comfortable assurance of privacy in their own mind.

# 7 Acknowledgements

We'd like to thank Dr Justin Pelletier who guided us from start to finish in this endeavor. He helped us to refine our research into a single, relevant topic, as well as provided support and mentorship throughout the duration of the process.

# 8 Appendix

## 8.1 Appendix A

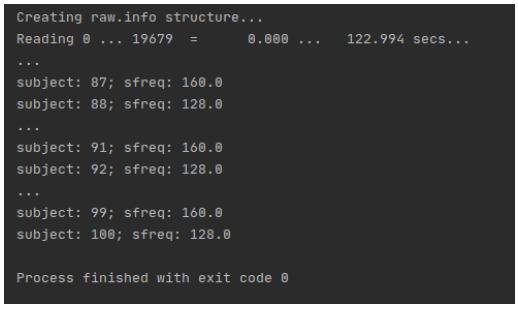



Figure 8: The sampling frequency used in the trials as well as documented with the dataset is 160 Hz. For subjects 88, 92, and 100, the sampling frequency is 128 Hz, as seen here. It is suggested to omit these subjects, as readings may be inaccurate. Nyquist rate states that a sampling frequency of 160 Hz can be used to sample data up to 64 Hz, well within the realm of brain waves. While the wave of interest for this study - beta waves - are within the Nyquist standards using a sampling rate of 128 Hz, consistency is unnecessarily sacrificed using data sampled at a lower rate.

## 8.2 Appendix B

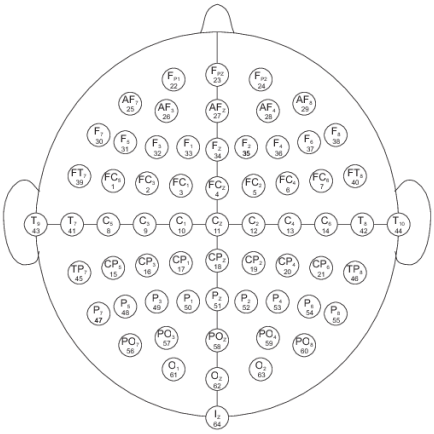


Figure 9: The original, un-altered montage provided with the Physio dataset. Electrode configuration during data collection followed international 10-10 standards, with 64 electrodes on the scalp. Electrodes Nz, F9, F10, FT9, A1, A2, TP9, TP10, P9, and P10 were excluded.

# References


[1] Kennedy PR, Bakay RA. Restoration of neural output from a paralyzed patient by a direct brain connection. Neuroreport. 1998;9(8):1707–1711.

[2] Mastinu SA Clemente, et al. Grip control and motor coordination with implanted and surface electrodes while grasping with an osseointegrated prosthetic hand. In: Journal of NeuroEngineering and Rehabilitation 16,[2019]; 2019. p. article 49.

[3] Rogers AJ, Miller JM, Kannappan R, Sethu P. Cardiac Tissue Chips (CTCs) for Modeling Cardiovascular Disease. IEEE Transactions on Biomedical Engineering. 2019 Dec;66(12):3436–3443.

[4] Van Erp J, Lotte F, Tangermann M. Brain-computer interfaces: beyond medical applications. Computer. 2012;45(4):26–34.

[5] Gladden ME. 3. In: Critical Challenges in Information Security for Advanced Neuroprosthetics. 2nd ed. Synthipnion Academic; 2017. p. 61–103.

[6] Foysal KR, Baker SN. Induction of plasticity in the human motor system by motor imagery and transcranial magnetic stimulation. The Journal of Physiology. 2020;.

[7] Speier W, Deshpande A, Pouratian N. A method for optimizing EEG electrode number and configuration for signal acquisition in P300 speller systems. Clinical Neurophysiology. 2015;126(6):1171–1177.

[8] Schalk G, McFarland DJ, Hinterberger T, Birbaumer N, Wolpaw JR. BCI2000: a general-purpose brain-computer interface (BCI) system. IEEE Transactions on biomedical engineering. 2004;51(6):1034–1043.

[9] for Adaptive Neurotechnologies NC. BCI2000 Mainpage; 2018. Available from: https://www.bci2000.org/mediawiki/index.php/Main_Page.

[10] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. circulation. 2000;101(23):e215–e220.

[11] Neurosky. What Is BCI and How Did It Evolve?; 2015. Available from: http://neurosky.com/2015/06/what-is-bci-and-how-did-it-evolve/.

[12] Shih JJ, Krusienski DJ, Wolpaw JR. Brain-computer interfaces in medicine. Mayo Foundation; 2012. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3497935.

[13] History of TMS;. Available from: http://www.brainefit.net/history-of-tms.html.

[14] Transcranial magnetic stimulation. Mayo Foundation for Medical Education and Research; 2018. Available from: https://www.mayoclinic.org/tests-procedures/transcranial-magnetic-stimulation/about/pac-20384625.

[15] Roelfsema PR, Denys D, Klink PC. Mind Reading and Writing: The Future of Neurotechnology. Trends in Cognitive Sciences. 2018 May;Available from: https://www.sciencedirect.com/science/article/pii/S1364661318300925.

[16] van Erp J, Lotte F, Tangermann M. Brain-Computer Interfaces: Beyond Medical Applications. Computer. 2012;45(4):26–34.

[17] Luis Fernando Nicolas-Alonso JGG. Brain Computer Interfaces, a Review; 2012. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3304110/.

[18] Martinovic I, Davies D, Frank M, Perito D, Ros T, Song D. On the Feasibility of Side-Channel Attacks with Brain-Computer Interfaces. In: Presented as part of the 21st USENIX Security Symposium (USENIX Security 12). Bellevue, WA: USENIX; 2012. p. 143–158. Available from: https://www.usenix.org/conference/usenixsecurity12/technical-sessions/presentation/martinovic.

[19] Kanaga EGM, Kumaran RM, Hema M, Manohari RG, Thomas TA. An experimental investigations on classifiers for Brain Computer Interface (BCI) based authentication. In: 2017 International Conference on Trends in Electronics and Informatics (ICEI). IEEE; 2017. p. 1–6.

[20] Ravi A, Heydari N, Jiang N. User-Independent SSVEP BCI Using Complex FFT Features and CNN Classification. In: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). IEEE; 2019. p. 4175–4180.

[21] Boonyakitanont P, Lek-uthai A, Songsiri J. Automatic Epileptic Seizure Onset-Offset Detection Based On CNN in Scalp EEG. In: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2020. p. 1225–1229.

[22] Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, et al. MEG and EEG data analysis with MNE-Python. Frontiers in neuroscience. 2013;7:267.

[23] Oliphant TE. A guide to NumPy. vol. 1. Trelgol Publishing USA; 2006.

[24] Lee G, Gommers R, Waselewski F, Wohlfahrt K, O'Leary A. PyWavelets: A Python package for wavelet analysis. Journal of Open Source Software. 2019;4(36):1237.

[25] Collette A, et al.. h5py/h5py 2.8. 0 Zenodo; 2018.

[26] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research. 2014;15(1):1929–1958.

[27] Nitsche MA, Paulus W. Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. The Journal of physiology. 2000;527(3):633–639.

[28] Weightman M, Brittain JS, Punt D, Miall RC, Jenkinson N. Targeted tDCS selectively improves motor adaptation with the proximal and distal upper limb. Brain Stimulation. 2020;.

[29] Chen J, McCulloch A, Kim H, Kim T, Rhee J, Verwey WB, et al. Application of anodal tDCS at primary motor cortex immediately after practice of a motor sequence does not improve offline gain. Experimental brain research. 2020;238(1):29–37.

[30] Reinhart RM, Cosman JD, Fukuda K, Woodman GF. Using transcranial direct-current stimulation (tDCS) to understand cognitive processing. Attention, Perception, & Psychophysics. 2017;79(1):3–23.

[31] Rawji V, Ciocca M, Zacharia A, Soares D, Truong D, Bikson M, et al. tDCS changes in motor excitability are specific to orientation of current flow. Brain stimulation. 2018;11(2):289–298.

[32] Saturnino G, Antunes A, Stelzer J, Thielscher A. SimNIBS: a versatile toolbox for simulating fields generated by transcranial brain stimulation. In: 21st Annual Meeting of the Organization for Human Brain Mapping (OHBM 2015); 2015. .

[33] Datta A, Bansal V, Diaz J, Patel J, Reato D, Bikson M. Gyri-precise head model of transcranial direct current stimulation: improved spatial focality using a ring electrode versus conventional rectangular pad. Brain stimulation. 2009;2(4):201–207.

[34] Bestmann S, Ward N. Are current flow models for transcranial electrical stimulation fit for purpose? Brain stimulation. 2017;10(4):865–866.

[35] Evans C, Bachmann C, Lee JS, Gregoriou E, Ward N, Bestmann S. Dose-controlled tDCS reduces electric field intensity variability at a cortical target site. Brain stimulation. 2020;13(1):125–136.

[36] Shirota Y, Terney D, Antal A, Paulus W. Influence of Concurrent Finger Movements on Transcranial Direct Current Stimulation (tDCS)-Induced Aftereffects. Frontiers in behavioral neuroscience. 2017;11:169.

[37] Kaczmarczyk I, Rawji V, Rothwell JC, Hodson-Tole E, Sharma N. Comparison

between conventional electrodes and ultrasound monitoring to measure TMS evoked muscle contraction. bioRxiv. 2020;.

[38] Sato A, Torii T, Iwahashi M, Iramina K. Alterations in Motor Cortical Excitability Induced by Peripheral Stimulation With Magnetic Stimulation. IEEE Transactions on Magnetics. 2018;54(11):1–4.

[39] van Elswijk G, Kleine BU, Overeem S, Eshuis B, Hekkert KD, Stegeman DF. Muscle imaging: mapping responses to transcranial magnetic stimulation with high-density surface electromyography. Cortex. 2008;44(5):609–616.

[40] Lapatki BG, Van Dijk JP, Jonas IE, Zwarts MJ, Stegeman DF. A thin, flexible multielectrode grid for high-density surface EMG. Journal of Applied Physiology. 2004;96(1):327–336.

[41] Dubuc B. The brain from top to bottom. 2002;.

[42] Tamara Bonaci HJC Ryan Calo. App Stores for the Brain : Privacy and Security in Brain-Computer Interfaces; 2015. Available from: https://ieeexplore-ieee-org.ezproxy.rit.edu/document/7128844.