Central Auditory Processing Disorder: Towards a Therapeutic EEG Neurofeedback Brain Computer Interface

Sean R. Chadwick
College of William and Mary

Follow this and additional works at: https://scholarworks.wm.edu/honorstheses

Part of the Other Analytical, Diagnostic and Therapeutic Techniques and Equipment Commons

Recommended Citation
https://scholarworks.wm.edu/honorstheses/134

This Honors Thesis is brought to you for free and open access by the Theses, Dissertations, & Master Projects at W&M ScholarWorks. It has been accepted for inclusion in Undergraduate Honors Theses by an authorized administrator of W&M ScholarWorks. For more information, please contact scholarworks@wm.edu.
Central Auditory Processing Disorder: Towards a Therapeutic EEG Neurofeedback

Brain Computer Interface

A thesis submitted in partial fulfillment of the requirement
for the degree of Bachelor of Science in Computational Neuroscience
from The College of William and Mary

by

Sean Robert Chadwick

Accepted for Honors

Paul Kieffaber, PhD, Director

Christy Porter, PhD

Chris Ball, PhD

Greg Smith, PhD
Abstract

Central auditory processing (CAP) refers to the process of integrating and processing auditory signals in the central auditory nervous system. Problems with CAP are thought to underlie central auditory processing disorder (CAPD) which is associated with specific populations of adults and children who demonstrate poor performance on tasks. CAPD is typically diagnosed in individuals with poor auditory perception who also show no physical problems with their inner ear, outer ear, or cochlea (Keilman et al., 2013; Koravand et al. 2013). CAPD is characterized by an impaired ability to filter out background noise and distinguish between different auditory stimuli, and is often comorbid with other neurological disorders (Kim & Chung, 2013; Strauss et al., 2008). Exciting new research has shown improvements those identified with CAPD-like disorders can improve speech comprehension, harmonic recognition, and sound localization, just by engaging in behaviors which are associated with CAP (Alain et al., 2014; Anderson et al., 2013). The overarching aim of this research was to create a online electroencephalography (EEG) brain computer interface (BCI) that could be used by anyone, not just those who show CAPD symptomatology, to gain increased performance on central auditory processing tasks.
Central Auditory Processing Disorder:

Towards a Therapeutic EEG Neurofeedback Brain Computer Interface

There are many examples of organisms whose nervous systems process auditory information from stereo sources (biaurally). An example of this is the use of echolocation by the fruit bat (Kossel et al., 2014) or the mammalian auditory systems wherein differences in the time-course and frequencies of reflected sound waves are essential for interacting with surroundings (Grothe et al. 2010). In humans, the ability to compare and process information from two sources is essential to understanding speech, detecting cues from the environment, and perceiving stereo musical effects (Kim & Chung, 2013; Naatanene et al., 2011; Strauss et al., 2004; Zendel & Alain, 2012). Central auditory processing (CAP) is generally robust for healthy young adults but tends to decline naturally with aging (Anderson et al., 2013; Kielman et al., 2013; Quaranta et al., 2015). CAP disorder (CAPD) in children and younger adults is often comorbid with other disorders such as attentional disorders (e.g. ADHD), autism spectrum disorders, and other learning disorders in general (Strauss et al., 2008). Interestingly, there is a wealth of new evidence suggesting long term improvement of CAPD can be achieved through occupational therapy (Alain et al., 2014; Anderson et al., 2013).

CAPD therapy typically involves having patients practice tasks which require some aspect of CAP; these include tone source localization or discriminating speech sounds, which are difficult for people with CAPD. Alain et al., (2014) showed that simply having people with CAPD focus on, (and differentiate between), harmonics and instrument sounds in stereo music can create long term improvement in CAP. This does not come as a surprise since Zendal and Alain previously demonstrated in 2012 that musicians experience slower onset of age related CAPD, and this is most likely due to their exposure and attention to musical sounds throughout
the course of their profession. Due to the comorbidity of CAPD with other disorders, Strauss et al., (2008) created measures of CAPD that do not rely on behavioral auditory discrimination by the participant. These techniques can help overcome limiting factors in the treatment of CAPD resulting from issues related to attentional deficits or simple failure to follow instructions in these patients.

The primary aim of this thesis was to examine the possibility of using an EEG neurofeedback system for training in CAP. Such a system could be used to augment conventional occupational therapy for CAPD. Similar to the CAPD identification task introduced by Strauss et al., (2008) such a system would work without patient knowledge of what was being tested, and would be robust despite problems with task attention or participant cooperation. Chang et al., (2014) published evidence that they had created one such EEG neurofeedback system, claiming that it could improve auditory discrimination even without participants being aware of such changes in their neural responses. Chang et al., (2014) used measures of the mismatch negativity (MMN) event-related brain potential, which had previously been demonstrated to be positively correlated with high performance in oddball auditory discrimination tasks, and provide feedback to the participant about the strength of this neural response. Changing the size of a centrally presented visual feedback stimulus (a green circle) in accordance with the amplitude of the MMN response, they were able to demonstrate robust auditory learning in participants, despite telling participants to ignore the auditory stimuli and focus only on the visual feedback stimulus. Chang et al.’s system is efficient and effective, despite still suffering from the same limitation as other neurofeedback systems because the MMN-based BCI can only provide meaningful feedback after the first 20 trials have occurred. At least this many trials is generally needed to get a good baseline of the participant’s normal neural
activity, which is a main input to their BCI algorithm. The value of this strategy is inline with ideologies of neurofeedback training in other contexts (e.g. ADHD therapy): participants do not need to be aware of what they are learning and the training can be implemented in the form of a variety of arbitrary video games or other tasks which may help reduce boredom and enable children to benefit more than standard occupational training. Thus, one of the first goals of the present research was to determine if other EEG-based measures of brain activity could be used to drive a BCI algorithm. We hypothesized that we could analyze steady state evoked potentials and beta-wave activity as measures of improvement in CAP, and provide feedback of this analysis in real time to participants during neurofeedback training.

**Steady state evoked potentials.** The steady state evoked potential is a measure of brain electrical activity that is elicited by any periodic sensory input. The procedure for measuring the steady state evoked potential is sometimes called “frequency tagging” because the neural activity will be dominated by the frequency of the periodic stimulus. The stimulus could potentially consist of any sensory modality, although for EEG experiments this is most often implemented as a flashing visual stimulus or an oscillating auditory sound, depending on the experimental task (Bharadawj et al., 2014; Gaume et al., 2014; Mahajan et al., 2014). This is well characterized for visual stimuli, where this effect is described as a visual steady state response or visual steady state evoked potentials (VSSR and VSSVEP respectively). Recently characterized is the auditory steady state response (ASSR and ASSVEP respectively), which has been demonstrated at a variety of frequencies, with established responses ranging from 16Hz to 80Hz (Bharadwaj et al., 2014; Mahajan et al., 2014). Noise tagging is a similar procedure to frequency tagging, but power is estimated over a spectrum of frequencies as opposed to a narrow band ($\approx 1$Hz) (Farquhar et al., 2008).
These effects can be exploited in a variety of ways to shed some light on the event related processes. For instance using VSSR an experimenter could determine precisely what times a participant’s gaze was directly looking or not looking at a particular stimulus on the screen. By employing multiple stimuli oscillating at different frequencies, an experimenter could determine which stimulus the participant is attending to with high temporal resolution. In this sense frequency tagging can be used to investigate how specific attentional mechanisms are associated with event related potentials and task performance in a variety of contexts. Farquhar et al., demonstrated in 2008 that noise tagging can indeed be used to create a BCI which can improve performance on an oddball discrimination task. This suggests that dual-ear ASSVEPs could be a useful measure to incorporate when building the neurofeedback system conceived for this project, as the task is very similar to Farquhar et al., except for the greater variance in range of possible stimuli. Therefore the auditory stimuli used in this experiment were amplitude modulated at 30 and 40Hz or at 20 and 30Hz to investigate noise tagging as a possible feature in the creation of the BCI.

**Overview of brain computer interfaces**

In general, a brain computer interface is a system that allows the user to interact with a machine with little or no physical action, only relying on information derived from the brain via neuroimaging techniques. Many interfaces are built to aid those who have been physically disabled in some way and require alternate methods of interacting with machines and humans, such as prosthesis control or BCI spellers (Perseh & Kiamini, 2008; Wang, 2013). Other paradigms of BCI are built with the intention of altering neural activity in ways that have been hypothesized to be beneficial for a particular disorder, such as ADHD (Lubar, 1995). Other implementations of BCIs include experiments seeking to use neurofeedback to enhance or
Elucidate some property of some event related neural activity (Chang et al., 2014; Nijboer et al., 2008; Strauss et al., 2004; Shibata et al., 2011).

In an interesting proof of concept, Shibata et al., (2011) showed that decoded fMRI neurofeedback could be used to train participants to have increased task performance compared to control groups when performing a simple visual processing learning (VPL) task involving discrimination of the angular orientation of Gabor patches. With no prior knowledge of the task, VPL induction was achieved over the course of several days with a task where participants attempted to make a circle larger by “thinking”. Over the course of 6 seconds fMRI measurements were taken, followed by 2 seconds of visual feedback where a circle was made larger or smaller, corresponding to measurements of change of fMRI activity in regions of interest (ROI) (Shibata et al., 2011). Research such as that by Shibata et al., (2011) raises interesting questions, including whether central auditory process learning can be induced in the same way as VPL. Given the accessibility and affordability of EEG, it also becomes important to know whether learning can be similarly induced by EEG neurofeedback.

Functional differences between EEG and fMRI

fMRI and EEG are both valuable neuroimaging techniques which present different limitations in their ability to provide information about neurological processes. MRI is determined by changes in the electron spin of water molecules and fMRI is a measure of the blood-oxygen-level-dependant (BOLD) response, a contrast between the concentrations of oxygenated and deoxygenated hemoglobin (Logothetis, 2008). EEG uses scalp electrodes to measure changes in the electrical field potentials generated by the post-synaptic activity of populations of neurons. The primary trade-off between fMRI and EEG is one of spatial versus temporal resolution. Although fMRI provides high resolution imaging of the localization of
changes in oxygen concentration, it is only capable of recording these changes over the course of 1-2 seconds. Thus, while it provides good spatial resolution, fMRI is unable to discriminate the millisecond-millisecond changes in brain activity that underlie brain function (Logothetis, 2008). Juxtaposed to this, EEG possesses high temporal resolution, providing information about the distribution of electrical potentials on the cortex on the scale of 2000 samples per second. However, this high temporal resolution comes at the cost of spatial resolution. In some cases, researchers have created protocols using both fMRI and EEG techniques simultaneously. Although capitalizing on the strengths of each technique, this procedure is costly, and requires a great deal of expertise to implement appropriately (Goldman et al., 2002). Historically EEG has been the most popular technique for creating brain computer interfaces largely because of its high temporal resolution, low cost, and high mobility relative to fMRI (Delorme et al., 2011).

**Model EEG brain computer interface paradigm**

The simplest example of a brain computer interface is comprised of several connected components. A neuroimaging system (such as a EEG cap and amplifier) is connected to a computer which performs operations on some event related time course of a stream of information. In many cases this is a window of 4-6 seconds of continuous imaging data taken during stimulus presentation. The information is added to a memory buffer and a computer performs operations on the data samples to determine some feedback value from the information. The feedback value could be the relative size of a circle, or the loudness of a tone, or even just a range of integers (Chang et al., 2014; Nijboer et al., 2008; Shibata et al., 2011). Typical operations include bandpass filtering of data to remove noise or decimation followed by characterization of the phase or frequency of particular potentials occurring in the cortex. There is a variety of software available for creating and implementing EEG BCI (see Delorme et al.,
2011); notable is the BCILAB software which acts as a toolbox extension to EEGLAB, both of
which were created by Christian Kothe and Scott Makeig at the Swartz Center for Computational
Neuroscience at University of San Diego. BCILAB is particularly useful because it allows
combination of analysis techniques of EEGLAB to enhance BCILAB’s streamlining of learning
algorithm development, segmenting, filtering, offline simulation, and implementation.

**Current limitations for EEG brain computer interfaces**

Many attempts have been made to create brain computer interfaces and creating a BCI
can be very difficult. When attempting to use EEG as a brain computer interface there are many
complicating factors which must be taken into account. Compared to fMRI, spatial resolution is
poor for EEG imaging. Therefore EEG BCIs must rely on signals generated by cells in the
cortex. As with any BCI or probabilistic simulation the system is also limited by the power of the
computational process which is being used to generate the neurofeedback model. The more
complicated the solution, the more processing time is required. Processing time is thus a very
important consideration because the goal in BCI is often to give participants timely feedback,
which can improve their ability to learn a task or control a device. Thus, there is usually a
tradeoff between accuracy and processing power, whereby more powerful probabilistic
conclusions often require more computational events and thus more time to implement (Perseh
and Kiamini, 2013).

One extreme example of this issue can be found in current attempts to model complete
mammalian nervous activity on a neuron by neuron basis, requiring unprecedented
computational performance and benchmarks of memory volume and processing speed. Such a
‘connectome’ model would require fast access to and processing of terabytes of information per
second to create a functional simulation of an individual's nervous system (Seung, 2011). There are currently no machines capable of accomplishing such a feat in the context of BCIs.

**Learning algorithms: statistical techniques and training solutions**

Creation of the learning algorithm can be the most difficult and elusive part of designing a BCI. Due to the lack of a complete model of the molecular mechanisms and neural connectivities underlying most cognitive processes, neuroscientists can only ‘relate’ certain stimulus events to the activity of localized populations of neurons. This implies that EEG and fMRI BCIs rely on reverse inference (Poldrack, 2006), and furthermore means that the basis of many BCIs are indeed arbitrary; there are often many different ways to implement a BCI that are ultimately functionally identical, but cause different and unexplained changes in patterns of brain activity (Perseh and Kiamini, 2013; Sanchez et al., 2014; Zuberer et al., 2015). Despite some of our lack of understanding of BCIs, neurofeedback algorithms have been known to cause real functional changes in measures of a person's performance on a variety of cognitive tasks, as well as long lasting changes in the actual connectivity and functional networking in the brain. Recently, Megumi et al., (2015) used fMRI to induce long lasting changes in connectivity, between brain regions that are generally unassociated and show low connectivity. These changes sometimes lasted for months, and occasionally the connectivity increased over a long time-course (weeks) after ending the training program (Megumi et al., 2015).

**Basic psychometric measures used when implementing EEG BCIs**

There are some general methods that have been identified as good approaches to BCI and these include: identifying brain ROI and restricting data processing to information relevant to those areas (Kropotov et al., 2005; Shibata et al., 2011), filtering EEG signals to restrict data to particular frequencies or frequency bands (Bharadwaj et al., 2014; Mahajan et al., 2014; Zuberer
et al., 2015), comparing data from later neurofeedback trials to data from earlier trials (Baharadwaj et. al., 2014; Megumi et al., 2015; Sanchez et al., 2014; Shibata et al., 2011), comparing online information from one or multiple trials to aggregated information collected from multiple participants who previously performed the performance task (Megumi et al., 2015; Sanchez et al., 2014; Shibata et al., 2011), and analysis of connectivity between particular ROI (Shibata et al., 2011; Zuberer et al., 2015). All of these can be combined using various statistical techniques to create some arbitrary feedback value, which generally represents predictive change or maintenance in some associated neural process (Baharadwaj et al., 2014; Delorme et al., 2011; Megumi et al., 2015; Sanchez et al., 2014; Shibata et al., 2011).

**Machine learning and support vector classification**

When building a BCI, there are generally two stages (offline algorithm training and online execution) which combine elements described above. Machine learning is implemented to classify both the offline and online data of interest collected from the participant. Classification of offline data can take days depending on the volume of data and the algorithm and hardware used to analyze example data sets. Algorithm training consists of deep analysis of multiple data sets, where the goal is to tune the parameters of the solution to the particular process in question. During online processes, which must happen on a time course of seconds, a more static approach is used based on an algorithm trained in the offline analysis. The online process must use a machine learning technique to classify changes in the ongoing EEG data over the course of the neurofeedback regimen. For example, Strauss et al., (2008) reported that using a new classification technique called a support vector machine (SVM) could be used to classify offline data for auditory brainstem responses (ABR) and beta-wave detection in children with CAPD.
They tested the hypothesis that a diagnostic SVM could be used to treat individuals with CAPD, and with measures of beta-wave activity, ABR, ASSEVP, LRP, single trial ERPs.

Some combination of these measures were suggested for inclusion when developing the novel neurofeedback model of CAP learning. Thus, the primary aim of the present research project was to expand on earlier research using EEG for neurofeedback and to evaluate the utility of event-related potentials (ERPs) and SSVEPs for the construction of a BCI for the treatment of CAPD.

**Method**

**Participants**

Participants were male and female undergraduates ages 18-30 years old (mean = 20.4), recruited or solicited from psychology classes at the college for credit or payment. Inclusion criteria included normal hearing and no history of seizure disorder. Participants were compensated with 1 SONA credit or $8 per hour of participation. All participants had hearing better than 20dB for 1000Hz and 2000Hz pure tones in each ear, except for two participant (where both ears were 25dB or 30Hz for 2000Hz).

**Privacy and Confidentiality**

The data participants contributed to this research is anonymous, identifiable only by a number assigned by the experimenter. Once participants left the lab, there was no way to connect their responses with their personal identity. Moreover, all data and records were stored on password-protected computers in a locked laboratory.

**Materials and Procedure**

Due to unanticipated factors, three different EEG bioamplifiers and electrode caps were used. Three different versions of the behavioral task were also created and tested on three
unique groups of participants. Due to amplifier availability and maintenance, each iteration of the task was evaluated with a different EEG system (three different amplifiers).

**First Amplifier (N=7).** The first amplifier used was a 24 channel direct current (DC) amplifier made by Neuro Assessment Systems (NAS). This amplifier was loaned to the cognitive psychophysiology lab by Dr. John Kimura of Sensorium DBPA and NAS. 20 electrode channels were recorded at a 2000Hz sample rate, using AgCL sintered electrodes in caps which did not contain facial electrodes. Photodiodes were used to mark stimulus events in the data file. After substantial testing, it was determined that participants may have been susceptible to an electrical shock using this amplifier and use was immediately discontinued.

**Second Amplifier (N=6).** The second amplifier used was a 24 channel Sensorium DBPA DC amplifier. 20 electrode channels were recorded at a 2000Hz sample rate, using the same caps as for the first system, which did not contain facial electrodes. After recording data from six participants, this amplifier was moved to the laboratory at Eastern Virginia Medical School for use with older adults at the Center for Geriatrics and Gerontology.

**Third Amplifier (N=10).** The third amplifier used was a 144 channel Sensorium DPBA-1 alternating current (AC) amplifier. 36 channels including 8 facial electrode channels were recorded at a 2000Hz sample rate, using a 72 channel cap made with AgCL sintered electrodes. The data were recorded using a bandpass filter of .001 to 500Hz low and highpass respectively. Impedance was adjusted to approximately 20KOhms before the participant began the task.

**General protocol for EEG data collection**

Participants reviewed and signed a consent form and were briefed on the EEG procedure. To control for differences in hearing, participants took a short, self-administered hearing test with Bose AE2 noise-cancelling over-ear headphones using a Samsung touchscreen tablet which
ran a 2014 version of the Temasek Polytechnic non-diagnostic hearing test table application (see Figure 22 for hearing test results). The hearing test asked participants to slide a button on a loudness scale to indicate the lowest dB level (lowest possible threshold was 20dB) at which they could hear any sound, testing responses for 250, 500, 1000, 2000, 4000, and 8000 Hz tones in each ear. In order to reduce impedance and increase data robustness, participants first used a brush with stiff bristles to lightly abrade the top, back, and sides of their head, which helped remove dead skin cells and decrease resistance between the electrodes and the scalp. Participants were then fitted with a 24 or 74 channel electrode cap; for the 74 channel cap electrode locations on their face were cleaned with NuPrep gel and alcohol wipes or electrode prep wipes. For caps with facial electrodes, a reference electrode was attached to the right side of the nose, a ground electrode was attached to the middle of the forehead, and ocular electrodes were placed above, below, and to the sides of both their eyes. For each cap 24 electrodes and any facial electrodes were filled with saline electrode gel applied with a blunt syringe. Participants were seated in a Faraday chamber and their cap was connected to the amplifier. The electrodes in the cap were filled with small amounts of electrode gel with the blunt syringes, which were used to gently abrade the scalp if needed to ensure a good connection. The task instructions were reiterated to the participants, after which they were fitted with 10Ohm 3A insert headphones with disposable in-ear front tubes made by E.A.R. Auditory systems. The in-ear headphones provided tones at a range of 65 to 75 dB. Participants were asked to make an effort to stay vigilant and minimize body movements during the task and respond as quickly and accurately as they were capable of. No guidance on was provided on keeping eyes open or closed during the task, but participants were told to blink normally. All participants performed task in the same location, which was a
copper RFI/EMI enclosure created by Universal Shielding Corporation. Written instructions and visual stimuli were provided using a 1-INC LCD Monitor (Model #il272DPB).

**Data Cleaning and Component Activations**

Data was analyzed with independent component analysis (ICA) to investigate ERPs of interest. EEGLAB, ERPLAB and custom matlab scripts were used to analyze the data. The data was first band first bandpass filtered at 0.1Hz 100Hz respectively. Events were aligned to photodiode triggers using measurements of latency provided by the graphics and audio cards at the onset of each pair of tones. Statistical analysis was performed using STATA.

**Method: First task**

The first experimental group of participants (N=8) performed the amplitude modulated tone discrimination task (see Figure 4 for diagram of this task). This task consisted of 10 blocks of 10 trials. During each trial, 6 seconds of a pair of amplitude modulated pure tones was presented, with one tone modulated at 30Hz and other tone at 40Hz, respectively. These tones were selected because of numerous studies indicating gamma frequencies as good target range for observing ASSR (Bharadwaj et al., 2014; Mahajan et al., 2014) and were presented with one unique AM tone in each ear. These AM tones were created using the Matlab wavwrite function and a custom script which controlled range of the tones generated and the sampling rate of sound files in the tone library. The script contains parts adapted from code used in AM exercises found in Dr. William Park’s Matlab web tutorials (Park, 2000). Each trial, participants were told their goal was to indicate, by pushing the left or right arrow key on the keyboard (all participants used a Razor Arctosa low latency keyboard), which ear was being presented with a higher pure tone. For this version of the discrimination task, participants were not informed to use two-hands to complete the response. In order to investigate possible thresholds of tone differentiation, after
each block the difference in frequency between the pair of modulated pure tones was reduced. On the first block, the distance between the tones presented to left and right ear was 10Hz, and this was decreased by 1Hz each trial such that tones presented in the 10th block (on trials 91-100) was only 1Hz difference in frequency (e.g. a 1000 and 1001 Hz pair of modulated tones might be presented in the last block). The lateralization of the pairs of tones (left vs right ear), the tone modulation itself (30Hz vs 40Hz), and block presentation sequence (forward vs reverse) of the pure tone sequences was counterbalanced across 8 participants in the group (2x2x2). The motivation for this task was to determine whether the ASSR elicited by the amplitude modulated tones could be used to to develop the proposed neurofeedback system.

Results: First Task

Behavioral Data

Participant learning appeared very poor or non-existent for the first task. Evidence of learning was tested by comparing performance in the first and last block of trials. Mean accuracy for the first task was low for the first block (M=0.525, SD=0.176) and the last block (M=0.513, SD =0.136) (see Figure 8 and 9). There was no significant difference between blocks, indicating that performance did not improve with experience in the task, t(6)=0.1238, p>.05. Participants reported that they would occasionally hear tones that were exceedingly different in frequency (i.e., tones that sounded further apart than 10Hz difference in pure tone frequency). Participants remarked that sometimes their response would be accurate during these events, but often their feedback would indicate incorrect answer despite very obvious differences in pitch. Reaction times were highly variable between participants and ranged from 0.7 to 3.5 seconds, although reaction times did not vary significantly for each participants themselves (see Figure 7).

Imaging results
EEG data was very poor for all data taken with this amplifier. Due to noise in from the environment and hardware failure of the amplifier, only one data set was suitable for analysis. Therefore the ERPs and frequency analysis was not conducted for this data set.

**Method: Second Task**

A second group of participants (N=7) performed a modified version of the amplitude modulated tone discrimination task (see Figure 5 for diagram of this task). The task was modified in order to address difficulties cited by participants that performed first the task, including claims of hearing tone pairs that were very different. Further investigation of this issue revealed that aliasing errors had occurred when creating the original tones, and that tones were being clipped due to sampling the signals at rate that was well under the maximum for the stimulus computer’s audio processing card (Creative SB X-Fi Audio [0001]; max sample rate 96,000 Hz). Attention to sampling rates and decimating algorithms is an important part of maintaining data integrity and stimulus resolution. Experimenters can avoid undersampling issues by sampling at rates higher than the Nyquist frequency (for a short review on aliasing, see book on EEG methods by Dickter and Kieffaber, 2013, pp. 27-29).

The second version of the discrimination task also consisted of 10 blocks of 10 trials, where each trial consisted of a 6s exposure to a pair of amplitude modulated pure tones. This task was very similar to the original task, except for creation of a new tone library with tones of the max sample rate. However, in addition to increasing the sampling rate of the sound library, we also reduced the frequencies at which the tones were AM, (such that one tone was modulated at 20Hz and other tone at 30Hz). This reduction was prompted by evidence that ASSR can also be attained at 16-23.5Hz as well as 32.5 and 40Hz (Mahajan et al., 2014), as well as by experiments where beta wave activity was targeted for investigation of CAPD (Straus et al., 2008). Each trial,
participants were told their goal was to indicate, by pushing the left or right arrow key on a keyboard, which ear was being presented with a higher pure tone. For this version of the discrimination task, participants were not asked to use two-hands to complete this response. Just as in the previous task, we attempted to find thresholds of tone differentiation, and each block the difference in frequency between the pair of modulated pure tones was reduced. On the first block, the distance between the non-AM aspect of the pure tones presented to left and right ear was 10Hz, and this was decreased by 1Hz each trial such that tones presented in the 10th block (on trials 91-100) was only 1Hz difference in frequency (e.g. a 1000 and 1001 Hz pair of modulated tones might be presented in the last block). A pseudorandom Matlab RAND process was used to generate a new presentation sequence, with the range of possible tones between 1000Hz and 2000Hz. The lower range of tone frequency was chosen to help increase the Nyquist frequency which we could sample just below. The lateralization of the pairs of tones (left vs right ear), the tone modulation itself (20Hz vs 30Hz), and block presentation sequence (forward vs reverse blocks) of the pure tone sequences was to be counterbalanced across 8 participants similar to the first task (2x2x2). The stimulus presentation sequence was identical to the initial task with the exception of the sampling rate of tones and the change in AM. Imaging data from this updated task was to be used to develop the neurofeedback platform.

Results: Second Task

Behavioral Data

Learning for this task was poor. Accuracy remained poor for participants throughout the task (see Figure 11 and 12) although minor decrease in accuracy was observed for the last block. Mean accuracy for the second task was low for the first block (M=0.431, SD=0.040) and the last block (M=0.593, SD =0.201). There was no significant difference between blocks, indicating
that performance did not improve with experience in the task, $t(5)=0.942$, $p>.05$. This may have been a result of the task being so difficult at this threshold (1Hz difference) that participants were simply guessing or unable to perceive any differences. Reaction times ranged from 1.4 to 4.5 seconds and participants did not appear to have any significant changes in reaction time over throughout the experiment.

**Frequency/Time domain analysis**

EEG Data were analyzed for power spectrum effects using the Fast Fourier transform (FFT). Plotting the grand mean of magnitude of the driving frequencies distributed over the entire electrode grid revealed 20/30Hz signal power was increased over C3 (see Figure 17). Plotting grand mean of magnitude of the driving frequencies for each trial showed 20Hz signal was relatively stronger, and that both 20 and 30Hz signal power decreased over the course of the experiment (see Figure 16). This effect may demonstrate a habituation to the 30/20Hz signal over time.

**Method: Third Task**

A third group of participants ($N=7$) performed a modified version of the amplitude modulated tone discrimination task (see Figure 6 for diagram of this task). This task was markedly different from the first two tasks because we removed the amplitude modulation of the tones and did not reduce the distance between pure tones as the stimulus schedule proceeded. Just as before, participants were told that on each trial their goal was to indicate, by pushing the left or right arrow key on a keyboard, which ear was being presented with a higher pure tone. For this version of the discrimination task, participants were asked to use two-hands to complete their response to try to analyze lateralized readiness potential (LRP) as a potential measure of improved task performance. Every block the distance between the frequencies of the pure tones
presented to left and right ear was 10Hz, and this was maintained for all blocks such that tones presented in the 10th block (on trials 91-100) were also 10Hz difference in pure tone frequency (e.g. a 1000 and 1010 Hz pair of non-AM tones might be presented in any block). The rationale for this change was prompted by discovery of a very recent study by Chang et al., (2014) who used a similar but not identical oddball paradigm difference of 8Hz (1000Hz and 1008Hz tones, presented without AM, and as coherent with both ears receiving 1000Hz, or one ear 1000Hz and the other ear 1008Hz). They were able to train participants to unconsciously notice deviant stimuli and cited enhanced auditory mis-match negativity as evidence for this claim (Chang et al., Oct. 2014). A pseudorandom Matlab RAND process was used to generate a new presentation sequence for the reduced range of pure frequencies (1000-2000Hz range, no AM). The lateralization of the pairs of tones (left vs right ear) and block presentation sequence (forward vs reverse order of trials in block) of the pure tone sequences was also counterbalanced across 10 participants with two participants per condition, (2x2 design). Imaging data from this updated task was meant to be used to develop the neurofeedback platform. An effort was made to randomly distribute participants by gender into two full 2x2 groups, as scant evidence exists that female and male participants demonstrate different neural oscillatory behavior (Güntekin & Başar, 2007), although due to time constraints and SONA participant availability this was not able to be implemented.

Results: Third Task

Behavioral Data

Participant learning was also poor for this data. Evidence of learning was tested by comparing performance in the first and last block of trials. Mean accuracy for the second task was low for the first block (M=0.560, SD=0.205) and the last block (M=0.453, SD =0.139).
There was no significant difference between blocks, indicating that performance did not improve with experience in the task, $t(9)=0.758$, $p>.05$. Reaction times were highly variable and ranged between 1.1 and 4 seconds. Participants did not generally react faster as the experiment proceeded and generally took about the same time to respond on earlier trials as later trials (see Figure 13).

**Lateralized readiness potential (LRP)**

Of particular interest in this third iteration of the experiment was the lateralized readiness potential (LRP). The LRP is a reflection of activity in the primary motor cortex and is thought to be useful for predicting the point at which a person was initiating the motor response related to a particular decision about the tone stimulus. This ERP occurs in the hemisphere contralateral to the side of the body where the motor output is engaged. LRP is identified by looking at activity localized over the motor cortex of each hemisphere and its strength is determined by an asymmetrical increase in the potential (Eimer, 1998). LRP was analyzed according to response locked analysis described in literature (Mordkoff & Gianaros, 2000). LRP was well characterized for the left hand, but not for the right hand (see Figure 20 & 21).

**Positive Going Potentials**

Grand averaged data showed possible P200 effects at a latency of 150-270 ms. P200 ERP was analyzed at the FZ, CZ and OZ electrodes for differences between correct trials and incorrect trials for the onset of the tones at that latency range. Paired two-tailed T-test revealed differences between P2 amplitude for incorrect or correct trials were not significant for any of the FZ ($p=.18$), Cz ($p=.59$), or OZ ($p=.91$) electrodes. ERP scalp maps for P200 are visible in Figure 19.

**Negative Going Potentials**
In accordance with literature surrounding the N100 waveform, mean negative amplitude between the first and second positive peaks tends to be larger when participants attend to the tone versus when they are not attending. For auditory attention, N100 effects could be more predominant on trials where the subjects had increased attention to the auditory stimuli (Hansen & Hillyard, 1980). N100 was analyzed for the differences between correct and incorrect trials at the FZ, CZ, OZ electrodes to look for this effect at 90-150ms. No significant difference was found for FZ (p=0.678), CZ (p=.216), or OZ (p=.167) electrodes. ERP scalp maps for N100 are visible in Figure 18.

**Discussion**

**Original Proposal of Experimental Neurofeedback Experiment**

The overarching goal of this project was to evaluate the utility of the steady-state auditory response and auditory ERPs for use in the design of a BCI appropriate for use in the treatment of CAPD. Data were collected over the course of three tasks. There was little evidence of learning in any of the tone discrimination tasks. Furthermore, neither steady-state responses nor ERPs could be used to differentiate task performance. Thus, we did not reach the stage of the project wherein neurofeedback could be provided to participants “online”.

For the online procedure, it was initially proposed that additional participants (separate from these tasks reported on above) would be recruited to undergo neurofeedback or null-neurofeedback (i.e., “sham”) training, for the purposes of evaluating the effectiveness of the BCI system on task performance enhancement. We originally intended to have one experimental group (n=8) undergo neurofeedback training over 4 days and then perform the same auditory discrimination task as performed by the first group. The neurofeedback platform was to consist of an AC EEG amplifier system, an EEG signal acquisition and filtering computer, and a
neurofeedback and stimulus presentation computer connected to the monitor and headphones (see figure 4 for schematic of proposed system). A learning algorithm was to be implemented in the neurofeedback task which was to provide feedback to participants, based on computer analysis of covert electrical activity of their cortex. Participants were to be presented with a short (4-6s) pure tone in one of their ears. This would follow a similar fashion as described in Shibata et al. (2011), where participants would have been asked to, “focus on the tones and somehow regulate activity in their central auditory cortex to maximize the magnitude of the feedback stimulus.” Feedback stimuli indicating performance were to be presented in the form of the size of a circle on a screen, where a bigger circle is indicative of better performance on the neurofeedback training. This is inline with a recent neurofeedback study which showed visual feedback stimuli are effective for training discrimination of auditory stimuli (Chang et al., 2014). After several days of the neurofeedback training, these participants were meant to perform the auditory differentiation task as described above. A third group of participants (N=8) was to be exposed to the same set of stimuli as the neurofeedback group, but no incentive or instruction was to be provided and no neurofeedback process was to be engaged (feedback stimuli was intended to be random noise and not actual EEG measurements). After this ‘null-training’ this group of participants was to take the same auditory differentiation task as other participants. Statistical analysis was to be used to evaluate differences in task performance between the three groups of participants (control, neurofeedback, null-neurofeedback) as a measure of the efficacy of the neurofeedback BCI. There were a number of technical and environmental factors that contributed to the challenges associated with this project.
Limitations

**Hardware availability and functionality.** One of the most limiting aspects of this study was the availability and functionality of the amplifiers used. The first amplifier used was a temporary DC amplifier loaned to Dr. Paul Keiffaber by Dr. John Kimura of Sensorium DBPA (whose company created all amplifiers used in this research). This amplifier was a modified DC amplifier that did not contain a serial port for triggers from the stimulus computer. This made attempts at analysis harder, as the photo-diode input was the only way to timelock stimulus events, and specificity of events (e.g. trial ‘left high frequency 30Hz AM and right low frequency 40Hz AM’) had to be added in after the fact, using custom matlab scripts in conjunction with the stimulus presentation sequences originally generated to provide the correct stimuli on each trial for that condition. Furthermore, during the collection of the data for the first (and second task), severe artifacts were found in the data, where artificial signals of approximately 80Hz were observed, and electrodes were frequently observed to disconnect from the amplifier. We attributed this noise to the proximate drilling associated with construction of the expansion to the William and Mary Integrated Science Center. The drilling often occurred within 10 meters of the recording/task room, and was clearly the cause of these artifacts as not only could we feel and hear the vibrations inside the Faraday cage (at ~80Hz), but these artifacts were not present during times when the drilling was not occurring. Compounding this, drilling would not rule out hardware problems with the first and second amplifier, as with both DC amplifiers electrodes would occasionally disconnected due to DC voltage drift. The first amplifier was immediately removed after it was deemed unsafe for human imaging due to one participant reporting that she could feel a very mild shock. The second amplifier was a custom built DC amplifier purchased by Dr. Kieffaber for eventual use in the hospital at EVMS. In August 2014, this research lab sent
the primary AC amplifier (third/fourth task) to a hospital to run participants for a separate study in a clinical setting. This amplifier was meant to go to the hospital sooner, but hardware problems resulted in the amplifier being sent back to Dr. Kimura for service (hence why the first amplifier was used instead of the lab's standard AC amplifier). Data collection with this amplifier was also cut short (n=4) because of the amplifier being sent to EVMS. Even then, the data collected using the second amplifier was poor due to drilling activities which occurred up until January 2015. Data collected with the third amplifier was robust, and no data sets were omitted from analysis (N=10).

**Issues of Proprietary Software.** Another, and probably the most significant obstacle to implementing the BCI, was the proprietary nature of the software used by Sensorium Inc. to communicate with the amplifier. Simultaneously recording a full dataset while streaming portions of the data into a buffer was not possible because the recording software was required to calibrate the amplifiers each session. Following lengthy discussions with Sensorium, it was ultimately determined that we would not be allowed access to their API in order to interact directly with their amplifier. As a result we were unable to successfully stream data from the amplifier in a way that would be necessary to provide participants with direct, online neurofeedback.

**Logic Errors when Scripting.** Other significant challenges that were only recognized following the third implementation of the program were programming logic errors. One specific error already mentioned was under-sampling (aliasing) of the tones in the first experiment. Investigation of improvements in auditory discrimination proved difficult because it appeared people were not learning the task. This was most likely due to the fact that while participants were making correct responses, the aliasing produced by the audio card manifested in tone
frequencies that were not present in the original stimuli. A more serious error that was only discovered following the third implementation of the task was that the stimulus presentation program was providing participants with inaccurate accuracy feedback. This unintentionally resulted in the feedback reinforcing the exact opposite of what was intended to be learned. (e.g. ‘Correct’ for selecting lower tone, ‘Incorrect’ for selecting higher tone. This effect was not noticed until the final stages of analysis, where the LRP was identified for group 3. Originally LRP was found, but ipsilaterally for right hand responses. This was not inline with literature which shows LRP only occurs contralaterally to the lateralization of the motor action (Mordkoff & Gianaros, 2000). Following investigation of the coding error, it was determined that this error was present for all three of the task implementations. Because of this unfortunate error, it is impossible to determine whether participants were in fact unable to discriminate the tone frequencies, or simply frustrated by the seemingly inappropriate feedback from the stimulus program.

**Future Directions**

**Task 4.** A fourth stimulus presentation should be implemented to examine differences in participants who received erroneous feedback with those who receive correct feedback on trials (Task 4). Data collection is currently proceeding for this task, and analysis should look at differences in accuracy and event related potentials between the groups.

**Error Related Negativity.** The Error related negativity (ERN) may also be a fruitful ERP to consider when developing a BCI. The ERN can be used to identify trials where the participants were either conscious of their error, unconscious of their error, or if participants were answering trials incorrectly despite knowing that they were correct. The ERN is a negative-going deflection of the ERP that occurs when participants make an incorrect response to a
stimulus. This component is robust for trials where the participant is aware they have made a mistake in their answers, and is also present in cases where the participant is unaware of their response (Nieuwenhuis et al., 2001). In this experiment ERN was not available for as the feedback received by participants was always opposite of the intended feedback. Going forward two modifications could be implemented to investigate this. First, implementing time locking of the onset of the feedback stimulus to the actual response, so individuals are reinforced immediately after they press a keyboard response (whereas here they were presented feedback 6 seconds after the task). This would also decrease the amount of time required for each trial, and allow for more trials for each participant and more data for analysis. Second, running a population where the feedback system is corrected (Task 4) would allow analysis of differences in ERN between groups who had correct versus inverted feedback provided on this task.

Mismatch negativity as a metric of auditory learning. After the initiation of this study, Chang et al., (2014) published an article where they report their creation of a auditory BCI very similar to the kind this research sought to intended to investigate. They accomplish this by examining mis-match negativity as the main measure for enhancement of performance on an oddball task, without the participant’s knowledge. Mismatch negativity was not a target for this auditory discrimination task, which involves small differences in pitch for a wide range of frequencies. A modified task, with congruent and deviant stimuli (e.g. 1000Hz pairs 80% and 1010Hz in one ear 20%) should be investigated as another potential target for neurofeedback.

Summary and Conclusions

The design and implementation of a BCI is highly complex and multifaceted in nature. Creation of an effective BCI requires advanced knowledge of computer programming, including coding of stimulus generation and presentation, ethernet UDP/TCP network protocols, EEG...
signal processing, machine learning algorithms, BCI software implementation, and optimization of computational processes in Matlab. This thesis showed the importance and difficulty associated with detection of logic errors early in the coding process; just one incorrectly assigned variable can cause catastrophic degradation of the intended experiment and radically change the implication of empirical results. Also important in the research process is isolating uncooperative environmental factors, where differences in time of day and uncontrollable events in proximity to the imaging area can potentially compromise data integrity. Consistency and observation at all times throughout the experimental process is key to overcoming situations where events related to the participant or environment may taint data. Furthermore, participant availability must be considered when estimating the costs of the experiment, (i.e. participation time, compensation), as it can be hard to find participants willing to return 4 or 5 days in a row for the repeated training required for BCIs.

Creating a BCI also requires a keen understanding of the instrumentation and hardware, not just in the case of collecting and interpreting imaging data, (e.g. functional differences between DC & AC amplifiers), but also in terms of understanding intellectual property and the limits software distributors may face when providing information about the inner-workings of their products. To implement a BCI thus requires access to amplifier systems which can be calibrated and integrated with the chosen 3rd party BCI software; careful consideration must be given to the hardware and software employed.

Going forward this research should build on previous work by improving the experimental task. By syncing the onset of the feedback to the participant response, this would allow more trials per participant in the same amount of time, as well as more robust analysis of well established ERPs such as the ERN. Further research should seek to implement an offline
machine learning algorithm that utilizes reliably established and behaviorally independent EEG measures of CAPD, such as those described by Strauss et al., (2008).

Most importantly, researchers reporting the efficacy of a BCI system must be careful to avoid misrepresenting positive results as indicative of actual improvement of a disorder at a level of biochemical mechanism. Such mistakes violate the empirical process by potentially engaging in weak reverse inference and undermine the validity of results (Poldrack, 2006). An example of this problem from outside the realm of EEG and neuroimaging can be found in pharmaceutical development and testing of new drugs. If a drug shows therapeutic potential researchers must be very careful to screen for side effects, such as complications that may only present under rare or extreme conditions, such as preexisting disorders or contraindications with other drugs or chemicals in the patient. The current science of neurofeedback has very little understanding of the long-term changes in brain activity that result from BCIs, and the observed benefit of a BCI must be weighed against the risk that the process may cause aberrant neural effects that are not as readily observed. Neurofeedback has exciting and real potential as a therapeutic medium for many disorders, including CAPD. Despite the success of some BCIs, more research is required to elucidate mechanisms of EEG neurofeedback at the deepest levels of analysis, and researchers must make efforts to protect participants during all parts of the empirical process.
References


Näätänen, R., Kujala, T., & Winkler, I. (2011). Auditory processing that leads to conscious


Figure 1. Simple mode of an EEG brain computer interface. Recordings taken from electrodes on the scalp are passed from the amplifier to a session of Matlab on the neurofeedback computer. After acquiring data for number of trials (during which null or false feedback is presented), the neurofeedback computer begins to provide the participant with feedback about changes in measures of brain activity compared to the participants baseline activity taken during the initial trials.
Figure 2: Example of amplitude modulated tones. Black lines represent the unmodulated pure tone. The red waveform is overlaid on the black waveform and represents the amplitude modulated version of the original tone. Not to scale.
Figure 3: Schematic of proposed CAPD EEG neurofeedback system. Data from all channels is sent from the amplifier to the recording computer. A session of matlab on the recording computer filters a specific subset of the data (select channels) and passes this information over ethernet TCP to the stimulus computer where each set of new trial data is added to a memory buffer for online analysis. A session of matlab on the stimulus computer performs the BCI loop process, which further filters the data and executes the neurofeedback algorithm to determine the magnitude of the neurofeedback stimulus. Finally the result of the neurofeedback process is displayed on the screen for the participant to see. The process repeats for each trial until training is completed.
Figure 4. Diagram illustrating stimulus scheduling for task 1. The lateralization of specific amplitude modulation of each tone was varied between experimental conditions.
Figure 5. Diagram illustrating stimulus scheduling for task 2. The lateralization of specific amplitude modulation of each tone was varied between experimental conditions, and the modulating frequencies were changed to 20Hz and 30Hz respectively.
Figure 6. Diagram illustrating stimulus scheduling for trials in task 3. Tones were not amplitude modulated for this iteration of the task.
Figure 7. Mean reaction time (seconds) for each participant (Task 1, n=7) during each of the 10 blocks which each contain 10 trials.
Figure 8: Initial calculations of mean accuracy for each participant (Task 1, n=7) during each of the 10 blocks, which each contain 10 trials.
Figure 9. This data represents the corrected mean reaction time (seconds) for each participant (Task 1, n=7) during each of the 10 blocks which each contain 10 trials. The data was corrected by subtracting each measure of accuracy from 1 to give the actual accuracy.
Figure 10. Mean reaction time (seconds) for each participant (Task 2, n=6) during each of the 10 blocks which each contain 10 trials.
Figure 11: Initial calculations of mean accuracy for each participant (Task 2, n=6) during each of the 10 blocks, which each contain 10 trials.
Figure 12. This data represents the corrected mean reaction time (seconds) for each participant (Task 2, n=6) during each of the 10 blocks which each contain 10 trials. The data was corrected by subtracting each measure of accuracy from 1 to give the actual accuracy.
Figure 13. This data represents the corrected mean reaction time (seconds) for each participant (Task 3, n=10) during each of the 10 blocks which each contain 10 trials.
Figure 14: Initial calculations of mean accuracy for each participant (Task 3, n=10) during each of the 10 blocks, which each contain 10 trials.
Figure 15. This data represents the corrected mean reaction time (seconds) for each participant (Task 3, n=10) during each of the 10 blocks which each contain 10 trials. The data was corrected by subtracting each measure of accuracy from 1 to give the actual accuracy.
Figure 16. Grand means of power (microvolts) for 20Hz (blue) and 30Hz (red) signal character averaged for each block (n=4). Data for each trial consists of stimulus onset locked window of -500 to 7000 ms. Two data sets were unavailable for due to data quality.
Figure 17: Topographical plot of 20Hz and 30Hz power in each channel, averaged across all participants. Peak power (dark read) can be seen around the C3 electrode, and is laterally localized above the auditory cortex. Data was calculated for -500ms to 7000ms, time-locked to auditory stimulus onset. Two data sets were unavailable for due to data quality.
Figure 18: Topographical plots of N1 mean ERP for incorrect (top) vs correct (bottom) trials. ERPs were averaged across all participants (Task 3, n=10). Mean amplitude between two peaks was calculated for 90ms to 150ms, time-locked to the auditory stimulus onset.
Figure 19: Topographical plots of N1 mean ERP for incorrect (top) vs correct (bottom) trials. ERPs were averaged across all participants (Task 3, n=10). Mean amplitude between two peaks was calculated for 150ms to 270ms, time-locked to the auditory stimulus onset.
Figure 20. Topographical plots for grand mean of LRP calculated for left and right button pushes. Right hand button pushes reveal no lateralization of activity. Left hand button pushes revealed expected lateralization of the potential over the contralateral (right) hemisphere. LRP was calculated using epoch which were time locked to the participant response (button press) at -120ms to 0ms (task 3, n=10).
Figure 21. LRP waveforms for left and right keyboard presses at C3 and C4 electrodes. Averaged across all participants for task 3 (n=10).
Figure 22. Example of the interface participants interacted with when taking the self-administered hearing test (left). Results of one participants hearing test (right). 20dB was the lowest possible power threshold that the test was able to assay, so hearing range for frequencies rated at 20dB may have been lower (e.g. 15dB). The grading of the scale occurred in 5dB increments, range of scores was (20db, 25dB, 30dB, 35dB). No participant scored higher than 30dB for any tone in the 1000-2000Hz range.