# FINAL REPORT

*Authors: Adrian KC Lee, Mark Wronkiewicz*

**Report Date: December 10, 2015**

**AFOSR Grant FA9550-12-1-0466**

# YIP -- An integrated neuroscience and engineering approach to classifying human brain-states

**Funding period: September 15, 2012 to September 14, 2015**

**Principal Investigator: Adrian KC Lee, ScD**

**Institute for Learning & Brain Sciences; Department of Speech & Hearing Sciences,**

**University of Washington**

## 1. Abstract

Harnessing the capability to read and classify brainwaves into the myriad of possible human cognitive states (referred to as brain-states) has been a long-standing engineering challenge. Brain signals are generally captured non-invasively by electroencephalography (EEG) – a cheap and portable brain imaging tool with time resolution fine enough to track the dynamic changes of different brain-states. While the scientific quest to map human brain function has exploded in the last two decades, the ability to link patterns in EEG signals to specific cognitive states remains elusive, owing perhaps to limited crosstalk between the fields of neuroscience and engineering. Here, we report a framework we developed that leverages the latest neuroscience knowledge to transform the current engineering approach to brain-state classification. We used inverse imaging techniques and surface-based spatial normalization algorithms to interpret brain signals across a large pool of subjects and cross-validated our findings with simulated and actual brain data. We concluded that decoding brain signals in the brain (a.k.a. source-space approach) confers two major benefits compared to classifying brain signals directly on the EEG sensor readings (a.k.a. sensor-space approach): i) it provides a principled method to transfer data from one subject to another, thereby reducing BCI calibration time; and ii) it increases classification accuracy regardless of which dimensionality-reduction techniques were used to preprocess the data. Overall, this innovative approach establishes a formal integrated neuroengineering framework that allows us to capitalize on the similarity in brain function across subjects (a traditional neuroscience approach) and optimally incorporate *a priori* information to maximize classification algorithm performance at an individual level (a traditional engineering goal) that ultimately improves our ability to classify human brain-states.

## 2. Highlights from the previous year

In the previous year, we published a manuscript addressing the goals of Aim 1 (*Test the hypothesis, using modeled EEG data, that implementing a spatial soft-prior in a Bayesian approach that incorporates anatomical information will improve the ability of the classification model to account for the inter-subject EEG patterns due to variability in the geometry of cortical folding across subjects*). Specifically, we found that knowing the spatial location of an active brain region (and incorporating that information through a soft prior) provided a noticeable increase in classification accuracy. We also found that this increase varied depending on the brain region in question. Toward Aim 3 (*Conduct a cognitive experiment using a multimodal imaging approach to assess the detectability of different brain-states*), we successfully classified different brain-states associated with maintaining and switching attention in the context of competing speech streams. This classification was carried out at a single-trial level and also highlighted the benefit of including neuroscience information. We are currently preparing a manuscript describing this work. We have also successfully installed one of the recently developed EEG electrode localization systems (Brainvision) that is photometry based. We expect that this technology will accelerate our ability to extend our findings here to eventual BCI deployment by dramatically reducing co-registration time (a necessary step that takes brain signals from sensor to source space).

## 3. Current state-of-the-art technology

The current BCI field focuses mostly on optimization of existing BCI paradigms. These include P300, motor-imagery, and steady-state visually evoked potentials (SSVEPs), which we will describe here. P300 speller BCIs rely on the "oddball" response that elicits a positive event-related potential (ERP) around 300 ms after a rare stimulus occurs—hence the "P300" name. Users attend to a letter in a matrix of flashing characters to spell one letter at a time and the system attempts to detect P300 responses to the letter attended. In motor-imagery BCIs, the system usually attempts to detect a change in power within the mu cortical-rhythm band (~10 Hz), which is modulated by overt or imagined movement (especially, hand movement). In SSVEP BCIs, the user makes choices by attending to one of several flashing stimuli on a computer screen. These stimuli all flash at slightly different frequencies; by calculating the power in each band over occipital cortex (and primary visual area), it is possible to reliably detect an increase in power of the frequency associated with the attended flashing item.

The field has almost exclusively focused on optimizing these three main BCI paradigms for the past two to three decades. There is good rationale for this optimization: fast and accurate classification improves the communication rate for the user, which is paramount in clinical situations where this might be the patients' only means of interacting with the world. That said, adoption of BCI technology is still not widespread in the clinical (or commercial) world despite a great deal of optimization effort. In this funded work, we specifically address some of the issues that are impeding the progress in this field.

The focus on optimization of a few paradigms also means that the BCI field as a whole reaches a local maximum. While there are likely other systems based on different brain states that could theoretically outperform any of the systems being focused on, the field has become inward looking and concentrated on incremental research. This point is quite salient when reviewing the age of the neuroscience literature underlying current paradigms. The P300 ERP was first described by Chapman & Bragdon, 1964 and Sutton et al., 1965. The mu rhythm used in motor-imagery BCIs was described in the early days of EEG, but it was shown to be reliably manipulated by thumb movement by Pfurtscheller & Aranibar, 1979. Finally, the early implementation of visually evoked potential BCIs was carried out in two parts by Vidal, 1973 and Vidal, 1977. Since then, we have accumulated a wide array of new findings regarding cortical processing, which could serve as priors for new BCI paradigms, but they have not been adequately explored.

While the limited scientific crosstalk is a bottleneck for BCI development, there are also practical issues that require solutions before we can expect BCIs to gain more widespread adoption in both clinical and commercial applications. For instance, long training times limit the amount of time a user has to operate a BCI. Typically, the first 20-30 minutes are spent in this training (or calibration) time where the system builds up the statistical model required to later predict user's intent. In an hour-long session, that translates to one-third or even one-half of the time drained for calibration, which is not tenable for real-world deployment (especially in clinical applications). Another obstacle stems from inter-subject variation. Between different users, we can expect that the signal composition recorded with EEG electrodes will be varied to some degree. This inconsistency arises from variation in both the brain and head tissue (i.e., skull and scalp) anatomy as well as spatial jitter in the exact location of the EEG electrodes in each experiment. All of this variation means that it is difficult to target common patterns of brain activity, or to recycle data across subjects as a way to, for example, cut down the 20-30 minutes of required training.

## 4. Objectives of the overall project

The primary objective of this proposal was to establish an engineering framework that can incorporate neuroscience insight into any brain-state classification algorithm. We view this focused direction as a necessary first step in transforming current BCI approaches so that collectively as a field, we can move towards leveraging latest neuroscience and engineering discoveries to accelerate BCI research. We focused on non-invasive EEG-based BCI approaches because we believe that this will be more easily deployable (compared to other invasive approaches, e.g., electrocorticograms, that require extensive surgical procedures).

Almost all BCI systems decode information directly from EEG sensor readings (known as operating in "sensor-space"). In this project, we utilized a framework developed in the neuroimaging domain to efficiently estimate neural currents in the brain (known as "source-space") from sensor readings. Critically, this approach provided a formal engineering framework to allow neuroscience findings to constrain BCI design. Specifically, we investigated how the inter-subject variability in the geometry of cortical folding affects brain-state classification. Furthermore, we explored whether we can extrapolate brain-state classification algorithms from a group of trained subjects and apply them to untrained subjects (known as the Transfer Learning Problem). We validated these findings by using both simulated and real neural data.

## 5. General methods

The technical basis for this work focuses mainly on the application of source imaging to BCIs. Source imaging is concerned with estimating activity at the surface of the cortex from noninvasive methods. This requires a head model for each subject to construct forward and inverse models, as well as co-registration of EEG electrodes. These methods are described extensively in our paper (Wronkiewicz, Larson, & Lee, 2015), but we describe them briefly below.

*MRI and EEG co-registration*

Structural MRI information was recorded for each subject using a multi-echo magnetization prepared rapid gradient echo (MEMPRAGE) scan as well as two multi-echo multi-flip angle (5° and 30°) fast low-angle shot (FLASH) scans. We then constructed a three-layer boundary element model (BEM) for each subject using this structural information and FreeSurfer software (Dale, Fischl, & Sereno, 1999). This permits a mathematical characterization of how activity from the neural sources propagates through the different tissue layers of the head to the EEG electrodes (Baillet S., 2010). We used a 3Space Fastrak (Polhemus) to record the locations of EEG electrodes, cardinal landmarks (nasion, left/right preauriculars), and other points on the scalp to aid in the co-registration of the electrodes with the BEM's coordinate frame.

*Source imaging*

We modeled the source space using about 7000 current dipoles distributed evenly (spaced approximately 7mm apart) across the entire cortex. Each dipole approximates primary neural current in a small patch of cortex (Hämäläinen, Hari, Ilmoniemi, Knuutila, & Lounasmaa, 1993). The forward solution is then constructed by calculating the effect of each modeled current dipole on each EEG electrode and storing these vectors in the forward (or gain) matrix. To compute the inverse solution (allowing estimation of cortical current from EEG recordings), we first required an estimate of sensor (EEG) noise. Sensor noise covariance matrices were calculated using 200 ms periods preceding the beginning of each trial on a subject-by-subject basis for trial-based analyses (empty-room recordings can also be used to calculate noise covariance of MEG for asynchronous activities such as resting-state analyses). The calculation of an inverse solution is ill-posed as there are many more current dipoles than EEG electrodes (~7000 vs. 60, respectively). Therefore, an extra constraint is required to generate a unique solution (Baillet, Mosher, & Leahy, 2001; Hämäläinen, Hari, Ilmoniemi, Knuutila, & Lounasmaa, 1993). Many techniques have been developed to accomplish this, and we employed the Minimum-Norm Estimate approach using MNE-Python (Gramfort, et al., 2014), which minimizes the difference between the expected and real EEG measurement profile (for a given solution) while also minimizing the solution vector's L2-norm (Hämäläinen, Lin, & Mosher, 2010). This results in a unique, stable, and linear estimate of source activity (Hämäläinen & Ilmoniemi, 1984).

*Behavioral task*

We used EEG data from a previously recorded attentional switching task (Larson and Lee 2014). For details on the experimental paradigm, see the previous manuscript. Briefly, 10 subjects took part in a task to selectively listen to one of two competing auditory streams. Each stream was comprised of spoken letters and delayed by 300 μs in either the right or left ear to create a leftward or rightward spatial percept. Subjects were cued before the trial began to either maintain attention to the same talker (left or right) throughout a trial or to switch their attention to the other stream halfway through during a 600 ms gap period. The task required subjects to report the number of spoken "Es" in the target stream after each trial was complete. We only kept trials with correct behavioral responses for analysis (mean = 38 per condition). Before each trial, listeners were cued to either maintain attention to the same auditory stream (left or right) throughout the trial or switch streams halfway through during a gap period. We used activity from the 600 ms gap period to classify whether the subject switched or maintained his or her spatial auditory attention. The aforementioned paper found that a specific brain region—the right temporoparietal junction (RTPJ)—is significantly more active (in terms of cortical current density) when switching auditory attention (Larson & Lee, 2014). This data was used for two experiments as described below.

*Transfer learning*

Our first goal was to find a way to reduce the 20-30 minute training period. There is some previous work in the BCI field focused on transfer learning, which is the general procedure of recycling data across subjects to reduce or eliminate this training time. While these previous studies used sensor-space information, we instead transferred data through the source space to reduce common sources of anatomical and electrode variability. To accomplish this, we first obtained source estimates for all trials and all subjects. Next, we morphed these source estimates across subjects to a subject of interest. This is a common technique from neuroimaging often referred to as "spatial normalization" because it attempts to account for spatial variance in cortical folding patterns across subjects. For each subject, we then trained a support vector machine (SVM) classifier using only transferred data and tested the classifier using data only from the subject of interest. We used the subject-specific accuracy as our benchmark (i.e., a classifier both trained and tested using data from the subject of interest) as this is the standard method used by the BCI field. We repeated with different amounts of transferred data (using 2-9 subjects) to explore how the size of the training data pool affected performance.

*Single-trial attention switch prediction*

We also used this neuroimaging finding as a prior to define the RTPJ as cortical region of interest (ROI). We then selected cortical activity only in this region to use as a feature in a classification task to predict (on a single-trial level) whether or not subjects switched attention. Unlike the transfer learning experiment, we trained and tested on each subject individually. Features were classified using an SVM as it is robust to highly dimensional data, and we used 10-fold cross validation to obtain stable results. We used a radial basis function and a wide range of reasonable values for C and gamma parameters (controlling for complexity of the decision surface and influence of each sample, respectively). To obtain classification accuracies for each subject, we averaged the 10 cross validation scores and selected the maximum across the two SVM parameters on a subject-by-subject basis (similar to a real-world implementation).

## 6. Results generated from this grant period

*Transfer Learning*

In the transfer learning experiment, our aim was to develop a BCI approach that trained a classifier from pre-existing data belonging only to other subjects (i.e., a "zero-training" approach as no training data is collected from the subject of interest). We found that, as the amount of data transferred from other subjects increases, the classification performance slowly increased and eventually recover the subject-specific benchmark accuracy (at approximately 7 subjects worth of data). Once 8 or 9 subjects were included, the mean accuracy exceeded the subject-specific accuracy (Figure 1). This suggests that we can transfer data across subjects through the source space to meet or exceed the accuracy expected from a standard BCI approach. Importantly, our experimental strategy used no training data from the subject whose data was classified, so this also has the practical advantage of eliminating the 20-30 minute training period if it was implemented in an online experiment.

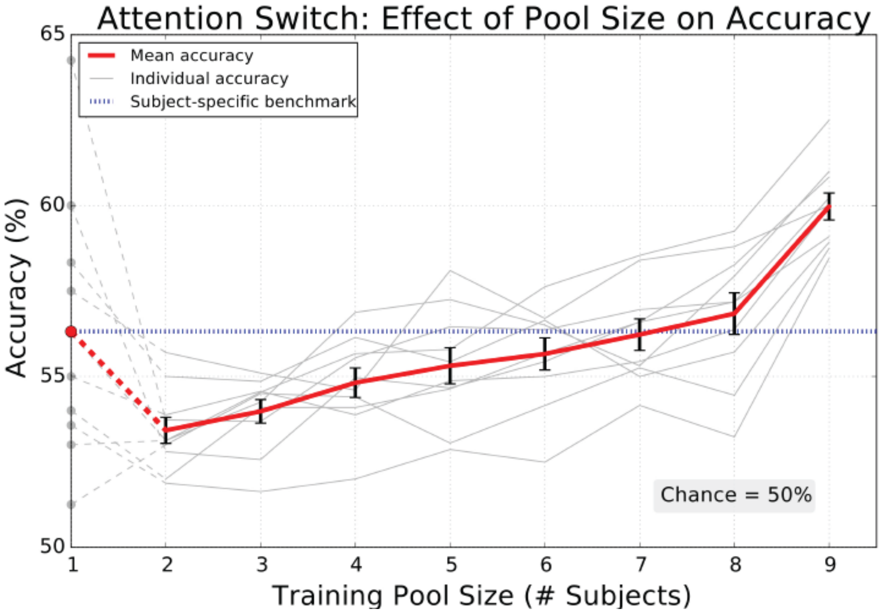

**Figure 1.** A classifier using transfer learning gave incremental performance gains in classifying attentional state and eventually surpassed the subject-specific benchmark. Mean ± SEM (red, black error bars) for 10 individual (gray) subject accuracies. Subject-specific classifier accuracy is indicated by the gray circular markers (left) and the mean is traced across all pool sizes for comparison (blue dotted line). Chance accuracy is 50 percent. Figure published in our previous work (Wronkiewicz, Larson, & Lee, 2015).

*Single-trial attention switch prediction*

To compare sensor- vs. source-based classification approaches, we classified whether or not a subject switched attention during an auditory task. In both approaches, the same dimensionality reduction techniques (ICA, PCA, and CSP) were used with a reasonable range of parameters. We found that selecting activity specifically from the RTPJ for classification yielded a significant boost in classification accuracy (Figure 2).

We found that a source-space classification approach (using individualized MRI information) gave a significant increase in performance over sensor-space in this BCI paradigm. This is likely due to the inclusion of neuroscience priors, which indicated that the RTPJ has significantly higher activity when switching attention to a different sound source. Including neuroscience priors is a pragmatic alternative to a strategy based solely on machine learning. Given an infinite amount of high quality data, we

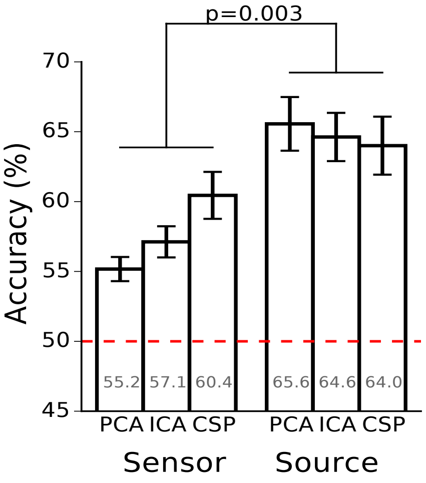

**Figure 2.** Single-trial classification of the same attention switching data using sensor and source space features. The individualized head model performed significantly better than the sensor-based approach (p=0.0032). Error bars are ± SEM and the red line indicates chance.

would expect that certain computational techniques would converge on similar (or more ideal) spatial weights that target the most informative cortical region. However, in BCI research, we are constrained by only having a few dozen trials per class and these trials are usually contaminated with noise. Under this low SNR and low trial-count regime, source imaging is a principled method to target regions of interest and provide the most informative data for efficiently training a classifier in real-world situations.

## 7. Impact in the community

We expect that the work investigated in this YIP award will have a significant long-term impact on the BCI community as well as the neuroscience field at large. We have successfully applied our source-based BCI approach to two different generic problems: transfer learning and single-trial task-related classification, and validated our substantial performance gain via simulated data as well as actual neural recordings. Furthermore, our results point to the ability to obtain reliable results for brain-state classifications (e.g., switching of auditory attention) that depart from the timeworn paradigms (e.g., P300, SSVEP etc.). This allows the BCI community to start thinking about developing new BCI paradigms that are more cognitively based, which ultimately could attract a wider user-base (e.g., serving hearing-aid users as opposed to targeting specifically to locked-in patients). Furthermore, we believe that our work here will inspire engineers to be more versed in brain sciences, as opposed to merely concentrating on the development of machine learning algorithms, thereby putting the “brain” back in the development of brain-computer interfaces.

The formal theoretical neuroengineering framework put forward in this grant is also expected to have ripple effects in the neuroscience community. Thousands of papers are published in the neuroimaging domain every year, yet only a handful would even consider transferring the neuroscience knowledge gain to the advancement of engineering, especially in the BCI domain. By putting forward a framework that allows engineers to formally leverage neuroscience findings as meaningful priors, we are hopeful that our work will truly serve as the interface between brains and computers.

## 8. Future research directions

While the work in this grant has primarily been focused on maximizing signal characterization (across and within subjects) related to the user performing a task (e.g., switching of auditory attention), another way to improve the SNR of our BCI approach is to better characterize the noise characteristics of the brain signals associated with the user not performing a task (e.g., at rest). Investigation of brain regions that are synchronized at rest has grown rapidly in the past decade. This phenomenon was first published by Biswal et al. in 1995, before resting-state network (RSN) research formally began with the discovery of the default mode network (DMN; Raichle, MacLeod, Snyder, Powers, Gusnard, & Shulman, 2001). Until this publication, much of modern human neuroscience was focused on neural responses evoked by stimuli or task conditions. By instead using the brain’s intrinsic activity, RSN research has shown that anatomically separate regions of the brain are functionally connected (or synchronized; van den Heuvel & Hulshoff Pol, 2010; Rosazza & Minati, 2011), giving important insight into the brain’s broader communication scheme (Fries, 2005; Snyder & Raichle, 2012).

Positron emission tomography (PET) and functional MRI (fMRI) have served as the primary imaging techniques for much of the RSN work to date (van den Heuvel & Hulshoff Pol, 2010). Electrophysiological exploration (e.g., M/EEG) of RSN activity is a more recent development in contrast to the hemodynamic methods that dominated early RSN work. Because RSN research using M/EEG is still in its infancy, there is no general consensus on what connectivity patterns (and at which cortical rhythm) correspond to the converging networks characterized by PET and fMRI studies. Furthermore, there can be specific RSNs associated with specific cortical rhythmic bands that can only be revealed by M/EEG due to its higher temporal sampling rate (revealing cortical rhythm well above 100 Hz) compared to PET and fMRI (being able to track rate typically of 0.05 Hz or below).

A future direction stemming from our current work is to accelerate our neuroscience understanding of RSN activity at different cortical rhythmic bands. This can then be used as *a priori* information that can be incorporated into the neuroengineering framework developed in this grant. Specifically, the source-space imaging approach can be further improved if the synchronization across brain regions at rest can be better characterized, resulting in a further SNR gain in brain-state classification.

Another future direction is to investigate how our neuroengineering framework approach to BCI can be implemented for real-time applications. Some recent technological developments make this line of engineering development work promising: 1) our lab has acquired a recently developed fast photometric-based EEG electrode localization system that will accelerate the co-registration process (a necessary step to transform brain data from sensor to source space), thereby making it more feasible for real-time applications; 2) wireless EEG electrodes make it more ergonomical for eventual field deployment of BCI technology.

## 9. Bibliography

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

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

Chapman, R. M., & Bragdon, H. R. (1964). Evoked responses to numerical and non-numerical visual stimuli while problem solving. *Nature , 203* (495), 1155-7.

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

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

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Sciences , 9* (10), 474-80.

Gramfort, A., Luessi, M., Larson, E., Engemann, D., Strohmeier, Brodbeck, C., et al. (2014). MNE software for processing MEG and EEG data. *NeuroImage , 86*, 446-60.

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

Hämäläinen, M., Hari, R., Ilmoniemi, J., Knuutila, J., & Lounasmaa, O. (1993). Magnetoencephalography - theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics , 65* (2), 413-97.

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

LA, F., & E, D. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology , 70* (6), 510-523.

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

Pfurtscheller, G., & Aranibar, A. (1979). Evaluation of event-related desynchronization (ERD) preceding and following voluntary self-paced movement. *Electroencephalography and Clinical Neurophysiology , 46* (2), 138-146.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences , 98* (2), 676-82.

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

Snyder, A. Z., & Raichle, M. E. (2012). A brief history of the resting state: The Washington University Perspective. *NeuroImage , 62* (2), 902-910.

Sutton, S., Braren, M., Zubin, J., & John, E. (1965). Evoked-potential correlates of stimulus uncertainty. *Science , 150* (3700), 1187-8.

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

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

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

Wolpaw, J. R., McFarland, D. J., Neat, G. W., & Forneris, C. A. (1991). An EEG-based brain-computer interface for cursor control. *Electroencephalography and Clinical Neurophysiology , 78* (3), 252-259.

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