A Brain-Computer Interface for Body Turning Movements

Master’s Thesis

Presented to

The Faculty of the Graduate School of Arts and Sciences
Brandeis University
Department of Psychology
Paul DiZio, Advisor

In Partial Fulfillment
of the Requirements for the Degree

Master of Arts
in
Psychology

by
Wanli Cheng

May 2014
Acknowledgement

First and foremost, I would like to express my sincere gratitude to my advisor Professor Paul DiZio and Professor Jim Lackner, who provided funding, space, equipment, training, suggestions, and all resources necessary to complete this research. I would also like to thank Professor Pengyu Hong for his advice on machine learning algorithms.

I will always remain grateful to Alberto Pierobon for his assistance on running the experiments and guidance on signal processing. I am also grateful to Bernie Gabin for his work on the legacy methods.

I would like to acknowledge all members in the Ashton Graybiel Spatial Orientation Laboratory for the help they provided during the course of my study.

This research was supported by the Office of Naval Research under grant ONR SBIR N00014-12-G-0546 0006.
ABSTRACT

A Brain-Computer Interface for Body Turning Movements

A thesis presented to the Department of Psychology

Graduate School of Arts and Sciences
Brandeis University
Waltham, Massachusetts
By Wanli Cheng

This report describes progress toward developing a brain-computer interface for decoding human navigational movements from scalp electrode recordings. In our paradigm, subjects (n=11) sat in a swivel chair, viewed a virtual environment displaying non-spatial information specifying a right or leftward navigational turn, and rotated their body by pushing against the floor with their feet. Signals monitoring chair rotation, electroencephalographic (EEG) activity, and instructed direction were recorded. Our first goal was to apply legacy methods previously used for synchronous decoding of the direction of planar arm movements to decoding the direction of body turning movements. The legacy method 1) extracted 34,020 features from data epochs of -.5 to +1.5 seconds relative to turn specification using combinations of EEG electrodes, band powers, and sliding windows of different sizes, 2) selected the best 100 features using linear discriminant analysis, 3) classified direction using a support vector machine, and 4) computed classification accuracy on movements not used for building the classifier. The average accuracy for decoding left-right movement direction was 82%. Our next goal was to apply a more theoretically motivated information gain criterion for selecting the best features from the same
set of extracted features. We found that classification accuracy asymptoted at 86% with an average of 255 features per subject. We then used the information gain criterion to investigate how to reduce the size of the necessary feature set without sacrificing classification accuracy. We found that EEG mean amplitude, a .5-second window size, and the electrodes in the right frontal quadrant carried more information than any other band power, window size, or quadrant. We then extracted this reduced set of 13 features from just the .5-second epoch preceding movement onset and found a decoding accuracy of 78%. This is the first time navigational turning movements have been decoded from surface EEG. The introduction of an information gain criterion for feature selection has advanced our ability to speed data processing by limiting the number of features needed to achieve classification of movement intent with pre-movement data, both of which are essential steps in moving toward a real time BCI application.

_Keywords:_ Brain-Computer Interface, Electroencephalography, Linear Discriminant Analysis, Information Gain, Support Vector Machine, Virtual Navigation
# Table of Contents

Introduction ............................................................................................................. 1

Method ..................................................................................................................... 8
  Participants ........................................................................................................... 8
  Apparatus .............................................................................................................. 8
  Procedure ............................................................................................................. 10
  Data Processing and Analysis .............................................................................. 11

Results ....................................................................................................................... 14

Discussion ............................................................................................................... 17

Appendices ............................................................................................................. 21

References ............................................................................................................. 31
List of Figures

Figure 1 ................................................................................................................................. 21
Figure 2 ................................................................................................................................. 22
Figure 3 ................................................................................................................................. 23
Figure 4 ................................................................................................................................. 24
Figure 5 ................................................................................................................................. 25
Figure 6 .................................................................................................................................. 26
Figure 7 .................................................................................................................................. 27
Figure 8 .................................................................................................................................. 28
Figure 9 .................................................................................................................................. 29
Figure 10 ................................................................................................................................. 30
Introduction

The question of how to control a robot was posed when the first robot was made, and engineers are still struggling to answer it. Thanks to the advance of brain sensing techniques, brain-robot interfaces have been developed to tap into biological controllers that we don’t understand but that have nevertheless proven to be capable of driving a robotic actuator (Dennis, 2008). A great deal of research has been directed toward decoding reaching movements for the control of robotic arms (Carmena et al., 2003; Chapin, Moxon, Markowitz, & Nicolelis, 1999; Contreras-Vidal, Grossberg, & Bullock, 1997; Fetz, 1999; Hammon, Makeig, Poizner, Todorov, & de Sa, 2008; Kim et al., 2006; Vargas-Irwin et al., 2010; Velliste, Perel, Spalding, Whitford, & Schwartz, 2008), but much less effort has targeted decoding whole-body navigational movements in order to drive remote real or virtual vehicles (Friedman et al., 2007; Galan et al., 2008; LaFleur et al., 2013; Millan Jdel, Renkens, Mourino, & Gerstner, 2004; Rea et al., 2014). Our focus here is on developing a non-invasive brain-computer interface (BCI) that would allow people to drive a virtual or mechanical mobile robot by decoding their movement intentions. Our specific goal is to decode body-turning movements in azimuth. The current project focuses on decoding real, ongoing movements as close to real time as possible.

Our approach is to begin with machine learning methods previously used for decoding reaching movements and to improve these legacy methods by employing a new information theory approach (Schlogl, Neuper, & Pfurtscheller, 2002). The overall improvement we are aiming for with the information theory approach is better decoding accuracy. We also aim to achieve the best possible decoding accuracy with brain signals preceding movement onset, which
is critical both for achieving real-time performance and, ultimately, for decoding intention to move during imagined rather than real movements. We will also use the information theory approach to reduce the feature space and thereby decrease the processing time for the decoding algorithms, which is also important for achieving real time performance as well as, ultimately, for simplifying the signal acquisition side of the BCI.

A number of paradigms, protocols, methods, and algorithms have been used in studies trying to decode planar direction of reaching movements. The signal acquisition systems employed range from invasive microelectrode arrays and electrocorticography (ECoG) to non-invasive electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS) (Besserve et al., 2007; Eric, Gerwin, Jonathan, Jeffrey, & Daniel, 2004; Hochberg et al., 2006; Mellinger et al., 2007; Tomita et al., 2014; Waldert et al., 2007; Wolpaw, McFarland, Neat, & Forneris, 1991). In non-invasive approaches, EEG has high temporal resolution, while fNIRS has better spatial resolution. Several EEG-based BCIs have been developed for decoding reaching movements (Hammon et al., 2008; Waldert et al., 2007).

Another dimension along which BCIs vary is asynchronous versus synchronous (Borisoff, Mason, Bashashati, & Birch, 2004; Diez, Mut, Avila Perona, & Laciar Leber, 2011; Galan et al., 2008). In asynchronous BCIs, spontaneous voluntary movements are decoded from brain signal records that are analyzed with no temporal reference to external stimuli or to the movements themselves. An asynchronous BCI system “reads the users mind” for a specific function with any constraining information, and this BCI mode is sought because it would be as fluid and intuitive as natural movement and it could be used by locked-in individuals with spinal cord injury and amyotrophic lateral sclerosis (ALS). Asynchronous BCIs have proven much harder to build than synchronous BCIs. The implementations that do work are slow and
cumbersome to use (Obermaier, Neuper, Guger, & Pfurtscheller, 2001). They are the only option for locked-in patients, but a better option is needed for general purpose. The work described here develops a synchronous BCI that is temporally bound to real movement onset, but the training paradigm and algorithms can be extended to an asynchronous or semi-synchronous BCI decoding imagined movements in the future.

BCI’s also vary in whether their purpose is to decode discrete or continuous movements (Gribble, Doud, Lucas, Pisansky, & He, 2011; Kim et al., 2006; Royer & He, 2009). Researchers have developed EEG-based systems for continuous control of a computer cursor (McFarland, Kruisienki, Sarnacki, & Wolpaw, 2008; Wolpaw & McFarland, 2004; Wolpaw et al., 1991). This system requires the user’s full attention to the computer display and it remains to be determined whether these systems can be used to accomplish tasks in the real world. BCIs have also been built for classifying continuous real movements like walking (Leeb et al., 2006; Presacco, Forrester, & Contreras-Vidal, 2011), but changing the direction of walking has not been addressed. Our approach to BCI two-dimensional navigation starts with solving a binary, discrete movement classification problem - decoding when a seated individual swivels leftward or rightward by pushing on the floor with both feet. Our BCI training paradigm is extendable to situations where the individual concatenates sequential turns to navigate a complex real environment.

Our approach builds on common protocols and algorithms employed for decoding discrete planar reaching movements from EEG recordings. The neural basis for classifying real movements is event-related desynchronization (ERD). During a resting state, EEG signals as well as other large scale neural signals classically present as oscillatory, which is thought to reflect the synchronized activity of large ensembles of single neurons. ERD is an interruption of
the normal EEG rhythm during an event, such as a movement (Chatrian, Petersen, & Lazarte, 1959; Neuper, Wortz, & Pfurtscheller, 2006; Pfurtscheller & Aranibar, 1977; Pfurtscheller & Lopes da Silva, 1999; Toro et al., 1994). EEG mu (8-12 Hz) and beta (13-30 Hz) band powers are reduced during real movements as well as during imagined or viewed movements after training. The direct current (DC) component (mean level) of the EEG signal has also been shown to decrease prior to voluntary movement, a phenomenon known as readiness potential (RP) (Jasper & Penfield, 1949; Kornhuber & Deecke, 1965; Shibasaki & Hallett, 2006). The RP is seen over the primary and supplementary motor cortices. It is at least tenfold smaller than the amplitude of mu EEG and is usually seen with event-triggered averaging. The RP has a small amplitude component from -1200 to -500 milliseconds and a larger component from -500 to 0 milliseconds, before movement onset.

ERDs and RPs have provided a starting point for a class of BCI approaches to decoding reaching movements. These approaches, which we started with here, use statistical machine learning algorithms to tease apart movement-specific EEG modulation from irrelevant EEG components (Hammon et al., 2008; Millan, Franze, Mourino, Cincotti, & Babiloni, 2002). Our machine learning algorithms used mean amplitude and four EEG frequency bands recorded from 54 electrodes covering the entire scalp, because ERD happens in different frequency bands and locations for different tasks. We evaluated the EEG signals in 4 different sizes of sliding windows within the epoch spanning a movement, because the temporal grain and temporal evolution of ERD, RP, and other possible signals are not well understood. We followed the customary method of extracting all combinations of these features from the recorded signals. Such a feature set is very large, and building a classifier with so many features with limited computer memory and CPU resources can be very slow, which is an obstacle to real-time
performance. So, after these features were extracted, a feature selection phase was employed to create a subset of the most important features before the classifier was built. In the end, the classifier would be built on the hyper-plane of the feature subset.

Why did we think that legacy methods that had been established for decoding arm movements would also work for leg movements? Existing navigational BCIs used P300 or steady state visual evoked potentials (SSVEPs) (Diez et al., 2011; Friedman et al., 2007), but we ruled out using these protocols because they are attention demanding and non-intuitive (Obermaier et al., 2001). We explored the possibility of using an ERD and RP-based protocol, because it is a more intuitive and natural neural interface for operating an external device. ERD and RP are motor phenomena, and both reaching and turning have motor components. Control of an arm is located laterally within the motor cortex, so signals emanating from it are easy to pick up with scalp electrodes. Leg motor cortex is partly buried between the hemispheres but also extends onto the centro-medial surface of the cortex, accessible to surface electrodes. A previous study has shown that left or right leg movements can be decoded with EEG signals (Boord, Craig, Tran, & Nguyen, 2010). We thought there might be some more rostral contributions to decoding turning direction, because spatial embodiment modulates time-frequency activities in the right temporo-parietal junction (Arzy, Thut, Mohr, Michel, & Blanke, 2006).

An issue raised above concerning the legacy methods involved the selection of features to be used in classification. The number of features is usually limited by processing speed, and the criteria for deciding which features are the best is usually arbitrarily set. For example, a commonly used legacy method of feature selection is linear discriminant analysis (LDA) (Duda, Hart, & Stork, 2001; Fisher, 1936; Muller et al., 2008; Parra, Spence, Gerson, & Sajda, 2005). LDA is trying to find a linear combination of features that has the maximum separation, which is
the ratio of between-class variance to within-class variance. However, building an LDA classifier on one single feature did not deal with separation maximization. It was a measure of discriminability of that feature, with the evaluation based on classification accuracy rather than separation. In this case, stochastic fluctuation on a small dataset may cause LDA accuracy to change a lot, so an arbitrary criterion for selection is not necessarily reliable.

One of our main goals here is to explore the use of information theory for feature selection. In information theory, entropy is a measure of uncertainty (Shannon, 1948). The entropy $H(X)$ is defined as

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log_b p(x_i)$$

where $p(x_i)$ is the probability of observing the outcome $x_i$ of the random variable $X$ and $b$ is the base of the logarithm. When base 2, $e$, or 10 is used, the units of entropy are in bits, nats, and dits, respectively. In a binary turning direction classification problem, when all turns are left turns, there is no uncertainty, so the entropy here is zero; when half of the turns are left turns, the uncertainty is maximized, so the entropy is one in this case. Information gain (also called Kullback-Leibler divergence) measures the reduction of entropy by taking into account additional information (Kullback & Leibler, 1951). The information gain $IG(Y|X)$ of a random variable $Y$ given a conditional random variable $X$ is defined as

$$IG(Y|X) = H(Y) - H(Y|X)$$

and conditional entropy $H(Y|X)$ is defined as

$$H(Y|X) = \sum_{i=1}^{n} p(x_i)H(Y|x_i)$$

In the same binary classification problem with entropy equal to one, when electrode Cz is high for all left turns and low for all right turns, the entropy of turning direction is reduced to zero.
given all observations of electrode Cz, so the information gain is one in this scenario. In a word, information theory provides evaluators such as information gain to assess the amount of uncertainty we could reduce by taking into account one specific feature. It has fewer statistical assumptions about the data than LDA and is widely used in other non-BCI classification tasks.

In summary, our goals are first to decode turning intent with legacy methods used for decoding reaching movements, and second to apply information theory to the selection of the best features, with the ultimate goad of using as few signals and features as possible for a successful real time navigational BCI.
Method

Participants

Eleven participants (5 males and 6 females) in an age range of 20 to 70 years, with no known neurological disorders, were recruited from students and staff at Brandeis University. The protocol was approved by the Brandeis IRB, and informed consent was obtained at the beginning of participation.

Apparatus

A g.tec 64-channel EEG system was employed in the experiment. It was composed of 4 multi-modal biosignal amplifiers (model: g.USBamp) with appropriate electrode interface boxes (model: g.GAMMAbox) and power supply, 54 active biosignal electrodes (model: g.LADYbird) mounted on standard 10-20 EEG positions on a fitted cap (model: g.GAMMAcap), 2 active electrooculography (EOG) electrodes positioned near the lateral canthi of both eyes to monitor horizontal eye movements, and 1 ear clip electrode as the reference for the whole system.

A Sparco racing seat (model: EVO VTR CS 918 98) was mounted on a bearing that swiveled freely (see Figure 1). A participant seated in the seat could turn it by pushing against the floor with both feet. The analog angular displacement signals of the seat were measured by a potentiometer installed on the base. Turn magnitude was limited to 45 degrees on each side.

All electrical signals were sampled at 256 Hz. EEG and EOG channels were low-pass filtered for anti-aliasing and notch filtered between 55-65 Hz. EEG channels were also filtered with a high-pass filter at 0.1 Hz and a Laplacian spatial filter with 8 nearest-neighbor electrodes (McFarland, McCane, David, & Wolpaw, 1997). Electrode Afz over the prefrontal cortex was
used as ground. These sampled signals were synchronized with digital signals demarcating events in a navigational display that the participant viewed by means of a customized control and data acquisition program based on BCI2000 (Schalk, McFarland, Hinterberger, Birbaumer, & Wolpaw, 2004).

The navigational display was presented on a ViewSonic 17-inch LCD display (model: VS10047) placed directly in front of the seat and mounted on a cantilevered support attached to the base of the seat. It could maintain its relative position and angle to the participant while the participant was turning the seat. The resolution of the screen was set to 1280x1024.

The screen presented a virtual planar navigation task representing an avatar on a forked path (see Figure 2). The screen was .6 meters from the participant’s eyes, and the forked path fit inside a circle subtending 5° of visual angle, straight ahead. The fork was made up of three straight paths, one vertical and the other two joined to its top end at 120° angles. At the beginning of every trial, the vertical path was gray and the left and right forks were red or green. Which fork appeared in which color varied randomly across trials, under the constraint that the two color combinations appear an equal number of times throughout the entire experiment. An avatar representing the location of the participant on the path always started at the bottom of the vertical path, moved upward along the path, and stopped at a small gap right before the fork junction. Avatar speed was set so that vertical travel time would be 2 seconds. When the avatar stopped at the gap, the vertical path color changed from gray to either red or green. A marker indicating the time of color change and avatar stop was synchronized with EEG signals. The participant then turned the seat toward the fork whose color matched the vertical path. After a 2-second delay, the avatar turned and moved along the path corresponding to either the actual seat turn direction in training blocks or the direction decoded from EEG signals in test blocks.
(described further below). The 2-second delay was included for two reasons: 1) the avatar would not be in motion during the epoch we planned to decode, ensuring that we would be decoding movement executed by the participant rather than visual motion of the avatar; 2) we anticipated computational delays of unknown length in real time classification during test blocks, so we imposed the maximum expected delay during training blocks in order to keep the two conditions equivalent.

Procedure

In the experiment, there were 5 training blocks followed by 5 test blocks for each participant. Each block included 24 trials. For each trial, the participant sat in the swivel seat and guided an avatar onto the color-coded fork of a virtual path by rotating left or right with their feet pushing against the floor. A classifier was built for each participant on EEG signals from training blocks in an offline procedure that took about 10 minutes, while the participant rested in the swivel seat. The classifier was then used to decode EEG signals in test blocks. Seat movement drove the avatar in training blocks, while classification result of EEG signals drove the avatar in test blocks.

At the beginning of each trial, one of the two fork branches was colored red or green, and the avatar started moving upwards from the lower end of the vertical path. As soon as the avatar reached the gap, it stopped moving and the color of the vertical path randomly changed to either red or green. At this point, the participant was instructed to look at the end of the tilted fork whose color matched the vertical path and turn the seat accordingly at the same time. The instructions were to turn as soon as ready, but no emphasis was placed on doing it as fast as possible. The avatar resumed moving two seconds after it had stopped, and turned in whatever
direction the participant had turned. In the end, an instruction would tell the participant to turn the seat back to its original position.

Data Processing and Analysis

Three stages of data processing and analysis will be described. The first stage involved using legacy methods to build classifiers for the direction of turning movements from EEG signals and to determine the accuracy of decoding on an independent set of test data. The second stage involved comparing the legacy method to a new method in which an information gain criterion rather than LDA was used to select features for classification, for the same data set. The third stage of analysis involved using the information gain evaluation criterion as a guide for assessing the degree to which the amount of brain signals acquired and the number of features extracted could be reduced without sacrificing classification accuracy.

Legacy methods: feature extraction. After the training blocks were finished, epochs of data were extracted for each trial including .5 seconds before the stop of the avatar to 1.5 seconds after it, for all 54 EEG electrodes, 2 EOG electrodes, and the seat channel. The mean amplitudes and band powers on theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and low gamma (30-55 Hz) bands were calculated on 126 sliding windows at 4 window sizes (1, .5, .25, and .125 seconds) with .05-second step size, for each EEG electrode in each epoch. In total, 34,020 EEG features (5 bands x 126 windows x 54 electrodes) for 120 movements were extracted from each epoch. No trials were removed as artifacts.

Legacy methods: feature selection and classifier building. For each individual participant, the accuracies of classifiers for turning direction using each feature were calculated with LDA, and the features were placed in rank order from best to worst. Afterwards, the top-ranked 100 features were used to build a Support Vector Machine (SVM) classifier (Lotte,
The SVM classifier was then used to perform classification on test blocks.

**Legacy methods: classification accuracy.** There were three types of classification assessment applied in the experiment: 75/25 percentage-split, 10-fold cross-validation, and training-test. The classifiers of 75/25 percentage-split were built on the first 75% of the training data and assessed on the last 25% of the training data. In 10-fold cross-validation, the training data were randomly split into 10 portions. In each fold, the classifier was built on 9 portions and assessed on the remaining portion. For training-test, the classifiers were built on the training data and assessed on the test data. Percentage split and cross validation were used to estimate the performance of the classifier in upcoming test blocks before actually going through these blocks. The training-test assessments were the ultimate method for evaluating system performance.

After human participant experiments were completed, multiple post hoc analyses were performed on these training and test data. The goals of subsequent stages of analysis were to cull and reduce the number of brain signals acquired and the number of features extracted as much as possible without sacrificing classification accuracy. Since we expected reducing features to decrease classification accuracy, we decided ahead of time that subsequent analysis should eliminate any participants whose turns could not be classified better than chance with the legacy methods.

**Feature selection with LDA versus information gain.** First, the LDA classifier was replaced by an information gain evaluator in feature selection. In the legacy methods, we had rather arbitrarily chosen to use the best 100 features because for most participants in a pilot study the classification accuracy decreased when additional features were used, suggesting the possibility that the data were being overfit. In the information gain method, features with zero
information gain were excluded, because, presumably, removing features with low information gain should have a small impact on classification accuracy. Classification accuracies with incrementally larger sets of top-ranked features were assessed. Finally, pairwise comparisons were made between classification accuracies of the information gain approach and the legacy methods.

**Assessment of best features using information gain.** The features extracted from the data can be characterized along four dimensions: electrode location, frequency band, sliding window size, and sliding window epoch. First we counted the number of features with non-zero information gain within sub-ranges of each dimension, for each participant, and conducted analyses of variance (ANOVAs) or t-tests to identify electrode subsets, frequency subsets, window sizes, and epochs that were significantly more likely than others to have high information gain. Then we used ANOVAs or t-tests to compare the classification accuracy with the best subset to classification accuracy with all features.
Results

The legacy feature extraction, feature selection, and classification approach that was successful in decoding the direction of planar arm movements in a previous experiment worked for decoding turning direction. The average online classification accuracy for all 11 participants was .788 (SD=.105), which was significantly higher than chance (t(10)=9.117, p<.001). The criterion for better-than-chance accuracy was derived with a Monte Carlo method. First, we randomly shuffled the time series of data in the -.5 to 1.5 seconds epoch for every trial, for every electrode, for every participant, in both the training and test blocks. We then built and evaluated classifiers for each participant and computed the mean accuracy across participants. This process was repeated 2,000 times to produce an empirical sampling distribution of mean accuracies. The upper bound of the .95 confidence interval for this distribution was .593 accuracy. This was depicted as a horizontal black line in Figure 3. Only one participant, MC, was below chance, with training (75/25 percentage split) and test accuracies of .5 and .517 respectively.

Several offline analyses were then performed to determine the optimal combination of features that could improve classification accuracy and reduce processing time. Because we wanted to explore improvements related to various feature selection methods, we excluded participant MC for whom the legacy method performed below chance level.

The information gain evaluator, in replacement of the LDA classifier, was used to rank each feature during feature selection. Different numbers of top-ranked features were used to build classifiers, and the optimal number was determined according to their classification accuracies (see Figure 4). Classification accuracy peaked at 255 features selected according to
the information gain criterion, whereas classification accuracy in the legacy methods had peaked at about 100 features selected with LDA. The classification accuracy of the new SVM classifier (M=.858, SD=.061) built on 255 features was significantly higher than the accuracy of the old SVM classifier (M=.816, SD=.057) built on 100 features with the legacy methods (t(9)=2.505, p=.034). It was also highly correlated with the accuracies of 75/25 percentage split (r=.769, p=.009) and 10-fold cross validation (r=.677, p=.032).

To determine the most useful subset of features in each dimension, the percentages of features that had non-zero information gain in various sub-ranges were calculated. For frequency bands (see Figure 5), the percentage of mean amplitude features was significantly higher than that of band power features of the low gamma band (F(1,9)=9.815, p=.012), which, in turn, was significantly higher than the percentage of all other band power features (F(1,9)=15.072, p=.004). For electrode locations (see Figure 7), the percentage of features that came from electrodes over the right hemisphere was significantly higher than those over the left hemisphere (F(1,9)=95.304, p<.001), but electrodes over the frontal lobe were not significantly different from electrodes over the parietal lobe (F(1,9)=.154, p=.704). There was a significant interaction between left/right and frontal/parietal directions (F(1,9)=44.670, p<.001). The different between right and left frontal electrodes was larger than the difference between right and left parietal electrodes. For different window sizes (see Figure 6), the percentage of 0.5-second window features was significantly higher than that of features in other window sizes (F(1,9)=60.126, p<.001). Using features from only the .5-second windows reduced the number of features by approximately 75% relative to using windows of all sizes, and the resulting classification accuracy (M=.894, SD=.051) was significantly higher than the classification accuracy of the legacy methods (t(9)=4.860, p=.001).
Lastly, subsets with the most non-zero information gain features in each dimension (.5-second windows, mean amplitude, right frontal electrodes) were examined with classification in different segments of the entire 2-second data epoch for each trial (see Figure 8). Classification accuracy using only the .5-second pre-movement epoch of each trial was compared to classification accuracy with the entire 2-second epoch. Note that -.5 seconds here is referenced to real seat movement and not the moment when the color-coded directional cue appeared, as in Figure 6. The average latency to seat movement following the directional cue was .796 seconds (SD=.278), and the comparable eye movement latency was .449 seconds (SD=.226). The result of a 3 (feature sets) x 2 (epoch ranges) repeated measures ANOVA indicated that both feature sets (F(2,18)=9.132, p=.002) and epoch ranges (F(1,9)=10.985, p=.009) had main effects, but there was no interaction between these two factors. With increasingly fewer features in the feature set, the classification accuracy dropped significantly both from using all bands (M=.845, SD=.089) to using only mean amplitude (M=.838, SD=.092) (F(1,9)=5.433, p=.045), and from using all electrodes to using only right parietal electrodes (M=.784, SD=.092) (F(1,9)=8.388, p=.018). Although the average classification accuracy using the .5-second pre-movement epoch was lower, there were much fewer features extracted from the EEG signals.
Discussion

These results provided the first successful decoding of a binary class of voluntary body turning movements from surface EEG signals for a seated participant. No information about turning amplitude, speed, acceleration, or force was collected or used in classification, so it could be inferred that the participant’s high-level intention to turn left or rightward, rather than trajectory details of movements, was decoded. The legacy methods that had worked for decoding reaching movements also worked here for decoding turning movements. By replacing the LDA classifier with the information gain evaluator, we were able to achieve higher classification accuracy of turning movements than the legacy methods. Moreover, the information gain criterion allowed us to identify the most important subset of features in this task. Reducing the ratio of features to classification accuracy decreased computation time, which was important for future development of real time BCI systems.

Our method for culling features using the information gain criterion required two steps. First we calculated which subset of features along a given feature dimension (electrode, frequency, window size, window location) had the highest information gain. However, it was not mathematically valid simply to sum up the information gain of a subset of individual features to get the information gain of the whole subset, because individual features could be correlated and carry some redundant information. Calculating the percentage of features with non-zero information gain was based on the idea that since we had extracted many overlapping features, important information would span across many features. Thus we confirmed the usefulness of
potentially good subsets by using them to build classifiers and then ultimately evaluating the accuracy of these classifiers on a test set of data.

Using the information gain criterion allowed us to determine that the mean amplitude of EEG signals was more important than any of the band powers. The low gamma band also provided high information gain and above-chance classification accuracy, but combining the mean amplitude and low gamma band power did not result in higher accuracy than using the mean alone. Thus, a strategy of using just the mean amplitude provided the best ratio of accuracy to number of features. The signals and algorithms used in legacy methods were designed to detect ERD, which was usually observed in specific frequency bands, especially low frequencies, and past BCI experiments had observed discriminating information in these bands. We found discriminating information in the low gamma band, but we found more discriminating information in the mean amplitude than other studies had found. Our finding in the mean amplitude was not unexpected because the readiness potential was a DC phenomenon that had previously been shown correlated with the onset of voluntary movement, though not the direction of movements. The evaluations we conducted with the information gain criterion also showed that a window size of .5 seconds for computing mean amplitude and band power was better than smaller or larger window sizes. The .5-second optimal window size was consistent with the known time courses of ERD and RP (Chatrian et al., 1959; Jasper & Penfield, 1949; Kornhuber & Deecke, 1965; Neuper et al., 2006; Pfurtscheller & Aranibar, 1977; Pfurtscheller & Lopes da Silva, 1999).

We also used the information gain criterion to show that classification accuracy was the highest when all electrodes were used relative to when electrodes in any single quadrant were used, equating for the window size, temporal location, and frequency band. On the other hand,
the right frontal quadrant of electrodes provided above chance classification accuracy on their own, at a slightly low accuracy than all electrodes and at a better accuracy than any other individual quadrant. The sufficiency of the right frontal electrode subgroup for classification of turning direction ruled out the possibility that we were just decoding which leg was used, because leg-related motor signals would be expected to be bilaterally modulated. The right hemisphere was suited for encoding turning directions because of its specialization for spatial functions. These considerations supported our contention that we were decoding task-level intention to turn left or right rather than low-level signals about which leg was used. The sufficiency of the right quadrant electrodes was also important because it provided a further culling of the number of features needed for classification and a further improvement of the accuracy to feature ratio.

Finally, we found that the pre-movement epoch alone was unexpectedly sufficient for decoding the intended direction of turning movements better than chance, although decoding accuracy was better using the whole epoch. Being able to decode with pre-movement epoch alone was very important, because it meant that we were decoding the planning of the movement instead of the movement execution. It had practical significance because it further reduced number of features per electrode needed for classification. It also helped rule out the possibility that we were decoding eye movements rather than body turning movements. In most of the trials in our experiment, saccadic eye movements preceded seat movements. A majority of the points in Figure 10 were above the black diagonal through the origin, in an area where eye movements preceded seat movements. However, the points in Figure 10 that were above the upper gray diagonal were trials where the eye movements occurred more than .5 seconds before the seat movements and were therefore not included in the epoch that provided sufficient classification.
Several participants (e.g. BH, EK, and ZG) whose turns were classified successfully (see Figure 9) fall mostly in the region of Figure 10 where eye movements were outside the .5-second window.

The investigation conducted here using the information gain criterion involved offline, post-processing of data collected while the participant was performing the turning task. The cumulative result of these post-processing analyses showed that we began by extracting 34,020 features for every trial and identified 13 features that were sufficient for decoding body turning movements better than chance - 13 electrodes, just the mean amplitude for each electrode, in a single .5-second window preceding movement onset. The implication was that in future experiments, we could use just electrodes in the right frontal quadrant and extract a total of 13 features per trial. Feature extraction was a major time limiting step in BCI operation, and this 2000-fold improvement in the accuracy to feature ratio would greatly facilitate real time operation. For example it might eliminate the need for a synchronizing signal from a real movement, making it possible to decode imagined movements. With just 13 features it would not be computationally prohibitive to run the classification algorithm for every single .5-second window in a data stream. In addition to the feature set being small, the critical features we found were in the pre-movement epoch feature, which raised the possibility of classifying intended movement and controlling a remote actuator or visual avatar in true real time.
Appendices

Figure 1
The racing seat and the LCD display mounted on a cantilevered support attached to the base of the racing seat.
Figure 2

The virtual planar navigation task that the participant saw in the experiment. This screenshot was taken when the avatar had stopped at the gap and the vertical path had changed its color. In this trial, the participant should turn the seat leftward.
Figure 3

The accuracy of online movement predictions for each subject. The training data were split into two parts - the first 75% and the last 25%. The classifiers were built on the first 75%. The training accuracy was assessed on the last 25%, and the test accuracy was assessed on the full test data.
Figure 4

The average accuracy of movement predictions across ten subjects when using different numbers of top-ranked features in classification. Each feature was ranked according to its information gain.
Figure 5
The percentage of features belonging to each frequency band for all features with non-zero information gain. The values are averaged across ten subjects, and the error bars represent standard deviations.
Figure 6

The percentage of features belonging to each time window for all features with non-zero information gain. The values are averaged across ten subjects. Zero on the time axis is the moment that information was provided about which direction to turn. The average time of movement onset was .796 seconds.
Figure 7

The percentage of features belonging to each electrode for all features with non-zero information gain. The values are averaged across ten subjects. Larger circles represent higher means; darker color represents lower standard deviations.
The classification accuracies with different feature sets (all bands with all electrodes, mean amplitude with all electrodes, and mean amplitude with right frontal electrodes) and epoch ranges (2-second epoch and .5-second pre-motion epoch). The classification accuracy of the legacy methods is also added for reference. The values are averaged across ten participants, and the error bars represent standard errors. The number near each data point is the amount of features that need to be extracted from EEG signals.

Figure 8
Figure 9

Classification accuracy on test data with the legacy methods (in blue), and using right frontal electrodes (Af4, F2, F4, F6, F8, Fc2, Fc4, Fc6, Fc8, C2, C4, C6, and C8), mean amplitude, and .5-second window size in the 2-second epoch (in green) and in the .5-second pre-movement epoch (in cyan).
Figure 10

Scatterplot of eye movement latency versus chair movement latency on training data. Both latencies are referred to the moment when the avatar stopped at the gap and the direction of the turn was specified. Points above the black diagonal through the origin represent cases where eye movements preceded chair movement. Points above the gray diagonal are cases where eye movements happened more than .5 seconds before chair movement.
References


