

A Feedback Information-Theoretic Approach to the Design of Brain-Computer Interfaces

**Cyrus Omar, Abdullah Akce, Miles Johnson, Timothy Bretl,
Rui Ma, Edward Maclin, Martin McCormick, and
Todd P. Coleman**
University of Illinois, Urbana-Champaign

This article presents a new approach to designing brain-computer interfaces (BCIs) that explicitly accounts for both the uncertainty of neural signals and the important role of sensory feedback. This approach views a BCI as the means by which users communicate intent to an external device and models intent as a string in an ordered symbolic language. This abstraction allows the problem of designing a BCI to be reformulated as the problem of designing a reliable communication protocol using tools from feedback information theory. Here, this protocol is given by a posterior matching scheme. This scheme is not only provably optimal but also easily understood and implemented by a human user. Experimental validation is provided by an interface for text entry and an interface for tracing smooth planar curves, where input is taken in each case from an electroencephalograph during left- and right-hand motor imagery.

1. INTRODUCTION

A brain-computer interface (BCI) is a direct communication pathway between a human user and an external device (Hatsopoulos & Donoghue, 2009; Wolpaw, 2007). It has three principal components: a sensor that measures the user's neural activity, an algorithm that maps these measurements to control signals for the external device, and a mechanism that provides feedback to the user about the device's resulting state. The measurements may come from noninvasive

This research has been sponsored in part to Timothy Bretl and Todd P. Coleman by a Seed Grant from the Center for Healthy Minds, funded through NIH/NIA under Award No. P30-AG023101; to Timothy Bretl by awards NSF-CNS-0931871 and NSF-CMMI-0956362-EAGER; to Todd P. Coleman by the AFOSR Complex Networks Program via Award No. FA9550-08-1-0079; and to Cyrus Omar by the NSF Graduate Research Fellowship.

Correspondence should be addressed to Rui Ma, Coordinated Science Laboratory, University of Illinois, Urbana-Champaign, 1308 West Main Street, Urbana, IL 61801. E-mail: ruima2@illinois.edu

sensors like an electroencephalograph (EEG) that observes the gross electrical activity of many neurons (Fabiani, McFarland, Wolpaw, & Pfurtscheller, 2004), or they may come from more invasive sensors like an electrocorticograph (Leuthardt, Miller, Schalk, Rao, & Ojemann, 2006) or intracortical electrodes, often placed in the primary motor cortex, which can observe ensemble spiking of individual neurons (Carmena et al., 2003; Kemere, Shenoy, & Meng, 2004; Schwartz, 2004; Serruya, Hatsopoulos, Paninski, Fellows, & Donoghue, 2002). The feedback may be provided by a graphical display, by vibrotactile arrays (Chatterjee, Aggarwal, Ramos, Acharya, & Thakor, 2007), by physical coupling with the external device, or by direct cortical stimulation (Dhillon & Horch, 2005; Inmann & Haugland, 2004; Sinkjaer, Haugland, Inmann, Hansen, & Nielsen, 2003). BCIs have been used to control a growing array of external devices that include computer cursors (Fabiani et al., 2004), text spellers (Sellers, Kubler, & Donchin, 2006), artificial limbs (Kuiken et al., 2009), humanoid robots (Bell, Shenoy, Chalodhorn, & Rao, 2008), and wheelchairs (Iturrate, Antelis, Kubler, & Minguez, 2009).

As implied by the sampling of work just given, the past two decades have seen tremendous excitement and progress in the study of BCIs. In addition to fundamental advances in neuroscience and to improvements in brain imaging technology, this progress has been driven in part by the application of statistical signal processing and optimal estimation algorithms that more systematically compute the mapping from neural activity to control signals. For example, it has become more common to model the desired control signal (imagined by the user) as a random process (e.g., a Markov process) and to model neural activity as a noisy measurement of this process. This approach allows the use of recursive estimation algorithms (e.g., Kalman, particle, or point process filters) to infer the desired control signal, which is then used to drive the external device (Brockwell, Rojas, & Kass, 2004; Eden, Frank, Barbieri, Solo, & Brown, 2004; Wu, Gao, Bienenstock, Donoghue, & Black, 2006). This approach can also be extended to consider higher levels of user intent, for example, as might be associated with goal-directed reaching movements (Srinivasan & Brown, 2007; Srinivasan, Eden, Willsky, & Brown, 2006).

However, these existing approaches fail to consider how desired control signals change in response to sensory feedback. Assume that the user has some high-level intent (e.g., to type the word *raindrop*), which is unknown to the external device. Because mistakes will inevitably be made in the interpretation of neural activity, the operation of the external device may deviate from this high-level intent (e.g., by instead producing the word *ruin*). At this point, the user's control strategy may change and subsequent control signals may follow a completely different pattern (e.g., by trying to correct the error before proceeding). In this context, it is critically important that the user and the external device agree on a protocol that specifies both what sensory feedback is provided to the user and how the user should react to this feedback in pursuit of his or her high-level intent.

A general way to design such a protocol begins by modeling a BCI as two agents cooperating to achieve a common goal. The user is one agent and the external device is the other. Both agents have a common objective (e.g., move the cursor to a goal position in minimum time) but have access to different pieces of information (e.g., only the user knows where the goal position actually is). This type of

“sequential, non-classical, team decision problem” (Ho & Chu, 1971) is notoriously difficult to solve (Papadimitriou & Tsitsiklis, 1986).

There are many commonly used BCIs, however, for which this difficult team decision problem can be expressed more simply as the problem of communication over a noisy channel with feedback. The following two BCIs are representative members of this class with broad applicability, and are used as examples throughout this article: (a) an interface for text entry, where an entire string of text is specified before it is expressed by, for instance, a speech synthesizer, and (b) an interface for tracing smooth planar curves, where the entire curve is specified before it is followed by, for instance, a powered wheelchair.

For both of these BCIs, we derive communication protocols—saying exactly how to choose desired control signals in response to sensory feedback—that not only are provably optimal but also are easy for a human user to implement. These protocols are based on a recently discovered principle in the information theory literature called *posterior matching* (Coleman, 2009; Shayevitz & Feder, in press), which we have already suggested may be applicable to BCI (Akce, Johnson, & Bretl, 2010; Omar, Johnson, Bretl, & Coleman, 2008a, 2008b). A key property of the posterior matching scheme is that it requires the user only to compare their intent with the external device’s best estimate of this intent—presented as sensory feedback—to choose a desired control signal.

It is important to emphasize that BCIs are complex systems that involve a number of design choices. It is often difficult to understand how these choices interact and in particular how to decide which choices are most critical to performance. The appeal of our approach is that it leads to a systematic procedure for making design choices that are, at least in principal, provably optimal. Even if the two BCIs we present here (interfaces for text entry and for tracing smooth curves) do not yet provide superior performance, we hope that our underlying philosophy—making design more systematic and theoretically well grounded using tools from stochastic optimal control and information theory—is adopted and generalized by the broader community.

In this article we first present the theory behind our approach to designing BCIs (section 2), then describe a set of experiments with human subjects that validate this approach for two example BCIs (sections 3 and 4), and finally address a number of questions raised by these experiments that may impact future studies (section 5).

2. THEORY

In this section we reformulate the problem of BCI design as the problem of deriving an optimal communication protocol. This approach will tell us, in a systematic way, exactly how measurements of neural activity should be mapped to control signals for the external device and exactly what type of sensory feedback should be provided to the user. It has application when (a) the communication of intent can be separated from the execution of intent and (b) intent can be modeled as a string in an ordered symbolic language. These two conditions are satisfied by a wide class of BCIs, including the ones for text entry and for tracing smooth planar curves that are considered in this article.

2.1. Notation

In the remainder of this section we use standard notation from the theory of probability and random processes, which we collect here for convenience (Grimmett & Stirzaker, 2001). We are given a probability space $(\Omega, \mathcal{F}, \mathbb{P}(\cdot))$ that consists of a sample space Ω , a set \mathcal{F} of subsets (events) of Ω , and a probability measure $\mathbb{P}(\cdot)$ that maps events in \mathcal{F} to points on the $[0, 1]$ line. With respect to this probability space, we denote random variables by uppercase letters X taking values in some set \mathcal{X} , and we denote specific realizations of these variables by lowercase letters $x \in \mathcal{X}$. The cumulative distribution function (CDF) F_X of a random variable X is

$$F_X(x) = \mathbb{P}(X \leq x).$$

The probability mass function $P_X(x)$ for a discrete random variable and the probability density function $f_X(x)$ for a continuous random variable are

$$P_X(x) = \mathbb{P}(X = x) \quad \text{and} \quad f_X(x) = \lim_{\Delta \rightarrow 0} \frac{\mathbb{P}(X \in [x, x + \Delta))}{\Delta},$$

respectively. Given another random variable Y , the conditional probability mass function and probability density function are

$$P_{X|Y}(x|y) = \mathbb{P}(X = x|Y = y) \quad \text{and} \quad f_{X|Y}(x|y) = \lim_{\Delta \rightarrow 0} \frac{\mathbb{P}(X \in [x, x + \Delta)|Y = y)}{\Delta},$$

respectively. We denote an infinite sequence of random variables by

$$\underline{X} = (X_1, X_2, \dots)$$

and a finite subsequence by

$$X_i^j = (X_i, \dots, X_j).$$

We also abbreviate

$$X^n = X_1^n = (X_1, \dots, X_n).$$

We use this same notation for a sequence of specific realizations, e.g., \underline{x} , x_i^j , and x^n . Finally, we recall the following formula from the chain rule for conditional probability:

$$P_{X^n}(x^n) = \prod_{i=1}^n P_{X_i|X^{i-1}}(x_i|x^{i-1}). \quad (1)$$

2.2. Problem Formulation

In both of the BCIs we consider, it is reasonable for the external device to wait until the user’s intent is clear before taking action. If the device is a speech synthesizer, *intent* is a string of text that must be understood by the interface before it is spoken. If the device is a powered wheelchair, *intent* is a desired path that must be specified before it is followed. We therefore view the purpose of these BCIs as communication and proceed to show how they can be modeled formally as communication channels.

Model intent as a discrete random process. We describe intent as a sequence $\underline{Z} = (Z_1, Z_2, \dots)$, where each Z_i lies in a finite alphabet $\Sigma = \{\sigma_1, \dots, \sigma_m\}$. It is clear that any string of text can be expressed as a sequence of symbols from the usual finite alphabet in which $\sigma_1 = a, \sigma_2 = b$, and so on. Similarly, any smooth two-dimensional curve can be approximated by a sequence of arcs with fixed length d and constant curvature κ_i chosen from a finite set $\{c_1, \dots, c_m\}$, where d and m are parameters (Figure 1a). If we associate the symbol σ_i with the arc generated by

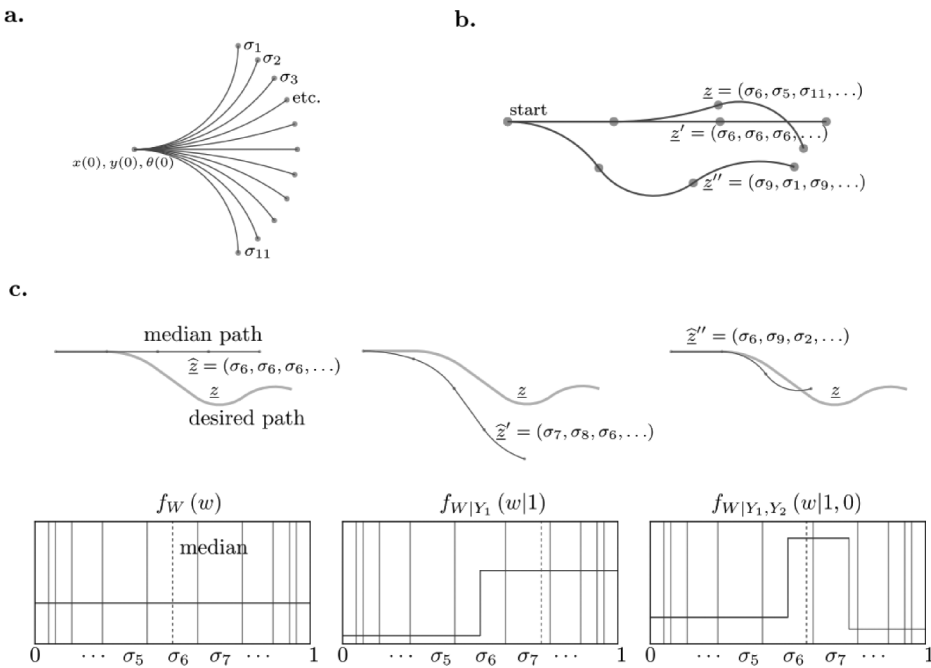


FIGURE 1 Illustration of the interface for tracing smooth planar curves. *Note.* (a) The alphabet used to express smooth planar curves. (b) Three example curves, $\underline{z} < \underline{z}' < \underline{z}''$, to illustrate the lexicographic ordering on curves. (c) Three snapshots of the interface. The interface maintains a posterior distribution over the unit interval and displays the path corresponding to the median of this distribution. The user responds by comparing the desired path to the median path using lexicographic ordering. See text for details.

each c_i , then this curve—just like text—can be expressed as a sequence of symbols from the resulting alphabet Σ .

To capture the fact that not all strings of text and not all curves are equally likely, we further assume that the user's intent \underline{z} is a random process generated by a given statistical model $P_{\underline{z}}(\underline{z})$. Following from (1), this model provides the length- n statistics

$$P_{Z^n}(z^n) = \prod_{i=1}^n P_{Z_i|Z^{i-1}}(z_i|z^{i-1}) \quad (2)$$

for any n . For convenience, we often assume a k th-order Markov model in which

$$P_{Z_i|Z^{i-1}}(z_i|z^{i-1}) = \prod_{i=1}^n P_{Z_i|Z_{i-k}^{i-1}}(z_i|z_{i-k}^{i-1}). \quad (3)$$

In practice, we compute such a model from data using prediction by partial matching (Begleiter, El-Yaniv, & Yona, 2004; Cleary & Witten, 1984; Moffat, 1990). Note that \underline{z} only appears random to the interface—to the user, intent is a specific, known realization \underline{z} to be communicated (i.e., a particular string of text or desired curve).

A key property of the alphabets Σ we consider—which is very important in the implementation of our solution approach—is that they admit a natural lexicographic ordering on sequences \underline{z} . In particular, given the ordering $\sigma_1 < \sigma_2 < \dots < \sigma_m$ on Σ , we define the usual lexicographic ordering on \underline{z} as follows:

$$\begin{aligned} \underline{z} < \underline{z}' & \text{ if } z_i < z'_i \text{ at the first index } i \text{ for which } z_i \neq z'_i \\ \underline{z} > \underline{z}' & \text{ if } z_i > z'_i \text{ at the first index } i \text{ for which } z_i \neq z'_i \\ \underline{z} = \underline{z}' & \text{ otherwise.} \end{aligned}$$

For text, this ordering corresponds to alphabetization. For curves, this ordering has an equally natural interpretation (Figure 1b)—we say $\underline{z} < \underline{z}'$ if the curve specified by \underline{z} turns left at the first point at which it differs from the curve specified by \underline{z}' .

Model motor imagery in EEG as a binary symmetric channel. The BCIs we consider use EEG signals to get input from the human user. A binary classifier attempts to distinguish between left- and right-hand motor imagery in the brain based on these signals, with some chance of making a mistake. We model this process as a binary symmetric channel, a standard communication channel model from information theory (Cover & Thomas, 2006). The k th input to this channel is a random variable $X_k \in \{0, 1\}$ that reflects the user's motor imagery, where $x_k = 0$ corresponds to left motor imagery and $x_k = 1$ corresponds to right motor imagery. The k th output is a random variable $Y_k \in \{0, 1\}$ that reflects the classifier's attempt to infer the user's motor imagery, where $x_k = y_k$ is a correct inference and $x_k \neq y_k$ is an incorrect inference. The probability of error is given by

$$P_{Y_k|Y^{k-1}, X^k}(y_k|y^{k-1}, x^k) = P_{Y_k|X_k}(y_k|x_k) = \begin{cases} 1 - \varepsilon & \text{if } y_k = x_k \\ \varepsilon & \text{otherwise,} \end{cases}$$

where ε is a parameter that can be computed from training data. Note that the channel inputs (X_1, X_2, \dots) are not the same as the sequence of symbols (Z_1, Z_2, \dots) that describe intent. The former are the means for communicating the latter.

Model the graphical display as noiseless feedback. The BCIs we consider provide visual feedback with a graphical display. This feedback is helpful because it allows the user both to correct errors made by the BCI in the classification of motor imagery and to avoid needless redundancy if errors were not made. We will say precisely how the user should do this in the sequel (*Solution Approach*), but for now we simply point out that knowledge of the previous channel outputs y_1, y_2, \dots, y_k provides sufficient information to choose the next channel input x_{k+1} . Even though EEG classification errors cause each output y_k to be a noisy reflection of the input x_k , it is reasonable to assume that the graphical display can perfectly convey this output to the user—for example, simply by showing a left- or right-facing arrow. In practice, the information provided by y_1, y_2, \dots, y_k is difficult for the user to interpret, so we show a candidate sequence \underline{z} instead (e.g., a candidate string of text or a candidate curve) that reflects the BCI's current belief about the user's intent \underline{z} . In either case, the feedback—in contrast to the classification of EEG signals—can be assumed to be causal and noiseless.

2.3. Solution Approach

For both BCIs of interest in this article, we have shown that it is reasonable to communicate intent before taking action, to model intent as a discrete random process, and to model the EEG sensor and the graphical display as a binary symmetric channel with noiseless feedback. The problem of interface design has therefore been recast as the problem of constructing an optimal communication protocol (Figure 2). We can think of this protocol as having two parts, a source code and a channel code, which jointly describe how the user should select motor imagery (i.e., inputs X_k) in response to feedback from the display (i.e., some reflection of the outputs Y_k) in order to most efficiently convey their underlying intent (i.e., the sequence \underline{Z}).

Source code. Although we have seen that \underline{Z} can represent both a string of text and a two-dimensional curve in the same way, it might have very different statistical properties in each case. The purpose of a source code is to provide a universal representation of \underline{Z} with a common statistical model. In particular, our source code puts the random process \underline{Z} , distributed according to $P_{\underline{Z}}(\underline{z})$, in one-to-one correspondence with a single, continuous random variable W that is

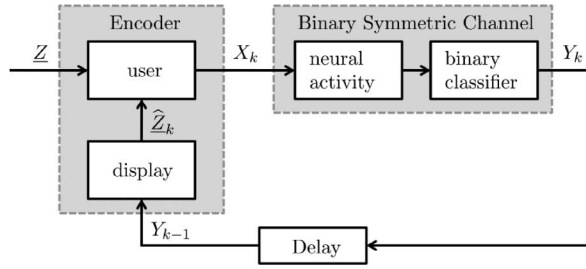


FIGURE 2 Feedback information-theoretic paradigm for designing brain-machine interfaces.

Note. It models the user's intent as a sequence \underline{Z} , the motor imagery in EEG as a binary symmetric channel with input X_k and output Y_k , and the graphical display as a noiseless feedback providing a candidate sequence to the user.

uniformly distributed on the interval $[0, 1]$. It is characterized by the invertible mapping $W = \phi(\underline{Z})$.

We construct ϕ using a method of lossless data compression called arithmetic coding (Witten, Neal, & Cleary, 1987). This method is both optimal (i.e., it guarantees that W is indeed uniformly distributed on $[0,1]$) and has the useful property that it preserves ordering, so that $\underline{z} < \underline{z}'$ if and only if $w = \phi(\underline{z}) < w' = \phi(\underline{z}')$. An arithmetic code maps each length- i sequence z^i to a subinterval $[l_{z^i}, r_{z^i}) \subset [0, 1]$ that is computed recursively from the one $[l_{z^{i-1}}, r_{z^{i-1}}) \supset [l_{z^i}, r_{z^i})$ associated with its prefix. Given that the alphabet Σ is of length m , then $[l_{z^{i-1}}, r_{z^{i-1}})$ is divided into m subintervals arranged in the same order as Σ and with size proportional to $P_{Z_i|Z^{i-1}}(z_i|z^{i-1})$. By a similar recursion we may construct the inverse ϕ^{-1} that maps W to its representation as a sequence \underline{Z} .

The source code described by ϕ allows us henceforth to ignore differences in the structure and statistics of \underline{Z} , making it much easier to derive an optimal channel code. User intent is given equivalently by W , which always has the same statistical distribution (uniform) and the same domain (the interval $[0, 1]$). This distinction, however, is completely transparent to the user. Because ϕ preserves ordering and because ϕ^{-1} allows us to reconstruct \underline{Z} for the purposes of display, our source code does not in any way change how the user interacts with the BCI.

Channel code. The purpose of a channel code is to specify the sequence of channel inputs X_k that the user should generate to ensure a vanishing probability of error in the BCI's reconstruction of W based on the channel outputs Y_k . A rate R is achievable if the resulting estimate \hat{W}_k computed by the BCI after k inputs satisfies

$$\mathbb{P}(|W - \hat{W}_k| > 2^{-kR}) \rightarrow 0. \quad (4)$$

A channel code is optimal if any rate $R < C$ is achievable, where C is the information capacity of the channel.

In the absence of feedback, forward error correction (i.e., adding a special pattern of redundant inputs) is a standard approach to deriving an optimal channel code (Cover & Thomas, 2006). However, this type of channel code is very hard for a human user to implement. In the presence of feedback—that is, assuming the user knows the previous channel outputs Y^k when choosing X_{k+1} —it is possible to reduce the complexity of coding and to increase the rate at which the error probability decreases, even though C remains constant. Classical examples of feedback coding include Horstein (1963) and Schalkwijk and Kailath (1966). Recently, these approaches have been unified as the *posterior matching* scheme, which provides both an optimal channel code and a simple recursive framework for both analysis and implementation (Coleman, 2009; Shayevitz & Feder, in press). Because the posterior matching scheme requires only a comparison between W and the current estimate \hat{W}_k to generate the next input X_{k+1} , it is easy to implement by a human user and is thus an ideal channel code for BCI.

Posterior matching specifies the input at time $k + 1$ according to

$$X_{k+1} = F_X^{-1}(F_{W|Y^k}(W|Y^k)),$$

where $F_X(x)$ is the optimal input CDF for the channel (known a priori) and $F_{W|Y^k}(w|y^k)$ is the posterior conditional CDF over intent (computed by the BCI based on the channel outputs—note that \hat{W}_k is derived from this CDF). The essence of this scheme is that X_{k+1} is statistically independent of all previous Y^k . Hence, there exists no more informative channel input for reducing ambiguity in W .

For a binary symmetric channel—our model of EEG-based BCI in this article—the posterior matching scheme has a simple interpretation. Assume that after time step k , the BCI computes $F_{W|Y^k}(w|y^k)$ according to Bayes’s rule and selects the median of the posterior as the estimate \hat{W}_k . This estimate is presented as feedback to the user, who selects the next input according to the simple comparison

$$X_{k+1} = \begin{cases} 1 & \text{if } W \geq \hat{W}_k, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Remarkably, this simple comparison is not only optimal but also easy for a human user to implement. Recall that we can reconstruct an estimate of \underline{Z} using the inverse mapping

$$\hat{\underline{Z}} = \phi^{-1}(\hat{W}_k).$$

Further recall both that sequences \underline{Z} admit an intuitive lexicographic ordering and that ϕ^{-1} preserves this ordering. If the graphical display shows $\hat{\underline{Z}}$, manifested either as a string of text or as a smooth planar curve, then it is easy for the user to decide if $\underline{Z} \geq \hat{\underline{Z}}$ (i.e., if $W \geq \hat{W}_k$). If so, then the user thinks “right” (i.e., $x_{k+1} = 1$); if not, then the user thinks “left” (i.e., $x_{k+1} = 0$).

Figure 1c shows an example. Initially, the posterior is uniform and the estimate \hat{w}_0 corresponds to the straight path given by $\hat{\underline{z}}$. Because $\underline{z} > \hat{\underline{z}}$, the user thinks

“right” and $x_1 = 1$. After observing a channel output of $y_1 = 1$, the BCI recomputes the posterior—increasing the probability of all points to the right of \hat{w}_0 and decreasing the probability of all points to the left of \hat{w}_0 —and generates a new estimate \hat{w}_1 . In this case, $\hat{z}' = \phi^{-1}(\hat{w}_1)$ moves to the right of z , so the user thinks “left” and $x_2 = 0$. After observing a channel output of $y_2 = 0$, the BCI again recomputes the posterior and the process repeats. It is important to emphasize again that, during this entire process of communication, the user’s intent \underline{Z} is assumed to be constant and known a priori to the user (but not to the BCI). In contrast, the input sequence \underline{X} is a response to feedback and may change depending on the (noisy) output sequence \underline{Y} .

3. METHODS

This study has been approved by the Institutional Review Board of the University of Illinois, Urbana-Champaign. In total nine healthy, able-bodied subjects (two female) participated in this study. All subjects were right-handed, were between the ages of 20 to 30, and had normal or corrected-to-normal vision. Eight of the subjects had limited exposure to EEG motor imagery, ranging from 6 to 8 hr, as part of another study. Subject 500 had several more hours of experience from previous studies. In the current study, each subject participated in three to four experimental sessions.

3.1. Experimental Protocol

Figure 3 shows the stages of an experimental session. Each session consists of an optional practice phase followed by a training phase and then a series of tasks during the testing phase. In the practice phase, subjects were given time to practice using a keyboard to familiarize themselves with the task. In the training phase,

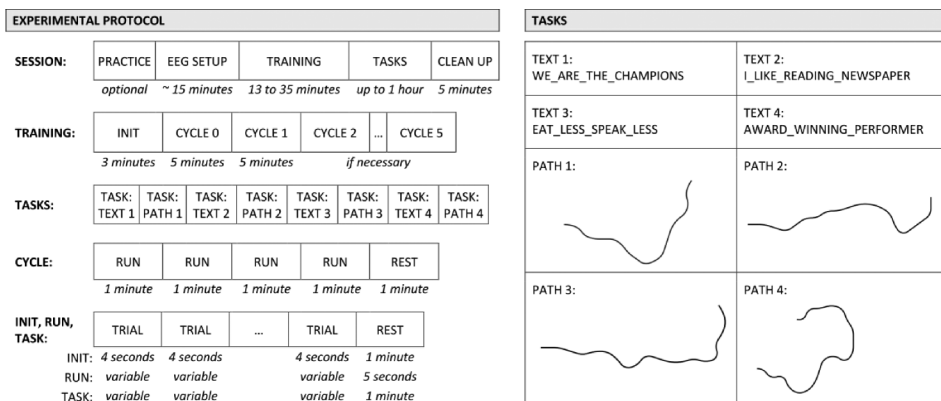


FIGURE 3 The schematic showing the stages of an experimental session and the target sequences for each task.

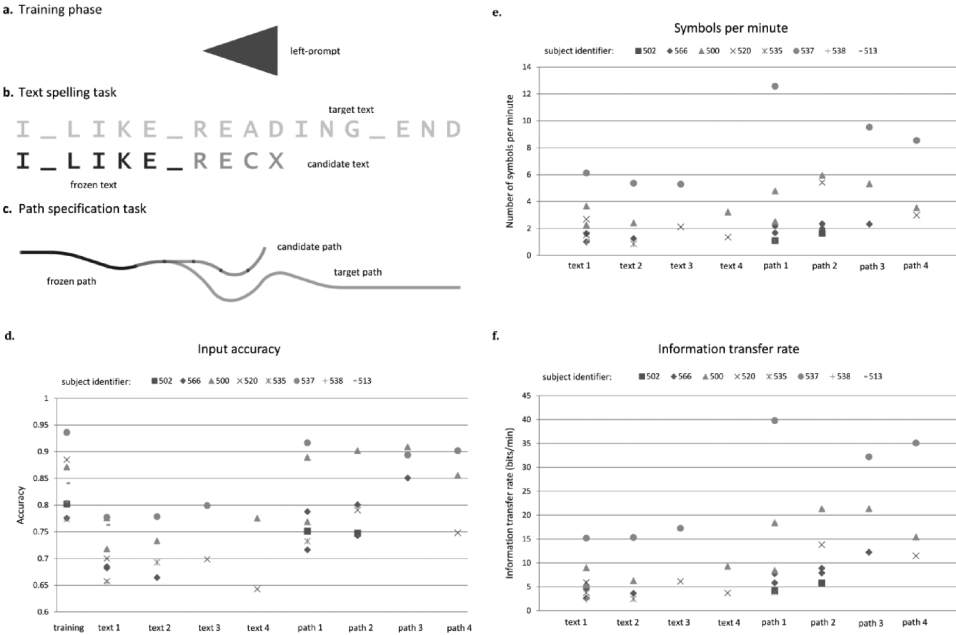


FIGURE 4 Experimental interface and performance results. *Note.* (a), (b), and (c) show screenshots of the interface taken during the training phase, the text spelling task, and the path specification task, respectively. (d), (e), and (f) show the performance obtained in successful tasks in terms of the input classification accuracy, the mean number of symbols specified per minute, and the information transfer rate, respectively.

the motor imagery classifier is trained. The subject was instructed to imagine movement of their left or right hand in response to directional prompts from the computer display (Figure 4a). After the classifier is initialized, audio feedback is provided following each presentation to indicate correct or incorrect classification. This phase lasts until the classifier has been trained to obtain an accuracy of at least 75%. In the testing phase, the subject is presented with two kinds of tasks. In text spelling tasks, the subject is asked to spell a target sentence, and in path specification tasks, the subject is asked to trace a target path. In each session, the subject is presented with at most four text spelling tasks and four path specification tasks, in a fixed alternating order. We allow the subject to leave a session early if the training phase or the execution of tasks cannot be completed within 15 min.

In addition to the target sequence, the subject is presented with a candidate sequence. The subject’s instructions are simple: If the target sequence is ordered to the left of the candidate sequence, the subject should provide a “left” command. Otherwise, the subject should provide a “right” command. Upon observation of the command, the interface updates its posterior estimate and produces a new candidate sequence and displays it to the subject.

There are two practical issues that might affect the behavior of the interface and the subject. First, the interface freezes a prefix of the candidate sequence permanently if its posterior probability is greater than a threshold T_{freeze} . Consequentially,

if the interface freezes a prefix that does not correspond to the target, the subject cannot modify it to correct his or her mistake. When this occurs, we say that the task has failed and the interface moves on to the next task. A successful task is one in which the entire target sequence has been frozen. Certainly, we can define success and failure in a task only when the target sequence is known to the interface.

The second issue is that the true candidate sequence can be computed to arbitrary length, but the displayed candidate sequence must be truncated. The displayed candidate sequence is chosen to be the longest prefix of the true candidate sequence with a posterior probability greater than a threshold, T_{decode} . Consequentially, if the displayed candidate sequence completely overlaps with, but is shorter than, the target, the subject has no basis for providing commands. If such a situation occurs, the subjects are instructed to provide a “right” command.

Text spelling task. The interface uses the alphabet $\Sigma = \{A, B, C, \dots, Z, _\}$, where ‘_’ represents a space and is ordered to the right of ‘Z’. The statistical language model was obtained from a corpus using the prediction by partial matching algorithm implemented in Begleiter et al. (2004). The training corpus was the one developed by Dasher (Ward & MacKay, 2002). The target sentences do not appear in the corpus, although each word appears at least once.

Figure 4b shows an intermediate state of the interface while the subject is spelling the target sentence “I_LIKE_READING”. The display appends the sequence “_END” to help subjects when candidate sequence overlaps with the target, but its posterior probability is not high enough to freeze the sequence. Just below the target sentence shown in gray, the interface displays the candidate sentence “I LIKE RECX” with the frozen prefix in black and the remaining in red. In this state, the subject performs left motor imagery since the target sentence comes before the candidate sentence in lexicographic ordering.

Path specification task. During a path specification task, the subject traces a smooth curve using an alphabet composed of 11 symbols, $\Sigma = \{\sigma_1, \sigma_2, \dots, \sigma_{11}\}$, where σ_i corresponds to a circular arc as shown in Figure 1a. The statistical language model was a fixed zeroth-order Markov model given by a discrete Gaussian kernel centered on the symbol σ_6 , corresponding to the notion that a straight arc has the highest probability.

Figure 4c shows an intermediate state of the interface while the subject is tracing the target path shown in gray. The candidate path is superposed on the target path, with the frozen prefix shown in black and the remaining in red. In this state, the subject performs right motor imagery because the target path is ordered to the right of the candidate path.

3.2. EEG Signal Processing

EEG signals were collected by an Electro-Cap through eight channels (F3, F4, C3, C4, T7, T8, P3, P4; Nunez & Srinivasan, 2006). Fpz was used as ground, and Cz was used as reference. Voltages were amplified by a low noise amplifier

(James Long Co., Caroga Lake, NY), low-pass filtered to 100Hz (antialiasing) and synchronously sampled at a 400 Hz sample rate by an Iotech Personal Daq 3000 A/D array. The exact mu-rhythm frequency varies among individuals, so the subject-specific mu-rhythm peak was estimated from the average power spectrum.

The feature extraction procedure, Common Spatial Analytic Patterns (McCormick, Ma, & Coleman, 2010) uses labeled training data to perform blind source separation and recover these source signals for a specific subject. The log-magnitude of each recovered source signal sample is modeled as a specific weighted sum of recently evolving binary-valued motor imagery. The classification step is based on a Hidden Markov Model of the binary-valued motor imagery (McCormick et al., 2010). The probability of the category of motor imagery is estimated and updated regularly for each new EEG sample. When the probability of left or right motor imagery exceeds a certain threshold T_{class} , a new classification is generated and passed to the BCI applications. This method significantly improves the information transfer rate (ITR) for BCI applications—from the highest rate of 37.5 bits/min previously documented (Blankertz, Dornhege, Krauledat, Müller, & Curio, 2007) to our highest rate of 60.9 bits/min.

3.3. Artifacts

We also simultaneously recorded electrooculography (EOG) and arm electromyography (EMG) on the subjects to statistically verify that eye or muscle movement signals are not correlated with the control signals. We used methods analogous to Vaughan, Miner, McFarland, and Wolpaw (1998) to discard data sets that had significant EOG or EMG artifacts. We computed the biserial correlation coefficients between the classifier outputs and the desired commands as well as that between the EOG/EMG and the desired commands. The biserial correlation coefficient is defined as

$$r = \frac{\sqrt{N^+ \cdot N^-}}{N^+ + N^-} \cdot \frac{\text{mean}[S^+] - \text{mean}[S^-]}{\text{std}[S^+ \cup S^-]},$$

where S^+ and S^- are the signal samples (EEG, EOG, or EMG) associated with left- or right-hand motor imagery, and N^+ and N^- are the number of samples of the corresponding classes (Blankertz et al., 2007). We have verified that the EOG and EMG artifacts were not significantly correlated with the classifier outputs. For example, to demonstrate the order of magnitude difference in r^2 for artifacts versus desired commands, subject 500 had r^2 values as follows: 0.0037 for left eye EOG, 0.0051 for right eye EOG, 0.0096 for left arm EMG, and 0.0062 for right arm EMG; in comparison, the r^2 value for the EEG classifier outputs was 0.2510.

3.4. Parameter Selection

- T_{class} , the threshold for classification decision, was empirically set to 0.99.
- $T_{accuracy}$, the minimum accuracy required to pass the training phase, was set to 0.75.

- ε , the crossover probability, was computed from the accuracy observed in the last training cycle individually for each experiment. We used a slightly larger ε for text spelling because of the additional cognitive load it requires as reported by our subjects.
- T_{decode} , the parameter that determines the length of the candidate sequence, was experimentally chosen to be 0.01 so that it is usually four to six symbols long.
- T_{freeze} , the parameter that determines when to freeze a prefix, was set to 0.995 to avoid frequent task failures.

3.5. Performance Measures

Averaged accuracy in a task \hat{P} . Suppose the subject conveyed $n_0 + n_1$ motor imageries in total during a task, n_1 of which were classified as the correct, desired control commands at the corresponding step, the accuracy is defined as the proportion of the correct control commands $\hat{P} = \frac{n_1}{n_0 + n_1}$.

ITR \hat{R} (bits/minute). The ITR is defined as the amount of information conveyed reliably per minute—as defined by the capacity of the associated binary symmetric channel with crossover probability given by our empirical accuracy \hat{P} :

$$\hat{R} = \frac{1 - H_{BSC}(\hat{P})}{d}$$

where average trial length is $d = T / (n_0 + n_1)$ and $H_{BSC}(\hat{P}) = -\hat{P} \log(\hat{P}) - (1 - \hat{P}) \log(1 - \hat{P})$.

Rate of spelling \hat{q} (symbols/minute). Suppose the subject spelled out a sequence of symbols of length N using in total T minutes in a task, the rate of spelling is defined as the number of symbols spelled per minute $\hat{q} = \frac{N}{T}$.

4. EXPERIMENTAL RESULTS

Figure 4d presents each subject's input classification accuracy during the last training cycle and during successfully completed tasks. The average classification accuracy was 0.71 during text spelling tasks, and 0.82 during path specification. The best classification accuracy was 0.80 during text spelling tasks, and 0.92 during path specification—both for subject 537.

Figure 4e shows system performance during successful tasks, as measured by the number of symbols specified per minute. The highest spelling rate was six characters per minute (char./min.), and the highest rate of tracing a planar curve

was 12.5 segments per minute (seg./min.), both achieved by subject 537 during the first task of each type.

Figure 4f presents the ITR in successful tasks. The average ITR was 6.7 bits/minute during text spelling and 15.2 bits/minute during path specification. The best ITR was 17.2 bits/minute during text spelling and 39.8 bits/minute during path specification—both performed by subject 537.

5. DISCUSSION

5.1. Comparison of System Performance

Another recently demonstrated BCI speller utilizes the Dasher interface, which also incorporates insights from information theory, to achieve a maximum rate of seven characters per minute (Felton, Lewis, Wills, Radwin, & Williams, 2007). Several fundamental distinctions hinder direct comparison with such a system. First, we were able to utilize a simpler discrete, binary control signal, as opposed to a continuous control signal modulated by motor imagery. Second, Felton et al. (2007) used a training text corpus consisting of only 200 short phrases. Target sentences were generated by choosing from among these phrases. Consequently, the log likelihood of the test sentences was much higher than those used in this study, which will result in higher average system performance. By employing a much larger corpus and not testing with complete sentences drawn directly from the training text, we better emulate usage in a free-spelling task. Despite these differences, our performance was comparable. This can be explained in part by our use of statistically sound methods for both source and channel coding to design our behavioral and feedback protocols. We expect future results from our lab and others to take this fundamental observation into consideration to obtain significantly higher (i.e., orders of magnitude) spelling rates in systems constrained by such noisy, low bandwidth signals.

5.2. Performance Differences in Text Spelling Versus Path Specification

According to the results, it was generally easier for the subjects to trace smooth planar curves than to spell text. The relative ease of path-tracing is unlikely to be caused by the difference in the likelihood of sequences, since the negative log-likelihoods were comparable in both cases: They ranged from 43 to 58 for sentences and from 57 to 73 for paths. As can be seen, the paths were actually less likely to be encountered given the statistical models used in our study. A plausible explanation for this is that the subjects can better perform direct lexicographic ordering on paths—based on spatial perception—as compared to alphabetization. The latter perhaps imposed heavier cognitive load on the subjects, requiring more reaction time to choose desired control commands. This implies that in the future researchers may want to develop more intuitive visual interfaces to reduce the difficulty in lexicographical ordering of texts.

5.3. Error Correction

Although feedback information theory for binary-symmetric channels with noiseless feedback guarantees exponential convergence onto the desired sequence, the experiments still showed that our system is susceptible to errors. The reasons for this are as follows: First, the assumption of a binary-symmetric channel may not hold strictly in practice. The system may have bias for one of the motor imageries. Second, the cross-over probability is hard to estimate accurately, and it may vary over time with the mental status of the subjects or changes in the sensor.

Due to these issues, the system can sometimes assign high probabilities to sequences that do not correspond to the target sequence. Although a backtracking mechanism may be useful for such situations, it is difficult to implement. The reason for this is that the representation of the probability distribution (a continuous function) requires arbitrary precision. To proceed in a computationally tractable way, we have to freeze the spelled sequence at some threshold, which means information about the history is intentionally discarded. Moreover, a system that uses “finished” sequences to drive a device, such as a voice synthesizer or motorized wheelchair, may be unable to reasonably backtrack. As a workaround, we chose a stringent threshold for freezing a prefix. As a future direction, detecting error-related potentials immediately after updating the frozen prefix may allow us to correct many such errors. An error-related potential could be incorporated probabilistically within our framework, for example, by decreasing the posterior for the prefix when such a signal is detected.

5.4. Future Directions

Learning and neurofeedback training. A learning effect is observed across different experiments on the same subject. Certain subjects (e.g., subjects 535 and 566) were not able to start spelling in the first couple of experiments. In addition, the daily condition of the subjects might have played an important role because the experiment requires a high level of concentration over a relatively long period and consumes significant mental energy. Better strategies for training subjects are needed to enhance the subjects’ fluency in generating motor imagery, such as the neurofeedback training suggested by Hwang, Kwon, and Im (2009).

Formulation in terms of decentralized control. It can be shown (Coleman, 2009) that the feedback communication problem is a simple instantiation of the more general class of sequential, nonclassical, team decision problems (Chu, 1971; Ho & Chu, 1971; Witsenhausen, 1971). We plan to address problems within this more general context—it will be critically important to do this, for example, to address BCIs where the external device has constrained dynamics. As these problems are notoriously difficult to solve (Papadimitriou & Tsitsiklis, 1986), we plan to carefully construct meaningful yet tractable problems that have closed-form optimal solutions and whose optimal policies are easy for the user to behaviorally implement.

6. CONCLUSION

This article has demonstrated an information-theoretic approach to design BCIs that explicitly takes feedback into consideration. This approach is particularly important for BCIs with inherently noisy, low bandwidth neural sensors (e.g., noninvasive EEG-based systems). What is remarkable about this approach is that, subject to statistical assumptions, it is both optimal from a communications perspective and easily implementable by the user. We have demonstrated this approach in two representative interfaces, one for text entry and the other for tracing smooth planar curves. Experimental results with nine human subjects, using a binary motor imagery EEG-based paradigm, showed promising results in performance as measured both by the information transfer rate and by the symbols per second conveyed. Perhaps more important than these initial results, we consider the merit of our work as pushing the broader community toward formal consideration of feedback protocols in BCIs. In particular, we hope that this underlying philosophy—making design more systematic and theoretically well grounded using tools from stochastic optimal control and information theory—are adopted and generalized.

REFERENCES

- Akce, A., Johnson, M., & Bretl, T. (2010). Remote teleoperation of an unmanned aircraft with a brain-machine interface: Theory and preliminary results. In *ICRA'09 Proceedings of the 2009 IEEE international conference on Robotics and Automation*. pp. 5322–5327.
- Begleiter, R., El-Yaniv, R., & Yona, G. (2004). On prediction using variable order Markov models. *Journal of Artificial Intelligence Research*, 22, 385–421.
- Bell, C. J., Shenoy, P., Chalodhorn, R., & Rao, R. P. N. (2008). Control of a humanoid robot by a noninvasive brain-computer interface in humans. *Journal of Neural Engineering*, 5, 214–220.
- Blankertz, B., Dornhege, G., Krauledat, M., Müller, K., & Curio, G. (2007). The non-invasive Berlin brain-computer interface: Fast acquisition of effective performance in untrained subjects. *NeuroImage*, 37, 539–550.
- Brockwell, A. E., Rojas, A. L., & Kass, R. E. (2004). Recursive Bayesian decoding of motor cortical signals by particle filtering. *Journal of Neurophysiology*, 91, 1899–1907.
- Carmena, J. M., Lebedev, M. A., Crist, R. E., O'Doherty, J. E., Santucci, D. M., Dimitrov, D. F., et al. (2003). Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS Biol*, 1 (2), e42.
- Chatterjee, A., Aggarwal, V., Ramos, A., Acharya, S., & Thakor, N. (2007). A brain-computer interface with vibrotactile biofeedback for haptic information. *Journal of NeuroEngineering and Rehabilitation*, 4 (1), 40.
- Chu, K. C. (1971, December). Team decision theory and information structures in optimal control problems: Part II. *1971 IEEE Conference on Decision and Control* (Vol. 10), 388–392.
- Cleary, J., & Witten, I. (1984, April). Data compression using adaptive coding and partial string matching. *IEEE Transactions on Communications*, 32, 396–402.
- Coleman, T. P. (2009, July). A stochastic control approach to 'posterior matching'-style feedback communication schemes. *IEEE International Symposium on Information Theory*, 1520–1524.

- Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). New York, NY: Wiley-Interscience.
- Dhillon, G. S., & Horch, K. W. (2005, December). Direct neural sensory feedback and control of a prosthetic arm. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *13*, 468–472.
- Eden, U. T., Frank, L. M., Barbieri, R., Solo, V., & Brown, E. N. (2004). Dynamic analysis of neural encoding by point process adaptive filtering. *Neural Computation*, *16*, 971–998.
- Fabiani, G. E., McFarland, D. J., Wolpaw, J. R., & Pfurtscheller, G. (2004, September). Conversion of EEG activity into cursor movement by a brain–computer interface (BCI). *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *12*, 331–338.
- Felton, E., Lewis, N. L., Wills, S. A., Radwin, R. G., & Williams, J. C. (2007, May). Neural signal based control of the Dasher writing system. *IEEE EMBS Conference on Neural Engineering*, 366–370.
- Grimmett, G., & Stirzaker, D. (2001). *Probability and random processes* (3rd ed.). New York, NY: Oxford University Press.
- Hatsopoulos, N. G., & Donoghue, J. P. (2009). The science of neural interface systems. *Annual Review of Neuroscience*, *32*(1), 249–266.
- Ho, Y.-C., & Chu, K. C. (1971, December). Team decision theory and information structures in optimal control problems: Part I. *1971 IEEE Conference on Decision and Control* (Vol. 10), 383–387.
- Horstein, M. (1963, July). Sequential transmission using noiseless feedback. *IEEE Transaction on Information Theory*, *9*, 136–143.
- Hwang, H., Kwon, K., & Im, C. (2009). Neurofeedback-based motor imagery training for brain–computer interface (BCI). *Journal of Neuroscience Methods*, *179*(1), 150–156.
- Inmann, A., & Haugland, M. (2004, 7). Implementation of natural sensory feedback in a portable control system for a hand grasp neuroprosthesis. *Medical Engineering & Physics*, *26*, 449–458.
- Iturrate, I., Antelis, J., Kubler, A., & Minguetz, J. (2009, June). A noninvasive brain-actuated wheelchair based on a P300 neurophysiological protocol and automated navigation. *IEEE Transactions on Robotics*, *25*, 614–627.
- Kemere, C., Shenoy, K. V., & Meng, T. H. (2004). Model-based neural decoding of reaching movements: A maximum likelihood approach. *IEEE Transactions on Biomedical Engineering*, *51*, 925–932.
- Kuiken, T. A., Li, G., Lock, B. A., Lipschutz, R. D., Miller, L. A., Stubblefield, K. A., & Englehart, K. B. (2009). Targeted muscle reinnervation for real-time myoelectric control of multifunction artificial arms. *Journal of the American Medical Association*, *301*, 619–628.
- Leuthardt, E. C., Miller, K. J., Schalk, G., Rao, R. P. N., & Ojemann, J. G. (2006, June). Electro-corticography-based brain computer interface—The Seattle experience. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *14*, 194–198.
- McCormick, M., Ma, R., & Coleman, T. P. (2010, March). An analytic spatial filter and a hidden Markov model for enhanced information transfer rate in EEG-based brain computer interfaces. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 602–605.
- Moffat, A. (1990, November). Implementing the PPM data compression scheme. *IEEE Transactions on Communications*, *38*, 1917–1921.
- Nunez, P., & Srinivasan, R. (2006). *Electric fields of the brain: The neurophysics of EEG*. New York, NY: Oxford University Press.
- Omar, C., Johnson, M., Bretl, T., & Coleman, T. P. (2008a). Policies for neural prosthetic control: Initial experiments with a text interface. *American Control Conference*, 4161–4166.
- Omar, C., Johnson, M., Bretl, T., & Coleman, T. (2008b). Querying the user properly for high-performance brain-machine interfaces: Recursive estimation, control, and feedback

- information-theoretic perspectives. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 5216–5219.
- Papadimitriou, C. H., & Tsitsiklis, J. (1986). Intractable problems in control theory. *SIAM Journal on Control and Optimization*, 24, 639–654.
- Schalkwijk, J., & Kailath, T. (1966). A coding scheme for additive noise channels with feedback–I: No bandwidth constraint. *IEEE Transactions on Information Theory*, 12, 172–182.
- Schwartz, A. (2004). Cortical neural prosthetics. *Annual Review of Neuroscience*, 27 (1), 487–507.
- Sellers, E. W., Kubler, A., & Donchin, E. (2006, June). Brain–computer interface research at the University of South Florida Cognitive Psychophysiology Laboratory: The P300 speller. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14, 221–224.
- Serruya, M., Hatsopoulos, N., Paninski, L., Fellows, M., & Donoghue, J. (2002). Brain–machine interface: Instant neural control of a movement signal. *Nature*, 416 (6877), 141–142.
- Shayevitz, O., & Feder, M. (in press). Optimal feedback communication via posterior matching. *IEEE Transactions on Information Theory*.
- Sinkjaer, T., Haugland, M., Inmann, A., Hansen, M., & Nielsen, K. D. (2003). Biopotentials as command and feedback signals in functional electrical stimulation systems. *Medical Engineering & Physics*, 25(1), 29–40.
- Srinivasan, L., & Brown, E. N. (2007). A state-space framework for movement control to dynamic goals through brain-driven interfaces. *IEEE Transactions on Biomedical Engineering*, 54, 526–535.
- Srinivasan, L., Eden, U. T., Willsky, A. S., & Brown, E. N. (2006). A state-space analysis for reconstruction of goal-directed movements using neural signals. *Neural Computation*, 18, 2465–2494.
- Vaughan, T., Miner, L., McFarland, D., & Wolpaw, J. (1998). EEG-based communication: analysis of concurrent EMG activity. *Electroencephalography and clinical Neurophysiology*, 107, 428–433.
- Ward, D. J., & MacKay, D. J. C. (2002). Fast hands-free writing by gaze direction. *Nature*, 418 (6900), 838–838.
- Witsenhausen, H. S. (1971). On information structures, feedback and causality. *SIAM Journal on Control*, 9, 149–160.
- Witten, I. H., Neal, R. M., & Cleary, J. G. (1987). Arithmetic coding for data compression. *Communications of the ACM*, 30, 520–540.
- Wolpaw, J. R. (2007). Brain–computer interfaces as new brain output pathways. *The Journal of Physiology*, 579, 613–619.
- Wu, W., Gao, Y., Bienenstock, E., Donoghue, J. P., & Black, M. J. (2006). Bayesian population decoding of motor cortical activity using a Kalman filter. *Neural Computation*, 18 (1), 80–118.

This article was downloaded by: [Ma, Rui]

On: 29 December 2010

Access details: Access Details: [subscription number 931650915]

Publisher Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



International Journal of Human-Computer Interaction

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t775653655>

A Feedback Information-Theoretic Approach to the Design of Brain-Computer Interfaces

Cyrus Omar^a; Abdullah Akce^a; Miles Johnson^a; Timothy Bretl^a; Rui Ma^a; Edward Maclin^a; Martin McCormick^a; Todd P. Coleman^a

^a University of Illinois, Urbana-Champaign

Online publication date: 28 December 2010

To cite this Article Omar, Cyrus , Akce, Abdullah , Johnson, Miles , Bretl, Timothy , Ma, Rui , Maclin, Edward , McCormick, Martin and Coleman, Todd P.(2011) 'A Feedback Information-Theoretic Approach to the Design of Brain-Computer Interfaces', International Journal of Human-Computer Interaction, 27: 1, 5 – 23

To link to this Article: DOI: 10.1080/10447318.2011.535749

URL: <http://dx.doi.org/10.1080/10447318.2011.535749>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.