Optimal Modality Selection Using Information Transfer Rate for Event Related Potential Driven Brain Computer Interfaces

Aziz Koçanaoğulları
Northeastern University
akocanaogullari@ece.neu.edu

Murat Akçakaya
University of Pittsburgh
akcakaya@pitt.edu

Barry Oken
Oregon Health & Science University
oken@ohsu.edu

Deniz Erdoğmuş
Northeastern University
erdogmus@ece.neu.edu

ABSTRACT

Individuals with reduced speech and motor ability due to injury or motor neuron diseases have difficulties in communication. Such events at the terminal stage lead to the loss of all muscular activity which is referred as locked-in state. In this condition, communication can only be performed with electroencephalogram (EEG) signals. Brain computer interfaces (BCI) typing systems provide people without muscular control a communication baseline. In BCI typing systems, user is presented stimuli and corresponding EEG evidence is used to detect the user intent among a pre-defined alphabet. Due to low signal-to-noise ratio (SNR) of EEG evidence, multiple stimuli sequences are required. Thus, to limit the time spent on typing the stimuli presented to the user should be optimized. BCI typing systems are designed to operate including different modalities where modalities surpass each other either in SNR of the respective signal or time spent to acquire the response. Therefore, it is a fundamental problem when to choose cheap in time - weak in response questions and when to choose expensive in time - strong in response questions. In this study we propose a modality selection mechanism for systems that rely on recursive evidence collections under Gaussian evidence model assumption. Specifically, we focus on BCI typing systems that operate with error related potentials (ERPs) and feedback related potentials (FRPs). We analytically derive a decision threshold to select each of these modalities. We also demonstrate the performance of the proposed method using a BCI typing system.

CCS CONCEPTS

• **Human-centered computing → HCI theory, concepts and models**; • **Theory of computation → Active learning**; *Bayesian analysis.*

KEYWORDS

modality selection, query selection, stimuli selection, active inference, brain-computer interface typing system, event related potential, feedback related potential

ACM Reference Format:

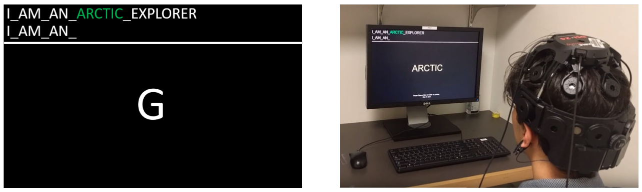
Aziz Koçanaoğulları, Murat Akçakaya, Barry Oken, and Deniz Erdoğmuş. 2020. Optimal Modality Selection Using Information Transfer Rate for Event Related Potential Driven Brain Computer Interfaces. In *PETRA '20: The Pervasive Technologies Related to Assistive Environments Conference June 30–July 03, 2020, Corfu, Greece.* ACM, New York, NY, USA, 7 pages. https://doi.org/XX.XXXX/XXXXXXX.XXXXXXX

PETRA '20, June 30–July 03, 2020, Corfu, Greece

ACM ISBN XXX-X-XXXX-XXXX-X/XX/XX…$15.00
https://doi.org/XX.XXXX/XXXXXXX.XXXXXXX

1 INTRODUCTION

Individuals with reduced speech and motor ability due to injury or motor neuron diseases, including cerebral palsy, multiple sclerosis (MS), amyotrophic lateral sclerosis (ALS), spinal muscular atrophy, locked-in syndrome, spinal cord injury, stroke, and traumatic brain injury, currently have assistive technology solutions to increase the quality of the communication. Unfortunately, after the loss of muscular activity resulting in locked-in state, communication can only be performed with EEG signals. Brain computer interfaces (BCIs) have shown promising capacity to mitigate the dependency on muscular activity, providing people with disabilities a communication baseline. Such systems typically rely on detection of event related potentials (ERPs), in which the user is expected to generate a signature change that allows the system to infer intent from the spatio-temporal time series collected [1, 18] as a result of stimuli presentation to the user. In this work we mainly focus on a gaze-independent BCI typing system called RSVPKeyboard as presented in Figure 1 [11]. The stimuli are flashed in a rapid fashion on a fixed pre-defined location as shown in Figure 1-(a). Different operation modalities and a detailed description of the system is presented in Appendix section (Section 6). In the figure the subject in Figure 1-(b) is performing a copy phrase task, in which he is tasked to complete the phrase 'I am an arctic explorer' as presented in Figure 1-(a). Implementation of RSVPKeyboard is accessible from https://github.com/BciPy/BciPy.



(a) (b)

Figure 1: EEG driven Rapid Serial Visual Presentation (RSVP) keyboard typing interface. (a) The stimuli is flashed in the middle of the screen while the user is informed with the text above. (b) The user is conducting copying the phrase task. The user is informed about the required phrase. EEG is collected on top of the scalp non-invasively.

Even being the most common evidence used in BCI typing systems, ERP is not the only possible modality. In addition to commonly used ERPs such as p300 [3], feedback related potentials (FRPs) have shown promising results in BCI community in typing systems to elicit respective responses [4]. FRP is often related with cognitive response to errors and hence also named as error related potentials (Errps) in the literature. In Chavariagga's work [2], FRP usage in non-invasive BCIs is discussed especially discussing high SNR values across users. Moreover, Gonzalez's work [6] discusses using two different response mechanisms at the same time to improve classification performance. In the presence of different modalities, where each stimuli selection method has a respective response time (computation time is negligible) and response characteristics (response signal to noise ratio (SNR), evidence distribution separability etc.), there exists a source selection problem to achieve better performance. In such cases, ideally the system actively alternates between modalities deciding either cheap but weak evidence collection or expensive but strong evidence collection. This decision is built on top of the posterior distribution evolution and observed evidences. Just to give a toy example; p300 driven ERP systems stimulate 6 letters approximately in a second compared to FRP systems span 700ms interval to stimulate a single letter. As the user is responding to a single letter instead of multiple rapid flashes, the signal quality and classification assessment for FRP is higher compared to ERP. Therefore, it is required to optimize an alternation algorithm between these two paradigms in estimation for faster user intent detection. Thus in our work we focus on designing a stimuli selection method that alternates between different modalities.

Query, sensor selection for brain computer interfacing is required to increase the speed of communication. It is argued in Moghammadfalahi's work [12] that showing the entire alphabet as a stimuli is not practical and the selection is optimized with maximizing expected posterior. In Mansiah's work [10] mutual information is used to select stimuli. In our previous findings [9] a new speed term is introduced to speed up the estimation process. All these methods show promising results. However, these methods stagger from not incorporating time required for stimuli selection. This ultimately leads to assumption of unit stimuli-response time which is not realistic in presence of different observation methods. Additionally, it is required to find relative EEG feature model performances of different methods for better assesment. Different distribution effect

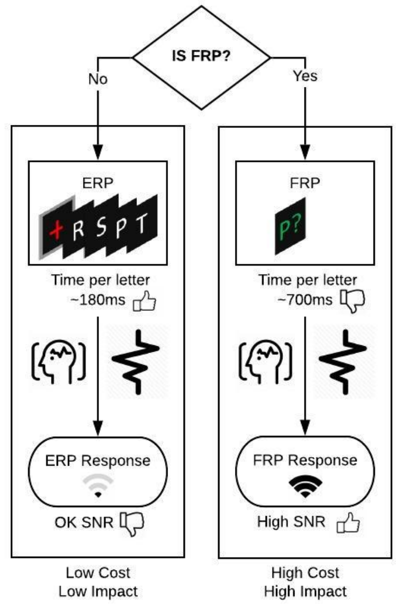


Figure 2: In this work we aim to find the optimal decision criteria (shown as *IS FRP?* in the figure) that allows us to optimally alternate between two different stimuli that result in two different performance characteristics. In BCI typing systems event related potentials (ERP) are cheaper in time to stimuli compared to feedback related potentials (FRP) however the response EEG SNR is much lower.

on BCI systems is also investigated from a distribution shift perspective. In Satti's work [17], a confusion matrix is learned over the trials to assess the performance of the EEG quality. In Raza's work [16], authors propose to update their model statistics over time based on recently observed data. Even though these models consider the evidence from single modality changes, they do not consider selecting different modalities with different performances. In this work we focus on active selection across modalities and stimuli in a BCI paradigm to optimize information gain per unit time.

Specifically, we assume the existence of two evidence sources; (1) cheap-less informative **ERP-evidence**, (2) expensive-more informative **FRP-evidence**. Both evidence sources are assumed to yield features that are randomly distributed in 1D. Also we assume each evidence modality employs respective negative and positive evidence distributions for positive and negative response to the stimuli by the subject respectively. In this paper we propose an algorithm to optimally alternate between these two query-evidence sources with a threshold. This decision for the BCI typing systyem is visualized in Figure 2. We make the following contributions: (i) We propose an objective that maximizes the information gain per unit time. (ii) We define a threshold on the recent posterior probability distribution over the alphabet. Once such threshold is passed, system queries to expensive but crisp evidence source. (iii) We evaluate the time-accuracy gain with such approach in a BCI typing system.

2 PRELIMINARIES

Let target letter denoted by σ being an element of a finite set (namely the alphabet) denoted by $\mathcal{A}$. Each step in the sequential decision making process is named *sequences* indexed by $s \in \mathcal{S}$. The system decides on a subset of queries with cardinality $N \in \mathcal{N}$ $\Phi_s \triangleq \{\phi_s^1, \ldots, \phi_s^N\}$ at the beginning of each sequence, where $\phi_s^i \in \mathcal{A}$ denotes a single letter flash. After querying, corresponding EEG evidence $\varepsilon_s \triangleq \{\varepsilon_s^1, \ldots, \varepsilon_s^N\}$ is observed. As the system proceeds through sequences, the decision and query selection will rely on previously asked queries and observed evidence. To preserve neat notation, we present $\mathcal{H}_s \triangleq \{\varepsilon_{1:s}, \Phi_{1:s}, \mathcal{H}_0\}$ to represent the task history and $\mathcal{H}_0$ to denote the information from the language model. Conventional query selection optimization for query selection in BCIs is centered about mutual information maximization, where the objective at sequence s as; $\arg\max_{\Phi \in Q^N} I(\sigma, \varepsilon_s | \Phi, \mathcal{H}_{s-1})$. Here $I(.,.)$ denotes the mutual information function between two arguments. In our previous work [8], mutual information objective and other surrogate objectives including posterior maximization and uncertainty reduction are shown to result in the same query type; presenting most likely letters.

However, as explained before, in the presence of different information channels it is plausible to maximize information rate. Otherwise the system, independent of time requirements selects the evidence source with highest separation between positive and negative response distributions. Therefore in this paper we focus on the following objective for query selection;

$$\Phi_s = \arg\max_{\Phi \in Q^N} \frac{1}{T_\Phi} I(\sigma, \varepsilon_s | \Phi, \mathcal{H}_{s-1}) \tag{1}$$

3 METHOD

As presented in the previous section, the aim of querying is to reduce the uncertainty during Bayesian classification task as fast as possible. As presented in the literature, the following equality holds;

$$\arg\max_{\Phi \in Q^N} \frac{1}{T_\Phi} I(\sigma, \varepsilon_s | \Phi, \mathcal{H}_{s-1}) = \arg\max_{\Phi \in Q^N} \frac{-1}{T_\Phi} H(\sigma | \varepsilon_s, \Phi, \mathcal{H}_{s-1})$$

Therefore one optimizes the objective presented in (1) by minimizing conditional entropy. In BCI typing systems, there exist two response channels; positive and negative responses. If shown, the subject reacts to the target of intent from the positive evidence channel (ERP-FRP) and the subject reacts to other non-target letters from the negative evidence channel (no ERP-no FRP). These responses indicate responses to target letter stimuli and non target stimuli respectively. Let us represent the positive and negative responses by $\ell = 1$ and $\ell = 0$ respectively defined as $\ell = \delta_\sigma(\phi)$. Here $\delta_.(.)$ represents the delta function and is equal to 1 if $\sigma \in \phi$. Hence, the label of a letter becomes 1 if shown in the stimuli. We first tailor the objective presented in (1) using the binary response mechanism as the following;

$$\begin{aligned} H(\sigma | \varepsilon_s, \Phi, \mathcal{H}_{s-1}) &= E_{\sigma|\mathcal{H}_{s-1}} E_{\varepsilon_s|\sigma,\Phi} \left(\log \frac{p(\varepsilon_s | \sigma, \Phi)}{p(\varepsilon_s | \Phi)} \right) \\ &= E_{\sigma|\mathcal{H}_{s-1}} E_{\varepsilon_s|\sigma,\Phi} \left(\log \frac{p(\varepsilon_s | \sigma, \Phi)}{\sum_{v \sim \sigma} p(\varepsilon_s | \Phi, v) p(v)} \right) \end{aligned} \tag{2}$$

Here v in the denominator is used a marginalization variable over the entire alphabet. We further make use of $p(\varepsilon_s | \phi, \sigma) = \sum_\ell p(\varepsilon_s | \ell) \delta_\sigma(\phi)$ introducing the label information into the equation yields the following;

$$E_{\sigma|\mathcal{H}_{s-1}} E_{\varepsilon_s|\sigma,\Phi} \left(\log \frac{p(\varepsilon_s | \ell = 1) \delta_\sigma(\Phi) + p(\varepsilon_s | \ell = 0)(1 - \delta_\sigma(\Phi))}{\sum_{v \sim \sigma} p(\varepsilon_s | \Phi, v) p(v)} \right)$$

In this equation, dividing and multiplying the term inside the logarithm with $p(\varepsilon_s | \ell = 0)$ yields us an equation with likelihood ratios as the following;

$$E_{\sigma|\mathcal{H}_{s-1}} E_{\varepsilon_s|\sigma,\Phi} \left(\log \frac{\frac{p(\varepsilon_s|1)}{p(\varepsilon_s|0)} \delta_\sigma(\Phi) + (1 - \delta_\sigma(\Phi))}{\sum_{v \sim \sigma} \frac{p(\varepsilon_s|1)}{p(\varepsilon_s|0)} \delta_v(\Phi) p(v) + (1 - \delta_v(\Phi)) p(v)} \right) \tag{3}$$

Inserting (3) into (2) and furthermore using in (1) yields us the following objective;

$$\begin{aligned} \Phi_{s+1} = \arg\min_{\Phi \in Q^N} \frac{1}{T_\Phi} E_{\sigma|\mathcal{H}_{s-1}} E_{\varepsilon_s|\sigma,\Phi} \Bigg(\\ \log \frac{\frac{p(\varepsilon_s|1)}{p(\varepsilon_s|0)} \delta_\sigma(\Phi) + (1 - \delta_\sigma(\Phi))}{\sum_{v \sim \sigma} \frac{p(\varepsilon_s|1)}{p(\varepsilon_s|0)} \delta_v(\Phi) p(v)} \Bigg) \end{aligned} \tag{4}$$

However, a general analytical solution to the term in (4) is not possible and hence one needs reasonable simplifications. In p300 and FRP driven BCI typing systems, it is widely assumed the relevant features extracted from the signals employ a Gaussian distribution. In inference, the system evaluates the likelihoods of new features using the distributions learned in the training time and proceed. In this work to approximate (4), we calculate mean and variance values for the likelihood ratio assuming evidence is distributed randomly for both positive and negative case and additionally, we make use of Teh's work [19] to approximate the conditional entropy.

3.1 Approximating conditional entropy

It is impractical to calculate the expected value of the log term in (4). However, the order of expectation and log is not interchangeable due to properties shown in Jensen's work [7]. However, Teh in the work [19] presents an approximation to the expected value of the log. Let x be a random variable the approximation is as the following;

$$E(\log(x)) \approx \log(E(x)) - (E(x^2) - E^2(x))/2E^2(x)$$

Equality holds when variance of the distribution is negligible to the mean-squared. From a practical standpoint, $E(\log(x)) \approx \log(E(x))$ eases the computation dramatically and usually preferred. In our work we employ this simplification following [19]. Please refer to Appendix (Section6.1) for the statistics of the likelihood ratios that allows us to obtain an approximation to (4).

3.2 Algorithm for Query Selection

In this section, we present a query selection algorithm starting a general solution and narrowing it down to the BCI query selection level. We employ the following score representation;

$$\text{sc} = E_{\varepsilon|1}(p(\varepsilon|1)/p(\varepsilon|0))$$

Algorithm 1 Query Selection Using ERP-FRP for BCI

learn $\mathcal{N}(\mathcal{M}_{0,\text{ERP}}, \sigma^2_{0,\text{ERP}}), \mathcal{N}(\mathcal{M}_{1,\text{ERP}}, \sigma^2_{1,\text{ERP}})$
learn $\mathcal{N}(\mathcal{M}_{0,\text{FRP}}, \sigma^2_{0,\text{FRP}}), \mathcal{N}(\mathcal{M}_{1,\text{FRP}}, \sigma^2_{1,\text{FRP}})$
initialize $T_{\Phi_{\text{ERP}}} \in \mathbb{R}^+, T_{\Phi_{\text{FRP}}} \in \mathbb{R}^+, N \in \mathbb{N}^+$
$\text{sc}_{\text{ERP}} \leftarrow E_{\varepsilon|_{\ell=1}}(p_{\text{ERP}}(\varepsilon|1)/p_{\text{ERP}}(\varepsilon|0))$ ▷ (9)
$\text{sc}_{\text{FRP}} \leftarrow E_{\varepsilon|_{\ell=1}}(p_{\text{FRP}}(\varepsilon|1)/p_{\text{FRP}}(\varepsilon|0))$ ▷ (9)
$\hat{c} \leftarrow$ s.t. $\text{Obj}_{\text{ERP}}(\hat{c}, \text{sc}_{\text{ERP}}) = \text{Obj}_{\text{FRP}}(\hat{c}, \text{sc}_{\text{FRP}})$ ▷ (7),(8)
$s \leftarrow 0$
while iterations **do**
  $p(\sigma|\mathcal{H}_s)$ gets updated
  $c \leftarrow \max_\sigma p(\sigma|\mathcal{H}_s)$
  if $c > \hat{c}$ **then**
    FRP Query Selection
    $\Phi_{s+1} \leftarrow \arg\max_\sigma p(\sigma|\mathcal{H}_s)$
  else
    ERP Query Selection
    for $i \in \{1, \cdots, N\}$ **do**
      $\Phi_{s+1} \leftarrow \Phi_{s+1} \cup \arg\max_{\sigma \notin \Phi_{s+1}} p(\sigma|\mathcal{H}_s)$
    end for
  end if
  $s \leftarrow s + 1$
end while

The value of the score function (expected value of the likelihood ratio under target distribution) assuming 1D-Gaussian distributions is provided in the Appendix (Section6.1). Following the score definition we rewrite the objective in (4) as the following;

$$\begin{aligned}\Phi_{s+1} = \arg\min_{\Phi \in Q^N} \frac{1}{T_\Phi} E_{\sigma|\mathcal{H}_{s-1}} \Bigg[&\log(\text{sc}\delta_\sigma(\Phi) + (1 - \delta_\sigma(\Phi)) \\ &- \log\left(\sum_v \text{sc}\ p(v)\delta_v(\Phi) + p(v)(1 - \delta_v(\Phi))\right)\Bigg]\end{aligned} \tag{5}$$

(5) is easily formed by computing the function value for each letter in the alphabet and solved greedily. This greedy solution yields a close to optimum solution as presented in Nemhauser's work [13]. The solution to objective (5) is selecting the most likely N letters from the alphabet [8].

Observe that in (5), $E_{\sigma|\mathcal{H}_{s-1}} \log(\text{sc}\delta_\sigma(\Phi) + (1 - \delta_\sigma(\Phi)) = \sum_{\sigma\in\Phi}(p(\sigma|\mathcal{H}_{s-1})) \log \text{sc} = \mathcal{P}_\Phi \log \text{sc}$. Applying this simplification and a similar algebraic manipulation in the denominator, it is possible to obtain the following simplification;

$$\Phi_{s+1} = \arg\min_{\Phi \in Q^N} \frac{1}{T_\Phi}(\mathcal{P}_\Phi \log \text{sc} - \log(\mathcal{P}_\Phi \text{sc} + (1 - \mathcal{P}_\Phi))) \tag{6}$$

Without loss of generality assume $A \in \mathcal{A}$ being the most likely letter (i.e. $p(A) = c = \max_i p(i)$). Let $\Phi_{\text{ERP}}, \Phi_{\text{FRP}}$ denote queries for ERP and FRP respectively and $N_{\text{FRP}} = 1$ denote FRP has only 1 letter as the query. From our previous discussion $A \in \Phi_{\text{ERP}}$ and $\Phi_{\text{FRP}} = \{A\}$. It is shown in our previous work [8] that the objective in (5) is monotonically decreasing wrt. $\mathcal{P}_\Phi$ which implies as the query set size $|\Phi|$ increases the objective decreases. Therefore the objective in (5) for ERP ($N_{\text{ERP}} > 1$) becomes the following;

$$\begin{aligned}&\text{Obj}_{\text{ERP}}(c, \text{sc}_{\text{ERP}}) = \\ &\frac{1}{T_{\Phi_{\text{ERP}}}}(c \log \text{sc}_{\text{ERP}} - \log(c\text{sc}_{\text{ERP}} + (1 - c))) \\ &\quad > \frac{1}{T_{\Phi_{\text{ERP}}}}(\mathcal{P}_{\Phi_{\text{ERP}}} \log \text{sc}_{\text{ERP}} - \log(\mathcal{P}_{\Phi_{\text{ERP}}}\text{sc}_{\text{ERP}} + (1 - \mathcal{P}_{\Phi_{\text{ERP}}})))\end{aligned} \tag{7}$$

Similar simplification is also doable for the FRP as the following;

$$\text{Obj}_{\text{FRP}}(c, \text{sc}_{\text{FRP}}) = \frac{1}{T_{\Phi_{\text{FRP}}}}(c \log \text{sc}_{\text{FRP}} - \log(c\text{sc}_{\text{FRP}} + (1 - c))) \tag{8}$$

Hence ERP querying is preferred if $\text{Obj}_{\text{FRP}} > \text{Obj}_{\text{ERP}}$. It is also possible to find numerically a threshold number $\hat{c}$ s.t. $\text{Obj}_{\text{ERP}}(\hat{c}) = \text{Obj}_{\text{FRP}}(\hat{c})$. Hence the BCI systems continues with FRP if $\max_\sigma p(\sigma|\mathcal{H}_{s-1}) > \hat{c}$. We visualize the algorithm in Algorithm1.

3.3 Algorithm Explanation

In this section we present a verbal explanation of the algorithm from a practical standpoint with the intent to simplify the approach and ease its usage.

Specifically for the BCI typing system, ERP requires multiple consecutive rapid stimuli to be elicited whereas FRP response appears to a single stimuli flash $N = 1$. Let us denote the queries selected for both paradigs as $\Phi_{s,\text{ERP}}$ and $\Phi_{s,\text{FRP}}$ respectively. Therefore decision between ERP and FRP relies only on the comparison of the objective in (5), whichever yields the minimum is selected. Ideally, showing multiple stimuli is expected to decrease uncertainty more than a single stimuli. However, in practice, FRP response of the user is more distinguishable compared to ERP yielding $\text{sc}_{\text{FRP}} >> \text{sc}_{\text{ERP}}$ moreover the time required for FRP is less than time required to elicit ERP $T_{\Phi_{\text{FRP}}} < T_{\Phi_{\text{ERP}}}$. Therefore in certain cases FRP reduces the uncertainty more than ERP.

The algorithm requires 4 Gaussian models for ERP positive-negative and FRP positive negative respectively. These distributions are required to calculate the necessary approximations. A threshold $\hat{c}$ is computed using objectives presented in (7) and (8). $\hat{c}$ satisfies the following; if the maximum value in the current posterior is $\hat{c}$, objective for ERP and FRP are equal. Hence, for a posterior probability with an element higher probability value than $\hat{c}$ FRP is preferred. ERP selects most likely N candidates for stimuli, whereas FRP selects the most likely letter only.

4 EXPERIMENTS

We use EEG data recorded when human users performed typing tasks using RSVP Keyboard, which is a noninvasive EEG-based brain computer interface for typing. The typing interface focuses on rapid serial visual presentation as the stimuli paradigm, however also supports other modalities including FRP, row-column flash matrix, single letter flash matrix etc. The system operates on the alphabet, which is called the state space $\mathcal{A}$ that includes English alphabet and additionally a space symbol and a backspace symbol (for deletion). Therefore $\mathcal{A} = \{A, B, C, ..., Z, \_, <\}$, where _ and < represent space and backspace, respectively. In this paper we make use of the implementation and the simulation environment named

BciPy in https://github.com/BciPy/BciPy [11]. This implementation is an updated version of Orhan's work [14].

Modules

Display: System employs a display module that controls the paradigm of visual stimuli presentation. In RSVP paradigm, a set of pseudo-randomly ordered stimuli are presented on a pre-fixed location of the screen in a rapid serial manner. Each stimulus is a trial. A set of trials which has been presented with no time gap in between, is called a sequence. Every ERP sequence flashes unique letters and hence enforced to contain at most only one single target stimulus. Time-series analysis is used to identify the stimulus on which the attention (target) is placed on. Every FRP sequence include only one element in the query set and flash the selected letter on the pre-defined location for a longer duration to obtain the user response.

Feature Extraction: The system collects EEG evidence. Due to low SNR values, it is mandatory to process and extract some features for signal reasoning. The system includes a drift removal filtering (frequencies «1Hz) and a bandpass filtering to remove artifact-related high frequency components. After filtering, EEG is decomposed into trials to obtain data for respective stimuli flash. Time-windowed data from different EEG channels is concatenated to obtain the EEG feature vector that has high dimensionality due to using all the channels. The dimensionality is reduced applying a channel wise principal component analysis (PCA) and EEG features are extracted from the reduced dimensional feature vectors using regularized discriminant analysis [5]. These features for both positive and negative responses are assumed to be randomly distributed.

Language Model (LM): Provides the prior distribution that is used for computing the posterior distribution of a symbol using an Online-Context Language Model (OCLM) that provides prior distributions given EEG evidence as part of our BCI system.

Operation Modes

Calibration mode: In the calibration mode, users are asked to pay attention to pre-defined target symbols within randomly-ordered sequences to collect labeled EEG data. This collection is specified for both ERP and FRP to learn class-conditional EEG evidence distributions and feature extraction pipeline parameters.

Copy phrase task mode: In this mode, users are presented with a set of pre-defined phrases. Each phrase includes a missing word and the users are asked to complete the phrase by typing the missing word. This mode is designed to assess the system and the typing performance in terms of speed and accuracy in the presence of a language model.

Free spell mode: The typing mode that allows the subject to type freely without any pre-defined task text.

Specifically, calibration data of 12 healthy participants (collected with Northeastern University IRB-130107) is used for our simulation results. During calibration, participants are presented with 100 sequences of symbols followed with a feedback symbol. A sequence contains randomly ordered ten symbols with a pre-defined target symbol and a randomly selected feedback symbol. EEG is acquired

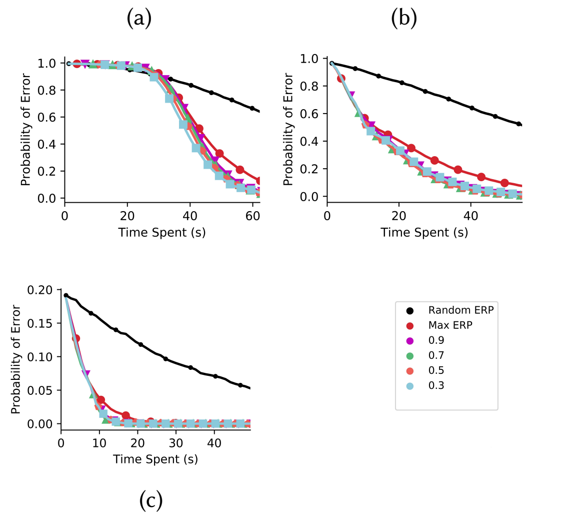


Figure 3: Probability of error vs time plots. Probability of error is calculated as $1-p(\sigma)$ where σ is the target letter. Observe that probability of error decreases faster if FRP query is used. *Random ERP* and *Max ERP* represent ERP presentation of randomly selected letters and top candidates respectively. FRP threshold $\hat{c}$ and respective error probability lines are color and marker coded. (a) represents a scenario where language model is adversarial (the target letter is one of the least likely). (b) represents where there exists no language model (uniform prior). (c) represents a scenario where the language model is supportive (target letter is one of the most likely).

from 16 channels using the International 10–20 configuration (Fp1, Fp2,F3, F4, Fz, Fc1, Fc2, Cz, P1, P2,C1, C2, Cp3, Cp4, P5, P6). At each sequence, respective evidence of consecutive rapid stimuli is treated as ERP evidence, whereas respective response to the feedback symbol is treated as FRP evidence. These signals are used to learn class conditional EEG evidence feature distributions.

Class conditional EEG evidence distributions are further used to simulate a copy letter task (copy phrase task where the task is a single letter) to assess the performance of the method with a simulated user (using the so called learned evidence distributions for ERP and FRP). During simulations when a target/nontarget letter is presented to the user in ERP/FRP paradigm, EEG features from corresponding class conditional distributions are sampled. These samples are used to estimate the user intent. For more information about our simulation framework, we refer the reader to Orhan's work [15]. In our experiments we categorize users based on their calibration performances. The performance is measured with area under receiver operating characteristics curve of the features samples of the training data both for ERP and FRP.

Our first set of experiments is to visualize how fast probability of error in letter estimation decreases with different values of $\hat{c}$ with an average performance user ($AUC_{ERP} = .75$, $AUC_{FRP} = .87$). In language model assisted BCI typing systems, effect of the language model is crucial. Statistically, since languages are structured, language models speed up the estimation process. However, in rare cases the user tries to type a statistically uncommon word (wrt. the training set of the language model) in which typing a letter task requires the user to overcome the information of language model. In our experiments we simulate an environment with uniform prior, supportive language model (where the target letter is one of the

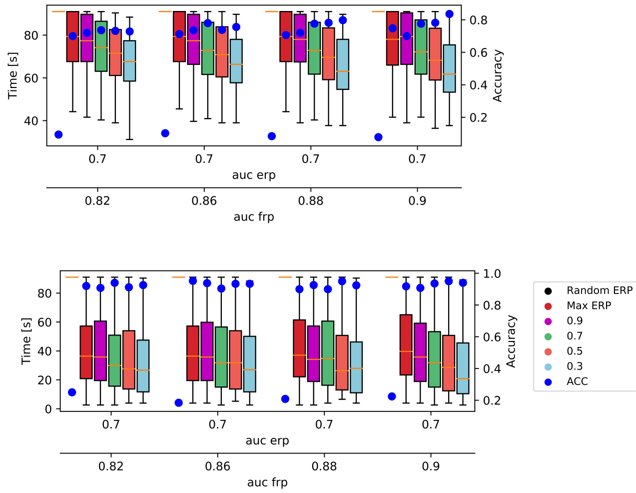


Figure 4: Effects of changing FRP model performance with a fixed ERP model performance on decision. Top figure represents from copy letter simulations where the language model is adversarial, bottom figure represents results from copy letter simulations without language model. Each bar plot and accuracy is a result of 1000 Monte Carlo simulations. The methods are color coded (from left to right on each block in the figure); Random, ERP only and $\hat{c} = [0.9, 0.7, 0.5, 0.3]$ It can be observed that FRP presence decreases time required till decision without sacrificing accuracy.

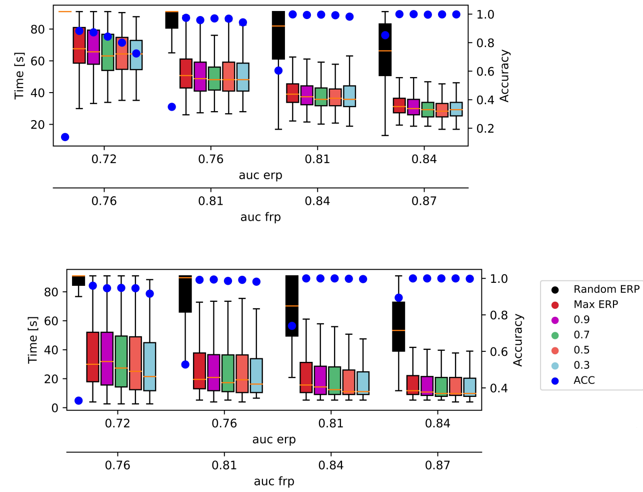


Figure 5: Effects of changing FRP model performance with a fixed ERP model performance on decision. Top figure represents from copy letter simulations where the language model is adversarial, bottom figure represents results from copy letter simulations without language model. Here we are changing ERP model performances and marginally better FRP model performance. Given these performance values and respective distributions, our algorithm results $\hat{c} > 0.9$ that results in ERP dominating the stimulus selection.

likely letters with respect to the language model) and adversarial language model (where the target letter is one of the least likely). We plot the change in probability of error over time as visualized in Figure 3 where probability of error represents ($1 - p(\sigma|\mathcal{H}_s)$ where σ is the target letter). We pick Random ERP querying as a universal baseline and we pick max ERP as the method baseline which selects the most N likely candidates and proceeds with ERP querying. Here in the figure different colors represent different values of thresholds $\hat{c}$s. As it it designed for, the system proceeds with FRP if there exists a letter with higher probability than $\hat{c}$ otherwise continues with ERP. It is observed that FRP in this scenario benefits the system significantly.

In our second set of experiments we investigate the performance effect of ERP and FRP models. We synthetically generate a set of performance intervals using real data from an average performing user. First we keep the ERP model performance constant and increase FRP model performance gradually. To report the accuracy and speed values we select a decision threshold of 95% confidence level (once a candidate letter achieves 0.95 posterior value), once received the system terminates. We visualize the results in Figure 4. Observe that as FRP performance increases, it is more beneficial to focus on FRP querying. This is due FRP response resulting higher impact evidence that allows faster inference. Additionally, comparing the uniform prior (no language model) and adversarial language model case, FRP querying is more impactful. This is due to overcoming the prior belief. In uniform case, the system selects many of the non-target letters as stimuli that causes a delay on FRP stimulus selection. Whereas in adversarial case, to be eligible being the query in FRP paradigm, the letter needs to be already up and coming noting it being the target letter. We also investigate the effect of ERP when there exists a marginal difference between ERP and FRP (by definition, we expect FRP to have higher performance, otherwise ERP is always preferable to FRP). We visualize our findings in Figure 5. It is observed that when there's a marginal difference between ERP and FRP performances the system does not gain a significant speed boost using FRP. Moreover, the optimal $\hat{c}$ for the scenarios presented in Figure 5 is close to 1, suggesting to continue with ERP.

Without such knowledge, it is possible to rule out FRP stimulus selection due to no significant improvement considering only a case in Figure 5, however it is apparent from Figure 4 FRP is beneficial and provides a speed up to the letter selection inference.

5 CONCLUSION

In this work we presented an optimal selection between two different sources; (1) cheap in time - low in inference impact evidence source, (2) expensive in time - high in inference impact evidence source. Specifically, we tuned our derivations under Gaussian distribution for evidence distributions conditioned on state and query. We derived an algorithm to optimally alternate between these two modalities that achieves faster inference without sacrificing accuracy. Moreover, we validated our findings with simulations that are based on human-in-the-loop experiment data.

As a feature work, we plan to investigate the sensitivity of the threshold parameter that adjusts selection between ERP and FRP to evidence model variance. Moreover, we would like to investigate the effects of distribution shift on different evidence modalities (specially for our work ERP and FRP). This will allow us to employ the fatigue effects on different evidence modalities and enable us to design an active update paradigm for the modality selection threshold $\hat{c}$.

ACKNOWLEDGMENTS

This work is supported by NSF (IIS-1149570, CNS-1544895, IIS-1715858, IIS-1717654, IIS-1844885, IIS-1915083), DHHS (90RE5017-02-01), and NIH (R01DC009834).

REFERENCES

[1] Rachel Atchley, Roger Ellingson, Daniel Klee, Tabatha Memmott, and Barry Oken. 2017. A cognitive stressor for event-related potential studies: the Portland arithmetic stress task. *Stress* 20, 3 (2017), 277–284.

[2] Ricardo Chavarriaga and José del R Millán. 2010. Learning from EEG error-related potentials in noninvasive brain-computer interfaces. *IEEE transactions on neural systems and rehabilitation engineering* 18, 4 (2010), 381–388.

[3] Reza Fazel-Rezai, Brendan Z Allison, Christoph Guger, Eric W Sellers, Sonja C Kleih, and Andrea Kübler. 2012. P300 brain computer interface: current challenges and emerging trends. *Frontiers in neuroengineering* 5 (2012), 14.

[4] Pierre W Ferrez and José del R Millán. 2008. Error-related EEG potentials generated during simulated brain–computer interaction. *IEEE transactions on biomedical engineering* 55, 3 (2008), 923–929.

[5] Jerome H Friedman. 1989. Regularized discriminant analysis. *Journal of the American statistical association* 84, 405 (1989), 165–175.

[6] Paula Gonzalez-Navarro. 2018. *Signal Models for Fusion of Multiple Brain Responses and Context Evidence Sources in EEG-Based BCIS.* Ph.D. Dissertation. Northeastern University.

[7] Johan Ludwig William Valdemar Jensen et al. 1906. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta mathematica* 30 (1906), 175–193.

[8] Aziz Koçanaoğulları, Deniz Erdoğmuş, and Murat Akçakaya. 2018. On analysis of active querying for recursive state estimation. *IEEE Signal Processing Letters* 25, 6 (2018), 743–747.

[9] Aziz Koçanaoğulları, Yeganeh M Marghi, Murat Akçakaya, and Deniz Erdoğmuş. 2018. Optimal query selection using multi-armed bandits. *IEEE Signal Processing Letters* 25, 12 (2018), 1870–1874.

[10] Boyla Mainsah, Dmitry Kalika, Leslie Collins, Siuyuan Liu, and Chandra Throckmorton. 2018. Information-based adaptive stimulus selection to optimize communication efficiency in brain-computer interfaces. In *Advances in Neural Information Processing Systems.* 4820–4830.

[11] Tab Memmott, Aziz Koçanaoğulları, Matthew Lawhead, Daniel Klee, Shiran Dudy, Melanie Fried-Oken, and Barry Oken. 2020. BciPy: Brain-Computer Interface Software in Python. *arXiv preprint arXiv:2002.06642* (2020).

[12] Mohammad Moghadamfalahi, Murat Akcakaya, Hooman Nezamfar, Jamshid Sourati, and Deniz Erdogmus. 2017. An active RBSE framework to generate optimal stimulus sequences in a BCI for spelling. *IEEE Transactions on Signal Processing* 65, 20 (2017), 5381–5392.

[13] George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher. 1978. An analysis of approximations for maximizing submodular set functions—I. *Mathematical programming* 14, 1 (1978), 265–294.

[14] Umut Orhan, Kenneth E Hild, Deniz Erdogmus, Brian Roark, Barry Oken, and Melanie Fried-Oken. 2012. RSVP keyboard: An EEG based typing interface. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* IEEE, 645–648.

[15] Umut Orhan, Hooman Nezamfar, Murat Akcakaya, Deniz Erdogmus, Matt Higger, Mohammad Moghadamfalahi, Andrew Fowler, Brian Roark, Barry Oken, and Melanie Fried-Oken. 2016. Probabilistic simulation framework for eeg-based bci design. *Brain-Computer Interfaces* 3, 4 (2016), 171–185.

[16] Haider Raza, Dheeraj Rathee, Shang-Ming Zhou, Hubert Cecotti, and Girijesh Prasad. 2019. Covariate shift estimation based adaptive ensemble learning for handling non-stationarity in motor imagery related EEG-based brain-computer interface. *Neurocomputing* 343 (2019), 154–166.

[17] Abdul Satti, Cuntai Guan, Damien Coyle, and Girijesh Prasad. 2010. A covariate shift minimisation method to alleviate non-stationarity effects for an adaptive brain-computer interface. In *2010 20th International Conference on Pattern Recognition.* IEEE, 105–108.

[18] Shravani Sur and VK Sinha. 2009. Event-related potential: An overview. *Industrial psychiatry journal* 18, 1 (2009), 70.

[19] Yee W Teh, David Newman, and Max Welling. 2007. A collapsed variational Bayesian inference algorithm for latent Dirichlet allocation. In *Advances in neural information processing systems.* 1353–1360.

6 APPENDIX

6.1 Statistics of likelihood ratio

Conditional entropy calculates the expected resulting entropy if the query set Φ was selected. Resulting distribution relies on the evidence distributions, specifically the likelihood ratio as presented in (4). Given two one dimensional Gaussian distributions $\varepsilon|_{\ell=1} \sim \mathcal{N}(\mathcal{M}_1, {\sigma_1}^2)$ and $\varepsilon|_{\ell=0} \sim \mathcal{N}(\mathcal{M}_0, {\sigma_0}^2)$, (4) takes the expected value of the likelihood ratio $p(\varepsilon|1)/p(\varepsilon|0)$ wrt. $\varepsilon|\sigma, \Phi$. Therefore we need to take two conditions into consideration; taking the expected value where $\varepsilon|_{\ell=1}$ and $\varepsilon|_{\ell=0}$. We derive the mean likelihood ratio for both cases using a little bit of algebra as the following;

$$\begin{aligned} E_{\varepsilon|_{\ell=0}} \frac{p(\varepsilon|1)}{p(\varepsilon|0)} &= \int_{-\infty}^{\infty} p(\varepsilon|1) d\varepsilon = 1 \\ E_{\varepsilon|_{\ell=1}} \frac{p(\varepsilon|1)}{p(\varepsilon|0)} &= \int_{-\infty}^{\infty} \frac{p^2(\varepsilon|1)}{p(\varepsilon|0)} d\varepsilon = \frac{\sigma_0^{1/2}}{2\pi^{1/2}\sigma_1} e^{c_1} e^{{b_1}^2/4a_1} \frac{\pi^{1/2}}{{a_1}^{1/2}} \end{aligned} \tag{9}$$

Here in (9), let sgn(.) denote the signum function, the parameters are the following;

$$\begin{aligned} a_1 &= k(1/{\sigma_1}^2 - 1/(2{\sigma_0}^2)), k = \operatorname{sgn}(1/{\sigma_1}^2 - 1/(2{\sigma_0}^2)) \\ b_1 &= k(2\mathcal{M}_1/{\sigma_1}^2 - \mathcal{M}_0/{\sigma_0}^2) \\ c_1 &= (-\mathcal{M}_1^2/{\sigma_1}^2 + \mathcal{M}_0^2/(2{\sigma_0}^2)) \end{aligned}$$

With a similar approach, we calculate the variances of the likelihood ratio distributions. To calculate the variance we require the second moment information of the ratio. With a little algebraic manipulation, second moments of the likelihood ratios are calculated as the following;

$$\begin{aligned} E_{\varepsilon|_{\ell=0}} \left(\frac{p^2(\varepsilon|1)}{p^2(\varepsilon|0)} \right) &= \int_{-\infty}^{\infty} \frac{p^2(\varepsilon|1)}{p(\varepsilon|0)} \delta\varepsilon = E_{\varepsilon|_{\ell=1}} \frac{p(\varepsilon|1)}{p(\varepsilon|0)} \\ E_{\varepsilon|_{\ell=1}} \left(\frac{p^2(\varepsilon|1)}{p^2(\varepsilon|0)} \right) &= \int_{-\infty}^{\infty} \frac{p^3(\varepsilon|1)}{p^2(\varepsilon|0)} \delta\varepsilon = \frac{\sigma_0}{2\pi^{1/2}\sigma_1^{3/2}} e^{c_2} e^{{b_2}^2/4a_2} \frac{\pi^{1/2}}{{a_2}^{1/2}} \end{aligned} \tag{10}$$

Here in (10), let sgn denote the signum function, the parameters are the following;

$$\begin{aligned} a_2 &= k(3/2{\sigma_1}^2 - 1/{\sigma_0}^2), k = \operatorname{sgn}(3/2{\sigma_1}^2 - 1/{\sigma_0}^2) \\ b_2 &= k(3\mathcal{M}_1/{\sigma_1}^2 - 2\mathcal{M}_0/{\sigma_0}^2) \\ c_2 &= (-3\mathcal{M}_1^2/2{\sigma_1}^2 + \mathcal{M}_0^2/{\sigma_0}^2) \end{aligned}$$