Sliding HDCA: Single-Trial EEG Classification to Overcome and Quantify Temporal Variability

Amar R. Marathe, Anthony J. Ries, and Kaleb McDowell, *Senior Member, IEEE*

***Abstract*—Patterns of neural data obtained from electroencephalography (EEG) can be classified by machine learning techniques to increase human-system performance. In controlled laboratory settings this classification approach works well; however, transitioning these approaches into more dynamic, unconstrained environments will present several significant challenges. One such challenge is an increase in temporal variability in measured behavioral and neural responses, which often results in suboptimal classification performance. Previously, we reported a novel classification method designed to account for temporal variability in the neural response in order to improve classification performance by using sliding windows in hierarchical discriminant component analysis (HDCA), and demonstrated a decrease in classification error by over 50% when compared to the standard HDCA method (Marathe *et al.*, 2013). Here, we expand upon this approach and show that embedded within this new method is a novel signal transformation that, when applied to EEG signals, significantly improves the signal-to-noise ratio and thereby enables more accurate single-trial analysis. The results presented here have significant implications for both brain–computer interaction technologies and basic science research into neural processes.**

***Index Terms*—Brain–computer interface (BCI), electroencephalography (EEG), hierarchical discriminant component analysis (HDCA) rapid serial visual presentation (RSVP), real-world environment, single-trial, sliding HDCA, temporal variability.**

I. Introduction

In RECENT years, various methods for single-trial electroencephalography (EEG) analysis have been developed [1]–[3]. Such single trial analysis methods have enabled the development of various brain–computer interaction (BCI) systems that use a person's EEG signals to directly interact with a computer or the external environment. BCI systems have been shown to work in controlled laboratory and medical settings; however, transitioning these approaches into more uncontrolled, complex, "real-world" environments presents several significant challenges. One such challenge is that dynamic, unconstrained environments often introduce additional temporal variability into the recorded EEG signal from several sources. The human brain is incredibly complex and the task of generating a response to a given stimulus involves several complex interrelated processes that can each be delayed for a variety of reasons. For example, temporal variability of the measured EEG signal has been shown to be influenced by endogenous process related to brain state (e.g., fatigue, attention), and exogenous factors such as stimulus properties [4]–[7].

Historically, trial to trial variability has been overcome by constraining experimental conditions and by averaging data across trials based on assumptions of signal stationarity. That is, data from time-locked events are averaged over time to capture an imperfect but representative sample of the activity that is time-locked to the stimulus of interest and remove activity that is not in phase with the stimulus. However, trial averaging does not account for or quantify any trial-to-trial temporal variability.

Single-trial EEG analysis methods have made significant progress in dealing with temporal variability in the neural signal by asking *if* a particular response occurred in response to a stimulus rather than *how* does the brain respond to a particular stimulus. Thus, by addressing this fundamentally different question, single-trial EEG analysis methods have fueled the development of a wide range of BCI technologies.

Methods for single-trial EEG classification can be divided into two basic categories. The first category of algorithms applies a spatial filtering algorithm to transform the multi-channel EEG signal into a new signal that contains more task-relevant information prior to applying a standard machine-learning classifier. Common spatial pattern (CSP) is an approach that looks for spatial filters that maximize the variance across two known conditions (e.g., target and nontarget) [8]. Variants of CSP, such as Common Spatio-Spectral Patterns [9], Common Sparse Spatio-Spectral Patterns [10], Common Spatio-Temporal Patterns [11], and Bilinear Common Spatial Patterns [12] are designed to look at combinations of spatial, temporal, and/or spectral features that aid in classification, and typically can outperform the basic CSP in specific circumstances. xDawn is another spatial filtering approach that seeks to maximize the difference in signal-to-noise ratio between the target class and nontarget class [13]. Best Lambda is another approach that uses a variance based spatial filter (similar to CSP) to reduce dimensionality, coupled with a linear discriminant analysis for classification purposes [14]. Importantly, Best Lambda forces an additional scarcity constraint which penalizes non-zero weights and typically leads to more robust classification.

The second category of classification algorithms are amplitude based time-domain approaches. These approaches began by using a simple logistic regression or Fischer linear discriminant to maximize the separation between classes [1], [15],

Manuscript received August 21, 2013; revised December 30, 2013 and January 23, 2014; accepted January 23, 2014. Date of publication February 12, 2014; date of current version March 05, 2014. This research was supported in part by an appointment to the U.S. Army Research Laboratory Postdoctoral Fellowship program administered by the Oak Ridge Associated Universities through a cooperative agreement with the U.S. Army Research Laboratory.

The authors are with U.S. Army Research Laboratory, Human Research and Engineering Directorate, Aberdeen Proving Grounds, MD 21005 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNSRE.2014.2304884

[16]. Hierarchical discriminant components analysis (HDCA), built upon these methods to include both spatial and temporal features of the signals to improve classification [16]–[19]; however, HDCA relies on temporal features that are consistent within a given time window.

Each of these existing methods has been shown to perform well on exemplar EEG data sets. Since temporal variability is inherent in all EEG data, it follows that each of these methods can handle temporal variability to some degree. In fact, many of the spatial filtering approaches typically work well in the face of temporal variability; however they are unable to quantify the temporal variability in the neural response.

Here we characterize the components of sliding HDCA (sHDCA), a recently developed method for single trial classification [20], and demonstrate their utility to quantify temporal variability in the neural response. Drawing on two previous studies [20], [21], we demonstrate how temporal variability of the neural response can negatively affect classification accuracy, and show that using a classifier designed to specifically account for temporal variability can significantly improve classification accuracy. Furthermore, we show that sHDCA includes a signal transformation that enables the quantification of trial-to-trial temporal variability through neural feature detection.

Participants performed a rapid serial visual presentation (RSVP) target detection task in which they were asked to identify specific targets from a sequence of rapidly presented visual stimuli. In a previous study, we reported that EEG recorded during this task resulted in large amounts of temporal variability in the peak latency of target-related neural activity that could largely be removed by aligning trials to the button response [21]. This study demonstrates that aligning the trials to the button to reduce the temporal variability decreases classification error by 75%. Previously, we showed that sHDCA reduces classification error by over 50% in the face of temporal variability in the neural response without needing information related to behavioral response time [20]. Here, we characterize the novel signal transformation contained within sHDCA and show that it enables single trial feature detection.

II. Methods

A. Participants

Fifteen participants (nine male, average age 39.5, 14 right handed) provided voluntary, fully informed consent in the current study as required by U.S. Army human use regulations [22], [23]. Participants reported normal or corrected-to-normal vision and reported no history of neurological problems.

The investigator adhered to the policies for the protection of human subjects as prescribed in U.S. Army human use regulations [22], and all procedures were performed in accordance to protocol ARL-20098-10025 approved by the U.S. Army Research Laboratory Institutional Review Board.

B. Stimuli and Procedure

As described previously, participants were shown short video clips in a RSVP paradigm [20], [21], [24] that either contained people or vehicles on background scenes, or only background scenes (Fig. 1). Observers were instructed to make a manual button press with their dominant hand when they detected a person or vehicle (targets), and to abstain from responding when a background scene (distracter) was presented. While a motor response was used in this task, previous work in RSVP based neural classification using HDCA has shown similar performance between motor and nonmotor tasks [16].

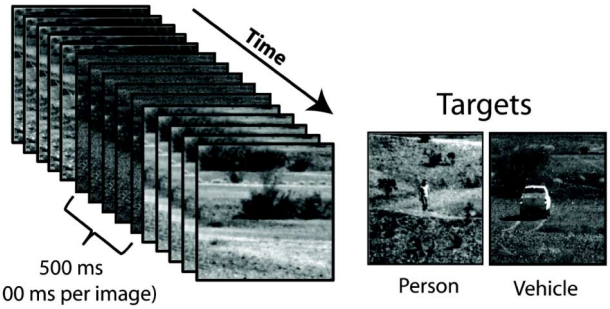


Fig. 1. Stimulus overview. Sequence of short videos RSVP. Stimuli were presented at 2 Hz. Participants were required to make a button press when a target was detected. Examples of target stimuli are shown to the right, and a nontarget stimulus is shown as the last frame in the sequence to the left.

Video clips consisted of five consecutive images each 100 ms in duration; each video clip was presented for 500 ms. This study treated the five 100 ms images as a single stimulus, thus while individual images were displayed at 10 Hz, videos consisting of five consecutive images were presented at 2 Hz. There was no interval between videos such that the first frame was presented immediately after the last frame of the prior video. If a target appeared in the video clip it was present on each 100 ms image. The distracter to target ratio was 90/10. RSVP sequences were presented in 2-min blocks after which time participants were given a short break. Participants completed a total of 25 blocks.

C. EEG Recording and Analysis

Electrophysiological recordings were taken from 64 scalp electrodes arranged in a 10-10 montage using a BioSemi Active Two system (Amsterdam, Netherlands) and sampled at 512 Hz. To record EOG, external leads were placed on the outer canthus and below the orbital fossa of both eyes. Continuous EEG data were referenced offline to the average of the left and right earlobes and digitally filtered 0.1–55 Hz. We removed EOG and EMG artifacts using independent component analysis (ICA) [25] to reduce muscle and ocular artifacts in the EEG signal and potential contamination with brain-based signals. After the ICA cleaning, we were left with 332 ± 27 (mean $\pm$ std) target epochs and 1753 ± 152 nontarget epochs.

1) ERP Analysis: ERP analyses were previously described in Ries and Larkin (2012) [21] and were used to illustrate the variability of the neural response. EEG data were processed and analyzed using EEGLAB [26] and ERPLab [27]. Continuous, artifact free data were epoched -1500 to 1500 ms around target and response onset. Target epochs followed by a button press within 200 to 1000 ms and nontarget epochs not followed by a response were included in the analysis. Averaging across all trials in a given condition may mask meaningful brain dynamics associated with performance especially in perceptually difficult tasks where the variance in ERP latency and reaction time (RT) increases [28]. Therefore, in order to assess the brain dynamics

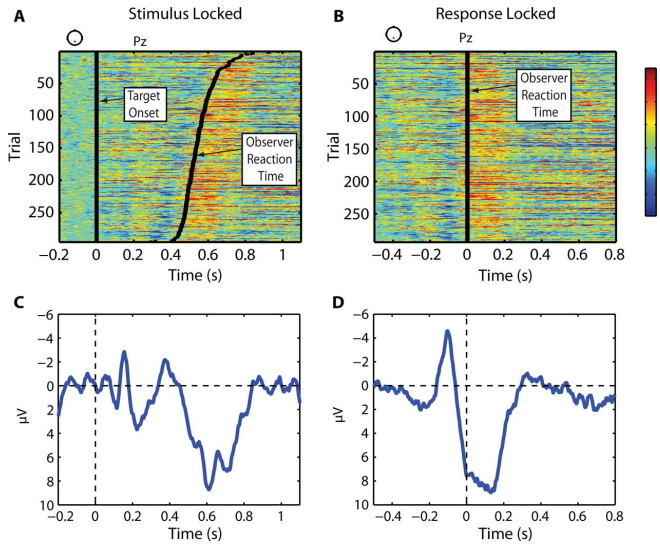


Fig. 2. Temporal variability in EEG of a single participant (S10). (A) Single trial EEG response at Pz when activity is aligned to the stimulus onset and sorted by response time. (B) Single trial EEG response at Pz when activity is aligned to the button press and sorted by stimulus onset time. Images (C) and (D) show trial averaged ERP wave forms calculated from (A) and (B), respectively. Adapted from [21] with permission.

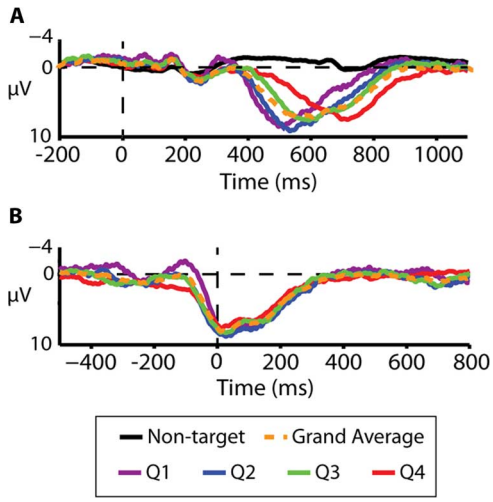


Fig. 3. Temporal variability in EEG across all participants. (A) Grand-averaged ERPs for data binned by reaction time quartiles for stimulus-locked data. The ERPs illustrate that the peak latency of target-related neural activity varies with reaction time. (B) Grand-averaged ERPs for data binned by reaction time quartiles for response-locked data. The ERPs illustrate that aligning to the response largely removes peak latency variability of target-related neural activity. Adapted from [21] with permission.

associated with varying levels of RT performance, target epochs were sorted into bins corresponding to an individual participant's reaction time quartile [29]. Grand averages across all subjects were then created for each stimulus and response-locked quartile.

For each subject, peak latency of target-related neural activity was calculated for stimulus-locked and response-locked averages in each quartile. Peak latency was measured at electrode Pz. Peak latency was designated as the time at which the waveform reached maximum positivity between 300 and 900 ms in stimulus-locked averages and between −200 and 400 ms in response-locked averages. Analyses from a previous study demonstrated the existence of temporal variability in the neural response as shown in Figs. 2 and 3. The results highlight temporal variability in the dataset, which makes it an appropriate dataset to demonstrate the effectiveness of sHDCA relative to other approaches.

2) Effect of Temporal Variability on Single Trial Classification: We extend the previous ERP analysis to demonstrate the effect of the temporal variability on single-trial classification. EEG data were binned into quartiles based on reaction time as described above. HDCA classifiers were then trained on target-evoked EEG data from one quartile and a random 75% of nontargets. The classifiers were then tested on data from each of the remaining three target quartiles and the remaining 25% of nontargets. This process was repeated four times such that each 25% of the nontargets were used as the test set. Performance measures were averaged across the four tests for each subject.

In this scheme, whenever the training quartile of the targets was not the same as the testing quartile of targets, all data in the test were completely novel to the classifier. In the case where the data were trained on the same quartile of targets as in the test set, the classifier performance was extraordinarily high as expected due to overfitting. These data, while presented for continuity of the graphs, were not used to demonstrate the effect of temporal variability on classification accuracy.

3) Single Trial Classification: The next step in the analyses was to apply single trial classification methods to the EEG data. The novel classification approach presented previously was a modification of HDCA. Consequently, HDCA served as an ideal baseline measure of classification performance for this study. Details of HDCA can be found in [16]–[19].

For classification purposes, EEG data were epoched −500 to 1600 ms around stimulus onset. Epoched EEG data were baseline corrected by removing the average of activity occurring between −500 and stimulus onset. Target epochs followed by a button press within 200 to 1000 ms and all nontarget epochs were included in the classification analysis.

a) Sliding Hierarchical Discriminant Components Analysis: As described in Marathe *et al.* 2012, sliding sHDCA builds upon the standard HDCA algorithm in an attempt to extract more information from temporally scattered events. sHDCA starts by using a standard HDCA classifier trained to discriminate targets versus nontargets (see Fig. 4). In this implementation of HDCA, the epoched data is divided into 50 ms time slices, and logistic regression (LR) classifiers are trained to discriminate targets from nontargets from each of the individual time slices between 300 and 800 ms after stimulus onset. In the standard HDCA algorithm, this initial classifier is typically a Fischer linear discriminant (FLD) classifier. We saw no significant differences between the FLD and LR implementation that we used. Each data point within the 50 ms time slice is treated as a repeated measure to train the individual classifiers. The output of each of the LR classifiers is then fed into another logistic regression classifier. In the standard HDCA, this second level LR classifier is used to make the final determination of target or nontarget. In sHDCA, rather than simply statically applying this classifier to each epoch, this initial classifier, which was created using data from 300–800 ms, is applied in a sliding fashion such that the leading edge of the HDCA classifier is applied at all time points between 100 and 1100 ms poststimulus. This sliding step

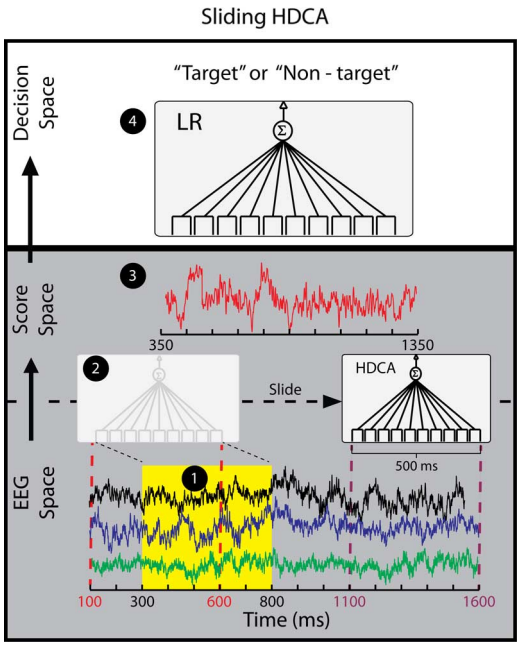


Fig. 4. Sliding HDCA method described previously [20]. Sliding HDCA starts with a standard HDCA classifier that uses ten 50 ms time windows that are trained on data that spans from 300 ms to 800 ms after stimulus onset as highlighted in yellow (step 1). That classifier slides in time such that the leading edge of the HDCA classifier is applied at all time points between 100 and 1100 ms to produce a 1000 ms sequence of scores (step 2–3). A logistic regression classifier is then applied to ten 100 ms time windows to produce a final interest score for the given epoch based on the 1000 ms score sequence (step 3–4). The final interest score is used to discriminate targets from nontargets. This study extends and validates the classification results of the entire sHDCA method and characterizes novel applications of the score space transformation shown in dark gray.

means that the sHDCA classifier is using epoch data from 100 ms poststimulus to 1600 ms poststimulus, which matches the data used by the standard HDCA algorithm.

Importantly, the parameter settings (e.g., number of temporal windows, window sizes, etc.) were not matched for the two methods. Instead, parameters were chosen that optimized performance for each method independently as a means to compare the best possible performance levels for each algorithm. For sHDCA, a limited sensitivity analysis was performed to identify parameters that maximized performance.

Since each application of the standard HDCA algorithm produces a single score, sliding the HDCA classifier in time produces a single score per application (per time point). When the sliding process is complete, we are left with a score signal that is 1000 ms in duration. From this score signal, a LR classifier is trained to discriminate targets versus nontargets based on the score signal. This LR classifier used for our second level classification divides the score signal into ten 100 ms time slices. The result of this LR classifier is the final score assigned to the epoch which is used to make the ultimate decision of whether the current epoch is a target or nontarget (see Fig. 4).

b) Alternative methods and comparison: Sliding HDCA classification performance was compared to other methods of single trial classification in an effort to validate the efficacy of this new method. Each method tested was applied to the exact same data as sHDCA, and we used specific parameter values that maximized the performance of each individual classifier across the population of participants. In addition to the standard HDCA, we chose two other alternative methods to compare against the sHDCA results. Cross correlation is a measure of similarity of two waveforms as a function of a time lag applied to one of them. In the context of single trial classification, cross correlation provides a simple means of pattern matching over time. A method for potentially addressing temporal variability is to simply use cross correlation to match the ERP pattern seen in a training set. For each subject, the response-locked ERP was calculated from the training set for all 64 channels. A cross correlation was calculated for each channel with its corresponding response-locked ERP, which resulted in 64 time series signals. HDCA was then applied to the 64 signals to discriminate targets from nontargets. The accuracy of this classification scheme was compared to both HDCA and sHDCA.

Another alternative method is to apply a standard HDCA to smoothed EEG data. Sliding an HDCA classifier in time might simply smooth the input space for the second level LR classifier used in sHDCA. If this were true, smoothing the raw EEG signals prior to applying a standard HDCA should achieve similar results to sHDCA. We applied a running average mean to the EEG data prior to training and testing and HDCA classifier to test this theory. The optimal amount of prior smoothing is unknown, therefore we experimented with running average means of 62.5, 125, 250, 500, and 1000 ms. Of these, 125 ms smoothing performed the best and was compared to the both standard HDCA and sliding HDCA.

4) Cross Validation: A 10-fold cross validation was used to determine the accuracy for all classification methods applied in this study. Data from each subject were divided into 10 equal sized blocks of trials. Classifiers were trained on nine of the 10 blocks, and then tested on the block left out. This process was repeated 10 times such that each of the 10 blocks of trials was used as the independent testing set once. Each training block was further divided into two equal and independent parts. The first part was used to train the initial HDCA algorithm. That HDCA algorithm is then applied (in sliding fashion) to the second part of the training set to produce a score signal, and the second level LR classifier that was then trained from that score signal. Finally, the entire two-step process was applied to the independent test set. Performance was evaluated based on the area under the receiver operating characteristic curve (AUC). Each participant's performance was calculated as the average AUC calculated across all 10 cross validation sets. Statistical analyses for each classification method were performed on the average AUC for each participant.

5) Signal-to-Noise Ratio: Sliding HDCA classification transforms multi-channel EEG signals into a single score signal before using a classifier to discriminate targets from nontargets. This transformation from a multi-channel EEG signal to a score signal serves to accentuate discriminating features of the EEG signal. This effect is best quantified by comparing the signal-to-noise ratio (SNR) for the EEG signal with the SNR for the score signal. In each epoch of data, 1 s of data was designated as "signal" and 1 s of data was designated as "noise." The SNR was calculated as follows:

$$\mathrm{SNR} = \frac{\mathrm{rms(signal)}}{\mathrm{rms(noise)}} \tag{1}$$

where rms is the root mean square as defined below

$$\mathrm{rms} = \sqrt{\frac{\sum_{i=1}^{n} a_i^2}{n}}. \tag{2}$$

SNR was calculated epoch by epoch for both the EEG signal at electrode Pz and the score signal generated by sHDCA. The SNR values were averaged for each signal type and compared using a Wilcoxon sign rank test. The SNR analysis was potentially confounded by the fact that we were directly comparing SNR from the score signal, which incorporates information from all electrode sites, to a single channel of EEG. To ensure that we were not unfairly biasing our results, we also calculated the SNR from the average signal across all EEG electrodes. No significant differences were found between the SNR of the average EEG signal compared to the SNR of the signal at Pz; thus, for purposes of reporting only information related to the SNR for the signal at Pz is presented.

6) Single Trial Feature Detection: As described in the previous section, the transformation of the multi-channel EEG signals into a score signal serves to accentuate discriminating features of the EEG signal to improve single trial classification. Accentuating these features also enables accurate detection of neural features from single-trial data. For all subjects, the reaction time was predicted from both the score signal and the EEG signal recorded at Pz. For EEG signals, data from time points ranging from 200 to 1000 ms after stimulus onset were used for prediction. For score signals, data from 550 to 1350 ms were used for prediction (note: the first score sample occurs at 350 ms (see Fig. 4), thus this time window is 200 to 800 ms relative to the start of the score signal). The score signals were produced using a 10-fold cross validation. Previous studies have shown that P300 peak latency can be highly correlated with reaction time for tasks that require simple stimulus discrimination and produce fast reaction times [5], [21], [30], [31]. Because the timing and scalp distribution of the target related neural activity in this data was indicative of P300-related processes, the time point at which the signal achieves its maximum was taken as the estimated reaction time. Future studies will fully explore alternative methods for predicting the peak latency.

An important consideration is the difference in the time scale between EEG space and score space. Each time point in the score signal is derived from 500 ms of data from the original EEG signal. Thus, an 800 ms score signal is derived from 1300 ms of the original EEG data. Thus comparing measures derived from 800 ms of EEG signal to measures derived from 800 ms of score signal as we have described above, presents a mismatch in terms of the time span considered by each method.

We quantified the accuracy of our single trial reaction time estimates by comparing them to the participants' reaction time on a trial by trial basis. The accuracy was quantified as follows.

Accuracy was measured as the percent error for each peak latency prediction as given below

$$\mathrm{PE} = \frac{|RT - pRT|}{RT} * 100 \tag{3}$$

where RT is the actual reaction time and pRT is the predicted reaction time. To quantify the overall accuracy of all reaction time predictions for a given subject, we calculated the distribution of prediction errors (DPE) for each subject

$$\mathrm{DPE} = \mathrm{PE} < X, \quad X \epsilon [0, 1, 2 \ldots 100\%]. \tag{4}$$

DPE was a 101 element vector that was calculated for each participant. The vectors were then averaged across all 15 participants and the mean and standard deviations of these vectors were reported as the accuracy of reaction time estimates from EEG and score signals. An area under the curve (AUC) measurement was used to quantify the overall accuracy of reaction time estimates. A Wilcoxon sign rank test compared the AUC between EEG and score signals.

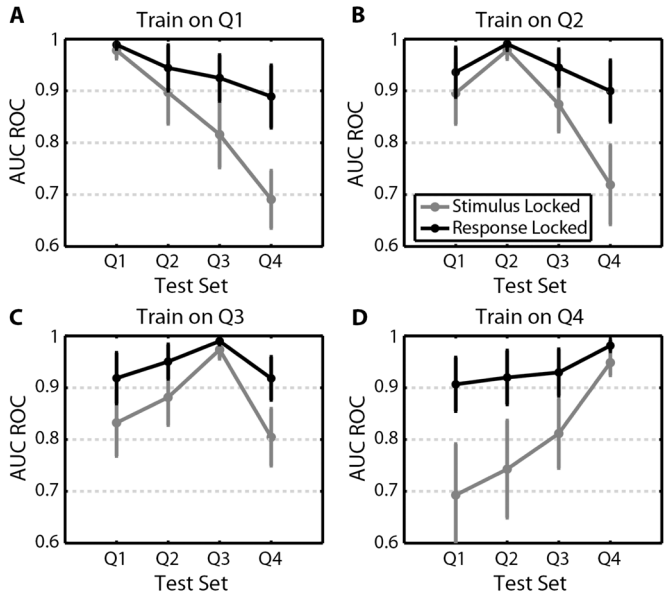


Fig. 5. Effect of response latency on classification performance for stimulus-locked (gray) and response-locked (black) data. An HDCA classifier was built on 25% of the trials based on response times, and then tested on each block of 25% of trials (sorted by response time). Classification performance declines as the response time of the test set deviates from the response time of the training set. Each panel shows results from training the classifier on data from the four quartiles of response time. Error bars show standard deviation across subjects.

III. Results

A. Effect of Temporal Variability on Classification

Fig. 5 illustrates that temporal variability causes a decline in classification performance using existing methods. In Fig. 5(a), an HDCA classifier was built on the 25% of the trials with the fastest response times, i.e., first quartile (Q1) for each participant, and then tested on each quartile of trials (sorted and binned by response times) independently. When the Q1 classifier is applied to Q1 trials, performance is predictably high since the classifier is tested and trained on the same data. However, comparing the results of applying the Q1 classifier to Q2, Q3, and Q4 trials, classification accuracy drops off linearly as the average reaction time of the test set deviates from the average reaction time of the training set when using the stimulus-locked epochs (gray lines). The response-locked epochs have largely reduced the temporal variability, and applying the Q1 classifier to Q2, Q3, and Q4 in these data results in a much smaller drop in performance than in the stimulus-locked data. It is important

to note, however; that classification performance does still drop when using the response locked data. This supports the idea that there exists neural variability that is not linearly related to or separate from the response time variability or that there is significant variability within each quartile.

A two-way analysis of variance (ANOVA) shows that the main effects of test set ($\mathrm{F}(3, 112) = 86.45$, $\mathrm{p} < 0.001$) and event locking method ($\mathrm{F}(1, 112) = 107.84$, $\mathrm{p} < 0.001$) are both significant. Furthermore, the interaction between test set and event locking method was also significant ($\mathrm{F}(3, 112) = 21.77$, $\mathrm{p} < 0.001$), which supports the idea that the classification performance declines more for stimulus-locked data than for response-locked data. This suggests that temporal variability in the neural response leads to a decline in classification performance. Fig. 5(b) shows results when Q2 is used as the training set. Again as the average reaction time of the test set deviates from the average reaction time of the training set—performance rapidly declines for the stimulus-locked epochs. Response locked epochs again show a similar, but much smaller decline in performance. Just as with Fig. 5(a), a two-way ANOVA shows that effects of test set ($\mathrm{F}(3, 112) = 63.61$, $\mathrm{p} < 0.001$), event locking ($\mathrm{F}(1, 112) = 70.86$, $\mathrm{p} < 0.001$) and the interaction term ($\mathrm{F}(3, 112) = 16.71$, $\mathrm{p} < 0.001$ are all significant. This pattern continues in Fig. 5(c) and (d) where the classifier was trained on Q3, and Q4 trials, respectively.

Fig. 6 provides further evidence that temporal variability causes dramatic declines in classification performance. The left two boxes show the classification performance for the standard HDCA classifiers under the stimulus-locked (Mean AUC ± Std: 0.8691 ± 0.0359) and response-locked epochs (Mean AUC ± Std: 0.9667 ± 0.0173), respectively. Reducing the temporal variability resulted in a 75% decrease in classification error.

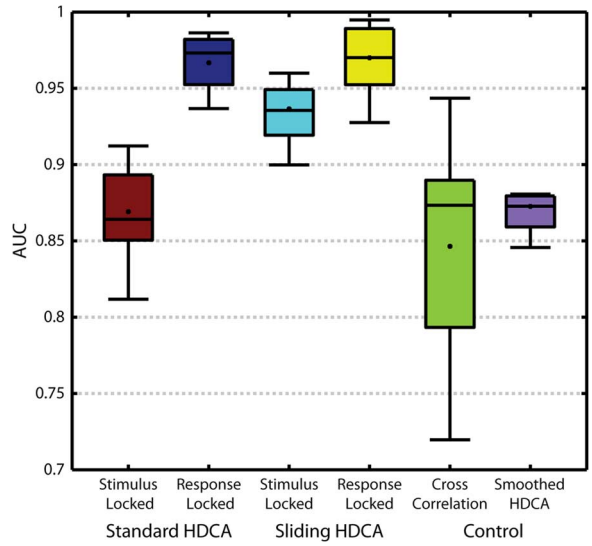


Fig. 6. Classification results across 15 subjects. Horizontal lines in each box represent the median and the dot represents the mean. The first (left-most) bar shows the classification accuracy when using the standard HDCA algorithm on stimulus-locked data. Reducing the temporal variability by aligning trials by the response time dramatically improves classification accuracy (second bar). Applying sHDCA to the stimulus-locked data improves classification accuracy over the stimulus-locked HDCA, but not to the same degree as when the response-locked data was used. Applying sHDCA to response-locked data (fourth bar) does not change the overall classification accuracy when compared to the HDCA on response-locked data (second blue). Using a cross correlation based classification (fifth bar) and a smoothed HDCA (sixth bar) does not improve classification performance over the standard HDCA on stimulus-locked data (see text for details on these methods). A portion of these results were presented at the HCII 2013 Conference [20].

B. Classification in the Face of Temporal Variability

Temporal variability exists in the neural signals and has a large negative impact on single trial classification accuracy. The question remains: is there a classification scheme that can overcome the loss of performance caused by the neural temporal variability without needing to align trials to a physical response? Sliding HDCA was able to regain most of the accuracy lost due to temporal variability of the neural response. Fig. 6 presents a box and whisker plot of the classification results across all 15 subjects. The maroon and blue boxes show the classification performance for the standard HDCA classifiers under the stimulus-locked and response-locked conditions as described above. The cyan box represents the classification accuracy seen using the sHDCA on stimulus-locked epochs. The classification accuracy of sHDCA was 0.9365 ± 0.0223 (mean ± std AUC). As reported before [20], this represents a 51.5% reduction of classification error over the standard stimulus-locked HDCA and the overall difference is statistically different (Wilcoxon Sign Rank Test $\mathrm{p} < 0.001$).

Several alternative classification schemes were also applied to these data. First, sHDCA was applied to response-locked epochs to evaluate the effect of sHDCA on data with little or no temporal variability. If sHDCA were to improve performance over the standard HDCA (on response-locked data), then sHDCA would be improving performance based on factors other than temporal variability. If sHDCA were to exhibit worse classification performance than HDCA, then the use of sHDCA would be limited to situations where temporal variability could be predicted *a priori*. Performance on the response-locked epochs was almost identical to the HDCA performance on response-locked data (sHDCA: 0.97 ± 0.02 AUC, HDCA: 0.9667 ± 0.01, Wilcoxon Sign Rank Test $\mathrm{p} > 0.05$).

Another method applied here was to use the cross correlation measure as a means of addressing temporal variability. The fifth box in Fig. 6 shows the classification accuracy of the cross correlation classification. Performance is statistically worse than sHDCA (Wilcoxon Sign Rank test, $\mathrm{p} < 0.01$) and not significantly different from HDCA on stimulus-locked data (Wilcoxon sign rank test, $\mathrm{p} = 0.25$). The final method was to apply a standard HDCA to smoothed EEG data. Sliding an HDCA classifier in time might simply smooth the input space for the LR classifier used in sHDCA. If this were true, smoothing the raw EEG signals prior to applying a standard HDCA should achieve similar results to sHDCA. The last box in Fig. 6 shows that the accuracy of smoothed HDCA actually is not significantly different from HDCA (Wilcoxon Sign Rank Test, $\mathrm{p} = 0.3$).

C. Effect of Sliding HDCA

The power of sHDCA stems from its ability to transform the highly variable EEG signal into a more stable score space signal

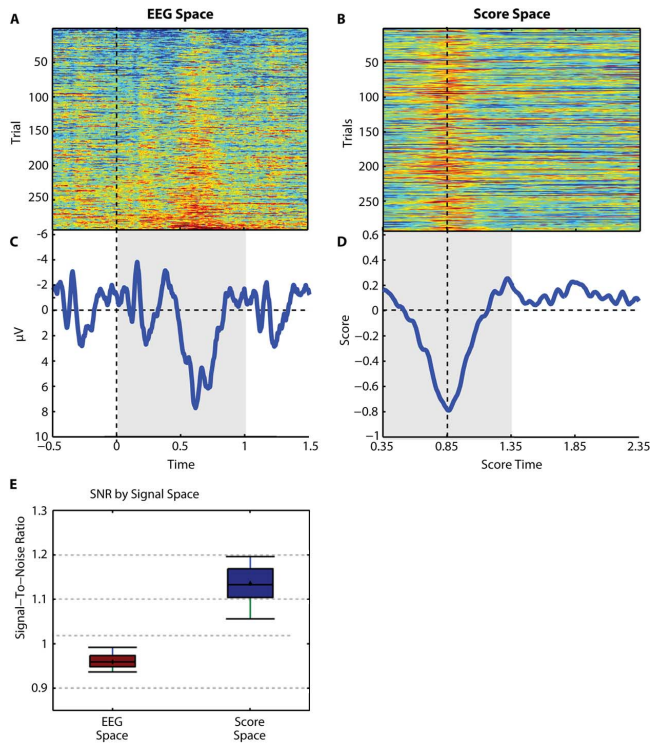


Fig. 7. Effect of sliding HDCA on EEG signal. (A) Shows EEG data at Pz sorted by the average amplitude calculated between 200 and 1000 ms post-stimulus onset for a single subject (S10). (B) Shows the score signal sorted in the same order as (A) for the same subject. (C) and (D) show the averaged ERP waveform of the trials shown in the (A) and (B) respectively. (E) Compares the signal to noise ratio across all 15 subjects of the raw EEG signal at Pz and the score signal obtained by sliding an HDCA classifier in time. SNR was calculated as the rms of the signal (gray regions in C and D) divided by the rms of the noise (white region in C and D).

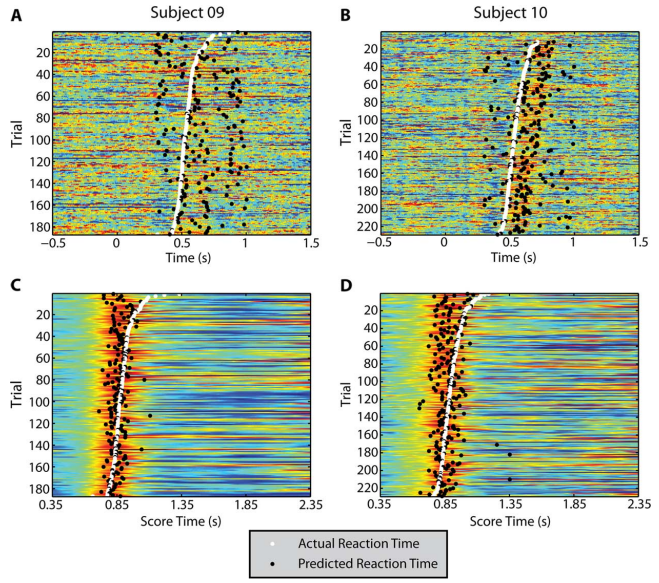


Fig. 8. Neural feature prediction for two participants. In all plots, the white dots represent the actual reaction time, and the black dots represent the predicted reaction time. (A) Shows reaction time predictions for subject 09 using EEG space. (B) Shows reaction time prediction for subject 10 using EEG space. (C) Shows reaction time predictions for subject 09 using score space. (D) Shows reaction time prediction for subject 10 using score space.

that accentuates the most discriminating features of the neural response (see Fig. 7).

Fig. 7(a) shows a raster plot of the EEG data collected from subject S10 from electrode Pz for all behaviorally correct target trials sorted by the average amplitude between 200 and 1000 ms poststimulus onset. Fig. 7(b) shows a raster plot of the score signal produced by sliding the embedded HDCA classifier in time. The trials are sorted in the same order as in Fig. 7(a). A comparison of these two plots illustrates two main points. First, the EEG signal contains a great deal of trial to trial variability in the peak amplitude, while the score signal has much more consistent peak amplitude across trials. Second, the score signal is more stable in the periods away from the discriminating regions (-0.5 to 0 and 1 to 1.5 s in EEG and 1.35 to 2.35 s in score time). These effects are more apparent in the trial averaged signals shown in Fig. 7(c) and (d). This effect can be quantified by calculating the overall signal-to-noise ratio (SNR) in the two signal spaces. Score space significantly (Wilcoxon sign rank test $p < 0.001$) increases SNR by an average of 18.4% across all 15 subjects [Fig. 7(e)]. To calculate the SNR, we defined the signal as the areas shaded in gray in Fig. 7(c) and (d) (epoch times corresponding to 0 to 1 s in EEG space and 0.35 to 0.85 s in score space). The region in white in those plots was taken to be the noise (epoch times corresponding to -0.5 to 0 and 1 to 1.5 s in EEG space and 1.3.5 to 2.35 s in score space).

One apparent difference between EEG space and score space is the shift in time. Recall that sHDCA slides a 500 ms HDCA classifier in time to produce score space. The HDCA classifier is first applied to data between 100 and 600 ms poststimulus onset. This application of HDCA produces a single score, which is assigned to 350 ms time point (midpoint of the 100 to 600 ms window). The HDCA classifier then slides in time sample by sample producing a single score for each application. For purposes of plotting rasters, the sliding step was carried out across 2 s of data; however only the first second of data was used for classification (see Fig. 4).

D. Peak Latency Estimation

In addition to the improved classification performance, sHDCA also enables estimation of reaction time from single trial data. Here we employ a simple peak-picking method to estimate the reaction time. Previous work has shown that neural measures of the peak amplitude is strongly correlated with reaction time for simple discrimination tasks that produce fast responses [5], [21], [30], [31]. The same analysis was done with both the EEG signal at electrode Pz and from the score signal obtained in sHDCA to compare the quality of estimates from each signal space. Fig. 8 provides a qualitative comparison of reaction time estimates for two participants. Subject 09 provides an example where EEG based estimates [Fig. 8(a)] are highly erratic from trial to trial, while score based estimates [Fig. 8(c)] are more closely coupled to the reaction times (see below for quantitative results). Subject 10 provides an example where both EEG and score based estimates are closely coupled to the actual reaction times. Fig. 9 provides a more quantitative assessment of the accuracy of our estimates. Fig. 9(a) quantifies the predictions seen in Fig. 8. For both participants (S09, S10),

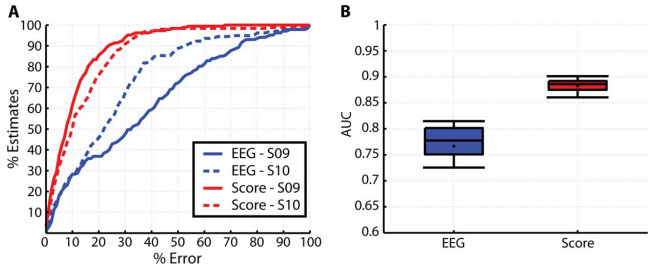


Fig. 9. Reaction time prediction accuracy is quantified by plotting the percentage of estimates (y-axis) that are within a given error window (x-axis). (A) Shows the accuracy for subjects S09 (solid lines) and S10 (dashed lines) individually for both EEG (blue) and score (red) based predictions. (B) Shows the area under the curve across all 15 subjects. Score space based preditions were significantly more accurate than EEG based predictions (Wilcoxon sign rank test $p < 0.05$).

the score based predictions had a larger proportion of estimates with less error. Fig. 9(b) shows that across all 15 subjects, score based predictions were significantly more accurate (Wilcoxon sign rank test, $p < 0.05$).

IV. Discussion

The current study employed a dynamic RSVP task in which participants were asked to identify the presence of people and vehicles in rapidly presented short-duration videos. Averaged ERP analyses demonstrated the existence of temporal variability in the neural response. Previously, we have demonstrated that a novel classification method that can overcome temporal variability in the neural response at the single-trial level [20]. Here, we extended and validated those results and demonstrated novel applications for components of the sHDCA classifier.

We showed that the temporal variability in the neural response was largely removed by aligning trials to the button response (see Figs. 2 and 3 [21]. State-of-the-art classification methods showed a dramatic decrease in classification accuracy with increased temporal variability in stimulus-locked data. Sliding HDCA classification is a novel classification method described here that reduces classification error by over 50% over a standard HDCA classifier using the same amount of data.

The improvement in performance was characterized using the parameter values described in Section II. In addition to those parameter settings, we also performed a limited sensitivity analysis to determine the optimal parameter set for this data set. In particular we varied the size and start time of the temporal window used by the embedded HDCA classifier. In this analysis, we compared classification performance by varying the temporal window start time between 0 and 400 ms, and temporal window end time between 600 and 1000 ms. Based on this analysis, we chose to present the results from using the smallest window that maximized performance (i.e., 300 to 800 ms). Similarly, the window size for the sliding step was varied in terms of its start time and end time. Performance plateaus with windows of 1 s or more. The smallest such window that maximized performance was chosen for analysis (-200 to 800 ms).

Finally, the sliding interval was also varied. We tested sliding the initial HDCA classifier by a single time point, every two time points, every five time points and every 10 time points. In this test, performance fell off dramatically when the sliding interval was increased beyond a single time point.

Based on our initial results, we presumed that sliding HDCA improved classification performance over standard HDCA on stimulus locked data by accounting for temporal variability in the neural response. We then applied sHDCA and HDCA to response locked data, where temporal variability was largely removed. In this test, had sHDCA performance been substantially worse than HDCA on the response locked data, then it would indicate that on data with little or no temporal variability, sHDCA would cause a drop in performance. Alternatively, if performance had been substantially higher with sHDCA on response locked data, then it would indicate that sHDCA was dealing with something other than temporal variability. Since sHDCA performance was neither better nor worse than HDCA on response locked data, we can conclude that the improvement seen in sHDCA is primarily due to its ability to address temporal variability in the neural response. It is important to note however; that response locked classification accuracy was at 0.97 (Az), thus there was little room for sHDCA to improve performance. Thus it is technically possible that there are other factors outside of temporal variability that enabled sHDCA to improve performance.

Another important issue is whether or not this method could work on data without overt motor responses. This issue is not addressed directly in this paper because the primary purpose was to demonstrate effective classification in the face of temporal variability. However, given that tradition BCIs are often expected to function without overt responses, it is important to consider whether this method could be effective in such applications. Previous work has shown that HDCA can effectively classify neural responses in the absence of motor responses [16]. Since sHDCA is based on, and more importantly directly uses, HDCA it follows that sHDCA *should* also be able to effectively classify neural responses in the absence of motor responses. Validation of this claim is warranted, and is currently underway in a separate study.

The process of transforming a signal from EEG space to score space by sliding an HDCA classifier in time served to improve the overall SNR of the signal by an average of 18.4%. This not only improved classification accuracy, but also enabled more accurate estimates of reaction time from single trial data using a simple peak picking method. The results obtained from applying this method in score space were compared against results obtained from applying the same method to the EEG signal recorded at Pz. Analysis revealed that score-based estimates were more accurate than EEG-based estimates.

The reaction time estimates performed in this study yielded accurate predictions by simply picking the maximum point in an epoch of score data. Approximately 40% of the estimates were within 10% of the actual reaction time, and nearly 70% of estimates were within 20%. Several previous studies have developed methods to pick the peak of the neural response [32]–[44]. However, most of these methods either only worked on simulated data or are not able to function on a single trial level. One previous study has demonstrated the ability to estimate stimulus onset time based on single trial data [45]. To our knowledge, no other study has estimated reaction times based on single

trial data. Others have correlated various features of the ERP to the response time [4] without explicitly estimating the reaction time.

The reaction time estimation performed in this study compared the accuracy of estimates derived from an 800 ms window in EEG space with an 800 ms window in score space. Since each time point in score space is calculated from 500 ms of data from the underlying EEG, there is an apparent mismatch in the amount of time used to estimate peak latency. A closer analysis of the time windows reveals that the excess time points allotted to the score signal is from 0 to 200 ms and 1000 to 1300 ms in EEG time. Both of these time spans do not contain the peak of the neural response, and thus serve as extra time which only serves to add noise into the peak prediction. This extra time only makes the decision harder for the score signal.

The alternative to this method would be to match the amount of time in the underlying EEG. One could argue that matching the temporal windows in such a manner would be a more correct method for comparing accuracy. Such a time matched analysis however; would compare a 1300 ms EEG time window against an 800 ms score space window (or some window size where the EEG window was 500 ms longer than the score window). All of the excess time would simply add data points to the EEG signal that do not contain the peak of the neural activity. These additional data points would only serve to add noise to the latency estimates, and thus make it more difficult to accurately pick the peak of the neural response. Based on these tradeoffs, we chose to tip the scales towards the EEG in order to show that the even when we tried to make it easier for the EEG based predictions; the score based predictions were still more accurate.

The results of the current study provide several avenues for future research. First, the score space transformation may enable the development of tools to enable a better understanding of neural processing. Current methods for analyzing EEG signals often rely on averaging several trials to overcome the variability in the EEG signal. The score space transformation removes a great deal of trial-to-trial variability from the underlying EEG signal as shown in Fig. 7(a) and (b) and improves the signal-to-noise ratio by 18.4%. This decreased variability and improved SNR may allow for a clearer understanding of how the brain functions in various situations. In particular, the improved SNR may enable the detection of novel low amplitude neural features, or may enable novel analyses to explore brain function based on a smaller number of trials. Furthermore, using the score space transformation, we were able to use an exceedingly simple peak picking approach to estimate the reaction time on a trial-by-trial basis which suggests that the score space transformation enables greater information extraction. Such a method may be generalized to identify other neural features from single-trial EEG.

In the current data set, there are examples of trials where it appears that the neural signals are fundamentally different for trials in which the behavior is the same. In most trials, the brain produces a prototypical response that corresponds to the data the classifier was trained on when a target is seen, and these trials usually lead to the participant pressing the button. However, there are also trials where this relationship fails. Furthermore, in some trials, the prototypical response is produced, enabling the classifier to correctly identify the target; however the participant fails to press the button. In other trials, the participant correctly presses the button, but the neural activity is drastically different from the "normal" activity and thus the classifier fails to detect the target. Developing single-trial analysis tools that can identify neural features will help investigate how the brain works in these scenarios.

A second avenue of future research will be to continue the development of single-trial classification methods that can overcome temporal variability in the neural response. We were able to develop a novel classification method that reduces classification error by over 50% in the presence of temporally variable neural signals. While this finding represents a significant advancement over current state-of-the-art methods, even further improvements may be possible. Initial analysis showed that using the response-locked data in which most of the temporal variability had been removed; classification error was decreased by over 75%. If we are able to develop a method to accurately identify neural features as described above, then we may be able to use that method to align the trials in real-time in order to reduce the temporal variability prior to classification. Such a method could potentially reach or even exceed accuracy levels of the response-locked classifier even on nonresponse tasks.

Finally, this study has broad implications for the future of human–machine interactions. Recent advances in the computer technology, signal processing, and neuroscience have led to the development of a wide range of BCIs, and have the potential for many future technologies [46]. In this study, we focus on RSVP-based BCIs for target detection as one specific BCI that has been developed. sHDCA may enable various advancements to current systems. By creating a method to enhance classification in the face of temporal variability without using information related to a behavioral button response, we enable the development of applications that either replace behavioral responses with neural classification or use neural classification to augment behavioral responses that may be unreliable. Furthermore, sHDCA may also enable identification of targets from a continuous stream of EEG data, thereby removing the need to perform all analyses time-locked to stimulus onset. Sliding the initial HDCA classifier in time produces a single score signal in time with a signal to noise ratio that is 18.4% greater than the original EEG signal. From this continuous score signal, it may be possible to identify the neural response to target stimuli without epoching the data relative to stimulus onset. Removing the need to time-lock analysis to specific events could enable future image analysis paradigms that follow a more natural search pattern (as opposed to an RSVP paradigm). Such advancements will be necessary to move BCI technologies into "real-world" environments.

The current study explicitly addressed classification in the face of temporal variability. In a data set where variability likely arose due to a combination of exogenous (varied stimulus difficulty) and endogenous (changing attention/fatigue levels through experiment) factors, we demonstrate that using a method designed to account for temporal variability can reduce classification error by over 50%. In more complex, "real-world" environments, temporal variability is also likely to occur as a result of system properties such as the inability to precisely

synchronize EEG and event data and the inability to precisely time the occurrences of events. While the source and amount of the variability may be different, our results indicate that employing methods specifically designed to account for temporal variability will dramatically improve classification accuracy in these situations. Furthermore, this approach may also enable an accurate prediction of neural features on a trial-by-trial basis, which has implications for basic neuroscience research and future BCI technologies.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for suggesting modifications that optimized the sHDCA algorithm and led to substantial improvement in this manuscript.

REFERENCES

[1] L. C. Parra, C. D. Spence, A. D. Gerson, and P. Sajda, "Recipes for the linear analysis of EEG," *NeuroImage*, vol. 28, no. 2, pp. 326–341, Nov. 2005.

[2] K.-R. Muller, M. Tangermann, G. Dornhege, M. Krauledat, G. Curio, and B. Blankertz, "Machine learning for real-time single-trial EEG-analysis: From brain-computer interfacing to mental state monitoring," *J. Neurosci. Methods*, vol. 167, no. 1, pp. 82–90, 2008.

[3] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 4, 2007.

[4] T. Kammer, L. Lehr, and K. Kirschfeld, "Cortical visual processing is temporally dispersed by luminance in human subjects," *Neurosci. Lett.*, vol. 263, no. 2–3, pp. 133–136, Mar. 1999.

[5] J. R. Folstein and C. Van Petten, "After the P3: Late executive processes in stimulus categorization," *Psychophysiology*, vol. 48, no. 6, pp. 825–841, 2011.

[6] S. K. L. Lal and A. Craig, "Driver fatigue: Electroencephalography and psychological assessment," *Psychophysiology*, vol. 39, no. 3, pp. 313–321, 2002.

[7] A. Craig, Y. Tran, N. Wijesuriya, and H. Nguyen, "Regional brain wave activity changes associated with fatigue," *Psychophysiology*, vol. 49, no. 4, pp. 574–582, 2012.

[8] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Dec. 2000.

[9] S. Lemm, B. Blankertz, G. Curio, and K.-R. Muller, "Spatio-spectral filters for improving the classification of single trial EEG," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 9, pp. 1541–1548, Sep. 2005.

[10] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K.-R. Muller, "Combined optimization of spatial and temporal filters for improving brain-computer interfacing," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 11, pp. 2274–2281, Nov. 2006.

[11] K. Yu, K. Shen, S. Shao, W. C. Ng, K. Kwok, and X. Li, "Common spatio-temporal pattern for single-trial detection of event-related potential in rapid serial visual presentation triage," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 9, pp. 2513–2520, Sep. 2011.

[12] K. Yu, K. Shen, S. Shao, W. C. Ng, and X. Li, "Bilinear common spatial pattern for single-trial ERP-based rapid serial visual presentation triage," *J. Neural Eng.*, vol. 9, no. 4, p. 046013, Aug. 2012.

[13] H. Cecotti, B. Rivet, M. Congedo, C. Jutten, O. Bertrand, E. Maby, and J. Mattout, "A robust sensor-selection method for P300 brain-computer interfaces," *J. Neural Eng.*, vol. 8, no. 1, p. 016001, Feb. 2011.

[14] J. Touryan, L. Gibson, J. H. Horne, and P. Weber, "Real-time measurement of face recognition in rapid serial visual presentation," *Front. Psychol.*, vol. 2, Mar. 2011.

[15] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley-Interscience, 2000.

[16] A. D. Gerson, L. C. Parra, and P. Sajda, "Cortically coupled computer vision for rapid image search," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 174–179, Jun. 2006.

[17] L. C. Parra, C. Christoforou, A. D. Gerson, M. Dyrholm, A. Luo, M. Wagner, M. G. Philiastides, and P. Sajda, "Spatiotemporal linear decoding of brain state," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 107–115, 2008.

[18] P. Sajda, E. Pohlmeyer, J. Wang, L. C. Parra, C. Christoforou, J. Dmochowski, B. Hanna, C. Bahlmann, M. K. Singh, and S.-F. Chang, "In a blink of an eye and a switch of a transistor: Cortically coupled computer vision," *Proc. IEEE*, vol. 98, no. 3, pp. 462–478, Mar. 2010.

[19] E. A. Pohlmeyer, J. Wang, D. C. Jangraw, B. Lou, S.-F. Chang, and P. Sajda, "Closing the loop in cortically-coupled computer vision: A brain-computer interface for searching image databases," *J. Neural Eng.*, vol. 8, no. 3, p. 036025, Jun. 2011.

[20] A. R. Marathe, A. J. Ries, and K. McDowell, "A novel method for single-trial classification in the face of temporal variability," in *Foundations of Augmented Cognition*, D. D. Schmorrow and C. M. Fidopiastis, Eds. Berlin, Germany: Springer, 2013, pp. 345–352.

[21] A. J. Ries and G. B. Larkin, Stimulus and Response-Locked P3 Activity in a Dynamic RSVP Task ARL-TR-6314, Technical Report 6314, 2012.

[22] U.S. Dept. Army, Use of volunteers as subjects of research Washington, DC, AR 70-25, 1990, Government Printing Office.

[23] U.S. Dept. Defense Office Secretary Defense, Code of Federal Regulations, Protection of Human Subjects Washington, DC, 32 CFR 219, 1999, Government Printing Office.

[24] J. Touryan, L. Gibson, J. H. Horne, and P. Weber, "Real-time classification of neural signals corresponding to the detection of targets in video imagery," presented at the Int. Conf. Appl. Human Factors Ergonom., Miami, FL, 2010, p. 60, unpublished.

[25] T.-P. Jung, S. Makeig, C. Humphries, T.-W. Lee, M. J. McKeown, V. Iragui, and T. J. Sejnowski, "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, vol. 37, no. 2, pp. 163–178, 2000.

[26] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004.

[27] S. J. Luck and J. Lopez-Calderon, *ERPLAB Toolbox*. : , 2010.

[28] S. J. Luck, *An Introduction to the Event-Related Potential Technique*, 1st ed. Cambridge, MA: Bradford Book, 2005.

[29] R. Poli, C. Cinel, L. Citi, and F. Sepulveda, "Reaction-time binning: A simple method for increasing the resolving power of ERP averages," *Psychophysiology*, vol. 47, no. 3, pp. 467–485, 2010.

[30] A. Magliero, T. R. Bashore, M. G. H. Coles, and E. Donchin, "On the dependence of P300 latency on stimulus evaluation processes," *Psychophysiology*, vol. 21, no. 2, pp. 171–186, 1984.

[31] J. Dien, K. M. Spencer, and E. Donchin, "Parsing the late positive complex: Mental chronometry and the ERP components that inhabit the neighborhood of the P300," *Psychophysiology*, vol. 41, no. 5, pp. 665–678, 2004.

[32] C. D. Woody, "Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals," *Med. Biol. Eng.*, vol. 5, no. 6, pp. 539–554, Nov. 1967.

[33] M. Kutas, G. McCarthy, and E. Donchin, "Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time," *Science*, vol. 197, no. 4305, pp. 792–795, Aug. 1977.

[34] P. D. Tuan, J. Möcks, W. Köhler, and T. Gasser, "Variable latencies of noisy signals: Estimation and testing in brain potential data," *Biometrika*, vol. 74, no. 3, pp. 525–533, Sep. 1987.

[35] J. Möucks, W. Köuhler, T. Gasser, and D. T. Pham, "Novel approaches to the problem of latency jitter," *Psychophysiology*, vol. 25, no. 2, pp. 217–226, 1988.

[36] J. J. Riera-Díaz, J. A. Carballo-González, R. Biscay-Lirio, and P. A. Valdés-Sosa, "The estimation of event related potentials affected by random shifts and scalings," *Int. J. Biomed. Comput.*, vol. 38, no. 2, pp. 109–120, Feb. 1995.

[37] L. Gupta, D. L. Molfese, R. Tammana, and P. G. Simos, "Nonlinear alignment and averaging for estimating the evoked potential," *IEEE Trans. Biomed. Eng.*, vol. 43, no. 4, pp. 348–356, Apr. 1996.

[38] R. Q. Quiroga and H. Garcia, "Single-trial event-related potentials with wavelet denoising," *Clin. Neurophysiol.*, vol. 114, no. 2, pp. 376–390, 2003.

[39] R. Li, A. Keil, and J. C. Principe, "Single-trial P300 estimation with a spatiotemporal filtering method," *J. Neurosci. Methods*, vol. 177, no. 2, pp. 488–496.

[40] C. D'Avanzo, S. Schiff, P. Amodio, and G. Sparacino, "A Bayesian method to estimate single-trial event-related potentials with application to the study of the P300 variability," *J. Neurosci. Methods*, vol. 198, no. 1, pp. 114–124, May 2011.

[41] C.-M. Wang, J.-C. Zhang, K. Yin, and L. Yao, "A multicomponent estimation method of single-trial ERPS for BCI applications," in *Proc. Int. Conf. Mach. Learn. Cybernet.*, 2008, vol. 6, pp. 3439–3444.

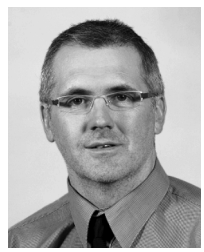
[42] W. D. Weeda, R. P. P. P. Grasman, L. J. Waldorp, M. C. van de Laar, M. W. van der Molen, and H. M. Huizenga, "A fast and reliable method for simultaneous waveform, amplitude and latency estimation of single-trial EEG/MEG data," *PLoS ONE*, vol. 7, no. 6, p. e38292, Jun. 2012.
[43] P. Jaskowski and R. Verleger, "Amplitudes and latencies of single-trial ERP's estimated by a maximum-likelihood method," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 8, pp. 987–993, Aug. 1999.
[44] P. Jaśkowski and R. Verleger, "An evaluation of methods for single-trial estimation of P3 latency," *Psychophysiology*, vol. 37, no. 2, pp. 153–162, 2000.
[45] A. Luo and P. Sajda, "Using single-trial EEG to estimate the timing of target onset during rapid serial visual presentation," in *Proc. 28th Annu. Int. Conf. IEEE EMBS*, 2006, pp. 79–82.
[46] B. J. Lance, S. E. Kerick, A. J. Ries, K. S. Oie, and K. McDowell, "Brain computer interface technologies in the coming decades," *Proc. IEEE*, vol. 100, pp. 1585–1599, May 2012.

Amar R. Marathe was born in Houston, TX, USA, on October 12, 1979. He received the B.S. degree in electrical engineering and computer sciences from the University of California, Berkeley, CA, USA, in 2001, and the M.S. and Ph.D. degrees in biomedical engineering from Case Western Reserve University, Cleveland, OH, USA, in 2008 and 2011, respectively.

Upon completion of his Ph.D. degree, he worked in the Brain Machine Interfaces Laboratory at the Cleveland Clinic, Cleveland, OH, USA, as both a Postdoctoral Research Fellow and a Senior Principal Research Engineer. He is currently a Postdoctoral Research Fellow in the Translational Neuroscience Branch at the U.S. Army Research Lab (ARL), Aberdeen Proving Grounds, MD, USA. Since joining ARL in May 2012, he has worked on detecting neural patterns associated with behavioral events from EEG through modern machine learning approaches.

Anthony J. Ries received the B.S. degree in psychology from Northwest Missouri State University, Maryville, MO, USA, in 2000, and the M.A. and Ph.D. degrees in cognitive neuroscience from the University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, in 2003 and 2007, respectively.

He is a Research Psychologist in the Translational Neuroscience Branch at the U.S. Army Research Laboratory, Aberdeen Proving Grounds, MD, USA. Before joining the Translational Neuroscience Branch, he worked as a Human Factors Engineer testing and evaluating the use of wireless EEG systems for HumanCentric Technologies, Cary, NC, USA, and for Trideum Corporation, a DoD-funded contractor. His current research focuses on the neural features of visual target recognition used in brain–computer interaction technologies.



Kaleb McDowell (SM'11) was born in Frederick, MD, USA, on July 10, 1970. He received the B.S. degree in operations research and industrial engineering from Cornell University, Ithaca, NY, USA, in 1992, and the M.S. degree in kinesiology and the Ph.D. degree in neuroscience and cognitive science from the University of Maryland, College Park, USA, in 2000 and 2003, respectively.

He is currently the Chief of the Translational Neuroscience Branch and Chair of the Neuroscience Strategic Research Area at the U.S. Army Research Laboratory (ARL), Aberdeen Proving Grounds, MD, USA. Since joining ARL as a Research Psychologist in 2003, he has contributed to over 30 reviewed publications, and has led several major research and development programs focused on neuroscience, indirect vision systems, and vehicle mobility. His current research interest focuses on translating basic neuroscience into applications for use by healthy populations in everyday, real-world environments.

Dr. McDowell received Department of Army Research and Development Achievement awards for technical excellence in 2007 and 2009 and the ARL Award for Leadership in 2011.