MIT Open Access Articles

*Classification of Imperfectly Time-Locked
Image RSVP Events with EEG Device*

The MIT Faculty has made this article openly available. **Please share**
how this access benefits you. Your story matters.

Citation: Meng, Jia, Lenis Mauricio Meriño, Kay Robbins, and Yufei Huang. "Classification of Imperfectly Time-Locked Image RSVP Events with EEG Device." *Neuroinformatics* 12, no. 2 (September 15, 2013): 261–275.

As Published: <http://dx.doi.org/10.1007/s12021-013-9203-4>

Publisher: Springer US

Persistent URL: <http://hdl.handle.net/1721.1/104077>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.

Classification of imperfectly time-locked image RSVP events with EEG device

Jia Meng · Lenis Mauricio Meriño · Kay
Robbins · Yufei Huang

Received: date / Accepted: date

Abstract Classification based on EEG data in an RSVP experiment is considered. Although the latency in neural response relative to the stimulus onset time may be more realistically considered to vary across trials due to factors such as subject fatigue and environmental distractions, it is nevertheless assumed to be time-locked to the stimulus in most of the existing work as a means to alleviate the computational complexity. We consider here a more practical scenario that allows variation in response latency and develop a rigorous statistical formulation for modeling the uncertainty within the varying latency coupled with a likelihood ratio test (LRT) for classification. The new model not only improves the EEG classification performance, but also may predict the true stimulus onset time when this information is not precisely available.

We test the proposed LRT algorithm on an EEG data set from an image RSVP experiment and show that, by admitting the latency variation, the proposed approach consistently outperforms a method that relies on perfect time-locking (AUC: 0.88 vs 0.86), especially when the stimulus onset time is not precisely

Jia Meng

Picower Institute for Learning and Memory, Massachusetts Institute of Technology
Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard
Tel.: 1-617-324-2471
E-mail: jmeng@mit.edu

Lenis Mauricio Meriño

Department of Electrical and Computer Engineering, University of Texas at San Antonio
Tel.: 1-210-458-7916
E-mail: lmauriciomb@gmail.com

Kay Robbins

Department of Computer Science, University of Texas at San Antonio
Tel.: 1-210-458-5543
E-mail: krobbs@cs.utsa.edu

Yufei Huang

Department of Electrical and Computer Engineering, University of Texas at San Antonio
Department of Biostatistics, University of Texas Health Science Center at San Antonio
Tel.: 1-210-458-6270
E-mail: yufei.huang@utsa.edu

available (AUC: 0.84 vs 0.71). Furthermore, the predicted stimulus onset times are highly enriched around the true onset time with $p\text{-value} = 5.2 \times 10^{-44}$.

Availability: The code is available at: <https://sites.google.com/site/rsvpntl/>

Keywords RSVP · Classification · time-locking · LDA · onset time prediction

1 Introduction

A brain computer interface (BCI) system provides a direct pathway for the brain to interact and even control an external device [1]. In this paper, we are concerned with the classification in a rapid serial visual presentation (RSVP) task [2–4]. RSVP specifies a process during which images (text or video) are displayed one-by-one in a fixed focal position for a subject, who is asked to search for the target images from a large collection of seemingly undesirable ones [5]. RSVP has been widely applied to improve individual reading rate, assist visual impairment and dyslexia, and study perceptual and cognitive psychology [6]. RSVP can also be used to improve the performance of sole human- or computer-based image recognition systems by reducing the human involvement in human-based systems and overcoming low classification accuracy of computer-based systems.

The decoding from EEG recording to reveal cognitive decisions in a RSVP experiment requires the extraction of signatures corresponding to neural responses, which are usually transient and time-locked to target image stimulus. These signatures are also known as event-related potentials (ERPs) or event-related spectrum perturbations (ERSPs). Compared with the noise, the neural signatures are usually very weak. The amplitude of a typical ERP is on the order of 1 to $10\mu\text{V}$, while the background EEG amplitude is on the order of $100\mu\text{V}$.

Currently, it is widely accepted that neural response is time-locked to stimulus. A well studied example is the P300 [7], which is a positive deflection in voltage with a latency (delay between stimulus and response) of roughly 300 to 600 ms elicited in decision making. Currently, the classification of EEG recordings requires perfectly time-locked events, where the stimulus onset time is precisely known, and the latency between stimulus and neural response is assumed to remain constant in different trials. Based on this assumption, classifiers for event detection, (e.g., detection of a target image in a RSVP experiment) can be trained by locating the relevant neural response relative to the stimulus onset time. However, such assumptions have significant limitations when applied to real scenarios. First, as shown in [8–10], the latency between the stimulus and its neural response is not constant across different trials of the same type of events. There are various factors that can affect the latency including a subject’s fatigue and concentration, stimulus’ strength, and environmental distractions. Presumably, in an RSVP experiment, some image clips may be more difficult to recognize than the others, and in such cases, the decision related neural response may be delayed. If a constant latency is assumed, the actual location of the neural response in EEG data would be missed. Since the optimal features cannot be extracted without the precise location of the response, the classification performance will be inevitably impaired. Second, the stimulus onset time may not be precisely available in practice. While it is possible to record this time information under laboratory conditions for RSVP, obtaining precise times for real applications is usually difficult since in practice, a

target may appear at any time (rather than at a pre-specified time point) and at any location (rather than a fixed focal position).

In response to the aforementioned limitations, the problem of classifying EEG events without perfect time-locking is considered. We propose here a statistical model that admits latency variation or imperfect stimulus onset time. Note that the latency variation and imperfect onset time are essentially two different sources of bias; however, they both contribute to the uncertainty in the neural response location and thus may be treated in the same way. The rest of the paper is organized as follows. In the section 2, the image RSVP experiment and EEG data preprocessing are introduced. In section 3, the proposed methods for modeling the latency variation and imperfect onset time are detailed. The results and concluding remarks are provided in section 4 and 5.

2 Experiment design and data preprocessing

2.1 Experiment design

The RSVP EEG recordings were obtained from [3]. During the experiment, participants are presented a series of bursts of small image clips under the RSVP paradigm. Each burst lasts for around 4.1s and consists of 49 image clips presented at a speed of 12 clips/second. Each burst may contain one target image clips, which shows an airplane that is not present in other non-target image clips. To ensure there is no interference from burst edges, the target clip is only presented after 500 ms from the onset and before 500 ms from the offset of the burst. EEG recordings were collected using BIOSEMI active view 2 system with 256 electrodes at 256 Hz sampling rate with 24-bit digitization. (Please refer to [3] for more details about the experimental design and data acquisition method.) The data set we adopted consists of 10 EEG recording sessions from 5 subjects (2 sessions per subject).

Data preprocessing was conducted using EEGLAB [11] under the MATLAB environment. After ICA transformation [12], frequency domain filtering was performed by applying 3 independent IIR Butterworth filters of order 3 including an IIR high-pass filter (2 Hz), an IIR low-pass filter (50 Hz), and an IIR band-rejection filter (40-80 Hz). Here, the high-pass filter (2 Hz) is mainly used to filter out slow artifacts, such as electrogalvanic signals and movement artifacts; while the low-pass filter (50 Hz) eliminates the high-frequency artifacts, such as electromyographic signals. The additional notch filter compensates for artifact noise caused by electrical power lines (60 Hz in the United States) [13]; however, the filtered frequency domain may contain information (gamma oscillation) that could be useful for the discrimination of target and non-target events. However, the focus of this work is to establish a method that can effectively perform classification without perfect time-locking

2.2 Time-frequency feature calculation and normalization

The target and non-target epochs were first isolated from the entire EEG sessions. As shown in Fig.1-A, a target epoch contains an EEG recording from 4s before

to 5s after the onset of a target image clips that has been correctly identified; similarly, a non-target epoch contains an EEG recording from 4s before to 5s after the onset of a random-selected non-target image clip that appears in a burst correctly identified by the test subjects. With the assumption that the target or non-target image is present at 0s of the epoch, each epoch records the EEG signal from - 4s to 5s. The raw EEG data are two dimensional electrical potentials in the space-time domain, where the spatio information naturally resides in the EEG recordings as the independent components of an ICA transformation. To capture the frequency characteristics, the Morlet wavelet [14] transformation was applied to each individual IC data separately over all epochs and the power at the following frequencies [2.0 2.6 3.3 4.3 5.6 7.2 9.3 12.0 15.5 20] Hz are calculated within [-4s, 5s] with 256 Hz sampling rate. The time-frequency transformed data were then down-sampled to 16 Hz to reduce sample dependence. The final transformed data of each epoch represents the power of an EEG recording distributed in 3 dimensions including the space (ICA components), the time, and the frequency dimension.

To eliminate cross-epoch variations, a log transformation of power was applied, and the power distribution over sampling time at each IC/frequency was then standardized. Specifically, let $y_{e,c,f,t}$ represent the log-power of the e -th epoch for the c -th IC, at the f -th frequency at time t . Then the normalized power $\hat{y}_{e,c,f,t}$ is calculated as

$$\hat{y}_{e,c,f,t} = \frac{y_{e,c,f,t} - \mu_{e,c,f}}{\sigma_{e,c,f}} \quad \text{when } -4 \leq t \leq 5 \quad (1)$$

where $\mu_{e,c,f}$ and $\sigma_{e,c,f}$ are the mean and the standard deviation of $\{y_{e,c,f,t} | -4 \leq t \leq 5\}$. For convenience, we refer to the normalized log-power as simply the power in the following.

After the preprocessing, the event-related potentials of target images are shown in Fig.1-B, from where we can see the strongest neural response comes around 0.2s-0.3s after stimulus onset, which is consistent previous study.

3 Methods

3.1 Discriminant feature identification

The discriminant features are defined as the power in a specific space, time, and frequency region that are distinct in target event and non-target event. The features are of 3 dimensions corresponding to ICA component, frequency, and time dimensions respectively, so a feature is represented by three indexes $\{c, f, t\}$.

Following the previous representation, let $\hat{y}_{e,c,f,t}$ define the normalized log-power in the e -th epoch for the c -th IC at f -th frequency at time t . Let l_e be the event type of the e -th epoch, with 1 representing the target event and 0 the non-target event. Then, the goal of identifying discriminant features is to find all the $\{c, f, t\}$'s whose powers are significantly different in target epochs vs. non-target epochs. Note that at this point we assume that the stimulus onset is fixed at time 0 in an epoch. The assumption will be relaxed in the next section. Then, the discriminant power of a single feature at $\{c, f, t\} \forall c, f, t$ is assessed by the Linear Discriminant Analysis (LDA) [15] classifier and quantified by the area under the receiver operating characteristic (ROC) curve, or AUC for convenience. Specifically, LDA assumes that the probability density functions (PDFs)

Fig. 1 A. Target and non-target epochs. The figure shows how a target or non-target epoch from the EEG recordings is isolated from the original EEG recording. A target epoch consists of the EEG signals from 4s before the target image clip onset until 5s after its onset, during which there is only one target image presented. Since a burst of images lasts only 4.1s, it takes only a part of the isolated epoch. A non-target epoch is defined similarly but devoid of target images. **B. Event-related potentials of target image clip.** The figure shows the event related potentials of the target image clips in RSVP paradigm in a 3s window. It can be seen that the neural responses last from 0.2s to 1s after the stimulus onset and peak around 0.2-0.3s, which is consistent with previous studies. The eye-related activities seem strongly correlated to the neural response. The numbers above the brain maps are latency time (ms).

of $\{\hat{y}_{e,c,f,t}|l_e = 1\}$ and $\{\hat{y}_{e,c,f,t}|l_e = 0\}$ are both normally distributed but with the different means and the same variance, or $\hat{y}_{e,c,f,t} \sim \mathcal{N}(\mu_{c,f,t,1}, \sigma_{c,f,t}^2) | l_e = 1$ and $\hat{y}_{e,c,f,t} \sim \mathcal{N}(\mu_{c,f,t,0}, \sigma_{c,f,t}^2) | l_e = 0$. Then, for some threshold constant h , the prediction of the event label for $\hat{y}_{e,c,f,t}$ is based on the following decision criterion

$$l_e = \begin{cases} 0 & \text{when } w\hat{y}_{e,c,f,t} < h \\ 1 & \text{when } w\hat{y}_{e,c,f,t} > h \end{cases}$$

where $w = (\mu_{c,f,t,1} - \mu_{c,f,t,0}) / \sigma_{c,f,t}^2$. For a binary classifier system, the ROC curve is a graphical plot of the sensitivity vs 1-specificity as its discrimination threshold h is varied, and the AUC represents the probability that a classifier would rank a randomly chosen positive event higher than a randomly chosen negative one [16].

To determine their discriminant power, the features were ranked by the AUCs and the top K most discriminant features selected as a set of discriminant features whose indices are represented by $\mathcal{D}$

$$\mathcal{D} = \{d_1, d_2, d_3, \dots, d_K\} \quad (2)$$

where $d_k = \{c_k, f_k, t_k\}$ represents the k -th most discriminant feature. In addition, among all t_k 's in $\mathcal{D}$, let t_{min} and t_{max} represent the smallest and largest t_k 's. Naturally, $\mathcal{T}_d = (t_{min}, t_{max})$ defines the time period where the discriminant features reside. The top K most discriminant features of the e -th EEG epoch can be extracted accordingly and denoted as

$$\mathbf{y}_e = [y_{e,1}, y_{e,2}, \dots, y_{e,K}]^\top \quad (3)$$

where $y_{e,k} = \hat{y}_{e,c_k,f_k,t_k}$. Furthermore, these features are assumed to follow a K -dimensional Gaussian distribution

$$\begin{cases} \mathbf{y}_e \sim \mathcal{N}(\boldsymbol{\mu}_1, \Sigma) & \text{when } l_e = 1 \\ \mathbf{y}_e \sim \mathcal{N}(\boldsymbol{\mu}_0, \Sigma) & \text{when } l_e = 0 \end{cases} \quad (4)$$

where $(\boldsymbol{\mu}_1, \Sigma)$ and $(\boldsymbol{\mu}_0, \Sigma)$ are the parameters (mean and covariance matrix) that can be estimated from the training data. This assumption (4) is essentially the same as that of Fisher's linear discriminant analysis (LDA) [15].

3.2 Modeling the cross-trial latency variation

In the previous section, when a stimulus is fixed at time 0 of an EEG epoch, a group of discriminant features were identified and the index of the k -th feature indicates that this feature locates at ICA component c_k , frequency f_k , and with the time latency between the stimulus and the k -th discriminant feature being t_k . In this section, we consider a more practical scenario where the neural response is not necessarily perfectly time-locked with the event, but rather, the latency between them may vary across different trials.

Conceivably, if the neural response on a particular trial gets delayed by $\tau \in \mathbb{R}$ compared with that of the time-locked events, then instead of appearing at time t_{min} , the true response will appear at time $(t_{min} + \tau)$. (Note τ may be negative as well, indicating a response appears earlier.) If we assume the relative locations of all the discriminant features do not change, then for an epoch, we may construct a discriminant vector $\mathbf{y}_{e,\tau}$ for any arbitrary delay τ of the neural response, such that,

$$\mathbf{y}_{e,\tau} = [y_{e,\tau,1}, y_{e,\tau,2}, \dots, y_{e,\tau,K}]^\top$$

where $y_{e,\tau,k} = \hat{y}_{e,c_k,f_k,t_k+\tau}$ is the delayed version of the original k -th discriminant feature $y_{e,k}$ in (3), and hence the vector $\mathbf{y}_{e,\tau}$ is a shifted version of the original discriminant vector $\mathbf{y}_e$ in (3). With particularly $\mathbf{y}_e = \mathbf{y}_{e,0}$. We may then naturally extend the model of $\mathbf{y}_e$ (4) to $\mathbf{y}_{e,\tau}$ as

$$\begin{cases} \mathbf{y}_{e,\tau} \sim N(\boldsymbol{\mu}_1, \Sigma) & \text{when } l_e = 1 \\ \mathbf{y}_{e,\tau} \sim N(\boldsymbol{\mu}_0, \Sigma) & \text{otherwise} \end{cases} \quad (5)$$

3.3 Classification of EEG epochs without perfect time-locking

We discuss in this section classification of target vs. non-target epochs when the neural response is not time locked and its latency τ is also unknown. We assume that the latency τ follows a uniform distribution within a small period of time

$(-w/2, w/2)$. Here, w indicates the size of the region that may contain event-related neural response. It should be large enough to capture most of the related neural response given the cross-trial latency variation, and at the same time be small enough to exclude most of random noise. It depends on the latency variation and noise level together to determine the proper choice of proper window size w for a specific experiment. For a particular trial, when the latency $\tau = -w/2$, the time period where discriminant features resides $\mathcal{T}_d = (t_{min} - w/2, t_{max} - w/2)$ and when $\tau = w/2$, $\mathcal{T}_d = (t_{min} + w/2, t_{max} + w/2)$. As a result, only the EEG samples from time $t_{min} - w/2$ to $t_{max} + w/2$ are considered for classification, from which the candidate discriminant vectors $\forall \tau \in (-w/2, w/2)$ can be extracted and denoted as $\mathbf{y}_{e,n}$ for $n = 1, \dots, N$, where $N = w/f_s + 1$ is the total number of the candidate discriminant vectors and f_s is the sampling rate. Given the fact that there is at most one target image in any epoch [3], there would be at most one candidate discriminant vector containing features for neural response, thus following the target model in (5), and the remaining would be background and follow the background model in (5). Let $\mathbf{y}_{e,1:N}$ be the collection of all N candidate vectors and let the s -th candidate vector $\mathbf{y}_{e,s}$ be the one corresponding to the real neural response, then s follows the uniform distribution on $[1, N]$ or $p(s = n) = 1/N \forall n$. To perform the classification, we proposed a likelihood ratio test (LRT) where the whole-epoch likelihood ratio R is calculated as

$$R = \frac{p(\mathbf{y}_{e,1:N} | l_e = 1)}{p(\mathbf{y}_{e,1:N} | l_e = 0)} \quad (6)$$

$$= \frac{\sum_{n=1}^N p(s = n) p(\mathbf{y}_{e,1:N} | l_e = 1, s = n)}{p(\mathbf{y}_{e,1:N} | l_e = 0)} \quad (7)$$

$$= \frac{\sum_{n=1}^N p(s = n) N(\mathbf{y}_{e,s} | \boldsymbol{\mu}_1, \Sigma) \prod_{i=1, i \neq n}^N N(\mathbf{y}_{e,i} | \boldsymbol{\mu}_0, \Sigma)}{\prod_{n=1}^N N(\mathbf{y}_{e,n} | \boldsymbol{\mu}_0, \Sigma)} \quad (8)$$

$$= \frac{1}{N} \sum_{s=1}^N \frac{N(\mathbf{y}_{e,s} | \boldsymbol{\mu}_1, \Sigma)}{N(\mathbf{y}_{e,s} | \boldsymbol{\mu}_0, \Sigma)} \quad (9)$$

$$= \frac{1}{N} \sum_{s=1}^N r_{e,s} \quad (10)$$

$$= \mathbb{E}(r_e) \quad (11)$$

where the position-specific likelihood $r_{e,s}$ is defined as

$$r_{e,s} = \frac{N(\mathbf{y}_{e,s} | \boldsymbol{\mu}_1, \Sigma_1)}{N(\mathbf{y}_{e,s} | \boldsymbol{\mu}_0, \Sigma_0)} \quad (12)$$

which represents the ratio for a specific choice of the latency s for epoch e . The result in (11) implies that the likelihood ratio of the entire epoch is equal to the average likelihood ratio of all the possible position-specific likelihood ratios for $\tau \in (-w/2, w/2)$, which makes sense intuitively. To predict whether the EEG epoch contains a target, the epoch log likelihood ratio R (11) is compared with a threshold, say 1 and a target epoch is predicted when $R > 1$. Note that this

proposed classifier collapses into the LDA when the latency variation $w = 0$, or there is only one candidate $N = 1$; in another word, LDA is a special case of this LRT classifier when no latency variation is assumed.

3.4 Prediction of response latency and stimulus onset time

As a side product of the LRT calculation, $r_{e,s} \forall s$ are indicative of the subject's neural response latency τ , when the stimulus onset time is perfectly time-locked at 0. Particularly, the estimated latency $\hat{\tau} = \{s | r_{e,s} > r_{e,i} \forall i \in \{1, 2, \dots, N\} \text{ and } s \neq i\}$, the time whose corresponding position specific likelihood ratio $r_{e,s}$ is the largest.

Under the RSVP paradigm, it is reasonable to assume the precise stimulus onset time is known, which is the same as target onset time. However, when the stimulus onset time is not precisely available as in many practical application, it can be estimated as $\hat{\tau}$, given there is no response latency. Knowing stimulus onset time is of interest in many applications. For instance, the model in (5) explains that, given that the target image onset is at 0s, the shifted neural response is due to cross-trial latency variation. However, when the stimulus onset time is not available and can vary in a large range, we can predict the stimulus onset time with the detected neural response.

4 Results

4.1 The discriminant features of RSVP events

In this section, we examine the identified discriminant features that may help to classify target vs. non-target epochs. After ICA and wavelet decompositions, the feature space is 3 dimensional with independent component (IC), time and frequency. The discriminant power of each individual feature at $\{c, f, t\}$ was estimated using the area under the curve (AUC) of an LDA classifier as detailed in section 3.

We then compared the discriminant power before (-2s to 0s) and after target image onset (0s to 2s). Ideally, the features before target image onset should follow the background distribution, representing the distribution when there is no target. The power distribution of those after image onset should span the region with higher discriminant power. We plotted the histograms of the features before and after target image onset in Fig.2, and as expected, the discriminant power after target onset is generally larger than that of before onset for most of the 10 tested EEG sessions. Please note that, since the proposed method assumes the stimulus onset time is unknown, the normalization procedure in equation (1) actually subtracts different amounts of baseline power in target and non-target epochs, which makes the resulting discriminant power relative to its epoch. As a result, some of the features before 0s can have large discriminative power but might not be truly discriminant. This is different from the perfect time locking case where the actual stimulus onset time is known so that the baseline stimulus was estimated using EEG signals before the stimulus onset [3, 17]. Given the fact that the features before 0s are not discriminant, the selected discriminant features are those whose

discriminant powers are larger than any of the features before 0s and correspond to a p-value $< 10^{-5}$.

Fig. 2 Discriminant power before and after the onset of target image clips. The figure shows the histograms of the discriminant power (AUC) of features before and after the onset of target images on the 10 tested EEG sessions from 5 subjects. Here, S2-1 indicates the 1st EEG session of the 2nd subject. In a causal system, the discriminant power before 0s should ideally be random and thus can be used as a reference. It can be seen from the figure that, the features after 0s have larger discriminant power on subject 1, 2, 5, whereas on subject 3 and 4 the two are similar, indicating that the proposed log-power features may perform better on subject 1, 2 and 5.

The largest AUC of each individual IC is shown in Fig.3-A. The ICA decomposition results in 256 ICs and the majority of them do not have significant discriminant power. The 4 most discriminant ICs (IC: 179, 183, 199, 215), as shown in Fig.3-B, are localized around eyes, which can be related to eye activities or Electrooculography (EOG) signals. This results suggests the EEG signals near the eyes are closely related to neural response in RSVP paradigm. This finding is also consistent with the previous studies [3, 17]. To investigate the distribution of the discriminant features ($p\text{-value} < 10^{-5}$) in time-frequency, we plotted the discriminant features in a 2-dimensional histogram in Fig.3-C. Two major response patterns that can be visually identified are a Theta band power increase around 300-500 ms after target image onset and a Delta band power increase after 1000 ms. This is again consistent with a previous study [17]. Please note that, since the proposed method assumes the stimulus onset time is unknown, the normalization procedure in equation (1) actually subtracts different amount of base line power in target and non-target epochs, which will lead to the features before 0s to be actually discriminative, meaning the most discriminative features should have p-values that are more significant than the estimated 10^{-5} . This is different from the perfect time locking case where the actual stimulus onset time is known, so the baseline stimulus was estimated using EEG signals before the stimulus onset [3, 17].

4.2 Exploration of cross-trial latency variation

In this section, we demonstrate the latency variation by examining the position-specific log likelihood ratio $\log(r_{e,s})$. As can be seen from Fig. 4, in a target epoch, the spike of the average $\log(r_{e,s})$ appears around the target image onset time (Fig. 4-A) with a slight increase as the number of the features used for classification increases from 1 to 16.

In contrast, in a non-target epoch, the average $\log(r_{e,s})$ stays almost flat (Fig.4-B), in dependent of the number of features used in the model. For individual EEG subjects/sessions (Fig.4-C, -D), the spike in target epochs is more pronounced for subjects 1, 2 and 5 (Fig.4-C), which is consistent with previous result (Fig.2). The distributions of the largest $r_{e,s} \forall s$ are further shown in Fig. 5 for different combinations of top features. It can be seen that, in a target epoch, the largest ratios mainly congregate around the target image onset time, as opposed to the rather dispersed distribution patterns in non-target epochs. However, they did not align always at time 0, indicating inherent variations in response latency. Taken together, this result suggest that the latency between the stimulus onset and neural responses is not constant and should be considered in classification. Note that the observed patterns are not as strong for subject 3 and 4 since the discriminant features were not as distinct for these two subjects.

To further understand whether the differences in discrimination power among the 5 subjects we examined the percentage of correct decisions by the 5 subjects in each session. As can be seen in Fig. 6, there is no significance difference in performance among the subjects. This fact indicates the difference in discrimination ability among the 5 subjects is not due to the difference in individual cognitive ability.

Fig. 3 A. Discriminant power of ICs. The ICA decomposition resulted in 256 Independent Components (ICs). This figure shows the largest AUC of all the features corresponding to a particular IC. It can be seen that some of them are apparently more discriminant than the others. The most discriminant ICs are IC: 179, 183, 199 and 215. Some features on these ICs can be highly discriminant against target and non-target epochs with AUC larger than 0.75. **B. The scalp map projection of the most discriminant ICs.** Figure shows 4 most discriminant ICs: 179, 183, 199 and 215. They are all apparently related to eye activities, indicating the EEG signal caused by eye activities are closely related to neural response under an image RSVP paradigm. Note that the figure was plotted with EEGLab [11] and the projection scale uses arbitrary units, indicating from strong positive correlation (red) to strong negative correlation (blue), and the channel locations below head center are shown as a “skirt” region outside the head. **C. Time-frequency locations of discriminant features.** The distribution of the discriminant features (p -value $< 10^{-5}$) in time-frequency are shown, where the white color indicates the time-frequency regions that are enriched with most discriminant features among the 256 ICs of the 10 subjects. Two major response patterns that can be visually identified are a Theta band power increase around 300-500 ms after target image onset, and a Delta band power increase after 1000 ms. This is consistent with a previous study [17].

4.3 Prediction of stimulus onset time

In this section, we investigate the prediction of the stimulus onset time given a target EEG epoch. In this case, we assumed that there was no response latency, which however might not be true for all trials. As a result, we regard the prediction successful as long as the difference between predicted and the true onset time (0) is less than 0.2s in a tested epoch. Note that because of the large variability of latency across trials [10], it is possible that the actual onset time can be delayed by more than 0.2s in some cases.

For comparison, we also tried to predict a target onset time on the non-target epochs. Since there is no target image in a non-target epoch, the predictions should be random and can also serve as a reference. When using different numbers of top features (K), the average successful prediction rates on all the 10 sessions is shown in Fig.7-A,B. It can be seen that, consistent with previous results, the predictions were better for Subjects 1, 2 and 5. We also showed the median and

Fig. 4 The average position-specific log likelihood ratio. The figures show the average $\log(r_{e,s})$ within an epoch. K represents the number of the top discriminant features used by the LRT algorithm. **A** In target epochs, the spike of the average $\log(r_{e,s})$ appears around the target image onset time (**A**). **B** In non-target epochs, the average $\log(r_{e,s})$ stays almost flat (**B**), regardless of the number of features used. **C**, **D** When we look at individual EEG subject/data, the spike in target epoch is more distinct for subject 1, 2 and 5 as opposed to subject 3 and 4 (**C**), which is consistent with the result in Fig.2.

mean performance in Fig.7-C. The proposed approach can successfully predict the onset time ($< 0.2s$) for more than 45% of the target image clips. This is a considerable improvement over that obtained by random (less than 10%). The best performance is obtained when using 36 features, indicating that the optimal number of features for predicting stimulus' onset time is 36. We then took a closer look at the prediction performance when using 36 top features. A histogram of the predicted target images onset time is shown in Fig.7-D,E. The predicted onset times mostly coincide with the true onset times. The regions larger than 0.2s or smaller than -0.2s indicate the prediction errors.

4.4 Classification of EEG signals without perfect time-locking

We showed that the latency between stimulus and neural responses may not always be constant (Fig.4, 5 and 7-D). Yet, many previous methods assume perfect time-

Fig. 5 Distribution of the largest $r_{e,s}$. The figure shows distribution of the largest $\log(r_{e,s})$ within an epoch when using top K discriminant features ($K = \{2, 4, 8, 16, 32, 64\}$). The largest log ratios mainly congregate around the target image onset time in a target epoch (labeled as “T”), as opposed to the dispersed distribution in non-target epochs (labeled as “N”). Again, this pattern is not as strong for subject 3 and 4.

Fig. 6 Percentage of correct decisions by the 5 subjects in the RSVP experiment. The figure shows that there is no significant difference between the group with strong discriminant features (subject 1, 2 and 5) and weak discriminant features (subject 3 and 4).

locking [3, 17]. Intuitively, classification performance could be improved if latency can be correctly addressed. In this section, we examine the performance of our proposed LRT test and investigate its ability to classify target images without perfect time-locking.

The latency delay τ is assumed to vary within a small time period $[-w/2, w/2]$ and follows a uniform distribution. We first tested the classification performance under various setting of w using 36 features, which can best capture the response latency (Fig.7-C). As expected (Fig.8-A,B), perfect time-locking ($w = 0$) does not give the best classification performance. In contrast, the best result based on the 36 most discriminant features is achieved when $w = 0.25$ s, and the performance degrades as w increases over more than 0.5s. The performance of the proposed LRT model was also compared with that of LDA with a perfect time-locking, which is equivalent to LRT with $w = 0$. Fig.8-C,D demonstrates the A_z performance of

Fig. 7 A-B. Successful prediction of target image onset time on 10 sessions In this test, we consider the prediction is successful as long as the difference between predicted time and true onset time is less than or equal to 0.2s. (A) shows the successful prediction of target image onset time on 10 EEG sessions when using different number of features. Figure (B) shows essentially the same metrics on non-target epochs, and for reference. Better performance is obtained on subjects 1,2 and 5, whose data have previously been shown to have more discriminant features compared with subject 3 and 4. It also can be seen from this figure that, the optimal number of features for target onset time prediction is different across 10 tested sessions, from less than 10 features on S3-1 to around 60 features on S1-1. **C. Prediction of target image onset time versus the number of features.** The best performance is obtained when using 36 features (45%). This is a considerable improvement over the results obtained by random (less than 10%). **D-E. Distribution of predicted target image onset times** Figure D shows the histogram of the predicted target image onset times. The predicted onset time are mostly coincide around the true onset time. The regions larger than 0.2 or smaller than -0.2 indicate the prediction errors. The cumulative distribution of the predicted target onset time is shown in E, where we can see that a drastic increase happens around the true onset time (0s).

LRT and LDA with different number of top features. With about 40 features, both LRT and LDA achieve the irrespective best AUCs but LRT outperforms LDA with the best AUC of 0.88 versus 0.86 for LDA.

4.5 Beyond RSVP: imperfect and unknown stimulus onset time

Under the RSVP paradigm, the stimulus onset time is considered to coincide with target image onset time due to the fixed focal position and virtually no disturbance. However, in more realistic BCI applications such as target detection in complex satellite imagery, imperfect or even no knowledge about stimulus onset is available. For the classification task in these more realistic scenarios, the algorithms assuming perfect time locking will fail. On the contrary, the proposed LRT algorithm can be employed since it does not need the precise knowledge of the stimulus onset except a loose time window of width w , in which the stimulus onset would occur. Note that in theory, no information about the stimulus onset time is required to implement

Fig. 8 A-B. Classification performance vs. assumed latency variation (w). The figure shows the classification performance versus different assumed latency variation windows, w . The LRT algorithm used the top 36 features as suggested by (Fig.??) with a 10-fold cross validation. Apparently, assuming perfect time-locking $w = 0$ s does not lead to the best classification performance. In contrast, at $w = 0.25$ s, LRT produces the optimal result. Note that the performance degrades significantly as w goes more than 0.5s, which is consistent with Fig.??, and suggests that the latency variation is rarely larger than 0.5s. **A** shows the mean and median performance among the 10 EEG sessions, **B** shows the performances of the individual sessions. **C-D. Comparison of LDA and LRT.** LDA assumes perfect time-locking, which is equivalent to the proposed LRT algorithm for $w = 0$. **C** shows the AUC of LDA and LRT ($w=0.25$) using different numbers of discriminant features and the best performances are (AUC: 0.88 vs 0.86), which were obtained from a 10-fold cross-validation. It can be seen from **D** that, LRT ($w=0.25$) outperforms LDA in performance regardless of the number of features used.

the proposed LRT; the time window is provided to make the computation more efficient and to improve the signal to noise ratio.

To demonstrate the ability of LRT to classify when the precise stimulus onset time is unknown, we intentionally shifted the EEG epochs by a random $t \in [-0.5s, 0.5s]$. This experiment imitates the scenario where error exists in the given onset time and follows a uniform distribution within the period $[-0.5s, 0.5s]$. Fig.9 examines the relationship between the window w and the LRT's classification performance. It is not surprising to see that the optimal time window size becomes much larger than that under perfect stimulus onset information. The optimal performance is reached at $w = 0.8$ (s); this increase accounts for roughly the originally 0.25s perfect window size under perfect stimulus onset time and 0.5s variation in error. The result is highly consistent across all the 10 tested sessions. Note that there appears to be no significant decrease in the mean AUC (0.86) as compared with that for noiseless stimulus onset time (AUC =0.88; Fig.8-C).

Fig. 9 Optimal LRT window with imperfect onset time. After imposing the additional perturbation, the optimal window in LRT model becomes much larger compared with Fig.?? , and the optimal performance is reached with $w = 1.0s$. This result is consistent across all the 10 tested sessions. The window $w = 1.0s$ accounts for the cross-trial latency variation (around 0.25s) and the noise in imperfect onset time (1.0s) .

In contrast, the performance of the perfect time-locking LDA, which wrongly assumes 0 being the stimulus onset, degraded significantly in all the tested EEG sessions (Fig. 10-A). Further investigation of the largest $\log r_{e,s}$ values in different sessions revealed a consistent sharp spike near the true target image onset time (Fig. 10-B), and the largest $\log r_{e,s}$ are enriched around the target onset time in target epochs when compared with non-target epoch (Fig.10- C). Taken together, these results confirm that the proposed LRT is robust against both cross-trial latency variation as well as imperfect stimulus onset time, scenarios more related with real applications.

4.6 Impact of Eye-Related Activities under Image RSVP Paradigm

As shown previously in Fig.3, many of the identified discriminant ICs are highly related to the regions surrounding the eyes, which may be related to eye activities. To assess the impact of eye-related features on classification, we manually identified a number of EEG channels that are physically located near the eyes (Fig.11-A), and calculated the normalized weight of these eye-related channels among all the channels, based on which we can rank the closeness of the ICs to eye-related activities (Fig.11-B,C). Consistent with Fig.3, we can see the most eye-related ICs have larger discriminant power than thoes least eye-related ICs (Fig.11-D). Next, we discarded the top 20 most eye-related ICs, and applied our approach to predict on-set time as well as target vs. non-target epochs as previously. It can be seen from Fig.11 that, while the general trends stay the same as previously, the classification performance of LRT was improved by filtering out those ICs that are likely associated with eye-related activities. In contrast, the target onset time predication did not benefit from this filtering; the performance degraded somewhat as a result of this filtering (Fig. 11-F). This observation indicates that under the image RSVP paradigm, the eye-related activities contain both information that can help precisely locate target image onset time, and noise that can impair the classification performance of the BCI systems. Therefore, removing those eye-related ICs can promote event classification but impairs target onset time prediction.

Fig. 10 Comparison of LDA and LRT with imperfect onset time: **A.** When using $w = 1s$, the performance of LRT significantly outperforms LDA on all the tested EEG sessions despite the number of features used. **B.** The average $\log r_{e,s}$ has a sharp spike near the true target image onset time. **C.** The blue curve shows the percentage of trials whose target onset times are successfully predicted subject to a tolerance setting, and the green curve shows the random performance obtained from non-target epochs. It can be seen that, the predicted target onset time is more enriched around the true target onset time in target epochs than in non-target epochs, especially for subjects 1,2 and 5.

5 Conclusion and Discussion

In this paper, we considered the EEG classification for RSVP experiments. Instead of assuming a perfect time-locking between the stimulus and the neural response as does most of the existing work, we considered a more practical scenario where the latency between them can vary and further extended to the scenario with imperfect stimulus onset time. We developed a rigorous statistical Gaussian mixture model

for modeling the uncertainty with the varying latency and proposed a likelihood ratio test (LRT) for classification of event types and predict the stimulus onset time. By admitting the latency variation and imperfect onset time, the proposed LRT algorithm not only improves the classification performance over those based on perfect time-locking, but can also predict the response latency for individual subjects.

The major contributions of this work include:

- A Gaussian mixture model was proposed to address the latency variation between the stimulus and neural response. Traditionally, it is assumed that neural response is perfectly time-locked with the stimulus onset time. However, this is not always true in practice because of fatigue, stimulus strength, distraction, etc. We demonstrated the existence of latency variations within the RSVP paradigm (Fig.4) and showed that, with the non-perfect time-locking assumptions, our LRT algorithm can improve the classification performance over the perfect time-locking LDA classifier.
- The proposed LRT algorithm can also be used to predict the response latency. Although this task is especially difficult given the extreme low signal to noise ratio, we showed that among the 5 tested subjects, the LRT was able to locate more than 40% target latency precisely on 3 subjects and it even achieves more than 80% success on the 2nd session of subject 5. The performance is well above that performance by random selection (less than 10%).
- The proposed LRT algorithm is also robust to noisy information regarding the stimulus onset time. As a result, the proposed LRT algorithm has advantages over the conventional LDA approach for practical applications because precise stimulus onset time is rather difficult to obtain in practice.

The advantages notwithstanding, limitations still exist in the proposed LRT algorithms, which need to be further investigated as part of our future work.

- In our LRT model, we assume the neural response latency τ follows a uniform distribution within a short period of time $[-w/2, w/2]$. However, based on Fig.4, 5 and 7-C, τ might more closely follow a Gaussian or Laplacian distribution. Although the extension to different distributions should be fairly straightforward under the framework of LRT, such investigation can shed light on the specific behavior of human cognition under different conditions.
- The proposed model did not address the distortion within the feature space due to cross-trial latency variation. Our model assumes the relative location of discriminant features will stay the same. However, a more suitable assumption would be that there exist several groups of discriminant features corresponding to several components of the neural response elicited in RSVP target events; while the relative locations of two groups may change, the relative locations of features within the same group should stay the same. Such assumptions may naturally rely on the part-based object detection [18, 19], i.e., each group of discriminant features are considered a part of the whole neural response; And the previous work [17, 20] in identifying discriminant components in RSVP paradigm may serve as its foundation.
- The computation time is relatively high for practical application. In addition to the already high computations associated the number of channels (256), sampling frequencies (256 Hz), resolution of time-frequency analysis (20 frequencies) and discriminant feature identification, evaluation of position specific

LRT for all possible positions impose more computational burden. It would be beneficial to explore effective feature dimension reduction approaches for this application.

Acknowledgements This project was supported by the Army Research Laboratory CAN-CTA initiative and the National Institute on Minority Health and Health Disparities (G12MD007591) from the National Institutes of Health.

References

1. J. Wolpaw, N. Birbaumer, W. Heetderks, D. McFarland, P. Peckham, G. Schalk, E. Donchin, L. Quatrano, C. Robinson, T. Vaughan, IEEE Transactions on Rehabilitation Engineering **8**(2), 164 (2000)
2. P. Sajda, A. Gerson, M. Philiastides, L. Parra, Towards brain-computer interfacing
3. N. Bigdely-Shamlo, A. Vankov, R. Ramirez, S. Makeig, IEEE Transactions on Neural Systems and Rehabilitation Engineering **16**(5), 432 (2008)
4. K. Hild, M. Pavel, D. Erdogmus, S. Mathan, in *Signals, Systems and Computers, 2009 Conference Record of the Forty-Third IEEE Conference on Asilomar* (2009), pp. 51–54
5. C. Eriksen, T. Spencer, Journal of Experimental Psychology **79**(2p2), 1 (1969)
6. J. Vidal, Annual review of Biophysics and Bioengineering **2**(1), 157 (1973)
7. S. Sutton, M. Braren, J. Zubin, E. John, Science **150**(3700), 1187 (1965)
8. A. Gerson, L. Parra, P. Sajda, Neuroimage **28**(2), 342 (2005)
9. A. Luo, P. Sajda, Frontiers in human neuroscience **3** (2009)
10. R. VanRullen, N. Busch, J. Drewes, J. Dubois, Frontiers in psychology **2** (2011)
11. A. Delorme, S. Makeig, Journal of neuroscience methods **134**(1), 9 (2004)
12. S. Makeig, A. Bell, T. Jung, T. Sejnowski, et al., Advances in neural information processing systems pp. 145–151 (1996)
13. E. Niedermeyer, F. Da Silva, *Electroencephalography: Basic principles, clinical applications, and related fields* (Lippincott Williams & Wilkins, 2005)
14. P. Goupillaud, A. Grossmann, J. Morlet, Geophysical Journal International **23**(1), 85 (1984)
15. R. Fisher, Annals of Human Genetics **7**(2), 179 (1936)
16. T. Fawcett, Pattern recognition letters **27**(8), 861 (2006)
17. J. Meng, L.M. Merio, N.B. Shamlo, S. Makeig, K. Robbins, Y. Huang, PLoS ONE **7**(9), e44464 (2012). DOI 10.1371/journal.pone.0044464
18. P. Felzenszwalb, D. Huttenlocher, International Journal of Computer Vision **61**(1), 55 (2005)
19. P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan, IEEE Transactions on Pattern Analysis and Machine Intelligence **32**(9), 1627 (2010)
20. M. Philiastides, R. Ratcliff, P. Sajda, The Journal of neuroscience **26**(35), 8965 (2006)

Fig. 11 Filtering Eye-Related Activities: **A.** We manually identified a number of channels that are physically located near the eyes, which are highlighted in red font, i.e., channel: [1:2, 234:254]. **B-C.** The figures show 9 ICs that are most/least related eye-activities by calculating the relative contribution of the channels surrounding the eyes in all the 256 EEG channels after the ICA decomposition. **D.** This is a similar test as that of Fig.9A. After filtering out the eye-related ICs, the optimal window in LRT model in case of the additional perturbation is still around 1s. Compared with using all the ICs, filtering eye-related ICs can clearly improve the classification performance regardless of the time window size. **E.** This is a similar test as that of Fig.10A. When using $w = 1$ s, filtering the eye-related ICs improves the performance of LRT but degrades that of LDA. **F-G.** These are similar test as those of Fig.10-C,D but with eye-related ICs filtered out. It can be seen that filtering out eye-related ICs slightly impairs the prediction of target image onset time.