Online Prediction of Driver Distraction Based on Brain Activity Patterns

Shouyi Wang, Yiqi Zhang, Changxu Wu, Felix Darvas, Wanpracha Art Chaovalitwongse

***Abstract*—This paper presents a new computational framework for early detection of driver distractions (map-viewing) using brain activity measured by electroencephalographic (EEG) signals. Compared to most studies in the literature, which are mainly focused on classification of distracted and non-distracted periods, this study proposes a new framework to prospectively predict the start and the end of a distraction period, defined by map-viewing. The proposed prediction algorithm was tested on a dataset of continuous EEG signals recorded from 24 subjects. During the EEG recordings, the subjects were asked to drive from an initial position to a destination using a city map in a simulated driving environment. The overall accuracies for the prediction of the start and the end of map-viewing were 81% and 70%, respectively. The experimental results demonstrated that the proposed algorithm can predict the start and end of map-viewing with a relatively high accuracy and can be generalized to individual subjects. The outcome of this study has a great potential to improve the design of future intelligent navigation systems. Prediction of the start of map-viewing can be used to provide route information based on a driver's needs and consequently avoid map-viewing activities. Prediction of the end of map-viewing can be used to provide warnings for potential long map-viewing durations. Further development of the proposed framework and its applications in driver distraction predictions are also discussed.**

***Index Terms*—Online adaptive predictions, driver distraction prediction, time series pattern recognition, EEG**

Shouyi Wang is with the Department of Industrial & Systems Engineering, University of Washington, Seattle, WA 98195, USA (E-mail: shouyi@uw.edu);

Yiqi Zhang is with Department of Industrial & Systems Engineering, the State University of New York (SUNY)-Buffalo, Buffalo, NY 14260, USA (E-mail:yiqizhan@buffalo.edu);

Changxu Wu is with Department of Industrial & Systems Engineering, the State University of New York (SUNY)-Buffalo, Buffalo, NY 14260, USA (E-mail: seanwu@buffalo.edu);

Felix Darvas is with Department of Neurological Surgery, University of Washington, Seattle, WA 98195, USA (E-mail: fdarvas@u.washington.edu);

Wanpracha Art Chaovalitwongse is with the Department of Industrial & Systems Engineering and the Department of Radiology, University of Washington, Seattle, WA 98195, USA. (E-mail: artchao@uw.edu).

I. Introduction

Distracted driving is one of the main causes of vehicle crashes. According to the statistics released by National Highway Traffic Safety Administration (NHTSA), 3,092 people were killed, and 416,000 people were injured in vehicle crashes involving distracted drivers in 2010 [1]. Driver distraction is defined as a form of inattention which "delayed the recognition of information needed to accomplish the driving task safely because some event, activity, object, or person within or outside the vehicle compels or induces the driver's shifting attention away from the driving task" [2]. In particular, with the wide application of electronic route navigation systems, the navigation-map-viewing behavior becomes an important source of driving distraction and vehicle accidents. The serious safety issue has directed many researchers' attention to distracted driving performance that is associated with map-viewing behaviors and navigation systems. The corresponding human factors research of such systems is believed to contribute to the development of safe, usable and acceptable assist systems to vehicle customers.

Navigation-map-viewing is a complex task, requiring the concurrent execution of various visual, motor, and cognitive skills, in addition to the normal driving task. Although map-viewing behavior is only a secondary task in driving, this activity is closely linked to the primary task of driving. A great number of studies have demonstrated that taking off eyes from road could result in driving performance decrement and raise significant safety issues [3], [4], [5]. Even when drivers have their eyes on the road, the cognitive distraction associated with in-vehicle devices can also have negative effects on driving performance [6], [7], [8], [9], [10]. This is because drivers may utilize too much cognitive capacity to process frequently received navigation information, and thus pay less attention to the driving task and road conditions. Therefore, both visual distraction (e.g. map-viewing) and cognitive distraction could lower driving performance and lead to dangerous situations [11], [12]. Moreover, another big problem of using the current navigation systems is that frequent redundant information may be provided in some periods. The navigation information provided for route assistance purpose may cause driving annoyance instead [13],

[14]. A driver would feel annoyed since processing such redundant information takes significant resources from the already limited cognition capacity for the primary driving task.

To reduce the risk of driving distractions and avoid dangerous situations, eye movement tracking, video camera recognition, and lane keeping performance are popular techniques employed in current driver distraction studies [28, 35]. These methods have been proved to be effective on classifying distraction and non-distraction periods. However, they have limitations in detecting navigation-related distractions in real-time. In recent years, many researchers have begun to use electroencephalographic (EEG) signals to study secondary tasks during driving [15], [16], [17]. EEG-based driving distraction studies offer a unique capability of real-time assessment of cognitive effort, engagement and workload through quantitative analysis of continuous EEG signals. In addition, EEG signals are generally unaffected by driving conditions and environments. Studies showed eye movement tracking techniques might become unstable in some driving environments [18]. For example, studies using eye trackers require an environment with dim illumination and low sunlight to minimize the glare and reflection [19]. An environment with strong glare and reflection could deteriorate the performance of eye movement tracking seriously [20]. Another advantage of EEG is its anonymous data in protecting the privacy for drivers. The methods using eye movement tracking and video recognition are hard to avoid the privacy issue of leaking drivers' personal information, such as faces, expressions, or even conversations. In contrast, EEG signals only detect a driver's electronic brainwaves without recording any other personal information. In addition, with recent advances of EEG technologies, the conventional wet EEG electrodes requiring skin preparation and conduction gels may be replaced by wireless dry electrodes, which enable remote acquisition of continuous EEG data conveniently [21], [22], [23]. Thus, EEG technologies provide a practical approach to study driver behaviors and handle driving distractions in the real world settings.

Table I. The survey of driving distraction data collection methods

Data collection method	**Studies**	**The period studied (Classification/Prediction)**
Eye movement tracking	Chisholm, Caird, & Lockhart, 2008; Donmez, Boyle, & Lee, 2007, 2008, 2010; Garay-Vega et al., 2010; Kaber, Liang, Zhang, Rogers, & Gangakhedkar, 2012; Liang & Lee, 2010; Metz, Schömig, & Krüger, 2011; Reyes & Lee, 2008; K. L. Young, Mitsopoulos-Rubens, Rudin-Brown, & Lenné, 2012; H. Zhang, Smith, & Witt, 2006; Y. Zhang et al., 2013 [3], [4], [5], [7], [9], [18], [19], [24], [25], [26], [27], [28]	Classification of distracted and non-distracted period
Video camera recognition	Stutts et al., 2005; Wege, Will, & Victor, 2012 [29], [30]	Classification of distracted and non-distracted period
Lane-keeping Performance	Alm & Nilsson, 1995; Greenberg, Tijerina, & Curry, 2003; Reed & Green, 1999; Young, Lenné, & Williamson, 2011 [31], [32], [33], [34], [35]	Classification of distracted and non-distracted period
EEG	C. Lin, Ko, & Shen, 2009; C.-T. Lin, Chen, Chiu, Lin, & Ko, 2011; Mouloua, Ahern, & Quevedo, 2012; Sonnleitner, Simon, Kincses, Buchner, & Schrauf, 2012 [16], [36], [37], [38]	Classification of distracted and non-distracted period
	Current work	Prediction of the start (Event I) and the end (Event II) of distracted period

Table I summarizes the four main groups in current driving distraction studies. It is found that current studies are mostly focused on the retrospective classification of distracted periods and non-distracted periods. There have been very few studies focusing on real-time, prospective prediction of driving distractions. With the help of EEG technologies, it may be

possible to move from retrospective off-line analyses to distinguish distraction and non-distraction periods to prospective online analyses to predict the initiation and the end of a distraction period. This study is aimed to address this transition and provide a new framework for online detection of map-viewing distractions using continuous EEG signals. To be more specific, two map-viewing actions (events) are studied: Event I when the driver is starting to look at a map and Event II when the driver is starting to look back at the road. In our experiments, drivers were asked to arrive at a predefined destination in an unfamiliar environment using a city map. The task of navigating in an unfamiliar environment was employed in this study because it is cognitive demanding and well suited to study driver distractions [39].

The proposed framework was designed to predict the start and the end of map-viewing periods through online monitoring of EEG recordings. Once the prediction of Event I is achieved, future navigation systems can subsequently provide verbal route instructions in advance. Driver's view will then stay on the road rather than looking at the map. As a result, the navigation system is able to prevent the incoming map-viewing actions rather than letting drivers read the map themselves. Likewise, by predicting the time when the driver finishes a map-viewing behavior (Event II), the navigation system can provide a warning to the driver when the predicted map-viewing duration is longer than a safe threshold. The proposed real-time prediction framework has a great potential to improve the design of future route assist systems, including e-map and other intelligent navigation systems. The capabilities of providing route information based on a driver's needs and warnings to potential long map-viewing actions would greatly improve the efficiency of a navigation system with reduced annoyance and enhanced safety.

The rest of the paper is organized as follows. The driving experiment is discussed in Section II. In Section III, the online prediction framework is presented, including feature extraction, feature selection, adaptive prediction scheme, and the evaluation metrics of prediction performance. The computational experiments are provided in Section IV. A comprehensive discussion of the impacts of this study is presented in Section V, and finally we conclude this study in Section VI.

II. Simulated Driving Experiment

A. Participant

Twenty-four participants (14 male and 10 female) took part in the current experiment with an average age of 23.3 (SD = 2.77). All the participants have normal or corrected-to-normal vision and valid driving license. They are also free of psychiatric or neurological disorders to limit potential confounds on the behavioral/cognitive aspects of driving performance and navigation activity.

Fig. 1. The experiment setup using a STISIM® driving simulator and a Neuroscan system including one Quik-Cap, Nuamps Express and Scan software to record and analyze EEG signals during driving (map for navigation is not included in this picture).

B. Apparatus

The driving task was completed using a STISIM® driving simulator (STISIMDRIVE M100K, Systems Technology Inc., Hawthorne, CA; See Figure 1). The driving simulator consists of a Logitech Momo® steering wheel with force feedback, a gas and a brake pedal (Longitech Inc., Fremont, CA). The driving scenario was presented on a 27-inch LCD with 1920×1200 pixels resolution.

An 8.5×6 inches map with designed route used for the map-viewing task was shown by a 19-inch Dell LCD display (1098FP model), which was 50cm from the right hand of subjects and 91cm from their eyes (See Figure 3). The visual angle of the touch screen was 13.1 degree. The screen was controlled by a Dell PC (OPTIPLEX 745) and connected with the driving simulator via a Labjack® system.

A Neuroscan system including one Quik-Cap, Nuamps Express and Scan software was used to record and analyze EEG in the study. The Nuamps Express is a 40 channel digital EEG and ERP recording system. There are four electrodes that were used for measuring eye movements to remove muscular artifacts. The rest 36 electrodes were mounted on the scalp and thus used for analyses in this paper. The placement of the 36 scalp electrodes is shown in Figure 2. The SCAN software, a research grade data processing tool, was employed to remove noise and artifacts or decompose complex signals. The EEG signals were amplified by NuAmps Express system (Neuroscan Inc, USA) and sampled at 1000Hz.

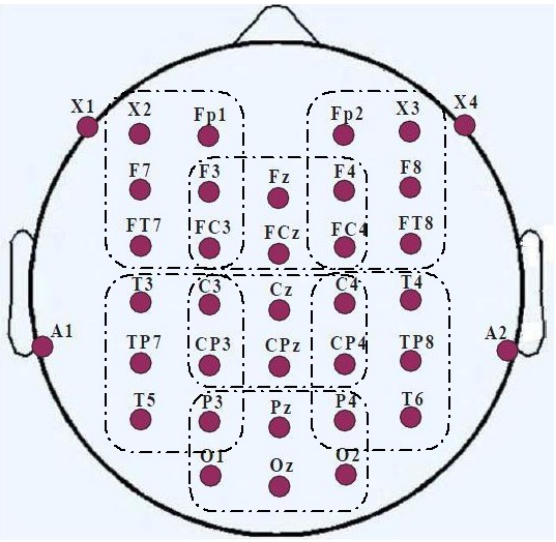

Fig. 2. The 36 EEG channels are divided into seven channel groups according to their spatial locations. In the feature extraction stage, features are first extracted from each single channel, and then averaged over each channel group.

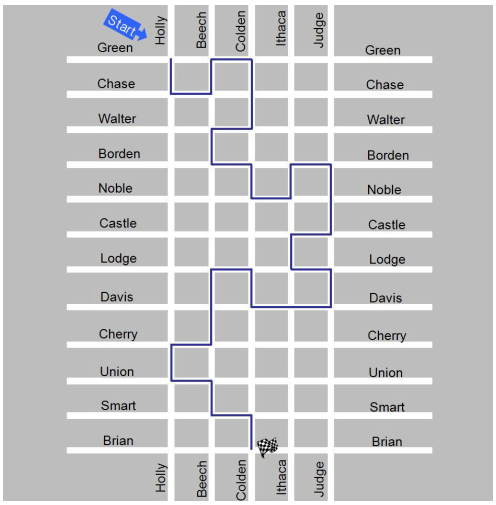

Fig. 3. The simulated city map employed in the driving experiment. Each subject was required to drive from an initial position to a destination pointed out in the simulated city map.

C. Experimental Procedure

Participants seated themselves comfortably in the chair and wore a fitted EEG cap with 40 channels. After setting up the EEG cap, participants were trained for the driving and navigation task by completing a practice block. The practice block allowed participants to learn how to operate the driving simulator including the steering wheel, speedometer, brake and gas pedal. Also, they could get familiar with the map-viewing task along with driving task. After the training segment, participants completed the test block by driving from an initial position to a destination pointed out by the simulated city map as shown in Figure 3. The designed route was highlighted on the static map. In order to control potential confounding factors, drivers were constrained to look at the map for any route information. There was no verbal navigation instructions provided to distract drivers. In both practice and test blocks, participants were asked to operate the driving simulator by following normal traffic laws as if they were driving a vehicle in the real world. Continuous EEG signals were recorded during the driving tasks. The start and the end time of the map-viewing periods were also recorded.

D. Target Event Definition

Figure 4 plots the statistics of the time lengths of the normal driving periods (inter-arrival periods of Event I), and the time lengths of map-viewing periods. The two target events to be predicted in this study were defined as follows.

- Normal Event I: Event I (starting to look at the map) after at least five seconds of continuous driving.
- Dangerous Event II: Event II (looking back to the road) after at least two seconds of map-viewing.

We are particularly interested in the prediction of 'normal Event I' because it is associated with the initiation of `uncertainty' of future route. The prediction of such event in driving is very insightful to provide navigation information timely based on a driver's need. It is noted that a driver may look at the map back and forth a number of times in a short period. These short-time frequent glances can be considered in one route-learning process. The corresponding Event Is are considered less typical to represent the initiation of cognitive uncertainty for future driving route after a relative long-term driving. Therefore, this study only focuses on the prediction of ***"Normal Event Is"***, which are defined as the Event Is that have at least five seconds of driving prior to map-viewing. The five seconds were chosen to eliminate the non-uncertainty-initiation- related Event Is [27], [40].

Also for the Event II, we only focus on the '***dangerous Event II***', since we consider the Event IIs with short map-viewing durations are much less dangerous than those with relative longer off-road glance durations. The prediction of 'dangerous Event IIs' is meaningful to provide warning in advance if the predicted time of looking back to the road exceeds a safety limit. According to [41], we employed two seconds as the criteria to identify the 'dangerous Event IIs'.

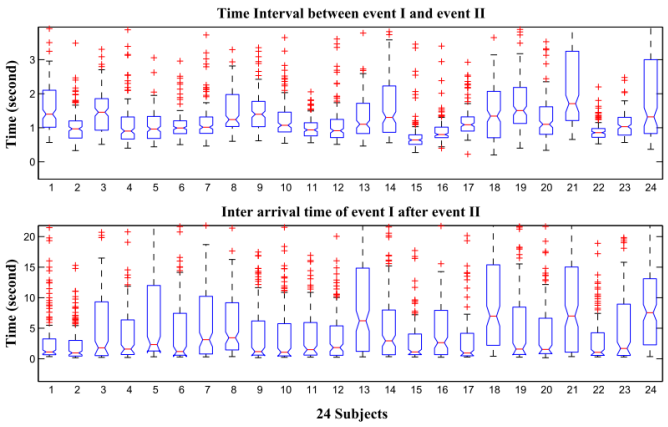

Fig. 4. Boxplots of the lengths of normal driving periods (inter-arrival intervals of event I) and the lengths the map-viewing periods. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points that are not considered as outliers, and outliers are plotted individually by the red 'plus' signs.

III. Adaptive-Threshold-Based Prediction Framework

In this study, we propose an adaptive-threshold-based (ATP) prediction framework to identify predictive patterns in EEG signals that are associated with Event I and Event II. The flowchart of the online prediction scheme is shown in Figure 5. A band-pass filter was employed to decompose EEG signals into four frequency bands; they are 8 to 13 Hz, 13 to 30 Hz, 2 to 50 Hz, and 1 to 100 Hz, respectively. Then quantitative pattern analysis for prediction was then performed in each frequency band, separately. In general, raw continuous EEG signals were converted to pattern clusters consecutively through a two-level feature extraction process using sliding windows. The probabilistic relationship between pattern clusters and interested event occurrences was then estimated. If a pattern cluster is more likely to be in the pre-event period, an event prediction is triggered. In following, the key components of the EEG-based prediction system are presented, including first level feature extraction of raw EEG, second-level feature extraction, feature selection, pattern-cluster formulation, probabilistic-rule-based prediction scheme, and prediction performance evaluation criterion.

A. First-Level Feature Extraction

As shown in Figure 5, the first-level features are extracted directly from raw EEG signals through a sliding window approach. Thirty-six EEG channels were divided into 7 channel groups according to their spatial locations (see Figure 2 for exact locations). Each feature was extracted from each of the 36 channels, and then averaged within each channel group. In particular, univariate, bivariate and time-frequency features from raw EEG signals were extracted as follows.

- *Univariate features:* For individual channel groups, nine univariate features were extracted, including *mean, variance, skewedness, kurtosis, signal power, curve length, number of peaks, average nonlinear energy,* and *variance to range ratio*. A detailed description of these EEG features can be found in [48]. Each feature was calculated from individual epochs of all 36 EEG channels and averaged among channels in the same group. Thus, 7 feature values (corresponding to 7 channel groups) were extracted for each univariate feature. In total, $7\times9 = 63$ features were extracted for the nine types of univariate features.
- *Bivariate features:* Three bivariate features were extracted including *pairwise Euclidean distance, pairwise T-statistics,* and *pairwise Pearson correlation*. The bivariate features were first calculated within each channel group, and subsequently averaged over all the pairs in each channel group. Each EEG epoch of 36 channels was transformed into 7 feature values (one for each group) for each bivariate feature. In total, $7\times3 = 21$ features were extracted for all three bivariate features.
- *Time-frequency feature:* Wavelet entropy was employed as a time-frequency feature. Specifically, discrete wavelet transform (DWT) was applied to each channel of EEG signal, and the wavelet entropy is computed based the relative energy associated with different frequency bands present in the EEG. The wavelet entropy provides information about the degree of order/disorder associated with a multi-frequency signal response [50]. Each EEG epoch of 36 channels was transformed into 7 wavelet entropy values for the 7 channel groups.

In the first-level feature extraction, each raw EEG epoch in a sliding window was converted to $63 + 21 + 7 = 91$ features. It should be noted that we applied different lengths of sliding window to monitor normal driving periods and map-viewing periods. This is because there were two events to predict. The prediction of Event I relied on the EEG signals in normal driving periods whereas the prediction of Event II relied on the EEG signals in map-viewing periods. The two periods do not have any overlap. In a real-time monitoring process, two independent prediction systems actually work interactively. If Event I is detected, the prediction system of Event II is turned on and the prediction system of Event I is turned off, vice versa. As a result, the two types of sliding window were applied alternately to monitor raw EEG signals. The length and step size of the Event I sliding window were 1 second and 100ms, respectively. The length and step size of the Event II sliding window were 250ms and 25ms, respectively.

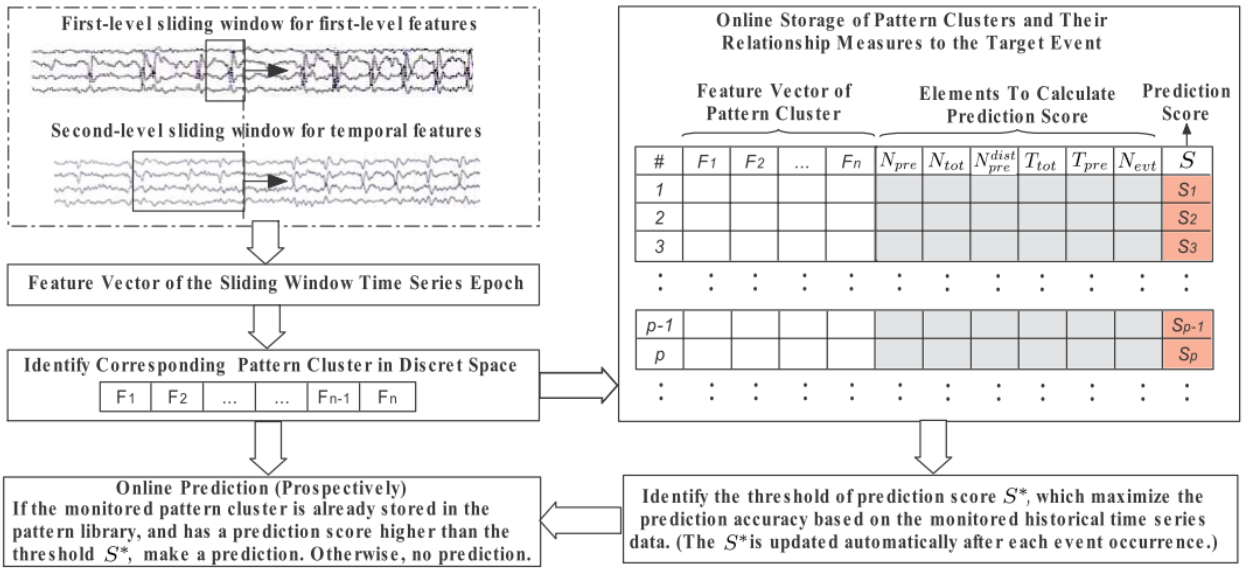

Fig. 5. Flowchart of the probabilistic adaptive-threshold-based online prediction scheme using the concept of pattern-cluster in discrete feature space.

B. Second-Level Feature Extraction

The second-level features were designed to characterize temporal patterns of the 91 first-level features. As shown in Figure 5, we employed a second sliding window to monitor the first-level features simultaneously. Sequences of the first-level features can be viewed as time series of features, each point representing one sliding window. Different lengths of the second-level sliding window were tested from 1 second to 5 seconds for Event I, and 500ms to 1250ms for the Event II. Given a pre-defined length of the second-level sliding window, we first applied a piecewise linear approximation algorithm to partition the time series into piecewise linear segments. Then we characterize temporal pattern of the time series by four features as demonstrated in Figure 6. In particular, for a time series $X = (x_1, x_2, \cdots, x_n)$, its key-turning points are shown in the figure. There are six sub-sections, three of which (segment a, c, e) showing an uptrend, and three of which (segment b, d, f) showing a downtrend. These trends indicate the degree of fluctuation of the time series. The following four important features are proposed to represent time series fluctuation patterns:

Feature 1: accumulated vertical increase in the segmented piecewise linear time series, which is calculated as

$$F_{II1} = H(a) + H(c) + H(e), \quad (1)$$

where the function $H(.)$ means the vertical distance from the starting point to the ending point of a sub-segment.

Feature 2: accumulated vertical decrease in the segmented piecewise linear time series, which is calculated as

$$F_{II2} = H(b) + H(d) + H(f). \quad (2)$$

Feature 3: percentage of the decreasing line segments, which is calculated as

$$F_{II3} = T(a + c + e)/T(X), \quad (3)$$

where $T(.)$ is the horizontal distance from the starting point to the ending point of a sub-segment.

Feature 4: range of the time series, which is calculated as

$$F_{II4} = max(X) - min(X), \quad (4)$$

where $max(X)$and $min(X)$ means the maximum and minimum values of the segmented time series, respectively.

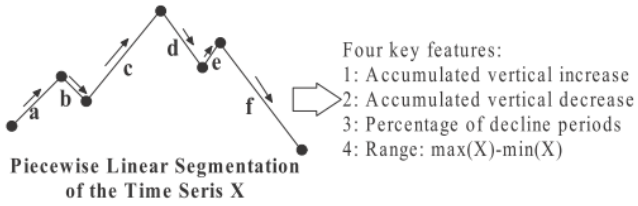

Fig. 6. Four skeleton-point-based features are employed to represent the temporal fluctuation pattern of a time series.

C. Feature Selection

For each of the 91 first-level features, there are 4 second-level features associate with it. That is, there are $91 \times 4 = 364$ feature candidates to represent each EEG epoch in the sliding window. However, not all of the extracted features are informative to the target events, and the high dimensional feature space makes the online learning task extremely difficult to capture predictive pattern in a short period. Feature selection has to be performed to achieve dimensionality reduction. In particular, we employed Pudil's sequential floating search [42], which is popular and fitted into our problem to select the most discriminative features to separate the two classes of EEG epoch (pre-event and non-event). There are two floating search schemes, called forward floating search (FFS) and backward floating search (BFS). Since FFS generally works faster than BFS when the expected number of selected features is much smaller than the

complete feature dimension. Thus, we employed FFS in this study. Starting from an empty feature set, the FFS is basically a bottom up search procedure, which includes new features and excludes the worst features in the current feature set sequentially to improve a class separability criterion. In this study, the 1-nearest neighbor classification error was used as the separability measure. The objective of SFFS is to select an optimal subset of features that minimize the 1-nearest neighbor classification error.

Let S_k be the feature subset of k features that have been selected from the complete feature set $F_d = \{f_1, f_2, ..., f_d\}$ using a criterion function $E(X_k)$. The values of $E(X_i)$ for all preceding subsets of size $i = 1, 2, ..., k-1$, are known and stored. The SFFS procedure can be summarized as follows:

- Step 1 (Inclusion): Select the most significant feature f_i from the available feature set $F_d - S_k$ to form the new feature set S_{k+1} by $E(S_k + f_i) \leqslant E(S_k + f_j) < E(S_k)$, where $f_i, f_j \in F_d - S_k$ and $i \neq j$.
- Step 2 (Exclusion): Find the least significant feature f_p in the subset S_{k+1} such that $E(S_{k+1} - f_p) \leq E(S_{k+1} - f_q)$ *for all* $f_q \in S_{k+1}$ and $p \neq q$. *If* $E(S_{k+1} - f_p) < E(S_{k+1})$, then exclude f_p from S_{k+1} *to form a new feature set* S'_k. *We have* $E(S'_k) < E(S_{k+1})$.
- Step 3 (Exclusion Continuation): Similar to step 2, continue to find the least significant feature f_m in the set S'_k. If $E(S'_k - f_m) \geq E(S'_k)$, then set $S_k = S'_k$, $E(S'_k) = E(S_k)$, and return to Step 1 for a new cycle of feature inclusion. If $E(S'_k - f_m) < E(S'_k)$, exclude f_m from S'_k to form a further reduced set S'_{k-1}. Set k=k-1. Repeat step 3 if k>2. If k=2, set $S_k = S'_k$, $E(S'_k) = E(S_k)$ and go to step 1.
- The FFS procedure stops when no features meet the criterion to be included in or to be removed from the current feature subset.

The floating search approaches take use of backtracking and are capable of correcting wrong inclusion/removal decisions. Floating search has become widely popular because it can often provide either the optimal or a close to optimal solution, and also require much less computational time than the traditional branch and bound method and most other currently used suboptimal strategies. However, in our experiments with 364 features, FFS still took long computing time due to the huge amount of possible feature combinations to be tested. And also based on the testing prediction accuracy, we found the obtained feature subset was suboptimal when compared with the features selected from a smaller pool of feature set.

To tackle the problem of "curse of dimensionality", we reduced the size of the feature candidates. In particular, we narrowed down the univariate features from nine to two: mean and curve length; and reduced the bivariate features from three to one: averaged pairwise Euclidean distance. The other two bivariate features were excluded mainly because their expensive cost of computing and they may not fit well to a fast online prediction task in this study. Together with the wavelet entropy, we narrowed the 1st level features to 4 candidates.

In addition, EEG signals are often described in terms of rhythmic activity and divided into frequency bands by using band-pass filters. To improve signal to noise ratio and also investigate EEG patterns in different brainwave bands, we analyzed EEG signals in four frequency bands: 8-13Hz, 13-30Hz, 2-50Hz, and 1-100Hz. The frequency bands 8-13 Hz and 13-30Hz generally correspond to the well-known alpha and beta bands of brainwaves, respectively. Alpha is considered to be an important brain frequency to learn and use information. When alpha is within normal ranges, one tends to experience good moods and have a sense of conscious and calmness. Beta waves represent some 'fast' cognitive activity of alert or anxious. The frequency band 2-50 Hz contains the five most useful brain frequencies that EEG researchers tend to follow: delta (below 4Hz), theta (4-7Hz), alpha (8-13Hz), beta (13-30Hz), and low gamma (30-45Hz). Finally, the EEG in frequency band 1-100Hz can be considered the cleaned raw EEG data for which the low frequency (<1Hz) and high frequency (>100Hz) noises are removed.

We performed FFS on the EEG of the four frequency bands separately and made online predictions for each frequency band separately. It is noted that the wavelet entropy was calculated based on the energy of different frequency bands and indicate the energy variations between bands. Thus, it is meaningless for single-band signals in 8-13 Hz and 13-30 Hz. For the two single-band EEG, the first-level feature candidates were three: mean, curve length and pairwise Euclidean distance. For EEG in 2-50Hz and 1-100Hz, we found that the features selection using 3 first-level features generated better prediction results than those using all the 4 candidates. The combination mean, pairwise Euclidean distance and wavelet entropy generated the best testing prediction performance. Thus, the experimental results of 2-50 Hz and 1-100 Hz reported in this paper used the selected features from a reduced feature set with these 3 first-level features. The FFS was performed on a reduced feature set with 84 features (7 channel groups×3 first level features×4 second level features).

Let the four first-level feature candidates denoted by: F_{I1} the averaged EEG signal within a channel group, F_{I2} the averaged curve length within a channel group, F_{I3} the averaged pairwise Euclidean distance within a channel group, and F_{I4} the averaged wavelet entropy within a channel group. The FFS selected features for event I and event II are summarized in Table II and Table III, respectively.

Table II. The FFS selected features for event I.

	Selected Feature	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}
8-13 Hz	Channel Group	1	1	4	5	6	4	1	5	2	4
	First Level	F_{I1}	F_{I2}	F_{I2}	F_{I1}	F_{I2}	F_{I1}	F_{I2}	F_{I2}	F_{I2}	F_{I2}
	Second Level	F_{II4}	F_{II1}	F_{II2}	F_{II4}	F_{II4}	F_{II4}	F_{II2}	F_{II4}	F_{II1}	F_{II1}
13-30 Hz	Channel Group	6	1	1	3	4	3	4	2	4	2
	First Level	F_{I1}	F_{I2}	F_{I2}	F_{I1}	F_{I2}	F_{I2}	F_{I1}	F_{I1}	F_{I1}	F_{I2}
	Second Level	F_{II4}	F_{II1}	F_{II2}	F_{II4}	F_{II2}	F_{II2}	F_{II4}	F_{II4}	F_{II2}	F_{II2}
2-50 Hz	Channel Group	7	4	2	6	6	7	1	2	-	-
	First Level	F_{I4}	F_{I4}	F_{I4}	F_{I3}	F_{I3}	F_{I3}	F_{I4}	F_{I1}	-	-
	Second Level	F_{II4}	F_{II4}	F_{II4}	F_{II1}	F_{II3}	F_{II1}	F_{II4}	F_{II4}	-	-
1-100 Hz	Channel Group	5	7	7	1	4	2	3	2	5	-
	First Level	F_{I4}	F_{I3}	F_{I4}	F_{I3}	F_{I3}	F_{I1}	F_{I1}	F_{I1}	F_{I1}	-
	Second Level	F_{II4}	F_{II2}	F_{II4}	F_{II2}	F_{II2}	F_{II2}	F_{II1}	F_{II4}	F_{II4}	-

Table III. The FFS selected features for event II.

	Selected Feature	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_10
8-13 Hz	Channel	1	1	6	4	2	3	-	-	-	-
	First Level	F_{I2}	F_{I2}	F_{I2}	F_{I2}	F_{I2}	F_{I2}	-	-	-	-
	Second Level	F_{II2}	F_{II1}	F_{II1}	F_{II1}	F_{II2}	F_{II2}	-	-	-	-
13-30 Hz	Channel	6	2	3	1	5	4	2	5	7	2
	First Level	F_{I1}	F_{I2}	F_{I2}	F_{I2}	F_{I2}	F_{I2}	F_{I2}	F_{I2}	F_{I2}	F_{I2}
	Second Level	F_{II4}	F_{II2}	F_{II4}	F_{II2}	F_{II1}	F_{II2}	F_{II4}	F_{II4}	F_{II4}	F_{II1}
2-50 Hz	Channel	6	1	4	6	4	7	5	7	-	-
	First Level	F_{I4}	F_{I3}	F_{I3}	F_{I3}	F_{I4}	F_{I3}	F_{I1}	F_{I4}	-	-
	Second Level	F_{II4}	F_{II2}	F_{II4}	F_{II3}	F_{II3}	F_{II4}	F_{II1}	F_{II3}	-	-
1-100 Hz	Channel	5	5	1	4	6	2	2	4	7	2
	First Level	F_{I4}	F_{I3}	F_{I3}	F_{I3}	F_{I3}	F_{I4}	F_{I1}	F_{I3}	F_{I3}	F_{I4}
	Second Level	F_{II4}	F_{II2}	F_{II1}	F_{II2}	F_{II2}	F_{II3}	F_{II1}	F_{II4}	F_{II1}	F_{II4}

As summarized in Table II and Table III, the selected features form a feature vector to represent the 'pattern' of each monitored EEG epoch in the sliding window.

D. Pattern Cluster in a Discrete Feature Space

For all N features that were selected, we partitioned each feature space into a number of non-overlapping intervals using linear equi-volume partitioning. The feature-vector patterns that fall into the same interval in all the 8 feature dimensions represent a set of close-by patterns with similar underlying cognitive activities. A set of such feature-vector patterns was considered as a pattern cluster. Using the pattern cluster, one can represent millions or billions of feature-vector patterns by a fixed number of pattern clusters representing groups of similar brain activities. In this study, we explored different numbers of partition bins ranging from 4 to 10. The partition with 8 intervals obtained relative better results and thus reported in this study. For example, for an 8-dimensional feature space with each dimension partitioned into 8 intervals. Then the total number of all possible pattern clusters was $8^8 = 16777216$, which accounted for numerous brain activities during driving. Although this was a very large number, our experiments showed that recorded pattern clusters were in a level of one thousand.

E. Probabilistic Prediction Score

A probabilistic prediction score was used to identify pattern clusters that are predictive to the two target events. In general, each EEG epoch in the sliding window was converted into a feature pattern cluster and then stored in a pattern-recording table. A prediction score that was associated with the likelihood in the pre-event period was calculated based on the appearance frequency of the pattern cluster in pre-event and non-event periods. A prediction was made if the prediction-score of the pattern cluster exceeded an adaptive score threshold, which was optimized after each occurrence of a target event.

1) Prediction Score: Given a pattern cluster indexed as the kth cluster in the pattern-recording table, its prediction score S_k is defined as follows:

$$S_k = \frac{N_{pre}/N_{tot}}{R_{pre}} \times \frac{N_{pre}^{dist}}{N_{evt}}, \quad (5)$$

where N_{pre} is the number of occurrences of the pattern cluster in all previously monitored pre-event periods; and N_{pre}^{dist} is the number of pre-event periods such that the pattern cluster appears at least once in each of them; N_{tot} is the total number of occurrences of the pattern cluster, and N_{evt} is the total number of events that have occurred. For example, if two events have been monitored, a pattern cluster occurs three times in the first pre-event period, 2 times in the non-event periods, and does not show up in the second pre-event period, then $N_{pre} = 3$, $N_{pre}^{dist} = 1$, $N_{tot} = 5$, and $N_{evt} = 2$. Finally, R_{pre} is the time ratio between pre-event periods and non-event periods. In particular, it is calculated as follows:

$$R_{pre} = \frac{T_{pre}}{T_{tot} - T_{pre}} = \frac{N_{evt} \times T_{hrzn}}{T_{tot} - N_{evt} \times T_{hrzn}}, \quad (6)$$

where T_{pre} is the total length of monitored pre-event periods, T_{tot} is the total length of monitored EEG time series; and T_{hrzn} is the length of prediction horizon.

The predictive score proposed in Eq. (5) indicates how strong a pattern cluster is associated with the target event. In particular, the first term of Eq. (5) is to evaluate if the pattern cluster occurs in pre-event periods at a random level. If the pattern is purely random in both pre-event periods and the non-event periods, then we have $E(N_{pre}/N_{tot}) = E(R_{pre})$. If the pattern occurs more frequently in pre-event periods than the non-event periods, we have $E(N_{pre}/N_{tot}) > E(R_{pre})$. The higher the ratio value, the more likely the pattern cluster is associated with the target event. If the pattern occurs less frequently in pre-event periods than the non-event periods, we have $E(N_{pre}/N_{tot}) < E(R_{pre})$. The second term of Eq. (5) is to evaluate if a pattern cluster occurs in many pre-event periods. We expect that an ideal candidate of predictive pattern should appear in most pre-event periods, not only in a few ones. That is $N_{pre}^{dist}/N_{evt} \approx 1$. In summary, Eq. (5) estimates the likelihood of a pattern cluster in the pre-event period and reduces the bad effects of some extreme situations. In general, the higher the prediction score, the higher probability the pattern cluster appears in the pre-event period, and thus the more prominent it is to predict events.

2) Score-Based Prediction Rule: *The pattern-recording table stores and summarizes the recorded pattern clusters as well as calculates their prediction scores according to Eq. (5). We employed an adaptive threshold on the prediction score to discriminate the pre-event and non-event pattern clusters.* Since for the recorded patterns we already know their class (event or pre-event) and prediction score, for any given score threshold, it is convenient to calculate the corresponding sensitivity and specificity retrospectively. We employed a heuristic search approach set the value of threshold S_*. In particular, we tried 30 values within the current prediction score range (S_{max}-S_{min}), starting from 1/30 of the range with an increment of 1/30 range each time. *The optimal threshold S_* is set to the value that maximized the overall prediction performance (sensitivity + specificity) based on the previously*

recorded patterns and their prediction scores. The threshold $S_*$was updated after each occurrence of a target event. The prediction rule works as follows. Each impending EEG epoch in the sliding window was first converted to a pattern cluster. Assume the pattern cluster was indexed as the kth cluster in the pattern-recording table, its prediction score is denoted as S_k. Then the prediction rule is defined by:

$$predictor = \begin{cases} 1, & if\ S_k > S_*\ (make\ a\ prediction) \\ 0, & otherwise\ (no\ prediction\ ) \end{cases}$$

F. Evaluation of Prediction Performance

The most commonly used prediction performance measures are specificity and sensitivity. However, the traditional definition of specificity and sensitivity only focus on the correctness of each individual prediction and do not consider the prediction horizon and event-specific information. They are inappropriate to measure prediction performance directly for our online prediction problem, which has to consider the effects of prediction horizon and the event-specific requirements. Thus, we proposed a modified version of sensitivity and specificity that are well suited to evaluate the online event-prediction problem. In particular, we introduced a time-block-based sensitivity, denoted by sen_{blk}, which is defined as portion of correctly predicted events in the total number of events. An event is considered to be correctly predicted if there is at least one true prediction within its preceding prediction horizon. According to Mormann et al. [43], we also employed a time-block-based specificity, denoted by spe_{blk}, which is defined as the portion of non-event time period that is not in false awaiting state. A demonstration of the sen_{blk} and spe_{blk} quantification is shown in Figure 7.

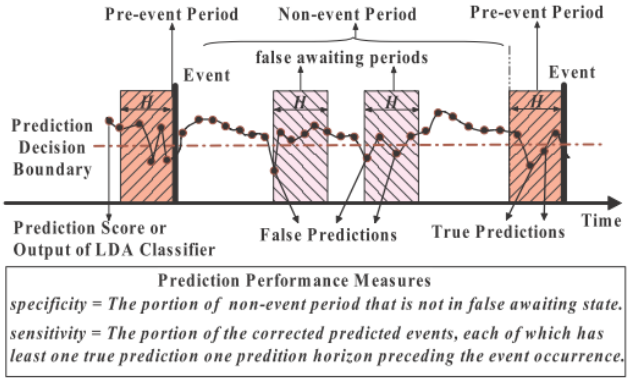

Fig. 7. A demonstration of the time-block-based sensitivity sen_{blk} and the time-block-based specificity spe_{blk}.

TABLE IV. Computational and parameter settings of our prediction framework.

Parameter Setting	Setting Choices
Prediction Horizon	Event I: 400, 600, 800, 1000 ms Event II: 300, 400, 500 ms
1st-level sliding window (monitor raw time series)	Event I: window size 1000 ms, moving step 100 ms Event II: window size 250 ms, moving step 25 ms
2nd-level sliding window (monitor feature time series)	Event I: window size 1000, 2000, 3000, 4000, 5000 ms, moving step 100, 200, 300 ms Event II: window size 500, 750, 1000, 1250 ms, moving step 25 ms
Online Prediction Scheme	Adaptive Threshold-Based Prediction Scheme
Feature Selection Method	Pudil's floating search based on 1-Nearest Neighbour leave-one-out classification performance.
1st-level features	Nine univariate features: mean, variance, skewness kurtosis, signal power, curve length, number of peaks average nonlinear energy, variance to range ratio.
	Three pairwise bivariate measures: Euclidean distance, T-statistics, Pearson correlation.
	One time-frequency measure: wavelet entropy (features are averaged over each channel group as shown in Figure 2.4)
2nd-level features (temporal pattern feature)	1. Accumulated vertical increase 2. Accumulated vertical decrease 3. Percentage of decline periods 4. Amplitude range

IV. Computational Experiments

A. Computational Settings

The proposed prediction framework was implemented and tested on the EEG recordings of 24 subjects for both 'normal Event I' and 'dangerous Event II'. The complete parameter settings of the prediction framework discussed in the previous section are summarized in Table IV.

B. Experimental Results

For the prediction of Event I, the averaged training and testing results over the 24 subjects for different prediction horizons and frequency bands are summarized in Table V. The best testing performance of our prediction algorithm was achieved with a sen_{blk} of 0.79 and a spe_{blk} of 0.83 using a prediction horizon of 400ms in frequency band 2-50 Hz. The average and standard deviation of prediction times are provided in Table VI. The detailed prediction results of Event

I for individual subjects based on the best prediction horizon and frequency band are shown in Table VII.

TABLE V. Training and testing prediction results for Event I averaged over the 24 subjects using the best parameter over different prediction horizons and frequency bands.

Frequency Band (Hz)	Horizon (ms)	Training		Testing	
		Sen_{blk}	Spe_{blk}	Sen_{blk}	Spe_{blk}
8-13 Hz	400	0.71	0.77	0.77	0.63
	600	0.72	0.72	0.76	0.56
	800	0.85	0.57	0.81	0.51
	1000	0.81	0.55	0.81	0.51
13-30 Hz	400	0.67	0.75	0.71	0.68
	600	0.71	0.71	0.77	0.62
	800	0.73	0.68	0.83	0.55
	1000	0.73	0.66	0.82	0.55
2-50 Hz	400	0.82	0.90	**0.79**	**0.83**
	600	0.87	0.82	0.83	0.75
	800	0.88	0.8	0.83	0.71
	1000	0.88	0.78	0.85	0.67
1-100 Hz	400	0.34	0.89	0.39	0.8
	600	0.48	0.77	0.52	0.68
	800	0.56	0.72	0.63	0.61
	1000	0.6	0.67	0.69	0.55

TABLE VI. Prediction time statistics of Event I averaged over the 24 subjects using the best parameter over different prediction horizons and frequency bands.

Frequency Band	Horizon (ms)	Statistics of Prediction Time for Event I			
		Training		Testing	
		mean (ms)	std.	mean (ms)	std.
8-13Hz	400	241.61	169.74	291.55	152.01
	600	385.98	249.94	475.08	201.89
	800	520.42	331.51	642.72	269.47
	1000	761.97	348.11	851.86	275.05
13-30Hz	400	81.29	147.60	81.27	149.59
	600	151.59	240.10	181.79	257.19
	800	208.50	319.52	274.66	345.14
	1000	304.95	414.72	410.37	445.55
2-50Hz	400	200.00	178.40	234.20	176.44
	600	336.65	266.40	381.65	244.33
	800	495.54	332.64	591.40	284.74
	1000	633.33	411.35	756.21	354.09
1-100Hz	400	244.81	171.24	297.91	152.13
	600	395.34	244.83	484.08	197.79
	800	553.26	317.99	651.32	266.50
	1000	737.25	358.06	847.44	277.44

TABLE VII. Training and testing results of Event I for 24 individual subjects using the prediction horizon of 400ms and the frequency band of 2-50 Hz.

Subject	Settings			Training		Testing	
	L_{mw} (ms)	L_{step} (ms)	Horizon (ms)	sen_{blk}	spe_{blk}	sen_{blk}	spe_{blk}
1	5000	100	400	0.77	0.95	0.75	0.88
2	4000	100	400	0.85	0.91	0.69	0.89
3	5000	100	400	0.86	0.88	0.86	0.75
4	5000	100	400	0.88	0.86	0.81	0.69
5	5000	100	400	0.71	0.91	0.79	0.89
6	3000	100	400	0.80	0.94	0.79	0.82
7	4000	100	400	0.79	0.93	0.72	0.83
8	5000	100	400	0.81	0.94	0.87	0.87
9	3000	100	400	0.75	0.97	0.80	0.94
10	3000	100	400	0.76	0.92	0.94	0.84
11	3000	100	400	0.93	0.85	0.85	0.73
12	3000	100	400	0.84	0.87	0.74	0.78
13	3000	100	400	0.87	0.88	0.86	0.81
14	4000	100	400	0.76	0.92	0.85	0.80
15	4000	100	400	0.75	0.90	0.82	0.85
16	3000	100	400	0.89	0.82	0.65	0.77
17	4000	100	400	0.88	0.96	0.71	0.95
18	4000	100	400	0.71	0.86	0.93	0.85
19	3000	100	400	0.79	0.91	0.56	0.90
20	4000	100	400	0.94	0.78	0.89	0.76
21	5000	200	400	0.82	0.74	0.50	0.87
22	4000	100	400	0.85	0.98	0.83	0.90
23	5000	100	400	0.79	0.87	0.77	0.73
24	4000	100	400	0.89	0.96	0.94	0.90
Ave.				**0.82**	**0.90**	**0.79**	**0.83**
PA				**0.86**		**0.81**	

Figure 8 illustrates an example of the prediction results of Event I in Subject 2 using the best training parameter settings. From the figure, it can be seen that our prediction algorithm yielded accurate prediction performance for the prediction of Event I. The high specificity achieved by our algorithm indicates that our framework is robust to signal noises because an alarm is only triggered when a monitored pattern cluster is already identified as a pre-event pattern with a higher than threshold prediction score in the pattern-recording table. All other noisy patterns cannot trigger any warning alarms.

For the prediction of Event II, the training and testing results averaged over the 24 subjects are reported in Table VIII. The best testing performance was achieved with a sen_{blk} of 0.96 and a spe_{blk} of 0.45 using the prediction horizon of 500ms and the frequency band of 13-30 Hz. The average and standard deviation of prediction times are provided in Table IX. Detailed prediction results for individual subjects are shown in Table X. Note that the prediction performance of Event II was worse than that of Event I. We postulate that, because in addition to the normal driving task the map-viewing process requires concurrent execution of various cognitive, visual and motor activities in a short period, EEG signals are more complex and it is harder to discover predictive patterns that are associated with the intention of looking back to the road.

TABLE VIII. Training and testing results Event II averaged over the 24 subjects with the best parameter over different prediction horizons and frequency bands.

Frequency Band	Horizon (ms)	Training		Testing	
		sen_{blk}	spe_{blk}	sen_{blk}	spe_{blk}
8-13 Hz	300	0.82	0.54	0.98	0.40
	400	0.80	0.51	0.99	0.37
	500	0.83	0.47	0.99	0.31
13-30 Hz	300	0.80	0.60	0.93	0.47
	400	0.82	0.57	0.96	0.44
	500	0.78	0.59	**0.96**	**0.45**
2-50 Hz	300	0.72	0.66	0.77	0.53
	400	0.76	0.63	0.82	0.49
	500	0.75	0.61	0.86	0.46
1-100 Hz	300	0.72	0.65	0.76	0.51
	400	0.74	0.63	0.80	0.49
	500	0.72	0.61	0.83	0.44

TABLE IX. Prediction time statistics of Event II averaged over the 24 subjects using the best parameter over different prediction horizons and frequency bands.

Frequency Band	Horizon (ms)	Statistics of Prediction Time for Event II			
		Training		Testing	
		mean	std.	mean	std.
8-13Hz	300	254.13	85.43	276.4	59.45
	400	338.28	106.56	370.57	75.57
	500	427.09	133.79	452.4	110.21
13-30Hz	300	213.83	108.95	240.6	88.47
	400	308.21	128.65	351.44	87.15
	500	382.48	156.6	429.78	117.51
2-50Hz	300	207.04	118.6	245.3	96.44
	400	297.98	141.5	330.07	123.9
	500	376.75	170.82	426.88	133.74
1-100Hz	300	261.71	84.84	278.93	58.86
	400	357.7	98.16	373.37	81.65
	500	443.72	120.96	478.78	71.67

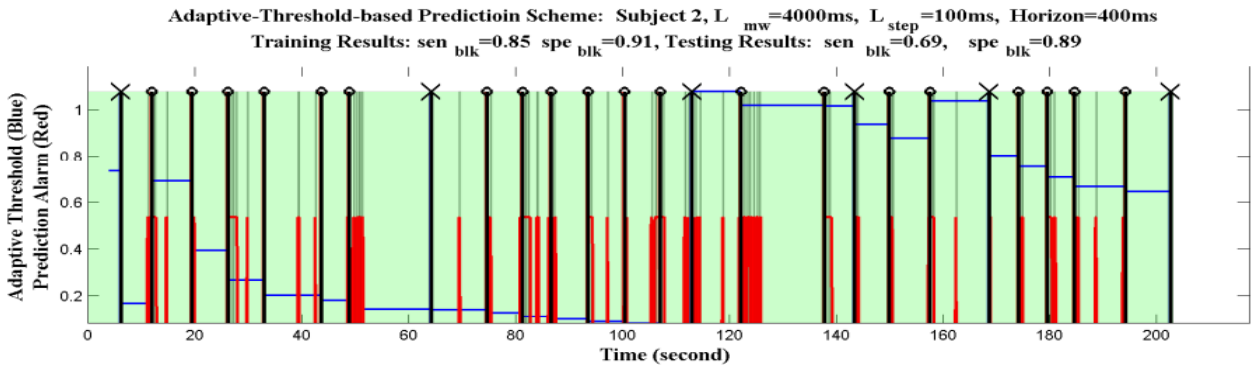

Fig. 8. A demonstration of the prediction outcome of Event I in Subject 2 using the prediction horizon of H= *400ms with* $L_{mw} = 4000\ ms$ *and* $L_{step} = 100$ *minute in the frequency band 2-50 Hz. The vertical black lines indicate the starting times of Event I. The cross sign on top of a vertical black line indicates the system failed to predict the corresponding event I within the prediction horizon of 400ms. The circle sign on top of a vertical black line indicates the corresponding event I was successfully predicted. The vertical red lines represent the timing of the predictions. The horizontal blue lines represent the threshold level, whose value is updated after each occurrence of event I.*

TABLE X. Training and testing results of Event II for 24 individual subjects using the prediction horizon of 500ms and the frequency band of 13-30 Hz.

	Settings			Training		Testing	
Subject	L_{mw} (ms)	L_{step} (ms)	Horizon (ms)	sen_{blk}	spe_{blk}	sen_{blk}	spe_{blk}
1	1250	25	500	0.60	0.64	0.87	0.41
2	750	25	500	0.97	0.27	1.00	0.28
3	1000	25	500	0.83	0.77	1.00	0.51
4	1000	25	500	0.75	0.57	1.00	0.29
5	750	25	500	0.50	0.65	1.00	0.49
6	1000	25	500	0.75	0.48	1.00	0.76
7	1000	25	500	0.60	0.59	1.00	0.42
8	1250	25	500	0.60	0.68	0.67	0.53
9	1250	25	500	0.86	0.64	1.00	0.62
10	1250	25	500	0.80	0.66	1.00	0.66
11	500	25	500	1.00	0.66	1.00	0.39
12	1250	25	500	0.67	0.69	0.60	0.54
13	1000	25	500	0.80	0.54	1.00	0.48
14	1250	25	500	0.86	0.50	1.00	0.30
15	500	25	500	0.96	0.29	1.00	0.12
16	1000	25	500	1.00	0.46	1.00	0.46
17	1250	25	500	0.67	0.82	1.00	0.80
18	1250	25	500	0.83	0.62	1.00	0.52
19	1000	25	500	0.87	0.40	0.93	0.28
20	1000	25	500	0.60	0.64	1.00	0.50
21	1250	25	500	0.85	0.31	1.00	0.38
22	750	25	500	0.80	0.81	1.00	0.15
23	750	25	500	0.88	0.84	1.00	0.62
24	1250	25	500	0.70	0.56	0.89	0.31
Ave.				**0.78**	**0.59**	**0.96**	**0.45**
PA				**0.68**		**0.70**	

C. *Effectiveness of Adaptive Updating*

The threshold level of prediction score was updated (optimized) after each event. To test the effectiveness of the adaptive threshold-updating scheme, we compared the prediction performances for different updating periods in the frequency band of 2-50 Hz. In particular for each subject, we updated the threshold in the first 0%, 10%, 30%, 50%, 70%, 90%, and 100% of the total target events, respectively. That is, 0% means that the initial threshold was unchanged throughout the prediction process, and 100% means that the threshold was updated for all events. Figure 9 plots the averaged prediction performances of the seven different parameters of updating periods. It can be seen that the overall prediction accuracies increased as the portion of EEG data used to update the threshold increased. The strong increase trend of prediction accuracy indicates that the adaptive threshold-updating scheme is truly effective in increasing the prediction performance over time

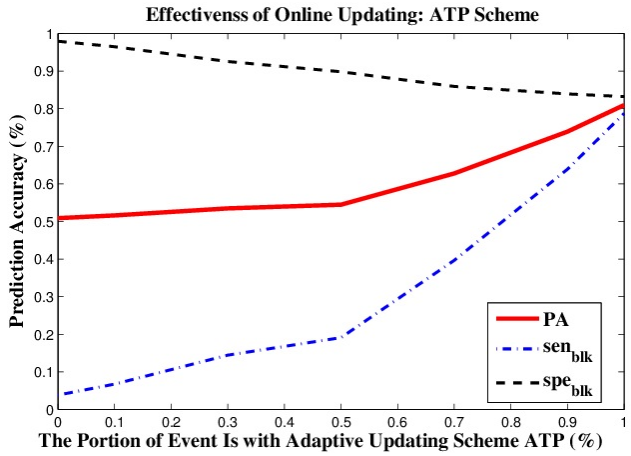

Fig. 9. The effectiveness of the adaptive threshold-updating scheme using the EEG data in the frequency band 2-50 Hz. The horizontal axis indicates the portion of events the threshold of the ATP scheme was actively updated. The point 0 indicates that the initial score threshold was unchanged throughout the prediction process; and the point 1 means that the threshold was updated for all events throughout the entire prediction process. The strong increase trend of prediction accuracy indicates that the adaptive updating scheme ATP is effective to increase online prediction performance over time.

D. *Effectiveness of Feature Selection*

As discussed in subsection III.C, we did experiments with different size of feature candidates. The Figure 10 and Figure 11 show the prediction performance using three different feature sets. The first feature set is the complete feature set

with all extracted features. The second feature set is the FFS-selected feature subset from the complete feature set. The third feature set is the FFS-selected feature subset from feature sets with reduced number of features. We discussed how we selected the reduced feature set in a heuristic manner in subsection III.C. The prediction performance boxplots clearly show that the complete feature set with around 300 features generated the worst prediction performance. The FFS-selected features from the complete feature set generally improved the prediction performance cross the 24 subjects. However, it clearly provided a suboptimal solution, since the FFS-selected features from a reduced feature set generated considerable better prediction results (the reduced feature set is a subset of the complete subset). Currently, the feature selection problem is still an open question for a high-dimensional data-mining task. Better feature selection frameworks are needed to search for the optimal feature subset in a high dimensional space while maintaining a good performance/speed ratio.

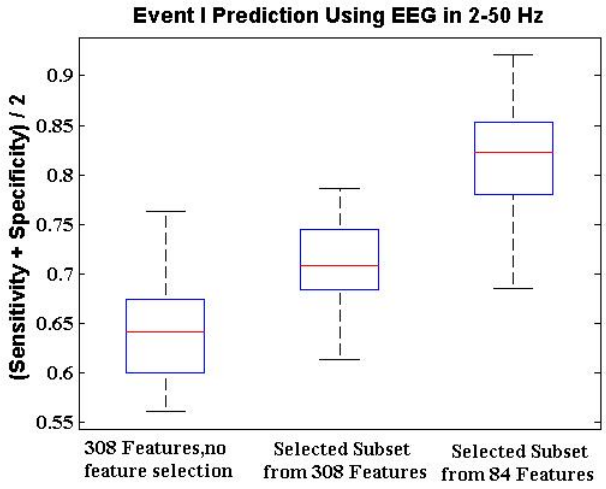

Fig. 10. Boxplot of the testing event I-prediction performance of the 24 subjects in 2-50 Hz using all 308 features, the selected features from the 308 features, and the selected features from a reduced set with 84 features, respectively. The feature set with 308 features was formed by 11 first-level features (9 univariate features + pairwise Euclidean distance+ wavelet entropy) × 4 second-level features ×7 EEG channel groups. The feature set with 84 features used four first-level features (mean, curve length, pairwise Euclidean distance, and wavelet entropy).

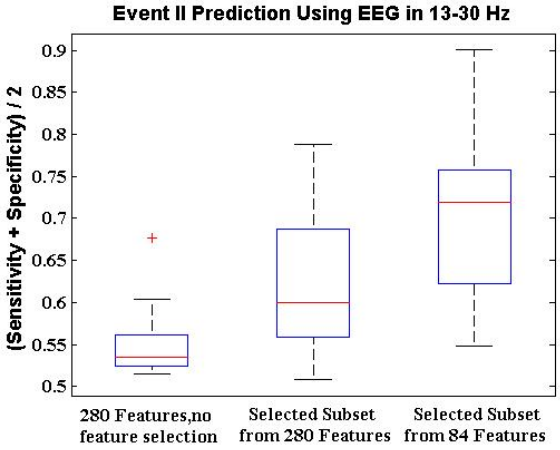

Fig. 11. Boxplot of the testing event II-prediction performance of the 24 subjects in 13-30 Hz using all 280 features, the selected features from the 280 features, and the selected features from a reduced set with 84 features, respectively. The feature set with 280 features was formed by 10 first-level features (9 univariate features + pairwise Euclidean distance) × 4 second-level features × 7 EEG channel groups. The feature set with 84 features used four first-level features (mean, curve length, and pairwise Euclidean distance).

V. Discussion

A. Prediction Performance

There has been an explosion research on detection of driver state for driving assistance systems. However, very few investigations have been performed on real-time prospective prediction of cognitive activity. Most of the current studies were concentrated on the detection or predictability of mental states. Recently, Stefan et. al. investigated the online prediction of a driver's intention to brake before any actions become observable [44]. A linear discriminant analysis (LDA) classifier was trained using the EEG signals between 260ms pre-response and the response onset. The detection rate was 80% using the combination of features of EEG, EMG, and several driver behavior features. The false alarm rate was 1.96 per hour, and the average alarm time was 167ms. Berka et. al. employed EEG data to monitor the levels of task engagement and mental workload continuously in an operational environment [45]. A second-by-second classification was applied to detect workload in different engagement levels. A problem of the existing approaches is that the trained online classifier cannot be adaptive to each person, and thus limit their prediction performance and application potentials. In this study, we develop an adaptive prediction framework to predict driver distraction prospectively. The proposed ATP scheme generated very promising prediction results based on a dataset of 24 subjects. Using the best training parameter settings, the average testing sensitivity and specificity of the ATP scheme

are 79% and 83%, respectively, to predict the intention of looking to the map.

The probabilistic ATP scheme constructs a pattern library for each subject specifically and gains the predictive pattern knowledge of each subject over time. With the pattern-cluster approach in discrete feature space, the size of the pattern library is limited. In real-life applications, the pattern-cluster library only takes a very small space and is extremely computational efficient. In this experiment, the total number of stored pattern clusters for each subject is at a level of one thousand, and the number of identified pre-event pattern-clusters is at a level of one hundred. Another significant advantage of the ATP scheme is that it is not sensitive to pattern noises and outliers in the online monitoring process. A prediction is only triggered if the monitored pattern cluster is an already identified as a pre-event pattern cluster in the pattern library. All other monitored patterns (including any pattern noises and outliers) cannot trigger any warning alarms. This makes the proposed probabilistic ATP prediction very attractive in real-life applications. The proposed ATP prediction framework provides a useful analytical tool for online monitoring and prediction of driver distraction using multichannel EEG signals.

B. The Application of Prediction of Two Events in Developing Information System

Previous work laid the foundation to differentiate distracted periods from non-distracted periods during driving. When drivers look at a map for route information, distraction is inevitable to some extent. Therefore, the proposed prediction approach of driving distraction has a potential to prevent the driver distraction in advance or at least help to reduce the extent of distraction and its resulting hazards. With such approach being applied, in-vehicle navigation systems can be improved by adopting a need-based design.

1) Application of Prediction of 'Normal Event I'

The prediction of 'normal Event I' can be applied to prevent head-turning distraction in advance. An improved design of navigation systems can be achieved by integrating the capability of predicting drivers' needs of route information. In other words, the route information would be presented only when a driver is uncertain about route and intends to look at the map for assistance. The route information provided in advance would prevent drivers from being distracted by reading a map passively, meantime, the redundant/unnecessary navigation information which leads to drivers' annoyance and takes up cognitive resources from the primary driving task, would be reduced.

The rules presented in Table XI with respect to 'Normal Event I' can be applied in the design of navigation and warning systems. T_{p1} is the prediction time ahead of the occurrence of Event I; and T_r is the average simple braking reaction time of auditory stimuli, which is 514ms [46].

TABLE XI. Application of prediction of "Normal Event I" in navigation systems design.

Hazard detection	Condition	System Operation
No/ Unknown	$T_{p1} > T_r$	Provide verbal route information in advance to avoid the map-viewing activity
	$T_{p1} < T_r$,	Switch on the prediction of 'dangerous Event II'
Yes	$T_{p1} > T_r$	Provide a verbal warning: Do not look away from the road
	$T_{p1} < T_r$	Provide a verbal warning: Look back to the road immediately.

When T_{p1} is longer than T_r, route information can be delivered before the driver looks at the map. If there is no hazard detected (or unknown), the driver will obtain the route information without being distracted from the road. Thus, the risk of eye-off-road can be largely reduced. However, when T_{p1} is shorter than T_r, the map-viewing action cannot be avoided, since the driver may already start looking at the map when the route information is presented. Thus, the system let the driver keep on reading maps rather than interrupting the driver by providing the route information to add on additional distraction. At this time, the prediction of 'dangerous Event II' is switched on, and the system will provide warnings to the driver in advance if the predicted map-view duration is longer than the safety threshold.

When there is a hazard event being detected and T_{p1} is longer than T_r, the system would alarm the driver to watch out the hazard and do not look away from the road. However, when T_{p1} is shorter than T_r, the driver may already look at the map when the warning message is provided. Therefore, the message should warn the driver to look back to the road and pay attention to the detected hazard.

2) Application of Prediction of 'Dangerous Event II'

The prediction of 'dangerous Event II' can be used to design a warning system to reduce the risk of long-time map-viewing activities in driving. When looking at a map, a driver loses attention to the driving task and road conditions. Thus, the map-viewing behavior would result in an impaired driving performance and may lead to dangerous situations. With the capability of predicting 'dangerous Event II', the system can predict the time duration of a map-viewing process ahead of time. If the predicted time duration of the map-viewing process exceeds a safe time length, the system can warn the driver ahead of time to look back and turn attention to road conditions.

The rules presented in Table XII with respect to 'dangerous Event II' can be applied to design navigation and warning systems. T_{p2} is the prediction time ahead of the occurrence of Event II and N is the number of prediction horizon. T_{hrzn} denotes the length of prediction horizon. The summation of T_{p2} and $N \times T_{hrzn}$ denotes the predicted time duration of map viewing task, which is the time interval between the predicted time of event II occurrence and the time of event I actual occurrence. The safety threshold is set to two seconds to select long map-viewing periods with high risk during driving [43].

TABLE XII. Application of prediction of "Dangerous Event II" in navigation systems design.

Hazard detection	Condition	System Operation
No/ Unknown	$T_{p2}+N\times T_{hrzn} > safety\ threshold$	Provide a warning in advance to prevent longer duration of maps-viewing and decrease the associated driving risks
	$T_{p2}+N\times T_{hrzn} < safety\ threshold$	Keeps on predicting the 'dangerous Event II' until the T_{p2}>safety threshold or the driver's attention return to the road.
Yes	$T_{p2}+N\times T_{hrzn} \leq or > safety\ threshold$	Provide warning: Look back to the road immediately.

A longer $T_{p2}+N\times T_{hrzn}$ than the safety threshold indicates that the duration of eye-off-road may exceed the safety limit. Therefore, when there is no hazard detected (or no hazard detection system is installed) the system should warn the driver to look back to the road without getting distracted for a long time. A shorter $T_{p2}+N\times T_{hrzn}$ than the safety threshold indicates a relatively safe map-glance action. In that case, when there is no hazard detected (or no detection system is installed), the system continues to monitor EEG signals and keeps on predicting the 'dangerous Event II' until the $T_{p2}+N\times T_{hrzn}$ is longer than the safety threshold or the driver return to the road.

However, when there is a hazard event being detected, a delay to return to the road may lead to less time to response to the hazard and cause dangerous situations. Therefore, the system should alert the driver to watch out the hazard immediately no matter the $T_{p2}+N\times T_{hrzn}$ is longer or shorter than the safety threshold.

C. Practical Applications of EEG Techniques

From a practical application standpoint, EEG technique applied in this study has its advantages in several situations compared to eye movement tracking used in previous studies. Firstly, it enables the system to obtain a driver's cognitive activities in real-time and provide appropriate feedbacks to assist the driver. The performance of the eye-movement-tracking-based methods can be serious deteriorated in the presence of strong light, while EEG signals are usually stable under various environmental conditions. This characteristic of EEG data is really attractive in the real world applications. In addition, EEG data are superior to protect driver privacy without recording any personal information. With higher and higher privacy standard in the current market, drivers would accept a system with techniques of high standard privacy protection more easily.

Although this study was carefully prepared, there are still several limitations. First of all, although the proposed computational prediction approach provides a promising tool to improve the current design of driving assistant systems, the influence of the driver-distraction prediction on driving behaviors is unknown. For example, the false predictions may also incur extra distractions and annoyance to drivers. This problem will be addressed in the future work. Secondly, the reaction period that allows drivers to response to the prediction outcomes was not considered in the current work. Reaction time can be used to evaluate whether a driver has enough time to react to an upcoming event or note. To enhance the practical utility of the proposed ATP prediction framework in the design of new navigation systems, we will consider the reaction time in the prediction system in future work. In addition, we employed the conventional wired wet EEG electrodes in the present experimental setup, which are difficult and inconvenient to be applied for real-world applications. In the further work, we will test EEG devices with dry electrodes and wireless data transfer function, and investigate the prediction power and stabilization of the remotely collected EEG data from dry electrodes.

VI. Conclusion

In this study, we investigated the online monitoring and prediction of driver distraction using EEG signals of 24 subjects. We propose an adaptive online prediction framework that is capable of capturing subject-specific predictive patterns autonomously by constructing a subject-specific pattern library, based on which a probabilistic prediction rule is established. The proposed online prediction system achieved promising prediction results with overall prediction accuracy of 81% for Event I and 70% for Event II. Under the best performance settings, the average prediction time of Event I is 234ms ahead of the real Event I occurrence, and the average prediction of Event II is 430ms ahead of the real Event II occurrence. The proposed methodology provides a practical tool to solve the challenging problem of online predicting of driver distraction using multivariate EEG signals. It has a potential to improve design of future intelligent navigation

systems from a novel perspective by preventing driver distractions in advance and the related safety risks.

Reference

[1]U.S. Department of Transportation, National Highway Traffic Safety Administration, "Distracted Driving 2010," Washington, DC, 2012. NHTSA'S National Center for Statistics and Analysis. Available from http://www-nrd.nhtsa.dot.gov/Pubs/811650.pdf

[2]T. Ranney, "Driver distraction: A review of the current state-of-knowledge." *National Highway Traffic Safety Administration Vehicle Research and Test Center*, Washington D.C, Tech. Rep., Apr. 2008.

[3]B. Donmez, L. N. Boyle, and J. D. Lee, "Differences in Off-Road Glances: Effects on Young Drivers' Performance," *Journal of Transportation Engineering*, vol. 136, no. 5, pp. 403–409, May 2010.

[4]B. Donmez, L. N. Boyle, and J. D. Lee, "Safety implications of providing real-time feedback to distracted drivers," *Accident; analysis and prevention*, vol. 39, no. 3, pp. 581–90, May 2007.

[5]Y. Liang and J. D. Lee, "Combining cognitive and visual distraction: less than the sum of its parts.," *Accident; analysis and prevention*, vol. 42, no. 3, pp. 881–90, May 2010.

[6]W. J. Horrey and C. D. Wickens, "Examining the Impact of Cell Phone Conversations on Driving Using Meta-Analytic Techniques," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 48, no. 1, pp. 196–205, Mar. 2006.

[7]S. L. Chisholm, J. K. Caird, and J. Lockhart, "The effects of practice with MP3 players on driving performance," *Accident; analysis and prevention*, vol. 40, no. 2, pp. 704–13, Mar. 2008.

[8]H.-H. Chiang, J.-W. Perng, B.-F. Wu, S.-J. Wu, and T.-T. Lee, "The Human-in-the-loop Design Approach to the Longitudinal Automation System for the Intelligent Vehicle, TAIWAN iTS-i," *2006 IEEE International Conference on Systems, Man and Cybernetics*, pp. 383–388, Oct. 2006.

[9]M. L. Reyes and J. D. Lee, "Effects of cognitive load presence and duration on driver eye movements and event detection performance," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 11, no. 6, pp. 391–402, Nov. 2008.

[10]J. D. Lee and D. L. Strayer, "Preface to the special section on driver distraction.," *Human factors*, vol. 46, no. 4, pp. 583–6, Jan. 2004.

[11]Y. C. Liu, "Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveller information systems.," *Ergonomics*, vol. 44, no. 4, pp. 425–42, Mar. 2001.

[12]G. Vashitz, D. Shinar, and Y. Blum, "In-vehicle information systems to improve traffic safety in road tunnels," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 11, no. 1, pp. 61–74, Jan. 2008.

[13]C. L. Baldwin, "Verbal collision avoidance messages during simulated driving: perceived urgency, alerting effectiveness and annoyance.," *Ergonomics*, vol. 54, no. 4, pp. 328–37, Apr. 2011.

[14]J. Fagerlönn, "Urgent alarms in trucks : effects on annoyance and subsequent driving performance," *IET Intelligent Transport System 5,* no.4, pp. 1–24, 2011.

[15]C.-T. Lin, I.-F. Chung, L.-W. Ko, Y.-C. Chen, S.-F. Liang, and J.-R. Duann, "EEG-based assessment of driver cognitive responses in a dynamic virtual-reality driving environment," *IEEE transactions on bio-medical engineering*, vol. 54, no. 7, pp. 1349–52, Jul. 2007.

[16]C.-T. Lin, S.-A. Chen, T.-T. Chiu, H.-Z. Lin, and L.-W. Ko, "Spatial and temporal EEG dynamics of dual-task driving performance.," *Journal of neuroengineering and rehabilitation*, vol. 8, no. 1, p. 11, Jan. 2011.

[17]M. M. Spapé and D. J. Serrien, "Prediction of collision events: an EEG coherence analysis," *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, vol. 122, no. 5, pp. 891–6, May 2011.

[18]B. Metz, N. Schömig, and H.-P. Krüger, "Attention during visual secondary tasks in driving: Adaptation to the demands of the driving task," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 14, no. 5, pp. 369–380, Sep. 2011.

[19]H. Zhang, M. R. H. Smith, and G. J. Witt, "Identification of Real-Time Diagnostic Measures of Visual Distraction With an Automatic Eye-Tracking System," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 48, no. 4, pp. 805–821, Dec. 2006.

[20]D. W. Hansen and A. E. C. Pece, “Eye tracking in the wild,” *Computer Vision and Image Understanding*, vol. 98, no. 1, pp. 155–181, Apr. 2005.

[21]J. Chiou, L. Ko, and C. Lin, “Using Novel MEMS EEG Sensors in Detecting Drowsiness Application,” in *Biomedical Circuits and Systems Conference. IEEE,* no. 95, pp. 33–36, 2006.

[22]C.-T. Lin, L.-D. Liao, Y.-H. Liu, I.-J. Wang, B.-S. Lin, and J.-Y. Chang, “Novel dry polymer foam electrodes for long-term EEG measurement,” *IEEE transactions on bio-medical engineering*, vol. 58, no. 5, pp. 1200–7, May 2011.

[23] T. J. Sullivan, S. R. Deiss, G. Cauwenberghs, “A Low-Noise, Non-Contact EEG/ECG Sensor,” In *Biomedical Circuits and Systems Conference,* pp. 154–157, 2007.

[24]B. Donmez, L. N. Boyle, and J. D. Lee, “Mitigating driver distraction with retrospective and concurrent feedback,” *Accident; analysis and prevention*, vol. 40, no. 2, pp. 776–86, Mar. 2008.

[25]L. Garay-Vega, a K. Pradhan, G. Weinberg, B. Schmidt-Nielsen, B. Harsham, Y. Shen, G. Divekar, M. Romoser, M. Knodler, and D. L. Fisher, “Evaluation of different speech and touch interfaces to in-vehicle music retrieval systems,” *Accident; analysis and prevention*, vol. 42, no. 3, pp. 913–20, May 2010.

[26]D. B. Kaber, Y. Liang, Y. Zhang, M. L. Rogers, and S. Gangakhedkar, “Driver performance effects of simultaneous visual and cognitive distraction and adaptation behavior,” *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 15, no. 5, pp. 491–501, Sep. 2012.

[27]K. L. Young, E. Mitsopoulos-Rubens, C. M. Rudin-Brown, and M. G. Lenné, “The effects of using a portable music player on simulated driving performance and task-sharing strategies,” *Applied ergonomics*, vol. 43, no. 4, pp. 738–46, Jul. 2012.

[28]Y. Zhang, E. Harris, M. Rogers, D. Kaber, J. Hummer, W. Rasdorf, and J. Hu, “Driver distraction and performance effects of highway logo sign design,” *Applied ergonomics*, vol. 44, no. 3, pp. 472–9, May 2013.

[29]J. Stutts, J. Feaganes, D. Reinfurt, E. Rodgman, C. Hamlett, K. Gish, and L. Staplin, “Driver’s exposure to distractions in their natural driving environment,” *Accident; analysis and prevention*, vol. 37, no. 6, pp. 1093–101, Nov. 2005.

[30]C. Wege, S. Will, and T. Victor, “Eye movement and brake reactions to real world brake-capacity forward collision warnings-A naturalistic driving study,” *Accident; analysis and prevention*, vol.58, pp.259-270, Sep. 2013.

[31]H. Alm and L. Nilsson, “The effects of a mobile telephone task on driver behaviour in a car following situation.,” *Accident; analysis and prevention*, vol. 27, no. 5, pp. 707–15, Oct. 1995.

[32]J. Greenberg, L. Tijerina, and R. Curry, “Driver distraction: Evaluation with event detection paradigm,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1843, no. 1, pp. 1–9, 2003.

[33]K. L. Young and P. M. Salmon, “Examining the relationship between driver distraction and driving errors: A discussion of theory, studies and methods,” *Safety Science*, vol. 50, no. 2, pp. 165–174, Feb. 2012.

[34]M. Reed and P. Green, “Comparison of driving performance on-road and in a low-cost simulator using a concurrent telephone dialling task,” *Ergonomics*, vol.42, no.8 , pp. 1015-1037, 1999.

[35]K. L. Young, M. G. Lenné, and A. R. Williamson, “Sensitivity of the lane change test as a measure of in-vehicle system demand,” *Applied ergonomics*, vol. 42, no. 4, pp. 611–8, May 2011.

[36]M. Mouloua, A. Ahern, and A. Quevedo, “The effects of iPod and text-messaging use on driver distraction: a bio-behavioral analysis,” *Work: A Journal of Prevention, Assessment and Rehabilitation*, vol. 41, pp. 5886–5888, 2012.

[37]A. Sonnleitner, M. Simon, W. E. Kincses, A. Buchner, and M. Schrauf, “Alpha spindles as neurophysiological correlates indicating attentional shift in a simulated driving task.,” *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, vol. 83, no. 1, pp. 110–8, Jan. 2012.

[38]C. Lin, L. Ko, and T. Shen, “Computational intelligent brain computer interaction and its applications on driving cognition,” *Computational Intelligence Magazine,IEEE,* vol.4, no.4, November, pp. 32–46, Nov. 2009.

[39]K. Young, M. Regan, and M. Hammer, “Driver distraction: A review of the literature,” *Distracted driving. Sydney, NSW:Australasian College of Road Safety*, pp. 379–405, 2007.

[40]M. Sodhi, B. Reimer, and I. Llamazares, "Glance analysis of driver eye movements to evaluate distraction," *Behavior research methods, instruments, & computers : a journal of the Psychonomic Society, Inc*, vol. 34, no. 4, pp. 529–38, Nov. 2002.

[41]S. Klauer, T. Dingus, and V. Neale, "The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data," *National Highway Traffic Safety Administration*, Washington D.C, Tech. Rep., Aug. 2006.

[42]P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern recognition letters*, vol. 15, no. 11, pp. 1119–1125, 1994.

[43]F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz, "Seizure prediction: the long and winding road," *Brain : a journal of neurology*, vol. 130, no. Pt 2, pp. 314–33, Feb. 2007.

[44]S. Haufe, M. S. Treder, M. F. Gugler, M. Sagebaum, G. Curio, and B. Blankertz, "EEG potentials predict upcoming emergency brakings during simulated driving," *Journal of neural engineering*, vol. 8, no. 5, p. 56001, 2011.

[45]C. Berka, D. J. Levendowski, M. N. Lumicao, A. Yau, G. Davis, V. T. Zivkovic, R. E. Olmstead, P. D. Tremoulet, and P. L. Craven, "EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks," *Aviation, space, and environmental medicine*, vol. 78, no. 5 Suppl, pp. B231–44, May 2007.

[46]F. Elliott and C. Louttit, "Auto braking reaction times to visual vs. auditory warning signals," *Proceedings of the Indiana Academy of Science*, vol.47, pp. 220–225, 2013.

[47]R. Kosinski, "A literature review on reaction time," *Clemson University*, Tech. Rep., September, 2008. Available from http://biae.clemson.edu/bpc/bp/Lab/110/reaction.htm

[48] S. Wang, C.J. Lin, C. Wu, and W. Chaovalitwongse, "Early Detection of Numerical Typing Errors Using Data Mining Techniques," *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans,* vol. 41, issue: 6, pp. 1199-1212, 2011.

Shouyi Wang received the B.S. degree in control science and engineering from Harbin Institute of Technology, Harbin, China, in 2003 and the M.S. degree in systems and control engineering from Delft University of Technology, Delft, the Netherlands, in 2005 and the Ph.D. degree in Industrial and Systems Engineering from Rutgers University, Piscataway, NJ in 2012. From 2012-2013, he was a research associate with the department of industrial and systems engineering and the Integrated Brain Imaging Center, University of Washington, Seattle. Currently, he is an Assistant Professor of Industrial and Manufacturing Systems Engineering with University of Texas at Arlington, Arlington, TX. His current research interests include data mining and pattern discovery, machine learning, intelligent decision-making systems, multivariate time-series modeling and forecasting.

Yiqi Zhang received the B.S. degree in psychology from Zhejiang University, Hangzhou, China, in 2011. She is currently working toward the Ph.D. degree with the Department of Industrial and System Engineering, State University of New York, Buffalo. Her current research interest includes the development of mathematical models of human response in the interaction with engineering systems and addressing human cognition with their application in the design of the intelligent transportation systems.

Dr. Changxu Wu (S'04-M'07) received the M.S. and the Ph.D. degrees in industrial & operational engineering from the University of Michigan-Ann Arbor, in 2004 and 2007 respectively. He is an Associate Professor of Department of Industrial and System Engineering at the State University of New York (SUNY), Buffalo, starting from August 2007. Dr. Wu directs the Cognitive System Lab at SUNY and he is interested in integrating cognitive science and engineering system design, especially modeling human cognition system with its applications in system design, improving transportation safety, promoting human performance in human-computer interaction, and inventing innovative sustainable and smart energy systems with human in the loop. He is Associate Editors for *IEEE Transactions on Intelligent Transportations Systems* and *Behaviour & Information Technology*. He has published more than 36 journal papers in the fields.

Felix Darvas received his B.S, M.S and Ph.D degrees in Physics in 1993, 1998 and 2002 from the Rheinische-Westfaelische

Technische Hochschule (RWTH) in Aachen, Germany. He was a post-doctoral research fellow and research faculty from 2002-2006 at the University of Southern California, Los Angeles, where he was working on imaging techniques and statistical analysis for MEG/EEG and molecular imaging. He is currently a Research Assistant Professor in the Department of Neurosurgery with the University of Washington, Seattle with research interests in brain machine interfaces, functional cortical networks and the human motor system.

W.A. Chaovalitwongse is an associate professor of industrial and systems engineering and radiology (joint), and a core member of the integrated brain imaging center (IBIC) at the University of Washington, Seattle. His research interests include optimization and machine learning in neurophysiological and imaging data. Chaovalitwongse has a PhD in industrial and systems engineering from the University of Florida, Gainesville. He is a senior member of IEEE.