Combined Linear Regression and Quadratic Classification Approach for an EEG-Based Prediction of Driver Performance

Gregory Apker, Brent Lance, Scott Kerick, and Kaleb McDowell

Human Research and Engineering Directorate
U.S. Army Research Laboratory
Aberdeen Proving Ground, MD 21005, USA
{gregory.apker.ctr,brent.j.lance.civ,
scott.e.kerick.civ}@mail.mil
Kgm8@cornell.edu

Abstract. Electroencephalography (EEG) has been used to reliably and non-invasively detect fatigue in drivers. In fact, linear relationships between EEG power-spectral estimates and indices of driver performance have been found during simplified driving tasks. Here we sought to predict driver performance using linear regression in a more complex paradigm. Driver performance varied widely between participants, often varying greatly within a single driving session. We found that a non-selective linear regression model did not generalize well between periods of stable and erratic driving, yielding large errors. However, prediction errors were significantly reduced by training a linear regression model on stable driving for each participant. To provide a confidence estimate for the stable driving model, a quadratic discriminate classifier was trained to detect the transition from stable to erratic driving from the EEG power-spectra. Combined, the regression model and classifier yielded significantly lower prediction errors and provided improved discrimination of poor driving.

Keywords: EEG, Regression, Driving, Fatigue, Power Spectral Density.

1 Introduction

Fatigue and drowsiness are among the primary contributors to vehicular accidents, being estimated to have contributed to between 40-90% of all accidents [1-2]. In fact, a 2005 poll conducted by the National Sleep Foundation found that about 60% of adult drivers admitted to getting behind the wheel in a drowsy or fatigued state [3]. As a result, the prevention of these accidents has become a major focus of driver safety research.

To date, many systems have been designed to detect driver fatigue. Typically, these systems have relied on vehicle mounted sensors which correlate certain behaviors, such as vehicle dynamics, driver posture, or eye-blinking characteristics [4-6]. However, recent research has argued that monitoring the neural correlates of fatigue using electroencephalography (EEG) may provide a more reliable estimate of driver

D.D. Schmorrow and C.M. Fidopiastis (Eds.): AC/HCII 2013, LNAI 8027, pp. 231–240, 2013.

fatigue [7-8]. Further, a number of these studies have found significant correlations between neural signals and fatigue (see Lal and Craig, 2001, for a review [9]). Intriguingly, the results of these studies have varied almost as widely as their respective tasks, suggesting many differing assessments of the influence of fatigue on neural signals [10], leading to the conclusion that the specific influence of fatigue is task dependent [11-12].

Nonetheless, the observation of measureable changes in brain activity with fatigue has led to the development of several methods for classifying driver fatigue spanning a wide variety of classification approaches to predict the onset of fatigue [13-16], as well as discriminate between multiple levels of fatigue within a given driver [17-18]. While these works have been entirely fatigued-based, in a series of recent works, Lin and colleagues avoided the fatigue construct entirely and described a linear relationship between indices of driver performance and power-spectral estimates of EEG data [19-21]. In fact, they have shown that this simple relationship can be used to directly predict driver behavior based solely on neural activity recorded during a driving task with minimal processing of the EEG data [19].

However, the driving simulation used in their task was highly simplified. It has been shown that increases in task complexity can have a significant effect on the onset and characteristics of driver fatigue, and may be partly responsible for the diverse findings of the neural correlates of fatigue [11-12, 23]. As a result, it remains unclear how well a simple linear regression approach to driver performance prediction would translate to more complex driving tasks.

To begin to address this question, we evaluated predictions of driver performance from two linear regression models similar to that described in Lin et al. (2005a) in a more realistic driving scenario requiring participants to not only control vehicle heading but also control the speed of the vehicle and abide by posted speed limit signs. One of these models was trained on the full set of driving data during the training period while the other model only considered those points which reflect stable driving for that participant. Both models were then evaluated for the same testing data. We found significantly better performance of a linear model trained only on reasonably stable driving versus a linear model trained on the full range of behavior. Ultimately, we determined that the application of this type of performance prediction model benefits when coupled with an additional measure to diagnose changes in the relationship between power spectral estimates of EEG and driving behavior, thereby providing a confidence measure of the model prediction and insight into the driver's state.

2 Methods

2.1 Experimental Design

Participants. Eleven participants (aged from 20 to 40 years) participated in a virtual reality-based highway driving experiment. Each participant was briefed on the experimental equipment and procedures and signed an informed consent form. The voluntary, fully informed consent of the persons used in this research was obtained as required by Title 32, Part 219 of the Code of Federal Regulations and Army

Regulations 70-25. The investigator has adhered to the regulations for the protection of human participants as prescribed in AR 70-25.

Driving Simulation. Participants completed two separate driving sessions: the first, an acclimation session, lasted 15 minutes, the second experimental session consisted of 45 minutes of continuous driving. Before each session, participants provided an estimate of their fatigue level via the Karolinska Sleepiness Scale (KSS) [23]. Additionally, participants were asked to verbally report their fatigue score on this scale every 15 minutes during the second experimental session without interruption of driving.

Participants drove down a straight, infinitely long highway and were instructed to keep their vehicle as close to the center of the right-hand lane as possible. Throughout the session, after participants had maintained the vehicle within the appropriate lane for 8-10 seconds, a lateral perturbation was applied to the vehicle, causing it to begin to veer off course. The strength of the perturbation increased until the participant made a corrective steering adjustment (defined as a steering wheel deflection of 1 degree in the opposite direction of the perturbation) at which point the perturbation ceased allowing the participant to return the vehicle to center of the driving lane. The perturbation would ramp down automatically after approximately 3 seconds if no correction was made, however the participant was still required to correct the vehicle's heading and position. If the participant did not perform a corrective steering adjustment, the vehicle would continue to veer out of the lane and off the road until the vehicle was 21.9 meters outside of the lane, at which point the participant would be alerted to regain control of the vehicle via an auditory cue.

In addition to maintaining control of the vehicle's direction, participants also maintained appropriate speed for the vehicle during the testing session via accelerator and brake pedals. Participants were instructed to obey posted speed limit signs which appeared on the right-hand side of the road during the driving session. The speed limit was 45 mph for the majority of the session; however at three different points during the 45minute driving session the posted speed limit was reduced to 25 mph.

Data Collection and Analysis. Vehicle, EEG, and eye-tracker data were collected simultaneously throughout the experiment.

Vehicle Status and Performance Metrics. Vehicle status (position and dynamics) was monitored throughout each session, sampled at 90 Hz for participants 1-7 and at 100 Hz. for participants 8-11. To estimate driving performance, the vehicle's lateral deviation was calculated for entire session as the difference between the vehicle's lateral position and the center of the driving lane. To account for the tendencies of some participants to consistently position the vehicle to the right or left of the center of the lane, the median of their offset was subtracted to minimize this bias. Lane Deviation (LD) was then calculated as the absolute value of the lateral deviation throughout the driving session. LD values over the entire session were smoothed using a 90 second moving average filter with 2 second increments [19].

Electroencephalography. EEG signals were collected using a 64-channel Biosemi Active Two EEG system (Amsterdam, Netherlands), sampled at 2048 Hz and down-sampled to 256Hz off-line. Electrode impedance was kept at or below 5 MΩ. Using the embedded timing pulses and event signals, the EEG time series was synchronized with the vehicle status and driving performance data. Following this, the data was bandpass filtered to remove signals greater than 50 Hz and less than 5 Hz. The power spectral density estimates (PSD) for each channel were calculated using a 750 point Hanning window with 250 point overlap. Each channel and frequency power estimate of the 1-40 Hz bands was then smoothed with the same 90 second moving average filter used to smooth the lane deviation data, reducing variance and preserving the temporal alignment of the PSD and LD data streams. As was described in Lin et al. (2005a), correlation between PSD estimates and LD were often strongest for channels Cz and Pz leading to their selection for regression analysis. The same general trend was observed in the present study and thus the same two locations were used for performance prediction.

Eye tracking. Eye position was monitored but was not used for the analysis.

2.2 Experimental Design

Cross-Validation Preparation. The aligned EEG and vehicle data from the experimental session were split into three 15 minute blocks to train and test each prediction approach. Three-fold cross-validation was conducted such that two blocks were used to train the prediction algorithm and the remaining block was used to assess prediction performance. To eliminate overlapping data between training and testing sets, 90 seconds of the training data that abutted the testing data was removed prior to each cross-validation iteration.

Full-Data Regression. Following channel selection, principle component analysis (PCA) was then performed on the combined PSD estimates of both channels of the training session. Using these eigenvectors, both training and testing PSD estimates were projected into the component space and only the scores from the top 50 components (based on their eigenvalues) were preserved. The projected PSD data of the training set were used to calculate the coefficients of a 51 parameter (50 component vectors + offset) linear regression model of lane deviation. These coefficients were then applied to projected PSD testing data to generate a prediction of LD over this period. The predicted and measured LD values were compared for each epoch to characterize the predictive accuracy of the algorithm. This was repeated three times-- once for each cross-validation block.

Stable-Driving Regression. Driving performance varied widely not only between participants, but also within a single driving session for several of them. This high degree of variability resulted in dramatically different regression coefficients for a single participant's driving behavior depending upon which period of the driving session is used to train the model. To generate a regression model that yields more stable performance across an experimental session, we attempted to fit a linear regression model to a narrower subset of the driving performance data that reflected the more consistent driving epochs. To accomplish this, we defined a behavioral threshold for stable regression based on each participants individual driving habits:

$$\text{"Stable-Driving" Threshold} = \widetilde{LD}_i + 0.5\sigma_{LD,i} \tag{1}$$

where $\widetilde{LD}_i$ and $\sigma_{LD,i}$ are respectively the median and standard deviation of lane deviation during the training period for participant i. A linear regression model was subsequently calculated using only the LD data and PSD estimates from the indices in which LD values were below this threshold. In essence, LD values below this threshold represents a regime of stable driving performance; henceforth we refer to this sub-threshold performance as "stable driving".

Performance Classification and Confidence Estimate. Given that the "stable-driving" regression model was trained on a subset of the data when performance was generally consistent, the reliability of the model is somewhat limited during the periods of less stable driving where the shift in the behavior may be accompanied by a shift in the natural relationship between PSD estimates and driving performance. In these cases, the predictions of the stable-driving model may be suspect. It would be useful to be able to predict when the participant's performance may be deviating from stable driving based on the patterns observed in the PSD estimates, thereby providing an estimate of the confidence in the predictions of the stable-driving model. To accomplish this, a Quadratic Discriminate Classifier (QDC) was developed by assigning sub-threshold epochs to one class and supra-threshold epochs as another class. The PSD data fed into the classifier was treated identically to that used for regression with the one exception that the PSD estimates were smoothed with only a 4 second sliding window with 2 second steps. This was done to preserve a higher degree of sensitivity of the classifier to more rapid changes in PSD. Based on the PSD data of the testing period, if the QDC predicted stable driving conditions, we considered the predictions of the stable-driving model to be valid. However, if a transition to supra-threshold driving (class 2) was predicted by the QDC with 95% confidence, we considered the stable driving model's estimates to be invalid. In this way, the QDC serves as binary estimate of our confidence in the stable driving model.

Statistical Analysis. To compare predictive performance between models, regression coefficients, mean squared error (MSE) of the prediction, and identification of supra-threshold LD values within a given participants and block were compared directly using a paired Wilcox test unless otherwise stated. Significance threshold was set to a p-value of 0.05.

3 Results

3.1 Driving Performance

Driving performance varied greatly between participants as some participants maintained a high level of control of their LD whereas others exhibited periods of large LD or highly variable driving performance. For instance, 3 of the 11 participants' average LD did not exceed 0.5 meters, whereas 3 other participants produced averaged LD in excess of 2 full lanes outside of the correct lane.

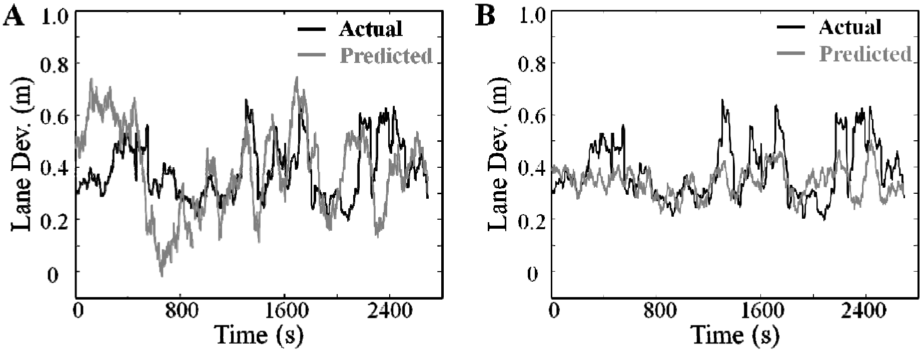

Fig. 1. Predicted Lane deviation for Full data and stable-driving data. Actual (black line) and predicted driver performance (grey lines) are drawn for a single driver for the full-data model **(A)** and the stable-driving model**(B)**.

3.2 Prediction of Driving Performance

Figure 1A illustrates the predicted LD for a single participant generated by the linear algorithm trained on the full set of driving data for each cross-validation block. For this participant, predicted performance varied widely across the driving session: In some portions of the session the predictions were generally accurate, while for other large portions, the prediction errors were very large. This trend was observed across all participants regardless of driving performance. The intra-session variability suggests that the linear relationship between the projected PSD data and driving performance was not consistent throughout the course of a single experiment.

The stable-driving model produced predictions of LD which were generally more accurate and less erratic across the entire session for each participant. Figure 1B illustrates the predicted LD from the stable-driving model for the same participant seen in Fig. 1A. In the figure, while the predictions of the model tended to under-estimate the full extent of driving error for large measured LD values, the predicted LD values appear to match the general trends and overall average LD levels. Importantly, this model did not produce large, inaccurate swings in predicted LD common in the full-data method. While predictions were generally more consistent across the experimental session using this approach, the correlation coefficients of model prediction did not significantly differ between approaches across the population.

The prediction errors of the full-data model increased as the variance increased in the training data. In fact, prediction errors of this model were significantly and positively correlated with the variability of the LD in the training data ($R=0.34$, $p< 0.05$), suggesting that model performance worsens when it is trained on a wider distribution of driving performance, perhaps because this additional variability cannot be explained by a single linear regression model. Interestingly, the prediction errors of the stable-driving model and the variability of training data were not found to be significantly correlated. Thus, in contrast to the full-data approach, the stable-driving model performance will not necessarily suffer for drivers who are more variable during the training period (or in general).

Comparing the prediction errors from each cross-validation block between models directly, the errors of the stable-driving approach were also significantly reduced relative to the original approach. The scatter plot shown in Fig. 2A indicates a significant bias for smaller prediction errors in the modified/stabilized algorithm. To quantify this, the relative improvement index (REI) was calculated as the ratio of the MSE of the full-data approach to that of the modified approach for each participant and block. An REI value greater than 1 indicates that the MSE of the full-data approach was greater than that for the stable-driving approach. As shown in Fig. 2B, a mean REI of 2.57 ± 2.83 and a median REI of 1.36 was observed and determined to be significantly greater than 1 (Wilcox signed rank test, $p<0.05$), indicating significantly smaller prediction errors in the stable-driving approach.

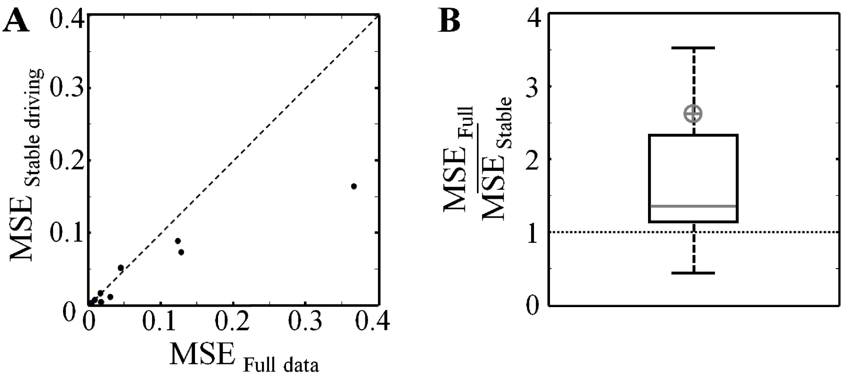

Fig. 2. Prediction error of Original vs. Modified algorithm. **A.** Scatter plot comparing the average MSE of the prediction for both linear algorithms for all participants. **B.** Relative Error Index as calculated from the MSE of each model, with mean (grey circle) and median (grey line) REI values.

3.3 Classification and Prediction of Driving Performance

In instances of large or variable LD values, the stable-driving model produced large prediction errors, indicating the linear relationship deteriorated during these periods. In these cases, the predicted LD from this model should no longer be trusted. A QDC was trained to detect changes in PSD which occurs beyond the stable-driving threshold to provide a confidence estimate for the stable-driving algorithm. Across all participants, the classifier was correctly identified the behavioral regime 88.3 ± 13% of the time. These class estimates serve primarily as an estimate of whether to trust the predictions of the stable-driving model. In the cases where the classifier predicted stable driving conditions from the PSD estimates, the average MSE of the stable-driving model predictions was 0.06 ± 0.14m across all participants. However, when the classifier predicted non-standard driving, the predicted LD of the stable driving model was always below the stable-driving threshold and had an average MSE of 0.17±0.3 m across the population, indicating the reduced confidence in the prediction was justified.

While the stable-driving model yielded significantly smaller predictions errors, the model often under-estimated large increases in LD and missed significantly more epochs of supra-threshold driving than did the linear model ($p<0.05$). This suggests the stable-driving model alone may not reliably predict when the participant begins to drive poorly. As previously described, the QDC output also serves as a predictor of supra-threshold driving. Combining the outputs of the classifier and stable-driving models may result in a more accurate estimate of sub- and supra-threshold driving epochs than the full-data model.

With respect to identifying periods of supra-threshold LD, the full-data model yielded predictive accuracy ranging from 56% to 100% across participants, with an average accuracy of 83.3 ± 15% for the population. The combined stable-driving and QDC system preformed better, yielding a range between 71% and 100%, with a significantly greater average accuracy of 89.8 ±11% ($p<0.05$). In addition, the number of false positives across the population, i.e. predictions of large LDs, was significantly greater in the full-data model compared to the combined approach ($p<0.05$), with no difference in the number of false negatives between approaches. Interestingly, the combined QDC and stable-driving predictions did not out-perform the QDC predictions of supra-threshold driving alone (88.6% average accuracy). This suggests that while the linear regression can provide a higher resolution estimate of the driving performance, the classifier was necessary to more reliably predict periods of the supra-threshold driving in this scenario, even when a more accurate linear model is used.

4 Discussion

In this study, we found that a linear algorithm indiscriminately trained on participants' driving data yielded larger predictions errors than one which was trained on a subset of driving data representative of stable-driving behavior. The stable-driving regression model was generally accurate during these periods of the testing data; however, performance deteriorated during periods of less stable driving. In this experiment, large or variable LDs were associated with a lack of vigilance on the part of the driver. Thus, the inability of the stable driving model to reliably predict far beyond the stable driving regime may be evidence of a shift in the natural relationship between PSD estimates and driving performance. This may in part explain why the model trained on the full set of data was less accurate, particularly in those cases where driving performance varied greatly.

Several researchers have recently applied non-linear algorithms to classify fatigue onset from EEG data with high degrees of accuracy [16-17], and in some cases discriminate between multiple levels of fatigue [18-19]; while others have shown broader network-based shifts in neural activity associated with fatigue driving performance [21]. Thus, it is possible that the onset of fatigue is accompanied by a more complex shift in the patterns of brain activity than can be characterized by a single linear algorithm. Another explanation for this is that additional processes or events not related to fatigue and drowsiness could have affected the relationship between PSD estimates and driver performance. Fatigue is only one of many physiological constructs which can affect driving behavior and alter neural activity. A system designed to predict driver performance in the real-world must be equipped to manage or anticipate these factors to allay their affects.

Given this, we hypothesized that the predictions of a linear model may be complemented by a secondary means to detect when a shift in the relationship may occur. Here, a quadratic discriminate classifier was trained to detect a change in the patterns of PSD data indicative of a transition in behavior (i.e. from stable- to errant-driving) in and out of a regime where a single linear model could not extrapolate to. Using this output, we were able to identify epochs where the stable-driving model produced vastly larger errors and thus indicating that the QDC provided a useful confidence estimate for the stable-driving model. In addition, we used the classifier output as an additional behavioral metric and combined those predictions with the stable-driving to accurately predict periods of poor driving at a significantly higher rate than the linear model as well as produce significantly fewer false positives. As a result, we conclude that while a linear model trained on a limited regime of stable driving behavior yields improved predictions of driving behavior, this approach is greatly benefitted by a complementary method to specifically identify non-stable driving.

A potential future application for a regression/classification system as described here is to use the classifier to toggle between regression models based on the predicted state of the driver. That is, if a different relationship between PSD and LD is found during poor driving performance (or N other behavioral regime), the information provided by the classifier could also serve to switch between regression models trained specifically for the classified driver states. Further, the accuracy and reliability of such an EEG-based system may be enhanced by leveraging other sources of information regarding the driver's state such as eye-tracking, posture etc. This multimodal approach has been shown to be effective for detecting fatigue onset in drivers [15, 24] and would be of great benefit to an automated system deciding to trust the output of one or multiple predictions of driving performance.

References

1. Treat, J.R., Tumbas, N.S., McDonald, S.T., Shinar, D., Hume, R.D., Mayer, R.E., Stanisfer, R.L., Castellan, N.J.: Tri-level study of the causes of traffic accidents. Report No. DOT-HS-034-3-535-77, TAC (1977)
2. Fletcher, K., McCulloch, S., Baulk, D., Dawson, D.: Countermeasures to driver fatigue: a review of public awareness campaigns and legal approaches. Aust. N.Z. J. Public Health 29, 471–476 (2005)
3. National Sleep Foundation. Sleep in America Poll, `http://www.sleepfoundation.org/article/sleep-america-polls/2005-adult-sleep-habits-and-styles`
4. Smith, P., Shah, M., da Vitoria Lobo, N.: Monitoring head/eye motion for driver alertness with one camera. In: Proc.15th International Conference on Pattern Recognition (ICPR 2000), Barcelona, Spain, vol. 4, pp. 636–642 (September 2000)
5. Perez, C.A., Palma, A., Holzmann, C.A., Pena, C.: Face and eye tracking algorithm based on digital image processing. In: Proc. IEEE International Conference on Systems, Man, and Cybernetics (SMC 2001), Tucson, Ariz, USA, vol. 2, pp. 1178–1183 (October 2001)

6. Popieul, J.C., Simon, P., Loslever, P.: “Using driver’s head movements evolution as a drowsiness indicator. In: Proc. IEEE International Intelligent Vehicles Symposium (IV 2003), Columbus, Ohio, USA, pp. 616–621 (June 2003)
7. Okogbaa, O.G., Shell, R.L., Filipusic, D.: On the investigation of the nerophysiological correlates of knowledge worker fatigue using the EEG signal. Applied Ergonomics 25, 355–365 (1994)
8. Lal, S.K.L., Craig, A.: Driver Fatigue: Electroencephelography and psychological assessment. Psycholphyiology 29(3), 313–321 (2002)
9. Lal, S.K.L., Craig, A.: A critical review of the pyshcophysiology of driver fatigue. Biological Psychology 55, 173–194 (2001)
10. Craig, A., Tran, Y., Witjesurya, N., Nguyen, H.: Regional brain wave activity changes associated with fatigue. Psychophysiology 49, 574–582 (2012)
11. Desmond, P.A., Matthews, G.: Implications of task-induced fatigue effects for in-vehicle countermeasures to driver fatigue. Accid. Anal. Prev. 29(4), 515–523 (1997)
12. Desmond, P.A., Matthews, G.: Task-induced Fatigue Effects and Simulated Driving. Quart. Journal of Experimental Psychology 55(2), 659–686 (2002)
13. Peiris, M.T.R., Davidson, P.R., Bones, P.J., Jones, R.D.: Detection of lapses in responsiveness from the EEG. Journal of Neural Engineering 8 (2011)
14. Stikic, M., Johnson, R.R., Levendowski, D.J., Popovic, D.P., Olmstead, R.E., Berka, C.: EEG-derived estimators of present and future cognitive function. Frontiers in Human Neuroscience 5 (2011)
15. Sandberg, D., Akerstedt, T., Anund, A., Kecklund, G., Wahde, M.: Detecting Driver Sleepiness Using Optimized Non-Linear Combinations of Sleepiness Indicators. IEEE Trans. on Intelligent Transportation Systems 12(1), 97–108 (2011)
16. Zhoa, C., Zheng, C., Zhao, M., Tu, Y., Liu, J.: Multivariate autoregressive models and kernel learning algorithms for classifying driving mental fatigue based on electroencephalographic. Expert Systems with Applications 38, 1859–1865 (2011)
17. Shen, K.Q., Ong, C.J., Li, X.P., Wilder-Smith, E.P.V.: A feature selection method multilevel mental fatigue classification. IEEE Trans. Biomed. Eng. 54(7), 1231–1237 (2007)
18. Shen, K.Q., Li, X.P., Ong, C.J., Shao, S., Wilder-Smith, E.P.V.: EEG-based mental Fatigue measurement using multi-class support vector machines with confidence estimate. Clinical Neurophysiology 119, 1524–1533 (2008)
19. Lin, C.T., Wu, R.C., Jung, T.P., Liang, S.F., Huang, T.Y.: Estimating Driving Performance Based on EEG Spectrum Analysis. EURASIP Journal on Applied Signal Processing 19, 3165–3174 (2005a)
20. Lin, C.T., Wu, R.C., Liang, S.F., Huang, T.Y., Chao, W.H., Chen, Y.J., Jung, T.P.: EEG-based drowsiness estimation for safety driving using independent component analysis. IEEE Trans. Circ. Syst. 52, 2726–2738 (2005b)
21. Chuang, S.W., Ko, L.W., Lin, Y.P., Huang, R.S., Jung, T.P., Lin, C.T.: Co-modulatory spectral changes of independent brain processes are correlated with task performance. Neyroimage 62, 1467–1477 (2012)
22. Pattyn, N., Neyt, X., Henderickx, D., Soetens, E.: Psychophysiological investigation of vigilance decrement: boredom or cognitive fatigue? Physiological Behavior 93(1-2), 369–378 (2008)
23. Akerstedt, T., Gillberg, M.: Subjective and objective sleepiness in the active individual. International Journal of Neuroscience 52, 29–37 (1990)
24. Yang, G., Lin, Y., Bhattacharya, P.: A driver fatigue recognition model based on information fusion and dynamic Bayesian network. Information Sciences 108, 1942–1954 (1942)