Jun 28th, 12:00 AM

PERCLOS: An Alertness Measure of the Past

Udo Trutschel
Circadian Technologies, Inc., Stoneham, MA

Bill Sirois
Circadian Technologies, Inc., Stoneham, MA

David Sommer
University of Applied Sciences Schmalkalden, Germany

Martin Golz
University of Applied Sciences Schmalkalden, Germany

Dave Edwards
Caterpillar, Inc., Peoria, IL

Follow this and additional works at: https://ir.uiowa.edu/drivingassessment

PERCLOS: AN ALERTNESS MEASURE OF THE PAST

Udo Trutschel[1], Bill Sirois[1], David Sommer[2], Martin Golz[2], & Dave Edwards[3]
[1] Circadian Technologies, Inc., Stoneham, Massachusetts, USA
[2] Faculty of Computer Science, University of Applied Sciences Schmalkalden, Germany
[3] Caterpillar, Inc., Peoria, Illinois, USA
Email: d.sommer@fh-sm.de

Summary: The growing number of fatigue related accidents in recent years has become a serious concern. Accidents caused by fatigue, or more precisely impaired alertness, in transportation and in mining operations involving heavy equipment can lead to substantial damage and loss of life. Preventing such fatigue related accidents is universally desirable, but requires techniques for continuously estimating and predicting the operator's alertness state. PERCLOS (percentage of eye closure) was introduced as an alertness measure. Some years later, it was claimed to be superior in fatigue detection to any other measure, including the general Eye-Tracking Signal (ETS) and even EEG recordings. This study will show that this is not the case. To put things into the prospective a fair and objective comparison between PERCLOS, the general ETS and EEG/EOG has to be established. To achieve this purpose, a protocol was established to investigate the fatigue detection capabilities of PERCLOS, ETS, and EEG/EOG in a simple two class discrimination analysis using an ensemble of Learning Vector Quantization (LVQ) networks as a classification tool. Karolinska Sleepiness Scale (KSS) and Variation of Lane Deviation (VLD) were used in order to obtain independent class labels, whereas KSS provided subjective alertness labels while VLD provided objective alertness labels. The general ETS and the fused EEG/EOG measures contain substantially greater amounts of fatigue information than the PERCLOS measures alone. These conclusions were found to be valid for all three commercially available infrared video camera systems that were utilized in the study. The data utilized in the discrimination analysis were obtained from 16 young volunteers who participated in overnight experiments in the real car driving simulation lab at the University of Schmalkalden.

INTRODUCTION

Impaired alertness is a concern for any operation that requires sustained vigilance and performance. While it is a challenging task to assess operator alertness objectively in the field, it is, nonetheless, a vitally important measure for transportation, mining operations and all other environments where lapses in attention and cognitive functioning can prove disastrous. A driver's alertness depends on time-of-day due to both homeostatic and circadian factors, on time-since-sleep (long duration of wakefulness), on time-on-task (prolonged work), inadequate sleep quality and quantity, and accumulated lack of sleep. Moreover, there are also psychological factors influencing the actual level of alertness, e.g. motivation, stress, and monotony, which is believed to play a major role in driving because it is primarily a simple lane-tracking task with a low event rate. Driver alertness is not always decreasing monotonically during driving. Rather, it shows slow waxing and waning patterns.

Many methods and technologies for objective alertness monitoring have been proposed based on measures of operator actions, such as PERCLOS, eye activity, and EEG. In 1994 PERCLOS was introduced by Wierwille and coworkers as alertness measure. Until the work of Dinges et al. several years later, PERCLOS had become only one measure for alertness among several others that had emerged. After the publishing of the Dinges report and the extensive promotion of PERCLOS using the PVT as a label, PERCLOS became the accepted standard for alertness measures. In a publication from the US Department of Transportation it was stated: "Nearly all of the technologies showed potential for detecting drowsiness by predicting lapses in at least one subject or a subset of subjects, but only PERCLOS correlated highly with PVT lapses both within and between subjects. PERCLOS not only had the highest coherence of the technologies tested, but correlated more highly with PVT lapses than did the subjects' ratings of their own sleepiness." This study concluded that PERCLOS performed better than eye blink measure, head movement, and even better than EEG. The latter claim that PERCLOS is able to detect performance lapses of the PVT better than the EEG should have raised some questions, especially since EEG is a direct measure, influenced by mainly cortical and to some degree also sub-cortical activities. PERCLOS on the other hand is a more indirect measure of eye and eyelid movements and reflects activation, such as a struggle against sleep.

An EEG-based system developed by Lal et al. (2003) was able to detect fatigue with an error rate of approximately 10%, which is remarkable considering the complexity of an alertness assessment. Similar EEG-based studies were also conducted by Golz et al. (2010) where microsleep detection errors in the vicinity of 10% were obtained, using adaptive methods with less predefined assumptions. The main drawback with physiological measurements is that electrodes have to be attached to the subjects, in this case the driver.

Concurrently, significant progress has been made in the improvement of eye-tracking technologies. Several infrared video camera systems are now commercially available (for an overview see Edwards et al. 2008). Unfortunately, most of the devices reduce the ETS to the PERCLOS measure as alertness assessment tool. This is regrettable because value of the information about the alertness state of the operator is lost. This study will show, by means of a discriminant analysis, that the fatigue information of the general ETS and in the combined EEG/EOG signals is in fact higher than that with PERCLOS.

To keep the concept of the discriminant analysis simple we introduced only two classes, 'alert' and 'fatigued'. The objective was to investigate which alertness measure was able to distinguish (with the smallest error) between the two classes. It is clear from this concept that any measures which contain no fatigue information cannot differentiate between the two classes, meaning the error would be in the vicinity of 50%, corresponding a random case. For the first type of labels, an orally spoken self-report of sleepiness on a continuous scale, the so-called Karolinska Sleepiness Scale (KSS), was recorded every two minutes during driving. The second type of labels was determined through analyzing driving performance. In previous studies it was particularly found that the Variation of Lane Deviation (VLD) correlates well with alertness and attention state of drivers (Pilutti et al. 1999). Learning vector quantization (LVQ) networks were used for the discrimination between the 'alert' and 'fatigue' classes. LVQ is a supervised neural network that was introduced by Kohonen in 1990. It allows specification of classes into which inputs can be classified.

DATA COLLECTION

The study was conducted in the Department of Adaptive Bio-Signal Analysis at the University of Applied Sciences in Schmalkalden, Germany. The study consisted of overnight driving simulation sessions as shown in Figure 1. Prior to the experiment volunteers were trained in the different tests, and were required to wear an activity monitor and to complete a sleep/wake log for at least 24 hours prior to the night time driving experiment. The subjects then arrived at the lab at 10 pm for wire-up. Logs and activity monitors were checked, retraining was conducted and the experiment sessions commenced at 11:30 pm. There were eight experiment sessions, each lasting one hour. The last session finished at 8:30 am. Volunteers were given a 1-hour break at 3:30 am. Each session included a 40-minute driving session, 10-minute CTT performance test, and 5-minute PVT. KSS were recorded as brief alertness assessments in 2-minute intervals during the driving task, as well as before and after the task. Further variables recorded included EEG (Fp1, Fp2, C3, C4, Cz, O1, O2, A1, A2, common average reference) and EOG (vertical, horizontal). In addition, ETS from three eye-tracking systems were recorded. We will not identify the vendors here since this paper is a review of the PERCLOS measure and not an evaluation of the devices. As an objective alertness label, the VDL was calculated.
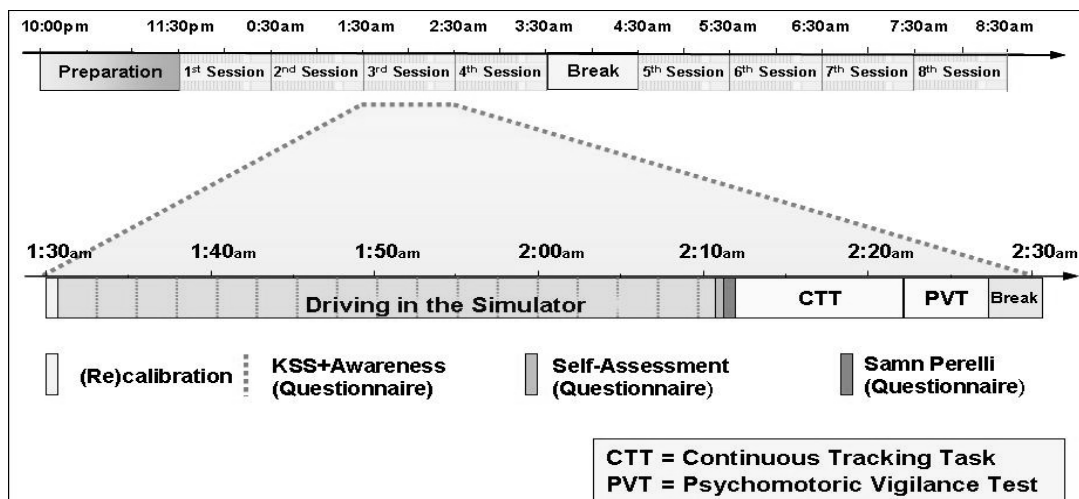

Figure 1. Experimental setup

METHODOLOGY

Feature Extraction of the PERCLOS, ETS, and EEG/EOG Data

To ensure a fair comparison of the three measures, PERCLOS, ETS, and EEG/EOG prepossessing, a feature extraction was performed according to a common concept (Golz et al. 2007). First, non-overlapping segmentation with variable time length was carried out, followed by linear trend removal and estimation of power spectral densities (PSD) utilizing the modified periodogram method. PSD values of all three signal types were averaged in spectral bands, applying optimized frequency bands. In case of ETS, EEG, and EOG data 1.0 Hz wide bands and a range of 1 - 23 Hz turned out to be optimal, whereas in case of PERCLOS, signals 0.2 Hz wide bands and a range of 0 to 4 Hz were used. All parameters were found empirically at the lowest classification errors in the test set. It is important for the EEG/EOG that the proposed frequency

features do not make any assumptions about predefined frequency bands, and that the alertness state of the operator is not determined using ratios or the power of certain frequency bands.

Labels of the Driving Session

KSS and VLD values were divided into categories 'alert' (class 1) and 'fatigued' (class 2). This allows getting labels with the corresponding data (PERCLOS, ETS, EEG/EOG) for the discriminant analysis. For the subjective label the threshold parameter was selected at KSS = 7 (Figure 2, left). For a better separation between class 1 and class 2 samples, the range of KSS = 6.9 - 7.1 was eliminated from the data set. This same procedure was applied to the objective label. Threshold was determined at VLD = 13.5%, and all samples in the range of VLD = 13.0% - 14.0% were eliminated (Figure 2, right).
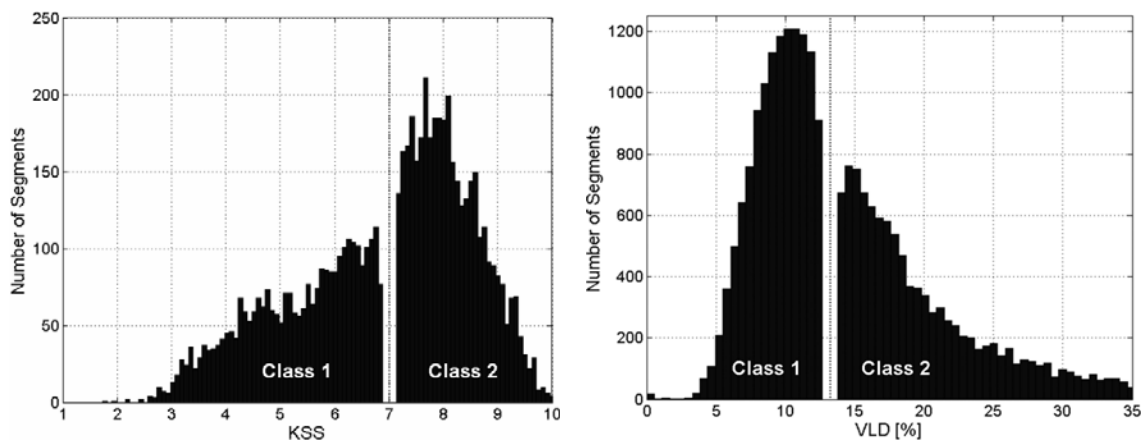

Figure 2. Distribution of the subjective label, Karolinska Sleepiness Scale (KSS) (left) and the objective label, Variation of Lane Deviation (VDL) (right) after elimination of values in the threshold region; separation leads to two classes: alert (Class 1) and strong fatigue (Class 2)

The segment length of the data was used as additional variable in order to determine the best time window for optimal fatigue detection.

Two Class Discrimination using LVQ Networks

Using an ensemble of Learning Vector Quantization (LVQ) networks for the discrimination task between the two classes is only one of several possible methods. There are some advantages to the LVQ networks which were introduced by Kohonen in 1990. LVQ neural networks are able to classify data with low error rates and good discrimination sensitivity. The LVQ output layer consists of one or several neurons which are designated to one of the two classes introduced for our task (Figure 3).
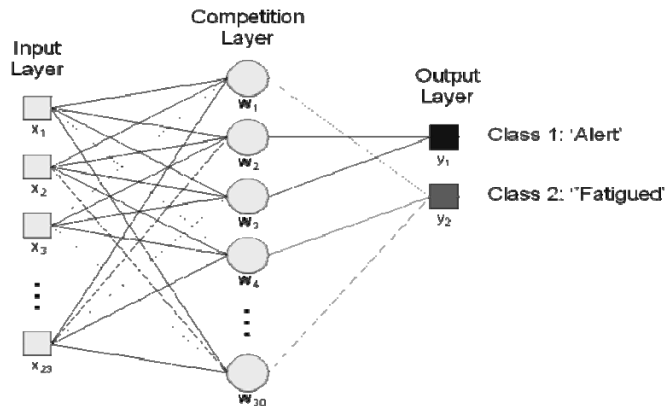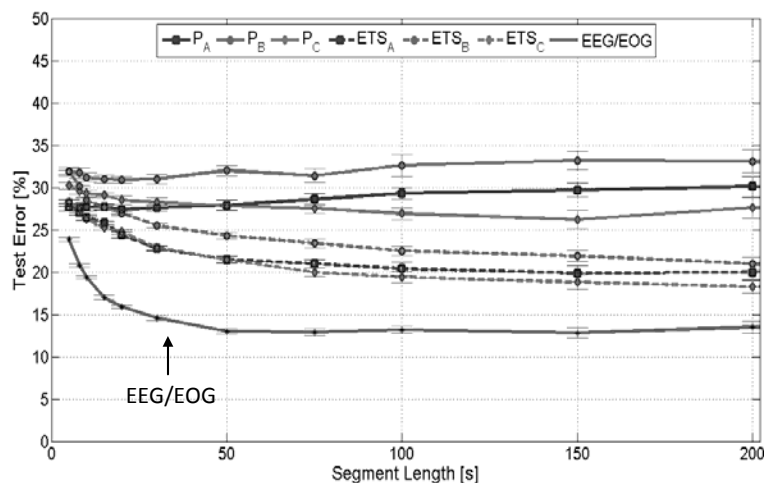

Figure 3. Architecture of the Learning Vector Quantization (LVQ) networks

This yields high sensitivity of the LVQ networks to small and unknown changes in the input feature vector. Input vectors are built from the frequencies of the signals. The supervised training procedure for one LVQ network utilized the data set (input vectors with corresponding labels) on a randomly divided ratio of 80% to 20% in each training and test set. For a given input vector, the neuron with the closest weight vector is selected as the "winner," then the winning neuron is rewarded if it belongs to the correct class (determined by the labels) and moved toward the input vector. On the other hand, if the winning neuron does not belong to the right class, it is "punished" and forced to move away the input vector. The test error obtained from the data test set provides the information about the degree of discrimination between the two classes ('alert' and 'fatigued') for the different signals (PERCLOS, ETS and EEG/EOG). A low test error represents good discrimination, and the fact that the signal contains high information about the alertness state of the operator. To get very reliable results and good comparison between the signals, the whole process is repeated for each of the 50 LVQ networks belonging to the classification ensemble. Figure 4 shows the results of the network ensemble (50 LVQ networks).

RESULTS

For simplicity, it was only asked how well different alertness measures may discriminate between the two classes ('alert' and 'fatigued'), or in other terms which signal provides the best ability to detect driver fatigue. The lowest test errors of classification are obtained for both KSS and VDL labels if features of EEG/EOG were fused together. Mean test errors of about 13% were obtained in a relatively broad range of optimal segment lengths between 50s and 300s for the subjective KSS label (Figure 4a). In case of the objective VDL label, the best mean test error achieved with EEG/EOG was 10% for segment lengths close to 150s. The PERCLOS measures of all three devices resulted in substantially higher test errors for the KSS as well as for the VLD labels. Mean test errors varied between 26% and 35% if segment lengths were larger than 50s. For both labels (KSS, VLD), the differences between the devices were roughly the same. Device 'A' performed slightly better for small segment lengths, whereas device 'C' performed slightly better for large segment lengths. The other important result of our investigation was directed to the vendors of the eye-tracking devices. They should not reduce the signal output of their devices to the PERCLOS measure. The ETS outputs of all three devices yielded lower test errors than the PERCLOS measure, but higher than the EEG/EOG. Furthermore, the differences between the devices were lower for the more general ETS measures than for the PERCLOS measures. One possible explanation would be that the PERCLOS measure is extracted slightly differently in each of the devices. Therefore, an alertness assessment algorithm based on the complete ETS output would be a substantial improvement.

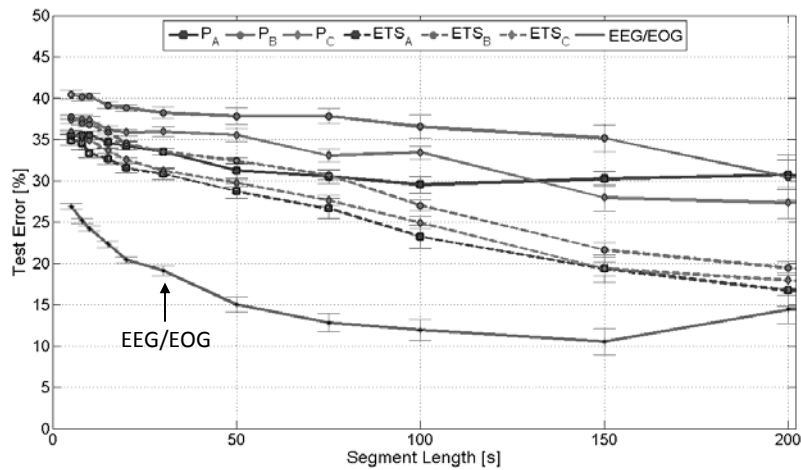**Figure 4a. Mean and standard deviation for subjective KSS label**

Figure 4b. Mean and standard deviation for objective VDL label

In general, it was found that test errors decreased with increasing segment length. This should be no surprise, because changes in the alertness level as well as in driving performance occur in the time frame of minutes. Therefore, it takes data information with segments lengths larger than 30 seconds to detect these changes. Accidents caused by microsleep events which occur on time resolution of seconds can thus not be prevented by the PERCLOS measure.

CONCLUSIONS

PERCLOS is, at the moment, the most often utilized measure of driver alertness in fatigue monitoring technologies, (e.g., infrared video cameras). The PERCLOS signal in general contains a fair amount of information about the fatigue state of an operator. One important application of these systems is the continuous alertness monitoring of operators in transportation and mining operations. The full potential of eye-tracking technologies can and should be used.

Results of experimental investigations and subsequent adaptive data analysis yielded substantial differences in the usefulness of electrophysiological signals (EEG/EOG), compared to an oculomotoric signal which was reduced to the PERCLOS measure. This main result is basically similar for the subjective label (KSS) as well as objective label (VLD). Mean test errors of 13% and 10% for subjective and objective labels, respectively, show that EEG/EOG has the potential to serve as a reference standard for use in evaluating fatigue monitoring technologies. Mean test errors between 26% and 32% for subjective and objective labels, respectively, show that the PERCLOS signals seems to carry considerably less information on driver alertness than EEG/EOG. It was shown that the reduction of ETS data to the PERCLOS measure diminishes substantially the potential of the devices for alertness assessment. These results clearly contradict those reported by Dinges et al. (1998), where PERCLOS was found to be the most reliable and valid method for determination of driver alertness levels, and that EEG results were worse than PERCLOS. Based on this statement, many questions from the scientific community should have been raised. Some doubts were expressed by Johns (2003). He pointed out that under the demands of sustained attention, some sleep-deprived subjects fall asleep while their eyes remain open. But PERCLOS does not include any assessment of eye and eye lid movements to capture this phenomenon. Important dynamic characteristics which are widely accepted, such as slow roving eye movements, reductions in maximal saccadic velocity, or velocity of eye lid reopening,

177

are ignored in the PERCLOS measure as well. However, their spectral characteristics were identified in our analysis which may account for the far better results of EEG/EOG data fusion presented here. Note that highly dynamical alterations are better reflected by ETS than by PERCLOS. Fortunately, this has been corrected in recent eye data algorithm (e.g. Schleicher et al. 2007). It is therefore necessary to analyze the complete ETS as fatigue detection tool in commercially available infrared video systems. The current results confirm the doubts stated in Johns publication (2003) and clearly show the limitations of PERCLOS. Some cautions should be considered when operator alertness is estimated by relying solely on PERCLOS.

In general, the goal of many researchers to reduce such a complex issue to a simple threshold parameter (Dinges et al. 1998) was clearly misguided. The era of alertness assessment based solely on a single measure such as PERCLOS should be in the past. To achieve the ultimate goal of reducing the number of costly driving accidents, other aspects such as mental workload, operator distraction, and of course inter-individual differences have to be taken into account. All of these considerations require a multi-sensor approach from appropriate hardware, along with the application of data fusion concepts from the software side. Therefore, additional measures beyond the ETS should be considered. In particular, operator based signals such as PERCLOS, ETS, heart rate, head motions, etc. should be combined with vehicle based measures such as steering wheel movements, time to lane crossing, and heading errors, etc.

REFERENCES

Dinges, D., Grace, R., (1998). PERCLOS: A Valid Psychophysiological Measure of Alertness As Assessed by Psychomotor Vigilance. TechBrief NHTSA, Publication No. FHWA-MCRT-98-006.

Edwards, D., Sirois, B., Davis, W., Dawson, T., Trutschel, U., Aguirre, A., Sommer, D., (2008). Operator Fatigue, Detection Technology Review. Caterpillar.

Golz, M., Sommer, D., Chen, M., Mandic, D., Trutschel U., (2007). Feature Fusion for the Detection of Microsleep Events. Journal of VLSI Signal Processing 49, 329–342.

Golz, M., Sommer, D., Trutschel, U., Sirois, B., Edwards, D., (2010). Evaluation of fatigue monitoring technologies, Somnologie 14, 187-199.

Johns, M., (2003). The Amplitude-Velocity Ratio of Blinks: A new Method for Monitoring Drowsiness. Sleep, vol. 26, pp. A51-52.

Kohonen, T., (1990). Improved versions of learning vector quantization, Int. Joint Conf. Neural Networks I, pp. 545-550.

Lal, S., Craig, A., (2002). Driver fatigue: electroencephalography and psychological assessment. Psychophysiology, 39, pp. 313–321.

Pandya, A.S., Macy, R.B., (1996). Pattern Recognition with Neural Networks in C++. CRC Press.

Pilutti, T., Ulsoy, G., (1999), Identification of Driver State for Lane-Keeping Tasks, IEEE Transactions on Systems, Man, and Cybernetic, Part A: System and Humans, vol. 29, pp. 486-502.

Schleicher, R., Galley, N., Briest, S., Galley, L., (2007), Looking Tired? Blinks and Saccades as Indicators of Fatigue. Ergonomics 51, 982-1010.

Wierwille, W., Ellsworth, L.A,, Wreggit, S., Fairbanks, J., Kirn, L., (1994). Research on vehicle-based driver status/performance monitoring: development, validation, and refinement of algorithms for detection of driver drowsiness. National Highway Traffic Safety Administration Final Report: DOT HS 808 247.