Biosignal Based Discrimination between Slight and Strong Driver Hypovigilance by Support-Vector Machines

David Sommer[1], Martin Golz[1,2], Udo Trutschel[2,3], and Dave Edwards[4]

[1] University of Applied Sciences Schmalkalden
Faculty of Computer Science, Schmalkalden, Germany
[2] Circadian, Stoneham, Massachusetts U.S.A.
[3] Institute for System Analysis and Applied Numerics, Tabarz, Germany
[4] Caterpillar Inc., Machine Research, Peoria, Illinois U.S.A.

Abstract. In the area of transportation research, there is a growing need for robust and reliable measures of hypovigilance, particularly due to the current volume of research in the development and validation of Fatigue Monitoring Technologies (FMT). Most of the currently emerging FMT is vision based. The parameter Percentage of Eyelid Closure (PERCLOS) is used for the fatigue detection. The development and validation of PERCLOS based FMT require an independent reference standard of drivers' hypovigilance. Most approaches utilized electrooculography (EOG) and electroencephalography (EEG) combined with descriptive statistics of a few time or spectral domain features. Typically, the power spectral densities (PSD) averaged in four to six spectral bands is used for fatigue characterization. This constricted approach led to sometimes contradicting results and questioned the validity of the EEG and EOG as gold standard for driver fatigue, wrongly as we will show. Here we present a more general approach using generalized EEG and EOG PSD features in combination with data fusion and advanced computational intelligence methods, such as Support-Vector Machines (SVM). Biosignal based discrimination of driver hypovigilance was performed by independent class labels which were derived from Karolinska Sleepiness Scale (KSS) and from variation of lane deviation (VLD). The first is a measure of subjectively self-experienced hypovigilance, whereas the second is an objective measure of performance decrements. For simplicity, two label classes were discriminated: slight and strong hypovigilance. The discrimination results of PERCLOS were compared with results from single and combined EEG and EOG channels. We conclude that EEG and EOG biosignals are substantially more suited to assess driver's hypovigilance than the PERCLOS biosignals. In addition, computational intelligence performed better when objective class labels were used instead of subjective class labels.

1 Introduction

Both distracted and fatigued driving accidents are thought to be underreported. Unlike alcohol related accidents where assessing the driver's state is relatively straight forward, there are no similarly objective means of ascertaining the driver's state of vigilance following a distraction or fatigue related accident. Unless the driver admits distraction or

J. Filipe, A. Fred, and B. Sharp (Eds.): ICAART 2009, CCIS 67, pp. 177–187, 2010.
© Springer-Verlag Berlin Heidelberg 2010

fatigue as a cause, one can only infer the driver's state from the physical evidence at the accident scene (Sirois et al 2007). Most drivers are reluctant to admit distraction or fatigue because they may fear being assigned blame for the incident. Therefore, the objective determination of driver's hypovigilance and distraction through the use of FMT systems could greatly reduce the occurrence of fatigue related incidents by informing drivers of their own level of vigilance. Vigilance describes the ability to sustain attention that is required for people to perceive and interpret random, relevant changes in the environment, in order to make effective decisions and perform precise motor actions. Hypovigilance is a deficit of vigilance. Two major causes of hypovigilance are central fatigue and task monotony. But, it is well known that several other factors influence driver's hypovigilance. It is a complex issue with several facets (Leproult et al 2002, Trutschel et al 2006).

Driver's hypovigilance depends for example on time-of-day due to the circadian rhythm, on time-since-sleep (long duration of wakefulness), on time-on-task (prolonged work), inadequate sleep, and accumulated lack of sleep. The last two factors may be caused by pathological sleepiness due to diseases, like sleep apnea or narcolepsy, or may be caused by intentionally sleep loss due to prolonged time awake. Moreover, there are also psychological factors influencing the actual level of vigilance, e.g. motivation, stress, and monotony. The last is believed to play a major role in driving, because it is mostly a simple lane-tracking task with a low event rate. Therefore, hypovigilance is considered as a psychophysiological variable not always decreasing monotonically during driving. It shows slow waxing and waning patterns, which can be observed in driving performance and repeatedly self-reported sleepiness.

There are many biosignals which contain more or less information on hypovigilance. Among them, EEG is a relatively direct, functional reflection of mainly cortical and to some low degree also sub-cortical activities. EOG is a measure of eye and eyelid movements and reflects activation / deactivation as well as regulation of the autonomous nervous system.

Until recently, for the assessment of driver's hypovigilance the analysis of EEG and EOG was based on a variety of definitions involving PSD summation in a few spectral bands which proved in clinical practice. The same applies to the location of EEG electrodes. Separate analysis of EEG of different electrodes and of alternative definitions of spectral bands led to inconsistent and sometimes contradicting results. Large inter-individual differences turned out to be another problematic issue.

Therefore, adaptive methods with less predefined assumptions are needed for comprehensive hypovigilance assessment. Here we propose a combination of different brain (EEG) and oculomotoric (EOG) signals whereby parameters of pre-processing and summation in spectral bands were optimized empirically. Moreover, modern concepts of discriminant analysis such as computational intelligence and concepts of data fusion were utilized. Using this general approach ensures optimal information gain even if unimodal data distributions are existent (Golz et al. 2007).

As a first step solution, we utilized SVM in order to map feature vectors extracted from EEG / EOG of variable segment lengths to two, independent types of class labels. For their generation a subjective as well as an objective measure was applied.

Both reflect different facets of hypovigilance: sleepiness and performance decrements, respectively.

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 found that especially the variation of lane deviation (VLD) correlates well with hypovigilance and attention state of drivers (Pilutti et al. 1999). For the discrimination task, a total of 10891 biosignal segments with the corresponding labels were selected. In 3611 cases the labels indicated low KSS and low VLD (class 1—class 1), in 3746 cases the labels indicated high KSS and high VLD (class 2—class 2). Here the labels of two selected facets of hypovigilance were in agreement. But for 1922 cases KSS was high and VLD was low (class 2—class 1) and for 1611 cases KSS was low and VLD was high (class1—class 2) showing that the two facets of hypovigilance are pointing in opposite directions. The disagreements between the subjective and objective labels are caused mainly by inter-individual differences of drivers regarding the ability to tolerate extreme fatigue and still keeping a safe performance level.

2 Methods

2.1 Experiments

16 participants drove two nights (11:30 p.m. - 8:30 a.m.) in our real car driving simulation lab. One overnight experiment comprised of 8 x 40 min of driving. EEG (FP1, FP2, C3, Cz, C4, O1, O2, A1, A2) and EOG (vertical, horizontal) were recorded at a sampling rate of 256 Hz. PERCLOS as another oculomotoric measure was recorded utilizing an established eye tracking system at a sampling rate of 60 Hz. Also several variables of driving simulation, like e. g. steering angle and lane deviation, were recorded at a sampling rate of 50 Hz. Especially, variation of lane deviation (VLD) is a good measure of driving performance and is used here as an objective and independent measure of hypovigilance as described below. VLD is defined as the difference between two subsequent samples of lane deviation normalized to the width of lane. For example, moving the car from the left most to the right most position of the lane results in VLD = 100 %. The KSS was mentioned above and is a standardized, subjective, and independent measure of hypovigilance on a numeric scale between 1 and 10. KSS was asked at the beginning and after finishing driving. During driving only relative changes in percent of the full range were asked because subjects are more aware of relative than on absolute changes.

2.2 Feature Extraction

To allow a comparison of the selected biosignals regarding hypovigilance, pre-possessing and feature extraction were performed due to the same concept for all biosignals (Golz et al. 2007). First, non-overlapping segmentation with variable segment length was carried out, followed by linear trend removal and estimation of power spectral densities (PSD) utilizing the modified periodogram method. Other estimation techniques,

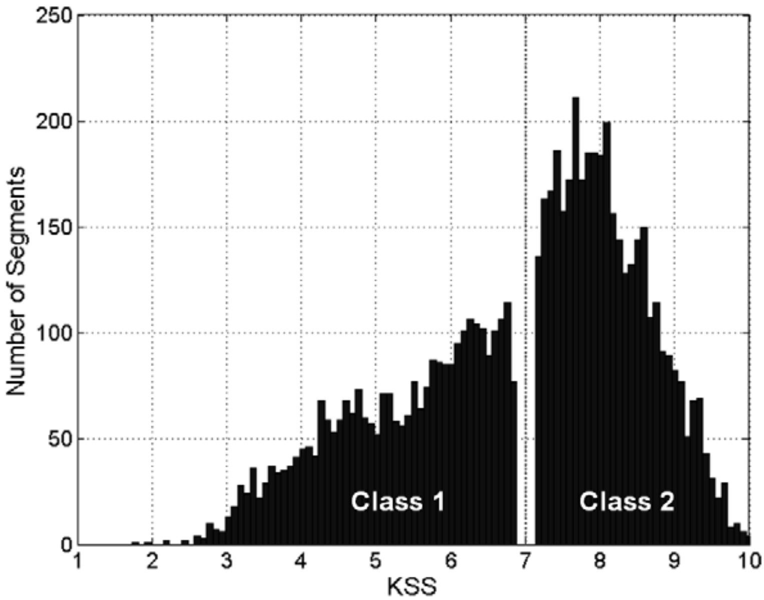

Fig. 1. Histogram of subjective ratings of sleepiness (KSS). Binarization leads to two classes: slight (class 1) and strong hypovigilance (class 2). Values in the immediate threshold region (around KSS=7) were eliminated.

such as Welch's method, the Multi-Taper method, and a parametric estimation (Burg method), were also applied, but resulted in slightly higher discrimination errors. It seems that these three methods failed due to reduced variance of PSD estimation at the expense of bias. In contradiction to explorative analysis, machine learning algorithms are not such sensitive to higher variances. Second, PSD values of all three types of signals were averaged in spectral bands. In case of EEG and EOG signals 1.0 Hz wide bands and a range of 1 to 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 optimal. All parameters were found empirically at lowest discrimination errors in the test set. Further improvements were achieved, but only in case of electrophysiological features, by applying a monotonic, continuous transform log(x).

2.3 Classification

KSS and VLD values were divided into categories 'slight hypovigilance' (class 1) and 'strong hypovigilance' (class 2). This was necessary to get labels for discriminant analysis (classification). For the subjective measure the threshold parameter was selected at KSS = 7 (Fig. 1). For a better visualization of separation between class 1 and class 2 samples in the range of KSS = 6.9 ... 7.1 were eliminated from data set. This step

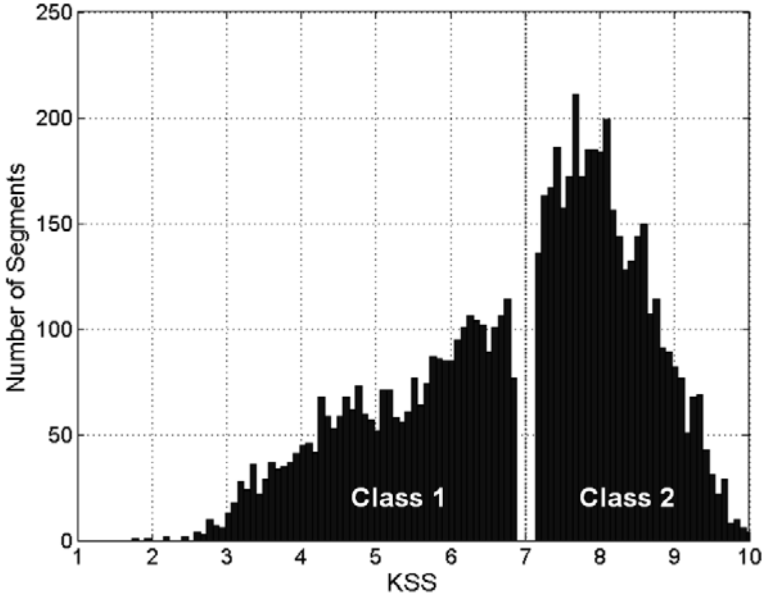

Fig. 2. Histogram of objectively measured performance (VLD). Binarization leads to two classes: slight (class 1) and strong hypovigilance (class 2). Values in the immediate threshold region were eliminated.

turned out to be not crucial. Results of classification (test set errors) showed not much of a difference.

The same binarization was applied also to the objective measure. Threshold was determined at VLD = 13.5 % and all samples in the range of VLD = 13.0 % ... 14.0 % were eliminated (Fig. 2). This data elimination also turned out to be not crucial.

Segment length was always optimized (see below) in order to get minimal test errors. Test errors were estimated by multiple, random cross validation (80 % training / 20 % test set). Due to the relatively high dimensionality of the feature space a powerful machine learning method, the Support-Vector Machine (SVM), was applied. SVM adapts an optimal separating hyperplane without any presumptions on data distribution. To achieve nonlinear discriminant functions special kernel functions have to be applied. Among several others, kernel functions such as radial basis function $k(\mathbf{x}_1, \mathbf{x}_2) = \exp\left(-\gamma \left\|\mathbf{x}_1 - \mathbf{x}_2\right\|^2\right)$ and the Coulomb function $k(\mathbf{x}_1, \mathbf{x}_2) = \left(1 + \gamma \left\|\mathbf{x}_1 - \mathbf{x}_2\right\|^2\right)^{-d}$ performed best in our application. Three SVM parameters (slack variable, two kernel parameters) were optimized carefully which requested high computational load (Golz et al. 2007). For each of the selected biosignals the segment length was varied in the range of 10 to 300 seconds to find an empirical optimum of the discrimination test error utilizing multiple hold-out cross validation. In general, small segment lengths lead to a high number of input vectors following to higher complexity presented to the discrimination algorithms and therefore to higher error rates for all signals.

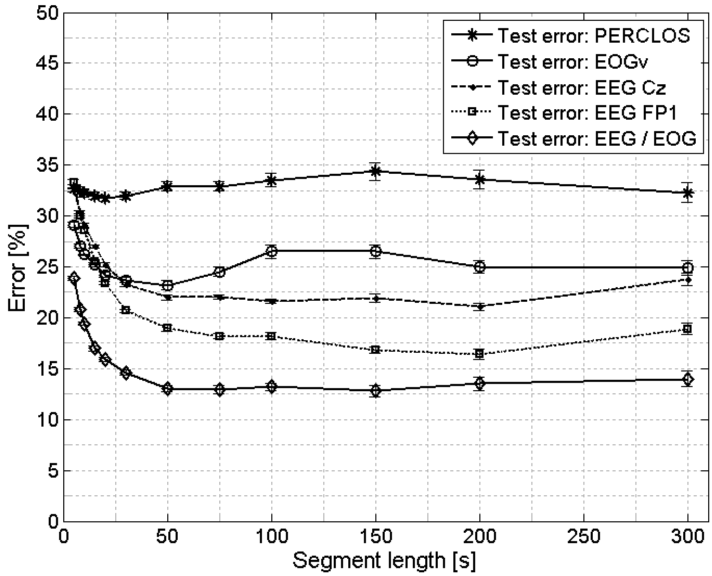

Fig. 3. Mean and standard deviation of test set errors for selected biosignals. PSD of PERCLOS had the lowest discrimination ability (largest errors), whereas PSD feature fusion of EEG and EOG performed best (lowest errors). Class labels were subjective (KSS).

3 Results

Discriminant analysis (classification) of different biosignals resulted in different errors for KSS labels (Fig. 3) and for VLD labels (Fig. 4). For the KSS labels, the PERCLOS signal and the vertical component of EOG (EOGv) showed relatively high errors and depend in similar manner on segment length. EEG at location 'Fp1' showed lower errors for all segments length compared to EEG at location 'Cz'. The feature fusion of EEG at all 7 locations and of both EOG components resulted in lowest errors. This confirms our previous finding (Golz et al. 2007) that feature fusion of EEG and EOG lead to significant improvements in the discrimination between two classes utilizing SVM. Mean errors of about 13 % yielded in a relatively broad range of optimal segment lengths between 50 and 150 seconds. Similar results for EEG / EOG signals were found in a previous study (Golz et al. 2005). In this study, which was based on different data sets, the optimal segment length was as well between 50 to 150 seconds. Learning Vector Quantization was used instead of SVM as classification method. PERCLOS features resulted considerably worse (Fig. 3). Mean errors varied between 32 and 34 % in the whole range of segment lengths.

Slightly better, but basically comparable results yielded if the objective measure (VLD) was used as class labels. Lowest errors resulted if features of EEG and EOG were fused together (Fig. 4). Mean errors of about 10 % yielded at optimal segment lengths of about 150 seconds. PERCLOS results were considerably worse. Mean errors varied between 26 and 30 % if segment lengths were larger than 50 seconds. The

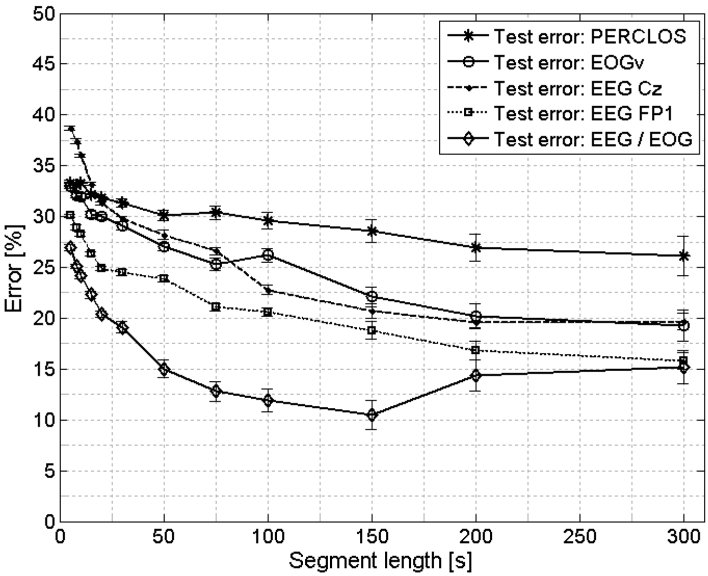

Fig. 4. Mean and standard deviation of test set errors for selected biosignals. PSD of PERCLOS had the lowest discrimination ability (highest errors), whereas PSD feature fusion of EEG and EOG performed best (lowest errors). Class labels were objective (VLD).

characteristics of the other signals EOG (vertical), EEG (Cz) and EEG (Fp1) as function of segment length is clearly more complex for the VLD labels than for KSS labels. In terms of mean errors, the achieved improvements due to feature fusion were considerable. The results (Fig. 3, 4) disclose two things: First, the driver hypovigilance detection is best for fused EEG / EOG biosignals when PSD features were utilized. This measure is suited as a benchmark to evaluate FMT devices. Second, the PERCLOS biosignal is able to detect driver hypovigilance with medium reliability.

The question arises if machine learning algorithms are able to find properties of driver hypovigilance, generally valid for all subjects under investigation. This was checked out by cross validation on the subject level. Learning algorithms were tested on all data of only one subject after they were trained on all data of all other subjects. This was repeated for every subject.

Results show high inter-individual variability (Fig. 5 , 6) not only between subjects, but also between the PERCLOS and the fused EEG/EOG biosignals, indicating that common characteristics were rarely found. Overall the inter-individual variability is larger for subjective labels (KSS) than for objective labels (VLD). This can be explained in that the subjects in our lab study were not professional drivers and could have difficulties to assess their own subjective sleepiness level using KSS. The classification errors between slight and strong hypovigilance are clearly biosignal and subject specific. Overall, the discrimination ability between the two classes is close to the optimal results only for subject '3' for EEG / EOG and for subjects '8, 15' for PERCLOS in case of subjective measures (KSS) as labels (Fig. 5). The picture differs completely when

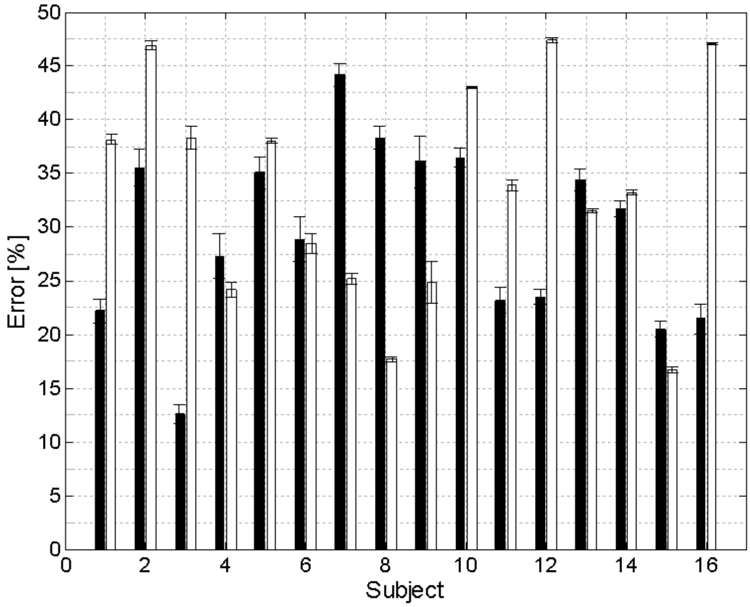

Fig. 5. Inter-individual differences of test set errors for the feature fusion of EEG and EOG (black bars) and for PERCLOS (white bars). Class labels were subjective (KSS).

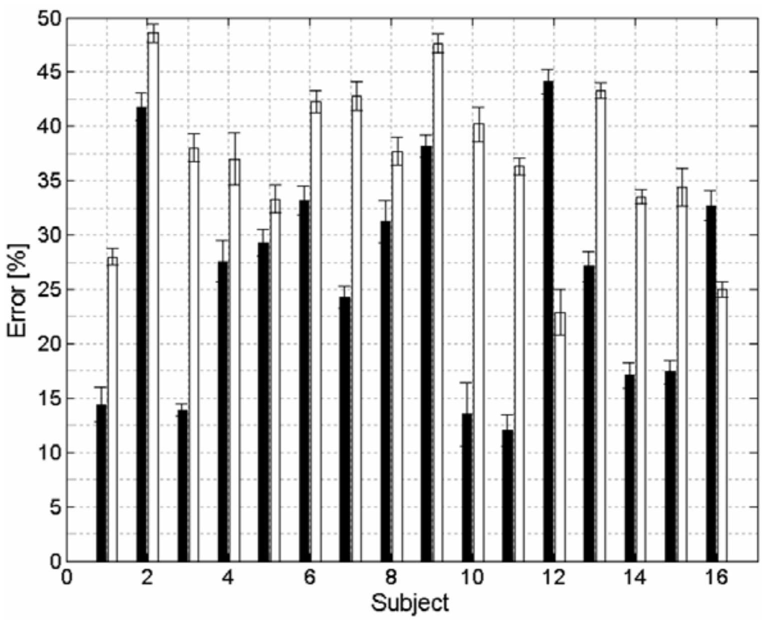

Fig. 6. Inter-individual differences of test set errors for the feature fusion of EEG and EOG (black bars) and for PERCLOS (white bars). Class labels were objective (VLD).

objective measures (VLD) were utilized as labels (Fig. 6). In this case driver hypovigilance detection is acceptable for subjects '1, 3, 10, 11' if EEG / EOG were processed. But there is no subject where PERCLOS reaches an acceptable low error level. This is somehow unexpected and questioned the validity of the application of stand alone vision based for FMT systems. In general, for most of the subjects the discrimination is limited. Test set errors are typically ranging between 15 % and 45 %, which is too high for practical applications. At the moment, it is necessary to have data of each subject also in the training set. In consequence, signal processing has to be specialised to any individual in order to get better results. Figures 5 and 6 also depict that in 9 out of 16 subjects for KSS labels and in 14 out of 16 subjects for VLD labels, the EEG / EOG feature fusion performed better than PERCLOS features.

4 Conclusions

Model free approaches are used in many different fields. Hence, it would be appropriate for the fatigue and performance research community to reach out and explore alternative algorithms beyond rule based statistical analysis of biosignals. This could help to advance the complex issue of driver hypovigilance which has eluded researchers for a long time.

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 (PERCLOS) which is at the moment the most often utilized measure of driver's hypovigilance in fatigue monitoring technologies, such as infrared video camera systems. This main result is regardless of the definition of hypovigilance, considering that subjective (KSS) as well as objective (VLD) labels has been utilized. Results were robust to different variations in parameters such as segment length which controls temporal resolution and amount of information to be involved. Mean test errors of 13 % and 10 % for subjective and objective labels, respectively, show that feature fused EEG and EOG has the potential to account for a reference standard (gold standard) to evaluate fatigue monitoring technologies (FMT). Mean test errors between 26 % and 32 % for subjective and objective labels, respectively, show that the PERCLOS signals seems to carry less information on driver's hypovigilance than fused EEG and EOG.

Our results contradict results of other authors (depicted in table 1 in Dinges et al 1998), where PERCLOS was found to be most reliable and valid for determination of driver's hypovigilance level. There, based on complete other data analyses, different measures of hypovigilance were compared. EEG resulted worse than PERCLOS, whereas measures of head position and of eye blink behaviour led to contradictory results between subjects. As a reference standard of hypovigilance they utilized measures of the well-known psychomotor vigilance task (PVT). Results are based on the fact that PERCLOS varies simultaneously with attention lapses in PVT which was repeated during 42 hours of sustained wakefulness. However, some doubts were raised (Johns 2003). It was pointed out that contradictions are possible, e. g. under demands of sustained attention some sleep-deprived subjects fall asleep while their eyes remain open. Unfortunately, PERCLOS does not include any assessment of eye and eye lid movements.

Important dynamic characteristics which are widely accepted, such as slow roving eye movements, reductions in maximal saccadic speed, or in velocity of eye lid re-opening, are ignored. Their spectral characteristics were picked up in our study through EOG and may account for the far better results of EEG / EOG data fusion presented here. Note, that highly dynamical alterations are better reflected by EOG than by PERCLOS. Our results support doubts stated in (Johns 2003) and clearly show limitations of PERCLOS. Some serious cautions should be considered when driver's hypovigilance is estimated relying solely on PERCLOS. In general, the aim of many researchers on driver's hypovigilance in the 90's to reduce such complex issue to a simple threshold parameter (Dinges et al 1998) was presumably misguiding. Fortunately, this has been corrected in recent projects. Different approaches were investigated Schleicher et al. 2007, among them also data fusion concepts (AWAKE 2004).

In addition, our previous findings (Trutschel et al. 2006, Golz et al. 2005, Golz et al. 2007) have shown that results on the assessment of driver states differ from subject to subject, as well as to some limited extent also from driving session to driving session. This was confirmed in the current investigations as well. This is a problematic issue for FMT systems. Individualization will be needed for reliable detection of driver's hypovigilance. To find practical solutions in order to address intra-individual differences in discrimination of slight and strong hypovigilance future investigations are required. For example, it could be futile to master group-average model predictions before exploring means of predicting individual hypovigilance. Due to large inter-subject variability in subjective alertness (KSS) and driving performance (VLD), it may turn out to be easier to develop reliable and accurate models of individualized measures of hypovigilance on the basis of an individual's data fusion concept than group-average vigilance models based on a single data stream.

References

1. AWAKE, System for Effective Assessment of Driver Vigilance and Warning According to Traffic Risk Estimation. In: Road Safety Workshop, Balocco, Italy (2004), `http://www.awake-eu.org/index.html`
2. Dinges, D., Grace, R.: PERCLOS: A Valid Psychophysiological Measure of Alertness As Assessed by Psychomotor Vigilance. TechBrief NHTSA, Publication No. FHWA-MCRT-98-006 (1998)
3. Golz, M., Sommer, D.: Detection of Strong Fatigue During Overnight Driving. In: 39th Annual Congress of the German Society for Biomedical Engineering (BMT 2005), Part 1, Nürnberg, Germany, pp. 479–480 (2005)
4. Golz, M., Sommer, D., Chen, M., Trutschel, U., Mandic, D.: Feature Fusion for the Detection of Microsleep Events. The Journal of VLSI Signal Processing 49, 329–342 (2007)
5. Johns, M.: The Amplitude-Velocity Ratio of Blinks: A new Method for Monitoring Drowsiness. Sleep 26, A51–A52 (2003)
6. Leproult, R., Coleccia, E., Berardi, A., Stickgold, R., Kosslyn, S.M., Van Cauter, E.: Individual differences in subjective and objective alertness during sleep deprivation are stable and unrelated. Am. J. Physiol. 284, R280–R290 (2002)
7. Pilutti, T., Ulsoy, G.: Identification of Driver State for Lane-Keeping Tasks. IEEE Transactions on Systems, Man, and Cybernetic, Part A: System and Humans 29, 486–502 (1999)

8. Schleicher, R., Galley, N., Briest, S., Galley, L.: Looking Tired? Blinks and Saccades as Indicators of Fatigue. Ergonomics 51, 982–1010 (2007)
9. Sirois, W., Dawson, T., Moore-Ede, M., Aguirre, A., Trutschel, U.: Assessing Driver Fatigue as s Factor in Road Accidents. In: Proceedings of the Fourth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, pp. 527–533 (2007)
10. Trutschel, U., Sommer, D., Aguirre, A., Dawson, T., Sirois, B.: Alertness Assessment Using Data Fusion and Discrimination Ability of LVQ-Networks. In: 10th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems (KES 2006), Bournemouth, UK, pp. 1264–1271 (2006)