Consecutive Detection of Extreme Central Fatigue

D. Sommer[1], M. Golz[1,3] and J. Krajewski[2]

[1] Faculty of Computer Science, University of Applied Sciences Schmalkalden, Schmalkalden, Germany
[2] Institute of Work and Organizational Psychology, University of Wuppertal, Wuppertal, Germany
[3] Institute for System Analysis and Applied Numerics, Tabarz, Germany

Abstract— **In order to establish fatigue monitoring technologies a valid method for automatic detection of extreme central fatigue is needed. At present, acquisition of biosignals and their analysis by computational intelligence methods are most promising. We present experiments during which 10 volunteers drove overnight in our real-car lab following a partial sleep deprivation design. Based on several biosignals (EEG, EOG) recorded during microsleep events a classifier was constructed. We have shown earlier that spectral power densities of EEG and EOG averaged in narrow bands performed best as signal features and that carefully parameterized Support-Vector Machines perform best for classification. Afterwards, classification of approximately 1.5 million consecutively segmented biosignals was performed in order to check utility for real detector application. The independent validation of this step is shown to be crucial. Two different methods based on a subjecttive and an objective measure are presented. A methodological problem remains open in how to proceed with suspicious periods where some behavioral signs point to extreme fatigue, but driving seems still to be possible.**

Keywords— **EEG, EOG, Support-Vector Machines, Fatigue, Microsleep.**

I. Introduction

Human factors are the outstanding cause of road accidents worldwide. GIDAS database (German in Depth Accident Study) reveals that 93.5 % of all accidents are caused by human factors [1]. One of the most important factors is extreme central fatigue which is estimated to be the main cause of 10 % up to 40 % of all fatalities.

Electrophysiology has the potential for establishing a reference standard needed for validation purposes of industrial devices for driver state monitoring. Results of adaptive signal processing and pattern recognition have shown earlier that it is possible to setup a detector for microsleep events (MSE) [2]. MSE are widely accepted as clearly visible behavioral states which appear under extreme central fatigue. They are defined as short intrusions of sleep into wakefulness under demands of sustained attention [3].

The main concern of this contribution is as follows. Up to now, MSE detector setup has been based on biosignals of doubtless MSE and of the same amount of counter-examples (Non-MSE). But there are a lot of periods where doubtless scoring of driver's state is not possible. This paper aims at processing of consecutively segmented biosignals. Therefore all periods will be classified. It will be shown that validation of their classification is difficult to estimate.

This is needed to validate if a real detector application is possible. At present, this methodology would be interesting as a reference laboratory standard in order to validate industrial fatigue monitoring devices which are mostly based on video analysis of oculomotoric activity and face expression.

II. Experiments

10 healthy young adults completed 7 overnight driving sessions (1 - 8 a.m.) in our real car driving simulation lab. Each session had duration of 40 min and was preceded and followed by vigilance tests and responding to sleepiness questionnaires. Reports of them will be given elsewhere. Time since sleep was at least 16 hours which was checked by wrist actometry. Subjects have been prepared beforehand by simulator training.

Several biosignals were recorded: EEG (C3, Cz, C4, O1, O2, A1, A2, com.av.ref.), EOG (vertical, horizontal), ECG, EMG (m. submentalis). In addition, three video recordings (driver's head & pose, driver's eyes, driving scene) were stored. Also several variables of the car, like e.g. steering angle and lane deviation, were sampled, but their analysis is not considered here. Further experimental details have been published elsewhere [3, 4].

III. Analysis

A. Scoring of MSE

A first judgment of ongoing MSE was done immediately during the experiments by two operators who watched the video streams. Typical signs of MSE are prolonged eyelid closures, roving eye movements, head noddings, major driving incidents and drift-out-of-lane accidents. Several other signs were observed, but we have been decided to rely not solely on them. Some examples are bursts of alpha and theta activity in the EEG, spontaneous pupil contractions and stare gaze. In all, we have found 2,290 MSE (per subject: mean number 229 ± 67, range 138 - 363).

For the detection of such events on a second by second basis a careful determination of the point in time where MSE is starting will be needed. Therefore, all recorded video material and biosignals underwent off-line scoring made by independent and trained raters. They refined results of online scoring into clear MSE, clear Non-MSE (red and blue dots, Fig. 1).

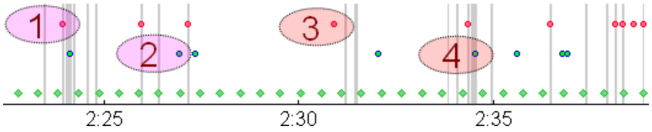


Fig. 1 An example of the outcome of visual scoring during a time span of approx. 15 min. MSE (red dots, upper row), Non-MSE (blue dots, middle row) and time points of additional scores (green dots, lower row) are visual ratings of experts. The outcome of automatic MSE classification is indicated by grey bars. Marked events are examples of conformities (1, 2) and nonconformities (3, 4) between subjective and objective detection, i. e. scoring and classification, respectively.

Later a third visual scoring performed by other trained raters was done. They labeled all periods (every 30 seconds) by different scores of the drivers state (green dots in Fig. 1). This is needed as a sample set for validation of consecutive classification (see below).

B. Pre-Processing

Empirical investigations performed earlier showed that segmentation is very sensitive to classification accuracy. Segment length should range between 4 and 12 sec [2], here we took 6 sec and 0.1 sec as step size of segmentation. Therefore, approximately 24.000 segments per driving session and 168.000 per night were obtained. Artifacts turned out to play a minor role when computational intelligence algorithms are applied for classification. Unpublished investigations utilizing Independent Component Analysis (ICA) to eliminate eye blink artifacts from EEG resulted in no improvement of MSE detection compared to the case of no artifact elimination.

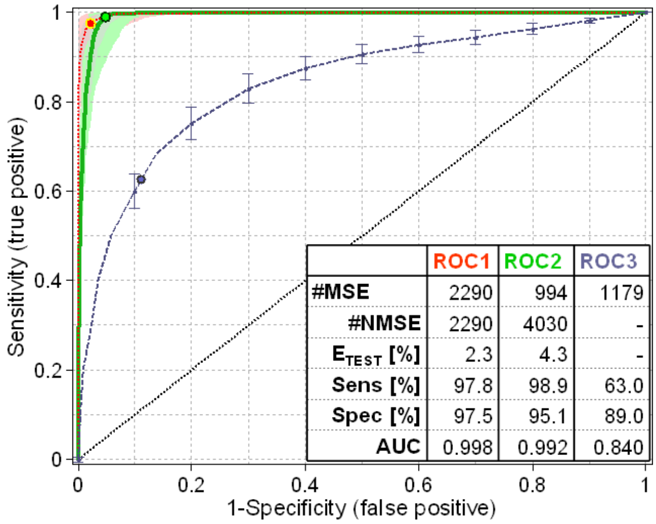
C. Feature Extraction

Several methods for extraction of translation invariant features in time, spectral and wavelet domain as well as in state space were utilized earlier. It turned out that spectral power densities estimated by the modified periodogram method are most useful for MSE classification [2]. Logarithmic scaling and summation in narrow spectral bands (width 1 Hz, range 0.5 to 23 Hz) are necessary to minimize classification errors further on. The Delay-Vector Variance, which is a state space method, is useful as a complement, but is not highly important [2]. Therefore, we utilized only band-averaged log power densities.

D. Classifier Setup

Support-Vector Machines incorporating radial basis functions as kernel were most optimal in terms of minimizing empirical errors of a test set. But this has only been found correct if the hyperparameter and the slack variable are optimized carefully [4]. It must be emphasized that this preparation of a classifier was based on test set of clear examples of MSE and Non-MSE. They cover about 15 % of the whole time only. In contrast, the consecutive recall of the classifier covers 100 % of the whole time of driving.

IV. Results



	ROC1	ROC2	ROC3
#MSE	2290	994	1179
#NMSE	2290	4030	-
E_{TEST} [%]	2.3	4.3	-
Sens [%]	97.8	98.9	63.0
Spec [%]	97.5	95.1	89.0
AUC	0.998	0.992	0.840

Fig. 2 Receiver Operating Characteristic (ROC) of one subject. ROC1 was based on evident MSE and Non-MSE examples; whereas ROC2 was based on visual scores given every 30sec. 95 % confidence intervals are marked (light red, light green). ROC3 represents results of other authors. Minimal test errors (E_{Test}) are indicated (open circles). The table contains number of examples, sensitivity, specificity and area under ROC.

A. Classificator performance

Conflict-free cases (true positives, true negatives), where SVM output is MSE or Non-MSE and rater's opinion is the same (Fig. 1: marked events 1 and 2, respectively), occurred most often. Conflicts arise when SVM output is Non-MSE and rater judges MSE (false negatives; marked event 3), or SVM output is MSE and rater judges Non-MSE (false positives; marked event 4).

This result is quantified by ROC analysis (Fig. 2, red). Lowest classification error rate (conflicts in test set) is 2.3 % in the mean. True positives and true negatives were found in 97.7 % of all events. Errors increased when data of the validation set were applied (Fig. 2, green). For a given specificity, sensitivity is lower. But note that this data set is unbalanced. That's why an optimistic bias due to prior probability has to be taken into account.

For comparison, results of other authors [5] are presented (Fig. 2, blue). MSE investigations performed also during a continuous tracking task resulted in lower sensitivity and lower specificity as well as in higher variance. This performance decrease could be due to other methods for feature extraction and classification, but should mostly account to the definition of MSE. These authors defined MSE by large tracking errors which are due to performance decreases. But MSE is only one cause among other psychophysiological factors, such as lack of concentration, or aversions against the monotonous task. We believe that visual scoring of only evident examples of MSE is reliable because behavioral signs of extreme central fatigue are complex and differ largely between subjects. Therefore, it is important to observe visually the temporal development of the many behavioral signs as well as to check the driving scene video for large performance decreases.

B. Consecutive Classification

So far, we have shown that processing EEG and EOG and utilizing computational intelligence methods is successful in regard to classify MSE and Non-MSE. But as mentioned above (Sect. III D), this is true for clearly scored events. They cover only 15 % of the total length of experiments.

For consecutive classification we utilized the optimized SVM classifier (Fig. 2, red) in recall mode, which means that no further adaptation (training) was done. As described in Sect. III B, signals were segmented consecutively; afterwards features were extracted and fed as input variables to the SVM classifier. This led to a binary output variable indicating MSE or Non-MSE at a sampling rate of 10 s^{-1}. Each row in Fig. 3 shows this variable over the span of one driving session and contains 24,000 samples. The first session (upper subplot) started at 1 a.m. after at least 16 hours of wakefulness. Mainly at the end of this session a number of MSE were detected. This time-on-task effect is common in tracking experiments. At the fifth session and later (lower two subplots) all subjects suffered from extreme fatigue. A lot of MSE were detected. This over-all increase is also due to the time-since-sleep effect. The impression from Fig. 3 that most of the time in the last driving sessions subjects experienced microsleep is misleading and is due to the compressed plot. As shown in Fig. 4 mean MSE durations were not larger than 60 % of the whole driving session.

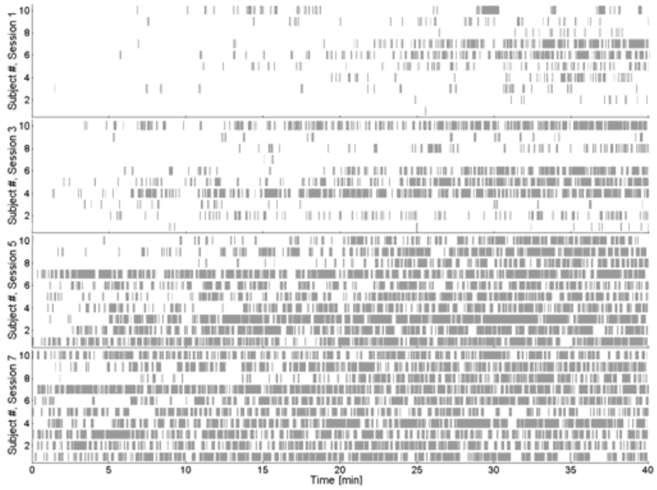


Fig. 3 Binary output of consecutive classification: MSE (grey) or Non-MSE (white) was detected. Subplots represent 5 of all 7 driving sessions. Each subplot show results of all 10 subjects (rows).

Descriptive statistics over all results of consecutive classification (Fig. 4) clearly confirms time-on-task and time-since-sleep effects. Within and between driving sessions mean durations of MSE (red bars) are increasing.

The only exception is that a slight decrease arises between 6th and 7th driving sessions. This could be caused by the time-of-day effect which is due to the habitual sleep-wake-cycle. Another effect could be a motivational one: subjects expect the end of experiments after the 7th session. Increasing mean of MSE durations has two aspects: increasing frequency and slight extensions of MSE periods.

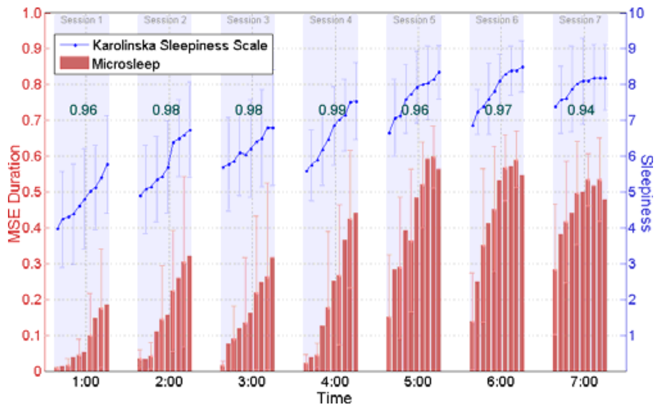


Fig. 4 Mean and standard deviation of MSE duration (red) and subjective sleepiness (blue). Averaging interval was 4 min. Strong correlations between objective and subjective measure are indicated.

In addition, subjectively experienced fatigue was asked every two minutes on the Karolinska Sleepiness Scale (1 = no fatigue, 10 = extreme fatigue, not able to stay awake) [6].

Mean and standard deviations (blue) of this purely subjective measure is strongly correlated with MSE duration which is a purely objective measure. Pearson's correlation coefficients are always greater than 0.95 except for the 7^{th} session. Note that for both measures large standard deviations emerge which is mainly due to large inter-individual differences.

V. Conclusions

We have shown that optimizing a classification algorithm empirically on a data set of only evident examples of MSE and non-MSE is successful for sensor applications where a consecutive sequence of signal segments has to be processed. The main difficulty of consecutive detection was the lack of validation because a lot of examples were to be processed where it was not clear how to label them. Many examples seemed to be Non-MSE, but some behavioral signs gave reason to doubts, e.g. stare gazes or slow sliding head movements. Otherwise, not every prolonged eye-lid closure must be MSE.

Two methods were proposed to validate the methodology. First, we engaged a further coworker to score visually at given points in time (every 30 sec) if MSE or Non-MSE appears. This way, a fully independent validation set was generated. ROC analysis (Fig. 2, ROC2) resulted in only slightly lower sensitivity and specificity compared to the test set of evident examples. Results of a comparable study of other authors showed that this match is not a matter of course.

Second, every consecutively detected MSE was counted and mean duration of MSE was computed. This was considered as an objective measure of extreme central fatigue. Well known effects in sleep psychophysiology were confirmed by this measure. Moreover, a subjective measure, the self-reported sleepiness on the standardized Karolinska Sleepiness Scale [6], correlated always strongly to this objective measure. Therefore, we conclude that consecutive and reliable detection of MSE during periods of extreme central fatigue is possible despite large intra- and inter-individual differences in behavior and in EEG and EOG characteristics [2] [7].

The question remains open if such detector application would also work during driving on real streets. Our aim was to verify MSE detection under controlled laboratory conditions. For applicative reasons this is valuable to online driver monitoring technology. Their improvement and validation necessitates an independent reference standard of microsleep detection and extreme central fatigue.

Acknowledgment

Part of this work was supported by the German Federal Ministry of Education and Research as part of the program “Applicative Research and Development at Universities of Applied Sciences”.

References

1. Gründl M (2005) Fehler und Fehlverhalten als Ursache von Verkehrsunfällen und Konsequenzen für das Unfallvermeidungspotenzial. PhD Thesis (in German), Regensburg, Germany, 2005.
2. Golz M, Sommer D, Chen M, Trutschel U, Mandic D (2007) Feature Fusion for the Detection of Microsleep Events. J VLSI Signal Proc Syst 49: 329-342
3. Overview at http://www.microsleep.de
4. Golz M, Sommer D, Holzbrecher M, Schnupp T (2007) Detection and Prediction of Driver's Microsleep Events, Proc 14th Int Conf Road Safety on Four Continents; Bangkok, Thailand, 2007
5. Davidson PR, Jones RD, Peiris MT (2007) EEG-Based Behavioral Microsleep Detection with High Temporal Resolution. IEEE Trans Biomed Engin 54: 832-839
6. Åkerstedt T (1990) Subjective and Objective Sleepiness in the Active Individual. Int J Neurosci 52: 29-37
7. Ingre M, Åkerstedt T, Peters B, Anund A, Kecklund G (2006) Subjective Sleepiness, Simulated Driving Performance and Blink Duration: Examining Individual Differences. J Sleep Res 15: 47-53

- Authors: Sommer, David; Golz, Martin
- Institute: Faculty of Computer Science, University of Applied Sciences – Schmalkalden
- Street: Blechhammer 4
- City: Schmalkalden
- Country: Germany
- Email: d.sommer@fh-sm.de; m.golz@fh-sm.de