Fusion of State Space and Frequency-Domain Features for Improved Microsleep Detection

David Sommer[1], Mo Chen[2], Martin Golz[1], Udo Trutschel[3], and Danilo Mandic[2]

[1] Univ. of Appl. Sciences Schmalkalden, Dptm. of CS, 98574 Schmalkalden, Germany
[2] Dptm.of Electr. & Electr. Engin, Imperial College, London, SW7 2BT, UK
[3] Circadian Technologies, Inc. 24 Hartwell Avenue, Lexington, MA 02421, USA

Abstract. A novel approach for Microsleep Event detection is presented. This is achieved based on multisensor electroencephalogram (EEG) and electrooculogram (EOG) measurements recorded during an overnight driving simulation task. First, using video clips of the driving, clear Microsleep (MSE) and Non-Microsleep (NMSE) events were identified. Next, segments of EEG and EOG of the selected events were analyzed and features were extracted using Power Spectral Density and Delay Vector Variance. The so obtained features are used in several combinations for MSE detection and classification by means of populations of Learning Vector Quantization (LVQ) networks. Best classification results, with test errors down to 13%, were obtained by a combination of all the recorded EEG and EOG channels, all features, and with feature relevance adaptation using Genetic Algorithms.

1 Introduction

One of the main problems associated with data fusion for real-world applications is related to combining the information coming from heterogeneous sensors, acquired at different sampling rates and at different time scales. Data/sensor fusion approaches dealing with combining data from homogeneous sensors are normally based either in the time domain, or in some transform domain, for instance on features coming from the frequency representation of signals, their time-frequency, or state-space features [1].

Notice that in this framework we deal with multivariate and multimodal processes, for which either there are no precise mathematical relationships, or if they exist they are too complex. Such is the case with the detection of lapses of attention in car drivers, due to fatigue and drowsiness, the so-called Microsleep Event. Their robust detection is a major challenge. Recent developments in this field have shown that most promising approaches for this purpose are based on a fusion of multiple electrophysiological signals coming from different sources together with Artificial Neural Networks in the detection and prediction [2-4].

In general, there are two standard approaches to combine multiple electroencephalogram (EEG) and electrooculogram (EOG) signals. In the first approach, called **Raw Data Fusion** the sensor data are merged without prior preprocessing or dimensionality reduction. Despite its simplicity, the major disadvantage here is the potentially vast amount of data to be handled. In the second approach, the so-called **Feature Fusion**, features extracted from signals coming from different sources and/or

W. Duch et al. (Eds.): ICANN 2005, LNCS 3697, pp. 753 – 759, 2005.

extracted by different methods are fused. In our investigations, the frequency domain features obtained from the Power Spectral Density (PSD) [4] and state space features obtained from the Delay Vector Variance (DVV) [5, 6], are combined in order to show whether such a combination of different features shows improvement in MSE detection over the standard approaches using only one of the signals and one class of features. The motivation for such an approach is as follows: PSD estimation is a "linear" frequency domain method which can be conveniently performed using the periodogram. This has been shown to perform particularly well in applications related to EEG signal processing, [2-4]. It is, however, natural to ask ourselves whether such an approach, based solely on the second order statistics conveys enough information to provide fast and reliable detection of such a complex event as the MSE.

On the other hand, the recently introduced DVV approach [5, 6] is a method based on the local predictability in the state space. The virtue of the DVV approach is that it can show both qualitatively and quantitatively whether the linear, nonlinear, deterministic or stochastic nature of a signal has undergone a modality change or not. This way, the DVV methodology represents a complement to the widely used linear PSD estimation. Notice that the estimation of nonlinearity associated with the DVV method is intimately related to non-Gaussianity, and we also set ourselves to investigate whether this additional information, which cannot be estimated by PSD, contributes to the discrimination ability, and if so, to estimate its importance level, as compared to the PSD based discrimination.

The purpose of this paper is therefore to provide a theoretical and computational framework for the combination of the two classes of features (PSD and DVV) and to show whether such a combination has the potential benefits for multivariate and multimodal signals over standard approaches. This is illustrated on a practical problem of detection of MSE in car drivers. Reliable methods to detect MSE in continuously recorded signals will be an important milestone in the development of drowsiness warning systems in real car cockpits. At present, however, achieving highly reliable MSE detection [3] is still a major issue to be resolved.

2 Data Fusion Architecture

To achieve the detection of MSE in real-world car driving situations, both estimated feature sets are merged by an adaptive feature weighting system (Fig. 1). The error of the training set is used to optimize the parameters of feature extraction based on PSD and DVV and also serves as fitness function in a genetic algorithm that examines the relevance of the different features employed. Subsequent multiple hold-out validation of LVQ networks yields the mean test set error for the evaluation of MSE detection ability. Consequently, test set errors were not used, directly and indirectly, for any step of optimization.

Fig. 1 shows the block diagram of the proposed data fusion system, which allows for the solution of the extensive data management problem in real time processing of multivariate and multimodal signals. Fusion based on features provides a significant advantage by means of a reduction in the dimensionality of the space in which the information to be processed resides. There is a trade-off associated with this strategy, since in principle, feature fusion may not be as accurate as raw data fusion because portions of raw signal information could have been eliminated.

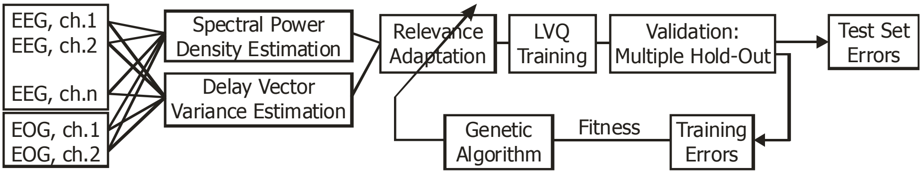


Fig. 1. Proposed Microsleep detection system based on feature fusion

In general, it is not known a priori which features within the different sets of features (EEG-FFT, EEG-DVV, EOG-FFT, EOG-DVV) are best suited for detection of MSE. It is intuitively clear that the obtained features differ in their importance level with respect to the classification accuracy. We therefore combine all different feature sets obtained from EEG and EOG by means of PSD and DVV. To prove whether our hypothesis that a combination of features coming from two different sources will indeed improve classification accuracy, we propose to use Genetic Algorithms (GA) to determine a scaling factor for every single feature coming from the four different sets. The scaling factors are used as gene expressions and the training error rate as fitness function. The sensitive adaptation of scaling factors by GA leads to a weighted Euclidean metric in the feature space, and can be interpreted as relevance factors [12]. For the purpose of comparison, the classification task is also performed without application of the relevance adaptation step (Fig. 3).

3 Experimental Setup

Our experimental setup was similar to the one presented in [4]. Seven EEG channels from different scalp positions and two EOG-signals (vertical, horizontal) were recorded from 23 young subjects (age range: 19 - 35 years) during driving simulation sessions lasting for 35 minutes. These sessions were repeated every hour between 1 a.m. and 8 a.m. This way, the likelihood of the occurrence of MSE was gradually increasing due to at least 16 hours without sleep prior to the experiment.

MSE are typically characterized by driving errors, prolonged eye lid closures or nodding-off. Towards automatic detection, two experts performed the initial MSE scoring, whereby three video cameras were utilized to record i) drivers portrait, ii) right eye region and iii) driving scene. For further processing, only clear-cut cases, where all the experts agreed on the MSE, were taken into account. Despite providing enough test data to tune our algorithms, the human experts could not detect some of the typical attention lapses, such as the one with open eyes and stare gaze. The number of MSE varied amongst the subjects and was increasing with time of day for all subjects. In all 3,573 MSE (per subject: mean number 162±91, range 11-399) and 6,409 NMSE (per subject: mean number 291±89, range 45-442) were scored. This clearly highlights the need for an automated data fusion based MSE detection system, which would not only detect the MSE also recognized by human experts, but would also offer a possibility to detect the critical MSE cases which are not recognizable by human experts.

4 Feature Extraction

In our experimental set-up, we varied two preprocessing parameters, the segment length and temporal offset. Evaluating test errors of our processing cue (Fig. 1) without relevance adaptation yields an optimal offset of 3 and 1sec and optimal segment length of 8 and 4 sec for the PSD and DVV, respectively. This means that EEG and EOG segments are beginning 3/1 sec (PSD/DVV) before and are finishing 5/3 sec after the onset of (N)MSE.

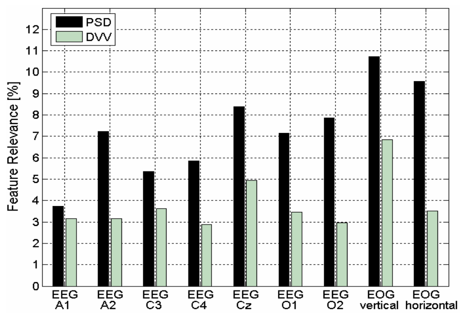
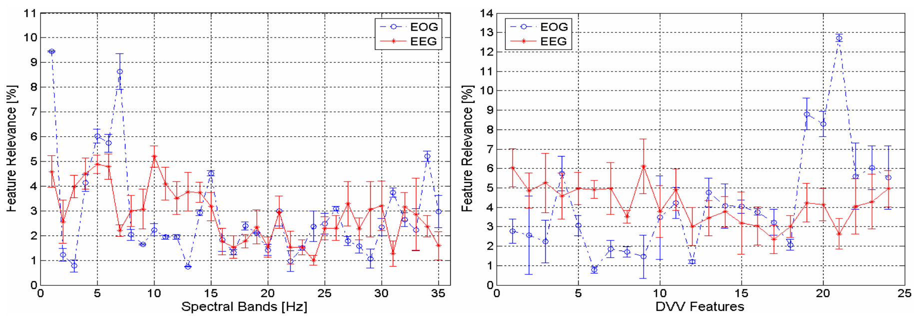


Fig. 2. Feature relevances for MSE detection estimated using GA
Fig. 2a. (left top): Normalized feature relevances for different data channels of EEG and EOG
Fig. 2b. (left bottom): PSD feature relevances over all EEG and EOG signals
Fig. 2c. (right bottom): DVV feature relevances over all EEG and EOG signals



The preprocessing involves linear trend removal and applying the Hanning window to the data segments. PSD estimation was performed by the discrete Fourier transform. The so calculated PSD coefficients were averaged within 1.0 Hz wide bands. Further improvements in classification were achieved by applying a monotonic continuous transformation log (x) to the PSD [7].

The linear feature extraction method PSD was accompanied by a feature extraction method originating from the nonlinear dynamics, the DVV. The DVV features were calculated with the embedding dimension (m=3). Basically, they are variances of distances between delay vectors calculated on original and on surrogate data; further details are presented elsewhere [5, 6]. In contrast to PSD features, classification results did not improve by applying log(x) to DVV features.

Before examining the MSE detection performance in such a feature fusion setting, we perform a rigorous analysis of the feature relevance for the different EEG and EOG signals. This was achieved by means of GA. The relevance scores for the single EEG and EOG signals (Fig 2a) were calculated and normalized using the sum over all

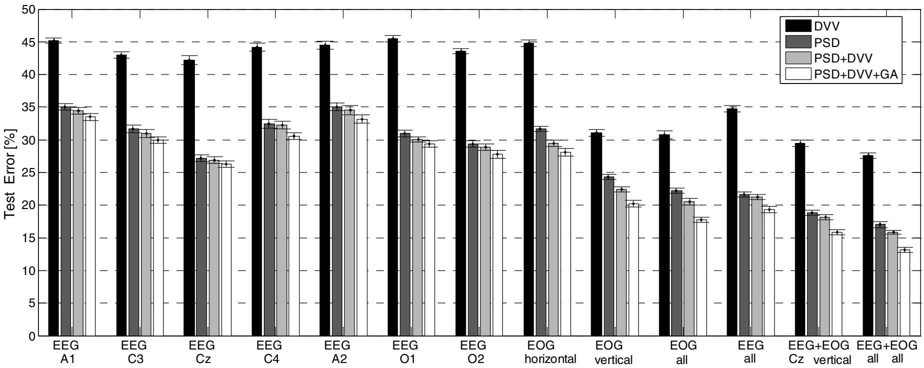


Fig. 3. Mean values and standard deviations of test errors for different single signals (first 9 groups) and for different combination of fused signals (last 4 groups)

feature relevance coefficients (35 for PSD, 24 for DVV). The normalized relevances for each PSD feature (Fig. 2b) and each DVV feature (Fig. 2c) were determined by averaging relevances of all single EEG / EOG signals.

5 Discriminant Analysis

Feature sets extracted by both methods (PSD and DVV) and of each of 7 EEG and of 2 EOG signals are merged in their multiple combinations both without and with an adaptive feature scaling system (GA). For each feature vector a label “MSE” or “NMSE” was assigned, thus introducing a two-class classification setting. Networks utilizing the OLVQ1 learning rule were used for analysis. Multiple hold-out validation [8] of the LVQ networks yields the mean test set error depicted in Fig 3. The test error rate was estimated as the ratio between the number of false classifications and the number of all classifications.

The error bars in Fig. 3 represent the standard deviation, which is caused by different initializations of LVQ networks and by the nature of the training progress due to randomly applied input vectors. To avoid the possibility of excellent results for some arbitrary settings, we repeated random partitioning 50 times, following the paradigm of multiple hold-out validation. For each partition, training and testing were repeated 25 times with different weight matrix initializations. The LVQ network was trained and tested by different selections of signals. Every signal was first selected alone for both training and testing (Fig. 3, first 9 groups). The feature extraction methods, PSD and DVV, were applied individually and in combination. The best single channel detection result was achieved with a combination of PSD, DVV and GA for the EOG channel ‘vertical’ followed by the EEG channel ‘Cz’.

In an earlier work we pronounce that a combination of EEG and EOG measures should be most successful in predicting MSE [4]. Our results (Fig. 3, right) lend further support to this statement, independent from the feature extraction method used. Our simulations on DVV and PSD features achieved mean test error rates of 28 % and 17 % respectively. We judge the standard deviations of 1.4 % as moderate. The fusion of DVV and PSD features from all signals, which yields 531 features (7 EEG + 2 EOG signals) x (35 PSD + 24 DVV features), gained only a small improvement in the test error rates, namely from 17 % to 16 %. This result is not satisfactory enough and can be corrected by applying a GA where the DVV and PSD features had to compete with each other regarding their relevance for the MSE detection. After multiplying each feature with the estimated relevance factor obtained by GA, the training of LVQ was repeated. This way the best test error rate of 13 % was achieved.

6 Conclusions

We have presented an adaptive system for the analysis of Microsleep events (MSE), where several combinations of feature fusion were used for MSE detection and classification by means of populations of Learning Vector Quantization (LVQ) networks. Best results, with test errors down to 13 %, were obtained by a combination of all the recorded EEG and EOG channels, all features, and with feature relevance adaptation using Genetic Algorithms (GA).

Due to their complementing abilities to represent the linear and nonlinear nature of the EEG and EOG signals [13], simple feature extraction methods, PSD and DVV, were applied before and during an onset of a MSE. The results showed PSD to be more effective as a feature extraction method. This was also confirmed by our feature relevance results using GA, which detects features that were most relevant for the MSE detection. The relevances of the PSD features were similar to other findings [2-4], but for the understanding of the DVV feature relevance more research is needed. Furthermore, there are large inter-individual differences of the EEG- and EOG- characteristic [9, 10]. It would be interesting to ascertain whether the found feature relevance distribution can be confirmed or whether the DVV features play more significant role in certain cases. In general, there are strong indications that the role of the DVV features as compared to PSD features increases for the EOG signals. Another issue to be investigated is the fusion of EEG- / EOG- features and other oculomotoric features such as pupillography [11] using a greater variety of feature extraction methods. This is likely to improve and stabilize the discrimination of MSE, an issue of important real world applications.

References

1. Hall, D., Llinas, J.; An introduction to multisensor data fusion, Proceedings of the IEEE, 85; (1997)
2. Polychronopoulos, A., Amditis, A., Bekiaris, E.; Information data flow in AWAKE multi-sensor driver monitoring system, Proc. IEEE Intell. Vehicles Symp. (2004), 902-906
3. Sagberg, F., Jackson, P., Krüger, H-P., Muzet, A., Williams, A.J.; Fatigue, sleepiness and reduced alertness as risk factors in driving, Project Report, Transport RTD (2004)
4. Sommer, D.; Hink, T.; Golz, M.; Application of Learning Vector Quantization to detect drivers dozing-off, European Symposium on Intelligent Technologies, Hybrid Systems and their implementation on Smart Adaptive Systems (2002), 119-123

5. Gautama, T., Van Hulle, M.M., Mandic, D.P.; On the characterization of deterministic/stochastic and linear/nonlinear nature of time series, Technical Report (2004)
6. Gautama, T., Mandic, D.P., & Van Hulle, M.M. (2004). A Novel Method for Determining the Nature of Time Series. *IEEE Trans. Biomedical Engineering*, **51**(5), 728-736
7. Gasser, T., Bächer, P., Möcks, J.; Transformations Towards the Normal Distribution of Broad Band Spectral Parameters of the EEG, Electroenceph. clin. Neurophysiol. 53, (1982), 119-124
8. Devroye, L., Gyorfi, L. & Lugosi, G.; A probabilistic theory of pattern recognition; Springer, New York; 1996
9. Golz, M., Sommer, D., Seyfarth, A., Trutschel, U., Moore-Ede, M.; Application of vector-based neural networks for the recognition of beginning Microsleep episodes with an eyetracking system. In: Kuncheva, L. I. (ed.) Computational Intelligence: Methods & Applications, (2001), 130-134
10. Jung, T.-P., Stensmo, M., Sejnowski, T., Makeig, S.; Estimating Alertness from the EEG Power Spectrum. IEEE Transactions on Biomedical Engineering 44, (1997), 60-69.
11. Wilhelm, B., Giedke, H., Lüdtke, H., Bittner, E., Hofmann, A., Wilhelm, H.; Daytime variations in central nervous system activation measured by a pupillographic sleepiness test. J Sleep Research 10, (2001), 1-8
12. Hammer, B., Villmann, T.; Generalized relevance learning vector quantization. Neural Networks, 15, (2002), 1059–1068
13. Fell, J., Röschke, J., Mann, K., Schaffner, C.; Discrimination of sleep stages: a comparison between spectral and nonlinear EEG measures. Electroencephalogr. Clin. Neurophysiol. 98 (1996), 401-10