BIOSLEEP: A COMPREHENSIVE SLEEP ANALYSIS SYSTEM

N.McGrogan[1], E.Braithwaite[1], L.Tarassenko[2]
[1]Oxford BioSignals Ltd, Magdalen Centre, Oxford Science Park, Oxford, UK
[2]Department of Engineering Science, University of Oxford, UK

Abstract-**Traditionally the analysis of sleep has used two distinct manual EEG analysis methods: one for general structure, the other for short time-scale events. Both methods suffer from high levels of inter-expert variability.**

In this paper we present a system which uses a neural network classifier to analyse each second of sleep. Post-processing techniques are described which result in outputs which mimic both of the traditional manual analysis methods. This combination of methods results in a comprehensive sleep analysis system providing information on both the macro and microstructure of sleep.

Our results show that it is possible to use a combined approach to sleep analysis and that there is strong correlation between expert scoring and the post-processed neural network output.

Keywords **- neural networks, sleep analysis, intelligent signal processing**

I. INTRODUCTION

Quality of sleep directly affects quality of life. It has been shown that fatigue and excessive daytime sleepiness are associated with increased numbers of automobile and other accidents [1], cardiovascular disease [2], and asthma [2]. Clinicians make use of a variety of diagnostic methods in order to determine a patient's level of sleep deprivation. These include sleep diaries, home sleep monitoring without using the electroencephalogram (EEG) and sleep lab (in-patient) monitoring including the EEG. It is only through using the EEG that a night's sleep can be fully assessed.

This paper describes a system that has been developed to analyse a single channel of EEG and, through further processing, extract information that is associated with a number of different clinical conditions.

A. Electroencephalographic monitoring

The Electroencephalogram (EEG) has been used to study sleep since the 1930s. In 1968 Rechtschaffen and Kales [3] (R & K) documented an analysis method which was developed by a committee with the aim of standardising the analysis of sleep. The R & K method requires the use of at least one channel of EEG in combination with two eye channels (EOG) and a chin electromyogram (EMG).

The R & K classification system divides the EEG and other signals into contiguous signal segments of, typically, 30 seconds duration. Each epoch is categorised using a set of rules and assigned a stage representing the depth of sleep. The available sleep stages are W (wake), R (REM or dreaming sleep), 1–4 (progressively deeper non-REM sleep) and M (movement artefact making classification impossible). When this manual was originally produced it was designed to be pro-forma for a standard rather than a gold standard.

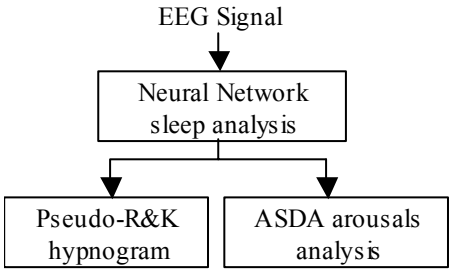

Fig. 1: BioSleep system structure

The R & K method is now recognised as suffering from a number of limitations [4], in particular, inter-expert variability (differences in analysis when different experts review the same data), intra-expert variability (the lack of consistency between analyses performed by a single expert when repeatedly presented with the same data), and the inability of the system to identify short-time-period structures within the EEG. It has, however, never been updated.

In 1992 the American Sleep Disorders Association (ASDA) [5] produced a second guidance document that was designed to standardise the identification of microarousals in sleep. The described methods were, however, too complex and time consuming to carry out on an entire night's sleep. In addition they suffer from a high level of inter- and intra-expert variability [6].

B. Current analysis methods

Currently, in order to analyse a complete eight hour sleep record two independent analyses must take place: a Rechtschaffen and Kales staging and an ASDA arousals analysis. It should also be noted that a number of signals, for example EMG, EOG, EEG, are required in order for full evaluation to take place.

Both methodologies analyse the EEG and it can therefore be suggested that the only real difference between the two approaches is that of the time periods of EEG analysed. One deals with the macro and the other the microstructure of sleep. This paper presents a system that analyses the EEG using a unified method and then, through post-processing, produces outputs that mimic results from both Rechtschaffen and Kales and microarousals analyses.

II. METHODOLOGY

In order to develop a system that is capable of identifying both the macro and microstructure within the sleep EEG a three stage process has been used as shown in Fig. 1.

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
25OCT2001	N/A	-

Title and Subtitle Biosleep: A Comprehensive Sleep Analysis System	**Contract Number**
	Grant Number
	Program Element Number
Author(s)	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) Oxford BioSignals Ltd, Magdalen Centre, Oxford Science Park, Oxford, UK	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es) US Army Research, Development & Standardization Group (UK) PSC 802 Box 15 FPO AE 09499-1500	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)

Distribution/Availability Statement
Approved for public release, distribution unlimited

Supplementary Notes
Papers from the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, October 25-28, 2001, held in Istanbul, Turkey. See also ADM001351 for entire conference on cd-rom., The original document contains color images.

Abstract

Subject Terms

Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** UU

Number of Pages
4

A. Sleep analysis

Previous work by Pardey et al [7] has demonstrated the use of a multi-layer perceptron neural network [8] in the analysis of sleep EEG. In Pardey's work, data from a single channel of scalp EEG (recorded from a central electrode) together with R & K scores agreed by three experts were available. The frequency characteristics of each second of EEG were described by the coefficients of a 10^{th} order autoregressive (AR) model and a neural network was trained to classify these coefficients. The sleep analysis carried out in this work builds on Pardey's techniques.

In neural network classification problems, the features derived from the input data and presented to the network are key to the classification performance. The use of AR coefficients suffers from two main problems:

1. The coefficients are dependent on the amplitude of the input signal so careful calibration is required.
2. The coefficients are not normalised so must be mapped to a 0-mean distribution before being applied to the neural network classifier. The parameters for this mapping must be derived from the training data.

To counter these problems we have replaced the auto-regressive coefficients used in [7] with reflection coefficients. Reflection coefficients provide a compact, parameterised estimation of a signal power spectrum in the same way as AR coefficients but are independent of signal amplitude and are normalised to the range [-1,1].

The neural network classifier is trained, as before, using sections of EEG which have been consensus scored with the R & K sleep stages W, R, and 4. The reflection coefficients for each second of the thirty second R & K epoch are calculated and presented to the network during the training process with a target output of 1 for the appropriate classification.

The output from the neural network provides an estimate of the probability that a single second of EEG represents each of three expert classified regions of sleep: wake (P(W)), REM/light (P(R)), or deep (P(S)). These three probabilities may be viewed separately or combined to provide a single value representing the depth of sleep: the *BioSleep hypnogram* calculated as P(W)-P(S). During periods of wakefulness the value of P(W)-P(S) will be close to 1; during deep sleep the value will approach –1; during light sleep both P(W) and P(S) will be small leaving P(W)-P(S) close to 0. The process of combination reduces the three-dimensional output from the classification process to a one-dimensional time series with little loss of information due to the physiological constraints of the sleep process.

Studies [7] have shown that the P(W)-P(S) graph correlates strongly with an expert scored R & K hypnogram but that it offers a much finer timescale.

The BioSleep hypnogram may be used directly to study the macrostructure of sleep. In addition the increased temporal resolution over traditional R & K analysis suggests that it may be possible to apply this method of sleep analysis to the analysis of the microstructure of sleep; this topic shall be covered in Section II.C.

B. Pseudo Rechtschaffen and Kales analysis

The BioSleep hypnogram, P(W)-P(S), described above can be used to study the macrostructure of sleep directly. However, clinicians are more familiar with the discrete output of an R & K analysis.

Since the BioSleep and R & K hypnograms correlate well with each other it is possible to classify each 30 second section of output from the neural network analysis to give a sleep stage. This classification uses the R & K stage names, but does not follow the traditional scoring rules, hence we shall refer to it as a *pseudo*-R & K hypnogram.

Thirty seconds of output from the neural network sleep analysis results in 90 individual values characterising the sleep over that epoch. This is too much data to classify directly, instead the values must be combined to form a more compact representation.

Since the outputs from the neural network analysis form estimates of the probability of class membership (for the three classes wake, REM / Light, and deep) they may be combined to give bulk probabilities of class membership for longer time periods. The winner-takes-all philosophy behind many of the R & K rules motivates the calculation of "majority probabilities" — the probability that the *majority* of a 30 second epoch (i.e., more than 15 seconds) represents each of the three classes. This combination will give a three-dimensional vector of probabilities, $\mathbf{p_{30}}$, for each 30 second epoch.

A total of 8502 epochs of consensus R & K scored EEG across 9 subjects were available to develop the pseudo-R & K classifier. Both neural network and nearest cluster mean classifiers were developed. The neural network classifier did not show any improvement in performance over the nearest mean method so the latter was selected for use since it is less complex.

Mean, $\boldsymbol{\mu}$, and covariance, $\boldsymbol{\Sigma}$, values were calculated to characterise the distribution of $\mathbf{p}_{30}$ values for each of the R & K classes. Pseudo R & K output is generated for new data by calculating $\mathbf{p}_{30}$ values for each 30 second epoch and selecting the sleep stage with the closest mean. The distance metric used in this comparison is the Mahalanobis distance, defined as:

$$d^2 = (\mathbf{p_{30}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{p_{30}} - \boldsymbol{\mu}).$$

This metric allows for the variance of each class in the calculation of distance.

Since the neural network sleep analysis is unable to distinguish between REM and light sleep we have also merged the pseudo-R & K stages 1 and R.

C. ASDA Microarousal detection

The definitions of the ASDA rules for scoring arousals are in terms of frequency shifts in the EEG [5]. Since the BioSleep sleep analysis system operates in the frequency domain, changes in the three class membership probabilities, and hence in the BioSleep hypnogram, correspond to changes in the EEG frequency characteristics. This connection motivates the use of the BioSleep output to isolate the EEG

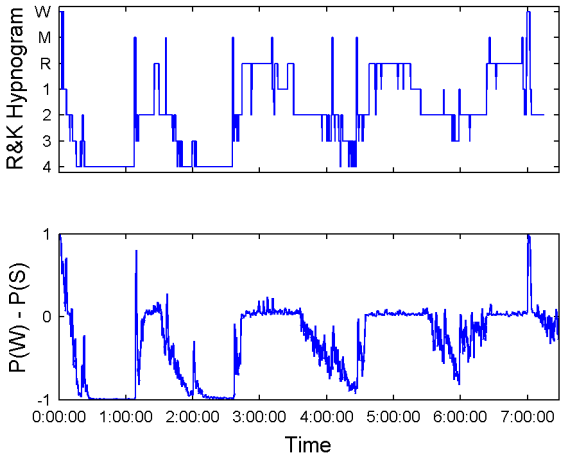

Fig. 2: Expert scored hypnogram and corresponding BioSleep hypnogram.

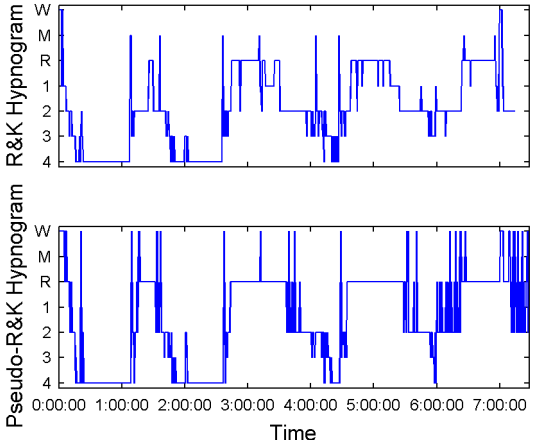

Fig. 3: Expert scored and BioSleep pseudo-R&K hypnograms.

frequency shifts which would be identified as arousals by the ASDA scoring rules. A set of filtering rules may then be applied to provide arousal detection [9].

According to the ASDA rules[5] a microarousal is defined as a shift in EEG frequency lasting for three seconds or more. In addition, two arousals separated by less than ten seconds of intervening sleep are treated as the same arousal event. By applying a threshold to the BioSleep hypnogram, periods of frequency shift are identified and the resulting signal indicates whether the subject is aroused or asleep. Periods of arousal or sleep which last for less than three seconds are then removed and any remaining periods of arousal with less than ten seconds of intervening sleep are merged to give the final output.

III. Results

A. Sleep analysis

Training of the neural network classifier took place in accordance with the protocol laid down in [7]. Applying the trained classification network to the data file used for illustrations in [7] results in the output seen in Fig. 2.

Qualitative comparison of the outputs from the new classification method against those from the old show that the structural information remains the same. Several improvements are also noted: the new method shows increased saturation of the P(W), P(R) and P(S) values during periods of known wake, light, and deep sleep respectively; in addition the gradual reduction in depth of sleep towards the end of the night is more evident.

B. Pseudo Rechtschaffen and Kales analysis

Nine full-night recordings of sleep EEG are available, together with R & K stages from a consensus of three experts. Due to the amount of data available a leave-one-out strategy has been adopted for the generation of results: for each of the nine available subjects cluster means and covariances are calculated using the data from the remaining eight. Test results are produced by classifying the single unused subject.

Validation of results is performed by comparison of the pseudo-R & K output with both the consensus and one of the expert R & K scores. Over the nine tests the pseudo-R & K results matched the consensus scores for a mean of 72.2% of the thirty second epochs; matches against the single expert scores showed a mean of 63.3%. These results compare reasonably well with the inter-expert matches reported in [10] of 74.6%.

Fig. 3 shows a comparison between an expert scored R & K hypnogram and the output from the pseudo-R & K analysis for one of the available subjects.

C. ASDA Microarousal detection

Three recordings of disturbed sleep EEG are available with expert arousal scoring. Each recording is 20 minutes long. A threshold of 0.25 has been selected arbitrarily as representing a point slightly above the normal level of REM and light sleep seen in a BioSleep hypnogram. The three recordings contain a mean of 25.7 arousals each. Comparison of the BioSleep arousal detection with the expert scoring shows a mean sensitivity (percentage of arousals detected) of 72.7% and a mean positive predictive accuracy (the percentage of detected arousals which are correct) of 96.0%.

Fig. 4 shows the results of applying the arousal detection method to one of the available EEG recordings. The expert arousal scoring is shown in the top graph followed by the BioSleep hypnogram output from the first stage of sleep analysis (in addition the position of the 0.25 threshold is shown). The bottom graph shows the arousal scores obtained after the threshold and clean-up processing has been performed.

IV. Discussion

The results presented above show that the system outputs exhibit strong correlation with consensus expert scores, both for the macro and microstructure of sleep. None of the three analyses described (the BioSleep hypnogram, pseudo-R & K, and microarousal detection) attempt to mimic the manual techniques directly but instead demonstrate that signal-processing methods may be used to provide a principled alternative to subjective assessment.

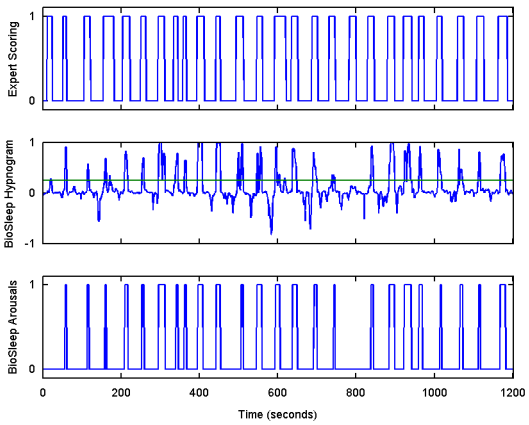

Fig. 4: Expert arousal scoring, BioSleep hypnogram, and BioSleep arousal detection.

One limitation of the use of a single EEG channel is the inability of the analysis process to distinguish between REM and light (R & K stage 1) sleep. The clinical significance of this shortcoming is, however, small since many sleep disorders may be discerned in the *structure* of sleep.

V. Further Work

The sleep analysis system presented in this paper is currently in use by a number of clinical researchers. As a result of these collaborations we hope to develop further improvements to the method to increase its clinical utility.

A prime area for advance is in the choice of electrode site. Currently the analysed EEG recording is taken from a central electrode, a choice motivated by convenience (this location is already recorded for traditional sleep analysis) and the reduced artefact seen on these channels. Using the signal processing techniques presented in this paper we hope to be able to perform sleep analysis using a single channel of EEG from an alternative site, e.g., mastoid to contra-lateral mastoid, which may prove more acceptable to the subject but cannot be analysed using traditional methods.

In addition we hope to calculate standard measures of sleep performance, such as sleep latency, from the analysis output in order to improve the quality of information available to the clinician.

Further areas for research include methods for reducing reflection coefficient variability and innovative filtering techniques to improve the post-processing methods.

VI. Conclusions

Despite the existence of standardised manual methods for the analysis of sleep EEG previous work [4][6] has shown the large inter-expert variability in both R & K and ASDA scoring. This variability illustrates the need for automated systems capable of consistent and accurate scoring of the sleep EEG.

The three-stage sleep analysis process presented here demonstrates that a unified approach to the analysis of sleep *is* possible using only a single channel of EEG with no requirement for extra EOG or EMG channels. This system draws inspiration from both standard clinical techniques and modern signal-processing methods in order to "score" sleep both in terms of its R & K stages and its arousal microstructure.

References

[1] P.Philip, F.Vervialle, P.LeBreton, J.Taillard, J.A.Horne, "Fatigue, alcohol, and serious road crashes in France; factorial study of national data", *BMJ Volume 322, 7th April 2001* pp 829-830

[2] J.Schneerson, *Handbook of Sleep Medicine,* Blackwell Science, 2000

[3] A.Rechtschaffen and A.Kales, *A manual of Standardized Terminology, techniques and scoring system for sleep stages of human subjects* U.S. Department of Health, Education and Welfare- NIH, Bethesda Maryland 20014, 1968

[4] J.Hasan "Limitations of Rechtschaffen & Kales" *Towards a new standard of Modeling sleep based on Polysomnograms* ESRS proceedings 1998

[5] Atlas Task Force of the American Sleep Disorders Association, "EEG Arousals: Scoring Rules and Examples", *Sleep* Vol 15 No 2, 1992, pp174-184

[6] Michael J. Drinnan et al., "Interobserver Variability in Recognizing Arousal in Respiratory Sleep Disorders," *Am J Respir Crit Care Med*, Volume 158, 1998 pp358–362

[7] J.Pardey, S.Roberts, L.Tarassenko, J.Stradling, "A new approach to the analysis of the human sleep/wakefulness continuum", *J. Sleep Research*, (1996) 5, 201-210

[8] C.M.Bishop, *Neural Networks for Pattern Recognition,* OUP, 2000

[9] Mayela E. Zamora, "The Study of the Sleep and Vigilance Electroencephalogram Using Neural Network Methods", PhD Thesis, University of Oxford, Department of Engineering Science, 2001.

[10] H Danker-Hopfe et al., "Interrater reliability between scorer from eight European sleep labs in subjects with different sleep disorders"