Online Single EEG Channel Based Automatic Sleep Staging

Gary Garcia-Molina[1,2], Michele Bellesi[2], Sander Pastoor[3], Stefan Pfundtner[3], Brady Riedner[2], and Giulio Tononi[2]

[1] Philips Research North-America, Briarcliff, NY, USA
[2] Department of Psychiatry, University of Wisconsin, Madison, WI, USA
[3] Philips Research Europe, Eindhoven, The Netherlands
gary.garcia@philips.com, bellesi@wisc.edu

Abstract. Recent evidence supports the positive effects of external intervention during specific sleep stages (e.g. enhanced memory consolidation and depression relief). To enable timely intervention, online automated sleep staging is required and preferably with short latency. In this paper, we propose an approach to achieve this based on the analysis of spectral features of a single electroencephalogram (EEG) channel and the use of Gaussian Mixture Models. We compare among several choices for the EEG signal location, the type of spectral features, and the duration of the signal segment (epoch) that is required to automatically identify the sleep stage. The performance metric used for comparison purposes is the kappa statistic, which measures the agreement between the automatic and manual sleep staging. The performance is higher when central EEG locations (C3, C4), longer epochs, and the power in five frequency bands are used. However, good results (kappa=0.6) can also be obtained for an epoch duration of 12 seconds.

Keywords: automatic sleep staging, online, single EEG signal, spectral features, GMM.

1 Introduction

Sleep is a state of reversible disconnection from the environment characterized by quiescence and reduced vigilance. Although the precise function of sleep remains to be elucidated, it appears that sleep primarily benefits the brain. For example according to the synaptic homeostasis hypothesis (SHY) [1], plastic processes occurring during wakefulness result in a net increase in synaptic strength in many cortical circuits, and sleep is needed to renormalize synaptic strength in a way that is beneficial to nerve cells and to memory. SHY also emphasizes that when cortical neurons begin oscillating at low frequencies during non-rapid eye movement sleep, these oscillations become synchronized. This is why the electroencephalogram (EEG) exhibits high power in the delta band (0.5-4 Hz) particularly at the beginning of a sleep episode. SHY also claims that slow waves in sleep do not merely reflect synaptic strength, but

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 333–342, 2013.

also play a functional role in “renormalizing” synaptic strength to a baseline level that is energetically sustainable and beneficial for performance [1].

Recent research evidence indicates that modulating sleep activity patterns, specifically slow waves via sensory, magnetic, or electric stimuli at specific sleep stages can be beneficial in a wide range of contexts including memory acquisition and consolidation [2][3] and relief from depression [4][5]. To verify the validity of such interventions in practice requires conducting research in a larger population using automated means for online sleep staging with low latency to allow timely intervention.

Conventional sleep staging relies on various bio-signals (polysomnography) for human experts to decide on sleep stages typically on the basis of 30-second long segments. Real-time sleep staging is proposed in [6] using EEG, electro-oculogram (EOG), and electro-myogram signals (EMG). We consider here the option of achieving online automatic sleep staging on the basis of a single channel (or signal). EOG and EEG electrodes are considered for this purpose. Using a single signal permits to simplify the research setup and increases the subject comfort.

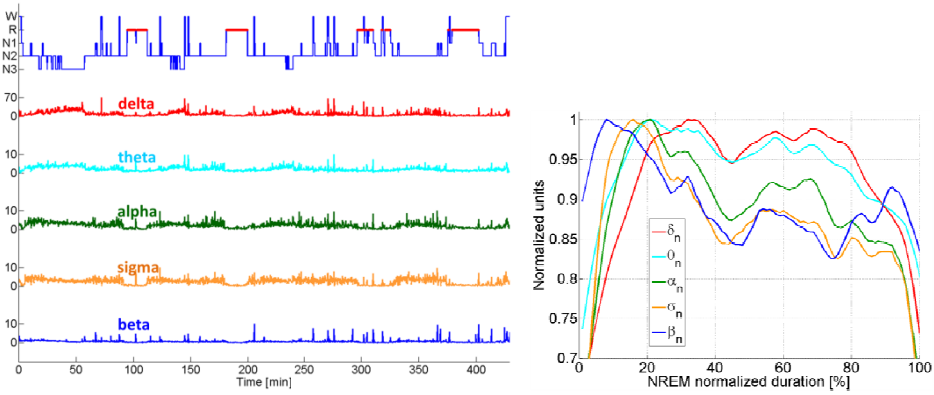
This paper is organized as follows. Section 2 presents an overview of the time course of the EEG signal during sleep. Section 3 describes the methods used for automatic sleep staging. The results in terms of performance are presented in Section 4. Section 5 concludes the paper.

2 EEG during Sleep

Two distinct types of sleep occur in humans: rapid eye movement (REM) sleep, and non–REM sleep. Compared to the low voltage, high frequency patterns appearing in the awake EEG, non–REM (NREM) sleep is associated with a synchronized EEG pattern. NREM is subdivided into stages N1, N2, and N3. During REM, the EEG exhibits a pattern similar to that observed during wakefulness [7].

Global trends for the EEG power during sleep in the classical frequency bands: delta (0.5–4 Hz), theta (4-8 Hz), alpha (8-12 Hz), sigma (11-15 Hz), and beta (15-30 Hz) are analyzed in detail in [8]. Fig. 1a summarizes the typical time courses of the power in those bands depending on the sequence of sleep stages as indicated by the hypnogram (top graph in Fig. 1a). The power in Fig. 1a is reported in RMS units which are obtained by calculating the square root from the average energy (in a window of a given duration) of the signal band-pas filtered in the frequency of interest. As sleep deepens the power in the delta and theta bands increase whereas the power in the alpha, sigma, and beta bands follow a quasi-opposite trend.

A more detailed view of the power trends during non-REM (NREM) can be observed in Fig. 1b where the average power in all frequency bands over 10 subjects is portrayed. The time was normalized to percent units to account for the difference in NREM duration for different subjects. The power was also normalized (by scaling w.r.t. the highest value) to facilitate the comparison. The power in delta and theta increase at the beginning of the NREM episode, remains high for most of the NREM duration, and decreases at the end of NREM (i.e. following an inverted-U trend). The power in alpha, sigma, and beta rapidly increases at the beginning of the NREM



(a) (b)

Fig. 1. a) Top: hypnogram of a representative night (red segments correspond to REM sleep). Time courses of the power, in RMS units, in the delta, theta, alpha, sigma, and beta frequency bands. b) (Average) normalized time course of the normalized power in delta, theta, alpha, sigma, and beta throughout a NREM episode.

episode, reaches its maximum, and decreases afterwards. The power in beta slightly increases by the end of the NREM episode. These observations are used to guide the spectral feature extraction as described in Section 3.2.

3 Methods

3.1 Dataset

The data from ten healthy, medication-free, right-handed subjects (five female; age 21.9 ± 0.5 years; mean ± std.) who participated in a previous study run at the University of Wisconsin [9] was used in this paper. The data includes right/left electro-oculogram signals (EOGR and EOGL), six bipolar EEG signals (F3-A2, F4-A1, C3-A2, C4-A1, O1-A2, and O1-A2), other polysomnography signals (details in [9]), and the manually annotated hypnogram done by an expert according to the standard rules in [7] and on the basis of 30-second long segments. To permit the analysis of segment durations as described in section 3.2, sleep stages were assigned to each 6-second long segment by simple extrapolation from the 30-second long based staging. Five sleep stages were considered: 1) Wake (W), 2) REM (R), 3) N1, 4) N2, and 5) N3.

3.2 Automated Sleep Staging

The procedure used to automatically score a hypnogram from the data of a single signal, is designed in such a way so as to accommodate online operation. A signal segment (*epoch*) of a given duration is presented to a *classifier* which decides the sleep stage to which this epoch belongs independently from previous decisions. In this

paper, we consider epochs with durations: 6, 12, 18, 24, and 30 seconds. The classifier is *trained* with ground truth data, i.e. epochs for which the sleep stage is known (annotated by the human expert).

Spectral Features per Epoch

Spectral features from the epochs are first extracted by calculating the log of the estimated power (in RMS units) in the delta, theta, alpha, sigma, and beta frequency bands (δ, θ, α, σ, and β are used to refer to the log of the power in RMS units). The logarithm is extracted to equalize the feature range across frequency bands (especially relevant for delta because the power in this band is higher than in other bands, see Fig. 1a) as this prevents the estimation of singular covariance matrices [10]. A combination of spectral features (e.g. δ and β) defines a *feature vector* associated with each epoch. For a sleep night recording, the number of feature vectors corresponds to the number of epochs in the recording.

The spectral combinations considered are:

1) Two-band spectral combinations between i) the bands where the power follows an inverted-U trend during NREM (i.e. δ and θ according to Fig. 1b) and ii) the bands where the power trend quasi-mirrors that of δ and θ during NREM (i.e. α, σ, and β according to Fig. 1b). This results in 6 two-band spectral combinations where the feature vectors have two components.
2) All possible four-band spectral combinations. This amounts to five combinations: δθασ, δθαβ, δθσβ, δασβ, and θασβ where the feature vectors have four components.
3) The combination of all five bands which results in feature vectors with five components.

Classifier

The classifier decides on the sleep stage for a given epoch on the basis of that epoch's feature vector (referred to as x, d is the feature vector dimension). The classification approach in this paper is based on the estimation of the probability that a given feature vector belongs to each sleep stage, referred to as $p(x|stage)$, and select the sleep stage (s_x) for which the likelihood is the largest (see Eq. 1).

$$s_x = \underset{stage=\{W,R,N1,N2,N3\}}{\arg\max} \left\{ p(x|stage) \right\} \tag{1}$$

The probability density function $p(x|stage)$ is estimated using a Gaussian Mixture Model (GMM) [11] and can be written as:

$$p(x|stage) = \sum_{m=1}^{M} a_m p\left(x|stage, m\right), \tag{2}$$

where a_m is the *m*-th mixture coefficient, $\sum_{m=1}^{M} a_m = 1$, and $p\left(x \middle| stage, m\right)$ is the *m*-th multimodal Gaussian mixture which can be written as in Eq. 3 (the "stage" sub-indices were removed in this equation for convenience of notation).

$$p(x|m) = \frac{1}{(2\pi)^{d/2}\left|\Sigma_m\right|^{1/2}} \exp\left\{-\frac{1}{2}\left(x-\mu_m\right)^T \Sigma_m^{-1}\left(x-\mu_m\right)\right\}, \tag{3}$$

where Σ_m is the $d \times d$ covariance matrix, and μ_m is the mean vector having d elements. The mixture model can approximate any probability density function arbitrarily closely provided that it contains enough components[12].

The GMM parameters are estimated using data from a *training set*. Given the order of the model *M*, the mixture coefficients, covariance matrices, and the mean vectors are estimated using the expectation maximization algorithm. This algorithm is described in detail in[10]. The order of the model is chosen in such a way so as to balance the model complexity (the higher *M* is, the higher the complexity is) and the fit with the data. For the data considered in this paper, we use *M=2*.

3.3 Performance Evaluation

Evaluation of the automated staging performance was made following a leave-one-subject-out procedure where the data from all the subjects but one in the dataset is used to estimate the GMMs for each sleep stage. The data from the remaining subject (e.g. subject *k*) is used to test the models and estimate a 5×5 *confusion matrix* C_k where the *i*, *j* element represents the number of times an epoch manually scored as belonging to sleep stage *i,* was classified as belonging to sleep stage *j*. The correspondence between numeric indices and sleep stages is as follows: 1) W, 2) R, 3) N1, 4) N2, and 5) N3. A visualization of the performance evaluation is illustrated in Fig. 2, where *N* is the number of subjects (10 in this paper), and $F_{k,1}, F_{k,2}$ are the feature vectors of the epochs of subject *k* corresponding to the first and second nights respectively. The training data consists of the epochs for both nights of all other (*N*-1) subjects in the dataset (i.e. $k_1, \ldots, k_{N-1} \neq k$). The training data is organized per sleep stage and the corresponding GMMs are estimated.

Although the confusion matrix provides sufficient information to assess the automatic staging performance, comparing the performance for different feature selections and electrode locations is facilitated by extracting a single quantity that reflects the information contained in the confusion matrix. This can be accomplished by estimating the so-called *kappa statistic* [13] from the global confusion matrix $C = \sum_k C_k$.

Kappa measures the agreement between the manual and automatic sleep staging and is defined as: $\kappa = (P_a - P_e)/(1 - P_e)$, where P_a is the relative observed agreement, and P_e is the hypothetical probability of chance agreement. P_a and P_e can be obtained

from the confusion matrix as: $P_a = \sum_{i=1}^{5} C_{i,i} / \sum_{i=1}^{5}\sum_{j=1}^{5} C_{i,j}$ and $P_e = \sum_{i=1}^{5} S_{:,i} S_{i,:} / \left(\sum_{i=1}^{5}\sum_{j=1}^{5} C_{i,j} \right)^2$, where $S_{:,i}$ and $S_{:,i}$ are respectively the sums across columns and rows of the confusion matrix C.

As discussed in [14], kappa values greater than 0.80 represent almost perfect, between 0.61 and 0.80 substantial, between 0.41 and 0.60 moderate, between 0.21 and 0.40 fair and between 0 and 0.21 slight agreement. The average kappa value characterizing the agreement between two human experts is 0.87 [15]. For illustration, the automatically generated hypnograms when κ=0.6 and κ=0.4 are represented in Fig. 5.

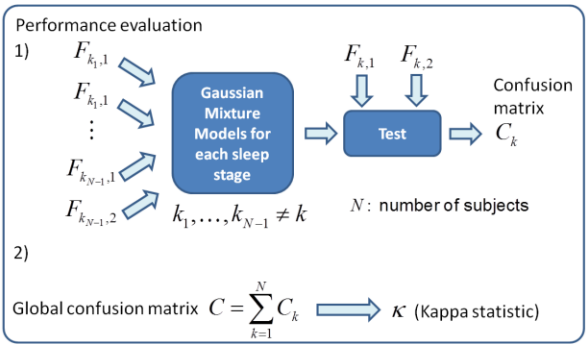


Fig. 2. Visualization of the performance evaluation method. A leave-one-subject-out procedure is used and the corresponding κ is estimated.

4 Results and Discussion

4.1 Two-Band Spectral Combinations

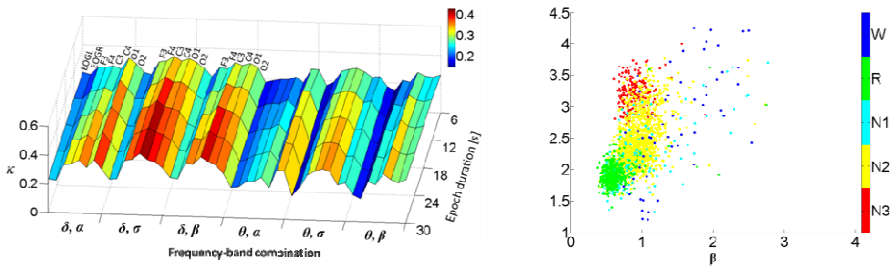
The results in terms of kappa for the six two-band spectral combinations (δα, δσ, δβ, θα, θσ, θβ) are represented in Fig. 3a. The electrode location and the epoch duration are also considered as independent variables.

As expected, kappa increases when the epoch duration increases. As for the electrode locations, the highest kappas are reached on central locations (C3 and C4), the lowest kappas are for EOG electrodes, and intermediate kappa values are for frontal and occipital locations. The maximum kappa is 0.42 (moderate agreement) for location C3, epoch duration 24 seconds, and the combination δσ.

If θ is used instead of δ, kappa decreases which suggests that δ plays a key role in defining the sleep stages. An illustration of the mapping of sleep stages in the δ-β plane is depicted in Fig. 3b.

4.2 Four-Band and Five-Band Spectral Combinations

The kappa results for all possible four-band spectral combinations and the five-band spectral combination are represented in Fig. 4. The electrode location and the epoch



(a) (b)

Fig. 3. a) Values of κ for: the 2-band spectral combinations for all considered electrode locations and epoch durations. b) Mapping of sleep stages in the δ-β plan for subject 1.

duration are also considered as independent variables. As in the case of 2-band spectral combinations, the value of kappa increases with the epoch duration. Kappa is larger for frontal and central electrode locations, low for EOG signals, and intermediate for occipital locations.

Using all the five bands improves kappa as compared to the four-band combinations. The maximum kappa is 0.63 (substantial agreement with the manual sleep staging) for location C4, epoch duration 24 seconds, and using the five frequency bands. When using only four frequency bands, the maximum kappa is 0.60 (moderate agreement) for location C4, epoch duration 24 seconds, and the combination δθαβ. For the four-band combinations, the one that is associated with lower kappas is: θαβσ. This confirms the results in Section 4.1 where δ was found to be essential for the automatic staging.

Table 1. Kappa for short epoch durations, frontal & central locations, and using δθασβ

	Location			
Epoch Duration	**F3**	**F4**	**C3**	**C4**
6 seconds	0.53	0.55	0.54	0.56
12 seconds	0.57	0.60	0.58	0.60

For the sake of low latency online sleep staging, it is useful to consider the performance for shorter epoch durations (i.e. 6 seconds and 12 seconds). This is reported in Table 1 where it can be observed that a moderate agreement with manual sleep scoring can be achieved for 12-second long epochs. An illustration of the automatically scored hypnogram using 12-second long epochs is shown in Fig. 5.

On the basis of 30-second long epochs using a single EEG signal, with a smoothing step (deciding on the stage by considering the results on adjacent epochs), [16] reports a kappa of 0.72 for 5 sleep stages. The smoothing step is of course not suitable for online operation. Using 30-second long epochs a kappa value of 0.79 is reported in [15] using a single EEG signal. The advantage of our approach relies on the fact that shorter epochs can be used, maintaining a moderate level of agreement, which ensures low latency, favoring real-time operation.

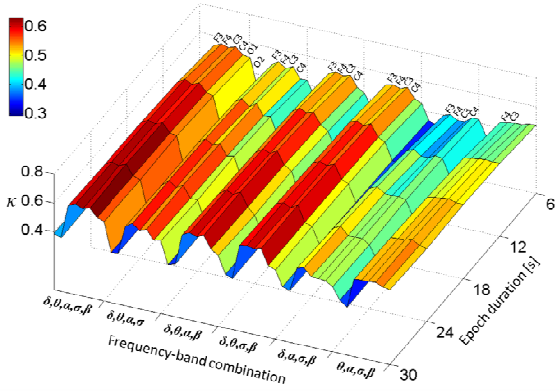


Fig. 4. Values of κ for: all 4-band spectral combinations and the 5-band spectral combination versus all considered epoch durations

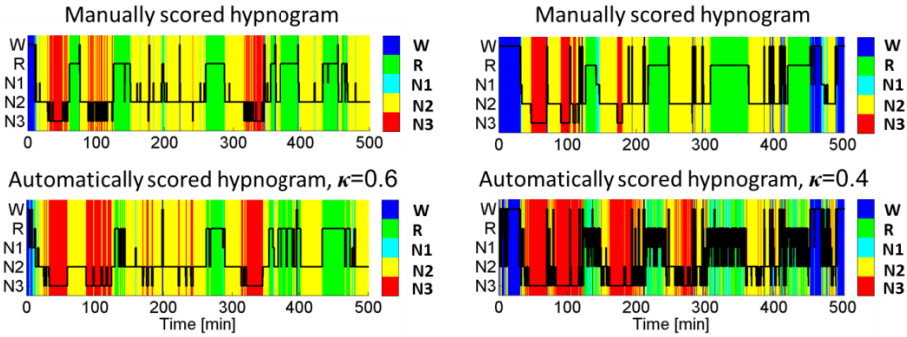


Fig. 5. Examples of automatically generated hypnograms; κ=0.6 (left) and κ=0.4 (right). (a) Top: manually scored hypnogram (reference). Bottom: automatically scored hypnogram (κ=0.6) using spectral features δ, θ, α, σ, and β extracted from the signal at electrode C4 and on the basis of 12-second long epochs. (b) Top: manually scored hypnogram. Bottom: automatically scored hypnogram (κ=0.4) using spectral features δ, and β extracted from the signal at electrode C3 and on the basis of 12-second long epochs.

5 Conclusions

In this paper, we propose an approach to automatic sleep staging on the basis of a single channel using spectral features and estimating the probability density function of feature vectors per sleep stage through GMMs. The performance was evaluated using the kappa statistic which can be calculated from the confusion matrix.

The performance was compared across different selections for the spectral features, epoch durations, and electrode location. Higher performance can be reached if central electrodes (C3, or C4), longer epochs, and all frequency bands are used. If only two

frequency bands were to be used (e.g. delta and sigma), the performance is moderate (kappa=0.4).

Interestingly, shorter epochs (6 and 12 seconds) can be used at an acceptable level of performance, kappa=0.56 (6 seconds, C4) and kappa=0.6 (12 seconds, C4). The sleep stage can then be decided for each 12-second long epoch independently from the stage of the previous epochs. This permits almost real-time operation and timely intervention during sleep.

Since the decision on the sleep stage to which an epoch belongs is taken independently from previous epochs, an obvious performance improvement step would consist in smoothing the results by considering the sleep stages from previous epochs. This step however, could increase the latency which may adversely impact online operation.

The use of standard manually scored 30-second long epochs as reference for the staging of shorter epochs may not be the best choice because this is a standard for offline analyses that was established in the clinical framework.

References

1. Tononi, G., Cirelli, C.: Sleep function and synaptic homeostasis. Sleep Medicine Reviews 10(1), 49–62 (2006)
2. Marshall, L., Helgadóttir, H., Mölle, M., Born, J.: Boosting slow oscillations during sleep potentiates memory. Nature 444(7119), 610–613 (2006)
3. Marshall, L., Mölle, M., Hallschmid, M., Born, J.: Transcranial direct current stimulation during sleep improves declarative memory. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 24(44), 9985–9992 (2004)
4. Landsness, E.C., Goldstein, M.R., Peterson, M.J., Tononi, G., Benca, R.M.: Antidepressant effects of selective slow wave sleep deprivation in major depression: a high-density EEG investigation. Journal of Psychiatric Research 45(8), 1019–1026 (2011)
5. Vogel, G.W., Vogel, F., McAbee, R.S., Thurmond, A.J.: Improvement of depression by REM sleep deprivation. New findings and a theory. Archives of General Psychiatry 37(3), 247–253 (1980)
6. Kuwahara, H., Higashi, H., Mizuki, Y., Matsunari, S., Tanaka, M., Inanaga, K.: Automatic real-time analysis of human sleep stages by an interval histogram method. Electroencephalography and Clinical Neurophysiology 70(3), 220–229 (1988)
7. Iber, C., Ancoli-Israel, S., Chesson, A.L., Quan, S.F.: The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, First. American Academy of Sleep Medicine (2007)
8. Merica, H., Fortune, R.D.: State transitions between wake and sleep, and within the ultradian cycle, with focus on the link to neuronal activity. Sleep Medicine Reviews 8(6), 473–485 (2004)
9. Hulse, B.K., Landsness, E.C., Sarasso, S., Ferrarelli, F., Guokas, J.J., Wanger, T., Tononi, G.: A postsleep decline in auditory evoked potential amplitude reflects sleep homeostasis. Clinical Neurophysiology 122(8), 1549–1555 (2011)
10. Bishop, C.M.: Neural Networks for Pattern Recognition. Oxford University Press (1995)
11. Reynolds, D.A., Rose, R.C.: Robust text-independent speaker identification using Gaussian mixture speaker models. IEEE Transactions on Speech and Audio Processing 3(1), 72–83 (1995)

12. McLachlan, G.J., Basford, K.E.: Mixture models: Inference and applications to clustering. CRC Press (1987)
13. Cohen, J.: A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 20(1), 37–46 (1960)
14. Landis, J.R., Koch, G.G.: The measurement of observer agreement for categorical data. Biometrics 33(1), 159–174 (1977)
15. Jensen, P.S., Sorensen, H.B.D., Leonthin, H.L., Jennum, P.: Automatic sleep scoring in normals and in individuals with neurodegenerative disorders according to new international sleep scoring criteria. Journal of Clinical Neurophysiology 27(4), 296–302 (2010)
16. Berthomier, C., Drouot, X., Herman-Stoïca, M., Berthomier, P., Prado, J., Bokar-Thire, D., Benoit, O., Mattout, J., D'Ortho, M.-P.: Automatic analysis of single-channel sleep EEG: validation in healthy individuals. Sleep 30(11), 1587–1595 (2007)