On the Continuous Evaluation of the Macrostructure of Sleep

Diego Alvarez-Estevez¹⁺, José M. Fernández-Pastoriza¹, Elena Hernández-Pereira¹, Vicente Moret-Bonillo¹

¹Laboratory for Research and Development of Artificial Intelligence (LIDIA), Dept of Computer Science, University of A Coruña, A Coruña, Spain
{diego.alvareze, chema.fernandez, vicente.moret, elena.hernandez}@udc.es

Abstract. Sleep staging is one of the most important tasks on the context of sleep studies. For more than 40 years the gold standard to the characterization of patient’s sleep macrostructure was the set of rules proposed by Rechtschaffen and Kales (R&K) recently modified by AASM rules. Nevertheless the resulting map of sleep, the so-called hypnogram, has several limitations such as its low temporal resolution and the unnatural characterization of sleep through assignment of discrete sleep states. This study reports an automatic method for the characterization of the structure of the sleep. The method is based on the use of fuzzy inference in order to provide soft transitions among the different states. Main intention is to overcome limitations of epoch-based sleep staging by obtaining a more continuous evolution of the sleep of the patient.

Keywords: Sleep Studies, Hypnogram, Fuzzy Reasoning

1 Introduction

One of the most important tasks within the analysis of the sleep is the characterization of the patient’s sleep macrostructure. This leads to the construction of the hypnogram in which the voluminous chart recordings (PSG) of electrical activities are summarized in a simple graph aimed at showing evolution of the different states of sleep throughout the night. The procedure for obtaining the hypnogram was initially proposed in 1968 by Rechtschaffen and Kales (R&K) [1]. They established a method based on a set of rules to assign to a time interval in the PSG a label representing certain state of sleep: Wakefulness (W), stages 1 to 4 (S1, S2, S3 and S4), and Rapid Eye Movement (REM) phase. Therefore the sleep recording was segmented into these classifiable intervals, called epochs, being its length arbitrary established to 30 seconds. R&K method has been the gold standard to the scoring of sleep macrostructure for more than 40 years, being only recently modified by the American Association of Sleep Medicine (AASM) [2]. In any case resulting hypnogram is still based on an epoch-based segmentation of sleep.
Sleep staging is in general a tedious task entailing too much time and effort for the physician. Accordingly, attempts to automate the epoch-based staging of sleep according to R&K/AASM methods are almost as old as the R&K rules. A detailed review of the literature can be found in [3], in which it is stated however that automation of hypnogram generation is still an open area of interest. On the other hand in the recent years several criticisms have been associated with such a method for sleep characterization [4-5]. Major drawbacks are associated with its low temporal resolution—one label for 30 seconds-and the unnatural classification of sleep based on fixed-duration discrete epochs. Effectively, evolution of biological processes rather occurs in a continuous manner in which a soft transition takes place between the different considered states.

On this context, there is an interest on exploring different approaches that could overcome such limitations. A first step in this line is the developing of computer approximations on the goal of achieving a more continuous marker of sleep depth [6-9]. However previous approaches suffer from the lack of information on the intra-sleep periods. A realization that could also account for a continuous characterization of NREM intra-states is desirable. In this respect Flexer et al. [10] developed a continuous probabilistic sleep stager (considering three states: wakefulness, deep sleep and REM) based on a single EEG signal. Nevertheless a problem with probabilistic approximations is that there is no possibility of differentiating between uncertainty about an event and the probability of its complement.

Our approximation proposes an alternative solution to the problem of the continuous sleep staging by using the fuzzy logic paradigm. Fuzzy logic allows us to quantify a decision in terms of a fuzzy degree of membership which avoids binary decisions based on categorical classifications. It also allows us to deal with uncertainty and imprecision, common aspects of medical diagnostic domains. Moreover given the membership $\mu_H(x)$ of a certain hypothesis $H$, the membership of the complementary hypothesis ($\mu_{\neg H}$) should not necessarily be $1 - \mu_H(x)$, and therefore overcoming the problems of the probabilistic approaches.

Our objective with the proposed approach is to present a system capable of performing a continuous characterization of sleep, with which to overcome the limitations of the discrete epoch-based methods of sleep classification. In [8] the authors presented a first prototype of such a system where implementation of the module to estimate patient’s sleep depth was performed. This paper presents an expanded system, where a continuous evolution of the degree of membership for each sleep state is accomplished. Finally and in order to allow the validation process, the continuous representation is used to go back over a classical discrete hypnogram. Validation over the discrete representation shows how the information contained in the classical hypnogram is preserved in the continuous representation.

2 Materials and Methods

The general approach is organized in three sequentially related processing steps. The first step is in charge of extracting relevant features over the biological signals in the
PSG. As for determining sleep structure, analysis of Electroencephalogram (EEG), electromyogram (EMG) and Electrooculogram (EOG) has to be performed. After relevant parameters have been extracted, information is fed into a second processing step where the reasoning process occurs obtaining as output, a degree of membership with respect to each considered state of sleep, in this work: awake (W), drowsy sleep (DS), deep sleep (DEEP) and REM. Note in the current work DS includes classical S1 and S2 sleep stages. The whole previous process is accomplished in a second-by-second granularity, thus with higher resolution in comparison to epoch-based procedures. This, together with the properties derived from the use of a fuzzy reasoning schema, allows us obtaining a new representation of the hypnogram in which current evolution of the different sleep states is individually characterized. Eventually (third step) this representation is used in order go back over the classical epoch-based hypnogram. Some post-processings are applied in this respect over the continuous hypnogram.

Detailed functionality concerning the above outlined approach is subsequently described.

2.1 Parameter Extraction

Detection of the Eye Movements. In the case of the EOG the interest is to characterize the eye movements. In order to achieve such a task, an overlapping moving window of 3 seconds is shifted second-by-second throughout both EOG_L and EOG_R channels, and computing the amplitude of the corresponding signal interval within the window. Amplitude is calculated as the difference between the maximum and the minimum values of the signal inside the window. Thereby a value for the amplitude of the signal is obtained for the current second. By repeating this process throughout the recording two amplitude signals—one for each channel— are obtained. Finally we construct a new signal $A_{EOG}$ (see Fig. 1a) by averaging the two previous amplitude signals obtained for each channel (left and right), thus obtaining a single parameter to represent the EOG amplitude independently of the channel. It can be shown in Fig. 1a that the amplitude of $A_{EOG}$ signal increases in presence of EOG movements while it is almost flat for a relaxed EOG.

Characterization of Muscle Tone. In the case of EMG, to distinguish between presence and absence of muscle tone, a similar amplitude-based analysis is performed. Using a window of 3 seconds and moving it second-by-second throughout EMG, a new $A_{EMG}$ signal is obtained. Differently from the amplitude computation in the EOG, this time each i-th sample of $A_{EMG}$ is calculated as the mean of the absolute value of the EMG samples included in the window (Fig. 1b). The main reason to do so is that it better supports the higher frequency nature of the EMG signal.

Processing of Electroencephalographic Activity. Regarding the EEG, the different sleep stages are characterized by the different proportion of characteristic frequencies in the most representative bands: alpha ($\alpha$, 8-12 Hz), beta ($\beta$, 13-30 Hz), theta ($\theta$, 4-7
Hz) and delta (δ, 0.5-3 Hz) [2]. Thus, Short-Time Fourier Transform (STFT) method (3-second window, 2-second overlapping) is used to compute spectra on every analysis window of the EEG. Then, within each window we quantify the power of the signal at each frequency band by using a band-pass filter in the corresponding range. The Power Spectral Density (PSD) is calculated by integrating the Fourier Transform of the resulting spectra from the filtered signal. Therefore, four measures of PSD are obtained at each time step through the previous process: PSD$_{a}$, PSD$_{b}$, PSD$_{c}$ and PSD$_{d}$, one for each kind of wave. Similarly to the case of EOG, average of parameters in the two EEG derivations –C3/A2, C4/A1–, is performed to obtain final values for the parameters. Fig. 1c shows the resulting four signals characterizing the activity of the EEG.

![Fig. 1](image)

**Fig. 1.** In the figure 3-epochs of 30 seconds are shown. Signal amplitudes are normalized in [-1,1]; a) EOG derivation with and AEOG signal superimposed; b) chin EMG with AEMG signal superimposed; c) EEG derivation with PSD$_{a}$, PSD$_{b}$, PSD$_{c}$ and PSD$_{d}$ signals superimposed.

**Analysis of Sleep Spindles.** Within the analysis of the EEG, a method for the detection of sleep spindle transients is also included. The main interest in the detection of this kind of events is that sleep spindles are one of the hallmarks of human S2 sleep stage and are also one of the few transient EEG events which are uniquely related to sleep [11]. Detection of these events is aimed at achieving better characterization of the sleep evolution during drowsy sleep periods.

The developed algorithm is based on the use of STFT to search increments in the sigma band (12-15 Hz) with a duration ≥ 0.5 seconds [2]. A hamming moving window of 2 seconds with 1.8 seconds overlapping is shifted throughout the EEG deriva-
tions, i.e. step shift is 0.2 seconds. This much finer resolution is needed in order to allow the detection of alterations in the spectral band on intervals with duration less than the minimum spindle duration.

Once average power in the sigma band \( \text{PSD}_{\sigma} \) has been established throughout the signal, the algorithm performs a second analysis over obtained values in order to establish a baseline (10 seconds) to which compare instant power values of spindle activity, thus allowing characterization of intervals with increments or decrements with respect to normal sigma activity in the signal. This process is repeated for each instant power value previously calculated, after which those intervals exceeding 2 times baseline value with duration ≥ 0.5 seconds are marked as possible spindle events.

A false positive detection process is finally run in order to discard false events caused by (1) interferences produced by increments in contiguous frequency bands such as alpha (8-12 Hz) and EMG artifacts (> 16 Hz) that can cause harmonics in the spindle band, and (2) events occurring during non-sleep periods such as clear stable intervals of alpha or beta activity. Previous detected activity in the EEG (\( \alpha, \beta, \theta, \delta \)) is used here to detect those false positives and to discard them as spindle events.

2.2 Fuzzy Reasoning Process

The second step of the analysis is divided into four submodules, each one being the responsible to accomplish the analysis regarding to one of the four considered sleep stages (awake, drowsy sleep, deep sleep and REM). Each submodule is implemented in the form of a Fuzzy Inference System (FIS) of type Mamdani [12]. This allows us to fulfill the requirement that knowledge should be accessible and extracted in form of human-like decision rules (fuzzy rules). Actually overall knowledge –extracted from medical expertise- is structured into four independent sets of rules, each one involving a particular sleep stage. Thus, the output of each submodule consists of a value \( \mu \) in the real interval \([0, 1]\), which represents the degree of membership for the current instant of time under analysis, i.e. \( \mu_W, \mu_{DS}, \mu_{DEEP} \) and \( \mu_{REM} \). In total, 93 fuzzy rules are implemented derived from medical expertise and further adapted based on empirical trials by the authors.

Input vector to these submodules is composed of the parameter information extracted on the previous analysis steps, composed of five values characterizing eye movements \( A_{\text{EOG}(i)} \), muscle activity \( A_{\text{EMG}(i)} \), EEG activity \( \text{PSD}_{\sigma}(i), \text{PSD}_{\beta}(i), \text{PSD}_\theta(i) \) and presence of sleep spindles (number of). This information is fed into the subsequent fuzzy reasoning modules to obtain the respective degrees of memberships for each state (W, DS, DEEP and REM).

In order to mimic humans’ procedure to capture dynamics of the signals and promote smooth transitions, parameters of the signals are averaged in the environment of the current instant under analysis. Specifically input for instant \( i \) is calculated by averaging input parameters on the interval \([i-15, i+15]\). Trapezoidal fuzzy sets are used for the partition of the input variables. For the \( A_{\text{EOG}} \) signal 3 fuzzy sets (low, medium and high) are used. Similarly 3 fuzzy sets (relaxed, medium and tense) are established for the parameter \( A_{\text{EMG}} \). In the case of the EEG, each of the corresponding frequency bands (\( \alpha, \beta, \theta, \delta \)) resulted in a variable partitioned again in 3 sets namely low, medium
and high. Output variables were partitioned by defining 5 fuzzy sets uniformly distributed along the interval [0, 1] (very low, low, medium, high, and very high). All the fuzzy sets were partially superimposed in order to exploit smoothing transitions and improve generalization capabilities of the FIS. For the configuration of the FIS the minimum was chosen as the T-norm operator for the conjunction and for the implication. On the other hand, the maximum was chosen as the S-norm operator for the disjunction and for the aggregation. Defuzzification was performed by using the center-of-gravity method.

Once all the seconds of the recording were analyzed, a continuous evolution of the degrees of membership for the different sleep stages is obtained. This output can be observed in Fig. 2, in which evolutions are represented for a full PSG recording. Note that this representation provides more information than the discrete hypnogram since (i) it is provided in a higher rate -for each second- compared to 30s epochs of R&K/AASM, (ii) the natural continuous evolution of biological processes is maintained and (iii) the information regarding each sleep state is also kept individually available.

Fig. 2. Evolution of sleep states throughout a full PSG recording estimated by their respective degrees of membership

### 2.3 Hypnogram generation

The proposed representation based on the evolution of the corresponding degrees of membership is used here to generate the classical hypnogram. The interest in going back to the epoch-based hypnogram from the continuous representation is diverse: (1) it can be a way to show how this new proposed representation preserves all the hypnogram information, (2) it shows how this fuzzy representation can be used as an alternative method within available literature regarding the problem of the automatic hypnogram generation in sleep. On the other hand (3), the only way to assess on the validity of (1) and (2) is to perform a validation process against experts’ manually generated hypnograms, only possible though discretization of the continuous representation and going back to the epoch-based hypnogram.

Thus, taking the continuous representation, some post-processings are performed:
1. An average of the second-by-second output of each subsystem within each epoch is performed to be used as the resulting degree of membership for the corresponding epoch. The epoch is finally assigned to a discrete stage (W, DS, DEEP, REM) by taking the corresponding maximum averaged degree of membership.

2. Previous processing could lead to noisy isolated epochs that break up the normal evolution of sleep according to medical scoring guidelines. Therefore a second step searches for unusual phase transitions such as direct transitions from W to DEEP or DEEP to REM, and then assigns the most possible one according to the normal sleep evolution [13].

3. In cases where the degree of membership regarding DS and REM is similar, the DS stage is assigned in case sleep spindles are detected within the epoch.

4. Additionally, in regions where similar degrees of membership for different stages are achieved, final labeling is decided based on the trend marked by the immediately previous and subsequent epochs.

Fig. 3 shows the result of applying such post-processing to the outputs presented at Fig. 2. Resulting hypnogram can now be used to perform a validation process against expert’s classification.

3 Results

In order to develop and validate the proposed method, a set of PSG recordings from real patients is used. The recordings are taken from the Sleep Heart Health Study (SHHS). This database, granted by the Case Western Reserve University, emerged from a multi-center cohort study implemented by the National Heart Lung & Blood Institute to determine cardiovascular and other consequences of sleep-disordered breathing. Details about the design of SHHS study can be found in [14]. Each recording comes with the annotations of physicians’ off-line scorings following R&K procedure [15].

A total of 26 patients (mean age ± std.deviation: 68.5±7.7, 8 females) from the SHHS database were randomly chosen for the evaluation of the system without previous knowledge of the clinical history of the participants. According to the AASM criteria for sleep staging, both EEG derivations C3/A2 and C4/A1, the submental EMG channel and both EOG derivations (left and right), were used from these recordings. EEG and EMG signals were sampled at 125 Hz whereas EOG was recorded at
50 Hz. In total the 26 recordings involve more than 16800 minutes of sleep. Annotations made by SHHS scorers are taken as reference for the validation process. It is important to remark that since the experts follow the R&K procedure -30 seconds labeling, not continuous- the epoch-based output resulting from the post-processings described on the previous section has to be used in order to allow the validation process. Table 1 shows the means and the standard deviations for the sensitivities, specificities and the values for the Area Under ROC Curve (AUC) obtained for the four considered stages of sleep and for the 26 patients.

**Table 1. Validation results between expert and system (mean ± std. deviation)**

<table>
<thead>
<tr>
<th></th>
<th>Awake</th>
<th>Drowsy</th>
<th>DEEP</th>
<th>REM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sens.</td>
<td>0.88±0.07</td>
<td>0.81±0.08</td>
<td>0.75±0.23</td>
<td>0.84±0.19</td>
</tr>
<tr>
<td>Spec.</td>
<td>0.96±0.04</td>
<td>0.89±0.08</td>
<td>0.97±0.03</td>
<td>0.95±0.04</td>
</tr>
<tr>
<td>AUC</td>
<td>0.92±0.04</td>
<td>0.85±0.05</td>
<td>0.86±0.11</td>
<td>0.89±0.11</td>
</tr>
</tbody>
</table>

As it can be seen from data the method works especially well in the discrimination of wakefulness and REM sleep (average sensitivity/specificity of 0.88/0.96 and 0.84/0.95 respectively). Sensitivity and specificity slightly decreases for DS (0.81/0.89). However, the lower sensitivity resides in the detection of stage DEEP (0.75) while on the other hand the highest specificity is achieved (0.97). Indeed, attending to AUC values, it can be said that best results are obtained regarding W and REM (AUCs of 0.92 and 0.89), followed by DEEP (AUC of 0.86) and DS (AUC of 0.85). Subsequently, Table 2 presents the accumulated contingency table over the 26 patients. Calculating Cohen’s kappa index over Table 2 results in $\kappa = 0.76$.

**Table 2. Accumulated contingency table. In parenthesis associated frequency ratio is indicated**

<table>
<thead>
<tr>
<th></th>
<th>W (0.31)</th>
<th>DS (0.03)</th>
<th>DEEP (0.01)</th>
<th>REM (0.01)</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>12208</td>
<td>1076</td>
<td>18 (&lt; 0.01)</td>
<td>576 (0.01)</td>
</tr>
<tr>
<td>DS</td>
<td>983 (0.02)</td>
<td>14357 (0.36)</td>
<td>1062 (0.03)</td>
<td>1152 (0.03)</td>
</tr>
<tr>
<td>DEEP</td>
<td>16 (&lt; 0.01)</td>
<td>666 (0.02)</td>
<td>2293 (0.06)</td>
<td>25 (&lt; 0.01)</td>
</tr>
<tr>
<td>REM</td>
<td>87 (&lt; 0.01)</td>
<td>749 (0.02)</td>
<td>36 (&lt; 0.01)</td>
<td>4509 (0.11)</td>
</tr>
</tbody>
</table>

### 4 Discussion and Conclusions

In this paper a new method for the evaluation of the macrostructure of sleep is presented. The method comprises mainly three processing tasks in which the first one acts over the raw signals in order to extract features that, subsequently, are fed into a reasoning stage organized as four fuzzy submodules, each one involving a different state of sleep (W, DS, DEEP, REM). A degree of membership representing the continuous evolution of the corresponding sleep stage is obtained as the output for each submodule. In this work, continuous means second-by-second. Of course, as long as
we are processing digital signals it would never be possible to achieve a real continuous output, but on the other hand, we can asymptotically approximate it by increasing the rate of the analysis (only limited by the sampling frequency of the digitalized signal).

The main objective of the proposed approach is to overcome limitations of epoch-based methodologies such as AASM and R&K regarding their low temporal resolution and the unnatural discrete classification of sleep. By using fuzzy logic categorical classifications can be avoided and soft transitions can be exploited. Such properties allow us approximating the continuous evolution of the sleep though its different states.

In addition to the advantages of the proposed continuous representation of the hypnogram (see Fig. 2), we have shown how this approach can also be used in order to obtain the classical epoch-based representation (see Fig. 3). At this respect, a validation process has been performed over 26 patients comparing discrete output from our method after some post-processings, to expert’s manually staging of sleep. Attending AUC indexes best results were achieved regarding discriminating capabilities on wakefulness and REM sleep. The highest specificity was obtained for the DEEP phase, however at the cost of lower sensitivity. On the other hand, drowsy sleep characterization showed slightly lower values of agreement with expert’s scorings while maintaining similar values of sensitivity and specificity. Further analysis by adjusting possible agreement occurring by chance on the 26 patients, established an overall inter-rater agreement for the method of $\kappa = 0.76$. This value is in accordance with recently reported literature on human inter-rater reliability on sleep scoring [16] (0.76 for AASM criteria and 0.68 for R&K standard).

These results seem to support the idea that it is possible to reconstruct the classic hypnogram from the continuous output –possible in the terms that method’s output is comparable to the classification made by a human expert. This suggests that this representation preserves the information contained in the discrete hypnogram, in fact constituting a superset. Therefore the method can be used as a supporting tool to assist physicians in the current sleep scoring task.

It has to be remarked that with our development we fuse S1 and S2 into a unique state representing drowsy sleep (DS), thus direct comparison of the results has to be carefully made. Another possible source of bias to be taken into account is given by our population age range (mean age ± std.deviation: 68.5±7.7) which could affect the results respect to other population sectors.

Future work will emphasize in these aspects and in a better characterization of the microstructure level by incorporating detection capabilities for additional transients such as k-complexes or vertex waves, expecting allowing a better characterization of sleep. It is also our intention to integrate this method into a system to diagnose sleep apnea syndrome [17] where sleep hypnogram will act as the neurophysiologic context in order to better classify respiratory apneic events.

**Acknowledgements.** This work has been funded by Xunta de Galicia under research project 09SIN003CT and program CN2011/007 partially supported with FEDER funds. We would like to thank members of Case Western Reserve University
and the Sleep Health Heart Study for their support with the PSG recordings database. The opinions expressed in the paper are those of the authors and do not necessarily reflect the views or the endorsement of The Sleep Heart Health Study.

References