Biomed Tech 2013; 58 (Suppl. 1) © 2013 by Walter de Gruyter · Berlin · Boston. DOI 10.1515/bmt-2013-4164

AUTOMATIC VALIDATION AND QUALITY BASED READJUSTMENT OF MANUALLY SCORED EEG AROUSAL

Dennis Lerch[1], Thomas Penzel[2] and Reinhold Orglmeister[1]

[1]Electronics and Medical Signal Processing, TU Berlin, Germany

[2]Interdisciplinary Sleep Medicine Center, Charité - Universitätsmedizin, Germany

dennis_lerch@gmx.de

Abstract: *A knowledge of arousals during sleep is important to attain a deeper understanding regarding cardiovascular diseases. Manual scoring is time consuming and not always accurate. Automatic approaches are even worse inter alia due to inaccurate learning data. This paper presents an algorithm to improve the accuracy of manually scored data. Also a measure of quality is introduced to judge the automatically estimated results.*

Keywords: *Arousal detection, validation, quality estimation*

Introduction

An arousal is characterized by an abrupt frequency shift in the EEG for from 3 s up to 15 s ([1]). Different studies (see [2]) have shown that manually scored events differ not only among scorers, but also between analyses of one scorer of the same data. Reliable data is required to set up an algorithm which determines the arousal as closely as possible. This paper shows an algorithm to readjust the manually scored starting time and to give a measure of its quality.

Methods

In this work, overnight polysomnographic recordings from the Daphnet-Project[1] are used. Different medical technicians scored the records, but each record was only scored once. Therefore, the following assumptions can be made:

1. The starting time of an arousal may differ from the true beginning, in some cases several seconds are possible.
2. Misclassification of non-existent arousals can occur.

$S = 446$ segments with manually scored arousals, extracted from 33 patients, are used. Each segment takes $L = 40\,\mathrm{s}$ and starts 20 s before and ends 20 s after the beginning of an arousal. According to [3], arousals are well visible in the frequency bands β_1 (16 to 24 Hz), β_2 (24 to 32 Hz) and γ (32 to 48 Hz). In total, instantaneous power of $B = 2 \cdot 3 = 6$ EEG bands (from derivation C3-A2 and C4-A1) are used.
The beginnings of arousals are detected by a threshold based algorithm. The thresholds are derived from the statistical properties of all segments.

[1]Daphnet-Project: Dynamical Analysis of Physiological Networks, EU-Project 2006-2009, Ref. 018474-2, examination of patients suffering from sleep apnoea, periodic leg movement and insomnia.

Determination of thresholds

Thresholds of power bands are used to ensure a qualitatively good data basis. The following procedure is performed for each power band b, where $1 \leq b \leq B$.
The 0.7-quantiles of the segments are determined and visualised in a histogram. A threshold $\tau_{band}(b)$ will be chosen to separate the two essentially visible intervals at lower and higher power. A 0.7-quantile implies that at least 70 % of all power values are less or equal than the given quantile value. Hence, no more than 30 %, or 12 s of the values of a 40 s segment, exceed the quantile value which can be caused by an arousal. Power bands for which the 0.7-quantile lies above their associated threshold will be classified as bad, meaning they are too much affected by artifacts.
Thresholds for arousal detection are empirically chosen from the histograms of the corresponding values of good power bands. The procedure is similar to the previous one. $\tau_{ar}(b)$ represents the found thresholds of the arousals in the power band b.

Arousal detection

The thresholds $\tau_{ar}(b)$ are applied to every segment in order to detect possible arousals. Unfortunately, the actual power during true arousals does not consistently exceed the threshold $\tau_{ar}(b)$. Therefore, threshold exceeding sections are combined to one interval if the temporal gaps are less than 2 s. Afterwards, all intervals with a duration of less than 3 s or more than 15 s are discarded. Remaining intervals, or so called blocks, must be tested if an arousal exists.
NON AROUSAL: If in all good bands of a segment no blocks exist, the segment will be marked as "No arousal occurred". $\mathbb{M}_0$ is the set of all concerning segments.
AROUSAL: An arousal in a segment will be found if at least one block in at least one good power band exists. These segments form the set $\mathbb{M}_1$. To estimate the beginning of the arousal, the median of all beginnings of blocks from all good power bands with exactly one block is determined. If only good power bands with more than one block exist, all existing blocks will be taken into consideration. But, in this case the quality of estimation is set to zero.

Quality of estimation

For a better evaluation of the estimated arousals, different quality measures will be introduced.
Quality of segments are expressed by Eq. 1

$$q_{seg.}\left(s\right) = \frac{n\_good\_band\left(s\right)}{B}, \quad (1)$$

where $1 \leq s \leq S$ and $n\_good\_band\left(s\right)$ equals the number of good power bands in segment s.

Quality of estimation as NON AROUSAL: The quality is described with $q_{non\,ar.}\left(s^{*}\right) = q_{seg.}\left(s^{*}\right)$, with $s^{*} \in \mathbb{M}_0$.

Quality of estimation as AROUSAL: The quality of a found arousal is the product of different quality measures (see Eq. 2).

$$q_{ar.}\left(s^{**}\right) = q_{seg.}\left(s^{**}\right) \cdot q_1\left(s^{**}\right) \cdot accuracy\left(s^{**}\right), \quad (2)$$

where $s^{**} \in \mathbb{M}_1$. $q_1\left(s^{**}\right)$ is the ratio between the number of good power bands with exactly one block and the number of all good power bands within the segment s^{**}.
The $accuracy$ weighs variations of possible beginnings of arousals in different power bands. It uses the standard deviation according to Eq. 3.

$$accuracy\left(s^{**}\right) = 1 - \frac{std\left(block\_begin\left(s^{**}\right)\right)}{S}. \quad (3)$$

$block\_begin$ is a vector containing all beginnings of blocks from power bands with exactly one block.

Results

The thresholds used are listed in Tab. 1. Tab. 2 shows the absolute frequencies of the qualities of the segments and of the not arousals detected. Two segments are completely unusable, and not all power bands in all segments are good enough for using. No arousals are found in 100 segments. The quality of the classification is very high. The other 364 segments contain arousals. Whereas the quality of 29 segments is $0\,\%$, that of 117 segments is $16.67\,\%$ and of 218 segments above $16.67\,\%$. A lot of segments contain only one power band with one block. The other power bands of the segments concerned contain more than one block in many cases. Therefore, these are currently not used to estimate the beginning of an arousal. This is the reason for a peak at $16.67\,\% \approx 1/6$. But a detailed investigation shows that very often blocks in different power bands exist which are located temporally close to each other. These blocks must be identified and should also be taken into consideration for estimating the beginning of an arousal. The quality $q_{is\,ar.}\left(s^{**}\right)$ will rise and the peak at $16.67\,\%$ will be shifted to a higher value.

In Fig. 1, the histogram of the differences between the manually scored and the automatically detected beginnings of the arousals are visualised. A large number of automatically detected arousals are located close by the manually scored arousals. Most of the estimated beginnings lie in the interval $[-5, \ldots, 5]\,s$ and correspond very well to the expected differences, determined from earlier spot tests. This allows the conclusion that the algorithm performs well.

Table 1: Thresholds for power band quality estimation and arousal detection of used bands β_1, β_2 and γ.

threshold	$\beta_1/\left(\mu V\right)^2$	$\beta_2/\left(\mu V\right)^2$	$\gamma/\left(\mu V\right)^2$
τ_{band}	35	30	50
τ_{ar}	11	40	50

Table 2: Number of segments with qualities for segment evaluation and the detection of NON AROUSAL.

quality / %	absolute freq. of $q_{seg.}\left(s\right)$	absolute freq. of $q_{non\,ar.}\left(s^{*}\right)$
0.00	2	0
50.00	9	2
66.50	3	1
83.00	6	1
100. 00	445	96

Discussion

The algorithm used is well suited to estimate the beginning of arousals of manually inaccurately scored EEG data. Furthermore, a quality measure exists to judge the results. In future work, the threshold decision will be replaced by a probability based approach. Moreover, an algorithm will be developed to accurately handle power bands with multiple blocks in order to identify the most likely blocks for estimating the beginning of an arousal.

Bibliography

[1] R. B. Berry and et al., "Rules for scoring respiratory events in sleep: Update of the 2007 AASM manual," *Clinical sleep medicine*, vol. 8(5), pp. 597–619, 2012.

[2] M. H. Bonnet and et al., "The scoring of arousal in sleep: Reliability, validity and alternatives," *Journal of Clinical Sleep Medicine*, vol. 3, pp. 133 – 151, 2007.

[3] H. P. Collin, *Detection of Cortical Arousals in Sleep EEG*. PhD thesis, University of Hawaii, 2010.

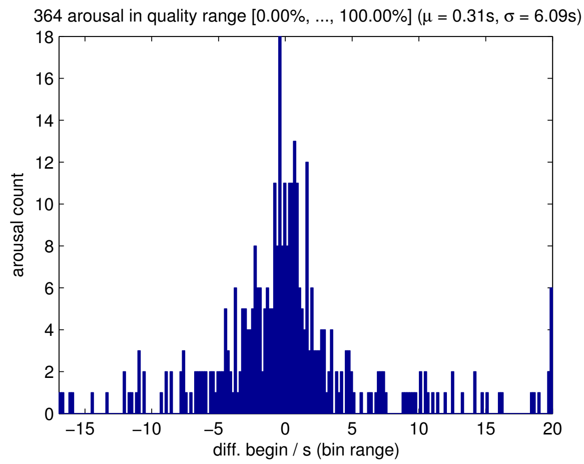

Figure 1: Differences of beginnings of arousals.