31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Sleep Stage Classification with Low Complexity and Low Bit Rate

Jussi Virkkala, *Member, IEEE,* Alpo Värri, Joel Hasan, Sari-Leena Himanen, Kiti Müller

Abstract—**Standard sleep stage classification is based on visual analysis of central (usually also frontal and occipital) EEG, two-channel EOG, and submental EMG signals. The process is complex, using multiple electrodes, and is usually based on relatively high (200-500 Hz) sampling rates. Also at least 12 bit analog to digital conversion is recommended (with 16 bit storage) resulting in total bit rate of at least 12.8 kbit/s. This is not a problem for in-house laboratory sleep studies, but in the case of online wireless self-applicable ambulatory sleep studies, lower complexity and lower bit rates are preferred. In this study we further developed earlier single channel facial EMG/EOG/EEG-based automatic sleep stage classification. An algorithm with a simple decision tree separated 30 s epochs into wakefulness, SREM, S1/S2 and SWS using 18-45 Hz beta power and 0.5-6 Hz amplitude. Improvements included low complexity recursive digital filtering. We also evaluated the effects of a reduced sampling rate, reduced number of quantization steps and reduced dynamic range on the sleep data of 132 training and 131 testing subjects. With the studied algorithm, it was possible to reduce the sampling rate to 50 Hz (having a low pass filter at 90 Hz), and the dynamic range to 244 μV, with an 8 bit resolution resulting in a bit rate of 0.4 kbit/s. Facial electrodes and a low bit rate enables the use of smaller devices for sleep stage classification in home environments.**

I. Introduction

SLEEP staging is an important instrument in clinical and research sleep studies. Based on recent standards, sleep stage classification is carried out visually based on frontal, central and occipital EEG, two-channel EOG, and submental EMG [1]. In the old standard, only a single central EEG channel was required with EOG and EMG [2]. The minimum required sampling rates are 200 Hz (500 Hz desired) and at least 12 bits per sample [1]. Usually, data is stored with a 16 bit resolution [3, 4]. This results in a total bit rate of at least 12.8 kbit/s. The recording process is also complex. In particular, the placement of multiple EEG electrodes within the hairline and visual analysis of signals are laborious tasks requiring trained personnel.

Various alternatives exist for estimating sleep structure, which are easier than the standard procedure based on this visual analysis of EEG, EOG and EMG. Some methods are based on single-channel central EEG [5], EOG [6-8] or forehead EEG analysis [9-11]. Sometimes only averaged features and calculated results are stored [12], and no raw data. Some limited analyses can also be performed without placing EEG electrodes. The measurement of movements of limbs - actigraphy - is a clinically accepted measure for sleep assessment [13]. Separation of wakefulness, SREM and NREM can only be performed using ECG information [14, 15] and wakefulness and sleep separation only using respiration [16].

There is an apparent gap in technologies between simple movement sensors and standard polysomnography. In other fields, established limited techniques exist, such as amplitude-integrated EEG for long term monitoring [17]. Sleep recording systems with low bit rates would result in lower power usage and thus lower size. Storage of raw data is however essential for applying different algorithms and for visual confirmation of the analysis. Bit rate can be reduced through various paths: 1) by reducing the number of recorded channels, 2) by reducing the sampling rate, and 3) by reducing the number of bits used to represent the selected dynamic range.

In our previous study [8] we developed single-channel analysis for separating wakefulness, SREM, S1/S2 and SWS epochs. The results obtained in validation data using a low dynamic range (300 Hz, 8 bit, ±260 μV) device were almost identical to those using a higher dynamic range (200 Hz, 16 bit, ±7.8 mV) device [8], which we had not anticipated. As the validation was made on different subjects in the previous study, we used the same data and simulated the effects of reducing sampling rate, reducing both quantization steps and the dynamic range in this study.

We also implemented a low complexity recursive infinite impulse response (IIR) filter instead of the originally used discrete Fourier transform (DFT) and inverse discrete Fourier transform (IDFT) approach. Low complexity analysis would be beneficial for the implementation of real time analysis in portable devices.

II. SUBJECTS

The single-channel algorithm was further developed and tested on data previously used for preliminary single-channel

Manuscript received April 23, 2009 and revised June 20, 2009. Jussi Virkkala is with Sleep Laboratory, Finnish Institute of Occupational Health, Helsinki, Finland and Department of Clinical Neurophysiology, Medical Imaging Centre, Pirkanmaa Hospital District, Tampere, Finland, phone: +358-40-5680360, fax: +358-9-5884759, e-mail: jussi.virkkala@ttl.fi, www: neuroupdate.com. Alpo Värri is with Department of Signal Processing, Tampere University of Technology, Tampere, Finland (e-mail alpo.varri@tut.fi). Joel Hasan and Sari-Leena Himanen are with Department of Clinical Neurophysiology, Medical Imaging Centre, Pirkanmaa Hospital District, Tampere, Finland (e-mail joel.hasan@pshp.fi, sari-leena.himanen @pshp.fi). Kiti Müller is with Finnish Institute of Occupational Health, Helsinki, Finland (e-mail kiti.muller@ttl.fi)

algorithms [8]. A total of 132 subjects' (age 26-61) data were used to train the algorithm and 131 different subjects' (age 28-60) data were used to test the trained algorithm.

III. METHODS

A. Equipment and visual scoring

Data was recorded in a sleep laboratory with Embla A10 (Embla, Broomfield, USA), at a bandwidth of 0.5-90 Hz and sampling rate of 200 Hz. The system has a 16 bit resolution, and a 238 nV/bit resolution was used. Thus the dynamic range was from $-(2^{15}\cdot 238)$ nV=-7.8 mV to $(2^{15}-1)\cdot 238$ nV=7.8 mV. Data was recorded according to the standard [2] with central EEG, two-channel EOG and submental EMG. The recordings were visually scored according to the common standard [2] by an experienced sleep technologist. Automatic single-channel analysis was based on left EOG (slightly lateral and 1 cm up from the outer canthus) referenced to the right EOG (slightly lateral and 1 cm down from the outer canthus).

B. Sleep stage determination

The applied sleep staging algorithm is an extension of a previously published method [8]. The analysis was carried out at 0.5 s intervals using overlapping 2 s segments for visually scored 30 s epochs. The reference analysis was performed using DFT (for 18-45 Hz beta) and IDFT (for 0.5-6 Hz peak-to-peak amplitude).

The new analysis was based on a Butterworth IIR filtered signal. The IIR filter order was 2 for the 0.5-6 Hz band and 4 for the 18-45 Hz band. For each segment, a 0.5-6 Hz filtered peak-to-peak amplitude and 18-45 Hz beta power were calculated after the Hann window was applied. With the IIR approach, the beta power was calculated as squared sum of filtered and window-applied signal segments. With 50 Hz resampled and aliased signal instead of 18-45 Hz band pass filter high pass filter with 18 Hz cutoff was used. The 30 s epoch stages were determined by the following criteria:

ST The number of segments (density) with below-threshold beta power values had to be above threshold. The beta power threshold was either fixed (online approach) or was based on the median beta value of the subject's data (this automatically obtained value was added to fixed threshold, offline approach). Used for separation of SREM, S1/S2, SWS from W.

NREM The maximum amplitude difference of the 0.5-6 Hz peak-to-peak amplitude values within a 30 s epoch had to be below threshold. Only amplitudes with a beta power below threshold were considered. If SREM was detected, 11 adjacent epochs (if not detected as awake) were also automatically marked as SREM. Used for separation of S1/S2, SWS from SREM.

SW_3T The number of segments (density) with an above-threshold 0.5-6 Hz peak to peak amplitude and beta below another threshold had to be above density threshold. Used for separation of SWS from S1/S2.

The original analysis had five parameters [8]. Here, two additional beta thresholds were added to the *NREM* and to *SW_3T* estimation, and the beta band was extended to 18-45 Hz from 18-30 Hz. SREM detection affected 11 adjacent epochs rather than the 3 in the original analysis. For each *ST, NREM* and *SW_3T* optimal thresholds were sought (for largest Cohen's Kappa) separately for the corresponding binary decisions, using the data of the training data set. All seven parameters (*ST*: beta threshold, segment density threshold; *NREM*: amplitude and beta threshold; *SW_3T*: amplitude and beta threshold, segment density threshold) were fixed across subjects. A simple, three-step, decision tree was used in the final sleep stage estimation [8].

C. Simulation

Dynamic range was modified by clipping data to be between $-(2^n\cdot 238)$ nV and $(2^n-1)\cdot 238$ nV, where n is an integer. Another reduction of bit rate was carried out by rounding values to the closest values of $2^n \cdot 238$ nV, where n is the integer. For instance, ±244 µV range with a 1.9 µV quantization step corresponds to 8 bit. Resampling both with and without anti-aliasing filter was performed before dynamic range modifications.

D. Statistical analysis

An automatic detection of 30 s epochs was compared to the visual scoring based on the standard method [2]. Agreement p_o and Cohen's Kappa κ [18] were used to evaluate the detection. Cohen's Kappa is the proportion of agreement after change agreement p_c is removed from consideration. Using probabilities p_{ij} from the agreement matrix, κ can be defined as.

$$\kappa = \frac{p_o - p_c}{1 - p_c} \quad p_o = \sum_i p_{ii} \quad p_c = \sum_i \left[(\sum_j p_{ji})(\sum_j p_{ij}) \right]$$

The Cohen Kappa values reported are from agreement matrix with all testing data pooled together. Statistical differences in Cohen's Kappa values between different methods are calculated using a pair-wise comparison of subject Cohen's Kappa values using The Wilcoxon Signed Ranks Test.

IV. RESULTS

All reported results are from the testing data set after optimization of parameters using different training data sets. With the extended DFT and IDFT method four-stage separation agreement with the fixed beta was 75% (Cohen's Kappa 0.61) and with the offline adjusted beta threshold, agreement was 77% (Cohen's Kappa 0.63), Table I. Using butterworth IIR filtering and the original 200 Hz data agreement, 74% and 77% (Cohen's Kappa 0.60 and 0.63) were obtained, as shown in Table II. In all tables, PPV

indicates positive predictive value.

TABLE I
AGREEMENT MATRIX FOR TESTING DATA SET USING DFT METHOD. ROWS REPRESENT STANDARD VISUAL ANALYSIS AND COLUMNS THE AUTOMATIC SINGLE CHANNEL ANALYSIS.

	Wake	SREM	S1/S2	SWS	Sensitivity
Wake	**14337**	1981	4800	263	67.1%
SREM	2285	**17377**	4303	56	72.3%
S1/S2	6278	4443	**57826**	4485	79.2%
SWS	481	372	4030	**11572**	70.3%
PPV	61.3%	71.9%	81.5%	70.7%	
Agreement					75.0%
Cohen's Kappa					0.61
WITH OFFLINE BETA THRESHOLDS					
	Wake	SREM	S1/S2	SWS	Sensitivity
Wake	**15378**	1629	4151	223	71.9%
SREM	1948	**17443**	4574	56	72.6%
S1/S2	5229	4408	**58934**	4461	80.7%
SWS	345	375	4179	**11556**	70.2%
PPV	67.2%	73.1%	82.0%	70.9%	
Agreement					76.6%
Cohen's Kappa					0.63

TABLE II
AGREEMENT MATRIX FOR TESTING DATA SET USING IIR METHOD. ROWS REPRESENT STANDARD VISUAL ANALYSIS AND COLUMNS AUTOMATIC SINGLE-CHANNEL ANALYSIS.

	Wake	SREM	S1/S2	SWS	Sensitivity
Wake	**15330**	1766	4100	185	71.7%
SREM	2893	**16398**	4664	66	68.3%
S1/S2	8403	3858	**55933**	4838	76.6%
SWS	974	225	3174	**12082**	73.4%
PPV	55.5%	73.7%	82.4%	70.4%	
Agreement					73.9%
Cohen's Kappa					0.60
WITH OFFLINE BETA THRESHOLDS					
	Wake	SREM	S1/S2	SWS	Sensitivity
Wake	**15392**	1668	4137	184	72.0%
SREM	1846	**16856**	5258	61	70.2%
S1/S2	5499	3927	**58736**	4870	80.4%
SWS	427	227	3654	**12147**	73.8%
PPV	66.4%	74.3%	81.8%	70.4%	
Agreement					76.5%
Cohen's Kappa					0.63

With resampling to 50 Hz, 8-bit sampling, and ±244 μV dynamic range agreements were 72% and 75% (Cohen's Kappa 0.57 and 0.61), Table III. With resampling to 50 Hz without an anti-aliasing filter, corresponding agreements were 75% and 77% (Cohen's Kappa 0.61 and 0.64), as Table IV shows. Cohen's Kappa using the 50 Hz resampled IIR method was lower than that of the 200 Hz, or the 50 Hz aliased IIR method.

V. Discussion

Many different methods exist for automatic sleep stage classification. Most methods are complex with multiple EEG, EOG and EMG electrodes, or only distinguish between wakefulness and sleep or wakefulness, SREM and NREM. In this study, an automatic single-channel EOG algorithm was further developed for separating wakefulness, SREM, light sleep (S1/S2), and deep sleep (SWS). Electrodes were placed at a standard location [2] to record EMG, EOG and EEG activity. Wakefulness was detected using EMG beta (18-45 Hz) activity, SREM based on high amplitude 0.5-6 Hz activity (assumed to be eye movements) and SWS based on medium amplitude 0.5-6 Hz activity.

TABLE III
AGREEMENT MATRIX FOR TESTING DATA SET USING RESAMPLED IIR METHOD. ROWS REPRESENT STANDARD VISUAL ANALYSIS AND COLUMNS AUTOMATIC SINGLE-CHANNEL ANALYSIS

	Wake	SREM	S1/S2	SWS	Sensitivity
Wake	**14287**	2458	4312	324	66.8%
SREM	3965	**15616**	4366	74	65.0%
S1/S2	9097	4448	**55121**	4366	75.5%
SWS	872	295	3444	**11844**	72.0%
PPV	50.6%	68.4%	82.0%	71.3%	
Agreement					71.8%
Cohen's Kappa					0.57
WITH OFFLINE BETA THRESHOLDS					
	Wake	SREM	S1/S2	SWS	Sensitivity
Wake	**14611**	2247	4245	278	68.3%
SREM	2948	**16237**	4765	71	67.6%
S1/S2	5377	4552	**58657**	4446	80.3%
SWS	269	299	3912	**11975**	72.8%
PPV	63.0%	69.6%	81.9%	71.4%	
Agreement					75.2%
Cohen's Kappa					0.61

TABLE IV
AGREEMENT MATRIX FOR TESTING DATA SET USING ALIASED IIR METHOD. ROWS REPRESENT STANDARD VISUAL ANALYSIS AND COLUMNS AUTOMATIC SINGLE-CHANNEL ANALYSIS

	Wake	SREM	S1/S2	SWS	Sensitivity
Wake	**14518**	1919	4820	124	67.9%
SREM	2274	**16973**	4731	43	70.7%
S1/S2	6394	4284	**58212**	4142	79.7%
SWS	593	343	3796	**11723**	71.2%
PPV	61.1%	72.2%	81.3%	73.1%	
Agreement					75.2%
Cohen's Kappa					0.61
WITH OFFLINE BETA THRESHOLDS					
	Wake	SREM	S1/S2	SWS	Sensitivity
Wake	**15015**	1717	4547	102	70.2%
SREM	1582	**17297**	5101	41	72.0%
S1/S2	4947	4305	**59650**	4130	81.7%
SWS	352	343	4045	**11715**	71.2%
PPV	68.6%	73.1%	81.3%	73.3%	
Agreement					76.9%
Cohen's Kappa					0.64

In our previous study, DFT and IDFT was used to calculate beta power and peak-to-peak amplitudes respectively. In this study, low order IIR filtering was used, resulting in lower complexity of the algorithm and making it more suitable for online applications.

With a 50 Hz sampling rate, Cohen's Kappa was smaller using the resampled IIR method but not with the aliased IIR method. Possible folding of powers >25 Hz resulted in optimal frequency weighting. This folding has been used by

e.g. Ehlert et al. in sleep staging [11] and more commonly in EMG analysis [19]. Excluding the anti-aliasing filter will change the time domain properties of the high frequency part of the signal and should be used with caution.

It has to be emphasized that the current setup using two facial electrodes is not optimal for recording lower amplitude EEG activity such as spindle and alpha activity. These activities are better recorded with an EEG electrode on the central and occipital region. With different electrode configurations and algorithms utilizing spindle and alpha activity, different results are likely to occur by lowering the bit rate or by excluding the anti-aliasing filter.

Despite this, the results obtained here suggest that low complexity (single-channel, self-applicable facial electrodes, simple algorithm) and low bit rate automatic sleep staging should be evaluated as a complementary technique between various movement sensors and ambulatory polysomnography. Other techniques based on, for instance heart rate and respiration, are complementary, enabling for instance studies of heart rate in different sleep stages.

VI. Conclusion

With the studied low complexity single-channel algorithm using two facial electrodes it was possible to reduce the bit rates without any significant effect on sleep classification agreement. Reduction was achieved by reducing the sampling rate from 200 Hz to 50 Hz, using only 8-bit quantization instead of 16-bit and by limiting the dynamic range to ±244 µV. The initial implementation of developed algorithms for online analysis has been carried out [20].

Wireless techniques exist for high bit rates, but lower complexity and lower bit rates will reduce power usage and thus also the system size. A smaller device would make it easier and more comfortable to record multiple nights, complementing the diagnostic laboratory sleep recordings and enabling large scale field sleep studies.

Acknowledgment

We thank Riitta Velin and Susan Pihl for their work in recording and scoring the material, and Alice Lehtinen for language editing.

References

[1] C. Iber, S. Ancoli-Israel, A. Chesson, and S. Quan, *for the American Academy of Sleep Medicine. The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology, and Technical Specifications*, 1st ed.: Westchester, Illinois: American Academy of Sleep Medicine, 2007.

[2] A. Rechtschaffen and A. Kales, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects," No. 204 in National Institutes of Health Publications. U.S. Government Printing Office, Washington DC 1968.

[3] B. Kemp and J. Olivan, "European data format 'plus' (EDF+), an EDF alike standard format for the exchange of physiological data," *Clin Neurophysiol,* vol. 114, pp. 1755-61, Sep 2003.

[4] T. Penzel, M. Hirshkowitz, J. Harsh, R. D. Chervin, N. Butkov, M. Kryger, B. Malow, M. V. Vitiello, M. H. Silber, C. A. Kushida, and A. L. Chesson, Jr., "Digital analysis and technical specifications," *J Clin Sleep Med,* vol. 3, pp. 109-20, Mar 15 2007.

[5] C. Berthomier, X. Drouot, M. Herman-Stoica, P. Berthomier, J. Prado, D. Bokar-Thire, O. Benoit, J. Mattout, and M. P. d'Ortho, "Automatic analysis of single-channel sleep EEG: validation in healthy individuals," *Sleep,* vol. 30, pp. 1587-95, Nov 1 2007.

[6] E. Werth and A. A. Borbely, "Recording the sleep EEG with periorbital skin electrodes," *Electroencephalogr Clin Neurophysiol,* vol. 94, pp. 406-13, Jun 1995.

[7] J. Virkkala, J. Hasan, A. Värri, S. L. Himanen, and K. Müller, "Automatic sleep stage classification using two-channel electro-oculography," *J Neurosci Methods,* vol. 166, pp. 109-15, Oct 15 2007.

[8] J. Virkkala, R. Velin, S. L. Himanen, A. Varri, K. Muller, and J. Hasan, "Automatic sleep stage classification using two facial electrodes," *Conf Proc IEEE Eng Med Biol Soc,* vol. 2008, pp. 1643-6, 2008.

[9] J. Toppila, P. Lapinlampi, T. Noponen, A. Urrila, M. Särkelä, M. Paloheimo, H. Viertiö-Oja, T. Porkka-Heiskanen, P. Meriläinen, and T. Salmi, "P36.22 EEG entropy in assessment of the depth of natural sleep in healthy volunteers," in *Clinical Neurophysiology*, 2008, p. 186.

[10] A.-M. Lapinlampi and S.-L. Himanen, "Sleep Staging with Frontopolar EEG Derivation," *Sleep and Hypnosis,* vol. 6, pp. 48-53, 2004.

[11] I. Ehlert, H. Danker-Hopfe, L. Höller, P. v. Rickenbach, R. Baumgart-Schmitt, and W. M. Herrmann, "A Comparison between EEG-Recording and Scoring by QUISI Version 1.0 and Standard PSG with Visual Scoring. A One-Channel Ambulatory EEG Recording Device Using Neural Network Techniques for Automatic Sleep Stage Classification," *Somnologie,* vol. 2, pp. 104-116, 1998.

[12] Oxford BioSignals, "BioSomnia," 2009.

[13] T. Morgenthaler, C. Alessi, L. Friedman, J. Owens, V. Kapur, B. Boehlecke, T. Brown, A. Chesson, Jr., J. Coleman, T. Lee-Chiong, J. Pancer, and T. J. Swick, "Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: an update for 2007," *Sleep,* vol. 30, pp. 519-29, Apr 1 2007.

[14] S. J. Redmond and C. Heneghan, "Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea," *IEEE Trans Biomed Eng,* vol. 53, pp. 485-96, Mar 2006.

[15] S. J. Redmond, P. d. Chazel, C. O'Brien, S. Ryan, W. T. McNicholas, and C. Heneghan, "Sleep staging using cardiorespiratory signals," *Somnologie,* vol. 11, pp. 245-256, Mar 2007.

[16] W. Karlen, C. Mattiussi, and D. Floreano, "Sleep and Wake Classification With ECG and Respiratory Effort Signals," *IEEE Trans Biomed Circuits and Systems,* vol. 3, pp. 71-78, 2009.

[17] L. Hellstrom-Westas and I. Rosen, "Continuous brain-function monitoring: state of the art in clinical practice," *Semin Fetal Neonatal Med,* vol. 11, pp. 503-11, Dec 2006.

[18] J. Cohen, "A coefficient of agreement for nominal scales," *Educational Psyhol Meas,* vol. 20, pp. 37-46, 1960.

[19] J. C. Ives and J. K. Wigglesworth, "Sampling rate effects on surface EMG timing and amplitude measures," *Clin Biomech (Bristol, Avon),* vol. 18, pp. 543-52, Jul 2003.

[20] J. Virkkala, "Wearable sleep recorder with online analysis," in *International Workshop on Wearable, Micro and Nano Technologies for the Personalised Health, pHealth 2009* Oslo, Norway, 2009.