Amplitude Normalization Applied to an Artificial Neural Network-Based Automatic Sleep Spindle Detection System

Errikos M. Ventouras, Maria Panagi, Hara Tsekou, Thomas J. Paparrigopoulos, and Periklis Y.Ktonas, *Senior Member, IEEE*

Abstract— **Sleep spindles are significant rhythmic transients present in the sleep electroencephalogram (EEG) of non-rapid eye movement (NREM) sleep. Automatic sleep spindle detection techniques are sought for the automation of sleep staging and the detailed study of sleep spindle patterns, of possible physiological significance. A deficiency of many of the available automatic detection techniques is their reliance on the amplitude level of the recorded EEG voltage values. In the present work, an automatic sleep spindle detection system that has been previously proposed, using a Multi-Layer Perceptron (MLP) Artificial Neural Network (ANN), was evaluated using a voltage amplitude normalization procedure, with the aim of making the performance of the ANN independent of the absolute voltage level of the individual subjects' recordings. The application of the normalization procedure led to a reduction in the false positive rate (FPR) as well as in the sensitivity. When the ANN was trained on a combination of data from healthy subjects, the reduction of FPR was from 42.6% to 19%, while the sensitivity of the ANN was kept at acceptable levels, i.e., 73.4% for the normalized procedure vs 84.6% for the non-normalized procedure.**

I. Introduction

Electroencephalography (EEG) is a basic tool used for investigating brain function during sleep. Sleep is composed of two major phases, repeating themselves a number of times during a normal night sleep period, non-rapid eye movement (NREM) sleep and rapid-eye-movement (REM) sleep. NREM sleep is in turn divided into stages 1 to 4, with each stage presenting varied characteristics of EEG waveforms, muscle activity and eye movement patterns [1,2]. Sleep spindles are important transient EEG waveforms appearing during the stages of NREM sleep. They are bursts of rhythmic EEG activity, with frequencies ranging from 11 to 16 Hz, characterized by a progressively increasing, then gradually decreasing amplitude, and a duration usually in the range of 1-2 sec. The amplitude is mostly below 50 μV peak-to-peak in an adult, although exceptions to both the amplitude and duration range are frequent [3,4]. Sleep spindles are a hallmark of stage 2 sleep and provide a major criterion for sleep staging [5].

The brain processes that generate sleep spindles, as well as the functional significance of spindles are active topics of research [6,7]. Quantitative investigations of sleep spindles have been used in the study of various pathological conditions such as affective disorders [8] and schizophrenia [9], including also the relation of spindles with cognitive mechanisms in psychotic patients [10], as well as in the study of dementia (e.g., Alzheimer's disease - AD) [11-13], where it has been shown that sleep spindles are poorly formed, of lower amplitude, shorter duration and much less numerous than in normal aging [11, 12].

The need for automatic sleep spindle detection has been closely related to the interest in research on sleep spindles, since the visual scoring of whole-night sleep EEG recordings for detecting spindles is a laborious and time-consuming task, prone to a high degree of intra- and inter-rater variability [14]. The task of automatic sleep spindle detection has proven to be a demanding one in the field of pattern recognition. This is at least in part due to the presence of low-amplitude spindles, the superimposition of much stronger slow-wave activity like delta rhythm waveforms, and the large inter-subject variability of spindle characteristics [1,4]. Additional difficulties are present in the automatic detection task due to the lack of a quantitatively strict definition of spindles as well as of a reliable "gold" standard, apart from visual inspection, for benchmarking the performance of the proposed systems [4,14,15]. In the pursuit of the automatic sleep spindle detection task, an extended variety of techniques have been proposed [15], probably due, at least in part, to the problems exposed above. Various pattern recognition techniques have been applied including Artificial Neural Networks (ANN) [14,16,17], Matching Pursuit (MP) and wavelet techniques [18-21], frequency and amplitude analysis [22-25], fuzzy detection [15,23], Support-Vector Machine (SVM) classifiers [16,26], switching linear Gaussian state-space models [27] and Bayesian algorithms [28].

Among the various problems that inter-subject variability presents to automatic detection systems, amplitude variability is a major cause for concern, especially for systems using amplitude features or general waveform morphological characteristics related to amplitude for producing the "template" with which subsequent novel recordings will be processed [14,23]. In previous work of our group [14,29], band-pass filtered EEG was used as input to a feed-forward Multi-Layer Perceptron (MLP) ANN. By training the network on characteristic examples of EEG segments with and without sleep spindles, acceptable classification results were provided, bypassing the feature selection stage. In the present work, we attempt to alleviate the problem of sleep spindle amplitude influence on the detection process, using a normalization procedure both in

E. M. Ventouras and M.Panagi are with the Department of Biomedical Engineering, Technological Educational Institute of Athens, Athens, 12210, Greece.

H. Tsekou, T. J. Paparrigopoulos, and P. Y. Ktonas are with the Sleep Study Unit, 1st Psychiatric Clinic, University of Athens Medical School, Eginition Hospital, Athens, 11528, Greece.

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

the training phase and in the testing phase of the ANN system. The aim of the normalization procedure was to implement system training and testing based not on amplitude values which might vary considerably among subjects, but on normalized amplitude values, i.e., values whose range is from -1 to 1.

II. Materials and Methods

The subject sleep recordings (polysomnograms) in the present study were provided by the Sleep Study Unit of the 1st Psychiatric Clinic of the University of Athens Medical School, at the Eginition Hospital in Athens. The polysomnograms were recorded with the Micromed/BrainQuick system. The set of recordings used, previously described in [14, 29], consisted of a whole-night sleep recording from 4 healthy control subjects (3 males of 26 years age each, denoted S1,S2 and S3, and 1 female of 27 years age, denoted S4 in the following). The night sleep record of each subject was divided into three consecutive parts of equal duration (thirds of the night). In each part, the longest stage 2 sleep period was selected, provided that it lasted for at least 10 minutes. Then a time frame (TF), starting 5 min after the start of stage 2 and lasting 5 min and 5 sec, was selected for visual analysis. This resulted in 3 time frames for each subject, TF_{Sij} i=1,…,4, j=1,2,3. From those TFs, electrode Cz recordings from subject S3 and S4 (sampling rate 256 Hz) and electrode C3 recordings (sampling rate 512 Hz) from subjects S1 and S2 were analyzed for sleep spindles, based on visual detection criteria described in [14]. The recordings of S1, S2 and S3 were analyzed by one experienced polysomnographer, while the recordings of S4 were analyzed with a consensus-detection methodology described in detail in [14].

After visual sleep spindle detection, the sampled EEG signal was band-pass filtered using a 128-coefficient FIR filter with 3dB cutoff frequencies set at 10.5 and 16 Hz. Then, the maximum value of the absolute voltage values of each filtered TF, $\max|TF_{Sij}|$, was computed and it was used to normalize the sampled data. The MLP was trained using the methodology described in [14]. Four normalized training sets (NTS_k, k=1,…,4) were used, producing, respectively, four sets of biases and weights (NTW_k, k=1,…,4). NTS_1 consisted of 2 sleep spindles and of 2 EEG segments not containing spindles, from each of TF_{S11},TF_{S12}, TF_{S13}. The filtered EEG that was used in NTS_1 was divided by the maximum of values $\max|TF_{S1j}|$, j=1,2,3. The same procedure was applied for producing NTW_2 and NTW_3, based on filtered EEG from subjects S2 and S3, respectively. For producing NTW_4, NTS_4 consisted of the merging of NTS_1, NTS_2 and NTS_3. Additionally, four non-normalized training sets (TS_k, k=1,…,4) were used, producing, respectively, four non-normalized sets of biases and weights (TW_k, k=1,…,4). Therefore, the total amount of training inputs (EEG segments) were 12, when training was based on data from one subject (either S1, S2, or S3), and 36 when training was based on data from all 3 subjects. The procedure used for producing the non-normalized training sets and the corresponding sets of biases and weights was the same as for the normalized ones, except that no division of the filtered EEG segments took place.

Performance evaluation of the ANN was accomplished using the output value O(t) of the ANN, $0 \le O(t) \le 1$, t corresponding to the time samples of the visually scored EEG recordings. Spindles of S4 were detected in TF_{S4j} j=1,2,3, using TW_k and NTW_k, k=1,…,4. For each NTW_k, the band-pass filtered EEG segments TF_{S41},TF_{S42} and TF_{S43} were divided by $\max|TF_{S41}|$, $\max|TF_{S42}|$ and $\max|TF_{S43}|$, respectively, producing normalized performance evaluation sets PENk/1, PENk/2 and PENk/3, k=1,…,4, corresponding to each of the training sets. In future applications, the normalization dividing value for a given recording will be extracted from the data available from that recording and will be the maximum of the absolute values of the voltages of the recording. Non-normalized performance evaluation sets, PEk/1, PEk/2 and PEk/3, k=1,…,4, were also produced for the corresponding TW_k, k=1,…,4, where no normalization of the band-pass filtered EEG segments TF_{S41},TF_{S42} and TF_{S43} took place.

The O(t) curve was divided into parts that had value greater or lower than a threshold value V_T. The parts that had a value greater than V_T were denoted as "peaks". We used two criteria for checking whether a sleep spindle presence was indicated by the ANN output. According to the "soft" criterion (SC), the ANN provided a spindle indication (SI) when a peak existed in the O(t) curve. According to the "hard" criterion (HC), a spindle presence was indicated only when the peak duration was greater than P_D sec. The SIs were automatically computed by the system. SIs that corresponded to visually detected spindles were termed "hits", while the other SIs were termed false positives (fps). The sensitivity of the network was computed as the percent ratio of the hits to the number of visually detected spindles and the false positive rate (FPR) as the percent ratio of the fps to the number of SIs given by the ANN.

III. Results

The absolute amplitude values used in the normalization procedure were $\max(\max|TF_{S11}|, \max|TF_{S12}|, \max|TF_{S13}|)$= 29.3μV, $\max(\max|TF_{S21}|,\max|TF_{S22}|,\max|TF_{S23}|)$= 26.1μV, $\max(\max|TF_{S31}|,\max|TF_{S32}|,\max|TF_{S33}|)$= 24.1μV, $\max|TF_{S41}|$ = 56.7μV, $\max|TF_{S42}|$= 72μV, $\max|TF_{S43}|$= 31.7μV. The consensus-based visual detection process for the time frames, used for testing the performance of the ANN, indicated 48, 62 and 72 spindles for TF_{S41}, TF_{S42} and TF_{S43}, respectively, resulting in a total amount of 182 spindles detected in 915 sec of EEG recording.

The performance of the ANN was investigated for V_T values 0.5, 0.6, 0.7, 0.8 and 0.9 for the SC (soft criterion), and for P_D=0.3 sec for the HC (hard criterion). Table I shows the mean (across thirds of the night) sensitivity and mean FPR values, for the normalized case (mean of performance evaluation sets PEN4/1, PEN4/2, PEN4/3) and for the non-normalized case (mean of performance evaluation sets PE4/1, PE4/2, PE4/3), for the various threshold values V_T and criteria types used.

TABLE I MEAN SENSITIVITY (SE) AND MEAN FALSE POSITIVE RATE (FPR) VALUES, ACROSS THIRDS OF THE NIGHT, WHEN THE ANN WAS TRAINED BASED ON MERGED DATA FROM SUBJECTS S1, S2 AND S3.

		Soft criterion (SC)					Hard criterion (HC), P_D=0.3 sec				
	V_T	0.5	0.6	0.7	0.8	0.9	0.5	0.6	0.7	0.8	0.9
Normalization procedure applied	SE	78,4	77,8	77,3	75,7	68,8	75,2	74,0	73,3	68,5	65,4
	FPR	21,6	21,1	19,5	19,8	19,8	19,7	18,6	17,9	17,0	15,1
Normalization procedure not applied	SE	83.8	85.2	85.9	85.0	84.1	83.8	84.8	85.0	84.5	83.6
	FPR	47.3	46.0	44.9	43.3	42.1	43.9	42.2	40.8	38.8	37.0

TABLE II MEAN (STANDARD DEVIATION) OF SENSITIVITY (SE) AND FALSE POSITIVE RATE (FPR) , FOR EACH THIRD OF THE NIGHT AND ACROSS THIRDS OF THE NIGHT, WHEN THE ANN WAS TRAINED BASED ON MERGED DATA FROM SUBJECTS S1, S2 AND S3.

	Normalization procedure applied				Normalization procedure not applied			
	1st third	2nd third	3rd third	Across thirds	1st third	2nd third	3rd third	Across thirds
SE	80.4 (5.8)	53.0 (6.3)	86.9 (1.6)	73.4 (15.7)	84.1 (1.6)	82.4 (0.8)	87.2 (1.4)	84.6 (2.4)
FPR	28.1 (3)	3.8 (1.5)	25.1 (2.4)	19 (11.3)	55.1 (2)	43.6 (3.3)	29.1 (4.5)	42.6 (11.3)

Table II shows the mean and standard deviation of the sensitivity and FPR values, across threshold values V_T and criteria types, for the normalized performance evaluation sets PEN4/1,PEN4/2,PEN4/3 and the non-normalized sets PE4/1,PE4/2,PE4/3, as well as the respective averages and standard deviations across thirds of the night.

Table III shows the mean and standard deviation (across thirds of the night, threshold values V_T and criteria types) of the sensitivity and FPR values when the ANN was trained based on data from subjects S1, S2 and S3 separately, i.e., using NTW_k, k=1,…,3 for the normalized case, and TW_k, k=1,…,3 for the non-normalized case.

Using the sensitivity values, averaged across thirds of the night, threshold values and criteria types, when the ANN was trained based on data from subjects S1, S2 and S3, separately and on merged data, there was a statistically significant difference between sensitivity values for the normalized and the non-normalized procedure, as attested by repeated-measures ANOVA, with the procedure being the between-subjects factor, with two levels (df=1, F=9.42, p=0.22). There was no statistically significant differentiation when the between-subjects factor was the training set used. Comparable results were obtained for the false positive rate values, where repeated-measures ANOVA, with the procedure being the between-subjects factor, showed significant differentiation (df=1, F=44.196, p=0.001). No significant differentiation existed when the between-subjects factor was the training set used.

TABLE III MEAN (STANDARD DEVIATION) OF SENSITIVITY (SE) AND FALSE POSITIVE RATE (FPR), WHEN THE ANN WAS TRAINED BASED ON DATA FROM SUBJECTS S1,S2 AND S3, SEPARATELY.

Subjects		Normalization applied	No normalization applied
1	SE	78.3 (10.3)	83.6 (3)
	FPR	26.6 (13.3)	48 (11.8)
2	SE	65.7 (19.3)	85 (2.2)
	FPR	15 (8.5)	43.7 (11)
3	SE	63.7 (16.7)	84.5 (2.1)
	FPR	15.9 (9.7)	39.8 (11.9)

IV. DISCUSSION

As can be seen by inspecting Tables II and III, when the ANN was used for detecting sleep spindles without the application of the normalization procedure, the average (across thirds of the night, thresholds and criteria types) FPR was not less than 39.8%. The application of the normalization procedure, by dividing the filtered EEG data of each third of the night to be inputted to the ANN by the maximum of the absolute voltage values of that third, and using weights and biases produced by normalized training, resulted in a notable reduction of FPR, with FPR ranging between 26.6% and 15%, for the various training sets used. This could be due to the fact that the filtered EEG recordings of subject S4, whose data were used for testing the ANN, had higher voltage values than the respective values of the filtered EEG data of subjects S1, S2 and S3, whose data were used for training the ANN. Therefore, it could be expected that, when the ANN was trained on non-normalized data from S1, S2 or S3, the higher filtered EEG voltage values that the ANN would encounter when detecting sleep spindles in subject S4 would result in a high rate of false positives. Apparently, the normalization procedure tended to mitigate this.

The performance of the ANN concerning sensitivity values, when the normalization procedure was applied, seemed to deteriorate for the 2nd third of the night, for which the maximum of the absolute voltage values of the testing record was highest (72μV), compared to the respective maximum values for the 1st (56.7μV) and 3rd (31.7μV) third of the night. This can be seen by inspecting Table II. This trend was also present when the ANN was trained based on data from subjects S1, S2 and S3 separately (not shown in the paper). It could be hypothesized that, in the case of the 2nd third of the night in the testing data, the (comparatively high) maximum value used for the normalization did not

represent adequately an "average" value for the envelope of the spindle activity present in the testing data. Therefore, the normalization could have resulted in the ANN having to cope with excessively low-value normalized data, leading to misses.

The main motivation for investigating the application of the normalization procedure studied in this work was the observation that various automatic sleep spindle detection systems rely on the amplitude level of the sleep spindle activity. Taking into consideration the preliminary findings presented in this work, we can conclude that amplitude normalization might alleviate, at least in part, amplitude sensitivity in system performance. Nevertheless, the definition of the normalization dividing factor seems to be important, and further research on this matter is needed. In addition, the normalization procedure should be applied to a larger number of subjects and recordings, both for training and testing purposes, in order to fully assess its usefulness.

REFERENCES

[1] M. Kryger, T. Roth, W. Dement eds., *Principles and Practice of Sleep Medicine*, 4th ed. Philadelphia, PA: Elsevier, Saunders, 2005.

[2] E. Niedermeyer, F. Lopes Da Silva eds., *Electroencephalography, Basic Principles, Clinical Applications, and Related Fields,* 4th ed. Baltimore: Williams & Wilkins, 1999, pp. 17-29.

[3] W. R. Jankel, and E. Niedermayer, "Sleep spindles," *J. Clin. Neurophysiol.,* vol. 2, pp. 1-35, 1985.

[4] L. Gennaro, and M. Ferrara, "Sleep spindles: an overview," *Sleep Med. Rev.,* vol. 7, no. 5, pp. 423-440, 2003.

[5] A. Rechtschaffen, and A. Kales, eds., *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects.* Washington, DC: Public Health Service, U.S. Government Printing Office, 1968.

[6] S. M. Fogel and C. T. Smith, C.T., "The function of the sleep spindle: a physiological index of intelligence and a mechanism for sleep-dependent memory consolidation," *Neurosci. Biobehav. Rev.*, vol. 35, pp. 1154–1165, 2011.

[7] S. Astori, R. D. Wimmer, A. Lüthi, "Manipulating sleep spindles – expanding views on sleep, memory, and disease," Trends Neurosci., vol. 36, no. 12, pp. 738-748, 2013.

[8] D. T. Plante, M. R. Goldstein, E. C. Landsness, M. J. Peterson, B. A. Riedner, F. Ferrarelli, T. Wanger, J. J. Guokas, G. Tononi, R. M. Benca, "Topographic and sex-related differences in sleep spindles in major depressive disorder: A high-density EEG investigation," *J. Affect Disorders*, vol. 146, no. 1, pp. 120-125, 2013.

[9] E. J. Wamsley, A. K. Shinn, M. A. Tucker, K. E. Ono, S. K. McKinley, A. V. Ely, D. C. Goff, R. Stickgold, D. S. Manoach, "The effects of eszopiclone on sleep spindles and memory consolidation in schizophrenia: a randomized placebo-controlled trial," *Sleep,* vol. 36, no. 9, pp. 1369-1376, 2013.

[10] M. S. Keshavan, D. M. Montrose, J. M. Miewald, R. D. Jindal, "Sleep correlates of cognition in early course psychotic disorders," *Schizophr. Res.*, vol. 131, no. 1-3, pp. 231-234, 2011.

[11] D. Petit, J.-F. Gagnon, M. L. Fantini, L. Ferini-Strambi, J. Montplaisir, "Sleep and quantitative EEG in neurodegenerative disorders," *J. Psychosom. Res.*, vol. 56, pp. 487-496, 2004.

[12] P. Y. Ktonas, S. Golemati, P. Xanthopoulos, V. Sakkalis, M. D. Ortigueira, H. Tsekou, et al., "Time-frequency analysis methods to quantify the time-varying microstructure of sleep EEG spindles: Possibility for dementia biomarkers?," *J. Neurosci. Meth.*, vol. 185, pp. 133-142, 2009.

[13] N. T. Economou, A. Kyrozis, I. Kritikou, P. Ktonas, A. Bonakis, D. Dikaios, et al., "Inter-hemispheric spectral coherence reduction in sleep spindle frequency activity in patients with cognitive decline associated with amnesic mild cognitive impairment and Alzheimer's disease," *J. Sleep Res.*, vol. 19 (Suppl.1), pp.1050, 2010.

[14] E. Ventouras, E. Monoyiou, P. Ktonas, T. Paparrigopoulos, D. Dikeos, N. Uzunoglu, C. Soldatos, "Sleep Spindle Detection Using Artificial Neural Networks Trained with Filtered Time-Domain EEG: A Feasibility Study," *Comput. Meth. Prog. Bio.,* vol. 78, pp. 191-207, 2005.

[15] L. Causa, C. Held, J. Causa, P. Estévez, C. Perez, R. Chamorro, M. Garrido, C. Algarín, P. Peirano, "Automated sleep-spindle detection in healthy children polysomnograms," *IEEE Trans. Biomed. Eng.,* vol. 57, pp. 2135-2146, 2010.

[16] N. Acır, and C. Gózelis, "Automatic recognition of sleep spindles in EEG by using artificial neural networks," *Expert Syst. Appl.*, vol. 27, no. 3, pp. 451–458, 2004.

[17] Gunes, S., Dursun, M., Polat, K., Yosunkaya, S., "Sleep spindles recognition system based on time and frequency domain features," *Expert Systems with Applications*, 38, 2455–2461, 2011.

[18] J. Zygierewicz, K. J. Blinowska, P. J. Durka, W. Szelenberger, S. Niemcewicz, W. Androsiuk, "High resolution study of sleep spindles," *Clin. Neurophysiol.*, vol. 110, pp. 2136-2147, 1999.

[19] S. V. Schonwald, E. L. de Santa-Helena, R. Rossatto, M. L. F. Chaves, G. J. L. Gerhardt, "Benchmarking matching pursuit to find sleep spindles," *J. Neurosci. Meth.*, vol. 156, no. 1–2, pp. 314–321, 2006.

[20] F. Duman, A. Erdamar, O. Erogul, Z. Telatar, S. Yetkin, "Efficient sleep spindle detection algorithm with decision tree," *Expert Syst. Appl.*, vol. 36, pp. 9980-9985, 2009.

[21] L. Zhang, H. Li, and Y. Wei, "Sleep spindle detection using a novel instantaneous frequency definition," *Math. Meth. Appl. Sci.*, vol. 35, pp. 2101–2110, 2012.

[22] P. Schimicek, J. Zeitlhofer, P. Anderer, B. Saletu, "Automatic sleep spindle detection procedure: aspects of reliability and validity," *Clin. Electroencephal.*, vol. 25, no. 1, pp. 26-29, 1994.

[23] E. Huupponen, G. Gomez-Herrero, A. Saastamoinen, A. Varri, J. Hasan, S.-L. Himanen, "Development and comparison of four sleep spindle detection methods," *Artif. Intell. Med.,* vol. 40, pp. 157-170, 2007.

[24] R. Bódizs, J. Körmendi, P. Rigó, A. Sándor Lázár, "The individual adjustment method of sleep spindle analysis: methodological improvements and roots in the fingerprint paradigm," *J. Neurosci. Meth.*, vol. 178, no. 1, pp. 205-213, 2009.

[25] A. Nonclercq, C. Urbain, D. Verheulpen, C. Decaestecker, P. Van Bogaert, P. Peigneux, "Sleep spindle detection through amplitude–frequency normal modelling," *J. Neurosci. Meth.*, vol. 214, no. 2, pp. 192-203, 2013.

[26] I. Mporas, P. Korvesis, E. Zacharaki, V. Megalooikonomou, "Sleep Spindle Detection in EEG Signals Combining HMMs and SVMs," in *Engineering Applications of Neural Networks, Communications in Computer and Information Science*, vol. 384, L. Iliadis, H. Papadopoulos, S. Jayne Eds. Berlin: Springer, 2013, pp. 40-49.

[27] T. A. Camilleri, K. P. Camilleri, S. G. Fabri, "Automatic detection of spindles and K-complexes in sleep EEG using switching multiple models," *Biomed. Signal Proces.*, vol. 10, pp. 117-127, 2014.

[28] B. I. Babadi, S. M. McKinney, V. Tarokh, J. M. Ellenbogen, "DiBa: a data-driven Bayesian algorithm for sleep spindle detection", *IEEE Trans. Biomed. Eng*., vol. 59, no. 2, pp. 483-493, 2012.

[29] E. M. Ventouras, N.-T. Economou, I. Kritikou, H. Tsekou, T. J. Paparrigopoulos, P. Y. Ktonas, "Performance Evaluation of an Artificial Neural Network Automatic Spindle Detection System", in *Proc 34th Int Conf IEEE-EMBS,* San Diego, 2012, pp.4328-4331.