34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Validation of a Novel Automatic Sleep Spindle Detector with High Performance During Sleep in Middle Aged Subjects

Sabrina L. Wendt, Julie A. E. Christensen, Jacob Kempfner, Helle L. Leonthin, Poul Jennum and Helge B. D. Sorensen

Abstract— **Many of the automatic sleep spindle detectors currently used to analyze sleep EEG are either validated on young subjects or not validated thoroughly. The purpose of this study is to develop and validate a fast and reliable sleep spindle detector with high performance in middle aged subjects. An automatic sleep spindle detector using a bandpass filtering approach and a time varying threshold was developed. The validation was done on sleep epochs from EEG recordings with manually scored sleep spindles from 13 healthy subjects with a mean age of 57.9 $\pm$ 9.7 years. The sleep spindle detector reached a mean sensitivity of 84.6 % and a mean specificity of 95.3 %. The sleep spindle detector can be used to obtain measures of spindle count and density together with quantitative measures such as the mean spindle frequency, mean spindle amplitude, and mean spindle duration.**

I. INTRODUCTION

Sleep spindles (SS) are generated from complex interactions between thalamic, limbic and cortical areas. In sleep microstructure they are one of the most important elements and they are hallmarks of non-REM stage 2 sleep (N2) [1]. The American Academy of Sleep Medicine (AASM) defines SS as EEG phenomena with sinusoidal spindle-like waveforms lasting 0.5-3 seconds having a frequency profile at 11-16 Hz and most conspicuous in EEG recordings from central deflections [2]. This definition of SS will be used throughout this study. As the AASM standard does not differentiate between slow ($<$ 13 Hz) and fast ($>$ 13 Hz) oscillations, neither will this study.

SS are believed to mediate many sleep related functions and considered to represent inhibition of sensory stimuli to cortex thus maintaining sleep [3]. They are hypothesised to have at least two functions in relation to cognition; being a physiological measure of cognition and fast spindles are probably involved in memory consolidation during sleep [4], [5]. Also they possibly play a role in attentional processing [6].

Manual scoring of SS is very time-consuming; even for experts who only tend to agree in 70 $\pm$ 8 % of scores [7]. To standardize the scoring of SS, an automatic SS detector is required to interpret sleep EEG.

S.L. Wendt (e-mail: sabrina_wendt@hotmail.com), J.A.E. Christensen, J. Kempfner and H.B.D. Sorensen (e-mail: hbs@elektro.dtu.dk) are with the Institute of Electrical Engineering, Technical University of Denmark, Kongens Lyngby, Denmark

S.L. Wendt, J.A.E. Christensen, J. Kempfner, H.L. Leonthin and P. Jennum are with Danish Center for Sleep Medicine, Glostrup University Hospital, Glostrup, Denmark

P. Jennum is with Center for Healthy Aging, University of Copenhagen, Copenhagen, Denmark

Several SS detectors using a bandpass filtering approach have been developed and applied to analyze data relating SS and cognition [8], [9], [10], [11], [12], [13], [14].

This study describes an SS detector based on the definitions stated by the AASM standard supplemented with knowledge from sleep scoring experts at Glostrup University Hospital, Denmark. The detector is thoroughly trained and tested to gain insight into the performance of the new method for automatic SS detection in sleep EEG.

II. METHODS

A. Subjects

The group used to train, test and validate the SS detector consisted of 13 healthy subjects including five males and eight females with a mean age of 57.9 $\pm$ 9.7 years. The same group was used previously by Christensen et al. to train, test and validate an SS detector using Matching Pursuit and Support Vector Machine [15].

B. Polysomnographic recordings

Polysomnography (PSG) EEG data were used in this study. All subjects underwent one night of full PSG recorded outpatient. The PSG equipment was fitted at the clinic and the subjects removed the equipment themselves the following morning. The PSG recordings were performed in accordance with the AASM standard using EEG electrodes located at F3, F4, C3, C4, O1 and O2 with reference to the mastoids according to the 10-20 system [16]. At the fitting of the equipment the impedances of the electrodes were below 10 kΩ for all channels. EEG was sampled with a frequency of 256 Hz. Furthermore, tibialis anterior muscle tonus, nasal flow, thorax abdominal respiratory movements, EOG, ECG, submental muscle tonus and blood oxygen saturation were recorded.

SS in the 13 subjects were manually prescored by visual inspection and approved by an experienced PSG technician using the program Nervus (V5.5, Cephalon DK, Nørresundby, Denmark) as described in [15]. As randomly selected sleep epochs were chosen for fully SS scoring, the procedure ensured data from several independent subjects without having to resort to the very time consuming task of scoring SS in a full night recording. In total this resulted in 882 manually scored SS from 375 sleep epochs corresponding to approximately three hours recording. The majority of sleep epochs included, were manually classified as N2. As described in [2], SS can be found in epochs scored as other stages than N2. Therefore, the dataset included some epochs

978-1-4577-1787-1/12/$26.00 ©2012 IEEE

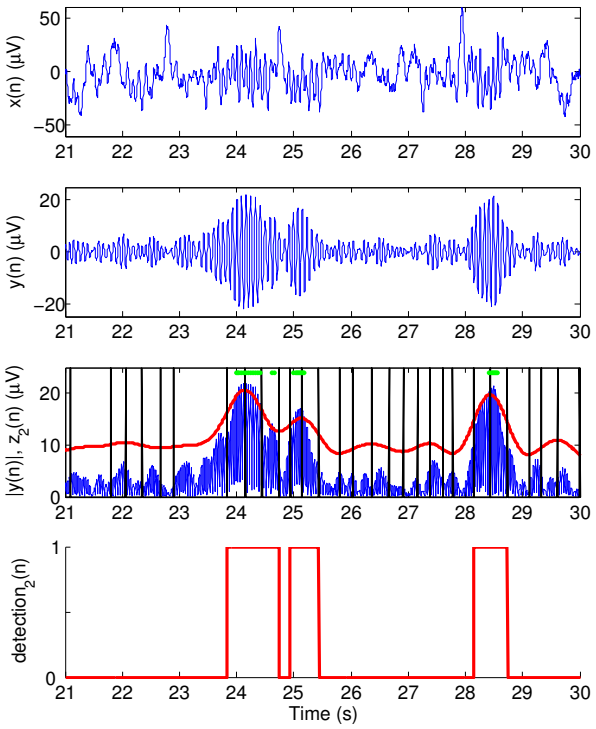

Fig. 1. Illustration of the important signal processing steps in the detection of SS. The top graph shows the raw EEG signal from C3-A2. The second graph shows the 11-16 Hz bandpass filtered version of the signal (1). The third graph shows the rectified signal in blue and the 1 Hz envelope with 8 μV offset in red (2). Vertical black lines indicate segment boundaries. A whole segment is classified as an SS candidate if the blue signal exceeds the red threshold at minimum one sample between two boundaries (marked with green dots), otherwise the segment is classified as background EEG. SS candidates become true SS if the segment is not covering an alpha intrusion or an artefact and has the correct duration. The bottom graph shows were SS are detected using the detector with the stated settings.

from N1, N3 and REM in order to make the detector work in these stages as well.

The raw sleep data, hypnograms and sleep events of the subjects were extracted from Nervus (V5.5, Cephalon DK, Nørresundby, Denmark) using the build-in export data tool. Further analysis of data was performed in MATLAB (R14 or R2008a, The MathWorks, Natick, MA., USA).

C. A novel sleep spindle detection algorithm

This study aims to develop an automatic SS detector during sleep. It is thus not intended for EEG from wake periods. If the detector should be fully automatic it would require a sleep-wake detector to initially locate sleep periods in the EEG. The SS detector will work in a standardized way classifying data as either SS or background EEG. The purpose of this is to obtain various quantitative measures describing the SS activity.

The algorithm relies on a bandpass filtering approach and uses decision fusion by combining the results from two individual detectors. Signals from one central deflection (C3-A2) and one occipital deflection (O1-A2) are used by the detector to detect SS in C3-A2. Both signals are bandpass filtered using an equiripple filter with passband between 11-16 Hz and stopbands at 10 Hz and 17 Hz with attenuation of 10^{-4} using a zero-phase digital filtering approach. The bandpass filtered signal can be expressed as:

$$y(n) = h_{11-16}(n) * x(n) \qquad n = 1, ..., N \qquad (1)$$

where $x(n)$ is the raw signal from C3-A2 and $h_{11-16}(n)$ is the impulse response of the 11-16 Hz bandpass filter. The bandpass filtered C3-A2, $y(n)$, is rectified by taken the absolute value of the signal. A time varying threshold is created by adding an offset to the envelope of the rectified signal. The envelope, $z_i(n)$, is calculated using an equiripple lowpass filter with stopband at the passband plus 1 Hz and a stopband attenuation of 10^{-4} using a zero-phase digital filtering approach. The first detector uses an envelope with passband at 2.25 Hz and an offset of 3 μV. The other uses an envelope with passband at 1 Hz and an offset of 8 μV. The i^{th} envelope can thus be described as:

$$z_i(n) = h_i(n) * |y(n)| \qquad i = 1, 2 \qquad (2)$$

where $|y(n)|$ is the rectified bandpass filtered raw signal and $h_i(n)$ is the impulse response of the i^{th} lowpass filter. The first and second derivative of the envelope is approximated using the central difference. The approximated first derivative of the i^{th} envelope is defined as:

$$z_i'(j) = \frac{z_i(j+1) - z_i(j-1)}{2} \; i = 1, 2 \; j = 2, ..., N-1 \quad (3)$$

where $z_i(n)$ is the i^{th} envelope. Points of zero crossings of the first and second derivative are located. A point at a zero crossing of the second derivative is removed if it originates from an almost stationary point of inflection on the envelope. Thus, the points of zero crossings reflect the local extrema of the envelope and its first derivative, and is used as boundaries for SS candidates. A SS candidate is detected in the whole interval between two boundaries when the rectified signal exceeds the time varying threshold at minimum one sample within these boundaries. A graphical illustration of the signal processing steps is presented in Fig. 1.

After identifying SS candidates some are discarded. If a candidate is more likely to be an alpha intrusion or if the amplitude of the original EEG is too high (>85 μV) according to the technicians at Glostrup University Hospital, it is discarded. An SS candidate is also discarded if it has a too short (<0.5 s) or too long (>3 s) duration according to the AASM [2].

A single detector uses one lowpass filter to find the envelope with a corresponding offset. The algorithm consists of a fusion of two detectors. It is assumed that two detectors working on different scales will compliment each other as they possess different qualities. If at least one detector has classified a sample as SS, then the sample is marked as SS in the final binary classification vector. If both detectors has classified a sample as background EEG, the sample is marked as background EEG in the final result.

The algorithm is presented in pseudocode in Algorithm 1.

Algorithm 1 SS algorithm using a fusion of 2 detectors

```
Bandpass filter signals from C3-A2 and O1-A2 in the 11-16 Hz band
for i = 1 to 2 {detector no.} do
  if i = 1 then
    f_passband ← 2.25 Hz and offset ← 3 μV
  else if i = 2 then
    f_passband ← 1 Hz and offset ← 8 μV
  end if
  Calculate envelope of rectified bandpass filtered C3-A2
  Find points of local extrema of the envelope and its first derivative (ignore almost stationary points of inflection on the envelope)
  if rectified filtered C3-A2 > envelope + offset then
    mark interval between surrounding points of extrema as SS candidate
  end if
  if SS frequency ≤ 13 Hz and power of bandpass filtered O1-A2 > power of bandpass filtered C3-A2 then
    remove SS candidate {alpha intrusion}
  end if
  if amplitude of rectified C3-A2 > 85 μV then
    remove SS candidate {artefact}
  end if
  if duration of SS < 0.5 s or duration of SS > 3 s then
    remove SS candidate {wrong duration}
  end if
  return detection_i(n)
end for
if Σ_{i=1}^{2} detection_i(n) ≥ 1 then
  result(n) ← 1
else
  result(n) ← 0
end if
```

D. Validation of the sleep spindle detector

Different statistical measures were calculated to validate the performance of the algorithm for SS detection: True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN). These numbers are calculated on sample basis. The obtained values were used to calculate the sensitivity, specificity, and the Matthews Correlation Coefficient (MCC) [17]. MCC was chosen because it is a balanced evaluation of the performance of a binary classifier and it is symmetric with respect to FP and FN [18]:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (4)$$

The SS algorithm was validated using the leave-one-subject-out method due to the small number of subjects. A subject with abnormal SS would greatly influence the performance if the 13 subjects were divided in separated training and test groups. Therefore, one subject was held out of the training set, the algorithm was optimized to fit the data from the remaining twelve subjects and the optimized model was then tested on the held out subject. This was done for all subjects.

The training phase included investigation of the performance of the detector with $11 \times 25 = 275$ combinations of lowpass filters and envelope offsets. The passband frequencies ranged from 0-2.5 Hz with a step size of 0.25 Hz and the envelope offsets ranged from 2-8 μV with a step size of 0.25 μV. The parameter setting yielding the highest MCC was chosen for the optimal SS detector. This SS detector, $\text{detection}_1(n)$, was then fused with all other detectors to find a fusion of detectors yielding even higher MCC. This process was repeated until the model used five detectors.

The SS detector in combination with $\text{detection}_1(n)$ yielding the highest MCC is denoted $\text{detection}_2(n)$ and so on. How to fuse the detectors depending on the number of independent detectors is explained in pseudocode in Algorithm 2; where an independent detector is denoted by k, j denotes how many independent detectors are fused, SS samples are given the value 1 and background EEG samples are given the value 0.

Algorithm 2 Fusion of SS detections

```
for j = 1 to 5 do
  if Σ_{k=1}^{j} detection_k(n) ≥ 1 then
    result_j(n) ← 1
  else
    result_j(n) ← 0
  end if
  return result_j(n)
end for
```

When testing the different models (using a fusion of 1-5 independent detectors) on the held out subjects, it became clear that little or no advantage was won by applying a fusion of more than two detectors. Therefore, during the training phase fusions of maximum five detectors were carried out.

III. RESULTS

The mean performance when testing the model using two detectors yielded a sensitivity of 84.6 % and a specificity of 95.3 %. The mean performance was affected by a single outlier with very low sensitivity (41.8 %) and thus the median performance of the algorithm should be mentioned. It yielded a sensitivity of 89.9 % and a specificity of 95.1 %. Further details of the performance are listed in table I.

TABLE I
PERFORMANCE OF THE SS ALGORITHM USING DECISION FUSION BY COMBINATION OF TWO SS DETECTORS

	Sensitivity	Specificity
min - max	41.8 - 96.6	91.6 - 98.9
mean ± std	84.6 ± 15.6	95.3 ± 2.2
25th percentile	83.1	94.1
50th percentile	89.9	95.1
75th percentile	93.7	97.3

On a standard PC with a 2.00 GHz processor using the presented algorithm for SS detection, it takes approximately five minutes to detect SS in a whole night of sleep ($\sim$8 hours of data).

IV. DISCUSSION

In this study a new SS detection algorithm based on the AASM definitions of SS is developed and validated. The algorithm includes a fusion of two detectors with complimentary qualities.

As the used data set only includes data from sleep stages, the algorithm for SS detection should not be applied on data from wake periods. Most of the validation data originate from N2, and thus the algorithm should be used with caution in other sleep stages than N2. The risk of false detections during wake is high due to artefacts from muscle tone and movement.

The automatic SS detector is validated on a data set with a high mean age. The detector in [15] achieved the approximate same mean sensitivity around 85 %. The current detector has a considerable higher mean specificity of 95 % compared to 85 % in [15]. One may argue that this result is due to over-fitting since the agreement between SS scorers is only 70 % [7]. Due to this fact, part of the detected FP could be true SS. Despite the low inter scorer agreement, visual inspection is still the golden standard when detecting SS.

Unfortunately, very few of the SS detectors used to analyze data relating SS and cognition are thoroughly validated. The papers describing the detectors do either not state the sensitivity or the specificity of the algorithm [9], [10], [11], or both [12], [13], [14]. It would be interesting to conduct a thorough validation of these algorithms on the same dataset, including the one in this study.

Compared to the presented SS detector, the SS detector presented by [8] is validated on a smaller number of subjects. Their performance has a lower mean specificity with a larger range (93.5 %, 88-100 %). On the contrary, their mean sensitivity is higher with a smaller range (96.2 %, 91-100 %). This reveals one of the biggest problems concerning automatic SS detectors; namely low specificities. Keeping in mind that if the specificity is only 90 %, 10 % of all background EEG could in fact be detected as SS.

V. CONCLUSIONS

This paper describes a simple, fast and thoroughly validated method for automatic SS detection in sleep EEG. The mean age of the control subjects is high, making this algorithm ideal to apply in middle aged subjects. The detector achieved a performance with a median sensitivity of $\sim$90 % and a median specificity of $\sim$95 %.

We believe that the here presented SS detector has achieved well enough performance in order to be used in a clinical investigation. In a future study this SS detector will be applied on sleep EEG from middle aged subjects.

REFERENCES

[1] L. De Gennaro and M. Ferrara, Sleep spindles: an overview, Sleep Med. Rev., vol. 7, no. 5, pp. 423440, Oct. 2003.

[2] C. Iber, S. Ancoli-Israel, A. Chesson, and S. F. Quan, The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications, American Academy of Sleep Medicine, 2007.

[3] T. T. Dang-Vu, S. M. McKinney, O. M. Buxton, J. M. Solet, J. M. Ellenbogen, Spontaneous brain rhythms predict sleep stability in the face of noise, Curr. Biol., vol. 20, no. 15, pp. 626627, Aug. 2010.

[4] S. M. Fogel and C. T. Smith, The function of the sleep spindle: A physiological index of intelligence and a mechanism for sleep-dependent memory consolidation, Neurosci. Biobehav. Rev., vol. 35, no. 5, pp. 11541165, Apr. 2011.

[5] J. Tamminen, J. D. Payne, R. Stickgold, E. J. Wamsley, M. G. Gaskell, Sleep spindle activity is associated with the integration of new memories and existing knowledge, J. Neurosci., vol. 30, no. 43, pp. 1435614360, Oct. 2010.

[6] G. Forest, J. Poulin, A.-M. Daoust, I. Lussier, E. Stip, and R. Godbout, Attention and non-REM sleep in neuroleptic-naive persons with schizophrenia and control participants, Psychiatry Res., vol. 149, no. 1-3, pp. 3340, Jan. 2007.

[7] J. Zygierewicz, K. J. Blinowska, P. J. Durka, W. Szelenberger, S. Niemcewicz, and W. Androsiuk, High resolution study of sleep spindles, Clin. Neurophysiol., vol. 110, no. 12, pp. 21362147, Dec. 1999.

[8] P. Schimicek, J. Zeitlhofer, P. Anderer, and B. Saletu, Automatic sleep-spindle detection procedure: aspects of reliability and validity, Clin. Electroencephalogr., vol. 25, no. 1, pp. 2629, Jan. 1994.

[9] S. Gais, M. Mölle, K. Helms, and J. Born, Learning-dependent increases in sleep spindle density, J. Neurosci., vol. 22, no. 15, pp. 68306834, Aug. 2002.

[10] R. Bódizs, T. Kis, A. S. Lázár, L. Havrán, P. Rigó, Z. Clemens, and P. Halász, Prediction of general mental ability based on neural oscillation measures of sleep, J. Sleep Res., vol. 14, no. 3, pp. 285292, Sep. 2005.

[11] P. Anderer, G. Gruber, S. Parapatics, M. Woertz, T. Miazhynskaia, G. Klösch, B. Saletu, J. Zeitlhofer, M. J. Barbanoj, H. Danker-Hopfe, S. L. Himanen, B. Kemp, T. Penzel, M. Grözinger, D. Kunz, P. Rappelsberger, A. Schlögl, and G. Dorffner, An E-health solution for automatic sleep classification according to Rechtschaffen and Kales: validation study of the somnolyzer 24x7 utilizing the Siesta database, Neuropsychobiology., vol. 51, no. 3, pp. 115133, 2005.

[12] F. Ferrarelli, R. Huber, M. J. Peterson, M. Massimini, M. Murphy, B. Riedner, A. Watson, P. Bria, and G. Tononi, Reduced sleep spindle activity in schizophrenia patients, Am. J. Psychiatry., vol. 164, no. 3, pp. 483492, Mar. 2007.

[13] E. J. Wamsley, M. A. Tucker, A. K. Shinn, K. E. Ono, S. K. McKinley, A. V. Ely, D. C. Goff, R. Stickgold, and D. S. Manoach, Reduced sleep spindles and spindle coherence in schizophrenia: mechanisms of impaired memory consolidation?, Biol. Psychiatry., vol. 71, no. 2, pp. 154161, Jan. 2012.

[14] T. O. Bergmann, M. Mölle, J. Diedrichs, J. Born, and H. R. Siebner, Sleep spindle-related reactivation of category-specific cortical regions after learning face-scene associations, Neuroimage., vol. 59, no. 3, pp. 27332742, Feb. 2012.

[15] J. A. E. Christensen, J. Kempfner, H. L. Leonthin, L. Arvastson, S. R. Christensen, P. Jennum, and H. B. D. Sorensen, Decreased sleep spindle density in patients with idiopathic REM sleep behaviour disorder and patients with Parkinson's disease, unpublished.

[16] H. H. Jasper, Report on the committee on methods of clinical examination in electroencephalography, Electroencephalogr. Clin. Neurophysiol., vol. 10, pp. 370375, 1958.

[17] B. W. Matthews, Comparison of the predicted and observed secondary structure of T4 phage lysozyme, Biochim. Biophys. Acta., vol. 405, no. 2, pp. 442451, Oct. 1975.

[18] P. Baldi, S. Brunak, Y. Chauvin, C. A. F. Andersen, H. Nielsen, Assessing the accuracy of prediction algorithms for classification: an overview, Bioinformatics., vol. 16, no. 5, pp. 412424, May 2000.