Development and comparison of four sleep spindle detection methods

Eero Huupponen a,*, Germán Gómez-Herrero b, Antti Saastamoinen a, Alpo Värri b, Joel Hasan a, Sari-Leena Himanen a

a Department of Clinical Neurophysiology, Medical Imaging Centre, Pirkanmaa Hospital District, P.O. Box 2000, Tampere, Finland
b Institute of Signal Processing, Tampere University of Technology, Korkeakoulunkatu 1, FIN-33101 Tampere, Finland

Received 15 September 2006; received in revised form 20 March 2007; accepted 11 April 2007

ARTIFICIAL INTELLIGENCE IN MEDICINE

http://www.intl.elsevierhealth.com/journals/aiim

KEYWORDS
Sleep EEG; Sleep spindles; Automated detection

Summary

Objective: The objective of the present work was to develop and compare methods for automatic detection of bilateral sleep spindles.

Methods and materials: All-night sleep electroencephalographic (EEG) recordings of 12 healthy subjects with a median age of 40 years were studied. The data contained 6043 visually scored bilateral spindles occurring in frontopolar or central brain location. In the present work a new sigma index for spindle detection was developed, based on the fast Fourier transform (FFT) spectrum, aiming at approximating our previous fuzzy spindle detector. The sigma index was complemented with spindle amplitude analysis, based on finite impulse response (FIR) filtering, to form of a combination detector of bilateral spindles. In this combination detector, the spindle amplitude distribution of each recording was estimated and used to tune two different amplitude thresholds. This combination detector was compared to bilaterally extracted sigma indexes and fuzzy detections, which aim to be independent of absolute spindle amplitudes. As a fourth method a fixed spindle amplitude detector was included.

Results: The combination detector provided the best overall performance; in S2 sleep a 70% true positive rate was reached with a specificity of 98.6%, and a false-positive rate of 32%. The bilateral sigma indexes provided the second best results, followed by fuzzy detector, while the fixed amplitude detector provided the poorest results so that in S2 sleep a 70% true positive rate was reached with a specificity of 97.7% and false-positive rate of 46%. The spindle amplitude distributions automatically determined for each recording by the combination detector were compared to

* Corresponding author. Tel.: +358 40 558 4283; fax: +358 3 3116 3018.
E-mail address: eero.huupponen@elisanet.fi (E. Huupponen).
1. Introduction

Sleep and sleep disorders are studied and treated in sleep medicine. Electroencephalographic (EEG), eye movement (EOG) and electromyographic (EMG) signals of the subject are recorded throughout the night. The visual sleep staging is performed with a 30-s time resolution, using six sleep stages: W (wake), REM, S1 (lightest sleep), S2, S3, S4 (deepest sleep), based on waveforms seen in all the signals, among them sleep spindles in the EEG [1]. Sleep spindles lasting from 0.5 to 2.0 s with frequencies ranging from 10.5 to 16 Hz are seen in sleep EEG occurring with the highest density in S2 sleep and with lower density in deeper sleep. Spindle frequency typically slows down along with deepening sleep and increases as the sleep gets lighter at the end of sleep cycles [2]. There are typically 200–1000 spindles during an all-night sleep recording. Spindles are considered sleep-maintaining events blocking the transfer of sensory information into the cerebral cortex at the level of thalamus [3–5]. Sleep spindles can be unilateral, occurring on only one brain hemisphere at a given time, especially at sleep onset [6], whereas bilateral spindles occurring simultaneously on both brain hemispheres in mirroring locations indicate a state of stabilized spindle generation.

Computer-based sleep analysis methods aim at providing an objective and reproducible description of sleep process and sleep EEG micro-events, like sleep spindles [7,8]. A further aim is to realize quantitative diagnosis of sleep EEG, providing detailed information of the state of the subject beyond visual sleep staging. For these purposes, automated analysis methods need to be developed, such as automated analysis of spindles on multiple EEG channels. In one work [9], spindles from 10 healthy subjects were studied using the 11.5–16 Hz band filtering based amplitude detector which was presented in [10] to analyse 18 EEG channels, applying an amplitude threshold of 12.5 μV. The central channels showed typically the largest number of spindles. In another work, sleep EEG with 21 EEG channels in 10 healthy subjects was studied [11]. Matching pursuit approach was applied in the detection, using a 15 μV amplitude threshold and a 0.5–2.0 s duration criterion for all EEG channels. Spindle amplitude distributions were presented, showing that largest amplitude spindles are located centrally and that they become somewhat reduced towards the temporal and frontal positions. Distributions of spindle amplitudes and frequencies were presented and thalamocortical origin of spindles was discussed.

The connection of verbal memory consolidation and spindle activity was examined in 19 subjects in one work [12]. The spindles were detected by filtering of 11–16 Hz band with a 0.5 s duration criteria. They report that a fixed amplitude threshold could not be used because of the high variability in spindle amplitudes between subjects. They adapted the detection threshold for each subject semi-automatically. First, the awake and EMG artefact segments were rejected, and then subject-specific threshold was obtained as 20% of the amplitude of the fifth largest spindle detected on a central EEG channel.

The present work has three objectives. First, to present a single sigma index to approximate our previous fuzzy detector. Second, to compare four methods in the detection of bilateral spindles: sigma index, fuzzy detector, a fixed amplitude detector and a combination detector utilizing both sigma index and spindle amplitude information. Third, to extract detailed characteristics of visually scored bilateral spindles, providing information facilitating the development of their automatic detection methods.

2. Methods and Materials

2.1. Recordings

All-night recordings from 12 healthy subjects (6 females and 6 males, with a median age of 40 years, ranging from 22 to 59 years) are included in the present analyses. None of the subjects used hypnotics or any other medication affecting the central nervous system. They had to be free from any sleep complaints or excessive daytime sleepiness. The subjects retired to bed between 10 and 12 p.m. and they were allowed to sleep maximally 8 h. Two consecutive nights were recorded with a 200 Hz sampling rate. EEG channels Fp1-A2, C3-A2, O1-A2, Fp2-A1, C4-A1, O2-A1, two EOG channels
and the submental muscle tonus were recorded. In addition tibialis anterior muscle tonus, body position, electrocardiogram, nasal airflow, thoracoabdominal respiratory movements and blood oxygen saturation were recorded.

The recordings are part of a larger study[13]. The recordings of the second night were used for the analysis and scored into sleep stages by the standard method[1]. The apnea-hypopnea index was calculated and it had to be less than 10 events/h. Spindles were visually selected by two scorers. The spindles had to present the typical waxing and waning pattern and a restricted duration (0.5—3 s). Inter-scorer agreement of spindles between the scorers was 81%. Bilateral spindles (occurring frontopolarly or/ and centrally) which were marked by both scorers were accepted into the study (Fig. 1). Scorings provide the start and end points of a total of 6043 bilateral spindles at 0.01 s time resolution.

2.2. Spindle detection methods

2.2.1. Sleep depth measure

A spectral mean frequency measure can be used to separate wakefulness from sleep [14]. Such a measure is used in all methods in the present work, aiming at identifying all those parts of the recordings containing sleep. Mean frequency is extracted for each second of the recording on the EEG channel C4-A1 with a frequency band of 0.5—12.5 Hz. It is determined based on a fast Fourier transform (FFT) amplitude spectrum \( S \) (5-s Saramäki windowing \( \beta = 3 \), centered on the second \( k \), zero padding to 2048 samples and scaled in the same way as presented below for sigma index of Method 1) as follows:

\[
c[f] = \frac{\sum_{f_l=0.5}^{f_{12.5}} S[f_l]}{\sum_{f_l=0.5}^{f_{12.5}} S[f_l]}
\]

Mean frequency \( f_{c,k} \) is then determined with linear interpolation as \( c[f_{mean}] = 0.5 \). Then, the measure is smoothed with 31-s median filtering to provide \( f_k \). As presented in our previous study on 15 healthy and 15 apnea subjects [15], such a sleep depth measure ranges roughly so that values exceeding 6 Hz indicate mostly wakefulness while smaller values indicate sleep.

2.2.2. EEG signal segmentation in spindle analysis

Four EEG channels are included in the present spindle analysis, the frontopolar channels Fp1-A2 and Fp2-A1 and the central channels C3-A2 and C4-A1, denoted shortly as Fp1, Fp2, C3, and C4 in the rest of the paper. When processing the all-night recordings, each EEG signal is segmented using three 67% overlapping 1.0 s long segments which are centered at sub-second time instances of 0.17, 0.5, 0.83 s at a full second \( k \), these segments are denoted as EEG\(_{k,i}\), \( i = 1, 2, 3 \) (Fig. 1). The time index \((k, i)\) is used in the rest of the paper to indicate the center point of the underlying 1-s EEG segment.

This segmentation scheme supports accurate subsequent combination of initial spindle detections across EEG channels, computed at respective time intervals \((k, i) \pm 0.5 \text{s}\). The present signal segmentation also allows minor timing differences

Figure 1   Bilateral sleep spindle, occurring most clearly in frontopolar brain positions (Fp1-A2, Fp2-A1), is seen in these EEG segments of 5.0 s. Four EEG channels are included in the present spindle analysis, Fp1-A2 on the left and Fp2-A1 on the right hemisphere frontopolarly and C3-A2, C4-A1, respectively, centrally. Visual spindle scoring is drawn with dashed vertical lines. The used 1-s EEG segmenting with 67% overlap is illustrated at a full second \( k \), the middle segment EEG\(_{k,2}\) is centered at \( k + 0.5 \text{s} \).
in spindle occurrences across EEG channels. Also, as the EEG segmentation is done the same way in all four methods (1–4), the same routine can be used to count the results.

2.3. Method 1, combination detector

2.3.1. Sigma index

A new simple feature for spindle detection is developed in the present work, called here the sigma index, to approximate our previous fuzzy detector [16], which is described below in Method 3. The EEG segment EEG_{k,i} of 1.0 s (200 samples) is made zero-mean and a 1.0 s long Hanning window, denoted as w_n, n = 1, ..., 200, is used to weight the signal samples, which are then zero-padded to a length of 512 and the FFT is taken of the entire sequence [17] to provide complex spectrum \( R(f) \). The frequency resolution of the spectrum \( R(f) \), denoted as \( \Delta f \), is then 0.39 Hz. The resulting FFT estimate is scaled to the corresponding amplitude spectrum as \( S(f) = 2 \times |R(f)|/\sum_{n=1}^{200} w_n \). Such a spectrum presents, e.g., a sinusoid with an (half of peak-to-peak) amplitude of 20 \( \mu \text{V} \) as a peak with a height of 20 \( \mu \text{V} \).

First, \( \text{max} = \max(S(f_{\text{spin}})) \) is extracted, where \( f_{\text{spin}} \) denotes the spindle frequency band, selected as 10.5–16.0 Hz in the present work. The quantity \( \text{max} \) reflects the average spindle amplitude during the signal segment EEG_{k,i}. Because 1.0 s long windowing is used, too short spindle activities, like 0.2 s, are averaged down. The mean amplitude level in the theta and low alpha frequency band is taken as \( \text{mean}_{\text{low}} = \max(S(f_{\text{low}})) \), where \( f_{\text{low}} \) is 4–10 Hz. Respectively, \( \text{mean}_{\text{high}} = \max(S(f_{\text{high}})) \), where \( f_{\text{high}} \) is 20–40 Hz. The sigma index, denoted as \( r_{k,i} \), is then obtained as

\[
 r_{k,i} = \frac{\text{max}}{(\text{mean}_{\text{low}} + \text{mean}_{\text{high}})/2}.
\]

A high sigma index value indicates a high probability of a spindle and vice versa, indicating how dominant the spindle peak is as compared to other EEG components. The sigma index is further complemented with alpha activity rejection so that if the maximum alpha peak in the 7.5–10 Hz band of spectrum \( S \) exceeds the spindle amplitude max, no spindle detection is allowed and \( r_{k,i} \) is set to zero. The \( \text{mean}_{\text{high}} \) helps against EMG type of interference, making the sigma index value smaller along increasing \( \text{mean}_{\text{high}} \). Based on the initial testing with six recordings of the work [16] where spindles were scored on one channel only, the sigma index values larger than 4.5 typically indicate spindles. The present work gives quantified evaluation of this feature.

One sigma index value per one signal segment EEG_{k,i} is obtained. As four EEG channels are analysed in parallel in the present work, four sigma index values are obtained simultaneously, denoted as \( r_{k,i}(C3) \), \( r_{k,i}(C4) \), \( r_{k,i}(Fp1) \), and \( r_{k,i}(Fp2) \).

2.3.2. Spindle amplitude

Also time-domain spindle amplitudes of the respective EEG segments are determined, using finite impulse response (FIR) filtering. A 401 tap FIR filter with a pass band of 10.5–16.0 Hz with 3 dB cut-off frequencies set at 9.9 and 16.6 Hz is used to band-pass filter the raw EEG signal. The spindle amplitude, or more generally sigma amplitude, is determined based on the filtered EEG as the peak amplitude (thus to positive or negative direction) of each 1.0 s signal segment centered at the time instance \( (k, i) \). Spindle amplitudes on the four parallel EEG channels are denoted as \( A_{k,i}(C3) \), \( A_{k,i}(C4) \), \( A_{k,i}(Fp1) \), and \( A_{k,i}(Fp2) \).

2.3.3. Detection of bilateral spindle segments

This framework for bilateral spindle detection was presented initially using a 1-s time resolution for the detector output [18]. In the present work, the output is obtained three times each second, in concordance to the underlying EEG segments centered at \( (k, i) \). Also, the new sigma index is used to replace the fuzzy detector. Method 1 uses two passes through each recording, the Preliminary run is done merely to find the appropriate amplitude thresholds that are used during the Analysis run.

2.3.4. Preliminary run

In spindle detection based on spindle amplitude, optimal amplitude threshold for an individual all-night recording can be estimated based on automatically determined spindle band amplitude distributions during spindles and non-spindle seconds [16]. In a later work, we presented a way to determine useful amplitude thresholds based on spindle amplitude distributions only [18] determined using the fuzzy detector. The new sigma index is used to achieve this in the present work.

In the Preliminary run of Method 1 through an all-night recording, the amplitudes of spindle segments detected by the sigma index (detection threshold \( \lambda_r \) set to 4.5) are collected from all EEG channels, as follows. Running \( k \) from 1 to the end of the all-night recording and at each value of \( k \), if one or more \( r_{k,i}(Fp1) > \lambda_r \) and \( f_k < 6 \), where \( j = 1, 2, 3 \), then maximum of respective amplitudes \( A_{k,j}(Fp1) \), where \( j = 1 \) or 1, 2, or 1, 3 or 1, 2, 3 or 2, 3 or 2 or 3, 3 or 2 or 3, is appended in the pool of collected amplitude values. Similarly is done for other channels also, Fp2, C3 and C4, adding all the time to the same pool. After that, possible amplitude values larger than three times
the median are removed from the pool. The mean value of the pool of amplitudes then provides a fairly large amplitude threshold, denoted as $l_{A,\text{fair}}$ for a particular all-night recording. Based on our previous experience [16], the magnitude of $l_{A,\text{fair}}$ is in any case smaller than the mean of the amplitudes of visually scored spindles of that recording would be. This is because not quite all true spindles are detected here and there are naturally also false-positive findings of borderline spindles, which are partly somewhat smaller in amplitude than visually scored spindles. Another amplitude threshold for Method 1, denoted as $l_{A,\text{low}}$, is obtained as $0.75 \times l_{A,\text{fair}}$ to estimate the low limit of true spindle amplitudes in that recording. The two thresholds $l_{A,\text{fair}}$ and $l_{A,\text{low}}$ are then used in the subsequent Analysis run of Method 1 through the same recording (in practice, just going through the computed values again, held in the RAM memory). To see what the pools of spindle amplitudes were like, please see Figs. 2 and 3.

### 2.3.5. Analysis run

A bilateral central spindle segment is detected if there are spindle detections on the central EEG channels on both brain hemispheres (C3, C4) at the same time. At least one of the two detections has to be based on the sigma index which includes the rough spindle duration analysis. Spindle amplitude $A_{k,i}$ is also used and thus there are three possible branches leading to a detection at a time instant $(k, i)$. The first branch is by sigma index based detection on both central EEG channels $(r_{k,i} > \lambda_{r})$ provided that also the simultaneous spindle amplitudes on both central EEG channels exceed the low amplitude threshold $(A_{k,i} > l_{A,\text{low}})$. In the second and third branches a sigma index based detection on one of the two channels is required to be accompanied by a spindle amplitude

![Figure 2](image1.png)

**Figure 2** Spindle amplitude distributions. Spindle amplitudes of visually scored bilateral spindles (a). Each curve represents one subject. There is a degree of inter-subject variability in spindle amplitudes. Spindle amplitudes automatically estimated in the Preliminary run of Method 1 (b), where intentionally a large number but at the same time partly somewhat smaller spindle amplitude values are collected. These are quite suitable for the determination of amplitude thresholds for each all-night recording separately.

![Figure 3](image2.png)

**Figure 3** Comparison of automatically determined amplitude thresholds of Method 1 (Table 2) to amplitudes of visually scored spindles. The low limit threshold $l_{A,\text{low}}$ and the "safe" amplitude threshold $l_{A,\text{fair}}$ of each subject are plotted with solid dotted lines, while the 3% and 22% percentiles of distributions of visually scored spindle amplitudes (seen in Fig. 2(a)) of each corresponding subject are plotted with dashed dotted lines, to see how they compare.
exceeding the larger amplitude threshold \( (A_{k,i} > \lambda_{A,\text{fair}}) \) on the mirroring central EEG channel, indicating visually noticeable spindle amplitude there. These three branches are designed to tolerate the inherent variability in bilateral spindle appearance.

The detection algorithm can thus be formulated as follows. Set \( C_{k,i} = 0 \) for all time instances \((k, i)\). A bilateral central spindle second is detected \((C_{k,i} = 1)\) if the following condition is met:

\[
C_{k,i} = 1 \quad \text{if} \quad \left\{ f_k < 6 \right\} \quad \text{and} \quad \left\{ \{ r_{k,i}(C3) > \lambda_r \right. \\
\left. \quad \text{and} \quad r_{k,i}(C4) > \lambda_r \text{ and } A_{k,i}(C3) > \lambda_{A,\text{low}} \right. \\
\left. \quad \text{and} \quad A_{k,i}(C4) > \lambda_{A,\text{low}} \quad \text{or} \quad \{ r_{k,i}(C3) > \lambda_r \right. \\
\left. \quad \text{and} \quad r_{k,i}(C4) > \lambda_{A,\text{low}} \} \quad \text{or} \quad \{ A_{k,i}(C3) > \lambda_{A,\text{fair}} \right. \\
\left. \quad \text{and} \quad A_{k,i}(C4) > \lambda_{A,\text{fair}} \left. \} \right. \\
\left. \text{and} \quad r_{k,i}(C4) > \lambda_r \right. \\n\right\}.
\]

A bilateral frontopolar (Fp) spindle segment is detected \((Fp_{k,i} = 1)\) in an identical way based on the detections on the two frontopolar channels \((Fp1, Fp2)\). Set \( Fp_{k,i} = 0 \) for all time indexes and

\[
Fp_{k,i} = 1 \quad \text{if} \quad \left\{ f_k < 6 \right\} \quad \text{and} \quad \left\{ \{ r_{k,i}(Fp1) > \lambda_r \right. \\
\left. \quad \text{and} \quad r_{k,i}(Fp2) > \lambda_r \text{ and } A_{k,i}(Fp1) > \lambda_{A,\text{low}} \right. \\
\left. \quad \text{and} \quad A_{k,i}(Fp2) > \lambda_{A,\text{low}} \} \quad \text{or} \quad \{ r_{k,i}(Fp1) > \lambda_r \right. \\
\left. \quad \text{and} \quad r_{k,i}(Fp2) > \lambda_{A,\text{fair}} \} \quad \text{or} \quad \{ A_{k,i}(Fp1) > \lambda_{A,\text{fair}} \right. \\
\left. \quad \text{and} \quad r_{k,i}(Fp2) > \lambda_r \right. \\n\right\}.
\]

The amplitude thresholds \( \lambda_{A,\text{low}} \) and \( \lambda_{A,\text{fair}} \) are recording-specific \((\text{Table 2})\) while \( \lambda_r \) is same for all recordings. We can see that by selecting the detection thresholds suitably we can reduce the detection Method 1 to only a sigma index based method as follows. Please note that no Preliminary run is done in Method 2.

### 2.4. Method 2, bilateral sigma indexes

Sigma index based detection on both brain hemispheres is obtained when selecting \( \lambda_{A,\text{low}} = 1 \) \((\text{which is always exceeded})\) and \( \lambda_{A,\text{fair}} = 10,000 \) \((\text{never exceeded})\), leaving only:

\[
C_{k,i} = 1 \quad \text{if} \quad \left\{ f_k < 6 \right\} \quad \text{and} \quad \{ r_{k,i}(C3) > \lambda_r \right. \\
\left. \text{and} \quad r_{k,i}(C4) > \lambda_r \} ;
\]

\[
Fp_{k,i} = 1 \quad \text{if} \quad \left\{ f_k < 6 \right\} \quad \text{and} \quad \{ r_{k,i}(Fp1) > \lambda_r \right. \\
\left. \text{and} \quad r_{k,i}(Fp2) > \lambda_r \} .
\]

Here \( \lambda_r \) is used as a normal threshold, same value for all recordings, as in Method 1.

### 2.5. Method 3, bilateral fuzzy detectors

The fuzzy detector was originally designed to provide a spindle detection output once per second, yet already focusing the detection on the time location of the spindle as well as possible \((16)\). In 2001 we modified the alpha rejection part to stand as reported here, in Step 4:

- **Step 1.** Ten successive spectra with a length of 0.5 s with 80% overlap are extracted, covering a 1.4 s stretch of EEG centered originally at mid-second, but here at the time instant \((k, i)\) of the present EEG segmentation. Each amplitude spectrum is calculated with a Saramäki window function \((\beta = 3)\) and zero padding to 512 samples, scaled the same way as in the sigma index of Method 1. In the present work only, the first three and last three amplitude spectra are set to zero to provide a direct comparison with Methods 1 and 2. This is because the 10 successive spectra make use of 1.4 s of EEG and Step 3 may make this even wider.

- **Step 2.** The spectrum showing the largest peak sigma \((10.5—16 \text{ Hz})\) activity is selected, let us index it as spectrum \( i \). The sigma amplitude of spectrum \( i \) is then estimated as the mean value of spectral samples surrounding the sigma peak in \( \pm 1.0 \text{ Hz} \) and assigned to \( A_1 \).

- **Step 3.** Four spectra surrounding the time location of spectrum \( i \) at \(-0.4, -0.2, 0.2\) and \(0.4 \text{ s}\) are then studied. Let us index them respectively as \(-2, -1, 1\) and \(2\). To consider the duration of the spindle, the idea is to track the maximum spindle amplitude in the direction in which it is diminishing least. To this end, one of the spectra \(-1\) and \(1\) showing the maximum sigma amplitude is selected and, using the method in Step 2, the sigma amplitude is determined and assigned to \( A_2 \). If the spectrum selected was \(-1\), then a further one of spectra \(-2\) and \(1\) showing the maximum sigma amplitude is selected and that sigma amplitude is assigned to \( A_3 \) and if the spectrum selected was \(1\) then a further one of spectra \(-1\) and \(2\) showing the maximum sigma amplitude is selected and that sigma amplitude is assigned to \( A_3 \) and if the spectrum selected was \(1\) then a further one of spectra \(-1\) and \(2\) showing the maximum sigma amplitude is selected and that sigma amplitude is assigned to \( A_3 \).

- **Step 4.** The final sigma amplitude \( A_s \) is obtained as the mean value of \( A_1, A_2 \) and \( A_3 \). The amplitude of a very short spindle, showing mostly in \( A_1 \), is thus attenuated. If peak alpha \((7.5—10 \text{ Hz})\) amplitude exceeds the sigma amplitude \( A_1 \), then \( A_s \) is set to zero. Three feature values are then obtained \((\text{if there would be division by zero, features are set to zero})\) as \( x_1 = A_s / A_{\text{th}}, x_2 = A_s / A_{\text{th}}, x_3 = A_s / A_{\text{th}}, \)

where \( A_{\text{th}}, A_{\text{th}}, \) and \( A_{\text{th}} \) are the mean theta \((4—7 \text{ Hz})\), alpha \((7.5—10)\), and gamma \((20—40 \text{ Hz})\) amplitudes of spectrum \( i \), respectively.

Membership functions are used to describe the feature values in fuzzy terms \((19)\). The input value ranges are \(0.5—3.0, 0.5—2.0\) and \(5.0—13.0 \text{ for input} \).
features 1, 2 and 3, respectively. All input ranges include three partly overlapping triangular membership functions, so that function "small" values 1 at the low end and 0 in the middle of the range, "medium" values 1 in the middle and 0 at both ends of the range and "large" values 0 in the high end and 1 at the high end of the range. The output ranges 0—10, with four triangular membership functions are selected so that function "small" values 1 at the low end and 0 at 3, "medium" values 0 at 3.5, 1 at 5.5 and 0 at 7.5, "large" values 0 at 5, 1 at 7 and 0 at 9, "very large" valued 0 at 7, 1 at 9 and 0.5 at 10.

The fuzzy rule set connects the membership functions of the inputs (features) and the output. The output value is designed to have a large value when indicating a spindle. The fuzzy rule set consists of 17 rules:

1. If (input 1 is large) and (input 2 is large) and (input 3 is large) then (output is very large);
2. If (input 1 is large) and (input 2 is large) and (input 3 is medium) then (output is very large);
3. If (input 1 is large) and (input 2 is medium) and (input 3 is large) then (output is very large);
4. If (input 1 is medium) and (input 2 is large) and (input 3 is large) then (output is very large);
5. If (input 1 is large) and (input 2 is large) and (input 3 is medium) then (output is very large);
6. If (input 1 is medium) and (input 2 is medium) and (input 3 is large) then (output is very large);
7. If (input 1 is medium) and (input 2 is large) and (input 3 is medium) then (output is very large);
8. If (input 1 is medium) and (input 2 is medium) and (input 3 is large) then (output is very large);
9. If (input 1 is medium) and (input 2 is large) then (output is very large);
10. If (input 1 is medium) and (input 2 is large) and (input 3 is large) then (output is very large);
11. If (input 1 is large) and (input 2 is large) then (output is very large);
12. If (input 1 is large) and (input 2 is medium) then (output is very large);
13. If (input 1 is medium) and (input 2 is medium) then (output is very large);
14. If (input 1 is medium) and (input 2 is medium) then (output is very large);
15. If (input 1 is medium) and (input 2 is medium) then (output is medium);
16. If (input 1 is large) then (output is small);
17. If (input 1 is medium) then (output is small).

The output defuzzification is done using the centroid method providing detection output \( F_{k,i} \) when analysing EEG segment centred at time instant \((k, l)\). With the four EEG channels, a total of four output values \( F_{k,i}(C3) \), \( F_{k,i}(C4) \), \( F_{k,i}(Fp1) \), \( F_{k,i}(Fp2) \) are obtained at time instant \((k, l)\).

### 2.5.1. Detection of bilateral spindle segments

The detection of bilateral spindle segments is then obtained as

\[
C_{k,i} = 1 \quad \text{if} \quad \{f_k < 6\} \quad \text{and} \quad \{F_{k,i}(C3) > \lambda_F \quad \text{and} \quad F_{k,i}(C4) > \lambda_F\};
\]

\[
F_{k,i}(Fp1) > \lambda_F \quad \text{and} \quad F_{k,i}(Fp2) > \lambda_F\}
\]

where \( \lambda_F \) is a normal threshold, the same for all recordings.

### 2.6. Method 4, bilateral fixed amplitude detectors

The present amplitude detector is a modification of the spindle detector in [10]. Three 401 tap long FIR filters are used to filter the raw EEG signal. The spindle band activity is obtained with a filter (identical to that in Method 1) with 10.5—16.0 Hz pass band with (3 dB cut-offs at 9.9 and 16.6 Hz), and respectively, alpha activity is obtained with 7.5—10 Hz pass band with cut-offs at 6.9 and 10.6 Hz and EMG content with 30—40 Hz pass band with cut-offs at 29.4 and 40.6 Hz.

The spindle amplitude of this detector, denoted as \( S_{k,i} \), of each 1-s long signal segment, centered at \((k, l)\), is determined based on the absolute filtered spindle band activity fulfilling the 0.5 s duration criteria. At each time instant, the value of \( S_{k,i} \) is first set to zero and then it is increased iteratively. At each iteration step, the first and last sample (and also at least one point in the middle third part) in the 1.0 s long signal segment exceeding \( S_{k,i} \) are found. The spindle duration is obtained as the duration between the first and last sample. The iteration is continued as long as duration exceeds 0.5 s.

Following the original idea [10], the root-mean-square values of the three filtered signals are also examined. \( S_{k,i} \) is set to zero if root-mean-square of alpha activity exceeds the root-mean-square of spindle activity \( \times \eta \) or the root-mean-square of EMG content exceeds 5 \( \mu \)V. With the four EEG channels included, a total of four amplitude values \( S_{k,i}(C3) \), \( S_{k,i}(C4) \), \( S_{k,i}(Fp1) \), \( S_{k,i}(Fp2) \) are obtained at each time instant \((k, l)\). Different values of \( \eta \) are tested to explore the capability of this detector.

### 2.6.1. Detection of bilateral spindle segments

The detection of bilateral spindle segments is then obtained as

\[
C_{k,i} = 1 \quad \text{if} \quad \{f_k < 6\} \quad \text{and} \quad \{S_{k,i}(C3) > \lambda_S \quad \text{and} \quad S_{k,i}(C4) > \lambda_S\};
\]

\[
S_{k,i}(Fp1) > \lambda_S \quad \text{and} \quad S_{k,i}(Fp2) > \lambda_S\}
\]
where $\lambda_S$ is a normal amplitude threshold, the same for all recordings.

2.7. Comparison of the detection methods to visual bilateral spindle scorings

Detections of bilateral spindle segments provided by all methods are compared to the visual spindle scorings in the same way. First, central and frontopolar bilateral spindle detections need to be combined to overall bilateral spindle detections, denoted as $B_{k,i} = 1$ if ($C_{k,i} = 1$ or $F_{p,k,i} = 1$). This is needed to enable the evaluation against the visual spindle scorings that are based on all four EEG channels.

In determining the receiver operating characteristics (ROC) curves, different threshold values are applied in the analysis of the 12 all-night recordings. In Methods 1 and 2 threshold $\lambda_r$ is varied while in Method 3 threshold $\lambda_r$ is varied and in Method 4 threshold $\lambda_S$ is varied. With each threshold value used, true and false-positive rates are calculated taking into account the fact that all the methods make use of the underlying EEG signal 0.5 s beyond the middle point ($k$, $i$), as well as the timing between method outputs obtained at 0.33 s steps and visual scorings at 0.01 s precision, creating a maximum of 0.17 s timing difference. A true positive spindle is counted if a bilateral spindle is detected (one or more $B_{k,i} = 1$) simultaneously with a visual spindle scoring or up to ±0.17 s beyond it. Whenever the detector output indicates a spindle ($B_{k,i} = 1$) but there is no visual scoring at that time instant ($k$, $i$) or up to ±0.5 s beyond it, a false-positive is counted after which during 1.0 s no additional false-positives are counted. A true negative second is counted if the all three detector output values ($B_{k,i} = 1$), 1, 2, 3, are zero and there is no visual scoring at those time instants or up to ±0.5 s beyond them.

The performance is examined in two types of ROC curves. In both types, the true-positive rate (sensitivity) is the number of true positive findings divided by the total number of scored spindles. In type 1, the specificity is the number of true negative findings divided by true negatives and false-positives together. In type 2, the false-positive rate is the number of false-positive findings divided by the number of true and false-positive findings together. It tells the proportion of false alarms of all detections.

### 3. Results

3.1. Comparison of the spindle detection performance

The sleep parameters of the subjects are shown in Table 1. The ROC curves of the four methods are presented in Figs. 4 and 5. Method 1 provided the best performance overall, followed by Methods 2 and 3. Method 4 provided the poorest and very variable performance depending on the $\eta$ value, in no case did the performance quite reach the other methods in any selection. With $\eta$ set to one, many true spindles were missed and only 50% sensitivity was reached, because of strict alpha rejection condition. With $\eta$ set to two, a higher sensitivity was obtained but with somewhat more false-positives also. An additional test was done with Method 4 ($\eta = 1.5$), making a preliminary run (like in Method 1) collecting the amplitudes $S_{k,i}$ to find $\lambda_{A, \text{fair}}$ threshold values for each recording. The method performance improved only marginally, for instance, providing a 53.3% sensitivity with a 36.1% false-positive rate in S2 sleep.

In all four methods, the results were better during S2 than SWS. In all cases, the same value of $\lambda_r$ provided a better performance in Method 1 than in Method 2, which shows the effect of amplitude thresholds ($\lambda_{A, \text{low}}, \lambda_{A, \text{fair}}$ Table 2) used in Method 1. However, if there were no overall improvement, this difference would not be important, as the value of $\lambda_r$ can be selected freely in both methods. Combining the outcome of ROC curves of type 1 and 2 in S2 for Method 1, sensitivities of 86.5% and 73.5% with 96.9% and 98.5% specificities and at the same time 49% and 34.7% false-positive rates were obtained with $\lambda_r = 4.5$ and 5.5, respectively. For method 2 in S2, sensitivities of 75.2% and 51.2% with 97.7% and 99.3% specificities and at the same time 45.0% and 26.4% false-positive rates were obtained with $\lambda_r = 4.5$ and 5.5, respectively.

### 3.2. Bilateral spindle properties

Mean spindle duration was 0.84 s and median duration was 0.80 s in the visually scored bilateral spindles in all 12 recordings (Fig. 6).

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Sleep parameters of the 12 subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median</td>
</tr>
<tr>
<td>TIB</td>
<td>8 h 0 min</td>
</tr>
<tr>
<td>SEI%</td>
<td>86.8</td>
</tr>
<tr>
<td>SREM%</td>
<td>18.7</td>
</tr>
<tr>
<td>S1%</td>
<td>9.5</td>
</tr>
<tr>
<td>S2%</td>
<td>51.4</td>
</tr>
<tr>
<td>S3%</td>
<td>10.4</td>
</tr>
<tr>
<td>S4%</td>
<td>5.5</td>
</tr>
<tr>
<td>SWS%</td>
<td>16.3</td>
</tr>
</tbody>
</table>

Median and range of time in bed (TIB), sleep efficiency index (SEI), SEI% = 100 × (TST/TIB), percentages of sleep stages SREM, S1, S2, S3, S4, and SWS (S3 + S4), referred to total sleep time (TST).
Amplitudes of visually scored bilateral spindles of the 12 subjects are visualised in Fig. 2(a). These were obtained so that in each spindle, first, central or frontopolar position was selected based on largest average of sigma amplitudes $A$ across brain hemispheres in each position. Then, amplitude values $A$ of both hemispheres were collected from that position.

Comparison of automatically determined amplitude thresholds ($\lambda_{A,\text{low}}$, $\lambda_{A,\text{fair}}$) to amplitudes of visually scored spindles is presented in Fig. 3. The thresholds $\lambda_{A,\text{low}}$ proved to estimate relatively well the low limit of amplitudes of visually scored spindles, as intended, while thresholds $\lambda_{A,\text{fair}}$ estimated roughly the 22% percentiles of the amplitudes of visually scored spindles. There is a good subject-specific correspondence between these quantities.

Spindle amplitude differences across brain hemispheres are depicted in Fig. 7. These were obtained

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Spindle amplitude thresholds of Method 1 determined automatically for each subject during the Preliminary run</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject no.</td>
<td>$\lambda_{A,\text{low}}$ ($\mu\text{V}$)</td>
</tr>
<tr>
<td>1</td>
<td>8.51</td>
</tr>
<tr>
<td>2</td>
<td>8.53</td>
</tr>
<tr>
<td>3</td>
<td>8.73</td>
</tr>
<tr>
<td>4</td>
<td>10.06</td>
</tr>
<tr>
<td>5</td>
<td>10.08</td>
</tr>
<tr>
<td>6</td>
<td>10.13</td>
</tr>
<tr>
<td>7</td>
<td>10.74</td>
</tr>
<tr>
<td>8</td>
<td>10.84</td>
</tr>
<tr>
<td>9</td>
<td>11.45</td>
</tr>
<tr>
<td>10</td>
<td>11.52</td>
</tr>
<tr>
<td>11</td>
<td>12.46</td>
</tr>
<tr>
<td>12</td>
<td>13.72</td>
</tr>
</tbody>
</table>

Sorted in increasing order, low limit threshold $\lambda_{A,\text{low}}$, followed by the “safe” amplitude threshold $\lambda_{A,\text{fair}}$. Amplitudes of visually scored bilateral spindles of the 12 subjects are visualised in Fig. 2(a). These were obtained so that in each spindle, first, central or frontopolar position was selected based on largest average of sigma amplitudes $A$ across brain hemispheres in each position. Then, amplitude values $A$ of both hemispheres were collected from that position.

Comparison of automatically determined amplitude thresholds ($\lambda_{A,\text{low}}$, $\lambda_{A,\text{fair}}$) to amplitudes of visually scored spindles is presented in Fig. 3. The thresholds $\lambda_{A,\text{low}}$ proved to estimate relatively well the low limit of amplitudes of visually scored spindles, as intended, while thresholds $\lambda_{A,\text{fair}}$ estimated roughly the 22% percentiles of the amplitudes of visually scored spindles. There is a good subject-specific correspondence between these quantities.
so that in each spindle, first, central or frontopolar position was selected based on largest average of sigma amplitudes $A$ across brain hemispheres. Then, the amplitude difference from left to right was collected from that position. The outcome shows that there is considerable variability in the amplitudes of bilateral spindles across hemispheres, typically $-5, \ldots, 5 \mu V$.

Figure 5  ROC curves of type 2 provided by Methods 1–4 in the detection of bilateral spindles. Label ”1” denotes combination detector, ”2” bilateral sigma indexes, ”3” bilateral fuzzy detectors, ”4” bilateral fixed amplitude detectors, with $\eta = 1, 1.5$ and $2$. (a) Total durations of the 12 all-night recordings are included. (b) S2 sleep stages are included. (c) SWS (S3 + S4) sections of the recordings are included. The performance obtained with threshold $\lambda_r = 4.5$ and $\lambda_r = 5.5$ of Methods 1 and 2 are denoted with an asterisk and a circle, respectively.

Figure 6  Distribution of the durations of visually scored bilateral spindles.

Figure 7  Distribution of differences in bilateral spindle amplitudes across brain hemispheres in visually scored bilateral spindles.
3.3. Sigma index properties

Maximum sigma index values during all visually scored bilateral spindles we collected, so that in each scoring either central or frontopolar sigma indexes were collected, those with largest mean amplitude. In non-spindle seconds we used maximum of feature values in 1-s steps, either central or frontopolar. The two distributions of sigma index values are quite separate, with only 11% overlap, showing usefulness of such a sigma index (Fig. 8).

To examine the dependency of sigma index values and spindle amplitudes, all the pairs of detected spindles and corresponding spindle amplitudes were collected in an extra test in the Preliminary run of Method 1 from all 12 subjects. There was minor correlation between these two quantities with a correlation coefficient of 0.25. Sigma index is, therefore, not totally dependent but nearly independent of absolute spindle amplitude.

4. Discussion

In the present work we developed and compared methods for the detection of bilateral spindles occurring centrally or frontopolarly on both brain hemispheres at the same time. The developed new sigma index worked well in spindle detection, as a part of the combination detector (Method 1) and as the core of Method 2. The combination detector was the best method overall. It combined the spindle index and absolute spindle amplitude information across the mirroring EEG channels which seems to offer a more flexible detection of bilateral spindles than achieved with Methods 2–4.

Sigma index value conveys information about how the EEG signal segment looks in visual inspection. This is because a large sigma index value tells that spindle activity is clearly standing out from the surrounding EEG activity. The sigma index proved to separate well spindles and non-spindle seconds (Fig. 8). Moreover, sigma index is nearly independent of absolute spindle amplitudes, which is an advantage because of inter-subject variability. Interestingly, sigma index can be used to estimate the amplitude distribution of spindles efficiently and to provide subject-specific amplitude detection thresholds, that were quite useful in Method 1 (Figs. 2 and 3). Also other features close to our sigma index have been presented, the closest one being spindle power as related to total power [20]. Very likely such a feature can be used for the same purposes. It might also be possible to determine the spindle amplitude thresholds separately at each symmetric electrode pair with the presented routine, provided there is a reasonable amount of spindles to be found in each position. On the other hand, this would produce a large collection of thresholds, especially, e.g., with 128 EEG channels, which might be cumbersome in reporting the outcome of spindle analysis.

In the present detection of bilateral spindles, the sigma index provided a slightly better performance than the fuzzy detector, both of them exceeding the fixed amplitude detector. The spindle duration analysis via 1.0 s FFT of this sigma index seems to be quite suitable and perhaps the longer spectrum works better in estimating the spectral background EEG than 0.5 s FFT of the fuzzy detector. We used there Hanning windowing in the sigma index, as it is a commonly used window. Saramäki windowing could also have been used.

The false-positive detections of all methods seemed to be very spindle-like waveforms. One explaining factor for some false detections might be that visual spindle scoring adapts with great ease to find small or large spindles in the all-night recordings. This requires a lot from computer analysis, so that detection sensitivity should be very high when the spindles are small but on the other hand not to provide an excessive number of false-positives over the night. Presented Methods 1–3 aim at flexible detection of spindles, irrespective of the amplitude.

Concerning amplitude-based spindle detection systems the higher the spindle amplitude variability is between the subjects the more important it is to use an adaptive spindle amplitude threshold. Regarding Method 4, it was a bit disappointing that
the additional test with adapted thresholds could not really improve the performance. One reason for this might be that the differences in the amplitude distributions of our data set were subtle. Furthermore, the three subjects with the smallest spindles had also rather few visually scored bilateral spindles. In that way their contribution to the results remain minor. In future studies this aspect could be examined so that each visually scored spindle is weighted in result calculus by a factor based on the inverse of the total number of visually scored spindles in the recording.

Sleep depth information was included in all four methods in the same way. Such a context measure in spindle detection seems reasonable, especially if a recording contains a lot of wake periods. An additional test was done ignoring sleep depth information and virtually no difference in the ROC curves was observed, however. This is due to two things: There were rather few wake periods in these recordings and the methods provided most false-positive findings during sleep.

The new spindle detection method is compared to a previous method in no report of the reference list, which shows how surprisingly rare such method comparisons are. In the present work, in order to allow a comparison of the four detection methods all methods had to be tuned to have a common spindle band, in this work it was 10.5—16 Hz. This is a wider band than the classical 12—14 Hz given in [1] and in those used in other work 11.5—16 Hz [10] and 11—16 [12]. The outcome of the present work shows (Fig. 8) that such a wide band was necessary and even a bit wider could have been used, like 10—16 Hz used in [11,16]. A wide spindle band may be necessary especially for sleep disorder patients, who may show spindles even well below 10 Hz, based on our experience [21]. Future work should address the aspect of how the slow spindles could be detected given the well-known 1/f characteristic of sleep EEG.

Comparing method performance between different reports has to be done with a degree of caution as there are subjects of variable age ranges included and variable ways to count the results, full details of which are not always given. Ten subjects of age 20—35 years were included in the work [10], where spindles were detected based on filtering of EEG band of 11.5—16 Hz, complemented with 0.5 s duration and 12.5 μV amplitude criteria. One hour of sleep of every subject was used for detector evaluation, providing 89.7% sensitivity with 6.5% false-positive rate. They discuss that method settings, like frequency limits, can be tuned. Adaptive autoregressive modeling was employed in feature extraction in [22] and subsequent neural networks in two stages, tested on six recordings of subjects with median age of 30 years. A relatively small number, a total of 264, visually scored spindles was used for method evaluation, resulting in a 94% sensitivity and 4% false-positive rate. They discuss the importance of methods to adapt to particular recordings. Nineteen subjects were studied in [12] of 26—54 years, filtering 11—16 Hz and 0.5 s duration and semi-automatically determined amplitude thresholds, reporting 81.3% true positive and 21.4% false-positive rates.

In the work [23] one subject of 25 years was included, a neural network approach was applied, resulting in 80% agreement with visual scoring. Spindle scorings were interesting in that work, separate scorings for well-defined and poorly defined spindles. Such scorings would be interesting for a larger data set and comparison of different methods.

Matching pursuit (MP) approach was used in [24] with nine young subjects, average age of 24.6 years. 11—15 Hz band for spindles, for method evaluation 8.9% of the total data was selected and noisy segments were removed, resulting in 81.2% sensitivity and 81.2% specificity. In [11], 10 subjects of average age of 39 years were analysed, also with MP approach. An agreement of 90% with visual scoring was reported with an amplitude threshold of 25 μV. A 15 μV threshold produced 70% agreement, which they state to correspond to the consent between sleep experts. Putting it all together, the results obtained in the present work are quite comparable to previous work.

Some of the subjects were middle-aged in the present work. Therefore, true spindle amplitudes were partly very small (Fig. 2(a)) and their automated detection is challenging. In one work [25] spindles were detected using a 5 μV amplitude threshold, 0.5 s duration combined with a subsequent visual inspection to confirm or reject the detections. They report a clear age effect on spindles, mean spindle amplitudes of 12.8 and 22.7 μV in the older and in younger subjects, respectively. Their spindle identification method is reasonable in terms of amplitude, so that also small spindles can be found. Obviously such a procedure is suitable only for research purposes. Fully automated methods are needed in the analysis of large amounts of all-night recordings.

Knowledge about the clinical significance of sleep spindles is continuously growing. The spindle density and the power in spindle band were found reduced in sleep disordered breathing patients as compared to healthy controls, as reported in [26]. Spindle density was found to be higher after learning as compared to non-learning task [27]. Spindle
amplitude was found to enhance as a result of sleep deprivation [28]. In apnea patients the spindle frequency may stay slower throughout the night as compared to healthy subjects [21]. Analysis of specifically bilateral spindles may provide relevant results on sleep process as they occur in a state of stable spindle generation [6]. Also the inter-hemispheric differences of bilateral spindles can be examined [29]. Automated spindle detection and analysis methods may offer useful possibilities to quantify the detailed spindle characteristics in the clinical data sets thus helping diagnosis and follow-up of treatment.

5. Conclusion

A new sigma index, which is nearly independent of absolute EEG amplitudes, was developed in the present work. The sigma index works slightly better than our previous more complex fuzzy spindle detector in the detection of spindles and can be used to replace it. There is a relatively large inter-subject variation in spindle amplitudes, which is a relatively well-known aspect. This seems to be accompanied with an inter-hemispheric spindle amplitude variation in bilateral spindles. Flexible automated spindle detection methods are therefore needed to study spindle characteristics in sleep EEG recordings.

Acknowledgements

This study was financially supported by the National Technology Agency of Finland project Microsleep ("Analysis method of studying the microstructure of sleep", 2003–2006), the Research fund of the Tampere University Hospital, EU project Biopattern, as well as Jenny and Antti Wihuri foundation.

References


