Journal of Neuroscience Methods 214 (2013) 192–203

Contents lists available at SciVerse ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Computational Neuroscience

Sleep spindle detection through amplitude–frequency normal modelling

Antoine Nonclercq[a,*], Charline Urbain[b], Denis Verheulpen[c], Christine Decaestecker[a], Patrick Van Bogaert[c], Philippe Peigneux[b]

[a] Laboratory of Image, Signal and Telecommunications Devices (LIST), CP165/51, Université Libre de Bruxelles, Avenue F. Roosevelt 50, 1050 Bruxelles, Belgium
[b] UR2NF – Neuropsychology and Functional Neuroimaging Research Unit at CRCN – Centre de Recherches en Cognition et Neurosciences and UNI – ULB Neurosciences Institute, CP191, Université Libre de Bruxelles, Avenue F. Roosevelt 50, B1050 Bruxelles, Belgium
[c] Department of Pediatric Neurology, Hôpital Erasme, Université Libre de Bruxelles, Route de Lennik 808, 1070 Bruxelles, Belgium

HIGHLIGHTS

- An automated model-based spindle detection algorithm is proposed.
- It models the amplitude–frequency spindle distribution with a bivariate normal distribution.
- It automatically adapts to each individual subject and derivation.
- It was tested in seven healthy children and six adult patients suffering from different pathologies, and performs similarly or better than sleep experts.
- Normal modelling enhances spindle detection quality compared to fixed amplitude and frequency thresholds.

ARTICLE INFO

Article history:
Received 27 August 2012
Received in revised form 17 January 2013
Accepted 18 January 2013

Keywords:
Spindle detection
Normal modelling
Self-organized learning
Intersubject variation
Intrasubject variation
Children sleep

ABSTRACT

Manual scoring of sleep spindles can be very time-consuming, and achieving accurate manual scoring on a long-term recording requires high and sustained levels of vigilance, which makes it a highly demanding task with the associated risk of decreased diagnosis accuracy. Although automatic spindle detection would be attractive, most available algorithms are sensitive to variations in spindle amplitude and frequency that occur between both subjects and derivations, reducing their effectiveness.

We propose here an algorithm that models the amplitude–frequency spindle distribution with a bivariate normal distribution (one normal model per derivation). Subsequently, spindles are detected when their amplitude–frequency characteristics are included within a given tolerance interval of the corresponding model. As a consequence, spindle detection is not directly based on amplitude and frequency thresholds, but instead on a spindle distribution model that is automatically adapted to each individual subject and derivation.

The algorithm was first assessed against the scoring of one sleep scoring expert on EEG samples from seven healthy children. Afterward, a second study compared performance of two additional experts versus the algorithm on a dataset of six EEG samples from adult patients suffering from different pathologies, to submit the method to more challenging and clinically realistic conditions. Smaller and shorter spindles were more difficult to evaluate, as false positives and false negatives showed lower amplitude and smaller length than true positives. In both studies, normal modelling enhanced performance compared to fixed amplitude and frequency thresholds. Normal modelling is therefore attractive, as it enhances spindle detection quality.

1. Introduction

Sleep spindles' detection is of major importance for staging sleep as well as in the field of sleep research. A spindle is commonly defined as a group of rhythmic waves characterized by progressively increasing, then gradually decreasing amplitude, that may be present in low voltage background EEG, superimposed to delta activity, or temporally locked to a vertex sharp wave and to a K complex (De Gennaro and Ferrara, 2003). Spindles are one of the hallmarks of Non-Rapid Eye Movement (NREM) stage 2 sleep, both in adults and children (Iber et al., 2007; Grigg-Damberger et al., 2007). They are affected by normal ageing, by pathological ageing (De Gennaro and Ferrara, 2003; Petit et al., 2004), as well as by brain pathology; therefore, the spindles could be used as a marker of normal brain functional development (Fogel and

* Corresponding author. Tel.: +32 2 650 30 86; fax: +32 2 650 47 13.
E-mail addresses: anoncler@ulb.ac.be, antoine.nonclercq@gmail.com (A. Nonclercq).

0165-0270/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jneumeth.2013.01.015

Smith, 2011; Petit et al., 2004). Furthermore, they have been proposed as ideal candidates to induce, in the neocortex, long-term synaptic changes which are necessary for learning and memory (Destexhe and Sejnowski, 2001; Peigneux and Smith, 2010), even after a daytime nap (Schmidt et al., 2006). Accordingly, recent evidence in adults suggests that spindles are highly correlated with intellectual ability (IQ tests) and with an overnight improvement in performance after learning new material (Fogel and Smith, 2011).

Sleep spindles rapidly develop during the first three months of age. Sizeable maturational changes can be observed throughout the developmental phase in terms of frequency, amplitude, and amount. These changes reputedly reflect the development of thalamocortical structures and the maturation of the physiological system that produces spindles (De Gennaro and Ferrara, 2003), indicating that spindles are longitudinal markers for the ontogenic evolution of brain functioning. Conversely, modifications in sleep and spindle variables are observed with age (De Gennaro and Ferrara, 2003). Sleep spindles may thus represent, at the electrophysiological level, an ideal mechanism that reflects long-term synaptic changes in the neocortex (Fogel and Smith, 2011). Therefore, modulations and alterations in sleep spindle activity should be explored in childhood developmental disorders with associated cognitive impairments and in brain pathologies (Reeves and Klass, 1998; Shibagaki and Kiyono, 1983), that could also be associated with impaired sleep-dependent consolidation processes (Urbain et al., 2011; Van Bogaert et al., 2012; Chan et al., 2011).

Manual scoring of spindles is time-consuming for recordings that typically show 1000 spindles (Acir and Güzeliş, 2004). Achieving accurate manual scoring on long-term recordings requires a high level of vigilance, resulting in a highly demanding task that augments the risk of decreased accuracy in the diagnosis, especially for sleep-related studies, for which precise information (such as spindle's amplitude, frequency, and length) is often required. Beside the laborious aspect of the task, visual analysis involves some subjectivity (inter-human agreement is estimated to be around 80–90% (Campbell et al., 1980), and degree of consent is $70 \pm 8\%$ (Zygierewicz et al., 1999)). A reliable spindle detection is therefore attractive, as it would enhance the speed, accuracy, and inter-rater agreement of spindle scoring.

Various spindle detection algorithms have been previously proposed. Recent ones are based on methods that include fuzzy logic (Huupponen et al., 2000a, 2003), neural network (Shimada et al., 2000; Huupponen et al., 2000b; Acir and Güzeliş, 2004; Ventouras et al., 2005; Güneş et al., 2011), bandpass filter (Clemens et al., 2005; Huupponen et al., 2007), fast time frequency transform (Knoblauch et al., 2003a, 2003b), Fourier transform (Huupponen et al., 2007; Duman et al., 2009), wavelet transform (Duman et al., 2009), Gabor transform (Schönwald et al., 2003) and matching pursuit (Durka et al., 2005; Schönwald et al., 2006; Ktonas et al., 2009).

Fewer studies have investigated spindle detection in children (Grigg-Damberger et al., 2007; Causa et al., 2010). However, assessing automated spindles detection, specifically in childhood, seems important as spindles show modification in amplitude, frequency, length, density, interspindle interval and topological distribution from infancy to adolescence (Nagata et al., 1996; Shinomiya et al., 1999; Scholle et al., 2007); therefore, automated algorithms should be able to adapt to those variations. Published spindle detection algorithms in children and in infants are based on empirical-mode decomposition and Hilbert–Huang transform (Causa et al., 2010), expert procedure and fuzzy logic (Held et al., 2004, 2006), peak identification (Estévez et al., 2002), merge neural gas model (Estévez et al., 2007), and neuro fuzzy approach (Heiss et al., 2002).

The majority of the proposed algorithms are – directly or indirectly – based on amplitude–frequency analysis, thus banking on spindle definition and mimicking visual analysis. One of the major difficulties encountered with these detection methods is the setting of suitable thresholds for the amplitude and the frequency. Spindle frequency is traditionally defined as 12–14 Hz (Rechtschaffen and Kales, 1968; Grigg-Damberger et al., 2007), but may often extend to both higher and lower frequencies. Therefore, the 'classical' 12–14 Hz spindle definition is believed to be too narrow (Jankel and Niedermeyer, 1985). The difficulty in finding the optimum frequency bounds has produced a large number of proposed values, among them: 11.5–15 Hz (Fish et al., 1988), 11.5–16 Hz (Zeitlhofer et al., 1997), 11–15 Hz (Ktonas et al., 2009), 11–16 Hz (Clemens et al., 2005), 10.5–16 Hz (Ventouras et al., 2005; Huupponen et al., 2007), and 10–16 Hz (Zygierewicz et al., 1999; Huupponen et al., 2000a; Estévez et al., 2002). Beside the often cited 12–14 Hz frequency range proposed by the National Institute of Neurological Diseases and Blindness of the U.S. Department of Health, Education, and Welfare (Rechtschaffen and Kales, 1968), various organizations have suggested other values to score spindles: 11–16 Hz by the American Academy of Sleep Medicine (Iber et al., 2007), 11–15 Hz by the International Federation of Clinical Neurophysiology (Noachtar et al., 1999), and 12–16 Hz by the Japanese Society of Sleep Research (Hori et al., 2001). Finally, the Paediatric Task Force published a visual scoring of sleep and arousal in infants and children (Grigg-Damberger et al., 2007), but did not propose a definite spindle frequency range, underlying the variety of criteria found in the literature. Similarly, the spindle's minimum amplitude is difficult to determine, independently of the chosen definition of the amplitude (Fish et al., 1988). Previous researchers have set arbitrary lower amplitude thresholds at 8, 10, 12, and 14 μV (Fish et al., 1988), 10 μV (Estévez et al., 2002; Ventouras et al., 2005), 15 μV (Zygierewicz et al., 1999), and 25 μV (Zeitlhofer et al., 1997).

Universal amplitude and frequency intervals are difficult to define for spindles because of the large variability across subjects and derivations. Indeed, there is a large variability in both amplitude (Huupponen et al., 2000a; Clemens et al., 2005) and frequency (Zeitlhofer et al., 1997) across subjects and in both amplitude and frequency across derivations (Zeitlhofer et al., 1997). Therefore, it seems difficult to design a spindle detection algorithm without adapting to this variability.

Consequently, we propose to consider that both amplitude and frequency ranges vary between subjects and derivations, and to tailor thresholds accordingly. We designed an algorithm that models the amplitude–frequency spindle distribution with a bivariate normal distribution (one normal model per derivation). Subsequently, spindles are detected when their amplitude–frequency characteristics are included within a given tolerance interval (TI) of the corresponding model. As a consequence, spindle detection is not directly based on amplitude and frequency thresholds, but instead on a spindle distribution model that is automatically adapted to each individual and each derivation. Furthermore, the TI may be adjusted for more sensitive or selective detection.

Taking this idea one step further, spindle detection may be considered *probabilistic in nature*. In this regard, a given spindle has a probability of detection, which may be high or low depending on its amplitude, frequency, morphology, background activity, and other attributes. An identical approach has previously been proposed for EEG spikes (Wilson et al., 1996). For instance, an event with a shape similar to a spindle, but with a low amplitude (compared to its background) or with a particularly high/low frequency, could be accepted as a spindle by some specialists and not by others, whereas a clear spindle would be accepted by all specialists. Our approach could also be seen as a probabilistic estimation of spindle events as a function of their amplitudes and frequencies.

To evaluate the ability of our algorithm to detect spindles, we have analyzed two data sets. The algorithm was first trained and tested against the scoring of one expert on EEG samples that we

obtained in seven healthy children (Study 1). The objective was to evaluate the ability of the algorithm to detect spindles for this category. Afterward, a second study compared performance of two additional experts versus the algorithm on six EEG samples from adult patients suffering from different pathologies (Study 2). The purpose was to submit the method to more challenging conditions, since pathologies may affect the shape of the spindles. Recording conditions in this database were suboptimal but closer to common clinical situations (i.e. only three derivations were available, sampling frequency was lower, artefacts were more frequent, and noise levels were higher).

2. Materials and methods

2.1. Recordings

2.1.1. Study 1

A group of 7 healthy children were included in the present analysis (4 girls and 3 boys, with a mean age of 10.1 years, ranging from 8.5 to 11.6 years). All subjects underwent scalp EEG with 250–256 Hz sampling frequency and 16-bit resolution. 10–20 electrode placement was used (with 32 electrodes). Data from children 1 and 2 were used to train the algorithm. The study protocol was approved by the local Committee (Hôpital Erasme Ethical Protocol P2008/338).

2.1.2. Study 2

Six polysomnographic recordings from adult patients (3 men and 3 women aged between 31 and 53 years) with various pathologies (restless legs syndrome, insomnia, apnoea/hypopnoea syndrome) were obtained from the DREAMS project database,[1] previously used in various clinical and research investigations, and more specifically in spindle detection (Devuyst et al., 2006, 2011). In these recordings only 3 derivations were available and sampling frequency ranged from 50 Hz to 200 Hz.

2.2. EEG sample selection

2.2.1. Study 1

A paediatric sleep expert (D.V.) performed sleep staging. One extract of each record, corresponding to the first occurrence of stage-2 sleep, was chosen for this study. Sleep staging was performed according to Grigg-Damberger et al. (2007). The extract durations ranged from 8.8 min to 29.8 min (16.3 ± 8.0 min), with an overall number of spindles of about 900.

2.2.2. Study 2

A segment of 30 min was extracted from each whole-night recording.

2.3. Instructions given to the experts and the algorithm

For Study 1, the expert was asked to mark all the spindles that he could find in the 7 extracts by analyzing all derivations of the recording. For Study 2, the samples were given independently to two experts for spindles scoring (data provided with the DREAMS project database).

The instructions given to the algorithm were identical to the ones given to the experts, i.e. the algorithm was programmed to detect spindles in all derivations of the recording. However, derivations with a mean amplitude lower than 5 μV_{RMS} (root mean square) were considered disconnected and therefore discarded.

2.4. Description of the algorithm

The detection algorithm presented in this paper follows these three major steps:

(a) amplitude–frequency extraction,
(b) spindle distribution modelling, and
(c) spindle detection.

Those three steps are detailed in the next subsequent sections.

2.5. Amplitude–frequency extraction

The signal was segmented using a 0.5 s moving window shifted by 125 ms steps along the EEG record. The window length was selected to be (i) small enough to detect short spindles and to divide the signal into quasi-stationary segments, and (ii) large enough to correctly estimate spindle amplitude and frequency (Barlow, 1985).

For each window:

- the RMS amplitude of the signal was calculated;
- the FFT spectrum was calculated with a Hamming window. Zero padding was used to obtain a 0.1 Hz resolution. This method has previously shown reliable spindle frequency estimation (Huupponen et al., 2006). The main frequency of the signal was evaluated as the one showing the highest power in the 5–35 Hz frequency range.

An illustration of the RMS amplitude and main frequency estimation is given in Fig. 1.

2.6. Spindle distribution modelling

Fig. 2a shows, in green, the spindle segments (i.e. the 0.5 s segments of the signal coinciding with the spindles scored by the expert) in terms of amplitude and frequency for child 1. To draw the figure, for each segment, the derivation with the highest energy in the 12–14 Hz bandpass, at the time coinciding with the segment, was chosen. The amplitude–frequency values were extracted, for the chosen derivation and corresponding time instant, and used to display each point of the figure. The 12–14 Hz frequency range (shown with blue lines in Fig. 2a) seemed to be too narrow, at least for this child, as this frequency range does not include all the spindle segments. Fig. 3 (bottom) illustrates, for the same child, the histogram of the spindle segments for a given derivation. Again, it can be seen that the spindle segments are not all included in the 12–14 Hz frequency range. This is not surprising as various authors consider this frequency range to be too strict for spindle detection (Jankel and Niedermeyer, 1985).

Hence, the segments included in the 12–14 Hz frequency range were supposed to be part of a larger population, corresponding to spindle segments. Therefore the statistics underlying spindle model should include the segments in the 12–14 Hz frequency range but should also consider that they represent a truncated set of the global histogram. The shape of the distribution of spindle segments, illustrated in Figs. 2a (in green) and 3 (bottom), motivated us to choose a bivariate normal distribution to model the amplitude–frequency behaviour of the spindles. Normal distribution of spindle amplitude and frequency has been previously reported (Zeitlhofer et al., 1997).

To build such a model, signal segments were first analyzed according to their amplitude–frequency characteristics. S1 segments were defined as segments for which the main frequency

[1] Stéphanie Devuyst and Thierry Dutoit, TCTS Laboratory, University of MONS and Myriam Kerkhofs, CHU de Charleroi Sleep Laboratory, Université Libre de Bruxelles. Freely available at http://www.tcts.fpms.ac.be/~devuyst/Databases/DatabaseSpindles/.

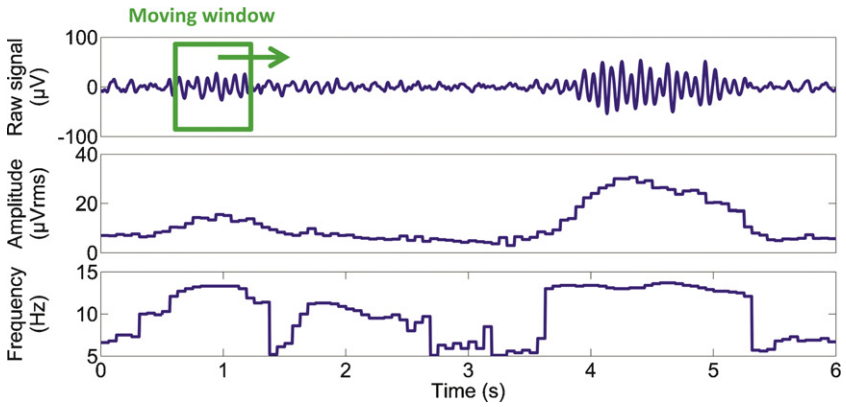

Fig. 1. Illustration of RMS amplitude and main frequency estimation.

ranged from 12 to 14 Hz with a RMS amplitude higher than one standard deviation from the mean. This spindle definition is directly derived from the 'classical' spindle definition (Rechtschaffen and Kales, 1968; Grigg-Damberger et al., 2007) since (i) the proposed frequency range is used and (ii) a spindle should be distinguishable from the background to be visually scored. Therefore, these signal segments are most likely parts of spindle segments. Fig. 2a shows S1 segments (in red) in terms of amplitude and frequency and Fig. 3 (top) illustrates the histogram of these segments for a given derivation.

Based on the S1 histogram, Maximum-Likelihood Estimation (MLE) (Pfanzagl, 1994) was used to estimate the normal model parameters. Again, since the segments with an amplitude and/or a frequency below or above the corresponding thresholds are excluded from S1, S1 is considered as being a truncated set of the spindle segments. In this regard, truncated normal distributions were used for MLE. Fig. 3 suggests that the resulting model (Fig. 3, middle) fit more closely the scoring of the expert (Fig. 3, bottom) compared to segments obtained with fixed thresholds (Fig. 3, top).

Subsequently, a second set of segments, labelled S2, was based on the spindle normal model. This second set was constituted by the segments belonging to the 90% TI of the amplitude–frequency normal model of the corresponding derivation. A 90% TI was chosen to focus on the 90% central population (discarding the 10% extreme values) and to better correspond to the scoring of the expert. Fig. 2b shows S2 distribution (in red) and compares it to the scoring of the expert (in green). Again, the figure suggests that the segments derived from the model (Fig. 2b) tend to fit more closely the scoring of the expert than the segments derived from fixed thresholds (Fig. 2a).

Figs. 2 and 3 also show that the majority of spindle segments are found within a frequency range of approximately 12–16 Hz. However, the expert scored some segments with much lower frequencies (as low as 5 Hz). This is not surprising as, for instance, it is difficult to visually define the beginning and the end of a spindle with a 125 ms resolution (i.e. the chosen window shift), and therefore, some segments may be included although they do not belong to a spindle. The proposed method allows those segments to be discarded, focusing only on segments belonging to spindles to build its model.

2.7. Spindle detection

Spindle detection is based on the S2 segments. Subsequent segments were merged together into S2 segment arrays. Two S2 segment arrays were merged together if the time that separated them was shorter than the length of the longest array. However, arrays more than 0.5 s away from each other were not merged.

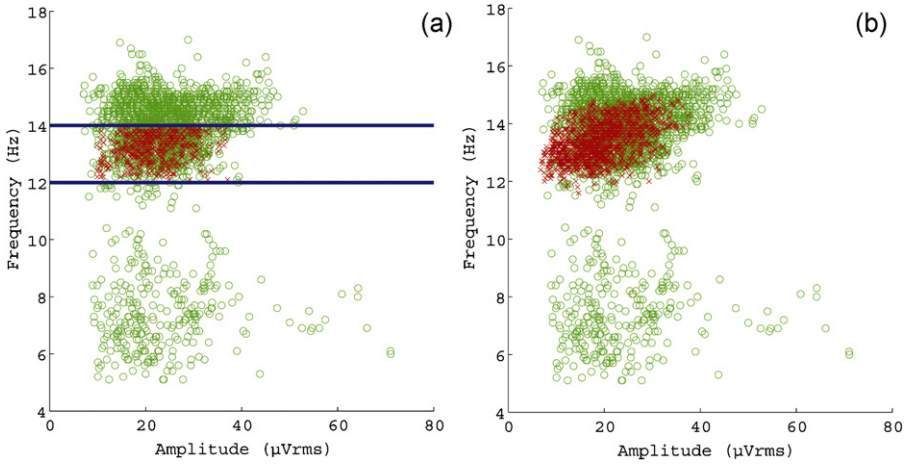

Fig. 2. Amplitude–frequency representation of events for child 1. Green: spindle segments, corresponding to the 0.5 s segments of the signal coinciding with the spindles scored by the expert. To draw the figure, for each segment, the derivation with the highest energy in the 12–14 Hz bandpass, at the time coinciding with the segment, was chosen. The amplitude–frequency values were extracted, for the chosen derivation and corresponding time instant, and used to display each point of the figure. Red: spindle events scored by the algorithm. (a) Segments belonging to the 12–14 Hz bandpass (shown with blue lines) with an amplitude higher than one standard deviation from the mean (i.e. S1 segments). (b) Segments belonging to the 90% TI of the amplitude–frequency normal model of the corresponding derivation (i.e. S2 segments). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

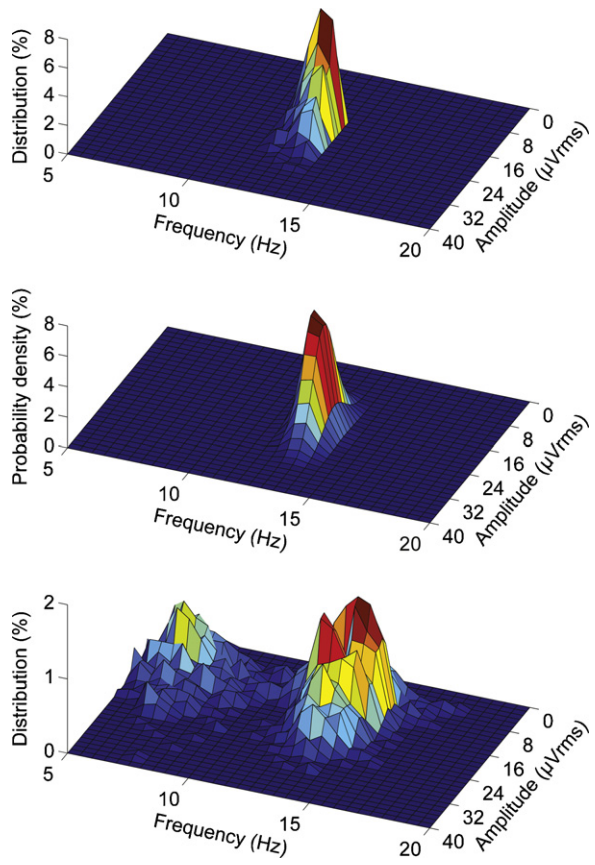

Fig. 3. Illustration of S1 histogram (top), the corresponding normal distribution model (middle), and the histogram of spindles scored by the expert (bottom) for a given derivation of child 1.

In order to be considered as a spindle detection event, a spindle array has:

- to be constituted of segments whose amplitude and frequency are included in the 90% TI of the model of the corresponding derivation (i.e. belonging to S2);
- to be at least 0.5 s long (as required in the definition of the spindle (Rechtschaffen and Kales, 1968; Iber et al., 2007)).

Furthermore, in Study 1, spindle arrays have to appear in at least two derivations to be considered as spindle detection events. As only three derivations were available in Study 2, spindles were considered even if they appeared in only one derivation.

Identical spindle detection was performed with the S1 events, i.e. before normal modelling, to evaluate the benefit of the normal modelling procedure.

2.8. Sensitivity to the initial frequency range

The normal modelling is based on a segment distribution limited to the 12–14 Hz frequency range (i.e. S1), as it clearly contains spindles. As spindle frequency is traditionally defined at least within the 12–14 Hz frequency range (Rechtschaffen and Kales, 1968; Grigg-Damberger et al., 2007), it seemed important to choose this frequency range for a first study. Since the method takes into account the truncated aspect of S1 segments, the initial frequency range (in this case 12–14 Hz) should have little influence on the statistical model. Nonetheless, it could still have some influence on the model's estimations. For instance, Figs. 2b and 3 suggest that the models could show lower mean spindle frequencies than the expert.

To evaluate the impact of the initial frequency range on the algorithm performance, all possible frequency ranges included in 10–16 Hz were computed (with 1 Hz step and 2 Hz minimum interval between the lower and upper frequencies).

2.9. EMG artefact and alpha wave rejection

For Study 2, EMG artefact and alpha wave rejection were implemented given the low quality of the recordings. EMG artefact rejection was based on a previous method (Devuyst et al., 2011). To make the distinction between spindles and EMG artefacts, the spindles power as related to the total power was taken into account. It is defined as:

$$\text{Relative spindle power} = \frac{\int_{11}^{15} S(f,t)df}{\int_{0.5}^{40} S(f,t)df} \quad (1)$$

where $S(f,t)$ is the spectrogram of the signal.

Epochs during which relative spindle power was inferior to a certain threshold were removed. Devuyst et al. (2011) proposed a threshold of 22%. In this study, threshold values from 0% to 50% (with 1% step) were evaluated.

Alpha wave rejection was also implemented, based on a previous method (Huupponen et al., 2003). Epochs with a duration of 3 s or more with a frequency below 12.5 Hz were rejected as probable alpha activity.

We assessed the improvement of both EMG artefact rejection and alpha wave rejection by successively adding those steps in the spindle detection algorithm while evaluating their impact on the detection.

2.10. Statistical analysis of automatically recognized spindles

For Study 1, the detection was tested against the scoring of the expert. For Study 2, the detection was evaluated against the union of the scoring of both experts, as proposed before for that database (Devuyst et al., 2011).

Various types of statistical analysis have been previously proposed in the literature. Some are based on the presence of one or more spindles in an epoch (Duman et al., 2009), which seems less accurate than the evaluation of spindle detection itself. In this work, statistics based on spindle detection events were applied. Similarly, considering the statistics based on spindle detection events, various definitions of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) are given, depending on the minimum overlap between a scored and a detected spindle. In this work, a TP was counted only if the scoring of the expert and the algorithm overlapped (i.e. if there was a time interval during which both scored and detected spindles occurred). A FN or a FP was counted otherwise, if either the expert alone or the algorithm alone scored a spindle, respectively. In this regard, the statistics may be considered quite strict, as they are based on spindle detection events rather than on an epoch, and as no time interval is allowed between a scored and a detected spindle in order to register a TP.

Various indicators were chosen to evaluate the algorithm and to compare our results to those reported in previous works: the true positive rate (or sensitivity), the false positive rate, the true negative rate (or specificity), and the detection correlation coefficient. The true positive rate is defined as:

$$\text{True positive rate} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

It measures the extent to which the algorithm is capable of detecting spindles. The false positive rate is defined as:

$$\text{False positive rate} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{3}$$

where FP + TN is equal to the number of all non-spindle seconds in the recording. It measures the proportion of false alarms of all detections. The specificity is defined as:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{4}$$

where TN is equal to the number of all non-spindle seconds in the recording where no FP was present. It measures the proportion of negatives that were correctly identified.

Furthermore, the detection correlation coefficient (Wilson et al., 1996), defined as:

$$\text{Detection correlation coefficient} = \sqrt{\text{Sensitivity} \cdot \text{Selectivity}} \tag{5}$$

was used to give an overall estimation of the agreement between the algorithm and the expert. The selectivity (also labelled positive predictive value or precision) is defined as:

$$\text{Selectivity} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{6}$$

A Receiver Operator Characteristics (ROC) curve was also used, to assess how the true-positive rate of the algorithm can be traded against its false-positive rate. It was obtained by increasing the TI from 0 to 99.9% (100 points, logarithmically spaced).

TP, FP, and FN were further analyzed to characterize the limits of our method. The amplitude and the main frequency of TP, FP, and FN were calculated. The derivation used to calculate the amplitude and the frequency of each spindle was the one showing the highest energy in the 12–14 Hz bandpass (for positives events, i.e. TP and FP, only derivations for which a spindle was detected were considered). The difference between the amplitude of the spindles detected by the algorithm and the amplitude expectation (location of the peak) of the normal model of the corresponding derivation was calculated for TP, FP, and FN. Similarly, the difference between the main frequency of the spindles detected by the algorithm and the frequency expectation of the normal model of the corresponding derivation was also calculated. Means and standard deviations of these differences were calculated. In this way, relative positioning of detected spindle amplitude and frequency with the model could be assessed in each group (TP, FP, and FN). Similarly, means and standard deviations of TP, FP, and FN lengths were calculated. Sine quality, defined as the relative power of the signal in a ±2 Hz frequency band, was also calculated. This indicator reflects the sine shape and rhythmic aspect of the spindle; a spindle with a perfect sine shape at a single frequency would show a sine quality of 100%. The overlap between the scoring of the expert and the one of the algorithm was calculated for TP. The overlap is defined as:

$$\text{Overlap} = \frac{t_{Exp+Alg}}{t_{Exp+Alg} + t_{Exp} + t_{Alg}} \tag{7}$$

where $t_{Exp+Alg}$ is the time period when spindles are jointly scored by the expert and the algorithm, t_{Exp} is the time period when spindles are scored by the expert alone and t_{Alg} is the time period when spindles are scored by the algorithm alone.

Finally, the inter-rater agreement was evaluated for the experts of the second study, through the average of all pairwise expert versus expert true positive rate, false positive rate, specificity, detection correlation coefficient and overlap values, as proposed by Wilson and Emerson (2002).

Table 1
Statistical evaluation of the algorithm for Study 1.

Child	Extract duration (min)	Before normal modelling					After normal modelling				
		True positive rate (%)	False positive rate (%)	Specificity (%)	Detection correlation coefficient (%)	Overlap (%)	True positive rate (%)	False positive rate (%)	Specificity (%)	Detection correlation coefficient (%)	Overlap (%)
1	8.8	32.9	0.8	99.3	54.5	49.1	63.5	3.8	96.4	70.5	62.6
2	14.8	65.7	2.9	97.2	66.2	66.9	81.4	7.4	91.9	63.5	69.1
3	29.8	87.3	2.8	98.1	85.2	70.0	91.2	6.5	94.2	79.3	76.1
4	25.1	68.5	1.8	98.6	78.8	58.1	90.3	6.7	93.9	83.4	70.8
5	9.2	48.6	2.0	98.5	55.9	58.3	75.7	7.8	91.5	55.8	58.9
6	13.8	55.3	1.2	99.1	68.7	53.1	67.1	4.2	96.0	67.1	70.0
7	12.7	36.0	1.5	99.0	52.9	36.7	80.2	5.0	95.4	74.8	56.4
Avg.	16.3	56.4	1.8	98.5	66.0	56.0	78.5	5.9	94.2	70.6	66.3

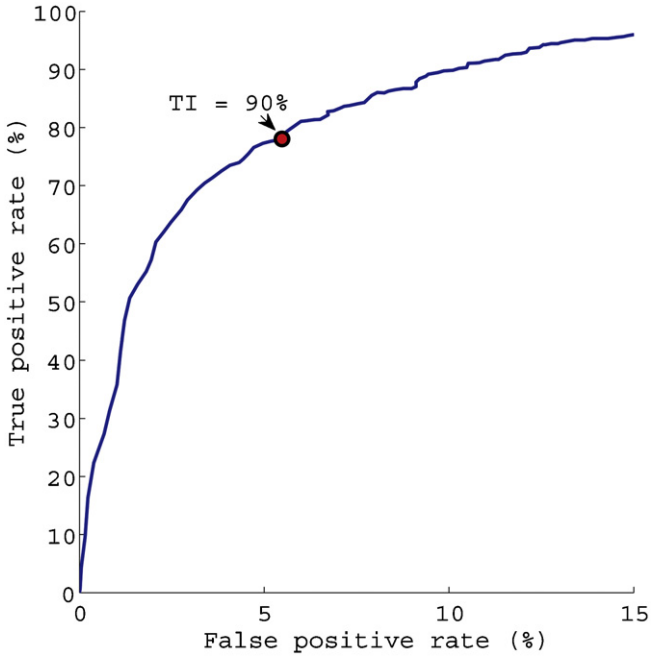

Fig. 4. Receiver Operator Characteristic (ROC) curve. Average true positive rate and average false positive rate are given for a TI ranging from 0 to 99.9% (100 points, logarithmically spaced).

3. Results

3.1. Algorithm performance in healthy children (Study 1)

Table 1 compares, in the first study, the true positive rate, false positive rate, specificity, detection correlation coefficient, and overlap for spindles detected with and without normal modelling for each child (i.e. based on S2 and S1 segments, respectively). Normal modelling enhanced overall detection performance, as the average detection correlation coefficient was increased by 4.6% using this method.

For each child, there was a large increase in the true positive rate (22.1% higher on average), which was achieved at the cost of a higher false positive rate (4.1% higher on average), and a lower specificity (4.3% lower on average). Focusing on TP, the overlap between the algorithm and expert's scoring was also higher (10.3% on average) when modelling was used.

3.2. True-positive, false-positive, and false-negative analysis (Study 1)

Spindle TP, FP, and FN were further analyzed in Tables 2–4, respectively. Table 2 indicates that both TP amplitudes and TP frequencies were lower than the mean of the corresponding derivation in the model (0.9 μV_{RMS} and 1.1 Hz on average, respectively). This may be due to the analysis performed on whole spindles, rather than solely on S2 events, and therefore also including signal portions that were not taken into account by normal models. Compared to Table 2, Tables 3 and 4 indicate that both FP and FN showed lower amplitudes and smaller lengths than TP. As expected, smaller and shorter spindles were more difficult to evaluate. Sine quality and frequency values suggest no specific trend.

3.3. Receiver Operator Characteristic curve (Study 1)

The ROC curve, in Fig. 4, shows that 75% and 95% of true positive rates were achieved with corresponding false positive rates of less than 5% and 13%, respectively. Furthermore, the algorithm is

Table 2
True-positive detection analysis in Study 1. Amplitude, frequency, sine quality, defined as the relative power of the signal in a ±2 Hz frequency band, and the length are given. The difference between the amplitude and the frequency of detected spindles and, respectively, amplitude and frequency expectation of the normal model of the corresponding derivation is also given.

Child	Extract duration (min)	Mean						Standard deviation					
		Amplitude (μV)	Spindle amplitude−amplitude expectation of the model (μV)	Frequency (Hz)	Spindle frequency−frequency expectation of the model l (Hz)	Sine quality	Length (s)	Amplitude (μV)	Spindle amplitude−amplitude expectation of the model (μV)	Frequency (Hz)	Spindle frequency−frequency expectation of the model l (Hz)	Sine quality	Length (s)
1	8.8	24.0	−2.0	12.9	−0.4	0.6	1.8	5.5	4.5	1.3	1.3	0.1	0.7
2	14.8	20.8	−1.1	12.0	−0.9	0.6	1.6	2.7	2.9	0.9	0.9	0.1	0.5
3	29.8	13.4	−0.7	11.9	−0.9	0.7	1.5	2.7	2.6	0.9	0.9	0.1	0.5
4	25.1	17.4	−1.9	11.9	−1.0	0.7	1.7	4.4	3.3	1.1	1.1	0.1	0.6
5	9.2	17.2	1.3	11.9	−0.9	0.6	1.1	3.0	3.2	1.3	1.2	0.1	0.4
6	13.8	15.9	−1.0	11.1	−1.7	0.6	2.2	3.3	3.3	0.9	0.9	0.1	0.9
7	12.7	21.4	−1.1	11.1	−1.5	0.7	2.0	4.1	3.7	0.8	0.8	0.1	0.7
Avg.	16.3	18.6	−0.9	11.8	−1.1	0.7	1.7	3.7	3.4	1.0	1.0	0.1	0.6

Table 3
False-positive detection analysis in Study 1. Amplitude, frequency, sine quality, defined as the relative power of the signal in a ±2 Hz frequency band, and the length are given. The difference between the amplitude and the frequency of detected spindles and, respectively, amplitude and frequency expectation of the normal model of the corresponding derivation is also given.

Child	Extract duration (min)	Mean						Standard deviation					
		Amplitude (μV)	Spindle amplitude−amplitude expectation of the model (μV)	Frequency (Hz)	Spindle frequency−frequency expectation of the model l (Hz)	Sine quality	Length (s)	Amplitude (μV)	Spindle amplitude−amplitude expectation of the model (μV)	Frequency (Hz)	Spindle frequency−frequency expectation of the model l (Hz)	Sine quality	Length (s)
1	8.8	17.7	−6.0	12.6	−0.7	0.7	0.9	1.7	3.7	0.8	0.9	0.1	0.3
2	14.8	17.3	−3.8	11.7	−1.3	0.6	1.1	2.8	2.6	1.1	1.1	0.1	0.6
3	29.8	11.8	−1.9	12.0	−0.8	0.7	0.9	2.7	2.4	1.0	1.0	0.1	0.4
4	25.1	16.5	−3.1	11.8	−1.1	0.6	0.9	6.9	2.6	1.4	1.4	0.1	0.4
5	9.2	14.8	−1.3	11.5	−1.1	0.6	1.1	2.0	2.2	1.1	1.0	0.1	0.6
6	13.8	13.0	−1.6	11.9	−0.9	0.7	1.0	2.1	2.0	0.9	0.8	0.1	0.5
7	12.7	19.0	−2.6	11.7	−0.8	0.7	0.9	4.1	3.4	1.2	1.2	0.1	0.4
Avg.	16.3	15.7	−2.9	11.9	−1.0	0.6	1.0	3.2	2.7	1.1	1.1	0.1	0.4

Table 4
False-negative detection analysis in Study 1. Amplitude, frequency, sine quality, defined as the relative power of the signal in a ±2 Hz frequency band, and the length are given. The difference between the amplitude and the frequency of detected spindles and, respectively, amplitude and frequency expectation of the normal model of the corresponding derivation is also given.

Child	Extract duration (min)	Mean						Standard deviation					
		Amplitude (μV)	Spindle amplitude−amplitude expectation of the model (μV)	Frequency (Hz)	Spindle frequency−frequency expectation of the model l (Hz)	Sine quality	Length (s)	Amplitude (μV)	Spindle amplitude−amplitude expectation of the model (μV)	Frequency (Hz)	Spindle frequency−frequency expectation of the model l (Hz)	Sine quality	Length (s)
1	8.8	21.3	−5.4	12.7	−0.7	0.6	1.1	7.7	8.0	2.0	2.0	0.1	0.6
2	14.8	18.8	−2.8	12.1	−0.9	0.5	1.2	5.4	4.7	1.5	1.5	0.0	0.5
3	29.8	11.3	−1.1	12.3	−0.7	0.7	0.8	2.3	2.8	1.6	1.6	0.1	0.3
4	25.1	18.7	−2.1	12.4	−0.5	0.6	1.0	19.5	19.8	2.3	2.3	0.1	0.5
5	9.2	19.8	3.4	12.4	−0.4	0.6	0.7	8.7	8.3	3.0	2.9	0.1	0.2
6	13.8	10.4	−4.8	12.1	−0.8	0.6	1.1	3.3	3.8	1.6	1.6	0.1	0.5
7	12.7	16.4	−4.6	11.1	−1.5	0.6	1.0	3.9	4.1	1.6	1.6	0.1	0.3
Avg.	16.3	16.7	−2.5	12.1	−0.8	0.6	1.0	7.2	7.3	1.9	1.9	0.1	0.4

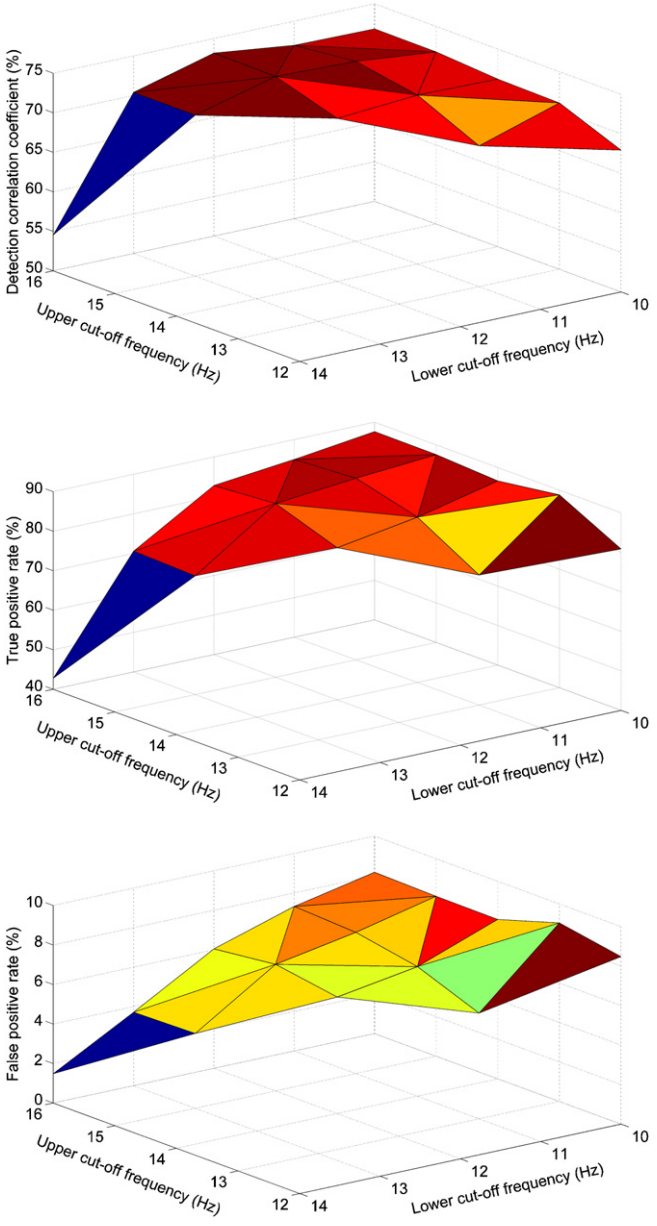

Fig. 5. Impact of the initial frequency range on the algorithm performance (detection correlation coefficient, and true and false positive rates). All possible frequency ranges included in 10–16 Hz are shown (with 1 Hz step and 2 Hz minimum interval between the lower and upper frequencies).

intrinsically adaptable: the true positive rate can be augmented by increasing the TI at the cost of a higher false positive rate, and vice versa.

3.4. Sensitivity to the initial frequency range (Study 1)

The impact of the initial frequency range on the detection outcome (the detection correlation coefficient, and the true and false positive rates) is shown in Fig. 5. The impact on the detection correlation coefficient was small (70–74% range) providing that the lower cut-off frequency was less than 13 Hz and the upper cut-off frequency was higher than 14 Hz. The true positive rate tended to be higher, at the cost of a higher false positive rate, when the lower cut-off frequency was decreased.

3.5. Algorithm performance in challenging recording conditions (Study 2)

Table 5 shows the algorithm's performance for the second study (without EMG artefact and alpha wave rejection). Normal modelling enhanced the overall performance in detection, even more than in the first study. Indeed, the average detection correlation coefficient increased by 20.6% with this method (13.2% when discarding patient 4; see below). As expected, this performance was lower than for the first study (e.g. the detection correlation coefficient dropped from 70.6% to 62.1% on average), since recordings were more challenging. The algorithm was then improved with EMG artefact and alpha wave rejection (see Table 6). A relative spindle power threshold of only 4% showed the best detection correlation coefficient (69.0% on average), leading to a performance close to the one achieved in Study 1 (70.6% on average). The alpha wave rejection had no impact on the detection. This could be due to the underlying spindle model that focused on the patient specific spindle frequency, hence rejecting alpha wave in an inherent way. Inter-rater agreement (see Table 7) showed that the performance of the experts was inferior to the one of the algorithm (e.g. the experts' detection correlation coefficient was only 44.2% on average).

The algorithm failed to converge for patient 4 (i.e. aberrant negative RMS amplitudes were included in the 90% TI range). A close examination of this patient's recording disclosed the largest artefacts, which implied a large variation on the RMS value of the samples used to generate the model (on average 5.6 times larger than for the other patients), eventually leading to aberrant models. The poor quality of EEG recording in this patient also made the visual scoring difficult, as the experts only agreed on 8.7% of the spindles (6 spindles out of 69) and the detection correlation coefficient was only 18.1%.

4. Discussion

We have developed an algorithm dedicated to the detection of spindles that adapts to intersubject and intrasubject variations in amplitude and frequency through normal modelling. Our analyses show that normal modelling enhances performance in terms of true positive rates, false positive rates, specificity, detection correlation coefficients, and overlap, compared to fixed amplitude and frequency thresholds. The algorithm is quite insensitive to the input frequency range used to model the spindle distribution, which is a major advantage as there is no consensus on a spindle frequency range.

Recordings from healthy children were used in the first study. Few data have been reported on automated spindle detection in children although they are of recent interest in the field of sleep research, which motivated our choice. Since spindles show important modifications in amplitude and frequency during maturation, the advantage of our algorithm includes an adaptation to those variations.

In the second study, our results were validated in more challenging conditions. Providing the addition of an EMG artefact rejection method, results were closely similar to those obtained in the first study (i.e. average detection correlation coefficients were 70.6% in the first study and 69.0% in the second study).

In both studies, some recordings were more difficult to score than others, leading to poorer performances. For instance, in Study 1, recordings 1 and 6 showed a true positive rate lower than 70%, and recordings 2 and 5 showed false positive rates higher than 7%. In Study 2, the algorithm could not achieve a meaningful model for patient 4. This patient's recording had a high level of artefacts and indeed was quite difficult to be scored by experts as they only agreed on 8.7% of the spindles. With such a low agreement

Table 5
Statistical evaluation of the algorithm for Study 2.

Subject	Before normal modelling					After normal modelling				
	True positive rate (%)	False positive rate (%)	Specificity (%)	Detection correlation coefficient (%)	Overlap (%)	True positive rate (%)	False positive rate (%)	Specificity (%)	Detection correlation coefficient (%)	Overlap (%)
1	33.6	0.3	99.8	55.0	63.2	82.8	3.8	97.0	72.7	64.3
2	40.3	0.5	99.6	55.9	61.2	68.8	3.6	97.1	56.1	57.6
3	25.0	0.3	99.8	40.2	50.8	77.3	7.0	92.6	40.9	58.3
4	1.6	0.4	99.7	4.5	51.7	–	–	–	–	–
5	32.0	0.9	99.4	46.5	65.7	80.6	2.5	98.3	72.9	64.9
6	25.6	0.3	99.8	46.9	64.0	65.8	2.0	98.7	67.9	66.8
Avg.	26.4	0.5	99.7	41.5	59.4	75.1	3.8	96.7	62.1	62.4

Table 6
Statistical evaluation of the algorithm enhanced with EMG artefact rejection, for Study 2.

Subject	Before normal modelling					After normal modelling				
	True positive rate (%)	False positive rate (%)	Specificity (%)	Detection correlation coefficient (%)	Overlap (%)	True positive rate (%)	False positive rate (%)	Specificity (%)	Detection correlation coefficient (%)	Overlap (%)
1	31.3	0.3	99.8	52.9	62.5	76.9	2.8	97.7	72.7	64.7
2	40.3	0.3	99.8	58.1	61.2	63.6	2.2	98.3	59.9	64.2
3	22.7	0.3	99.8	38.9	51.3	75.0	0.9	99.2	71.1	60.5
4	1.6	0.3	99.8	5.1	51.7	–	–	–	–	–
5	30.1	0.8	99.5	46.0	65.8	75.7	1.9	98.6	72.9	64.9
6	25.6	0.3	99.8	46.9	64.0	64.1	1.7	98.9	68.3	65.3
Avg.	25.3	0.4	99.8	41.3	59.4	71.1	1.9	98.6	69.0	63.9

rate between experts, their scoring could not be retained as a 'gold standard' for an accurate evaluation of the algorithm's performance. Therefore, recordings that include unclear spindles should be analyzed with caution before running the algorithm: when experts do not agree on numerous spindles (Campbell et al., 1980; Zygierewicz et al., 1999), the algorithm cannot, a fortiori, perform a high quality detection. In this regard, the user should be given the opportunity to evaluate the quality of the recording which results in the quality control of the spindle detection. For instance, the variation in amplitude of samples eligible for spindles could be used as a quality indicator; a large value would suggest that artefacts are present in the spindle bandpass. Since S1 samples are defined as segments belonging to spindles according to the 'classical' spindle definition (Rechtschaffen and Kales, 1968; Grigg-Damberger et al., 2007), the standard deviation of S1 segment amplitudes could be used as such indicator. This choice would be convenient, because it would also give an estimation of the quality of the normal model used in our algorithm, since it is based on S1 samples. Furthermore, a stopping criterion could be added to the algorithm to avoid building a model when the level of artefacts is believed to be too high. For instance, a stopping criterion ranging between 6 and 9 μV_{RMS} would have stopped the algorithm for all the derivations of patient 4 in Study 2 (i.e. the patient for which algorithm could not achieve a meaningful model) while performing the detection for all the derivations of the remaining patients (in both studies).

The heterogeneity of datasets and the means of assessing performance make the direct comparisons with other published spindle detection algorithms difficult. However, some comparisons were possible with works that published at least two statistic indicators common to the ones chosen in this paper.

Huupponen et al. (2000a) proposed an algorithm with a true positive rate ranging between 73.4% and 84.9% and a false positive rate ranging between 2.3% and 5.4%, hence with performance similar to our algorithm for the first study and slightly more sensitive and less selective for the second study. However, they evaluated their algorithm using the following criterion: "one second before or after a scored spindle or 1 s after a false-positive, no false-positives were counted." If such a criterion was used to evaluate our algorithm, it would have shown a 4.3% false positive rate for the first study and 1.7% for the second one, hence tending to be lower than the one from Huupponen et al. (2000a). The same group proposed another algorithm (Huupponen et al., 2007) with a true positive rate ranging between 51.2% and 86.5% and a false positive rate ranging between 26.4% and 46.0%. Their definition of the false positive rate differed from ours and from the one used in their previous work, as the false positive rate is set to the ratio of FP to the sum of TP and FP. Using this definition, our algorithm shows a false positive rate of 35.5% for the first study and 32.9% for the second one, therefore demonstrating similar performance.

Table 7
Statistical evaluation of the experts for Study 2.

Subject	True positive rate (%)	False positive rate (%)	Specificity (%)	Detection correlation coefficient (%)	Overlap (%)
1	44.7	3.0	97.2	41.3	61.8
2	62.8	1.2	99.0	62.7	67.8
3	55.7	1.1	98.9	33.7	63.1
4	18.8	1.6	98.3	18.1	68.6
5	57.5	1.8	98.3	56.2	68.4
6	53.3	2.2	98.1	53.1	66.5
Avg.	48.8	1.8	98.3	44.2	66.0

Ventouras et al. (2005) proposed an algorithm with a true positive rate ranging between 80.2% and 92.9% and a specificity ranging between 86.9% and 95.0%. Clemens et al. (2005) proposed an algorithm that showed true and false positive rates of 89.7% and 21.4%. Schönwald et al. (2006) evaluated another algorithm using 1 s epochs, which showed a true positive rate and a specificity equal to 80.6%. Therefore, these three published algorithms are more sensitive but less selective than ours.

Duman et al. (2009) chose to assess performance based on the presence of one or more spindles in a 30 s epoch, with a higher true positive rate (96.17%) and a specificity (95.54%) similar to the first study but lower than the second one. Furthermore, statistics were based on the presence of one or more spindles in an epoch, rather than on spindle detection itself, so it is likely that FP and FN balanced each other.

Finally, Devuyst et al. (2011) proposed an algorithm that was tested on the database used in the second study, and showed a sensitivity of 70.2%, a false positive rate of 1.38% and a specificity of 98.6%. In this regard, this study obtained similar performance levels compared to our algorithm. However, it should be noted that they did not exclude any patient from the dataset.

To our knowledge, only three algorithms that have assessed spindles detection in children and infants have been published so far. Causa et al. (2010) proposed an algorithm with 92.2% sensitivity and 90.1% specificity, hence more sensitive but less selective than ours. Estévez et al. (2007) proposed an algorithm, tested on a single infant, with 96.3% accuracy and an 89.7% detection rate. Held et al. (2004) proposed an algorithm, tested on two infants, with 87.7% expert agreement and 91.9% precision. Comparison with those two last algorithms seems more difficult since performance indicators differ and since the number of subjects used for assessment is low.

All together, the performance obtained in the present work is quite comparable to previously published algorithms. In some cases, it tends to be more selective at the cost of a lower sensitivity for the chosen TI (i.e. 90%). Higher sensitivity may be reached by lowering the TI at the cost of lower selectivity, as shown in Fig. 4.

A normal distribution was chosen to model the spindle distribution in amplitude and frequency, as this type of distribution has been previously reported (Zeitlhofer et al., 1997). However, as other studies have reported a non-Gaussian spindle distribution (Schönwald et al., 2003, 2006), further studies should investigate the impact of other types of distribution on the detection.

In conclusion, we have described an algorithm that automatically adapts to differences in spindle amplitude and frequency between subjects (intersubject spindle variation) as well as within a single subject (intrasubject spindle variation) – hence exhibiting improved performance metrics, without the need of a priori knowledge from an expert. Our algorithm compares positively with other spindle detection algorithms and could be used in common medical practice. Furthermore, as normal modelling is inherent to the algorithm, it automatically determines the statistical parameters of spindle amplitude and frequency, which is of great help in various research applications.

Funding

C.D. is a Senior Research Associate with the “Fonds National de la Recherche Scientifique” (FNRS, Brussels, Belgium). C.U. is currently supported by ARC grant “Pathophysiology of Memory Consolidation” at the Université Libre de Bruxelles (ULB), and was supported by a ULB grant from Fondation Vigneron.

Conflict of interest

None.

Acknowledgement

The authors would like to thank Dr. Rachel Leproult from the Neuropsychology and Functional Neuroimaging Research Unit for her comments and suggestions on this paper.

References

Acir N, Güzeliş C. Automatic spike detection in EEG by a two-stage procedure based on support vector machines. Comput Biol Med 2004;34:561–75.
Barlow JS. Methods of analysis of nonstationary EEGs, with emphasis on segmentation techniques: a comparative review. J Clin Neurophysiol 1985;2:267–304.
Campbell K, Kumar A, Hofman W. Human and automatic validation of a phase-locked loop spindle detection system. Electroencephalogr Clin Neurophysiol 1980;48:602–5.
Causa L, Held CM, Causa J, Estévez PA, Perez CA, Chamorro R, et al. Automated sleep-spindle detection in healthy children polysomnograms. IEEE Trans Biomed Eng 2010;57:2135–46.
Chan S, Baldeweg T, Cross JH. A role for sleep disruption in cognitive impairment in children with epilepsy. Epilepsy Behav 2011;20:435–40.
Clemens Z, Fabó D, Halász P. Overnight verbal memory retention correlates with the number of sleep spindles. Neuroscience 2005;132:529–35.
De Gennaro L, Ferrara M. Sleep spindles: an overview. Sleep Med Rev 2003;7:423–40.
Destexhe A, Sejnowski T. Thalamocortical assemblies: how ion channels, single neurons, and large-scale networks organize sleep oscillations. Oxford University Press: Oxford; 2001.
Devuyst S, Dutoit T, Didier JF, Meers F, Stanus E, Stenuit P, et al. Automatic sleep spindle detection in patients with sleep disorders. Conf Proc IEEE Eng Med Biol Soc 2006;1:3883–6.
Devuyst S, Dutoit T, Stenuit P, Kerkhofs M. Automatic sleep spindles detection – overview and development of a standard proposal assessment method. Conf Proc IEEE Eng Med Biol Soc 2011:1713–6.
Duman F, Erdamar A, Erogul O, Telatar Z, Yetkin S. Efficient sleep spindle detection algorithm with decision tree. Expert Syst Appl 2009;36:9980–5.
Durka PJ, Matysiak A, Montes EM, Sosa PV, Blinowska KJ. Multichannel matching pursuit and EEG inverse solutions. J Neurosci Methods 2005;148:49–59.
Estévez PA, Held CM, Holzmann CA, Perez CA, Pérez JP, Heiss J, et al. Polysomnographic pattern recognition for automated classification of sleep-waking states in infants. Med Biol Eng Comput 2002;40:105–13.
Estévez PA, Zilleruelo-Ramos R, Hernández R, Causa L, Held CM. Sleep spindle detection by using merge neural gas. In: Proc 6th Int Wkshp Self-Organizing Maps, vol. 1; 2007. p. 10–8.
Fish DR, Allen PJ, Blackie JD. A new method for the quantitative analysis of sleep spindles during continuous overnight EEG recordings. Electroencephalogr Clin Neurophysiol 1988;70:273–7.
Fogel SM, Smith CT. The function of the sleep spindle: a physiological index of intelligence and a mechanism for sleep-dependent memory consolidation. Neurosci Biobehav Rev 2011;35:1154–65.
Güneş S, Dursun M, Polat K, Yosunkaya S. Sleep spindles recognition system based on time and frequency domain features. Expert Syst Appl 2011;38:2455–61.
Grigg-Damberger M, Gozal D, Marcus CL, Quan SF, Rosen CL, Chervin RD, et al. The visual scoring of sleep and arousal in infants and children. J Clin Sleep Med 2007;3:201–40.
Heiss JE, Held CM, Estévez PA, Perez CA, Holzmann CA, Pérez JP. Classification of sleep stages in infants: a neuro fuzzy approach. IEEE Eng Med Biol Mag 2002;21:147–51.
Held CM, Causa L, Estévez P, Pérez C, Garrido M, Algarín C, et al. Dual approach for automated sleep spindles detection within EEG background activity in infant polysomnograms. Proc IEEE Eng Med Biol Soc 2004;1:566–9.
Held CM, Heiss JE, Estévez PA, Perez CA, Garrido M, Algarín C, et al. Extracting fuzzy rules from polysomnographic recordings for infant sleep classification. IEEE Trans Biomed Eng 2006;53(October):1954–62.
Huupponen E, Värri A, Himanen SL, Hasan J, Lehtokangas M, Saarinen J. Optimization of sigma amplitude threshold in sleep spindle detection. J Sleep Res 2000a;9:327–34.
Huupponen E, Värri A, Himanen SL, Hasan J, Lehtokangas M, Saarinen J. Autoassociative MLP in sleep spindle detection. J Med Syst 2000b;24:183–93.
Huupponen E, Himanen SL, Hasan J, Värri A. Automatic analysis of electroencephalogram sleep spindle frequency throughout the night. Med Biol Eng Comput 2003;41:727–32.
Huupponen E, De Clercq W, Gómez-Herrero G, Saastamoinen A, Egiazarian K, Värri A, et al. Determination of dominant simulated spindle frequency with different methods. J Neurosci Methods 2006;156:275–83.
Huupponen E, Gómez-Herrero G, Saastamoinen A, Värri A, Hasan J, Himanen SL. Development and comparison of four sleep spindle detection methods. Artif Intell Med 2007;40:157–70.
Hori T, Sugita Y, Koga E, Shirakawa S, Inoue K, Uchida S, et al. Proposed supplements and amendments to ‘A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects’, the Rechtschaffen & Kales (1968) standard. Psychiatry Clin Neurosci 2001;55:305–10.
Iber C, Ancoli-Israel S, Chesson A, Quan SF, For the American Academy of Sleep Medicine. The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications. Westchester: American Academy of Sleep Medicine; 2007.

Jankel WR, Niedermeyer E. Sleep spindles. J Clin Neurophysiol 1985;2:1–35.
Knoblauch V, Martens W, Wirz-Justice A, Kräuchi K, Cajochen C. Regional differences in the circadian modulation of human sleep spindle characteristics. Eur J Neurosci 2003a;18:155–63.
Knoblauch V, Martens WL, Wirz-Justice A, Cajochen C. Human sleep spindle characteristics after sleep deprivation. Clin Neurophysiol 2003b;114:2258–67.
Ktonas PY, Golemati S, Xanthopoulos P, Sakkalis V, Ortigueira MD, Tsekou H, et al. Time-frequency analysis methods to quantify the time-varying microstructure of sleep EEG spindles: possibility for dementia biomarkers. J Neurosci Methods 2009;185:133–42.
Nagata K, Shinomiya S, Takahashi K, Masumura T. Developmental characteristics of frontal spindle and centro-parietal spindle. No To Hattatsu 1996;28:409–17.
Noachtar S, Binnie C, Ebersole J, Mauguiere F, Sakamoto A, Westmoreland B. A glossary of terms most commonly used by clinical electroencephalographers and proposal for the report form for the EEG findings. The International Federation of Clinical Neurophysiology. Electroencephalogr Clin Neurophysiol Suppl 1999;52:21–41.
Peigneux P, Smith C. Memory processing in relation to sleep. In: Kryger M, editor. Principles and practice of sleep medicine. Elsevier: Amsterdam; 2010. p. 335–47.
Petit D, Gagnon JF, Fantini ML, Ferini-Strambi L, Montplaisir J. Sleep and quantitative EEG in neurodegenerative disorders. J Psychosom Res 2004;56:487–96.
Pfanzagl J. Parametric statistical theory. Walter de Gruyter: Berlin; 1994.
Rechtschaffen A, Kales A. A manual of standardized terminology, techniques and scoring systems for sleep stages of human subjects. Washington: Public Health Service, U.S. Government Printing Office; 1968.
Reeves AL, Klass DW. Frequency asymmetry of sleep spindles associated with focal pathology. Electroencephalogr Clin Neurophysiol 1998;106:84–6.
Schmidt C, Peigneux P, Muto V, Schenkel M, Knoblauch V, Münch M, et al. Encoding difficulty promotes postlearning changes in sleep spindle activity during napping. J Neurosci 2006;26:8976–82.
Scholle S, Zwacka G, Scholle HC. Sleep spindle evolution from infancy to adolescence. Clin Neurophysiol 2007 Jul;118:1525–31.
Schönwald SV, Gerhardt GL, de Santa-Helena EL, Chaves MF. Characteristics of human EEG sleep spindles assessed by Gabor transform. Physica A 2003;327:180–4.
Schönwald SV, de Santa-Helena EL, Rossatto R, Chaves ML, Gerhardt GJ. Benchmarking matching pursuit to find sleep spindles. J Neurosci Methods 2006;156:314–21.
Shibagaki M, Kiyono S. Duration of spindle bursts during nocturnal sleep in mentally retarded children. Electroencephalogr Clin Neurophysiol 1983;55:645–51.
Shinomiya S, Nagata K, Takahashi K, Masumura T. Development of sleep spindles in young children and adolescents. Clin Electroencephalogr 1999;30:39–43.
Shimada T, Shiina T, Saito Y. Detection of characteristic waves of sleep EEG by neural network analysis. IEEE Trans Biomed Eng 2000;47:369–79.
Urbain C, Di Vincenzo T, Peigneux P, Van Bogaert P. Is sleep-related consolidation impaired in focal idiopathic epilepsies of childhood? A pilot study. Epilepsy Behav 2011;22:380–4.
Van Bogaert P, Urbain C, Galer S, Ligot N, Peigneux P, De Tiège X. Impact of focal interictal epileptiform discharges on behaviour and cognition in children. Neurophysiol Clin 2012;42:53–8.
Ventouras EM, Monoyiou EA, Ktonas PY, Paparrigopoulos T, Dikeos DG, Uzunoglu NK, et al. Sleep spindle detection using artificial neural networks trained with filtered time-domain EEG: a feasibility study. Comput Methods Programs Biomed 2005;78:191–207.
Wilson SB, Harner RN, Duffy FH, Tharp BR, Nuwer MR, Sperling MR. Spike detection. I. Correlation and reliability of human experts. Electroencephalogr Clin Neurophysiol 1996;98:186–98.
Wilson SB, Emerson R. Spike detection: a review and comparison of algorithms. Clin Neurophysiol 2002;113:1873–81.
Zeitlhofer J, Gruber G, Anderer P, Asenbaum S, Schimicek P, Saletu B. Topographic distribution of sleep spindles in young healthy subjects. J Sleep Res 1997;6:149–55.
Zygierewicz J, Blinowska KJ, Durka PJ, Szelenberger W, Niemcewicz S, Androsiuk W. High resolution study of sleep spindles. Clin Neurophysiol 1999;110:2136–47.