



Combining time-frequency and spatial information for the detection of sleep spindles

Christian O'Reilly^{1,2,3*}, Jonathan Godbout⁴, Julie Carrier^{5,6} and Jean-Marc Lina⁴

¹ Montreal Neurological Institute, McGill University, Montreal, QC, Canada

² Département de Psychiatrie, Université de Montréal, Montreal, QC, Canada

³ Dream and Nightmare Laboratory, Center for Advanced Research in Sleep Medicine, Hôpital du Sacré-Coeur, Montreal, QC, Canada

⁴ Laboratoire PhysNum, École de Technologie Supérieure, Centre de Recherches Mathématiques, Montreal, QC, Canada

⁵ Département de Psychologie, Université de Montréal, Montreal, QC, Canada

⁶ Chronobiology Laboratory, Center for Advanced Research in Sleep Medicine, Hôpital du Sacré-Coeur, Montreal, QC, Canada

Edited by:

Simon C. Warby, Stanford University, USA

Reviewed by:

Erin J. Wamsley, Furman University, USA

Errikos-Chaim Michael Ventouras, Technological Educational Institution of Athens, Greece

Róbert Bódizs, Semmelweis University, Hungary

***Correspondence:**

Christian O'Reilly, Montreal Neurological Institute, McGill University Montreal, H3A 2B4 QC, Canada

e-mail: christian.oreilly@mail.mcgill.ca

EEG sleep spindles are short (0.5–2.0 s) bursts of activity in the 11–16 Hz band occurring during non-rapid eye movement (NREM) sleep. This sporadic activity is thought to play a role in memory consolidation, brain plasticity, and protection of sleep integrity. Many automatic detectors have been proposed to assist or replace experts for sleep spindle scoring. However, these algorithms usually detect too many events making it difficult to achieve a good tradeoff between sensitivity (Se) and false detection rate (FDr). In this work, we propose a semi-automatic detector comprising a sensitivity phase based on well-established criteria followed by a specificity phase using spatial and spectral criteria. In the sensitivity phase, selected events are those which amplitude in the 10–16 Hz band and spectral ratio characteristics both reject a null hypothesis ($p < 0.1$) stating that the considered event is not a spindle. This null hypothesis is constructed from events occurring during rapid eye movement (REM) sleep epochs. In the specificity phase, a hierarchical clustering of the selected candidates is done based on events' frequency and spatial position along the anterior-posterior axis. Only events from the classes grouping most (at least 80%) spindles scored by an expert are kept. We obtain $Se = 93.2\%$ and $FDr = 93.0\%$ in the first phase and $Se = 85.4\%$ and $FDr = 86.2\%$ in the second phase. For these two phases, Matthew's correlation coefficients are respectively 0.228 and 0.324. Results suggest that spindles are defined by specific spatio-spectral properties and that automatic detection methods can be improved by considering these features.

Keywords: sleep spindles, detection, electroencephalography, time-frequency, hierarchical clustering, machine learning, pattern recognition, sleep

INTRODUCTION

EEG sleep spindles are short bursts of oscillatory activity in the 11–16 Hz frequency band during NREM sleep, especially in stage 2 sleep. This sporadic activity is a topic drawing increasingly more attention as it is thought to have an important role in the protection of sleep integrity and in the consolidation of new learning (Steriade, 2006; Dang-Vu et al., 2010; Fogel et al., 2012). Usually, the study of sleep spindles is time consuming due to the manual processing it requires. Aside from preprocessing steps such as sleep staging and artifact rejection, a polysomnographic expert has to manually identify hundreds of spindle occurrences hidden in whole-night EEG recordings, a tedious and error-prone task. Over the years, many automatic detectors have been proposed to assist or replace the experts in this task. These can be roughly split in two classes. The first one transforms the recorded signal in a new function—the *detection function*—whose amplitude is related to the probability of spindle activity. A simple threshold (or a set of thresholds) is applied to this function to decide on the presence or absence of spindle activity. This operation is typically followed by some additional criteria such as rejection of

small duration events, generally <500 ms to follow standard definitions of sleep spindle (Rechtschaffen and Kales, 1968; Iber et al., 2007). Many systems following this general approach have been proposed (e.g., Schimicek et al., 1994; Huupponen et al., 2007; Devuyst et al., 2011; Babadi et al., 2012). In the second class of detectors, EEG signals are segmented in a sequence of events (i.e., epochs that are potentially associated with spindle occurrences). For each event, a set of features is extracted to better synthesize its key characteristics. Then, two approaches can be used to classify these events as spindles or non-spindles: supervised (guided by pre-annotated spindles) or unsupervised (clustering techniques finding regular subsets of events and selecting subsets that are most likely to be associated with spindle activity). Here again, many systems have been proposed in the literature (e.g., Acir and Güzelis, 2004; Olbrich and Achermann, 2005; Ventouras et al., 2005; Sinha, 2008; Ahmed et al., 2009).

However, the detection of an important proportion of false positives is a persistent problem observed with these automated detectors when compared to expert scoring. This issue has often been hidden by reports of apparently highly specific systems

which large numbers of false positives were masked by the important asymmetry between spindle vs. non-spindle events (O'Reilly and Nielsen, 2013; O'Reilly and Nielsen, in revision). Looking at the false detection rate (instead of specificity) reveals this important weakness. In this context, achieving a satisfactory tradeoff between sensitivity (Se) and false detection rate (FDr) proved to be challenging.

In this work, we propose a two-step detector which aims to decrease the FDr by combining a *sensitivity phase* based on well-established criteria to a *specificity phase* using spatial and time-frequency criteria. This approach mixes both types of classification approaches previously described. In the *sensitivity phase*, putative events are first detected from the wavelet representation of the EEG recordings and then selected as those with large sigma index—a measure proposed by Huupponen et al. (2007) as a ratio of specific spectral bands—and high amplitude in the spindle frequency band. The threshold used in this selection process is based on the rejection of a null hypothesis ($p < 0.1$) stating that the considered event is not a spindle. The non-parametric model of the null hypothesis is constructed from events occurring in spindle-free epochs, e.g., in REM stage. In the *specificity phase*, hierarchical clustering of detected events is performed using the spectral and the topographical (anterior vs. posterior localization) properties of spindles. This spatio-spectral classification is motivated by evidences of a dichotomy in sleep spindles: one class occurs in frontal regions and has lower frequencies; another class is characterized by higher frequencies and a more centro-parietal topography (Werth et al., 1997; Zeitlhofer et al., 1997; Anderer et al., 2001; De Gennaro and Ferrara, 2003; Martin et al., 2013). Then, classes grouping a large proportion of events scored as spindles by an expert are selected. In this phase, the detector tries to reject as many false positives as possible—hence effectively biasing the detection threshold toward specificity—without rejecting too many true positives. Interestingly, parameters for such clustering can be learned from a small sample of expert detections and then be generalized automatically to the whole night.

MATERIALS AND METHODS

PREPROCESSING

Signal mixture

A first preprocessing step consists in locally averaging the EEG signals to obtain one highly informative signal out of the N_c EEG channels available. This is made possible by the fact that the spindle activity is generally relatively synchronous across the scalp, with maximal apparition delays between sensors generally below 25 ms (O'Reilly and Nielsen, 2014b). We consider the following virtual channel:

$$s^{(m)} = m^T S \quad (1)$$

where m is a vector of N_c components specifying weights associated with every channel of this mixture. This vector is normalized with a L_1 norm (i.e., elements sum to unity) and defines what we call a *montage*. The S matrix has a dimension $N_c \times N_t$ and is obtained by simply stacking together the signals from the N_c channels, each one containing N_t time samples.

Time-frequency representation

Spindle activity was assessed in the time-frequency plane using the Continuous Wavelet Transform (CWT). This transform is defined as follows:

$$W(a, b) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \Psi_{\beta, \gamma}^*(a\omega) S(\omega) e^{i\omega b} d\omega \quad (2)$$

with a and b being parameters associated respectively with scale (i.e., inverse of frequency) and time. $S(\omega) = \int_{-\infty}^{+\infty} s(t) e^{-i\omega t} dt$ is the Fourier transform of the signal $s^{(m)}(t)$, $*$ indicates the complex conjugate, and $\Psi_{\beta, \gamma}(\omega)$ is a wavelet in the frequency domain. For this study, we used the Morse wavelet (Lilly and Olhede, 2009, 2010):

$$\Psi_{\beta, \gamma}(\omega) = H(\omega) c_{\beta, \gamma} \omega^\beta e^{-\omega^\gamma} \quad (3)$$

with $c_{\beta, \gamma}$ being an irrelevant normalization factor and $H(\omega)$ being the Heaviside function (null everywhere but for $\omega \geq 0$ where it is equal to 1). We set $\gamma = 20$ and $\beta = 10$. These values were found to provide the best tradeoff between time and frequency resolution for sleep spindle representation. See **Figure 1** for an example of time-frequency representation of a sleep spindle using this transform.

Wavelet ridge and temporal markers in the time-frequency plane

Computing (2) produces a matrix $W^{(m)}$ of CWT coefficients $w_{i,j}^{(m)}$ at time t_j and frequency $f_i = f_0/a_i$, f_0 being the main frequency of the wavelet $\Psi_{\beta, \gamma}(\omega)$. For each time sample, we considered the local maximal amplitude along frequencies of the spindle spectral band. We then computed the time course of those wavelet maxima, i.e.:

$$d(t_j) = \max_{1 \leq i \leq N_f} |w_{i,j}^{(m)}| \quad (4)$$

Named *ridge* (Delprat et al., 1992), this piecewise continuous path across the time-frequency map $W^{(m)}$ quantifies the power of instantaneous frequency in the signal. To be sensitive to the spindle frequency band, it was computed using frequencies sampled from 10 to 16 Hz with 0.1 Hz resolution, resulting in $N_f = 61$ frequencies per time sample.

To allow for a parsimonious assessment of spindle features, the ridge was first marked according to the local maxima of the $d(t_j)$ function:

$$t^{max} = \{t_j \in t : \dot{d}(t_j) \dot{d}(t_{j+1}) < 0 \text{ and } \dot{d}(t_j) > 0\} \quad (5)$$

with \dot{d} being the time derivative of d . These maxima are considered as time markers for the putative events (i.e., one event is counted for each item in the t^{max} set) in the time-frequency plane.

Feature computation for the sensitivity phase

Two features are computed for signal detection in the *sensitivity phase*. The first one is the ridge amplitude at the maxima: $\tilde{x}_n^{amp} = d(t_n^{max})$, with $n = 1, 2, \dots, N$ and N being the number of elements in the t^{max} set. The second feature is a spectral sigma ratio similar to what was proposed by Huupponen et al. (2007) but computed using the modulus of the activity in the time-frequency space ($|W^{(m)}|$) in the 4–40 Hz range:

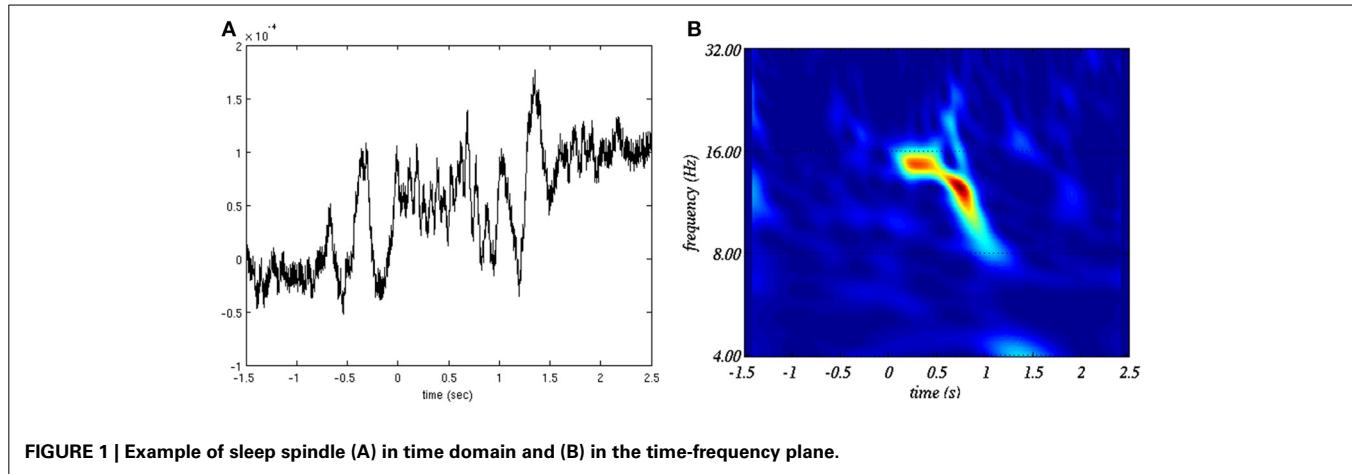


FIGURE 1 | Example of sleep spindle (A) in time domain and (B) in the time-frequency plane.

$$\tilde{x}_n^{\text{sigma}} = \frac{2a_\sigma}{a_\alpha + a_\beta} \\ = \frac{2\max\{|W|_{[10.5 - 16 \text{ Hz}; t_n^{\max}]}|\}}{\text{mean}\{|W|_{[4 - 10 \text{ Hz}; t_n^{\max}]}|\} + \text{mean}\{|W|_{[20 - 40 \text{ Hz}; t_n^{\max}]}|\}} \quad (6)$$

This index increases with narrow band activity having a peak in the 10.5–16 Hz band. Compared to the root-mean-square amplitude of the activity in the sigma band—a measure often used for spindle detection (e.g., Schimicek et al., 1994; Molle et al., 2002; Clemens et al., 2005; Schabus et al., 2007; Warby et al., 2014)—it has the advantage of penalizing muscular artifacts (20–40 Hz) and signs of arousal (4–10 Hz). It might, however, be adversely impacted by the increase of theta and beta activity associated with sleep spindles (Vyazovskiy et al., 2004). This measure was chosen since it represents a state-of-the-art approach for spindle detection and it has shown to perform reasonably well in previous studies (Huupponen et al., 2007, 2008; Sheng-Fu et al., 2012; O'Reilly and Nielsen, in revision). An example of corresponding values is shown in Figure 2.

Feature computation for the specificity phase

Two other features defined between 0 and 1 are computed in the *specificity phase*. The first feature assesses the main frequency mode of a putative spindle n :

$$\tilde{x}_n^{\text{freq}} = \left. \frac{f_n^{\max} - f_1}{f_{N_f} - f_1} \right|_{f_1=10, f_{N_f}=16} \quad (7)$$

where $f_n^{\max} = f_i$ with $i = \operatorname{argmax}_{1 \leq i \leq N_f} |w_{i,j}^{(m)}|$ and j is such that $t_j = t_n^{\max}$. Figure 3 summarizes important concepts introduced so far.

The second feature captures the location of spindle activity along the anteroposterior axis of the scalp. To compute this value, we consider the first principal component (PC) of a 500 ms window centered around t_n^{\max} . This spatial eigenvector represents a normalized topography over the channels, and its components correspond to the relative weight for each channel. Being the

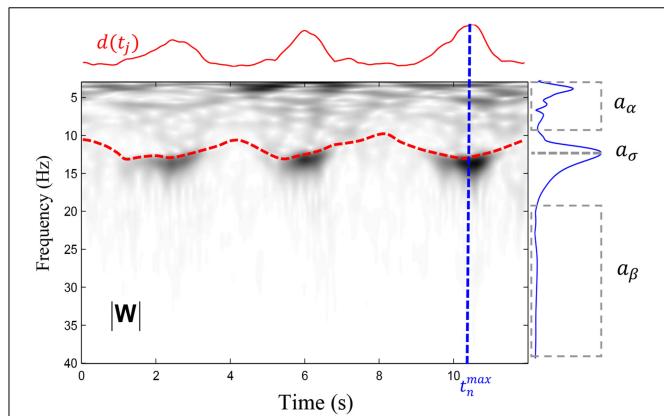
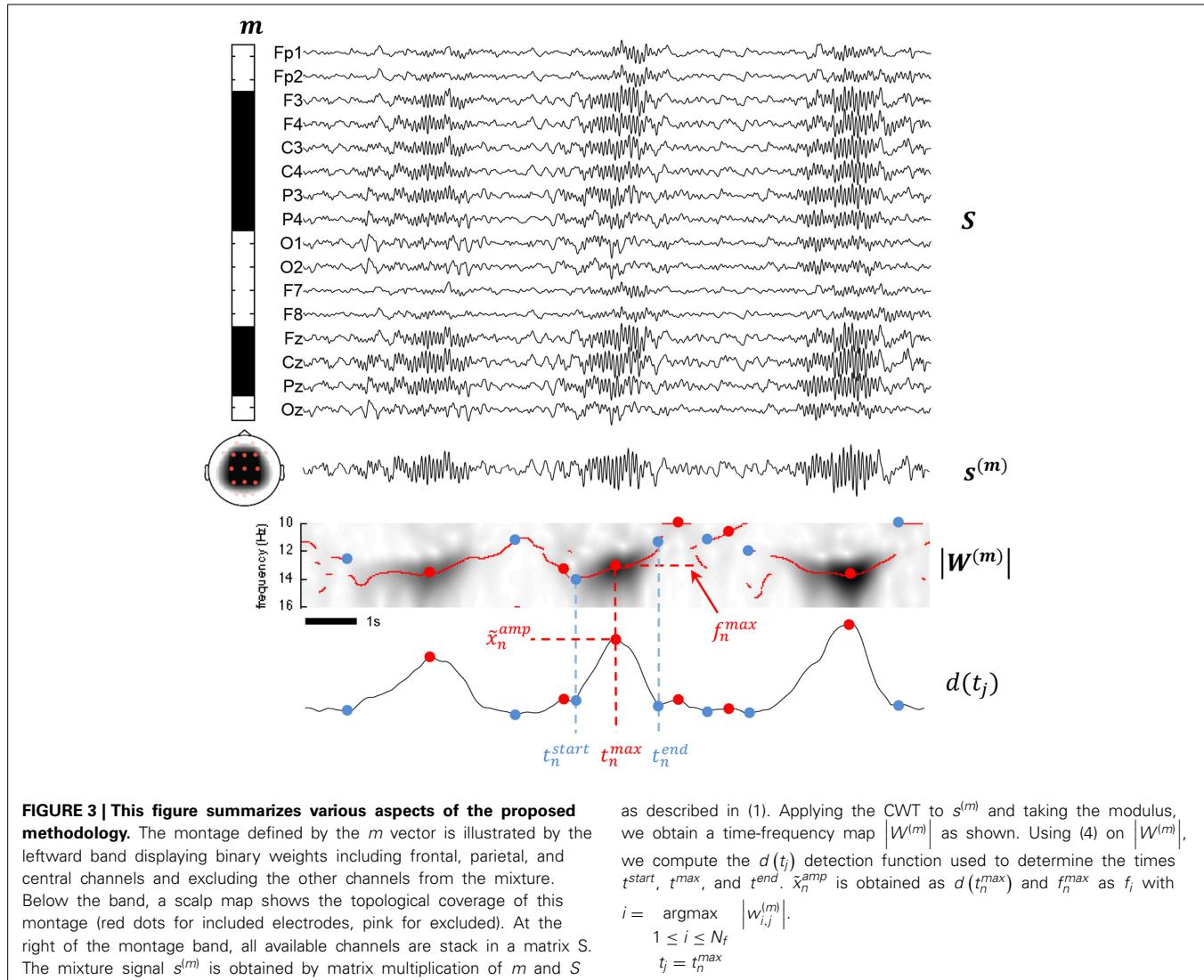


FIGURE 2 | Illustration of the variables entering in the computation of $\tilde{x}_n^{\text{sigma}}$ as defined in (6). The time-frequency plane ($|W|$), in gray levels, is represented by the modulus of wavelet coefficients computed for frequencies between 4 and 40 Hz with 0.1 Hz resolution. The red dashed line shows the ridge computed in the 10–16 Hz band whereas the solid red line shows the $d(t_j)$ function defined in (4). A maximum has been detected at time t_n^{\max} and the variation of the coefficients (i.e., the instantaneus spectrum) at the time t_n^{\max} is shown by the solid blue line. From this spectrum, we extract the average amplitude in two intervals (shown by dashed boxes comprising respectively the low and the high frequencies) to obtain $a_\alpha = \text{mean}\{|W|_{[4 - 10 \text{ Hz}; t_n^{\max}]}|\}$ and $a_\beta = \text{mean}\{|W|_{[20 - 40 \text{ Hz}; t_n^{\max}]}|\}$. We further take the maximal amplitude in the spindle band to obtain $a_\sigma = \max\{|W|_{[10.5 - 16 \text{ Hz}; t_n^{\max}]}|\}$.

first PC, this topography picks the larger variability of the multivariate signal over the analyzed window. Then, the position of the channel with maximal weight can be considered representative of the scalp localization of the event centered around t_n^{\max} . Channel positions are specified as (x_n, y_n) coordinates in the 10-5 system (Oostenveld and Praamstra, 2001) mapped to a flat top view of the scalp as specified in the EEG1005.lay montage file of the FieldTrip software (Oostenveld et al., 2011). Only the y_n value is used for spindle detection given the observation of different types of spindles in relation with their anteroposterior position (Dehghani et al., 2011; Martin et al., 2013; O'Reilly and Nielsen, 2014b). The feature for localization along the medial axis is defined as:



$$\tilde{x}_n^{\text{med}} = y_n + 0.5 \quad (8)$$

such that it is normalized to the $[0, 1]$ range.

Threshold computation

Two extra quantities are used to set thresholds needed by the algorithm. Both are derived from information related to the timing and space location of the spindles given *a priori* by a gold standard, typically an expert. The first one is a sleep stage related feature, $\tilde{x}_n^{\text{stage}}$, which value is an integer between 0 and 5 (0: awake; 1: NREM1; 2:NREM2, 3:NREM3; 4:NREM4; 5:REM). This value is defined on the current sleep stage at the moment of t_n^{max} . The second feature indicates whether an event occurred during a time window associated with a spindle also visually identified by an expert on channels Fz, Cz, or Pz. That is, $\tilde{x}_n^{\text{expert}} = 1$ if t_n^{max} is co-occurring with a spindle labeled on any of these three channels. Otherwise, a zero value is attributed.

It is worth highlighting that the proposed detection technique rests on “point” features (i.e., features evaluated at a given point

in time) and not on features computed on time windows. Thus, the detector set instantaneous markers for sleep spindles without explicit duration.

SENSITIVITY PHASE

The goal of this phase is to detect as many true spindles as possible, missing only a small proportion, at the cost of a relatively high amount of false positives. In this *sensitivity phase*, we test the null-hypothesis stating that $\tilde{x}_n^{\text{sensitive}} = [\tilde{x}_n^{\text{amp}} \tilde{x}_n^{\text{sigma}}]$ is not associated with a spindle. For this assessment, a sample of the null-hypothesis, i.e., non-spindle events, is built from $\tilde{x}_n^{\text{sensitive}}$ of all events with $\tilde{x}_n^{\text{stage}} = 5$. Although, it has been proposed that isolated spindles can occur in REM (Rechtschaffen and Kales, 1968), this is controversial. In the same line of thought, sleep spindles could also be present in transition pages marked as REM but containing some proportion of NREM sleep. Nevertheless, presence of spindles in pages marked as REM should be rare and should therefore have little impact on our statistics.

Decision thresholds are computed separately for both features. This implicitly postulate statistical independence, a reasonable hypothesis given the relatively low correlation reported (about 0.25 according to Huupponen et al., 2007) between these two features. Two thresholds— τ^{amp} and τ^{sigma} —are obtained as the value of \tilde{x}^{amp} and \tilde{x}^{sigma} at the $(1 - \alpha)$ percentile of the distribution of the non-spindle events. That is, we compute thresholds that should fail to reject at most a proportion α of false positives. As discussed in O'Reilly and Nielsen (2014a), such an approach sets the expected false detection rate (FDr; complete definition in **Table 1**, Section Performance Assessment) to:

$$FDr = \frac{\alpha\kappa}{P\%} \quad (9)$$

with κ being the proportion of false positives in the tested sample and $P\%$ the proportion of the tested sample not rejected by this threshold. Although we cannot compute the value for the FDr because we lack an estimate for κ , we can obtain an upper bound \widetilde{FDr} using:

$$\widetilde{FDr} = \frac{\alpha}{P\%} \quad (10)$$

With these thresholds, we can now define a subset X of selected candidates as:

$$X = \{X_m\} = \left\{ \tilde{x}_n \in \tilde{X} : \tilde{x}^{amp} \geq \tau^{amp} \text{ and } \tilde{x}^{sigma} \geq \tau^{sigma} \right\} \quad (11)$$

SPECIFICITY PHASE

Previous selection of events is used as input to the *specificity phase* which tries to keep only selected candidates corresponding

Table 1 | Definition of performance metrics used in this paper.

Measure	Formula
Sensitivity (Se)	$S_e = \frac{N_{TP}}{N_{FN} + N_{TP}}$
Specificity (Sp)	$S_p = \frac{N_{TN}}{N_{FP} + N_{TN}}$
False positive rate (FPr)	$FPr = \frac{N_{FP}}{N_{FP} + N_{TN}} = 1 - S_p$
False Discovery Rate (FDr)	$FDr = \frac{N_{FP}}{N_{FP} + N_{TP}}$
False positive proportion (FPr)	$FPr = \frac{N_{FP}}{N_{FN} + N_{TP}}$
Matthew's correlation coefficient (MCC)	$MCC = \frac{N_{TP} * N_{TN} - N_{FP} * N_{FN}}{\sqrt{(N_{TP} + N_{FP})(N_{TP} + N_{FN})(N_{TN} + N_{FP})(N_{TN} + N_{FN})}}$
F1 score	$F_1 = \frac{2N_{TP}}{2N_{TP} + N_{FP} + N_{FN}}$
Cohen κ	$\kappa = \frac{X - (N_{TN} + N_{TP})N_e}{X - N_e^2}$ with $X = (N_{TP} + N_{FP})(N_{TP} + N_{FN}) + (N_{FP} + N_{TN})(N_{FN} + N_{TN})$

with spindles, as identified by an expert. A partition of selected candidates in homogenous classes of events is performed using the *ascending hierarchical classification* (AHC) algorithm (Timm, 2002). This technique starts with every item of X being considered as a singleton class and iteratively regroups together the two most similar classes until only one class regrouping all items is left. The outcome of such a process can be represented as tree graph called a *dendrogram* (see **Figure 4** for an example). The AHC algorithm is defined by a *metric* and a *linkage criterion*. The former defines how we assess the distance between two items whereas the latter do the same for two classes of items. In our case, we used the Euclidean distance as metric:

$$d(x_n, x_m) = \sqrt{\sum_i (x_n^{(i)} - x_m^{(i)})^2} \quad (12)$$

where the i index iterates over elements of x_n and x_m vectors. For linkage criterion, we used the average distance $d(x_n, x_m)$ between items of two classes $A, B \in X$ defined as:

$$L(A, B) = \frac{1}{|A| |B|} \sum_{x_a \in A} \sum_{x_b \in B} d(x_a, x_b) \quad (13)$$

with $|A|$ and $|B|$ standing for the cardinality of classes A and B , respectively. **Figure 4** illustrates the use of the AHC algorithm.

The final clustering is obtained by cutting the dendrogram at the maximal value of inter-class dissimilarity subject to the inequality:

$$\frac{|B|}{|A|} \geq r \quad (14)$$

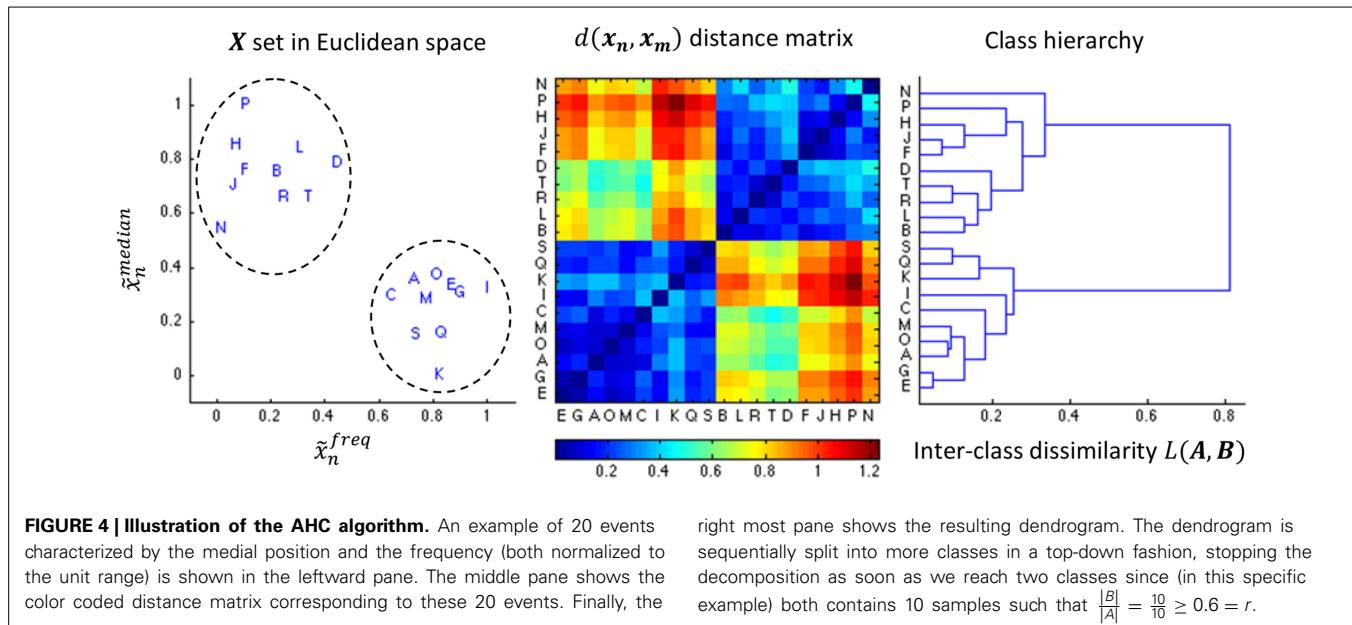
with A and B being respectively the largest and second largest classes. This criterion tends to favor homogeneity of class sizes. A value $r = 0.6$ was chosen in this study because it was found to be a good tradeoff between accepting only equally sized classes (i.e., $r = 1.0$) and allowing much disparate classes such as one big cluster associated with a very small outlier class (i.e., $r \rightarrow 0.0$). Classes obtained that way are then sorted in descending order according to their number of *expert events* (i.e., events scored as spindles by the expert). For the *specific detection*, only events belonging to the first N_{class} classes are labeled as spindles, with N_{class} being the smallest number of classes grouping at least 80% of the expert events.

PERFORMANCE ASSESSMENT

For assessing performances, we used a terminology borrowed from confusion matrices. Four classification outcomes can be encountered in the dual-class problem considered here: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). If we consider a variable $\tilde{x}_n^{selected}$ which takes 1 when the n th event is designated as a spindle by the algorithm and otherwise takes 0, these four cases are obtained as follow:

$$TP \Leftrightarrow \tilde{x}_n^{selected} = 1 \wedge \tilde{x}_n^{expert} = 1 \quad (15)$$

$$TN \Leftrightarrow \tilde{x}_n^{selected} = 0 \wedge \tilde{x}_n^{expert} = 0 \quad (16)$$



$$FP \Leftrightarrow \tilde{x}_n^{selected} = 1 \wedge \tilde{x}_n^{expert} = 0 \quad (17)$$

$$FN \Leftrightarrow \tilde{x}_n^{selected} = 0 \wedge \tilde{x}_n^{expert} = 1 \quad (18)$$

Counts of each outcome are labeled respectively N_{TP} , N_{TN} , N_{FP} , and N_{FN} and are the constitutive elements of the metrics used to score our algorithm (see **Table 1**). Here, we are measuring agreement using a “by-event” approach (Warby et al., 2014) where an agreement is marked if and only if a specific point (i.e., the local maximum of the ridge) is within one of the spindle windows scored by the expert. The total number of events N_e (i.e., $N_e = N_{TP} + N_{FP} + N_{FN} + N_{TN}$) is defined by the segmentation described in section Wavelet ridge and temporal markers in the time-frequency plane.

IMPLEMENTATION

The detector has been implemented as a “process” in Brainstorm (Tadel et al., 2011). The source code is available from the corresponding author.

SAMPLE

We tested our algorithm on polysomnograms recorded in a hospital-based sleep laboratory from 9 (7 women, 2 men) young (mean \pm standard deviation: 22.6 ± 2.4 years old) and healthy subjects. Recording was performed at 256 Hz using a Vita-port-3 System (low-passed at 70 Hz with 1-s time constant) and the data were recorded using the Columbus software from TEMEC Instruments (Kerkrade, Netherlands). We used a standard 10–20 EEG sensor grid (C3, C4, Cz, F3, F4, Fz, F7, F8, O1, O2, Oz, P3, P4, Pz, T3, T4, T5, T6, Fp1, Fp2) with a 10 kΩ ear-linked reference as well as bipolar chin EMG, ECG, and EOG. Sleep stages were scored by a certified polysomnographer with 15 years of experience according to modified rules of Rechtschaffen and Kales (1968) adapted for 20-s epochs. Muscle artifacts were automatically detected (Brunner et al., 1996) and visually confirmed. Sleep

spindles were scored by the same expert on Fz, Cz, and Pz channels in NREM sleep epochs. Spindle scoring was performed on raw signals according to the rules of the AASM (Iber et al., 2007). Sleep stage distribution per subject (Table S1) as well as number of spindles scored per derivation per subject (Table S2) are provided as Supplementary Documents.

Every recording was sanctioned by the ethics review board of the Hôpital du Sacré-Coeur de Montréal and participants gave informed consent.

RESULTS

SENSITIVE DETECTION

Montage selection

We tested six different montages to study their effect on the sensitive detection: m_1 corresponds to frontal channels Fp1, Fp2, F7, and F8; m_2 to occipital channels O1, O2, and Oz; m_3 to channels F3, F4, C3, C4, P3, P4, Fz, Cz, and Pz; m_4 , m_5 , and m_6 to only Fz, Cz, and Pz, respectively. To avoid biasing toward some of these selected channels, we used equal weights for every channel of the montages (i.e., weights equal to $1/N_i$ where N_i equals the number of channels included in the montage).

Performance of the sensitive detection depends on the capacity of the chosen montage to discriminate between the sleep spindles (in red in **Figure 5**) and the non-spindle events (in black). For example, the small overlap between these two sets of curves in m_3 indicates a good discriminative power. We note that some simpler montages (e.g., montages m_5 and m_6 using only Cz and Pz, respectively) also show similarly good performances. Lower discrimination is obtained using only Fz (m_4) or using in general only frontal and prefrontal (m_1) or occipital (m_2) scalp channels. Results presented subsequently are obtained using m_3 .

Performance evaluation

Results from a receiver operating characteristic (ROC) curve analysis are presented in **Figure 6**.

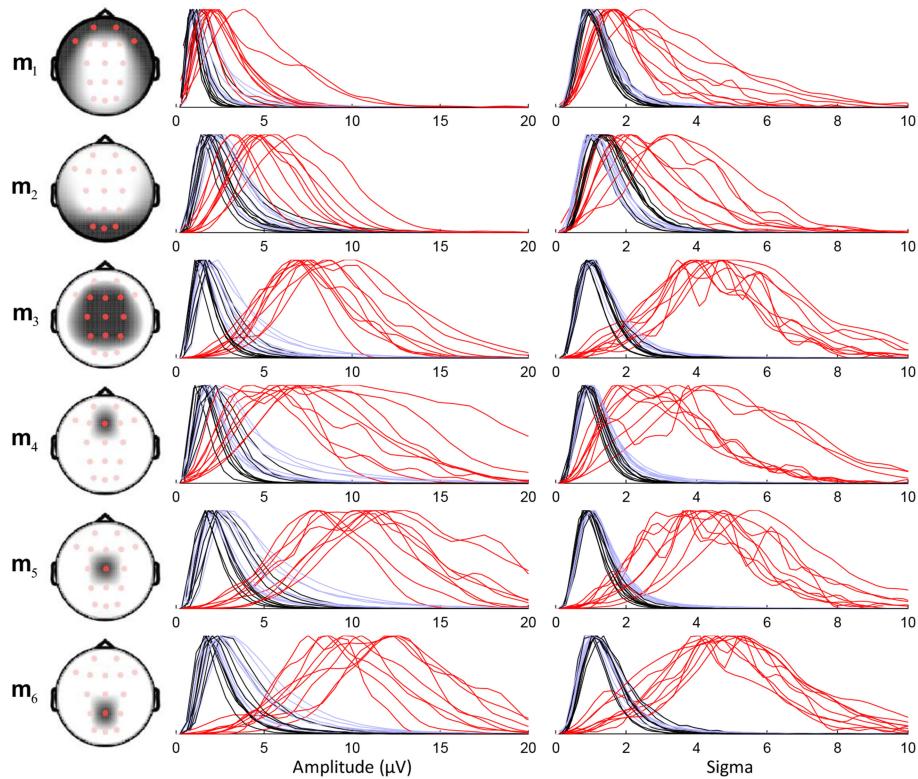


FIGURE 5 | Density plots with normalized maximal amplitude for the distribution of sensitive detection features (amplitude and sigma index) for six different montages. The head drawings show the topographical coverage of each montage. Feature distributions are plotted

as three separate lines per subject: red for detected spindles corresponding with expert scoring, black for events scored as spindles by the algorithm in REM epochs only, and light blue for detected events during REM and NREM sleep.

Averages and standard deviations (SD) of the performance statistics are reported in **Table 2** for the conditions $S_e = S_p$ and $\alpha = 0.1$, with the second condition focusing slightly more on sensitivity. One should note that this table do not report specificity since this statistic has little value in evaluating spindle detectors because it systematically takes high values given the small proportion of positive to negative cases (i.e., spindle vs. non-spindle) (O'Reilly and Nielsen, 2013). For the same reason, the reader should be cautious in interpreting the ROC curves in **Figure 6** since only the portion with large specificity is meaningful. Lower specificity are associated with prohibitively high FDr, something not visible in ROC curve (O'Reilly and Nielsen, 2013).

SPECIFIC DETECTION

Figure 7A shows the proportion of spindles scored by the expert (green) and the proportion of total events (black) contained in the four classes produced by the clustering algorithm. These classes are sorted in decreasing number of expert events. Lines of lighter color are used for individual subjects while darker lines are used for the median across subjects. As specified in the Materials and Methods section, events selected by the specificity phase are those belonging to the first classes regrouping at least 80% of the expert events. As can be seen, only one class is required to reach

this criterion. Except for S4, using only one class, we can keep more than 80% of the expert events while keeping about only 50% of the total number of events initially selected in the previous sensitivity phase. In **Figure 7B**, classification performances obtained with this criterion (white bars) are compared to the performance obtained before the application of this criterion (black bars).

It should be noted that results of **Figure 7** are obtained using all available expert scoring. This is in average 390 spindles per subject. We also tested whether the proposed algorithm could be used with a reduced number of sleep spindles sampled by the expert. Hence, bootstrapping over 500 repetitions has been performed using randomly selected subsets of 1, 2, 4, 8, 16, 32, 64, and 128 scored events. **Figure 8** shows the differential (partial minus exhaustive scoring) in sensitivity and specificity. Subject S4 was excluded from this analysis because the unusual clustering in four equal size classes for this subject produced unstable results when using small subsets of expert scorings. As can be seen, the performances are not significantly degraded by partial scoring using about 16 or 32 spindles visually scored by an expert.

CHARACTERISTICS OF DETECTED SPINDLES

This section compares automatically detected spindles with those identified by the expert.

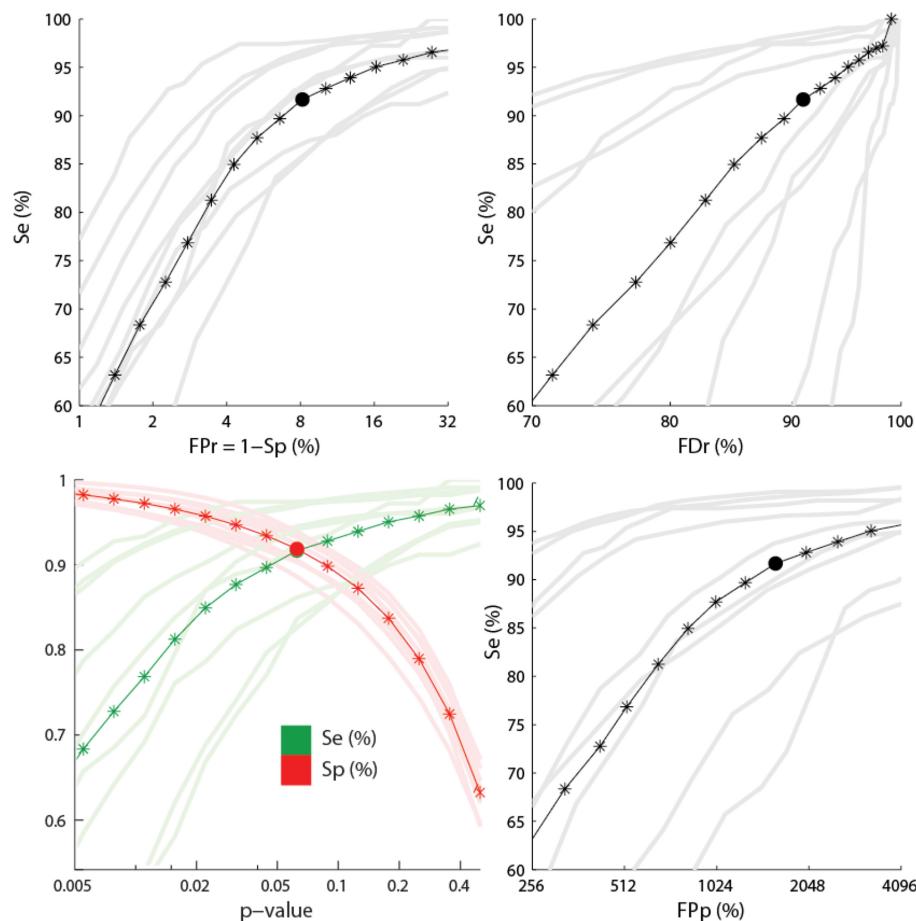


FIGURE 6 | Performance for the sensitive detection. The lower left panel shows the sensitivity (green) and the specificity (red) as function of the p -value used for feature thresholding. The three other panels show ROC-like curves using FPr , FDr , and FPp . Dark colors show average statistics whereas

light colors correspond to performances for every subject. Stars indicate the positions where the statistics have been computed. Segments between stars are obtained by linear interpolation. Filled circles show points where $Se = Sp$. The x-axes are plotted using a logarithmic scale.

Table 2 | Average \pm SD value for performance statistics when $Se = Sp$ and when $\alpha = 0.1$ for the sensitive and the specific phase.

Mesure	Sensitive phase		Specific phase
	$Se = Sp$	$\alpha = 0.1$	$\alpha = 0.1$
Se	$92.1 \pm 3.0\%$	$93.2 \pm 4.8\%$	$85.4 \pm 7.4\%$
FPr	$7.9 \pm 2.7\%$	$11.0 \pm 2.1\%$	$4.5 \pm 1.7\%$
FDr	$89.1 \pm 8.3\%$	$93.0 \pm 4.2\%$	$86.2 \pm 6.1\%$
FPp	$1779.6 \pm 1936.6\%$	$2177.1 \pm 2110.3\%$	$730.3 \pm 598.9\%$
MCC	0.28 ± 0.13	0.23 ± 0.08	0.32 ± 0.08
$F1$	0.19 ± 0.13	0.13 ± 0.07	0.23 ± 0.09
κ	0.17 ± 0.13	0.11 ± 0.06	0.22 ± 0.09

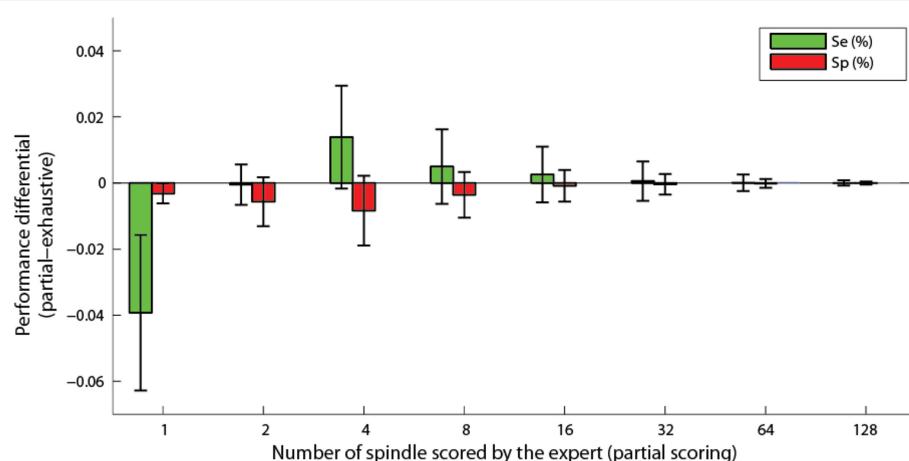
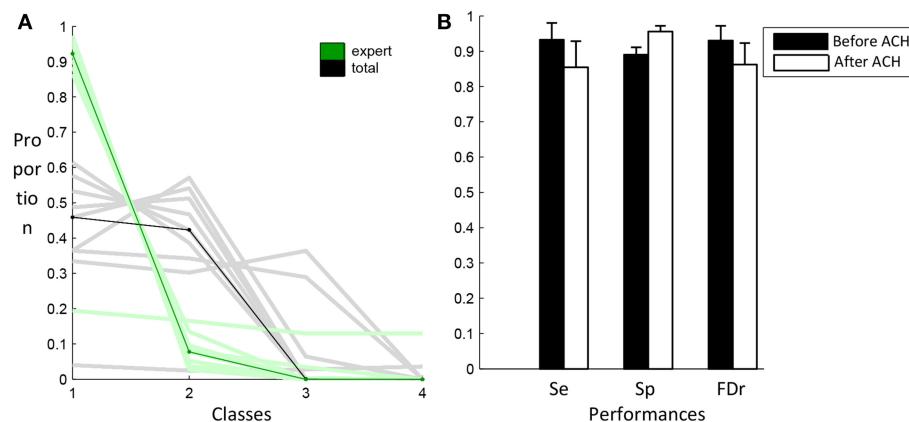
Frequency and medial position

Figure 9 shows the joint distribution of \tilde{x}_n^{freq} and \tilde{x}_n^{med} . The latter value varies between 0.15 (occipital) and 0.9 (pre-frontal). In general, distributions of features after the sensitivity phase suggest two classes of events, although the frontier separating these classes is blurry and varies from subject to subject. From the literature,

we would expect a fast (higher frequency) centro-parietal ($0.35 < \tilde{x}_n^{med} < 0.5$) class and a slow (lower frequency) frontal ($\tilde{x}_n^{med} > 0.5$) class. This behavior is observed for subjects S1, S3, S4, and S9, and to a lesser extent for S5 and S6. In S2 and S8, we do observe fast and slow classes, but both in centro-parietal region. In S7, the slow class is located in occipital region ($\tilde{x}_n^{med} = 0.2$) suggesting alpha rhythm contamination. Actually, most spindles scored by the expert tend to be in the fast centro-parietal class. Spindles automatically scored after the specificity phase follow this trend (comparison of results in second and third column of Figure 9).

Average spindle

Figure 10 shows the grand average for spindles scored by the expert, spindles selected by the sensitive detection, and events accepted during sensitivity phase but rejected by the specificity phase. In Figure 10A, the joint distribution for the frequency and the medial position is shown. In Figure 10B, the average signal for each channel is shown using a 5-s window centered around t_n^{max} . Averages are first computed within subjects and then between subjects. At each level, signals are time-aligned by maximizing the



cross-correlation of the central 500 ms of activity in the 10–16 Hz band. **Figure 10C** shows the first principal component (i.e., the component with the highest variability) computed on the central 500 ms window of the between-subjects averaged signal (band-passed in the 10–16 Hz band with a 5th order Butterworth filter). Finally, **Figure 10D** shows the time-frequency plot computed using the CWT (Morse wavelets with $\gamma = 20$ and $\beta = 20$) of the between-subjects average signal using the montage specified by the topographic vector of the first principal component (i.e., as shown in panel C).

Topographies in panel C and joint distributions in panel A both tend to support the existence of two classes of events with fast (13–14 Hz) centro-parietal activity and a slow (10–12 Hz) more diffuse activity generally located in more frontal areas. The expert visually scored mainly the first class and so did our specific selection. Spindles are shown to be in phase with a ~ 1 Hz component, reproducing the observations about slow wave/spindle

phase-amplitude coupling previously reported (Molle et al., 2002; Kokkinos and Kostopoulos, 2011).

Spindles across sleep cycles

Figure 11 shows how the proportion of spindles in each of the fours first sleep cycles evolves for 1) events selected by the expert, 2) events selected by the specific detection, 3) events rejected by the specific detection. Sleep cycles were defined according to Aeschbach and Borbely (1993): one cycle is a sequence of a NREM period followed by a REM period. The NREM period starts at the first epoch of NREM sleep and terminates at the first REM epoch. The REM period terminates only if the next 15 min are free of REM epochs. At least four cycles were present in every subject. As can be seen, in both expert scoring and detector expert class, spindles show a similar trend with an increasing density from the beginning to the end of the night. The non-expert class shows an inverse tendency.

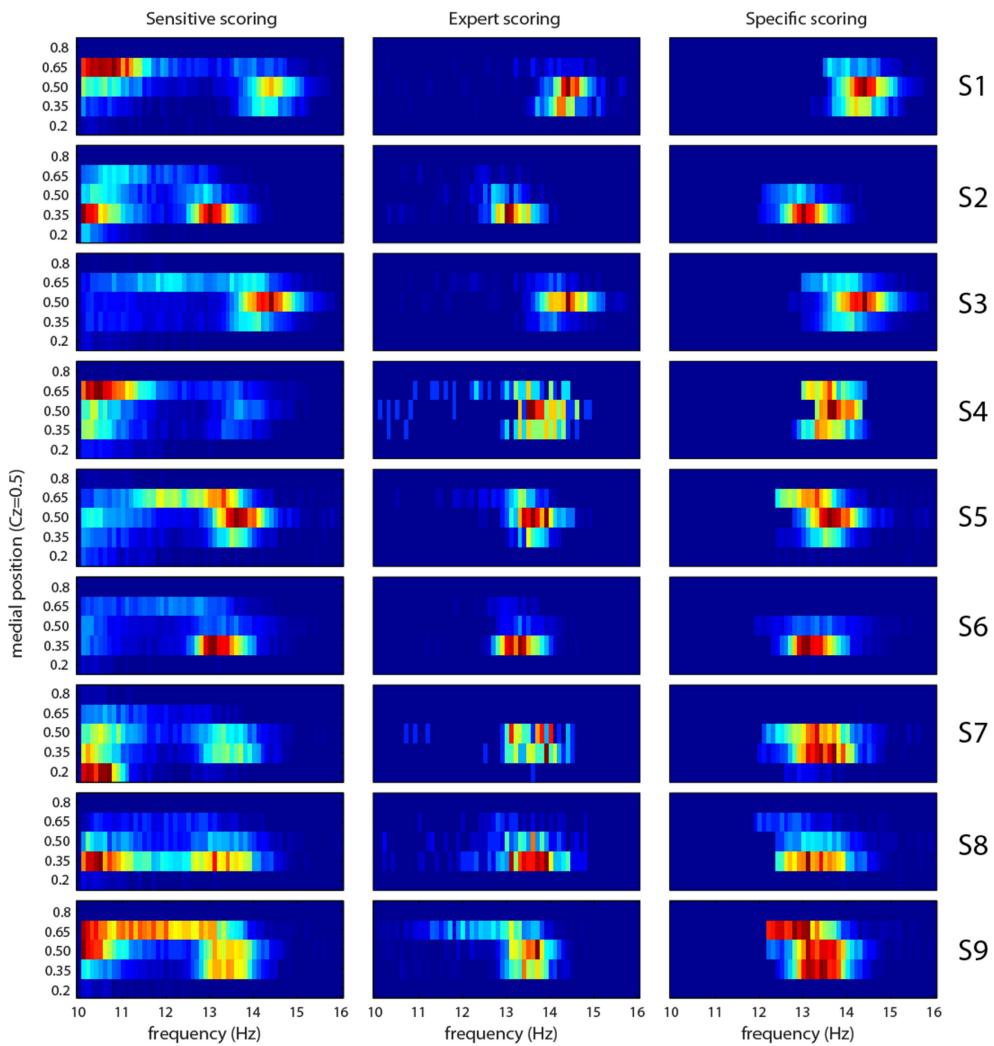


FIGURE 9 | Joint distribution of frequency and medial position (0.15 is occipital and 0.9 is prefrontal) features for each subject (S1–S9 rows) and for sensitive detection (first column), expert scoring

(second column), and specific detection (third column). Cold colors (toward blue) represent small amplitude, hot colors (toward red) large amplitudes.

DISCUSSION

The goal of this study was to tackle the problem of high false detection rates in sleep spindle scoring. The strategy adopted was to split the problem in two steps, a sensitive detection (unsupervised) and a specific detection (supervised). In the following sections, we discuss various aspects of our method and results.

Detection Montage

The approach described in section Montage selection provides the possibility of compressing a multivariate signal (coming from different channels) into a univariate signal using a specific montage. In this study, based on standard definition of spindles (Rechtschaffen and Kales, 1968; Iber et al., 2007) and on the current knowledge on spindle topography, we favored a montage weighting equally frontal (F3, Fz et Fz), central (C3, Cz, C4) and parietal (P3, Pz, P4) channels and excluding the others. This

montage failed to show a clear superiority compared to montages using single channels (e.g., Cz or Pz). One should note, however, that our gold standard (i.e., expert scoring) assessed spindles only on Fz, Cz, and Pz, a fact that could have contributed in favoring montage using only these channels. Also, further work is needed to confirm whether an improved detection can be achieved by tailoring more accurately the montage. Nonetheless, the approach has interesting applications for future developments as it provides a great flexibility to apply arbitrary montage to EEG signals, as shown for computation of Figure 10D.

ADAPTIVE SEGMENTATION AND TIME-FREQUENCY REPRESENTATION

In our method, we proposed an adaptive segmentation that split the whole night in a sequence of contiguous events. This segmentation was performed using the ridge of the continuous wavelet transform of the time series for the chosen montage.

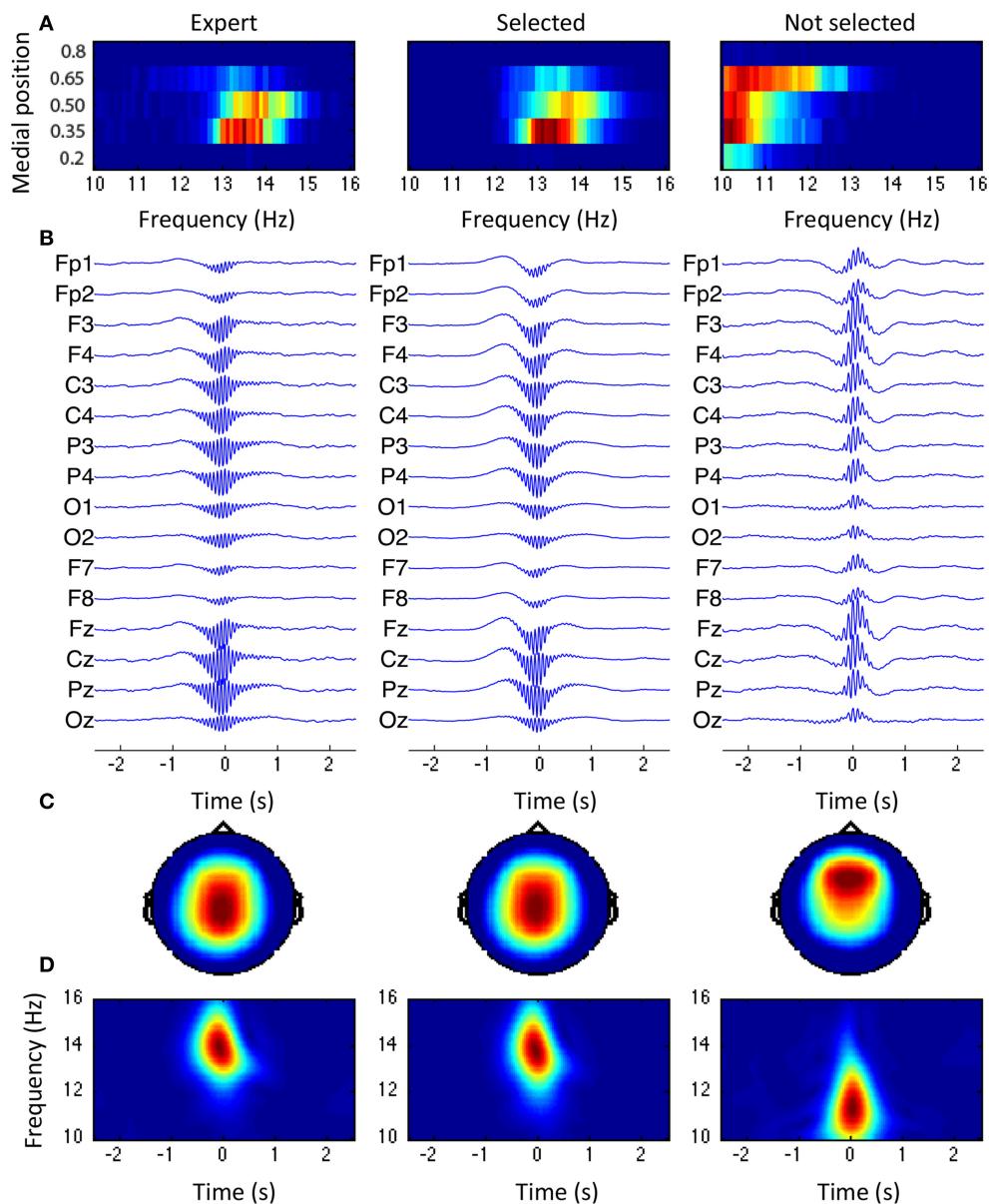


FIGURE 10 | Grand average across subjects. From left to right: spindle scored by the expert, events from the selected class, and events from the rejected classes. **(A)** Joint distribution of the frequency and the medial

position. **(B)** Average spindle. **(C)** Topography of the first principal component obtained through PCA. **(D)** Time-frequency plane for the average spindle, using the montage specified by the topography shown in **(C)**.

For simplicity, and because it provides a good tradeoff between temporal and spectral resolution, the Morse wavelet was used. Its parameters ($\gamma = 20$ and $\beta = 10$) were chosen using visual inspection. One should note, however, that β is the most sensitive parameter. Large values tend to over-smooth and reduce the temporal resolution whereas too small values tend to under-smooth resulting in appearance of amplitude modulation of the time-frequency plane at higher frequencies (closer to the spindle band). Higher values for β might be adequate in more noisy environments—such as for EEG signals collected during functional magnetic resonance imaging (fMRI)—to shift the tradeoff between temporal resolution and noise rejection.

CHARACTERISTICS OF SELECTED SPINDLES

Most spindles scored by the expert were rapid (>13 Hz) and in the centro-parietal region of the scalp. The usual slow/fast dichotomy was not observed (see Figure 9). This result could be attributed to a specific detection bias of this expert and needs to be corroborated by looking at scorings from other experts. Notably, however, this slow-fast dichotomy has mostly been reported in studies using automated spindle detections. Since experts score spindles with enough amplitude to be visually discriminated from background activity, part of the false positives could also be false negatives from experts. Also, in *post-hoc* investigations, we noted that spindles detected in Fz are simultaneously detected in

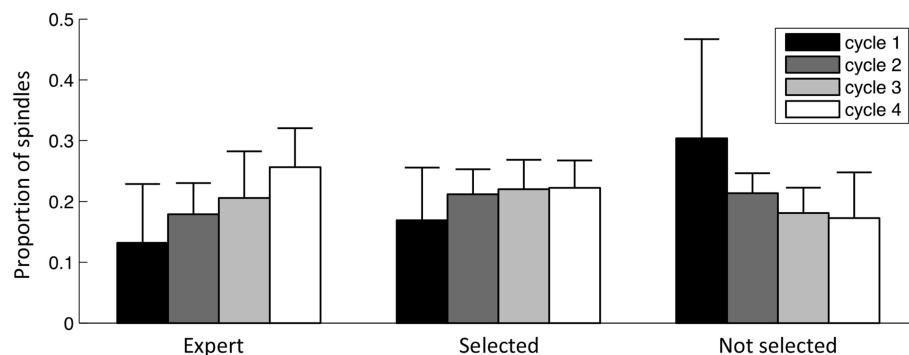


FIGURE 11 | Spindle distribution across sleep cycles. Bars indicate the mean value across subjects, whiskers indicate standard deviations. Values are normalized for subjects and classes (i.e., each of these three sets of bars sums to the unity).

the fast centro-parietal class, which tends to indicate that spindles detected in Fz are observations of the same phenomenon producing faster and stronger spindles in Cz and Pz.

The coupling observed here and elsewhere (Molle et al., 2002, 2011) between the phase of a slow ~ 1 Hz oscillation and the amplitude in the spindle band (see **Figure 10B**) warrants further investigation on spindle relationship with other frequency bands. Aside from slow waves, spindles have also been reported to be coupled with gamma (30–100 Hz) oscillations (Ayoub et al., 2012). This kind of features might be useful in increasing specificity of future detectors.

Spindle distribution across the first four sleep cycles behaved similarly for the class of spindles selected by the specific detection and the expert detection (fast spindles occurring in more posterior locations) and is shown to increase progressively across the night. This profile agrees with the evolution of the sigma band (12–14.75 Hz) reported by De Gennaro and Ferrara (2003). Events not selected at the specificity phase are generally slower with more anterior localization and have an inverse tendency: their density decreases across night. De Gennaro and Ferrara (2003) have reported that the power in the delta band (0.5–4.75 Hz) shows a similar trend, motivating the investigation of whether events in these classes are coupled with the activity in this lower frequency band. Also, since spindles have been detected on all NREM states, a more thorough analysis would be necessary to disambiguate the role of sleep stages in this trend.

DEPENDENCE ON EXPERT SCORING

The proposed system is semi-automatic, requiring an expert for stage scoring and partial spindle annotation. Stage scoring is a standard operation generally performed before manual or automatic spindle detection. However, if one does not want to score whole nights, only some spindle-free epochs (such as REM epochs) can be scored manually and fed to the algorithm. Alternatively, automatic sleep scoring algorithms can be used (Anderer et al., 2005). Although these algorithms do not perform as well as experts, they should be reasonably accurate to discriminate some classes of spindle-free epochs (wake, REM) vs. epochs possibly containing spindles (NREM stages).

As for partial spindle scoring, our results suggest that only 20 spindles per subject are sufficient to benefit from the supervised classification. Thus, the expert scoring burden is relatively small with this detector. Of course, as for any supervised system, the scoring will be as biased as the expert. Thus, using expert consensus (Warby et al., 2014) on small number of spindles instead of single-expert scoring is worth more investigation. Another future avenue is to automate the clustering using some *a priori* knowledge instead of expert scoring. To implement this, we could for example take advantage of the fact that events detected by the sensitivity phase naturally tend to show two classes plus some outliers. Using the k-mean clustering algorithm with $k = 2$ to extract the centroid of the two classes and reject outliers that are not close enough to these centers is likely to give interesting results.

GOLD STANDARD IN SPINDEL SCORING

It should be noted that the performance assessment reported in this study is limited by the relatively low reproducibility of our gold standard: expert scoring. With relatively low inter-rater agreement between expert scorers (around 86% in Campbell et al., 1980); $61 \pm 6\%$ and Cohen κ of 0.52 ± 0.07 (Wendt et al., 2014); around 0.2 and 0.4 Cohen κ in DREAMS and MASS open-access databases, respectively (O'Reilly and Nielsen, in revision), development of automated detectors will stay rather limited until the subjective assessment of spindle by experts is transcended and supplanted by a more robust, objective, and commonly agreed upon gold standard (O'Reilly and Nielsen, in revision).

CLUSTERING

The clustering algorithm has shown to be able to dichotomize sleep spindles in the fast/slow classes reported in the literature for all but one subject. Topography of spindles is not always stable across time and the clustering might be sensible to this inhomogeneity. The properties of the clustering process will require more investigation on larger samples to better understand when it fails, what it indicates, and how it can be corrected. Also, although both fast and slow classes are generally correctly identified, the slow class was rejected by our automated system because our expert ignored tentative spindles from this class. Whether this behavior is typical in expert scoring is still to be evaluated.

Similarly, whether some variables (e.g., the expert) impact on the minimal number of scored spindles needed to obtain a reliable clustering is still an open question. In our investigation, only a small number of spindle per subject (about 20) were shown to be sufficient.

Furthermore, given the somewhat low inter-rater agreement between experts reported in literature (Wendt et al., 2014), using an expert consensus measure could present great advantages (Warby et al., 2014). One should note, however, that such a strategy would probably bias scoring toward classes with high amplitude and high signal-to-noise ratio.

CONCLUSION

The principal contribution of this paper is to propose a two-step methodology to address first the sensitivity and second the specificity of spindle detection. For this last step, we proposed an unsupervised clustering using spectral and positional (along the medial axis) features to take into account the fast-posterior/slow-anterior spindle dichotomy followed by a supervised class selection. Some other original contributions proposed in this paper are: (1) the compression of channel arrays into a univariate signal using a fixed montage, (2) using the ridge of a time-frequency map to segment the signal and transform it into a detection function, and (3) using *p*-values for setting selection thresholds, based on a null-hypothesis elaborated from the spindle-free periods during sleep (e.g., REM).

Acceptable classification results have been obtained with $Se = 85.4\%$, $FDr = 86.2\%$, and $MCC = 0.32$. Although these results are similar to those available in literature, a more thorough comparison is not reported here since such an analysis would be unreliable due to the large confounding impact of using different expert scorings. For example, MCC has been shown to vary between 0.25 and 0.55 for a same detector depending on the database and the expert scoring (O'Reilly and Nielsen, in revision). Also, the assessment methodology would not be completely comparable because of the use of a particular segmentation paradigm impacting on the counts of positive/negative events. A more thorough assessment performed with comparison against other standard detection algorithms on an open-access database (e.g., O'Reilly et al., 2014) is warranted. Such an assessment is however outside of the scope of the present paper and is a topic for future investigations.

Nevertheless, it appears that there is room for improvement since the obtained agreement is below what is expected from experts. The proposed system might be enhanced by adding specific features that are known from literature to be associated with sleep spindles such as circadian and homeostatic influences (Knoblauch et al., 2003), phase coupling with slow oscillations (Molle et al., 2002), age (Martin et al., 2013), and so on. A thorough analysis of whether adding such features can indeed improve spindle detection would however be necessary since correlations between spindles and these other variables are emerging when averaging over a large number of events. Thus, they might prove not to be specific enough to improve detection of single events and can even have a detrimental impact on automatic detection, as formalized by the No Free Lunch theorem (Wolpert and Macready, 1997).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fnhum.2015.00070/abstract>

REFERENCES

- Acar, N., and Güzelis, C. (2004). Automatic recognition of sleep spindles in EEG by using artificial neural networks. *Expert Syst. Appl.* 27, 451–458. doi: 10.1016/j.eswa.2004.05.007
- Aeschbach, D., and Borbely, A. A. (1993). All-night dynamics of the human sleep EEG. *J. Sleep Res.* 2, 70–81. doi: 10.1111/j.1365-2869.1993.tb0065.x
- Ahmed, B., Redissi, A., and Tafreshi, R. (2009). “An automatic sleep spindle detector based on wavelets and the teager energy operator,” in *Conference Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2009*, (Minneapolis, MN), 2596–2599.
- Anderer, P., Gruber, G., Parapatics, S., Woertz, M., Miazynskaia, T., Klosch, G., et al. (2005). An E-health solution for automatic sleep classification according to Rechtschaffen and Kales: validation study of the Somnolyzer 24 x 7 utilizing the Siesta database. *Neuropsychobiology* 51, 115–133. doi: 10.1159/000085205
- Anderer, P., Klosch, G., Gruber, G., Trenker, E., Pascual-Marqui, R. D., Zeitlhofer, J., et al. (2001). Low-resolution brain electromagnetic tomography revealed simultaneously active frontal and parietal sleep spindle sources in the human cortex. *Neuroscience* 103, 581–592. doi: 10.1016/S0306-4522(01)00028-8
- Ayoub, A., Mölle, M., Preissl, H., and Born, J. (2012). Grouping of MEG gamma oscillations by EEG sleep spindles. *Neuroimage* 59, 1491–1500. doi: 10.1016/j.neuroimage.2011.08.023
- Babadi, B., McKinney, S. M., Tarokh, V., and Ellenbogen, J. M. (2012). DiBa: a data-driven Bayesian algorithm for sleep spindle detection. *IEEE Trans. Biomed. Eng.* 59, 483–493. doi: 10.1109/TBME.2011.2175225
- Brunner, D. P., Vasko, R. C., Detka, C. S., Monahan, J. P., Reynolds, C. F. III, and Kupfer, D. J. (1996). Muscle artifacts in the sleep EEG: automated detection and effect on all-night EEG power spectra. *J. Sleep Res.* 5, 155–164. doi: 10.1046/j.1365-2869.1996.00009.x
- Campbell, K., Kumar, A., and Hofman, W. (1980). Human and automatic validation of a phase-locked loop spindle detection system. *Electroencephalogr. Clin. Neurophysiol.* 48, 602–605. doi: 10.1016/0013-4694(80)90296-5
- Clemens, Z., Fabo, D., and Halasz, P. (2005). Overnight verbal memory retention correlates with the number of sleep spindles. *Neuroscience* 132, 529–535. doi: 10.1016/j.neuroscience.2005.01.011
- Dang-Vu, T. T., McKinney, S. M., Buxton, O. M., Solet, J. M., and Ellenbogen, J. M. (2010). Spontaneous brain rhythms predict sleep stability in the face of noise. *Curr. Biol.* 20, R626–R627. doi: 10.1016/j.cub.2010.06.032
- De Gennaro, L., and Ferrara, M. (2003). Sleep spindles: an overview. *Sleep Med. Rev.* 7, 423–440. doi: 10.1053/smrv.2002.0252
- Dehghani, N., Cash, S. S., and Halgren, E. (2011). Topographical frequency dynamics within EEG and MEG sleep spindles. *Clin. Neurophysiol.* 122, 229–235. doi: 10.1016/j.clinph.2010.06.018
- Delprat, N., Escudie, B., Guillemin, P., Kronland-Martinet, R., Tchamitchian, P., and Torresani, B. (1992). Asymptotic wavelet and Gabor analysis: extraction of instantaneous frequencies. *IEEE Trans. Inf. Theory* 38, 644–664. doi: 10.1109/18.119728
- Devuyast, S., Dutoit, T., Stenuit, P., and Kerckhofs, M. (2011). “Automatic sleep spindles detection—overview and development of a standard proposal assessment method,” in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2011*, (Boston, MA), 1713–1716.
- Fogel, S., Martin, N., Lafourche, M., Barakat, M., Debas, K., Laventure, S., et al. (2012). NREM sleep oscillations and brain plasticity in aging. *Front. Neurol.* 3:176. doi: 10.3389/fneur.2012.00176
- Huupponen, E., Gomez-Herrero, G., Saastamoinen, A., Varri, A., Hasan, J., and Himanen, S. L. (2007). Development and comparison of four sleep spindle detection methods. *Artif. Intell. Med.* 40, 157–170. doi: 10.1016/j.artmed.2007.04.003
- Huupponen, E., Kulkas, A., Tenhunen, M., Saastamoinen, A., Hasan, J., and Himanen, S. L. (2008). Diffuse sleep spindles show similar frequency in central and frontopolar positions. *J. Neurosci. Methods* 172, 54–59. doi: 10.1016/j.jneumeth.2008.03.019
- Iber, C., Ancoli-Israel, S., Chesson, A., and Quan, S. F. (2007). *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology*

- and Technical Specifications.* Westchester, IL: American Academy of Sleep Medicine.
- Knoblauch, V., Martens, W. L., Wirz-Justice, A., and Cajochen, C. (2003). Human sleep spindle characteristics after sleep deprivation. *Clin. Neurophysiol.* 114, 2258–2267. doi: 10.1016/S1388-2457(03)00238-4
- Kokkinos, V., and Kostopoulos, G. K. (2011). Human non-rapid eye movement stage II sleep spindles are blocked upon spontaneous K-complex coincidence and resume as higher frequency spindles afterwards. *J. Sleep Res.* 20, 57–72. doi: 10.1111/j.1365-2869.2010.00830.x
- Lilly, J. M., and Olhede, S. C. (2009). Higher-order properties of analytic wavelets. *IEEE Trans. Signal Process.* 57, 146–160. doi: 10.1109/TSP.2008.2007607
- Lilly, J. M., and Olhede, S. C. (2010). On the analytic wavelet transform. *IEEE Trans. Inf. Theory* 56, 4135–4156. doi: 10.1109/TIT.2010.2050935
- Martin, N., Lafontaine, M., Godbout, J., Barakat, M., Robillard, R., Poirier, G., et al. (2013). Topography of age-related changes in sleep spindles. *Neurobiol. Aging* 34, 468–476. doi: 10.1016/j.neurobiolaging.2012.05.020
- Molle, M., Bergmann, T. O., Marshall, L., and Born, J. (2011). Fast and slow spindles during the sleep slow oscillation: disparate coalescence and engagement in memory processing. *Sleep* 34, 1411–1421. doi: 10.5665/SLEEP.1290
- Molle, M., Marshall, L., Gais, S., and Born, J. (2002). Grouping of spindle activity during slow oscillations in human non-rapid eye movement sleep. *J. Neurosci.* 22, 10941–10947.
- Olbrich, E., and Achermann, P. (2005). Analysis of oscillatory patterns in the human sleep EEG using a novel detection algorithm. *J. Sleep Res.* 14, 337–346. doi: 10.1111/j.1365-2869.2005.00475.x
- Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011:156869. doi: 10.1155/2011/156869
- Oostenveld, R., and Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clin. Neurophysiol.* 112, 713–719. doi: 10.1016/S1388-2457(00)00527-7
- O'Reilly, C., and Nielsen, T. (2014a). Assessing EEG sleep spindle propagation. Part 1: theory and proposed methodology. *J. Neurosci. Methods* 221, 202–214. doi: 10.1016/j.jneumeth.2013.08.013
- O'Reilly, C., and Nielsen, T. (2014b). Assessing EEG sleep spindle propagation. Part 2: experimental characterization. *J. Neurosci. Methods* 221, 215–227. doi: 10.1016/j.jneumeth.2013.08.014
- O'Reilly, C., Gosselin, N., Carrier, J., and Nielsen, T. (2014). Montreal archive of sleep studies: an open-access resource for instrument benchmarking and exploratory research. *J. Sleep Res.* 23, 628–635. doi: 10.1111/jsr.12169
- O'Reilly, C., and Nielsen, T. (2013). “Revisiting the ROC curve for diagnostic applications with an unbalanced class distribution,” in *Proceedings of the 8th International Workshop on Systems, Signal Processing and Their Applications* (Algiers).
- Rechtschaffen, A., and Kales, A. (1968). *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects* (Bethesda, MD: U. S. National Institute of Neurological Diseases and Blindness, Neurological Information Network).
- Schabus, M., Dang-Vu, T. T., Albouy, G., Balteau, E., Boly, M., Carrier, J., et al. (2007). Hemodynamic cerebral correlates of sleep spindles during human non-rapid eye movement sleep. *Proc. Natl. Acad. Sci. U.S.A.* 104, 13164–13169. doi: 10.1073/pnas.0703084104
- Schimicek, P., Zeilhofer, J., Anderer, P., and Saletu, B. (1994). Automatic sleep-spindle detection procedure: aspects of reliability and validity. *Clin. EEG* 25, 26–29. doi: 10.1177/155005949402500108
- Sheng-Fu, L., Chih-En, K., Yu-Han, H., Chun-Yu, C., and Yu-Hung, L. (2012). “An adaptive neuro-fuzzy inference system for sleep spindle detection,” in *Fuzzy Theory and its Applications (iFUZZY), International Conference on 2012*, (Taichung), 369–373.
- Sinha, R. K. (2008). Artificial neural network and wavelet based automated detection of sleep spindles, REM sleep and wake states. *J. Med. Syst.* 32, 291–299. doi: 10.1007/s10916-008-9134-z
- Steriade, M. (2006). Grouping of brain rhythms in corticothalamic systems. *Neuroscience* 137, 1087–1106. doi: 10.1016/j.neuroscience.2005.10.029
- Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., and Leahy, R. M. (2011). Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput. Intell. Neurosci.* 2011:879716. doi: 10.1155/2011/879716
- Timm, N. H. (2002). *Applied Multivariate Analysis*. New York, NY: Springer. Available online at: <http://public.eblib.com/EBLPublic/PublicView.do?ptID=219437>
- Ventouras, E. M., Monoyiou, E. A., Ktonas, P. Y., Paparrigopoulos, T., Dikeos, D. G., Uzunoglu, N. K., et al. (2005). Sleep spindle detection using artificial neural networks trained with filtered time-domain EEG: a feasibility study. *Comput. Methods Programs Biomed.* 78, 191–207. doi: 10.1016/j.cmpb.2005.02.006
- Vyazovskiy, V. V., Achermann, P., Borbely, A. A., and Tobler, I. (2004). The dynamics of spindles and EEG slow-wave activity in NREM sleep in mice. *Arch. Ital. Biol.* 142, 511–523.
- Warby, S. C., Wendt, S. L., Welinder, P., Munk, E. G., Carrillo, O., Sorensen, H. B., et al. (2014). Sleep-spindle detection: crowdsourcing and evaluating performance of experts, non-experts and automated methods. *Nat. Methods* 11, 385–392. doi: 10.1038/nmeth.2855
- Wendt, S. L., Welinder, P., Sorensen, H. B., Peppard, P. E., Jennum, P., Perona, P., et al. (2014). Inter-expert and intra-expert reliability in sleep spindle scoring. *Clin. Neurophysiol.* doi: 10.1016/j.clinph.2014.10.158. [Epub ahead of print].
- Werth, E., Achermann, P., Dijk, D. J., and Borbely, A. A. (1997). Spindle frequency activity in the sleep EEG: individual differences and topographic distribution. *Electroencephalogr. Clin. Neurophysiol.* 103, 535–542. doi: 10.1016/S0013-4694(97)00070-9
- Wolpert, D. H., and Macready, W. G. (1997). No free lunch theorems for optimization. *Evol. Comput. IEEE Trans.* 1, 67–82. doi: 10.1109/4235.585893
- Zeitlhofer, J., Gruber, G., Anderer, P., Asenbaum, S., Schimicek, P., and Saletu, B. (1997). Topographic distribution of sleep spindles in young healthy subjects. *J. Sleep Res.* 6, 149–155. doi: 10.1046/j.1365-2869.1997.00046.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 12 September 2014; **accepted:** 27 January 2015; **published online:** 19 February 2015.

Citation: O'Reilly C, Godbout J, Carrier J and Lina J-M (2015) Combining time-frequency and spatial information for the detection of sleep spindles. *Front. Hum. Neurosci.* 9:70. doi: 10.3389/fnhum.2015.00070

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2015 O'Reilly, Godbout, Carrier and Lina. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.