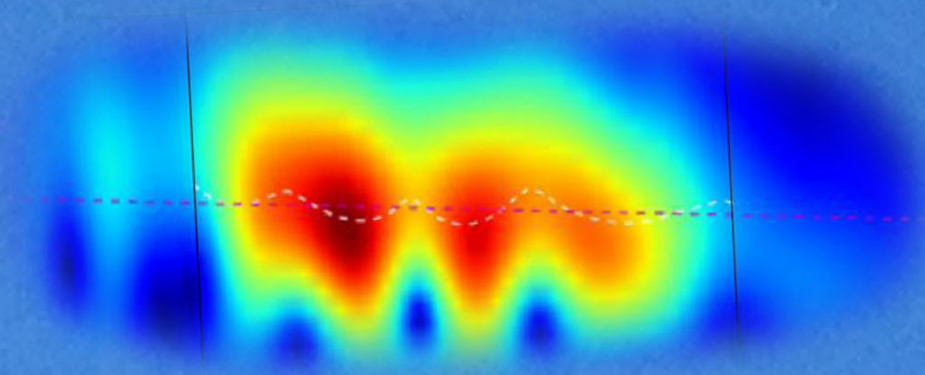


SLEEP SPINDLES: BREAKING THE METHODOLOGICAL WALL

EDITED BY : Christian O'Reilly, Simon C. Warby and Tore Nielsen

PUBLISHED IN: Frontiers in Human Neuroscience





frontiers

Frontiers Copyright Statement

© Copyright 2007-2017 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88945-116-6

DOI 10.3389/978-2-88945-116-6

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

SLEEP SPINDLES: BREAKING THE METHODOLOGICAL WALL

Topic Editors:

Christian O'Reilly, École Polytechnique Fédérale de Lausanne, Switzerland & Centre de Recherche de l'Hôpital du Sacré-Cœur de Montréal, Canada

Simon C. Warby, Centre de Recherche de l'Hôpital du Sacré-Cœur de Montréal & Université de Montréal, Canada

Tore Nielsen, Université de Montréal, Montreal & Centre de Recherche de l'Hôpital du Sacré-Cœur de Montréal, Canada



In the last decade, sleep spindles have attracted steadily increasing attention. This interest is motivated by the many intriguing relationships between spindles and various diseases (e.g., schizophrenia, Parkinson, Alzheimer, autism, mental retardation), recovery processes (e.g., post brain stroke), and cognitive faculties (e.g., memory consolidation, intelligence, dream recall, sleep preservation). Nonetheless, a methodological wall has impeded the study of sleep spindles. Their investigation rests heavily on our ability to reliably and consistently identify spindle patterns from background EEG activity, a task involving many obstacles, including: a fuzzy definition of spindles, low inter-expert agreement on their scoring, lack of consensus on standard techniques

for their automated detection, low reproducibility of observed characteristics and correlates, unavailability of large, standardized, high-quality databases, and inconsistencies in the methods used to evaluate the performance of automated detectors.

The primary aims of this research topic were to bring together world-class researchers on a project designed to facilitate exchanges on methodological difficulties encountered in assessing sleep spindles and to promote standardized spindle-related resources. In preparing their contributions, authors were encouraged to use existing – or to propose new – publicly available resources for assessing sleep spindles. To allow fair and accurate comparison of reported results, the authors were also encouraged to validate their tools on a common benchmark. A database containing expert spindle scoring (i.e., the Montreal Archive of Sleep Studies) was made publicly available for that purpose.

Citation: O'Reilly, C., Warby, S. C., Nielsen, T., eds. (2017). Sleep Spindles: Breaking the Methodological Wall. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-116-6

Table of Contents

Section 1

06 Editorial: Sleep Spindles: Breaking the Methodological Wall

Christian O'Reilly, Simon C. Warby and Tore Nielsen

Section 2: Advances in automatic spindle detection

09 Automatic Sleep Spindle Detection and Genetic Influence Estimation Using Continuous Wavelet Transform

Marek Adamczyk, Lisa Genzel, Martin Dresler, Axel Steiger and Elisabeth Friess

29 Spindles in Svarog: framework and software for parametrization of EEG transients

Piotr J. Durka, Urszula Malinowska, Magdalena Zieleniewska, Christian O'Reilly, Piotr T. Róžański and Jarosław Żygierewicz

41 Automated detection of sleep spindles in the scalp EEG and estimation of their intracranial current sources: comments on techniques and on related experimental and clinical studies

Periklis Y. Ktonas and Errikos-Chaim Ventouras

45 Sleep spindle and K-complex detection using tunable Q-factor wavelet transform and morphological component analysis

Tarek Lajnef, Sahbi Chaibi, Jean-Baptiste Eichenlaub, Perrine M. Ruby, Pierre-Emmanuel Aguera, Mounir Samet, Abdennaceur Kachouri and Karim Jerbi

62 Automatic sleep spindle detection: benchmarking with fine temporal resolution using open science tools

Christian O'Reilly and Tore Nielsen

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

Christian O'Reilly, Jonathan Godbout, Julie Carrier and Jean-Marc Lina

95 Expert and crowd-sourced validation of an individualized sleep spindle detection method employing complex demodulation and individualized normalization

Laura B. Ray, Stéphane Sockeel, Melissa Soon, Arnaud Bore, Ayako Myhr, Bobby Stojanoski, Rhodri Cusack, Adrian M. Owen, Julien Doyon and Stuart M. Fogel

111 Stage-independent, single lead EEG sleep spindle detection using the continuous wavelet transform and local weighted smoothing

Athanasios Tsanas and Gari D. Clifford

126 Corrigendum: A comparison of two sleep spindle detection methods based on all night averages: individually adjusted vs. fixed frequencies

Péter P. Ujma, Ferenc Gombos, Lisa Genzel, Boris N. Konrad, Péter Simor, Axel Steiger, Martin Dresler and Róbert Bódizs

127 *A comparison of two sleep spindle detection methods based on all night averages: individually adjusted vs. fixed frequencies*

Péter Przemyslaw Ujma, Ferenc Gombos, Lisa Genzel, Boris Nikolai Konrad, Péter Simor, Axel Steiger, Martin Dresler and Róbert Bódizs

Section 3: Modeling the spindle waveform

138 *Using a quadratic parameter sinusoid model to characterize the structure of EEG sleep spindles*

Abdul J. Palliyali, Mohammad N. Ahmed and Beena Ahmed

Section 4: Correlates of sleep spindles

152 *Sleep spindle and slow wave frequency reflect motor skill performance in primary school-age children*

Rebecca G. Astill, Giovanni Piantoni, Roy J. E. M. Raymann, Jose C. Vis, Joris E. Coppens, Matthew P. Walker, Robert Stickgold, Ysbrand D. Van Der Werf and Eus J. W. Van Someren

165 *Sleep spindling and fluid intelligence across adolescent development: sex matters*

Róbert Bódizs, Ferenc Gombos, Péter P. Ujma and Ilona Kovács

176 *Sleep spindle alterations in patients with Parkinson's disease*

Julie A. E. Christensen, Miki Nikolic, Simon C. Warby, Henriette Koch, Marielle Zoetmulder, Rune Frandsen, Keivan K. Moghadam, Helge B. D. Sorensen, Emmanuel Mignot and Poul J. Jennum

189 *Sleep spindles predict stress-related increases in sleep disturbances*

Thien Thanh Dang-Vu, Ali Salimi, Soufiane Boucetta, Kerstin Wenzel, Jordan O'Byrne, Marie Brandewinder, Christian Berthomier and Jean-Philippe Gouin

198 *Sleep spindle deficits in antipsychotic-naïve early course schizophrenia and in non-psychotic first-degree relatives*

Dara S. Manoach, Charmaine Demanuele, Erin J. Wamsley, Mark Vangel, Debra M. Montrose, Jean Miewald, David Kupfer, Daniel Buysse, Robert Stickgold and Matcheri S. Keshavan

214 *Correlations between adolescent processing speed and specific spindle frequencies*

Rebecca S. Nader and Carlyle T. Smith

222 *Age-related changes in sleep spindles characteristics during daytime recovery following a 25-hour sleep deprivation*

T. Rosinvil, M. Lafortune, Z. Sekerovic, M. Bouchard, J. Dubé, A. Latulipe-Loiselle, N. Martin, J. M. Lina and J. Carrier



Editorial: Sleep Spindles: Breaking the Methodological Wall

Christian O'Reilly^{1,2*}, Simon C. Warby^{2,3} and Tore Nielsen^{3,4}

¹ Blue Brain Project, École Polytechnique Fédérale de Lausanne, Geneva, Switzerland, ² Center for Advanced Research in Sleep Medicine, Centre de Recherche de l'Hôpital du Sacré-Cœur de Montréal, Montreal, QC, Canada, ³ Département de Psychiatrie, Université de Montréal, Montreal, QC, Canada, ⁴ Dream and Nightmare Laboratory, Center for Advanced Research in Sleep Medicine, Centre de Recherche de l'Hôpital du Sacré-Cœur de Montréal, Montreal, QC, Canada

Keywords: sleep spindles, methods, sleep, open access

Editorial on the Research Topic

Sleep Spindles: Breaking the Methodological Wall

Research on sleep spindles and their correlates has progressed steadily over the last decade. The subject has evolved from a simple topic of investigation to an emerging research field, as indicated this year by the first international conference on sleep spindles in Budapest, Hungary, as well as the launching of a scientific journal (i.e., *Sleep Spindles and Cortical Up States: A Multidisciplinary Journal*) on this topic. This increasing interest has been fueled by reports of associations of sleep spindle characteristics with diseases such as schizophrenia (Ferrarelli et al., 2007, 2010; Manoach et al.), Parkinson's disease (Christensen et al.), REM sleep behavior disorder (Christensen et al., 2014; O'Reilly et al., 2015), Alzheimer's disease (Montplaisir et al., 1995; Rauchs et al., 2008), autism (Limoges et al., 2005), and mental retardation (Shibagaki et al., 1982), with recovery processes following brain stroke (Gottselig et al., 2002), with cognitive faculties such as memory consolidation and intelligence (Fogel and Smith, 2011), and with sleep preservation (Landis et al., 2004; Dang-Vu et al., 2010; Schabus et al., 2012). Nonetheless, many methodological difficulties have been encountered in reliably detecting sleep spindles. Hence, this research topic was launched as a forum for proposing better practices in the study of sleep spindles and to provide new insights on spindle correlates. Authors were invited particularly to propose open-access resources that could help promote improved methods and support standardization in the field.

OPEN ACCESS

Edited and reviewed by:

Hauke R. Heekeren,
Freie Universität Berlin, Germany

*Correspondence:

Christian O'Reilly
christian.oreilly@epfl.ch

Received: 29 March 2016

Accepted: 16 December 2016

Published: 18 January 2017

Citation:

O'Reilly C, Warby SC and Nielsen T
(2017) Editorial: Sleep Spindles:
Breaking the Methodological Wall.
Front. Hum. Neurosci. 10:672.
doi: 10.3389/fnhum.2016.00672

CONTRIBUTIONS

A total of 17 papers were accepted for publication on the research topic, with 10 being focussed particularly on methodological issues such as spindle detection and the remaining seven providing new insights on sleep spindle correlates.

Methodological Advances

Different approaches were investigated for tackling the difficult task of detecting sleep spindles automatically, including the use of continuous wavelet transform (Adamczyk et al.; Tsanas and Clifford), complex demodulation (Ray et al.), matching pursuit (Durka et al.), and morphological component analysis of a sparse representation of EEG segments using the discrete tunable Q-factor wavelet transform (Lajnef et al.).

Among the developments proposed for sleep spindle detection, some concentrate on particular issues associated with clinical applications or with better control of factors impacting spindle variability. For clinical applications, Tsanas and Clifford propose a detector deployable with single-lead recordings that does not require prior sleep stage scoring, two arguably

important features for daily clinical use. From the perspective of better controlling factors impacting on the variability of spindle properties, Ray et al. propose an algorithm accounting for variability across the night, across derivations, and across subjects while keeping the number of user-defined parameters to a minimum. Ujma et al. propose arguments that support dynamically determining, for each subject, the threshold used for separating fast from slow spindles according to the spectral structure of the individual's EEG. Such individually defined thresholds are used in the detector proposed by Adamczyk et al. Some of the proposed detection techniques also aim at a more general detection framework, which could manage a larger set of sleep waveforms, e.g., including not only sleep spindles but also K-complexes (Durka et al.; Lajnef et al.).

In their contribution to the special issue, O'Reilly and Nielsen suggest modified versions of four standard detection algorithms to improve temporal resolution in determining spindling time windows. They also provide an in-depth analysis of the limitations and pitfalls associated with spindle detection assessment. Pitfalls and guidelines for spindle detection can also be found in an opinion paper by Ktonas and Ventouras.

O'Reilly et al. take a different approach and propose a semi-automated detector relying on machine learning. In this approach, sigma-band amplitude, and spectral ratio features are used in a first step followed by hierarchical clustering based on frequency and spatial position of the spindle along the anterior-posterior axis of the scalp, so as to capture differences between classes of slow and fast spindles. This proposal falls to some extent at the opposite end of a spectrum when compared to the proposal of Tsanas and Clifford; whereas the former tries to benefit from high-density grid recordings for research purposes, the latter focuses on obtaining reliable detections from minimal information for clinical uses. Related to the context of the former study are the comments from the Ktonas and Ventouras opinion paper on the estimation of intracranial current sources of sleep spindles, a topic that is likely to become increasingly important with the improvement of source localization algorithms, and the wider spread of EEG high-density sensor grids.

Targeted more toward developing an improved representation of sleep spindles than toward detection *per se*, Palliyali et al. propose to parameterize the structure of spindles using a quadratic parameter sinusoid. In their study, they provide a detailed analysis of the parameters' sensitivity and show, among other findings, that these parameters take distinct values for spindle vs. non-spindle epochs.

More closely related to the very definition of sleep spindles, Nader and Smith propose some controversial results that challenge the traditional view of sleep spindles by investigating sleep spindles in atypical stages (e.g., REM) and frequency bands (e.g., 16–18.5 Hz).

It is noteworthy that a significant number of contributed papers (Durka et al.; O'Reilly and Nielsen; Palliyali et al.; Tsanas and Clifford) include an evaluation of their detection algorithms on a common database (the second subset of the Montreal Archive of Sleep Studies; O'Reilly et al., 2014), thereby providing much better cross-study comparisons than if they had been evaluated using different expert scorings (O'Reilly and Nielsen).

Proposal of Open-Access Tools

A valuable outcome of this research topic is the release of many open-access resources for studying sleep spindles. This is the case for the matching pursuit detector of Durka et al. which is provided as part of the Signal Viewer, Analyzer, and Recorder On GPL (SVAROG) package available at <http://braintech.pl/svarog>; of the detectors evaluated in O'Reilly and Nielsen which are part of the open-source Python package SpynDle available at https://bitbucket.org/christian_oreilly/spyndle; and of the single-lead detector of Tsanas and Clifford available as a Matlab source code at <https://people.maths.ox.ac.uk/tsanas/>. Similarly, some other Matlab packages are available directly from the authors Adamczyk et al., Lajnef et al., and Ray et al. Finally, the detector from O'Reilly et al. has been implemented as a Brainstorm (Matlab) process for easy integration with neuroimaging pipelines implemented in this environment. It is also available from the authors.

Other Advances in the Study of Sleep Spindling

Although primarily targeted at discussing methodological issues related to the investigation of sleep spindles, other types of validation studies of sleep spindles were included to broaden the scope of this research topic. This includes two papers on the relationship between sleep spindles and mental faculties in adolescents, one examining how spindling frequency is related to processing speed as well as the relationship between performance on a motor task and sleep quality (Nader and Smith), the other assessing links between sleep spindles and fluid IQ, with a particular attention to sex as a modulating factor (Bódizs et al.). Similarly, Astill et al. studied links between performance on a motor task and sleep spindling in children; they found better performance with faster EEG, in accordance with what was reported for adolescents (Nader and Smith).

Two contributions examine how diseases are correlated with properties of sleep spindles, one focusing on Parkinson's disease (Christensen et al.), the other on schizophrenia (Manoach et al.). Others report correlates of sleep spindles including age-related impact of sleep-deprivation (Rosinvil et al.) and level of insomnia symptoms in response to a stressful situation (Dang-Vu et al.). Finally, Adamczyk et al. report on the influence of genetics on the variability of slow and fast sleep spindles.

These studies demonstrate once more that sleep spindling is an important physiological process that can be modulated by many conditions. They also further highlight the relevance of establishing the role of sleep spindles in the normal functioning of the brain.

CONCLUSION

With the publication of an e-book compiling all these contributions on sleep spindle correlates and methodological advancements for their study, another step has been taken in advancing the foundations of this emerging research field. It is the hope of its editors that these papers will support the continued enhancement of methods used to study sleep spindling, promote

the establishment of commonly used open-access research tools and, eventually, foster a better understanding of the mechanisms involved in sleep spindles and their role in neurophysiological and pathological processes.

AUTHOR CONTRIBUTIONS

COR wrote the first draft. All authors revised and edited the manuscript.

REFERENCES

- Christensen, J. A., Kempfner, J., Zoetmulder, M., Leonthin, H. L., Arvastson, L., Christensen, S. R., et al. (2014). Decreased sleep spindle density in patients with idiopathic REM sleep behavior disorder and patients with Parkinson's disease. *Clin. Neurophysiol.* 125, 512–519. doi: 10.1016/j.clinph.2013.08.013
- 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
- Ferrarelli, F., Huber, R., Peterson, M. J., Massimini, M., Murphy, M., Riedner, B. A., et al. (2007). Reduced sleep spindle activity in schizophrenia patients. *Am. J. Psychiatry* 164, 483–492. doi: 10.1176/ajp.2007.164.3.483
- Ferrarelli, F., Peterson, M. J., Sarasso, S., Riedner, B. A., Murphy, M. J., Benca, R. M., et al. (2010). Thalamic dysfunction in schizophrenia suggested by whole-night deficits in slow and fast spindles. *Am. J. Psychiatry* 167, 1339–1348. doi: 10.1176/appi.ajp.2010.09121731
- Fogel, S. M., and Smith, C. T. (2011). The function of the sleep spindle: a physiological index of intelligence and a mechanism for sleep-dependent memory consolidation. *Neurosci. Biobehav. Rev.* 35, 1154–1165. doi: 10.1016/j.neubiorev.2010.12.003
- Gottselig, J. M., Bassetti, C. L., and Achermann, P. (2002). Power and coherence of sleep spindle frequency activity following hemispheric stroke. *Brain* 125(Pt 2), 373–383. doi: 10.1093/brain/awf021
- Landis, C. A., Lentz, M. J., Rothermel, J., Buchwald, D., and Shaver, J. L. (2004). Decreased sleep spindles and spindle activity in midlife women with fibromyalgia and pain. *Sleep* 27, 741–750.
- Limoges, E., Motttron, L., Bolduc, C., Berthiaume, C., and Godbout, R. (2005). Atypical sleep architecture and the autism phenotype. *Brain* 128(Pt 5), 1049–1061. doi: 10.1093/brain/awh425

ACKNOWLEDGMENTS

The authors would like to thanks Julie Carrier, Nadia Gosselin, Sonia Frenette, Tyna Paquette, H  l  ne Blais, and Stuart Fogel for their help in setting up the MASS PSG database and/or in annotating its spindles; their work made it possible to provide this resource as a common benchmark to evaluate contributions to this research topic and for the continued growth of this new research domain.

- Montplaisir, J., Petit, D., Lorrain, D., Gauthier, S., and Nielsen, T. (1995). Sleep in Alzheimer's disease: further considerations on the role of brainstem and forebrain cholinergic populations in sleep-wake mechanisms. *Sleep* 18, 145–148.
- O'Reilly, C., Godin, I., Montplaisir, J., and Nielsen, T. (2015). REM sleep behaviour disorder is associated with lower fast and higher slow sleep spindle densities. *J. Sleep Res.* 24, 593–601. doi: 10.1111/jsr.12309
- 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
- Rauchs, G., Schabus, M., Parapatics, S., Bertran, F., Clochon, P., Hot, P., et al. (2008). Is there a link between sleep changes and memory in Alzheimer's disease? *Neuroreport* 19, 1159–1162. doi: 10.1097/WNR.0b013e32830867c4
- Schabus, M., Dang-Vu, T. T., Heib, D. P., Boly, M., Desseilles, M., Vandewalle, G., et al. (2012). The fate of incoming stimuli during NREM sleep is determined by spindles and the phase of the slow oscillation. *Front. Neurol.* 3:40. doi: 10.3389/fneur.2012.00040
- Shibagaki, M., Kiyono, S., and Watanabe, K. (1982). Spindle evolution in normal and mentally retarded children: a review. *Sleep* 5, 47–57.

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.

Copyright © 2017 O'Reilly, Warby and Nielsen. 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.



Automatic Sleep Spindle Detection and Genetic Influence Estimation Using Continuous Wavelet Transform

Marek Adamczyk^{1*}, Lisa Genzel², Martin Dresler^{1,3}, Axel Steiger¹ and Elisabeth Friess¹

¹ Max Planck Institute of Psychiatry, Munich, Germany, ² Centre for Cognitive and Neural Systems, University of Edinburgh, Edinburgh, UK, ³ Donders Institute for Brain, Cognition and Behaviour, Nijmegen, Netherlands

Mounting evidence for the role of sleep spindles in neuroplasticity has led to an increased interest in these non-rapid eye movement (NREM) sleep oscillations. It has been hypothesized that fast and slow spindles might play a different role in memory processing. Here, we present a new sleep spindle detection algorithm utilizing a continuous wavelet transform (CWT) and individual adjustment of slow and fast spindle frequency ranges. Eighteen nap recordings of ten subjects were used for algorithm validation. Our method was compared with both a human scorer and a commercially available SIESTA spindle detector. For the validation set, mean agreement between our detector and human scorer measured during sleep stage 2 using kappa coefficient was 0.45, whereas mean agreement between our detector and SIESTA algorithm was 0.62. Our algorithm was also applied to sleep-related memory consolidation data previously analyzed with a SIESTA detector and confirmed previous findings of significant correlation between spindle density and declarative memory consolidation. We then applied our method to a study in monozygotic (MZ) and dizygotic (DZ) twins, examining the genetic component of slow and fast sleep spindle parameters. Our analysis revealed strong genetic influence on variance of all slow spindle parameters, weaker genetic effect on fast spindles, and no effects on fast spindle density and number during stage 2 sleep.

OPEN ACCESS

Edited by:

Christian O'Reilly,
École Polytechnique Fédérale de
Lausanne, Switzerland

Reviewed by:

George Kostopoulos,
University of Patras, Greece
Simon C. Warby,
Stanford University, USA

*Correspondence:

Marek Adamczyk
marek.adamczyk84@gmail.com

Received: 31 January 2015

Accepted: 30 October 2015

Published: 19 November 2015

Citation:

Adamczyk M, Genzel L, Dresler M,
Steiger A and Friess E (2015)
Automatic Sleep Spindle Detection
and Genetic Influence Estimation
Using Continuous Wavelet Transform.
Front. Hum. Neurosci. 9:624.
doi: 10.3389/fnhum.2015.00624

Keywords: EEG, sleep spindle, automatic detection, twins, heritability

INTRODUCTION

Sleep spindles are one of the hallmarks in electroencephalographic (EEG) signal during non-rapid eye movement (NREM) sleep. They are characterized as bursts of rhythmical activity in the 10–16 Hz frequency range, with waxing and waning shapes lasting usually from 0.5–2.5 s. There are two types of sleep spindles. The so-called fast spindles are mainly present in parietal brain regions, whereas slow spindles predominate in frontal areas. Low-resolution electromagnetic tomography (LORETA) demonstrated a distributed slow spindle source in the prefrontal cortex and a fast spindle source in the precuneus (Anderer et al., 2001). However, both spindle types are generated via thalamic-cortical loops (Astori et al., 2013). The average slow spindle peak is 11.5 Hz and fast spindle peak is 13 Hz, with large inter-subject variation (Werth et al., 1997).

There is a mounting evidence for the role of sleep spindles in neuroplasticity. Increased spindle density and activity was observed after both declarative and procedural learning (Gais et al., 2002; Morin et al., 2008). Increases in spindle activity were also reported to positively correlate with memory retention (Clemens et al., 2005; Nishida and Walker, 2007;

Genzel et al., 2009; Cox et al., 2012). These oscillations provide excellent conditions for long-term synaptic changes (Buzsáki, 1989; Fogel and Smith, 2011), and the interplay of spindles and hippocampal ripples plays an important role in neuroplasticity (Clemens et al., 2007; Genzel et al., 2014). Specifically, spindles deafferent the cortex from the hippocampus, enabling local processing of increased firing rates in the cortex in response to hippocampal firing during ripples (Peyrache et al., 2009; Wierzyński et al., 2009; Genzel et al., 2014) and may additionally serve a role in cortical plasticity processes that are independent of hippocampal-led replay (Andrillon et al., 2011; Genzel et al., 2014). Sleep spindles have also been proposed to represent a biomarker of learning trait and intelligence (Fogel and Smith, 2011), however the strength of this association has recently been doubted (Ujma et al., 2014). Furthermore, impaired sleep spindle activity was shown in various psychiatric disorders (Astori et al., 2013). Reduced spindle activity was reported in patients with schizophrenia (Ferrarelli et al., 2007, 2010; Wamsley et al., 2012), affective disorders (de Maertelaer et al., 1987; Lopez et al., 2010) and Alzheimer's disease (Montplaisir et al., 1995), and these diseases also showed impaired sleep related memory consolidation (Dresler et al., 2010, 2011; Genzel et al., 2011, 2015).

In view of the putative potential of sleep spindles as biomarkers, their heritability is of interest. Previous studies showed that the NREM sleep power spectrum in the sleep spindles frequency range has finger-print characteristics (De Gennaro et al., 2005; Buckelmüller et al., 2006) and is heritable (Ambrosius et al., 2008; De Gennaro et al., 2008), suggesting that sleep spindle activity is also heritable. However, this “spindle-print” on the power spectrum is influenced by a number of mixed slow and fast spindle characteristics: their frequency, amplitude and amount. Therefore, we decided to investigate the heritability of sleep spindle basic characteristics in detail. For this purpose we developed, validated and applied a new spindle detection algorithm to our twin data.

A number of spindle detection algorithms are already published. One of the first was presented by Schimicek et al. (1994). This method uses a band-pass filter (pass-band: 11.5–16 Hz) and detects spindles with a fixed amplitude threshold (peak-to-peak amplitude of 25 μ V). Later algorithms proposed a diversity of solutions to better “extract” sleep spindles from the signal as well as to handle high inter-subject variability in sleep spindle frequency and EEG signal amplitude. One of the approaches to improve the extraction of spindle shapes from the signal is the application of a wavelet transform (WT) instead of a band-pass filter (Zygierewicz et al., 1999; Latka et al., 2005; Wamsley et al., 2012). The outcome of a WT depends not only on the power in a given frequency, but also on the shape of graphoelements in the signal, and therefore may be more specific than band-pass filtering (Addison, 2002). The other approach that considers waxing and waning shape of sleep spindles is the application of two thresholds, from which the higher one is used to localize activity bursts in sigma frequency and the lower one to estimate the duration of sleep spindles (Ferrarelli et al., 2007). Another challenge in sleep spindle detection is the

variation in EEG signal amplitude between subjects, but also channels. Reasons for this phenomenon can be of a technical nature (movements during the measurement period influencing electrode placement, differences in electrode impedance) as well as physiological. EEG signal decreases with age (Dijk et al., 1989b), and is higher in females compared to males (Dijk et al., 1989a). For this reason, spindle detection threshold in many algorithms is set individually according to various characteristics of analyzed EEG signal: for example through the average amplitude in individually localized spindle frequency range (Bódizs et al., 2009; Ujma et al., 2015) or the amplitude of pre-localized spindle candidates (Huupponen et al., 2007). Furthermore, inter-subject variation in slow and fast spindle frequency reported by Werth et al. (1997) suggests that these frequency ranges should be adjusted individually in order to discriminate between fast and slow spindles. Bódizs et al. (2009, 2012) proposed to estimate spindle frequency ranges using pre-computed average frequency spectra in the 9–16 Hz range. Slower and faster sigma peaks are usually dominant over the frontal and parietal derivations, respectively. For this reason, normalized frequency spectra for frontal and parietal EEG channels were compared and a peak higher in the frontal EEG spectrum was considered a slow spindle peak whereas a peak higher in the parietal EEG spectrum was considered a fast spindle peak.

Due to inter-subject variation in slow and fast spindle frequency, as well as in signal amplitude, spindle detection is a challenging task. It was shown recently that agreement between algorithms and humans is surprisingly low (Warby et al., 2014). Proper separation between slow and fast spindles seems to be very important, since these two types of spindles may play different roles in sleep-dependent memory processing (Möller et al., 2011). For this reason, our aim was to develop a spindle detector which acknowledges considerable inter-subject variability in sleep spindle activity. In our algorithm we combined previously published methodological solutions with our proposal of detection thresholds adjustment and estimation of spindle frequency ranges. We compared spindle detection of our new algorithm with both a human scorer and a commercially available SIESTA spindle detector (Anderer et al., 2005). Considerable detection differences between the algorithms raises the question on how different methods could influence the interpretation of previous findings. In order to investigate this further, we applied our algorithm to sleep-related memory consolidation data, which were already analyzed with the SIESTA algorithm and revealed a positive correlation between spindle activity and declarative memory consolidation (Genzel et al., 2009). Finally, we analyzed a twin study comparing slow and fast sleep spindle parameters: total count, density, amplitude, duration and frequency between healthy monozygotic (MZ) and dizygotic (DZ) twins.

MATERIALS AND METHODS

Almost all computations were performed using MATLAB 2014a. Only MANCOVA analysis was performed using SPSS v17. The source code is available from the corresponding author.

Validation Sample—Nap Recordings

Our algorithm was validated with data from an earlier study (Genzel et al., 2014). In brief, 20 participants (10 male, age 20–30 years) had two nap sessions in the sleep laboratory separated by at least 4 weeks, one with and one without previous learning experience. For more details regarding study design and participants please see Genzel et al. (2014). Eighteen naps from $n = 10$ subjects were randomly selected and our algorithm was compared with the SIESTA algorithm of Anderer et al. (2005) and with a human scorer. Sleep spindle scoring was performed by a trained research assistant and double-checked by an experienced sleep expert. The experimental protocol was approved by the Ethics Committee of the Ludwigs Maximilian University, Faculty of Medicine, Munich and written informed consent was obtained from the participants.

Sleep-Related Memory Consolidation Sample

The data of the memory consolidation study were described by Genzel et al. (2009). Recruited subjects were $n = 12$ healthy volunteers, six males and six females. Age ranged between 20–30 years. Prerequisites for inclusion and exclusion criteria as well as study protocol are described in detail elsewhere (Genzel et al., 2009). Briefly, the subjects spent six nights in our sleep laboratory, where three nights served as adaptation nights which were followed by study nights. Each experimental session consisted of adaptation night, learning before the study night (declarative memory: finger tapping task, procedural memory: verbal paired associates task), study recording with various experimental sleep conditions [REM sleep deprivation, slow wave sleep (SWS) deprivation and undisturbed night] and a retest after two nights of recovery sleep. EEG recordings from the undisturbed study night were used for sleep spindle analysis. The experimental protocol was approved by the Ethics Committee of the Ludwigs Maximilian University, Faculty of Medicine, Munich and written informed consent was obtained from the participants.

Twin Sample

We analyzed the data of the twin study described by Ambrosius et al. (2008). We recruited $n = 35$ pairs of MZ and $n = 14$ pairs of DZ twins. All twin pairs had been raised together. The twins underwent physical, psychiatric, and laboratory examinations to exclude acute and chronic diseases. Prerequisites for inclusion and determination of zygosity are described in detail elsewhere (Ambrosius et al., 2008). Due to technical reasons (high EEG amplitude differences in consecutive nights) 3 MZ pairs were excluded. All presented results have been obtained from the remaining 32 pairs of MZ twins (mean (SD): 23.8 (4.8) years; range: 17–43 years, 16 male pairs, 16 female pairs) and 14 pairs of DZ twins (22.1 (2.7) years; range: 18–26 years, 7 male pairs, 7 female pairs). Fifteen of thirty-two monozygotic and ten of fourteen dizygotic twin pairs were living together at the time of the examination. The experimental protocol was approved by the Ethics Committee

for Human Experiments of the Bayerische Landesärztekammer (Munich, Germany) and written informed consent was obtained from the participants. The subjects spent three consecutive nights in our sleep laboratory, where the first night served for adaptation and exclusion of sleep disturbances. Almost all twin partners were recorded at the same time. EEG data of the second and third recording night were used for spindle analysis.

EEG Recording

All polysomnographic recordings (Comlab 32 Digital Sleep Lab, Brainlab V 3.3 Software, Schwarzer GmbH, Munich, Germany) were performed according to the international 10–20 electrode system (high-pass filter at 0.53 Hz, low-pass filter at 70 Hz, sampling rate of 250 Hz). Electrooculographic (EOG) montage was done according to Rechtschaffen and Kales (1968). We recorded nap validation samples and memory samples with C3A2 and C4A1 EEG electrodes, whereas twin samples were recorded using 10 EEG electrodes: Fp1A2, Fp2A1, F3A2, F4A1, C3A2, C4A1, P3A2, P4A1, O1A2 and O2A1. Professional scorers scored sleep stages in 30 s epochs according to the standard guidelines (Rechtschaffen and Kales, 1968). Recordings of the twin partners were scored by the same rater.

SIESTA Algorithm

The SIESTA algorithm was described in detail by Anderer et al. (2005). This solution was created using a large database of visually detected sleep spindles (SIESTA database). Briefly, spindle criteria were based on sleep spindle characteristics from the database: length from 0.3–2 s, minimal peak-to-peak amplitude at least 12 μ V and frequency from 11–16 Hz. Authors introduced these criteria to a spindle detector described by Schimicek et al. (1994; briefly described in the introduction). Localized spindle candidates fulfilling minimal criteria were further evaluated with a classifier trained on the SIESTA database. Spindle classification was based on linear discriminant analysis and as an input used spindle duration and mean amplitudes in four frequency bands: spindle, theta, alpha and fast beta. The outcome of each spindle evaluation was a discriminant score, and the SIESTA detector offers three detection thresholds for discriminant scores. If a user chooses the lowest threshold, the algorithm accepts all “possible” spindles. This threshold resulted in 90% detection sensitivity in the SIESTA database. The middle threshold accepts all “probable” spindles. This threshold maximized the agreement with human scorers in the SIESTA database by maximizing the sum of sensitivity and specificity. The highest threshold accepts only “certain” spindles. This threshold resulted in detection specificity above 97% in the SIESTA database.

For both data sets, validation and sleep-related memory consolidation sample, we report results of SIESTA analysis performed with middle detection threshold (“probable” patterns), which seems to balance detection sensitivity and specificity.

Statistical Analysis

Algorithm Validation

Our validation data set consisted of detailed information about the exact placement of each detected sleep spindle for both SIESTA analysis and visual scoring. We compared spindles marked in time using 0.1 s windows to obtain the number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). The problem related to statistical analysis of spindle detection agreement is the fact that the majority of EEG signal usually does not contain spindles, which inflates strongly TN and mildly FP. Due to class imbalance, we report results of multiple agreement measures. First, we calculated sensitivity ($TP/[TP + FN]$), specificity ($TN/[TN + FP]$) and precision ($TP/[TP + FP]$). These measures are commonly used, so we report them for the sake of comparison with other published spindle detectors. However, due to the aforementioned bias, specificity outcomes tend to be strongly overestimated, and precision mildly underestimated. We also calculated the general scoring agreement using measures which should correct for the bias towards long fragments of signal, where there are no sleep spindles: adjusted geometric-mean (Batuwita and Palade, 2012), Matthews correlation coefficient and Cohen's kappa coefficient (equations can be found in Supplementary Material). Adjusted geometric-mean was developed to measure the agreement in imbalanced datasets, where the positive data examples are largely outnumbered by the negative data examples. It adjusts the impact of sensitivity and specificity according to the observed size differences between classes. Matthews correlation coefficient is a geometric mean corrected for chance agreement. It actually returns the same values as Pearson correlation of spindles marked in time between two scorers. Kappa takes the observed agreement and corrects it for a putative chance agreement. There are several benchmarks characterizing agreement based on Cohen's kappa values. According to Landis and Koch (1977) kappa values from 0–0.2 have been termed *slight*, between 0.21 and 0.40 *fair*, between 0.41 and 0.60 *moderate*, between 0.61 and 0.80 *substantial*, and between 0.81–1 as *almost perfect* agreement. In addition, we used Pearson's correlation to obtain subjects-wise spindle density agreement.

Human scorers marked sleep spindles only in stage 2 sleep, since in SWS it is much more difficult to visually detect spindles intermingled into delta waves. For this reason the agreement comparison for stage 2 sleep included visual scoring and automatic algorithms, whereas for SWS we compared only our detector and SIESTA algorithm. We analyzed the agreement of sleep spindles scored in the C3A2 EEG channel.

Sleep-Related Memory Consolidation

For the sleep-related memory consolidation data, we had only a general outcome from SIESTA spindle analysis about each subject, including average spindle density, amplitude and duration in sleep stage 2 and SWS. We used Pearson's correlation to obtain subjects-wise spindle density agreement between algorithms as well as between spindle density and declarative

memory performance. We analyzed sleep spindle activity in the C4A1 EEG channel.

Twin Study—Genetic Variance Analysis

We investigated MZ and DZ twins in order to separate the variance of sleep variables into environmental and genetic components according to Christian et al. (1974, 1987). Briefly, there are two independent estimates of genetic variance: the within-twin pair estimate (GWT), and the combined within- plus among-twin pair component estimate (GCT). GWT depends only on mean squares (MS) for within-pair variation, whereas GCT depends on MS of both within- and among-twin pair variation. A test of equality of variances (F' test) for MZ and DZ twins determines the selection of genetic variance estimate. We used the GCT test when MZ and DZ variances were not equal (the null hypothesis of equal variances was tested using $\alpha = 0.2$, as suggested by the authors). In the other case the GWT test was used. As a prerequisite for the analysis, each studied variable had to fulfill the assumptions of normal distribution (measured by a non-significant goodness-of-fit by the Kolmogorov-Smirnov test) in both twin samples and equal means between twin samples (t -test). The significantly unequal means between MZ and DZ twin samples indicate that the investigated variable could be associated with the type of twins being studied. In this case the estimation of genetic variance would be biased. Therefore, if there was an evidence for significantly unequal means between MZ and DZ twin samples, the GVA was not performed. The influence of covariates (age, sex and cohabitation) was analyzed by MANCOVA. Prerequisites were considered to be violated, if the appropriate test showed a significant result at the 5% level. GVA was performed on the mean results of two recording nights. We include a more detailed description of GVA in the Supplementary Material.

We estimated the genetic influence on the most basic parameters describing sleep spindle activity during the whole night: the absolute number of spindles, spindle density (average number of spindles per 30 s epoch), length, amplitude and mean frequency. In order to minimize the effects of possible covariates, we selected a subgroup of MZ twins closely matched for age, gender and cohabitation to DZ twins. GVA for matched MZ and DZ samples can be found in the Supplementary Material. We analyzed sleep spindle activity in left hemisphere. In the results section we present GVA from F3A2 and P3A2 EEG derivations, analysis from Fp1A2 and C3A2 channels can be found in the Supplementary Material.

Twin Study—ICC Analysis

We illustrate differences between within-twin pair resemblance and night-to-night stability with intraclass correlation coefficients (ICCs). In order to reveal the strength of observed ICC results, we applied bootstrapping analysis as well as providing the interpretation of computed correlations proposed by Landis and Koch (1977). To obtain levels of statistical significance for ICC results we applied bootstrapping analysis similarly to Tarokh et al. (2011). Each sample was recreated by choosing subject values randomly with repetitions up to the same

number as in the original set. For each bootstrapped sample ICC was computed. Only positive ICC values of bootstrapped samples were accepted. Bootstrapping was continued until 1000 positive ICC values were reached. For each investigated parameter we present ICC results of original sample together with the 1/100th top percentile (congruent with significance level $P = 0.01$) and median (congruent to positive ICC values obtained by chance) value of bootstrapped data. Bootstrapping was performed separately for each investigated sample. The sample for within-pair similarity estimation consisted of 64 values in MZ twins (32 twin pairs, 2 values for each pair) and 28 values in DZ twins (32 twin pairs, 2 values for each pair). The sample for stability estimation between consecutive nights consisted of 128 values in the MZ set (32 twin pairs, 2 subjects in each pair, 2 values for each subject) and 56 values in the DZ set (14 twin pairs, 2 subjects in each pair, 2 values for each subject). The smaller the sample size, the easier it is to obtain high ICC by chance. For this reason, bootstrapped ICC values are higher for samples with smaller sizes. According to Landis and Koch (1977), ranges of ICC values were designated as being in *slight* agreement (from 0–0.2), *fair* agreement (from 0.21–0.40), *moderate* agreement (from 0.41–0.60), *substantial* agreement (from 0.61–0.80), and *almost perfect* agreement (from 0.81–1). ICCs estimating within-pair resemblance were performed on mean results of two recording nights.

Automatic Sleep Spindle Detection: Description of the Algorithm

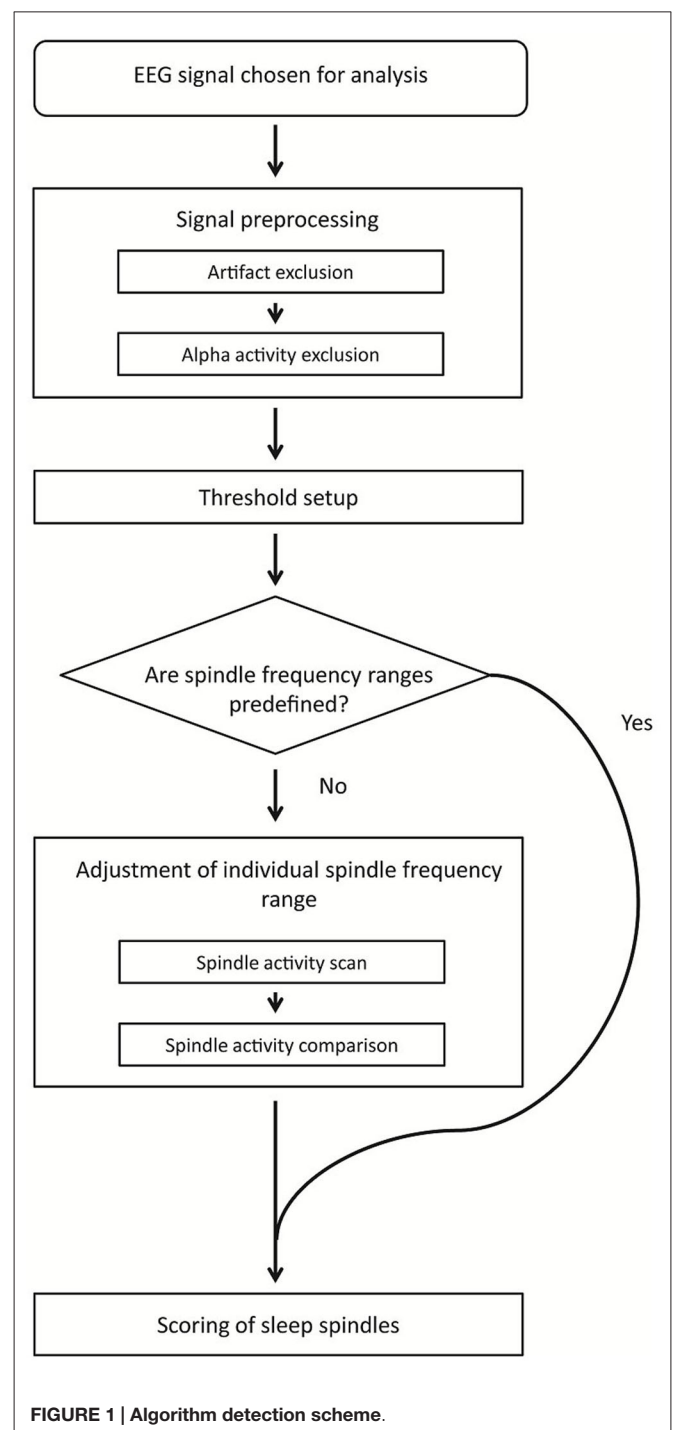
Figure 1 depicts the block diagram of spindle detection procedure. First, our method rejects artifacts and strong alpha activities. The signal chosen for spindle detection without excluded fragments is used in further analysis. The detection threshold is then set separately for each channel. If slow and fast spindle frequency boundaries are not predefined, an automatic adjustment procedure sets them individually for each subject using frontal and parietal EEG channels. When spindle frequency boundaries and detection threshold are set, the algorithm scores sleep spindles.

Preprocessing Before Spindle Detection

To decrease the computation load, algorithm re-samples the signal to 100 Hz. Therefore, the algorithm resolution is 0.01 s. The first part of the algorithm checks the properties of the signal and rejects periods of signal with high muscle contamination as well as segments dominated by alpha activity.

Artifact exclusion

In order to identify fragments with high frequency muscle artifacts, the EEG signal was band-pass filtered (FIR filter; -3 dB at 19.8 and 45.5 Hz). The standard deviation of the signal was computed over a 1 s sliding window (step: 0.5 s) and if it exceeded $5.75 \mu\text{V}$, a window of 7 s (fragment in which the threshold was exceeded ± 3 s) was excluded from further analysis.



Exclusion of segments with strong alpha activity

Alpha activity is present in the EEG signal mostly during wake when the eyes are closed, but can also be present in EEG during shallow sleep, after arousals and during REM sleep. The shape and frequency of alpha waves (long waxing and waning bursts of activity in the range of 8–12 Hz) is similar to sleep spindles and therefore may lead to false spindle detection. To exclude EEG fragments with probable strings of alpha waves, alpha activity was compared with delta activity on long signal fragments. First, the

signal was high-pass filtered (FIR filter; -3 dB at 1.4 Hz). Then, we computed the amplitude spectrum [Fast Fourier Transform (FFT) performed on a 4 s Hanning window; step: 1 s] and for each second mean amplitude was stored for 2 – 4 (delta) Hz and 8 – 12 (alpha) Hz frequency ranges. Alpha and delta activity were compared in a 15 s sliding window (step: 1 s). Fifteen values for both alpha and delta activity were weighted using a Hanning window and then averaged, resulting in α_{activity} and δ_{activity} . Due to the Hanning window, central values in an analyzed fragment had the strongest influence on the outcome. A 15 s fragment was excluded from further analysis if α_{activity} was higher than $1.1 \times \delta_{\text{activity}}$.

The reasoning behind our preprocessing methods is described more in detail in the Supplementary Material.

Threshold Setup

The threshold was computed using exactly the signal chosen for spindle detection, without fragments excluded due to artifacts or strong alpha activity. Our aim was to obtain a basic threshold (BT) value close to signal background activity. We therefore firstly focused on the 6 – 18 Hz frequency range, since frequencies below 6 and above 18 Hz are strongly influenced by sleep quality (amount and strength of delta waves), and could be strongly influenced by artifacts (for example muscle contamination). The signal was band-pass filtered (FIR filter; -3 dB at 5.5 and 18.2 Hz) and amplitude spectra were computed (FFT; 2 s sliding window; step: 2 s). Second, amplitude spectra were logarithm transformed (base 10). Due to this transformation, all peaks in activity had a lower influence on the final outcome. Third, the median over all amplitude spectra was computed in order to obtain the background activity for each frequency bin, since the median should be less influenced by temporary events than a mean. BT was set as a mean background activity in the 6 – 18 Hz range. Two thresholds were defined for spindle detection: minimum spindle activity threshold (SA) and minimum spindle peak threshold (SP). SA was set as 55 times BT, while SP was set as 80 times BT.

Detection of Spindle Events

In order to detect spindle events, we applied the continuous wavelet transform (CWT) to the signal. As a mother wavelet, we used the complex Morlet wavelet which follows the equation:

$$\psi\left(\frac{t-b}{a}\right) = \frac{1}{\pi^{1/4}} e^{i2\pi f_0[(t-b)/a]} e^{-[(t-b)/a]^2}$$

where t is time, a is scale parameter so the mother wavelet can be dilated according to the frequency of interest and shifted across the signal using the location parameter b . Central frequency f_0 influences the frequency of a complex sinusoid inside the wavelet envelope. For our mother wavelet we chose central frequency $f_0 = 2$, since it closely resembles a spindle shape. The example of the mother wavelet is shown on **Figure 2**. A spindle was identified, if the outcome of CWT exceeded SA by a period of at least half a second, and SP at least once. The spindle was marked over the signal fragment, where CWT exceeded SA.

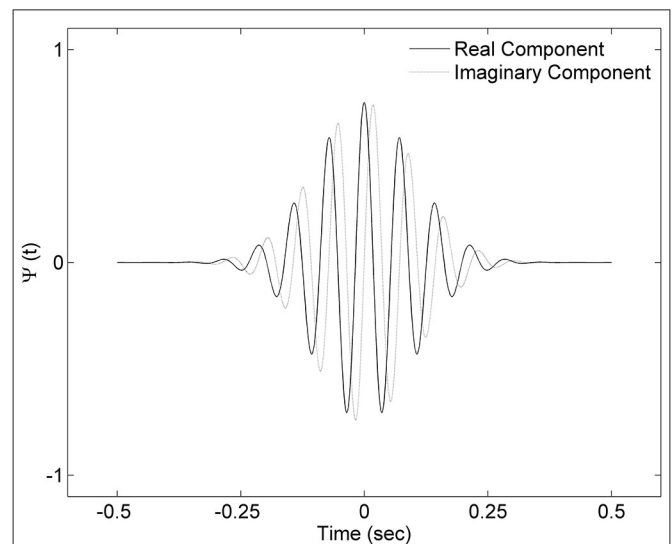


FIGURE 2 | Complex Morlet wavelet with central frequency $f_0 = 2$ used in the analysis. Presented wavelet corresponds to 14 Hz frequency.

Adjustment of Individual Spindle Frequency Range

Slow spindle activity is more prominent in frontal EEG channels and fast spindle activity is more prominent in parietal channels. In order to localize individual ranges of fast and slow spindle frequency, our algorithm scanned spindle events activity in the 9 – 16 Hz frequency range and compared the frequency distribution of spindles detected in frontal and parietal EEG channels. Individual spindle frequency range was computed using exactly the signal chosen for spindle detection, without fragments excluded due to artifacts or strong alpha activity. The example of spindle frequency estimation is illustrated in **Figure 3**.

Spindle activity scan

We performed a spindle activity scan using frontal EEG channel F3A2 and parietal channel P3A2. For each channel, CWT was computed with wavelets corresponding to the 9 – 16 Hz frequency range (step: 0.1 Hz). For the CWT outcome in each frequency bin (CWT_{bin}), fragments fulfilling spindle criteria were marked (outcome of CWT_{bin} exceeded SA by a period of at least half a second, and SP at least once). For each frequency bin, every marked fragment overlapping exactly with the signal section where CWT_{bin} exceeded threshold SA was then investigated. Mean CWT_{bin} over this fragment was computed for each 0.1 Hz frequency bin in the 9 – 16 Hz range (71 bins). A localized fragment was accepted as a spindle belonging to the currently analyzed frequency bin only if currently analyzed frequency was dominant. That is, only if mean CWT_{bin} over this fragment in this frequency was higher than every other mean CWT_{bin} over this fragment for each other 0.1 Hz frequency bin in the 9 – 16 Hz range. If other frequency than currently analyzed was identified as dominant, this fragment was rejected. For each 0.1 Hz frequency bin all accepted spindles were summarized and these sums were combined into a vector of spindle activity over frequency range separately for frontal channel F3A2 and parietal channel P3A2.

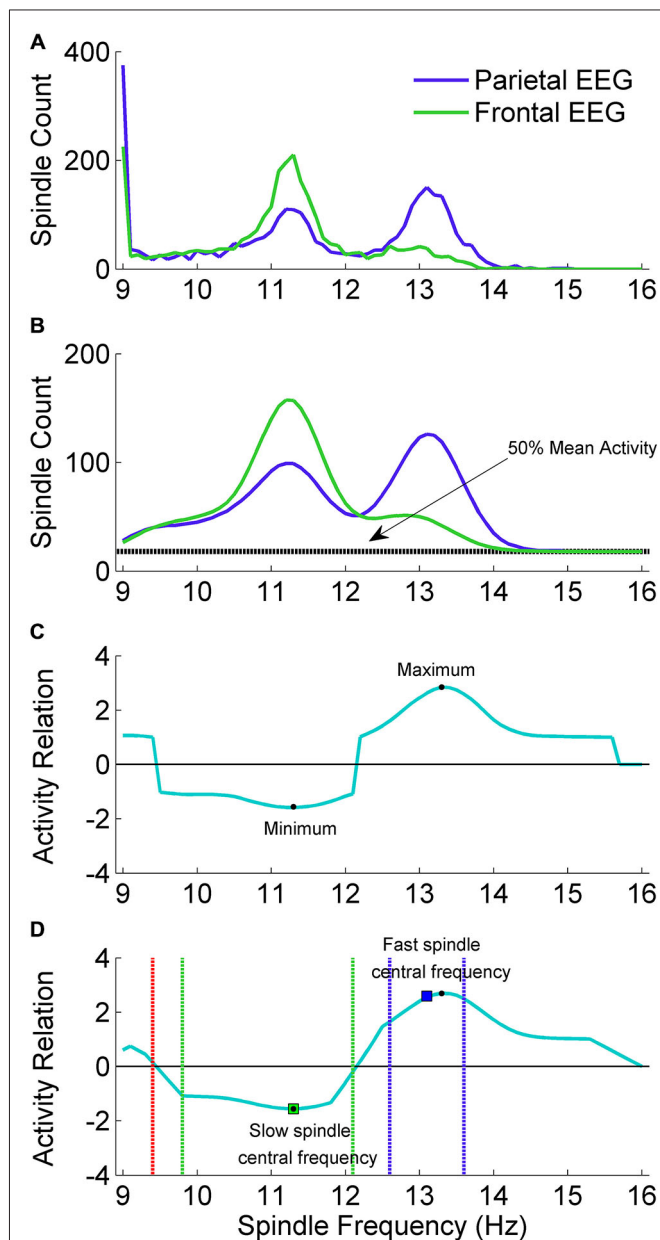


FIGURE 3 | The adjustment scheme of individual spindle frequency range. (A) The outcome of spindle activity scan which resulted in two vectors of spindle activity over frequency range separately for frontal channel F3A2 (vec_{slow} : green color) and parietal channel P3A2 (vec_{fast} : blue color). (B) In both activity vectors the value in 9 Hz was set to zero, vectors were smoothed and 50% of mean spindle activity (dashed black line) was added to both of them. (C) Vector (vec_{rel}) showing a relation of spindle activity between frontal EEG and parietal EEG, computed according to “Spindle Activity Comparison” Section. (D) Smoothed vec_{rel} . First, algorithm localized minimum and maximum (black dots). Localized minimum in vec_{rel} was set as slow spindle central frequency (green square). Localized maximum in vec_{rel} was a starting point to estimate fast spindle frequency ranges using vec_{fast} . Local maximum in vec_{fast} was set as fast spindle central frequency (blue square). Ranges of fast (dashed blue lines) and slow (dashed green lines) spindle frequency were estimated according to “Spindle Activity Comparison” Section. First frequency bin below slow spindle range in which spindle activity was higher in the parietal channel was set as frequency in which slow spindles are unlikely ($stop_{detect}$: red dashed line).

The example outcome of spindle activity scan is illustrated in Figure 3A.

Spindle activity comparison

Spindle activities estimated for frontal and parietal EEG signals were compared to find frequency ranges of slow and fast spindles. Slow spindle activity is more prominent in frontal EEG channels and fast spindle activity is more prominent in parietal channels. For this reason, vector with spindle activity data from frontal EEG channel is called vec_{slow} and vector with spindle activity from parietal EEG is called vec_{fast} . Since 9 Hz was the lowest frequency bin for which spindle activity scan was performed, frequency of spindles detected using wavelet in 9 Hz frequency bin was compared only to higher frequencies. Therefore the 9 Hz frequency bin in both spindle activity vectors (vec_{slow} and vec_{fast}) included spindle bursts in 9 Hz and possibly below. Sleep spindles in such a low frequencies are unlikely. For this reason, the value in both spindle activity vectors responding to 9 Hz was set to zero. Then a moving average (0.7 Hz window) was applied twice for each vector to smooth the data. The example of preprocessed spindle activity vectors is illustrated in Figure 3B.

The next step was to compute a vector (vec_{rel}) showing a relation of spindle activity between vec_{slow} and vec_{fast} . First, we calculated a grand mean ($mean_{act}$) over both activity vectors of average spindle activity for all frequency bins. vec_{rel} was computed according to the following rule:

```

for i = 1 to the number of frequency bins do
  if  $vec_{fast}(i) > vec_{slow}(i)$  do
     $vec_{rel}(i) = [vec_{fast}(i) + 0.5 \times mean_{act}] / [vec_{slow}(i) + 0.5 \times mean_{act}]$ 
  elseif  $vec_{slow}(i) > vec_{fast}(i)$  do
     $vec_{rel}(i) = -[vec_{slow}(i) + 0.5 \times mean_{act}] / [vec_{fast}(i) + 0.5 \times mean_{act}]$ 
  else
     $vec_{rel}(i) = 0$ 
  end if
end for

```

vec_{rel} is positive when there are more spindles in vec_{fast} and negative when there are more spindles in vec_{slow} . 50% of $mean_{act}$ was included to avoid cases when small spindle numbers in vec_{slow} and vec_{fast} produce very high results in vec_{rel} . The example of obtained vec_{rel} is illustrated in Figure 3D.

vec_{rel} was smoothed (moving average, 0.7 Hz window) before localizing slow and fast spindle frequency range. The minimum value in vec_{rel} shows the strongest relative spindle activity in frontal EEG when compared to spindle activity in parietal EEG. Frequency responding to this minimum value was taken as a putative central frequency of slow spindle activity ($slow_{ctr}$). To find a putative central frequency of fast spindle activity ($fast_{ctr}$) algorithm analyzed vec_{rel} in the frequency range between $slow_{ctr}$ and 16 Hz. Frequency responding to the maximum value in vec_{rel} within the $slow_{ctr}$ –16 Hz range was taken as a candidate for $fast_{ctr}$.

Fast spindle activity is usually clearly visible in vec_{fast} . Therefore, $fast_{ctr}$ was shifted from the maximum in vec_{rel} towards the local maximum in vec_{fast} . The range of fast spindle

frequency was estimated similarly to method presented by Bódizs et al. (2009): second derivative of vec_{fast} was computed and zero-crossing points encompassing $\text{fast}_{\text{cntr}}$ were taken as fast spindle frequency ranges.

Frequency ranges of slow spindle activity in vec_{slow} are often difficult to distinguish, so they were estimated using vec_{rel} . The higher boundary was extended from $\text{slow}_{\text{cntr}}$ to the highest frequency below fast spindle frequency range, in which spindle activity was higher in the frontal channel. The lower boundary of slow spindle activity was more difficult to establish, since it is important to avoid classification of alpha waves as sleep spindles. Therefore the lower boundary of slow spindle frequency range was extended cautiously from $\text{slow}_{\text{cntr}}$ to the first frequency bin in which vec_{rel} value was 40% higher than minimum in $\text{slow}_{\text{cntr}}$. In addition, the algorithm set a frequency $\text{stop}_{\text{detect}}$ in which slow spindles are unlikely and should not be detected. $\text{stop}_{\text{detect}}$ was set as the highest frequency below slow spindle frequency range, in which spindle activity was higher in the parietal channel. If such a frequency was not present above 9 Hz, $\text{stop}_{\text{detect}}$ was set at 9 Hz. The example outcome of spindle frequency estimation is illustrated in **Figure 3D**.

The minimum frequency range was set as at least 0.5 Hz around estimated central frequencies of fast and slow spindles. Spindle activity comparison between vec_{slow} and vec_{fast} was performed only if each vector included at least 30 spindles. Otherwise estimation of spindle detection ranges would have low reliability. If the amount of detected spindles was too low, frequency range was set at 13.1–15 Hz for fast, 11–12.9 Hz for slow spindles and $\text{stop}_{\text{detect}}$ at 9 Hz. The result of spindle

frequency estimation as well as spindle detection with applied individual frequency ranges for twin pair number 10 is illustrated in **Figure 4**.

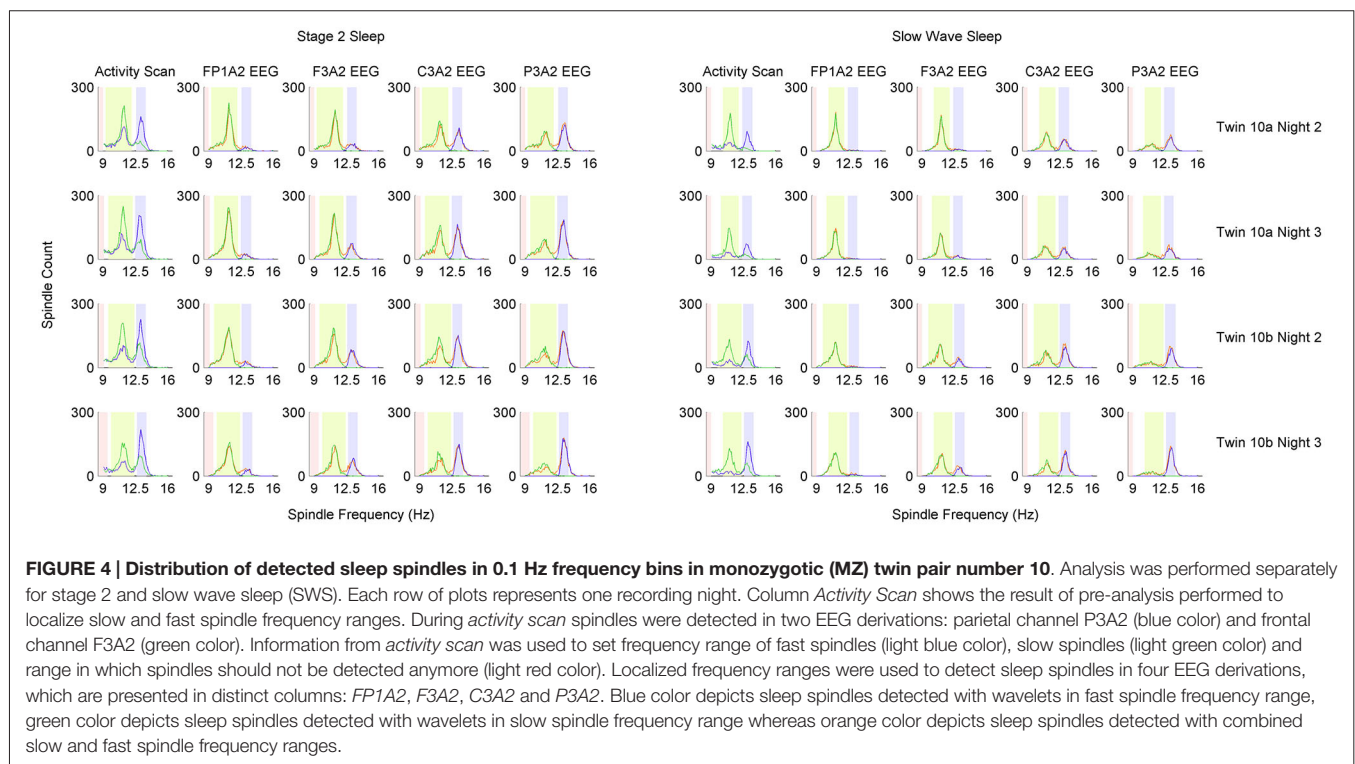
We applied automatic individual adjustment of spindle frequency range in the twin sample, since in this data set recordings include multiple EEG derivations along the antero-posterior axis. However, experiments in which sleep spindle analysis is of interest often include recordings with few EEG channels. Our validation sample and sleep-related memory consolidation sample included only central EEG derivations C3A2 and C4A1. Therefore, in our algorithm the user has the option to set the frequency range for slow and fast spindles. For all recordings in the validation sample and sleep memory consolidation sample, we set 11–12.9 Hz as slow and 13.1–16 Hz as fast spindle frequency range.

Scoring of Sleep Spindles

In order to score sleep spindles, the algorithm analyzed results of CWT computed with wavelets corresponding to $\text{stop}_{\text{detect}}$ frequency, slow spindle frequency range (CWT_{slow}) and fast spindle frequency range (CWT_{fast}). We computed CWT_{slow} in each time point as a maximum CWT value in this time point over slow spindle frequency range. CWT_{fast} was computed the same way. In addition to slow and fast spindles, sleep spindles without distinction between slow and fast ones were scored (all sleep spindles).

All sleep spindles

All sleep spindles were detected using the maximum of both CWT_{slow} and CWT_{fast} (CWT_{all}). Places fulfilling spindle criteria



(according to “Detection of Spindle Events” Section) for CWT_{all} were localized. A marked place was accepted as a sleep spindle, if over this place the mean CWT_{all} was higher than the mean CWT of $stop_{detect}$.

Fast sleep spindles only

Fast sleep spindles were detected using CWT_{fast} . Places in which CWT_{fast} was continuously higher than CWT_{slow} and spindle criteria for CWT_{fast} were fulfilled (according to “Detection of Spindle Events” Section), were classified as fast sleep spindles.

Slow sleep spindles only

Slow sleep spindles were detected using CWT_{slow} . The algorithm localized fragments in which CWT_{slow} was continuously higher than CWT_{fast} and spindle criteria for CWT_{slow} were fulfilled (according to “Detection of Spindle Events” Section). A marked fragment was classified as a slow sleep spindle, if over this place the mean CWT_{slow} was higher than the mean CWT of $stop_{detect}$.

Results of spindle detection in four EEG channels along the antero-posterior axis for twin pair number 10 are illustrated in **Figure 4**. We included such figures for all analyzed twin pairs in the Supplementary Material. An example of sleep spindle detection on EEG fragment is presented in **Figure 5**.

Twin data included frontal and parietal EEG channels, therefore we could apply our automatic individual spindle frequency adjustment and report results from fast and slow spindle detection. In contrast, the validation set as well as memory consolidation data included only central electrodes. For this reason, in these two datasets we used fixed spindle frequency ranges and analyzed results of all sleep spindles detected, without distinction between slow and fast ones.

Computation of Sleep Spindle Amplitude and Frequency

In order to estimate sleep spindle amplitude and dominant frequency, the signal was first band-pass filtered (FIR filter;

−3 dB at 8.7 and 18.5 Hz). Then, a Hanning window was applied to exact a fragment with a marked spindle, and an amplitude spectrum was computed similarly to Huupponen et al. (2006): the fragment was zero-padded to 10 s window and FFT was computed resulting in frequency resolution of 0.1 Hz. The maximum peak in the amplitude spectrum was taken as spindle amplitude and frequency.

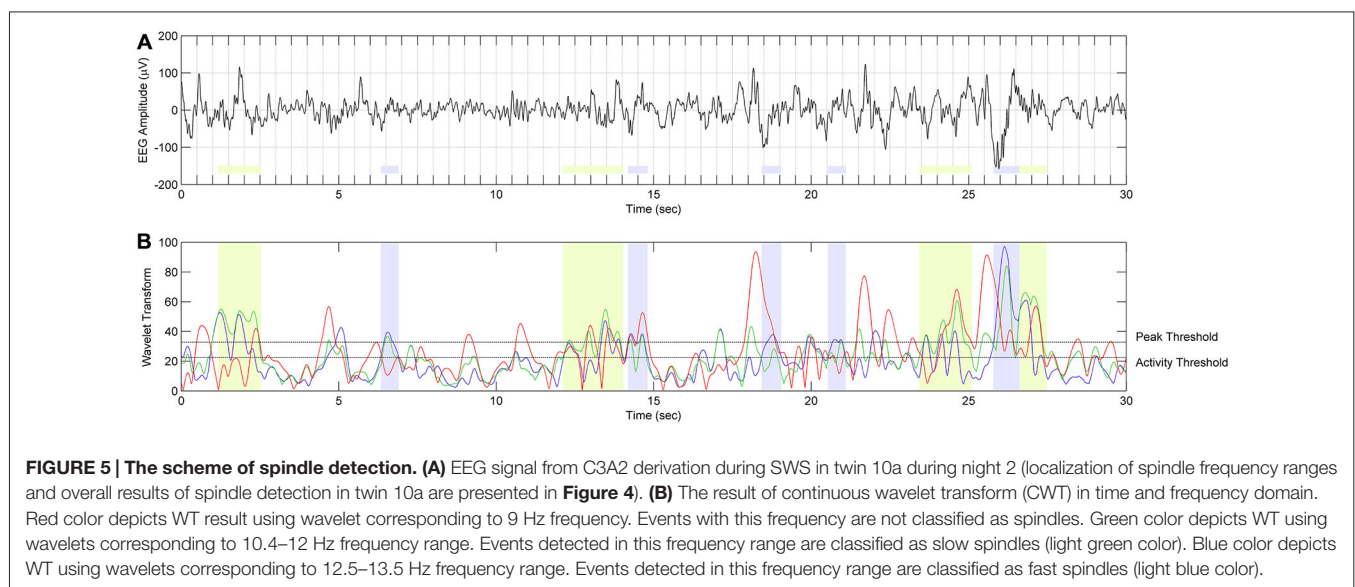
Average Spindle Detection Time

The time required to perform the spindle detection for the whole night EEG recording (around 8 h of sleep) in four EEG channels, with spindle detection ranges individually adjusted using one frontal and one parietal channel, was around 4 min 15 s. When spindle detection ranges were fixed, the CWT algorithm required around 2 min to perform spindle detection in four EEG channels. We performed the analysis using an Intel i5-4310M processor (2.7 GHz, 3 MB).

RESULTS

Algorithm Validation

Our choice of the mother wavelet as well as detection thresholds ratio (spindle activity threshold SA and minimum spindle peak threshold SP) was based on visual observation of sleep spindles and their CWT transforms. We set the actual values of detection thresholds on a level which matched detection sensitivity presented by the SIESTA algorithm. **Figure 6** shows the precision and sensitivity results from a validation dataset of the CWT detector vs. human and vs. SIESTA algorithm using a range of detection threshold levels. We always changed both thresholds percentage-wise, to keep their ratio intact ($SP = 1.45 \times SA$). Results show that a similar amount of detected spindles between our algorithm and SIESTA detector resulted in the highest possible combination of sensitivity and precision. Also, in order to maximize the agreement with a human scorer, we would



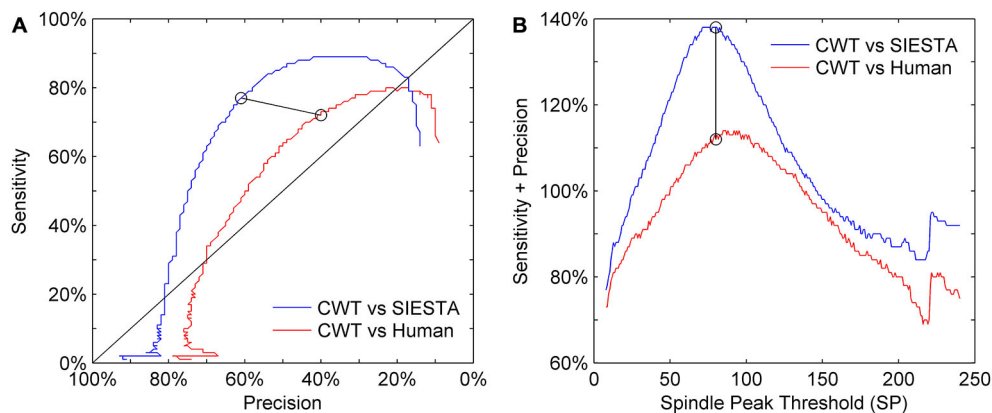


FIGURE 6 | Sensitivity-precision plot showing how these two measures depend on spindle detection thresholds. Sleep spindles were scored in C3A2 EEG channel. **(A)** ROC-like plot of sensitivity vs. precision, **(B)** the sum of sensitivity and precision according to detection thresholds variety. We had two detection thresholds in our algorithm: spindle peak threshold (SP) set as 80 times basic threshold (BT) and spindle activity threshold (SA) set as 55 times BT (calculation of BT is described in “Threshold Setup” Section). We illustrate how performance changes according to SP, where y axis shows multiplication rate of BT used to obtain SP, but for each iteration values of both thresholds were changed together to always keep the same ratio between them ($SP = 1.45 \times SA$). Black circles connected with black line mark sensitivity and precision obtained for thresholds chosen for our algorithm.

need to raise the thresholds by 10%. However, the agreement of our detector with a human would be still much lower than the agreement between two machines.

The agreement between our algorithm, human scorer and SIESTA algorithm on the validation data set is illustrated in **Figure 7** and summarized in **Table 1**. During stage 2 sleep, mean spindle density was 4.0 for our algorithm, 3.95 for SIESTA detector and 2.5 for human scorer. The subject-wise correlation of spindle density between our detector and SIESTA was $r = 0.86$. The correlation of spindle density between our detector and human scoring was $r = 0.73$, whereas the correlation between the SIESTA detector and human scorer was $r = 0.55$. Due to the fact that amounts of NREM sleep stages differed significantly between recordings, we computed our agreement measures using weighted averages, where weight for each recording was its number of investigated sleep epochs divided by the total number of investigated sleep epochs from all recordings. The agreement between our detector and SIESTA algorithm measured with kappa ranged from 0.31–0.74, with weighted average kappa of 0.62 (sensitivity: 0.77; specificity: 0.93; precision: 0.61). The kappa between our detector and human scorer ranged from 0–0.62 with weighted average of 0.45 (sensitivity: 0.72; specificity: 0.90; precision: 0.40) and kappa between the SIESTA detector and human scorer ranged from 0.08–0.54 with weighted average of 0.44 (sensitivity: 0.62, specificity: 0.92, precision: 0.43). We observed very similar results when using Matthews correlation, with high agreement between automatic detectors when compared to the agreement between algorithms and human scorer. Discrepancies between machines and human were smaller when the agreement was measured using adjusted geometric mean. The reason for that is the human scorer marked the smallest amount of spindles in the signal, resulting in the strongest imbalance between classes. As a result, specificity in this case

had the strongest influence on the outcome of the adjusted geometric mean (equations can be found in the Supplementary Material).

According to published benchmarks for kappa coefficient (Landis and Koch, 1977) the agreement between our algorithm and SIESTA detector was *fair* for two naps, *moderate* for nine naps and *substantial* for seven naps. The agreement between our algorithm and human scorer was *fair* for five naps, *moderate* for 11 naps, *substantial* for one nap and in the one case, there was no agreement between our algorithm and human scorer. The agreement between human scorer and SIESTA detector was *slight* for three naps *fair* for four naps and *moderate* for 11 naps.

SWS was not present in two nap recordings, so the validation set consisted of 16 naps from 9 subjects. During SWS, mean spindle density was 4.03 for our algorithm and 4.56 for SIESTA detector. The subject-wise correlation of spindle density between our detector and SIESTA was $r = 0.80$. According to kappa coefficient, the agreement between our detector and SIESTA algorithm ranged from 0.35–0.86, with weighted average kappa of 0.56 (sensitivity: 0.64; specificity: 0.94; precision: 0.66). The agreement between our algorithm and SIESTA detector was *fair* for three naps, *moderate* for four naps, *substantial* for eight naps and *almost perfect* for one nap.

Since the agreement between scorers was mostly *moderate*, we tried to reveal the reasons for disagreement between scorers by investigating in detail the group of consensus spindles, which were marked by all scorers, as well as distinct groups of spindles marked by only one scorer. To assume that scorers agreed on a spindle, at least 0.3 s consecutive marked fragment had to overlap. We chose this length since 0.3 s was the shortest spindle length marked by scorers. We analyzed spindles detected during stage 2 sleep. **Figure 8** shows overlap between scorers in marked spindles. All spindles were measured as described in “Computation of Sleep Spindle

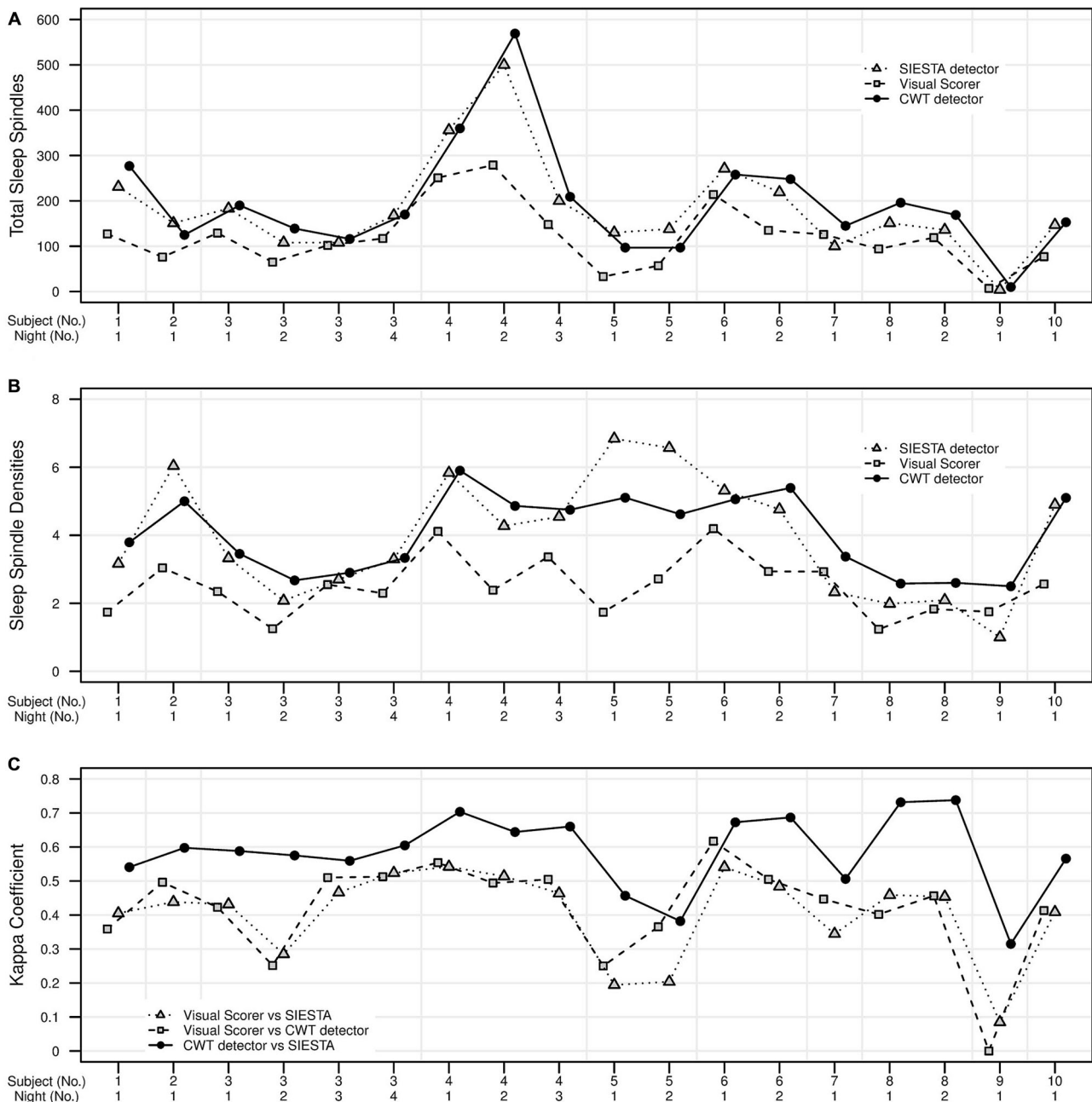


FIGURE 7 | Validation set of 18 nap EEG recordings. Agreement in sleep spindle detection during stage 2 in C3A2 EEG derivation between our algorithm, human visual scorer and SIESTA automatic spindle detector. On y axis there are presented: **(A)** total number of detected sleep spindles for each recording, **(B)** spindle density for each recording, **(C)** the kappa coefficient of scorers agreement for each recording. Subject id and number of nap recording are presented on x axis.

Amplitude and Frequency” Section. Amplitudes in other frequency ranges were computed using similar technique, however without pre-filtering of the signal. Results are presented in **Table 2**.

Consensus spindles could be characterized as the ones with high amplitude (12.51 μV in amplitude spectrum), high frequency (clearly above 13 Hz) and strong activity when

compared to the background. Spindles marked only by a single scorer, conversely, had significantly lower amplitudes, frequencies and spindle to background activity ratio. Our CWT detector marked the highest number of spindles not scored by the others ($N = 726$). It was 20% of all spindles marked by our algorithm. Spindles detected only by our detector had the lowest average frequency (11.94 Hz) and the highest activity

TABLE 1 | Spindle detection agreement between our CWT detector, SIESTA algorithm and human scorer.

Agreement measure	CWT vs. SIESTA		CWT vs. Human		SIESTA vs. Human	
	Stage 2	SWS	Stage 2		Stage 2	
Sensitivity	0.77	0.64	0.72		0.62	
Specificity	0.93	0.94	0.90		0.92	
Precision	0.61	0.66	0.40		0.43	
Kappa	0.62	0.56	0.45		0.44	
Adjusted geometric mean	0.88	0.85	0.85		0.83	
Matthews correlation	0.63	0.57	0.48		0.46	
Spindle density correlation	0.86	0.80	0.73		0.55	

Sleep spindles were scored in the C3A2 EEG channel. CWT, continuous wavelet transform; SWS, slow wave sleep.

in delta and theta frequency ranges. Only 11% of spindles marked by the human scorer were not detected by any automatic algorithm. The average frequency of these spindles was close to the ones marked only by the CWT detector (12.05 Hz). Furthermore, spindles marked only by the human scorer had the lowest amplitude in the amplitude spectrum when compared to automatic detectors (7.92 μV), and were the longest. It means that they often in those cases marked longer fragments than the actual spindle activity. The SIESTA algorithm marked 17% of spindles which were not detected by others. Spindles marked only by the SIESTA algorithm had high average frequency (above 13 Hz) as well as a relatively high amplitude and high activity when compared to the background. These spindles were the ones that on average resembled consensus spindles the most, so the question was: why they were not marked by both the human scorer and the CWT detector? The most important reason was that these spindles were on average the shortest (0.74 s). Spindles with this length should be detected, but

TABLE 2 | Characteristics of sleep spindles detected by all scorers (consensus) and of spindles detected only by a single scorer in stage 2 sleep.

	Consensus (N = 1556) Mean (SD)	Only CWT (N = 726) Mean (SD)	Only SIESTA (N = 583) Mean (SD)	Only Human (N = 243) Mean (SD)
Spindle amplitude (μV) ^a	12.51 (3.46)	8.54 (2.30)	9.24 (3.54)	7.92 (4.00)
Spindle frequency (Hz)	13.75 (1.15)	11.94 (1.74)	13.10 (1.23)	12.05 (2.70)
Spindle length (s)	0.88 (0.33)	0.81 (0.42)	0.74 (0.41)	0.99 (0.47)
Delta (μV) ^b	6.74 (4.16)	10.22 (7.44)	7.26 (4.74)	7.45 (6.38)
Theta (μV) ^b	4.61 (2.30)	6.56 (4.22)	4.82 (2.95)	5.95 (4.93)
Alpha (μV) ^b	3.80 (1.94)	5.10 (2.33)	4.34 (3.18)	6.27 (3.98)
Spindle to background ratio ^c	2.56 (1.15)	1.22 (0.58)	1.73 (0.72)	1.22 (0.64)

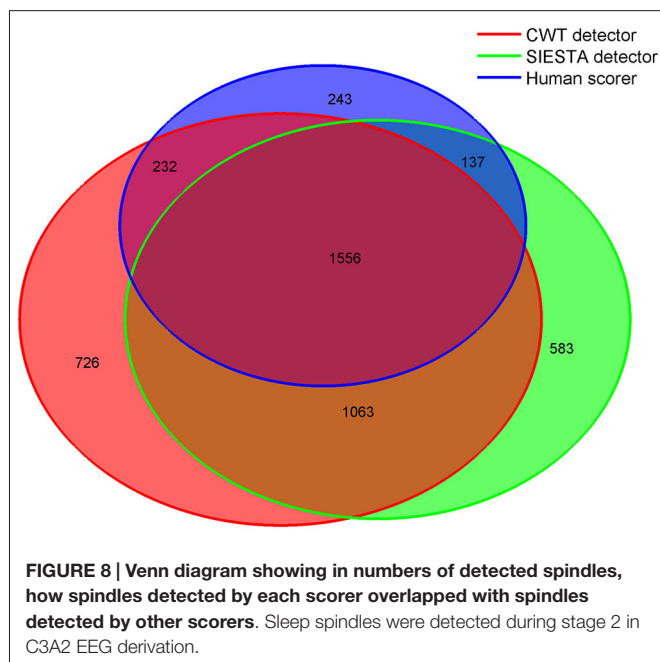
Sleep spindles were scored in the C3A2 EEG channel. ^aComputed from amplitude spectrum as described in "Computation of Sleep Spindle Amplitude and Frequency" Section. ^bMean amplitude in chosen frequency from amplitude spectrum (delta: 2–4.5 Hz, theta: 4.6–7.5 Hz, alpha: 7.6–11 Hz). ^cSpindle amplitude divided by mean amplitude in 2–11 Hz background frequency computed from amplitude spectrum.

24.7% of spindles detected only by the SIESTA algorithm were shorter than half a second. According to our rules, spindles shorter than 0.5 s were not detected. Furthermore, the shortest spindles marked by the SIESTA algorithm also had the highest amplitudes. There was a moderately strong negative correlation, in spindles detected only by the SIESTA detector, between spindle length and amplitude ($r = -0.53$ compared to $r = -0.24$ in spindles detected only by CWT detector). Due to the fact that many spindles marked by SIESTA were short and therefore could be missed, we analyzed just spindles whose length was at least 0.7 s and which were detected just by our algorithm or by the SIESTA detector. Results are presented in Table 3.

TABLE 3 | Characteristics of sleep spindles detected only by the CWT detector and the SIESTA detector, whose length was at least 0.7 s in stage 2 sleep.

	Only CWT (N = 398) Mean (SD)	Only SIESTA (N = 250) Mean (SD)
Spindle amplitude (μV) ^a	8.21 (2.44)	7.80 (3.41)
Spindle frequency (Hz)	11.90 (1.77)	13.02 (1.41)
Spindle length (s)	1.01 (0.48)	1.10 (0.38)
Delta (μV) ^b	9.91 (7.20)	6.28 (3.83)
Theta (μV) ^b	6.13 (3.84)	4.02 (2.12)
Alpha (μV) ^b	4.74 (2.21)	3.63 (2.45)
Spindle to background ratio ^c	1.23 (0.60)	1.73 (0.72)

Sleep spindles were scored in the C3A2 EEG channel. ^aComputed from amplitude spectrum as described in "Computation of Sleep Spindle Amplitude and Frequency" Section. ^bMean amplitude in chosen frequency from amplitude spectrum (delta: 2–4.5 Hz, theta: 4.6–7.5 Hz, alpha: 7.6–11 Hz). ^cSpindle amplitude divided by mean amplitude in 2–11 Hz background frequency computed from amplitude spectrum.



Characteristics of “long” sleep spindles detected only by our CWT detector (average spindle amplitude, frequency and background activity) were very similar when compared to all spindles marked only by our algorithm. In “long” sleep spindles detected only by the SIESTA algorithm we observed a 15% drop in spindle amplitude, whereas their average frequency remained high and the ratio of their activity to the background remained the same, when compared to all spindles marked only by the SIESTA detector. We conclude that spindles marked only by our algorithm were slower and/or intermingled into other frequencies while spindles marked by the SIESTA detector were either short or had too low an amplitude for other scorers. Spindles marked only by the human scorer were few, characterized by slower frequency and a length longer than the actual spindle activity.

We investigated the performance of our CWT detector using the validation set, which consisted of recordings with only central derivations available. For this reason, we used fixed spindle detection frequency ranges and we did not distinguish between slow and fast spindles, but we analyzed all sleep spindles only. Unfortunately, we could not directly evaluate the performance of the CWT detector with individually adjusted spindle frequency ranges vs. other scorers. To get the impression how adjusted frequency ranges would affect the detection, we compared the agreement of the CWT detector with itself when using fixed spindle frequency ranges vs. individually adjusted spindle frequency ranges. We analyzed the second recording night of our twin data. Our results include pooled detection agreement from Fp1A2, F3A2, C3A2, and P3A2 channels. Results are presented in **Table 4**.

The agreement was higher for stage 2 sleep when compared to SWS. The reason was that the algorithm with adjustable frequency ranges detected significantly more spindles during SWS when compared to fixed frequency ranges. The agreement was also high when we considered all sleep spindles together. Mean all spindle density was 4.02 during stage 2 and 4.39 during SWS for algorithm with adjustable frequency ranges compared to 3.96 during stage 2 and 3.47 during SWS for algorithm with fixed frequency ranges. During stage 2 the agreement was

almost perfect, however during SWS it dropped to *substantial*. The agreement dropped significantly when the CWT detector made a distinction between slow and fast spindles. Mean slow spindle density was 2.05 during stage 2 and 3.15 during SWS for the algorithm with adjustable frequency ranges compared to 2.46 during stage 2 and 2.51 during SWS for the algorithm with fixed frequency ranges. The agreement during both, stage 2 sleep and SWS was *substantial*. Mean fast spindle density was 1.64 during stage 2 and 0.88 during SWS for algorithm with adjustable frequency ranges compared to 1.22 during stage 2 and 0.58 during SWS for algorithm with fixed frequency ranges. During stage 2 the agreement was *substantial* and during SWS it dropped to *moderate*.

Sleep-Related Memory Consolidation Data

Mean sleep spindle density during stage 2 sleep was 4.46 for our algorithm and 4.0 for the SIESTA detector, whereas during SWS it was 3.43 and 3.56, respectively. Sleep spindle analysis performed with the SIESTA detector was already described by Genzel et al. (2009). Results returned by the SIESTA algorithm revealed significant Pearson's correlation between spindle density and declarative memory consolidation (stage 2 sleep: $r = 0.627$, $P = 0.015$; SWS: $r = 0.516$, $P = 0.043$). Results returned by our algorithm confirmed previous findings in terms of a significant relationship between spindle density and declarative memory consolidation (stage 2 spindles: $r = 0.579$, $P = 0.024$; SWS: $r = 0.585$, $P = 0.023$). **Figure 9** shows the relation between memory consolidation and spindle density. The subject-wise correlation of spindle density between our detector and SIESTA was $r = 0.93$ for stage 2 sleep and $r = 0.80$ for SWS.

As for spindle activity (absolute number of spindles per night \times mean spindle amplitude \times mean spindle duration), **Table 5** shows correlations between declarative memory consolidation and spindle parameters included in spindle activity calculations. During stage 2 sleep, spindle activity obtained from the SIESTA detector were significantly related to declarative memory consolidation ($r = 0.616$, $P = 0.017$; Genzel et al., 2009). However, the relationship of declarative memory consolidation and spindle activity computed using our algorithm was only marginally significant ($r = 0.468$, $P = 0.062$). In SWS, spindle activity obtained from both algorithms was in marginal relationship with declarative memory consolidation (our algorithm: $r = 0.420$, $P = 0.087$; SIESTA: $r = 0.419$, $P = 0.087$). The subject-wise correlation of spindle activity between our detector and SIESTA was $r = 0.94$ for stage 2 sleep and $r = 0.93$ for SWS.

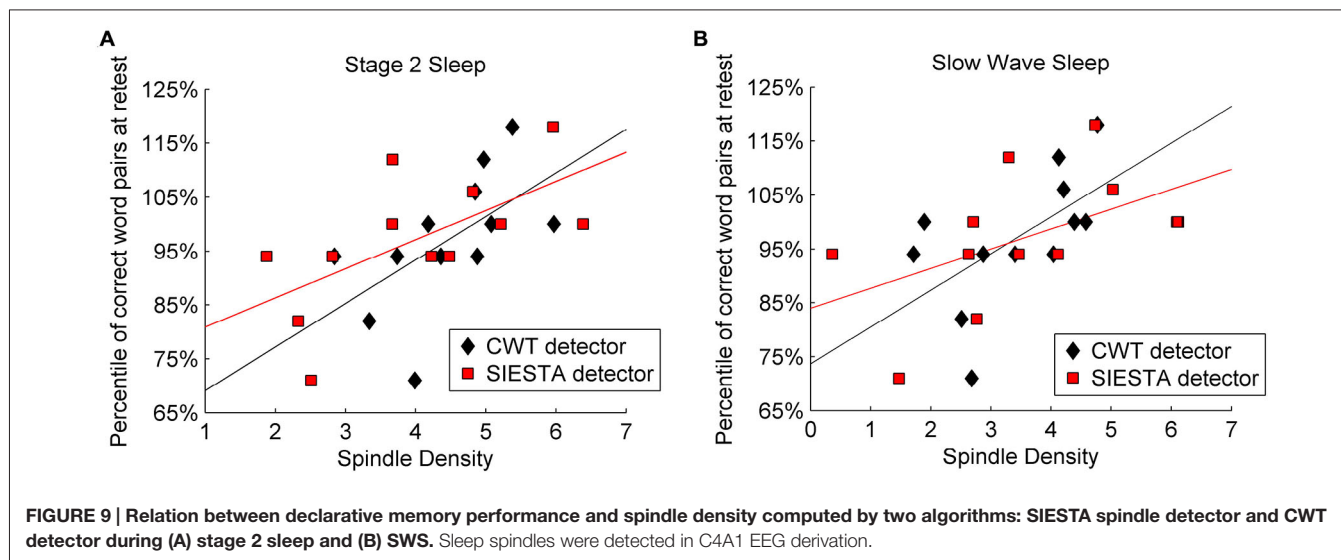
Genetic Influence on Sleep Spindles

Here we report the results of spindle detection with individually adjusted spindle frequency ranges. All estimated frequency ranges for each twin pair can be found in the Supplementary Material (Tables S1–S3 and Figures S1–S46). GVA of sleep spindles detected with fixed spindle frequency ranges are also included in the supplement. We applied individual adjustment of slow and fast spindle frequency ranges separately for stage 2 sleep and SWS. The average frequency of slow spindles detected during

TABLE 4 | Twin set, night 2.

Agreement measure	Slow spindles		Fast spindles		All spindles	
	Stage 2	SWS	Stage 2	SWS	Stage 2	SWS
Sensitivity	0.82	0.66	0.64	0.52	0.90	0.74
Specificity	0.97	0.98	0.99	0.99	0.99	0.99
Precision	0.67	0.76	0.80	0.78	0.92	0.95
Kappa	0.69	0.61	0.64	0.53	0.89	0.77
Adjusted geometric mean	0.93	0.87	0.87	0.82	0.96	0.91
Matthews correlation	0.70	0.64	0.67	0.58	0.89	0.80
Spindle density correlation	0.74	0.73	0.72	0.71	0.93	0.74

Spindle detection agreement between our CWT detector with fixed frequency ranges (slow spindle: 11–12.9 Hz, fast spindle: 13.1–16 Hz) compared to the same detector with individually adjusted spindle frequency ranges. Agreement was calculated from pooled channels Fp1A2, F3A2, C3A2, and P3A2. CWT, continuous wavelet transform; SWS, slow wave sleep.



stage 2 sleep was 11.43 Hz with inter-subject variability ranging from 10.04–12.37 Hz. During SWS, the average frequency of slow spindles was 10.99 with 9.62–12.27 Hz inter-subject range. The average frequency of fast spindles detected during stage 2 sleep was 13.59 Hz with inter-subject variability ranging 12.30–14.83 Hz. During SWS, the average frequency of fast spindles was 13.55 with 12.26–14.73 Hz inter-subject range.

The criterion of normal distribution was not fulfilled for the average slow spindle length during stage 2 sleep in the F3A2 EEG channel, therefore it was log transformed prior to all analyses. We observed that age, as a covariate, had a marginally significant effect on fast spindle density (higher spindle density in younger subjects), and sex, as a covariate, had a marginally significant effect on slow spindle number (higher spindle number in females). Sample means of averaged over-pairs measures revealed no significant night effects (Supplementary Material, Tables S4, S6, S8 and S10). However, in the F3A2 derivation, we observed significantly higher slow spindle amplitude in DZ twins during stage 2 sleep as well as significantly higher slow spindle absolute number and density in DZ twins during SWS (Supplementary Material, Tables S8 and S10). Therefore, for

these three slow spindle parameters GVA was not applicable. In both, stage 2 sleep and SWS, we identified a significant genetic influence on variance of all but one remaining slow spindle parameter. The exception was the average slow spindle frequency in the F3A2 channel during SWS, on which the genetic effect was only marginally significant. **Tables 6, 7** depict GVA of sleep spindle parameters during stage 2 sleep and SWS, respectively.

Considering fast sleep spindles, GVA revealed significant genetic control on variance of spindle length, amplitude and frequency during both stage 2 sleep and SWS. However, we found no genetic effects on fast spindle number and density during stage 2 sleep, whereas during SWS in the P3A2 channel genetic influence on variance was significant on fast spindle number (the effect was weak: $P = 0.049$), and only marginally significant on fast spindle density.

The mean ICC of all slow spindle parameters for night-to-night stability was similar in both groups: 0.91 in the MZ set compared to 0.88 in the DZ set. All these values were above the significance threshold ($P = 0.01$) set by bootstrapping analysis. According to the Landis and Koch (1977) benchmark, night-to-night stability in the MZ set was *almost perfect* for all but one slow spindle characteristic (it was *substantial* for spindle number in P3A2 channel during SWS). Night-to-night stability in the DZ set was *almost perfect* for all but four slow spindle parameters. It was *substantial* for spindle number in the F3A2 channel during SWS as well as for spindle amplitude in the P3A2 channel during stage 2 sleep, as well as for spindle amplitude in both channels during SWS. The mean ICC of all slow spindle parameters for within-pair resemblance was 0.91 in MZ twins and 0.35 in DZ twins. In the MZ set, within-pair similarity was always above the significance level, and according to the benchmark within-pair similarity was *almost perfect* for all slow spindle parameters. In the DZ set however, within-pair similarity was below the significance level for all parameters besides spindle frequency during SWS. In addition, within-pair similarity for

TABLE 5 | Pearson's correlation between declarative memory consolidation and spindle characteristics.

	CWT detector		SIESTA detector	
	Stage 2	SWS	Stage 2	SWS
Spindle density	0.58	0.59	0.63	0.52
Spindle absolute number	0.45	0.48	0.60	0.49
Spindle amplitude (μV) ^a	0.31	0.07	0.45	0.28
Spindle length (s)	0.22	0.36	0.50	0.28
Spindle activity ^b	0.47	0.42	0.62	0.42

Sleep spindles were scored in the C4A1 EEG channel. ^aComputed from amplitude spectrum as described in "Computation of Sleep Spindle Amplitude and Frequency" Section. ^bSpindle activity: absolute number of spindles per night \times mean spindle amplitude \times mean spindle duration.

TABLE 6 | Genetic variance analysis, type of estimate applied (GCT: combined among- and within-twin pair component estimate, GWT: within-pair estimate) and intraclass correlation coefficients (ICCs) for spindle parameters in stage 2 sleep.

Variable	Type	P	Analysis	ICC MZ	ICC DZ	ICC MZ cn	ICC DZ cn
<i>EEG channel: F3A2</i>							
Number of spindles	Slow	<0.001	GWT	0.91 (0.47, 0.12)	0.43 (0.62, 0.18)	0.92 (0.33, 0.09)	0.88 (0.49, 0.14)
	Fast	0.279	GWT	0.75 (0.44, 0.13)	0.49 (0.61, 0.18)	0.85 (0.32, 0.08)	0.92 (0.49, 0.12)
Spindle density	Slow	0.001	GCT	0.94 (0.45, 0.13)	0.24 (0.65, 0.19)	0.94 (0.31, 0.09)	0.91 (0.46, 0.13)
	Fast	0.164	GWT	0.78 (0.48, 0.11)	0.54 (0.67, 0.18)	0.86 (0.32, 0.08)	0.94 (0.50, 0.13)
Spindle length	Slow	0.002	GCT	0.96 (0.45, 0.12)	0.42 (0.65, 0.19)	0.96 (0.32, 0.09)	0.92 (0.48, 0.13)
	Fast	0.030	GWT	0.74 (0.46, 0.12)	0.40 (0.61, 0.17)	0.82 (0.33, 0.08)	0.91 (0.46, 0.13)
Spindle amplitude	Slow*	—	—	0.88 (0.46, 0.12)	0.19 (0.65, 0.18)	0.91 (0.34, 0.09)	0.88 (0.48, 0.13)
	Fast	<0.001	GCT	0.88 (0.45, 0.13)	0.10 (0.62, 0.16)	0.88 (0.32, 0.08)	0.74 (0.50, 0.13)
Spindle frequency	Slow	<0.001	GWT	0.94 (0.43, 0.12)	0.43 (0.62, 0.18)	0.93 (0.32, 0.09)	0.96 (0.48, 0.13)
	Fast	<0.001	GWT	0.93 (0.42, 0.12)	0.67 (0.64, 0.18)	0.96 (0.33, 0.09)	0.96 (0.48, 0.14)
<i>EEG channel: P3A2</i>							
Number of spindles	Slow	<0.001	GCT	0.96 (0.53, 0.11)	0.22 (0.69, 0.19)	0.93 (0.34, 0.08)	0.88 (0.48, 0.13)
	Fast	0.271	GWT	0.80 (0.45, 0.12)	0.70 (0.68, 0.19)	0.83 (0.34, 0.09)	0.86 (0.46, 0.13)
Spindle density	Slow	<0.001	GCT	0.96 (0.53, 0.11)	0.04 (0.68, 0.18)	0.94 (0.36, 0.08)	0.91 (0.48, 0.13)
	Fast	0.196	GWT	0.80 (0.44, 0.12)	0.67 (0.66, 0.19)	0.85 (0.32, 0.09)	0.88 (0.44, 0.12)
Spindle length	Slow	<0.001	GCT	0.94 (0.52, 0.12)	−0.19 (0.66, 0.19)	0.90 (0.37, 0.08)	0.81 (0.47, 0.13)
	Fast	0.002	GWT	0.78 (0.45, 0.12)	0.48 (0.62, 0.19)	0.89 (0.33, 0.08)	0.92 (0.47, 0.14)
Spindle amplitude	Slow	0.005	GWT	0.84 (0.46, 0.13)	0.50 (0.61, 0.17)	0.88 (0.31, 0.09)	0.79 (0.47, 0.12)
	Fast	0.047	GWT	0.82 (0.48, 0.13)	0.57 (0.67, 0.17)	0.88 (0.33, 0.09)	0.64 (0.48, 0.12)
Spindle frequency	Slow	<0.001	GWT	0.94 (0.44, 0.12)	0.41 (0.64, 0.18)	0.90 (0.34, 0.09)	0.95 (0.46, 0.12)
	Fast	<0.001	GWT	0.94 (0.47, 0.12)	0.70 (0.61, 0.18)	0.97 (0.32, 0.09)	0.98 (0.47, 0.13)

Results of genetic variance analysis, type of estimate applied (GCT: combined among- and within-twin pair component estimate, GWT: within-pair estimate) and intraclass correlation coefficients (ICCs). ICC MZ: ICCs of monozygotic (MZ) twins, ICC DZ: ICCs of dizygotic (DZ) twins, ICC MZ cn: ICCs of consecutive nights for each subject in MZ group, ICC DZ cn: ICCs of consecutive nights for each subject in DZ group. ICC results include: original sample ICC (upper percentile of bootstrapped data, median of bootstrapped data). *Analysis of variance not applicable (significant differences between the means in DZ and MZ twin set).

multiple parameters was below the bootstrapped median value, so it was lower than expected by chance. Within-pair similarity was *almost perfect* only once and *substantial* only twice. ICC estimations of slow spindle within-twin-pair resemblance as well as night-to-night stability were similar for sleep stage 2 when compared to SWS.

Considering fast spindles, the mean ICC for night-to-night stability was similar in both groups: 0.86 in the MZ set, compared to 0.85 in the DZ set. All these values were above the bootstrapped significance threshold ($P = 0.01$). Night-to-night stability in the MZ set was *almost perfect* for all fast spindle characteristics, whereas in the DZ set it was *almost perfect* for all but spindle amplitude parameters. Night-to-night stability of fast spindle amplitude in the DZ set ranged from *moderate* to *substantial*, therefore our finding of significant genetic influence on fast spindle amplitude should be treated with caution. The mean ICC of all fast spindle characteristics for within-pair resemblance was 0.76 in MZ twins and 0.45 in DZ twins. Within-pair similarity in the MZ set was below the significance level only for spindle number and density in F3A2 during SWS. According to the benchmark, in MZ twins within-pair similarity was seven times *almost perfect*, ten times *substantial* and three times only *moderate*. In DZ set within-pair similarity was at most *substantial* (six times) and only these values were above significance level. Again, some values were below the bootstrapped median, so they were lower than expected by chance.

Within-pair similarity in MZ twins was the lowest for fast spindle quantification parameters: total number and density,

especially in SWS. These lower ICC results were not influenced by night-to-night stability, which was always *almost perfect*.

DISCUSSION

In this study we present an automatic sleep spindle detection algorithm based on CWT, which carefully localizes fast and slow spindles frequency for each individual and estimates the signal amplitude for each investigated EEG channel. We used a validation data set of 18 naps and compared our solution against human scoring and a SIESTA spindle detector. While the SIESTA detector is a popular and well tested solution, it does not distinguish between slow and fast spindles. In addition, its detection threshold is not individually adjusted according to signal amplitude (see “SIESTA Algorithm” Section). During sleep stage 2, the agreement between human scorer and both detectors was *moderate*, whereas the agreement between detectors was *substantial*. During SWS, the agreement between detectors was *moderate*. Due to observed differences between spindles scored by each algorithm, we found it interesting to apply our algorithm to sleep-related memory consolidation data previously analyzed with the SIESTA detector and described in Genzel et al. (2009). This experiment did not significantly improve our knowledge about spindles and memory consolidation, but we saw how technical differences can influence the analysis outcome. We confirmed significant positive correlation between spindle density and declarative memory consolidation,

TABLE 7 | Genetic variance analysis, type of estimate applied (GCT: combined among- and within-twin pair component estimate, GWT: within-pair estimate) and intraclass correlation coefficients (ICCs) for spindle parameters in slow wave sleep.

Variable	Type	P	Analysis	ICC MZ	ICC DZ	ICC MZ cn	ICC DZ cn
<i>EEG channel: F3A2</i>							
Number of spindles	Slow*	—	—	0.94 (0.46, 0.12)	0.14 (0.61, 0.18)	0.92 (0.35, 0.09)	0.76 (0.46, 0.13)
	Fast	0.204	GWT	0.41 (0.53, 0.12)	0.45 (0.76, 0.14)	0.84 (0.41, 0.08)	0.93 (0.61, 0.11)
Spindle density	Slow*	—	—	0.96 (0.47, 0.12)	0.05 (0.62, 0.18)	0.96 (0.32, 0.08)	0.91 (0.46, 0.13)
	Fast	0.677	GWT	0.45 (0.55, 0.12)	0.51 (0.74, 0.17)	0.81 (0.35, 0.08)	0.89 (0.57, 0.11)
Spindle length	Slow	<0.001	GCT	0.98 (0.57, 0.11)	0.45 (0.69, 0.18)	0.92 (0.42, 0.08)	0.89 (0.49, 0.13)
	Fast	0.010	GWT	0.78 (0.44, 0.13)	0.42 (0.64, 0.19)	0.66 (0.35, 0.09)	0.85 (0.44, 0.13)
Spindle amplitude	Slow	<0.001	GCT	0.89 (0.45, 0.13)	0.16 (0.65, 0.20)	0.91 (0.34, 0.09)	0.78 (0.45, 0.13)
	Fast	<0.001	GCT	0.83 (0.48, 0.12)	−0.30 (0.61, 0.19)	0.88 (0.34, 0.08)	0.60 (0.47, 0.13)
Spindle frequency	Slow	0.052	GWT	0.91 (0.45, 0.12)	0.81 (0.66, 0.19)	0.93 (0.33, 0.08)	0.97 (0.48, 0.13)
	Fast	0.027	GWT	0.86 (0.49, 0.13)	0.73 (0.58, 0.18)	0.85 (0.35, 0.09)	0.90 (0.48, 0.13)
<i>EEG channel: P3A2</i>							
Number of spindles	Slow	0.005	GCT	0.88 (0.42, 0.12)	0.45 (0.66, 0.19)	0.89 (0.34, 0.08)	0.87 (0.46, 0.13)
	Fast	0.049	GWT	0.54 (0.49, 0.11)	0.15 (0.71, 0.19)	0.90 (0.35, 0.09)	0.83 (0.54, 0.12)
Spindle density	Slow	0.030	GWT	0.90 (0.53, 0.13)	0.64 (0.70, 0.20)	0.93 (0.36, 0.09)	0.95 (0.50, 0.13)
	Fast	0.071	GWT	0.68 (0.45, 0.12)	0.21 (0.59, 0.18)	0.88 (0.34, 0.08)	0.81 (0.50, 0.13)
Spindle length	Slow	<0.001	GCT	0.88 (0.57, 0.11)	0.46 (0.70, 0.20)	0.88 (0.39, 0.08)	0.90 (0.47, 0.13)
	Fast	0.020	GWT	0.74 (0.48, 0.13)	0.47 (0.63, 0.19)	0.80 (0.34, 0.09)	0.94 (0.50, 0.13)
Spindle amplitude	Slow	0.004	GWT	0.83 (0.44, 0.12)	0.41 (0.63, 0.19)	0.79 (0.33, 0.09)	0.73 (0.48, 0.12)
	Fast	0.004	GCT	0.80 (0.49, 0.12)	0.29 (0.62, 0.19)	0.84 (0.35, 0.09)	0.56 (0.48, 0.13)
Spindle frequency	Slow	0.049	GWT	0.82 (0.45, 0.12)	0.69 (0.63, 0.19)	0.88 (0.33, 0.08)	0.93 (0.48, 0.13)
	Fast	0.001	GWT	0.94 (0.47, 0.13)	0.77 (0.64, 0.18)	0.93 (0.32, 0.08)	0.96 (0.48, 0.13)

Abbreviations explanation as in **Table 6**. *Analysis of variance not applicable (significant differences between the means in DZ and MZ twin set).

but we did not reproduce a significant positive correlation between spindle activity and declarative memory consolidation. Finally, comparison of basic spindle parameters between a group of 32 healthy MZ and 14 DZ same-gender twins revealed strong genetic influence on the variability of all slow spindle parameters, fast spindle morphology, and a weaker genetic effect on variance of fast spindle quantification parameters.

In our algorithm, we detect spindles with CWT using the Morlet wavelet, since wavelets of this type were shown to catch sleep spindle characteristics very well (Zygierevicz et al., 1999). Our solution rejects periods of signal with strong muscle artifacts as well as segments dominated by alpha activity. Furthermore, our method of adjusting spindle detection threshold was designed to reflect background signal amplitude as independent of signal/sleep quality and temporary events as much as possible. For this reason, signal activity was filtered below 6 Hz to avoid the influence of delta waves and k-complexes, and above 18 Hz to exclude possible muscle artifacts. In addition, logarithm transformation of frequency spectra, combined with usage of median instead of a mean, should decrease the influence of temporary activity bursts and frequency peaks. However, thresholds computed with our algorithm during stage 2 sleep were on average 9% lower than thresholds computed for SWS, so our threshold adjustment method is still sleep quality/stage dependent. We are not aware how different sleep stages influence adjustable thresholds used in other algorithms, but our conclusion is that, to avoid unnecessary variance among sleep recordings, thresholds based on general signal amplitude should be computed using homogenous sleep stage.

Our automatic adjustment of sleep spindle frequency boundaries is based on comparison of parietal and frontal EEG signals, like that proposed by Bódizs et al. (2009, 2012), but instead of frequency spectra our method analyses the frequency of pre-localized spindle events. Since this approach filters out all unnecessary parts of the signal it may be more exact, especially when sleep spindle density is low. Furthermore, our solution is robust against possible amplitude differences between channels. We observed considerable inter-subject variation in both slow and fast sleep spindle frequency. In addition, the average frequency of slow sleep spindles during SWS was slower than during stage 2 sleep which suggests that spindle frequency ranges should be set separately for shallow and deep sleep. The frequency distribution of pre-localized spindle events as well as estimated spindle frequency ranges for each twin can be found in Supplementary Material.

We compared spindle detection of our new algorithm with a human scorer and the commercially available SIESTA spindle detector, which was developed using a large database with manually scored sleep spindles (Anderer et al., 2005). One limitation in this study is that we did not compare scorings on an independent test set. We set detection thresholds using the validation set in order to match the sensitivity of the SIESTA algorithm. Our comparison results could thus be inflated due to an overfitting problem. The comparison of our solution with other algorithms and human scorers using an independent dataset should be the next step in future work. According to published benchmarks for the kappa coefficient (Landis and Koch, 1977), during sleep stage 2 the agreement between a human scorer and both algorithms was *moderate*,

while both algorithms scored significantly more spindles. The agreement between algorithms was *substantial* during sleep stage 2 and dropped to *moderate* during SWS. In particular, the agreements with the human scorer seemed low and as presented in **Figure 6**, even manipulation of detection thresholds would not improve the agreement significantly. When we compared automatic algorithms we observed that spindles marked only by the SIESTA detector were either short or had the lowest amplitude, whereas spindles marked only by the SIESTA detector had a lower frequency, around 12 Hz, and higher activities in EEG background. Spindles marked only by the human scorer were the longest and had a very low amplitude in frequency spectrum. This low amplitude was problematic, since the human scored clearly the lowest amount of spindles, which means that the human detection threshold was the highest. The reason for the low average amplitude in the frequency spectrum was that marked events were often longer than activity in the sigma range. Spindles marked only by a visual scorer were rare, only 11% of total spindles scored. However, characteristics of these spindles show that visual scoring is prone to mistakes/inconsistencies. Since a sleep spindle is a very characteristic element of an EEG signal, this result seems to be disappointing. However, low spindle detection agreement is surprisingly a general phenomenon. Wendt et al. (2015) reported the average intra-expert agreement and inter-expert agreement measured with kappa at 0.66 and 0.52, respectively. Warby et al. (2014) reported that agreement between gold standard (consensus of human experts) and automatic algorithms measured with kappa ranged from 0.15–0.41 and pointed that the agreement between automatic detectors was generally lower than their agreement with the gold standard. Consistent high discrepancies between scorers indicate that even a small difference in detection approach results in a significantly different type of scored events. Unfortunately, simple differences in sensitivity between scorers only partially explain the problem. As Warby et al. (2014) observed: “*automated methods as a group were not consistent among themselves: they did not find the same ‘hidden’ spindles*”. Automatic detectors use various signal processing techniques, spindle frequency ranges and decision-making processes. All these variables add up to significantly different detection results. Whereas most human scorers seem to share the decision process, according to Warby et al. (2014), experts “*frequently rely on spindles being a ‘distinct train of waves’ that is clearly distinguishable from background*”. The general human tendency to score spindles with a clearly strong spindle activity compared to other frequencies is most likely the main reason why inter-expert agreement is higher than agreement between automatic methods as well as between automatic methods and human scorers. There are already algorithms which mimic this approach, including ones proposed by Huupponen et al. (2007), and the SIESTA detector used to validate our algorithm. However, firstly, the average inter-expert agreement is still only *moderate*, and second, human visual scoring is usually performed on “raw” EEG signal, while all automatic methods use filtering or various transformations to extract activity in the spindle frequency range. Since we

are not aware of any physiological data supporting the notion that spindles should dominate the frequency spectrum, our algorithm detects also spindles which are intermingled in other frequencies.

Low agreement between spindle detection methods combined with the highly individual character of sleep spindles (Werth et al., 1997), as well as the whole frequency spectrum (Buckelmüller et al., 2006), may lead to heterogeneous discrepancies in estimated spindle activity across subjects. As the result, the by-subject correlation of spindle activity estimated by different detection methods can be low. Warby et al. (2014) reported that the correlation between by-subject spindle density estimated from the gold standard and from the best automated detector was only $r = 0.62$. This fact leads to the question whether results of experiments are reproducible. For this reason, we re-analyzed sleep-related memory consolidation data, previously analyzed with the SIESTA detector and described in Genzel et al. (2009). The design of this project could be especially susceptible to these sort of discrepancies, since the idea was to correlate by-subject spindle activity estimations with memory retention. By-subject correlation between our algorithm and SIESTA detector did not fall below $r = 0.80$ neither for validation nor for memory consolidation data, and using our algorithm we reproduced almost all findings reported in Genzel et al. (2009). However, the highest discrepancy in by-subject correlation of spindle activity estimated by both algorithms and memory retention results was observed for spindle activity in sleep stage 2. It was surprising, since by-subject correlation of this parameter estimated by both algorithms was relatively high ($r = 0.94$). It shows that even small differences in spindle detection might lead to significantly different conclusions derived from an experiment. Significant discrepancies between spindle scorings increase the value of perfect reproducibility of the method and findings, provided by every automatic algorithm. For this reason, we conclude that the application of automatic algorithms for spindle detection in research projects should be encouraged.

Analysis of the twin data revealed high ICCs for night-to-night stability across investigated fast and slow spindle parameters during both sleep stage 2 and SWS, supporting previous reports about sleep spindle fingerprint characteristics (Werth et al., 1997; De Gennaro et al., 2005). Recently Eggert et al. (2015) reported ICC results for night-to-night stability of sleep spindles detected during stage 2 sleep with the SIESTA algorithm. The authors reported results from the central channel without distinction between slow and fast spindles. ICCs for night-to-night stability were also high for all spindle characteristics. The highest stability with ICC (0.92) was observed for spindle amplitude, and the lowest stability with ICC (0.84) was observed for spindle density. In our analysis we distinguished between slow and fast sleep spindles and we performed the analysis during stage 2 and during SWS separately. When comparing fast and slow spindles we observed that, besides frequency, stability of fast spindle parameters was moderately lower. This lower night-to-night stability of fast spindles dropped slightly further when we

looked into within-pair similarity of MZ twins. Interestingly, the drop off between night-to-night stability and within-pair similarity of MZ twins was not observed for slow spindle parameters, where ICC estimates were on average exactly the same. In DZ twins, within-pair similarity was clearly lower than their night-to-night stability for both fast and slow spindle parameters. As a result, GVA revealed genetic control on variance of all slow and most of fast spindle parameters during stage 2 sleep and SWS. However, the genetic component of fast spindle parameters, besides spindle frequency, was weaker, especially for fast spindle quantities. GVA did not show significant genetic determination of fast spindle number and density during sleep stage 2. Analyses repeated with a subgroup of MZ twins closely matched for age, gender and cohabitation to DZ twins confirmed our findings in the total twin sample (see Supplementary Material, Tables S20–S23). In addition, for matched MZ sample GVA could be performed on slow spindle amplitude in SWS as well as slow spindle quantities in SWS. For all these remaining parameters we found significant genetic influence.

The number of DZ twin pairs ($n = 14$) is a limitation of our study. It is the reason why there is high variability of within-pair similarity estimates between spindle parameters in DZ twins. With our sample size, strong dissimilarity within just one twin pair strongly affects ICC outcome for the whole group. Sometimes these values were very high, above the bootstrapped significance threshold ($P = 0.01$), but sometimes these values were below similarity expected by chance (median of bootstrapped data) or even below zero, which means that resemblance between DZ twins was lower than observed in the population. Such low similarities within DZ twins has little biological sense and most likely could occur due to the small sample size. If we would compute narrow sense heritability, error margins would be high due to the small size of the DZ sample. Therefore, we do not provide narrow sense heritability estimations. Furthermore, we did not correct our GVA results for multiple testing, so there is an increased probability of type 1 error.

The next limitation of our study is the fact that we compared our algorithm to the SIESTA detector and human scorer using only fixed spindle detection frequency ranges. While individually adjusted frequency ranges may improve the quality of spindle detection, this change in the algorithm could result in significant detection differences. To illustrate how such change influences the detection, we provided the comparison of our algorithm with itself, with and without adjustable frequency ranges. The agreement was *almost perfect* when we considered all sleep spindles together during stage 2 sleep. However, the agreement dropped during SWS, since the detector with individually adjusted frequency ranges marked more spindles. This was because individually adjusted frequency ranges in SWS were often lower than 11 Hz, which was the lower boundary in the detector with fixed ranges. The agreement dropped further when we analyzed slow and fast spindles separately, since individually adjusted

boundaries between fast and slow spindles varied and were rarely 13 Hz, like in fixed ranges approach. As a result, spindles classified as fast when using individually adjusted frequency ranges could be classified as slow when using fixed ranges. Ujma et al. (2015) compared a spindle detector with individually adjusted spindle frequency ranges vs. a different detector with fixed frequency ranges (slow spindles: 11–13 Hz, fast spindles: 13–15 Hz). The reported agreement was poor, especially for slow spindles. Differences between fixed and adjusted frequency ranges had a high impact on observed detection discrepancies. In many subjects individually adjusted fast spindle activity peak was approaching or fell below the 13 Hz threshold. Slow spindles seemed to be even more problematic. In approximately 25% of subjects the individually adjusted peak of slow spindle activity fell below 11 Hz, which is the commonly used lower boundary for spindle frequency. In order to compensate for the lack of validation of our adjustable frequency ranges vs. other methods, we provided detailed plots with detection results over frequency range for each twin in the Supplementary Material and in addition, we estimated genetic influence on sleep spindles also using fixed spindle detection frequency ranges. Results are included in Supplementary Material (Tables S12–S19). Due to fixed thresholds the separation between slow and fast spindles was less exact and therefore differences between two spindle types decreased. However, the outcome still supported our main observations about stronger night-to-night stability and stronger genetic influence on slow spindles when compared to fast ones.

Due to reported differences between spindle algorithms, as well as between human and automatic spindle scoring, spindle findings should be interpreted carefully. Our findings on strong genetic influence on spindle frequency, length and amplitude further promote the view that variability in the morphology of both slow and fast spindles is genetically driven. However, comparably weaker genetic effects on fast spindle quantity (density and total amount) may reflect stronger environmental influences on this spindle type (i.e., memory load). This is supported by a previous study on the role of fast spindles in sleep-dependent memory processing (Möller et al., 2011). A detection algorithm which considers the individual morphology of two types of spindles may be an important tool to identify environmental influences on this relevant sleep phenomenon.

ACKNOWLEDGMENTS

We would like to thank Richard Fitzpatrick for reading and commenting on the article.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fnhum.2015.00624/abstract>

REFERENCES

- Addison, P. (2002). *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance*, Bristol, Philadelphia: Institute of Physics Publishing.
- Ambrosius, U., Lietzenmaier, S., Wehrle, R., Wichniak, A., Kalus, S., Winkelmann, J., et al. (2008). Heritability of sleep electroencephalogram. *Biol. Psychiatry* 64, 344–348. doi: 10.1016/j.biopsych.2008.03.002
- Anderer, P., Gruber, G., Parapatics, S., Woertz, M., Miazhyńska, 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 × 7 utilizing the Siesta database. *Neuropsychobiology* 51, 115–133. doi: 10.1159/000085205
- Anderer, P., Klösch, 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
- Andrillon, T., Nir, Y., Staba, R. J., Ferrarelli, F., Cirelli, C., Tononi, G., et al. (2011). Sleep spindles in humans: insights from intracranial EEG and unit recordings. *J. Neurosci.* 31, 17821–17834. doi: 10.1523/jneurosci.2604-11.2011
- Astori, S., Wimmer, R. D., and Lüthi, A. (2013). Manipulating sleep spindles—expanding views on sleep, memory and disease. *Trends Neurosci.* 36, 738–748. doi: 10.1016/j.tins.2013.10.001
- Batuwita, R., and Palade, V. (2012). Adjusted geometric-mean: a novel performance measure for imbalanced bioinformatics datasets learning. *J. Bioinform. Comput. Biol.* 10:1250003. doi: 10.1142/s0219720012500035
- Bódizs, R., Gombos, F., and Kovács, I. (2012). Sleep EEG fingerprints reveal accelerated thalamocortical oscillatory dynamics in Williams syndrome. *Res. Dev. Disabil.* 33, 153–164. doi: 10.1016/j.ridd.2011.09.004
- Bódizs, R., Körmendi, J., Rigó, P., and Lázár, A. S. (2009). The individual adjustment method of sleep spindle analysis: methodological improvements and roots in the fingerprint paradigm. *J. Neurosci. Methods* 178, 205–213. doi: 10.1016/j.jneumeth.2008.11.006
- Buckelmüller, J., Landolt, H. P., Stassen, H. H., and Achermann, P. (2006). Trait-like individual differences in the human sleep electroencephalogram. *Neuroscience* 138, 351–356. doi: 10.1016/j.neuroscience.2005.11.005
- Buzsáki, G. (1989). Two-stage model of memory trace formation: a role for “noisy” brain states. *Neuroscience* 31, 551–570. doi: 10.1016/0306-4522(89)90423-5
- Christian, J. C., Borhani, N. O., Castelli, W. P., Fabsitz, R., Norton, J. A., Reed, T., et al. (1987). Plasma cholesterol variation in the National Heart, Lung and Blood Institute Twin Study. *Genet. Epidemiol.* 4, 433–446. doi: 10.1002/gepi.1370040605
- Christian, J. C., Kang, K. W., and Norton, J. J. (1974). Choice of an estimate of genetic variance from twin data. *Am. J. Hum. Genet.* 26, 154–161.
- Clemens, Z., Fabó, D., and Halász, 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
- Clemens, Z., Mölle, M., Eross, L., Barsi, P., Halász, P., and Born, J. (2007). Temporal coupling of parahippocampal ripples, sleep spindles and slow oscillations in humans. *Brain* 130, 2868–2878. doi: 10.1093/brain/awm146
- Cox, R., Hofman, W. F., and Talamini, L. M. (2012). Involvement of spindles in memory consolidation is slow wave sleep-specific. *Learn. Mem.* 19, 264–267. doi: 10.1101/lm.026252.112
- De Gennaro, L., Ferrara, M., Vecchio, F., Curcio, G., and Bertini, M. (2005). An electroencephalographic fingerprint of human sleep. *Neuroimage* 26, 114–122. doi: 10.1016/j.neuroimage.2005.01.020
- De Gennaro, L., Marzano, C., Fratello, F., Moroni, F., Pellicciari, M. C., Ferlazzo, F., et al. (2008). The electroencephalographic fingerprint of sleep is genetically determined: a twin study. *Ann. Neurol.* 64, 455–460. doi: 10.1002/ana.21434
- de Maertelaer, V., Hoffman, G., Lemaire, M., and Mendlewicz, J. (1987). Sleep spindle activity changes in patients with affective disorders. *Sleep* 10, 443–451.
- Dijk, D. J., Beersma, D. G., and Bloem, G. M. (1989a). Sex differences in the sleep EEG of young adults: visual scoring and spectral analysis. *Sleep* 12, 500–507.
- Dijk, D. J., Beersma, D. G., and van den Hoofdakker, R. H. (1989b). All night spectral analysis of EEG sleep in young adult and middle-aged male subjects. *Neurobiol. Aging* 10, 677–682. doi: 10.1016/0197-4580(89)90004-3
- Dresler, M., Kluge, M., Genzel, L., Schüssler, P., and Steiger, A. (2010). Impaired off-line memory consolidation in depression. *Eur. Neuropsychopharmacol.* 20, 553–561. doi: 10.1016/j.euroneuro.2010.02.002
- Dresler, M., Kluge, M., Pawlowski, M., Schüssler, P., Steiger, A., and Genzel, L. (2011). A double dissociation of memory impairments in major depression. *J. Psychiatr. Res.* 45, 1593–1599. doi: 10.1016/j.jpsychires.2011.07.015
- Eggert, T., Sauter, C., Dorn, H., Peter, A., Hansen, M. L., Marasanov, A., et al. (2015). Individual stability of sleep spindle characteristics in healthy young males. *Somnologie* 19, 38–45. doi: 10.1007/s11818-015-0697-x
- Ferrarelli, F., Huber, R., Peterson, M. J., Massimini, M., Murphy, M., Riedner, B. A., et al. (2007). Reduced sleep spindle activity in schizophrenia patients. *Am. J. Psychiatry* 164, 483–492. doi: 10.1176/appi.ajp.164.3.483
- Ferrarelli, F., Peterson, M. J., Sarasso, S., Riedner, B. A., Murphy, M. J., Benca, R. M., et al. (2010). Thalamic dysfunction in schizophrenia suggested by whole-night deficits in slow and fast spindles. *Am. J. Psychiatry* 167, 1339–1348. doi: 10.1176/appi.ajp.2010.09.121731
- Fogel, S. M., and Smith, C. T. (2011). The function of the sleep spindle: a physiological index of intelligence and a mechanism for sleep-dependent memory consolidation. *Neurosci. Biobehav. Rev.* 35, 1154–1165. doi: 10.1016/j.neubiorev.2010.12.003
- Gais, S., Mölle, M., Helms, K., and Born, J. (2002). Learning-dependent increases in sleep spindle density. *J. Neurosci.* 22, 6830–6834.
- Genzel, L., Ali, E., Dresler, M., Steiger, A., and Tesfaye, M. (2011). Sleep-dependent memory consolidation of a new task is inhibited in psychiatric patients. *J. Psychiatr. Res.* 45, 555–560. doi: 10.1016/j.jpsychires.2010.08.015
- Genzel, L., Bäurle, A., Potyka, A., Wehrle, R., Adamczyk, M., Friess, E., et al. (2014). Diminished nap effects on memory consolidation are seen under oral contraceptive use. *Neuropsychobiology* 70, 253–261. doi: 10.1159/000369022
- Genzel, L., Dresler, M., Cornu, M., Jäger, E., Konrad, B., Adamczyk, M., et al. (2015). Medial prefrontal-hippocampal connectivity and motor memory consolidation in depression and schizophrenia. *Biol. Psychiatry* 77, 177–186. doi: 10.1016/j.biopsych.2014.06.004
- Genzel, L., Dresler, M., Wehrle, R., Grözing, M., and Steiger, A. (2009). Slow wave sleep and REM sleep awakenings do not affect sleep dependent memory consolidation. *Sleep* 32, 302–310. doi: 10.1055/s-2007-991839
- Genzel, L., Kroes, M. C., Dresler, M., and Battaglia, F. P. (2014). Light sleep versus slow wave sleep in memory consolidation: a question of global versus local processes? *Trends Neurosci.* 37, 10–19. doi: 10.1016/j.tins.2013.10.002
- Haseman, J. K., and Elston, R. C. (1970). The estimation of genetic variance from twin data. *Behav. Genet.* 1, 11–19. doi: 10.1007/bf01067367
- Huupponen, E., De Clercq, W., Gomez-Herrero, G., Saastamoinen, A., Egiastian, K., Varri, A., et al. (2006). Determination of dominant simulated spindle frequency with different methods. *J. Neurosci. Methods* 156, 275–283. doi: 10.1016/j.jneumeth.2006.01.013
- 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
- Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174. doi: 10.2307/2529310
- Latka, M., Kozik, A., Jernajczyk, J., West, B. J., and Jernajczyk, W. (2005). Wavelet mapping of sleep spindles in young patients with epilepsy. *J. Physiol. Pharmacol.* 56, 15–20.
- Lopez, J., Hoffmann, R., and Armitage, R. (2010). Reduced sleep spindle activity in early-onset and elevated risk for depression. *J. Am. Acad. Child Adolesc. Psychiatry* 49, 934–943. doi: 10.1016/j.jaac.2010.05.014
- Mölle, M., Bergmann, T., 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
- Montplaisir, J., Petit, D., Lorrain, D., Gauthier, S., and Nielsen, T. (1995). Sleep in Alzheimer's disease: further considerations on the role of brainstem and forebrain cholinergic populations in sleep-wake mechanisms. *Sleep* 18, 145–148.
- Morin, A., Doyon, J., Dostie, V., Barakat, M., Hadj Tahar, A., Korman, M., et al. (2008). Motor sequence learning increases sleep spindles and fast frequencies in post-training sleep. *Sleep* 31, 1149–1156.
- Nishida, M., and Walker, M. P. (2007). Daytime naps, motor memory consolidation and regionally specific sleep spindles. *PLoS One* 2:e341. doi: 10.1371/journal.pone.0000341

- Peyrache, A., Khamassi, M., Benchenane, K., Wiener, S. I., and Battaglia, F. P. (2009). Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nat. Neurosci.* 12, 919–926. doi: 10.1038/nn.2337
- Rechtschaffen, A., and Kales, A. (1968). *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*. Los Angeles, CA: UCLA Brain Information Service.
- Schimicek, P., Zeitlhofer, J., Anderer, P., and Saletu, B. (1994). Automatic sleep-spindle detection procedure: Aspects of reliability and validity. *Clin. Electroencephalogr.* 25, 26–29. doi: 10.1177/155005949402500108
- Tarokh, L., Carskadon, M. A., and Achermann, P. (2011). Trait-like characteristics of the sleep EEG across adolescent development. *J. Neurosci.* 31, 6371–6378. doi: 10.1523/jneurosci.5533-10.2011
- Ujma, P. P., Gombos, F., Genzel, L., Konrad, B. N., Simor, P., Steiger, A., et al. (2015). A comparison of two sleep spindle detection methods based on all night averages: individually adjusted versus fixed frequencies. *Front. Hum. Neurosci.* 9:52. doi: 10.3389/fnhum.2015.00052
- Ujma, P. P., Konrad, B. N., Genzel, L., Bleifuss, A., Simor, P., Pótári, A., et al. (2014). Sleep spindles and intelligence: evidence for a sexual dimorphism. *J. Neurosci.* 34, 16358–16368. doi: 10.1523/jneurosci.1857-14.2014
- Wamsley, E. J., Tucker, M. A., Shinn, A. K., Ono, K. E., McKinley, S. K., Ely, A. V., et al. (2012). Reduced sleep spindles and spindle coherence in schizophrenia: mechanisms of impaired memory consolidation? *Biol. Psychiatry* 71, 154–161. doi: 10.1016/j.biopsych.2011.08.008
- Warby, S. C., Wendt, S. L., Welinder, P., Munk, E. G. S., Carrillo, O., Sorensen, H. B. D., 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. (2015). Inter-expert and intra-expert reliability in sleep spindle scoring. *Clin. Neurophysiol.* 123, 1548–1556. doi: 10.1016/j.clinph.2014.10.158
- Werth, E., Achermann, P., Dijk, D. J. and Borbély, 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
- Wierzynski, C. M., Lubenov, E. V., Gu, M., and Siapas, A. G. (2009). State-dependent spike timing relationships between hippocampal and prefrontal circuits during sleep. *Neuron* 61, 587–596. doi: 10.1016/j.neuron.2009.01.011
- Zygierewicz, J., Blinowska, K. J., Durka, P. J., Szelenberger, W., Niemcewicz, S., and Androsiuk, W. (1999). High resolution study of sleep spindles. *Clin. Neurophysiol.* 110, 2136–2147. doi: 10.1016/s1388-2457(99)00175-3

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.

Copyright © 2015 Adamczyk, Genzel, Dresler, Steiger and Friess. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and 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.

Spindles in Svarog: framework and software for parametrization of EEG transients

Piotr J. Durka^{1*}, Urszula Malinowska², Magdalena Zieleniewska¹, Christian O'Reilly^{3,4}, Piotr T. Róžański⁵ and Jarosław Żygierewicz¹

¹ Faculty of Physics, University of Warsaw, Warsaw, Poland, ² Department of Neurology, Epilepsy Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA, ³ McConnell Brain Imaging Center, Montreal Neurological Institute, McGill University, Montreal, QC, Canada, ⁴ Center for Advanced Research on Sleep Medicine, Centre de Recherche de l'Hôpital du Sacré-Cœur, Université de Montréal, Montreal, QC, Canada, ⁵ College of Inter-Faculty Individual Studies in Mathematics and Natural Sciences, University of Warsaw, Warsaw, Poland

OPEN ACCESS

Edited by:

Simon C. Warby,
Stanford University, USA

Reviewed by:

Christian Bénar,
Institut National de la Recherche
Médicale, France
Alpar S. Lazar,
University of Cambridge, UK

*Correspondence:

Piotr J. Durka,
Faculty of Physics, University of
Warsaw, ul. Pasteura 5, 02-093
Warsaw, Poland
durka@fuw.edu.pl

Received: 15 October 2014

Accepted: 21 April 2015

Published: 08 May 2015

Citation:

Durka PJ, Malinowska U, Zieleniewska M, O'Reilly C, Róžański PT and Żygierewicz J (2015) Spindles in Svarog: framework and software for parametrization of EEG transients. *Front. Hum. Neurosci.* 9:258. doi: 10.3389/fnhum.2015.00258

We present a complete framework for time-frequency parametrization of EEG transients, based upon matching pursuit (MP) decomposition, applied to the detection of sleep spindles. Ranges of spindles duration (>0.5 s) and frequency (11–16 Hz) are taken directly from their standard definitions. Minimal amplitude is computed from the distribution of the root mean square (RMS) amplitude of the signal within the frequency band of sleep spindles. Detection algorithm depends on the choice of just one free parameter, which is a percentile of this distribution. Performance of detection is assessed on the first cohort/second subset of the Montreal Archive of Sleep Studies (MASS-C1/SS2). Cross-validation performed on the 19 available overnight recordings returned the optimal percentile of the RMS distribution close to 97 in most cases, and the following overall performance measures: sensitivity 0.63 ± 0.06 , positive predictive value 0.47 ± 0.08 , and Matthews coefficient of correlation 0.51 ± 0.04 . These concordances are similar to the results achieved on this database by other automatic methods. Proposed detailed parametrization of sleep spindles within a universal framework, encompassing also other EEG transients, opens new possibilities of high resolution investigation of their relations and detailed characteristics. MP decomposition, selection of relevant structures, and simple creation of EEG profiles used previously for assessment of brain activity of patients in disorders of consciousness are implemented in a freely available software package Svarog (Signal Viewer, Analyzer and Recorder On GPL) with user-friendly, mouse-driven interface for review and analysis of EEG. Svarog can be downloaded from <http://braintech.pl/svarog>.

Keywords: sleep spindles, matching pursuit, EEG transients, time-frequency, sleep, Svarog, open source, free software

1. Introduction

Sleep spindles are defined in Rechtschaffen and Kales (1968); Ibert et al. (2007) as a train of distinct waves with frequency 11–16 Hz (most commonly 12–14 Hz) with a duration ≥ 0.5 s. Detection of these structures by human experts, trained in visual analysis of EEG, constitutes a gold standard.

Unfortunately, the inter-expert agreement in scoring sleep spindles is limited. This drawback undermines the idea of repeatability of experiments, which lies at the foundations of hard sciences: the same study of sleep spindles on the same dataset may yield different results, because of differences in the visual selections done by human experts.

Explosion of the applications of computerized signal processing methods resulted in a multitude of automatic detection algorithms. The most effective so far are based upon a common framework, introduced in Schimicek et al. (1994), reviewed e.g., in Warby et al. (2014):

1. EEG is band-pass filtered in the frequency range related to sleep spindles.
2. Signal from the previous step is subjected to amplitude thresholding in the time domain.
3. Epochs exceeding the threshold are filtered in the time domain to select those corresponding to sleep spindles.

Contrary to the visual detection by human experts, who concentrate directly and separately on relevant transient structures visible in EEG, each step of this sequential procedure implements only one aspect of the definition, and accumulates the bias from the previous steps. This drawback is the consequence of separate application of filters in the frequency and time domains. This turns our attention to the time-frequency methods of signal processing.

Classically, methods like short-time Fourier transform (STFT) and wavelet transform (WT) are used to compute the distribution of signal's energy in the time-frequency plane (Durka and Blinowska, 1997). Regions of increased energy correspond directly to signals transients, but their automatic selection still requires some kind of thresholding. Bias resulting from *a priori* choices of thresholds and further postprocessing becomes even more difficult to assess than in the spectral methods. Also, results depend significantly on prior choices of parameters like the duration of the time window in STFT or choice of the mother wavelet in WT.

Algorithm adapting the parameters automatically to the local content of the analyzed signal was introduced in Mallat and Zhang (1993). Matching pursuit algorithm (MP, Section 2.1) is an iterative procedure explaining the signal as a sum of Gabor functions (Figure 1), chosen optimally from a large and redundant set. Comparing to WT and STFT, analysis window and partly also the mother wavelet in this approach are chosen individually for each local transient structure present in the analyzed signal. Another unique feature of MP is the explicit parameterization of the structures fitted to the signal in terms of their time and frequency centers, duration and phase. This allows to perform detection directly in the space of these parameters in one step.

This approach has been successfully applied for the detection and parameterization of EEG transients including sleep spindles in different paradigms, mostly at the University of Warsaw. Additionally, MP-based detection of several types of EEG transients can be efficiently combined into an automatic sleep

stager, based explicitly upon the accepted criteria for stages (Malinowska et al., 2009). However, in spite of almost 20 years of publishing results (c.f. Durka and Blinowska, 1995; Żygierewicz et al., 1999; Malinowska et al., 2013 and many more) and free software for MP decomposition (our versions of the MP algorithm have been freely available since 2001, Durka et al., 2001), this approach to EEG analysis has been seldom applied outside our group. One of the reasons may have been a relative technical complexity of the whole procedure. To cope with this problem, this paper introduces a user-friendly and freely available multiplatform software for detection of sleep spindles (and other transients) in MP decompositions of EEG. This plugin is embedded in Svarog—Signal Viewer, Analyzer and Recorder On GPL.

Detection of sleep spindles presented in this paper relies on the correspondence of their shape (waxing and waning oscillations) to the Gabor functions used in MP decomposition (Figure 1), so finding corresponding structures among the Gabor functions fitted by the MP to EEG time series is straightforward and consists of setting the limits on their frequency centers, durations and amplitudes. Duration and frequency are taken literally from the definition of sleep spindles. As for the minimal amplitude, which is not directly defined, we adapt the common approach, which relates this parameter to the RMS of the signal filtered in the sigma band.

2. Materials and Methods

2.1. Matching Pursuit Algorithm (MP)

2.1.1. Matching Pursuit (MP)

MP was proposed by Mallat and Zhang (1993) as a suboptimal, iterative solution to the intractable problem of an optimal representation of a signal x in a redundant dictionary D , containing dense set of functions g_γ . In plain English, the gist of the MP procedure can be summarized as follows:

1. We start by creating a huge, redundant set D (called a dictionary) of candidate waveforms for representation of structures possibly occurring in the signal. For the time-frequency analysis of signals we use dictionaries composed of sines with Gaussian envelopes, called Gabor functions, which reasonably represent waxing and waning of spindle oscillations.
2. From this D dictionary we choose only those functions, which fit the local signal structures. In such a way, the width of the analysis window is adjusted to the local properties of the signal. Local adaptivity of the procedure is somehow similar to the process of visual analysis, where an expert tends to separate local structures and assess their characteristics. Owing to this local adaptivity, MP is the only signal processing method returning explicit time span of detected structures.
3. The above idea is implemented in an iterative procedure: in each step we find the “best” function, and then subtract it from the signal being decomposed in the following steps.

As for the mathematical description, denoting the function fitted to the signal x in the n -th iteration of MP as g_{γ_n} , and the

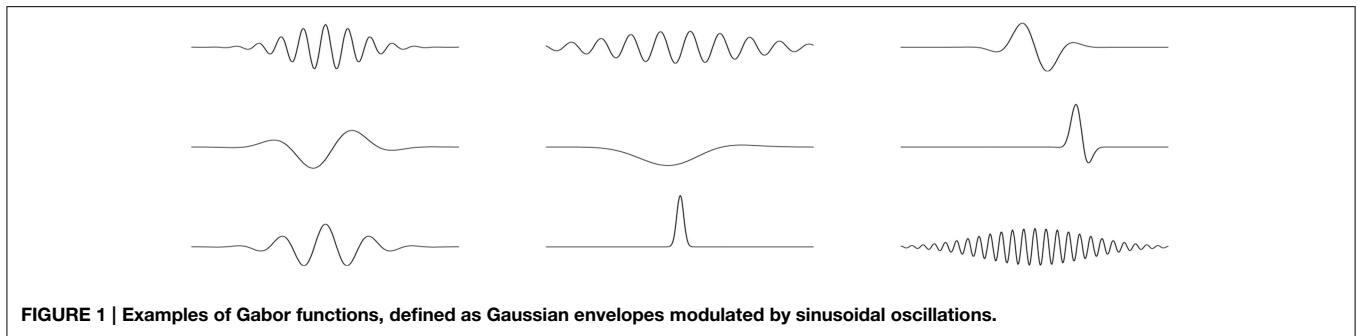


FIGURE 1 | Examples of Gabor functions, defined as Gaussian envelopes modulated by sinusoidal oscillations.

residual left after n -th iteration as $R^n x$, we can describe the procedure as:

$$\begin{cases} R^0 x = x \\ R^n x = \langle R^n x, g_{\gamma_n} \rangle g_{\gamma_n} + R^{n+1} x \\ g_{\gamma_n} = \arg \max_{g_{\gamma_i} \in D} |\langle R^n x, g_{\gamma_i} \rangle| \end{cases} \quad (1)$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product of signals and $|\cdot|$ the absolute value. As a result we get an approximate expansion:

$$x \approx \sum_{n=0}^{M-1} \langle R^n x, g_{\gamma_n} \rangle g_{\gamma_n} \quad (2)$$

where M equals the number of iterations of Equation (1). For a time-frequency analysis of real-valued signals, dictionary D is usually composed from Gabor functions:

$$g_{\gamma}(t) = K(\gamma) e^{-\pi \left(\frac{t-u}{s}\right)^2} \cos(\omega(t-u) + \phi) \quad (3)$$

where γ is a set of parameters such that $\gamma = (u, \omega, s)$ and $K(\gamma)$ is a normalization constant such that $\|g_{\gamma}\| = 1$.

The procedure is generic. The only major settings correspond to:

- quality of the decomposition, regulated mainly by the size of the dictionary D , which controls the parameterization accuracy of detected structures.
- number of iterations M , which regulates the accuracy of the overall approximation, the number of low energy structures included in the decomposition increases with M .

In both cases, higher settings result in higher accuracy.

2.1.2. Size of the dictionary D

Size of the dictionary D determines the number of candidate waveforms that will be fitted to the signal, and hence the resolution of the resulting decomposition. The resolution goes up with the number of functions in the dictionary. To make this setting independent of the size of the signal, we introduced one parameter regulating the density of the dictionary, related to the maximum distance between the dictionary's waveforms. This parameter is called in the Svarog interface (**Figure 2**) "energy

error" ϵ , since it relates to the maximum error that MP can make in a single iteration, as explained in details in Kuś et al. (2013)¹.

This special construction of the dictionary, ensuring a uniform distribution in the space of inner products, imposes non-uniform distribution of dictionary's functions in the space of their time positions, widths and frequencies (Kuś et al., 2013). For example, setting of $\epsilon = 0.04$ used for MP decompositions in this paper gives, for the frequency range of sleep spindles, possible time widths 0.53, 0.8, 1.21, and 1.82 s. That means that a spindle—or even a perfect Gabor function—with a width 1.5 s will be matched by a Gabor function from the dictionary with width either 1.21 or 1.82 s, and the leftover energy will be accounted for in the remaining iterations or will be left as a residual modeling noise if not accounted by the first M functions.

2.1.3. Number of iterations M

Number of iterations M is easier to assess, since the g_{γ_n} in Equation (2) are ordered by decreasing energy. That means that in two different decompositions differing only in the setting of the number of iterations, say 50 and 100, the first 50 waveforms will be the same (with small exceptions if stochastic decomposition was chosen), and iterations 51–100 will contain only structures of energy smaller than contributed by $g_{\gamma_{50}}$.

Increasing the number of iterations will not improve the quality of fit of any single waveform, so if we are interested in structures of relatively high energy, as is usually the case when looking for structures which are also visible for human expert, it makes no sense to increase M above the number which can be determined heuristically for a given problem and class of signals.

Described above MP decomposition is a purely mathematical procedure. In relation to EEG analysis, bad news are:

¹ ϵ relates to the maximum distance between two neighboring functions available for decomposition. The distance between two Gabor functions g_1 and g_2 from the dictionary D , proposed in Kuś et al. (2013), is measured in the space of inner products $\langle g_1 | g_2 \rangle$ related to the energy as

$$d(g_1, g_2) = \sqrt{1 - \langle g_1 | g_2 \rangle} \quad (4)$$

Dictionary is constructed in such a way that this distance is kept uniform across the neighboring functions. When fitting the dictionary's functions to a signal, the maximum error occurs when a signals structure falls exactly in between two functions available in the dictionary. In such a dictionary, this error will not exceed the distance between neighboring functions from the dictionary. In energy units it will be $d(g_1, g_2)^2$ —the (maximum) "energy error" ϵ .

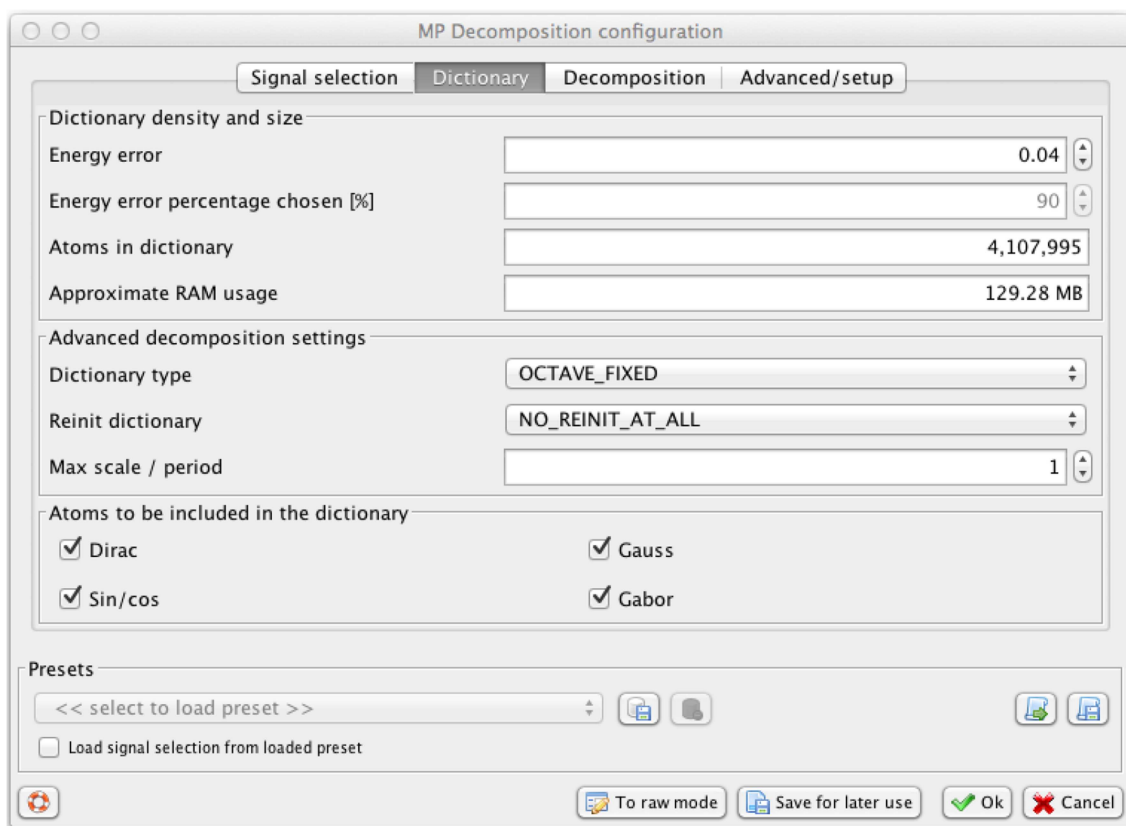


FIGURE 2 | Svarog window for setting the parameters of the MP decomposition, presenting values used in this study. Decreasing the parameter “Energy error”—in the text referred as ϵ —increases resolution and

the number of functions in the dictionary and usage of RAM, calculated automatically in the lower panels. For explanation of other parameters see Kuš et al. (2013).

1. Computation of the MP decomposition of a signal is relatively time-consuming even on a modern PC.
2. Settings of the energy error and number of iterations may require some consideration in case of limited computational resources, as discussed above in Sections 2.1.2 and 2.1.3.

Good news are:

1. Unlike most of the time-frequency methods of signal processing, setting of parameters is not a tradeoff between different aspects of the quality of decomposition, but a tradeoff between the quality and speed.
2. MP decomposition is generic, and once performed, the same decomposition of given epoch can be used to investigate the presence of different structures (c.f. **Figures 7, 8**). That’s where the weight is switched from mathematics to neuroscience.

2.1.4. Software implementation

Program computing the actual MP decomposition of given epoch is implemented in C and compiled separately for each platform. It is a command-line program, taking input from a config file and writing output to a binary file containing parameters of the fitted functions (a “book” *.b). To facilitate its application, we created a wrapper/GUI module for Svarog, which is a multiplatform

EEG review system. After installation and configuration of the system (Section 4.4), user can perform MP decompositions of the epoch selected by mouse, setting the decomposition parameters in tabs of the window displayed in **Figure 2**. Svarog then writes the selected (referenced and filtered) epoch to disk and calls the MP binary, which computes its decomposition and saves results to disk. These results can be then explored as an interactive time-frequency map as shown in **Figure 3**, or used for construction of summary reports on selected structures, as discussed in Section 3.3. For those who want to design their own post-processing, we provide scripts for reading the results of MP decomposition in Matlab and Python (Section 4.4).

2.2. Experimental Data

Data comes from the first cohort/second subset of the Montreal Archive of Sleep Studies (MASS-C1/SS2) (O’Reilly et al., 2014). It includes whole-night recordings from 19 young and healthy participants (8 male and 11 female; 23.6 ± 3.7 SD years old) with expert scoring of sleep stages according to the rules of Rechtschaffen and Kales (1968). For the gold standard, we used scoring of spindles from expert #1 available on MASS website. This scoring was performed for epochs of non-rapid eye movement stage two sleep, on C3 channel (linked-ear reference),

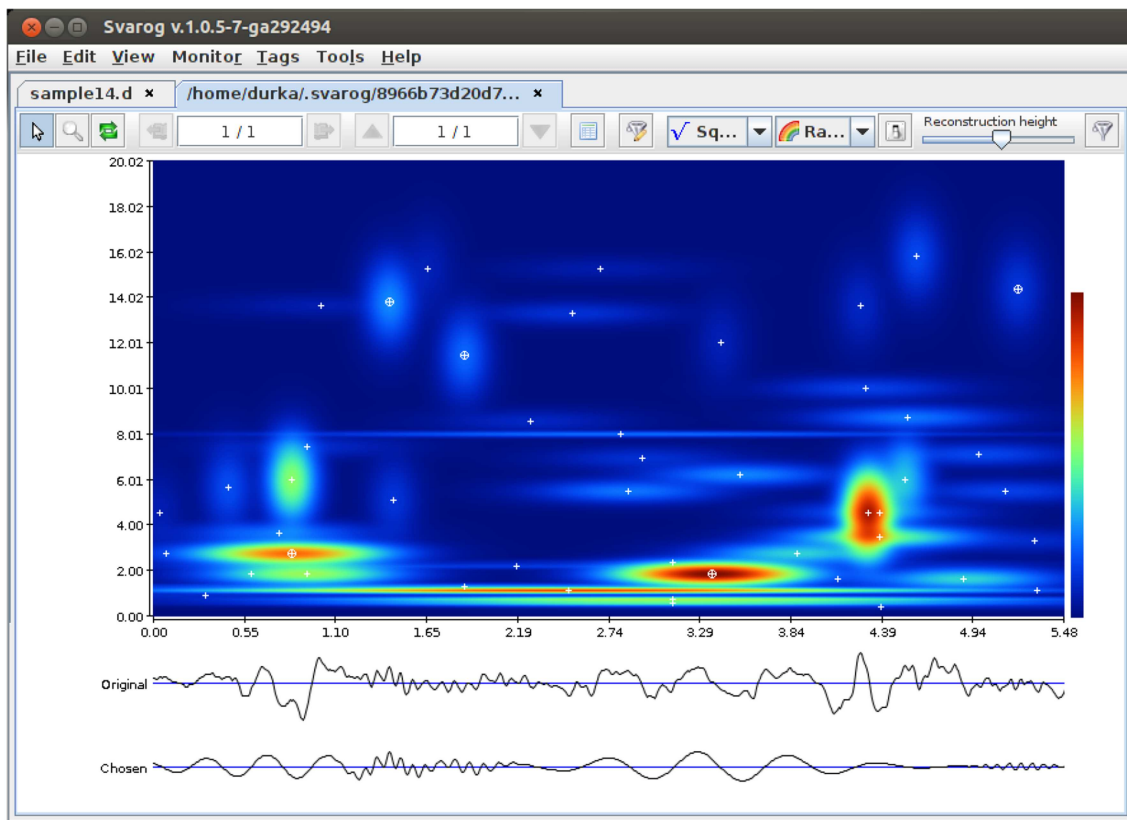


FIGURE 3 | Results of MP decomposition displayed as an interactive time-frequency map of signal's energy density in Svarog. Clicking center of a blob (marked by white cross) adds the corresponding function to the reconstruction (bottom signal). From Kuš et al. (2013)

and following AASM rules (Ibert et al., 2007). This database was chosen as it is open for sleep research and therefore facilitate reproducibility (see Section 4.4).

2.3. Measures of Performance of Detection

We based the assessment of efficiency of the detector on the markings with the accuracy of the EEG sampling, as proposed in O'Reilly et al. (in revision). In such an approach, at each sample (in our case 256 samples per second), there are four well-defined outcomes of comparison of expert's and detector's scorings: spindle present according to both expert and detector (true positives; TP), spindle absent according to both expert and detector (true negatives; TN), spindle present according to expert, but absent according to detector (false negative; FN), spindle absent according to expert, but present according to detector (false positives; FP). Counts of each type of outcome can be used to formulate various measures of detector performance:

$$\text{sensitivity} = \frac{TP}{TP + FN} \quad (5)$$

Positive predictive value² (PPV):

²PPV is related to False Discovery Rate as: $PPV = 1 - FDR$.

$$PPV = \frac{TP}{FP + TP} \quad (6)$$

Matthews coefficient of correlation (MCC):

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{P * P' * N * N'}} \quad (7)$$

where $P = TP + FN$, $P' = TP + FP$, $N = FP + TN$, $N' = FN + TN$.

Cohens κ :

$$\kappa = \frac{\frac{TN + TP}{P + N} - P_e}{1 - P_e} \quad (8)$$

where P_e is the probability of random agreement defined as:

$$P_e = \frac{P'P + N'N}{(P + N)^2} \quad (9)$$

F₁-score:

$$F_1 = 2 * \frac{PPV * \text{sensitivity}}{PPV + \text{sensitivity}} \quad (10)$$

2.4. Detection of Sleep Spindles

Division between the purely mathematical MP decomposition of signals and further neuroscience research is clearly reflected in the

structure of the Svarog software package. The first step, briefly covered in Section 2.1, consists of a generic approximation of the signal by a linear sum of Gabor functions. The second step, which is selection of the structures corresponding to sleep spindles, constitutes the main topic of this article.

MP offers explicit parameterization of signal structures in terms of their time and frequency positions, widths and amplitudes. Detection of sleep spindles within the proposed framework can be perceived as filtering out irrelevant structures from a database containing all the waveforms fitted by MP to a given signal epoch. Settings of the filter can be directly based upon the classical definition(s) mentioned in the Introduction. We choose frequency range 11–16 Hz and duration exceeding 0.5 s. Duration and time center of each detected spindle are returned explicitly by the MP algorithm, as parameters u and s from Equation (3), which gives us the time extent of the spindle from $u - s/2$ to $u + s/2$. Duration is taken here explicitly as the half-width of the Gaussian envelope of the Gabor function, but it can be adjusted by a multiplicative factor e.g., to optimize the concordance with visual detection. In general, using the setting window presented in **Figure 7**, one can easily test the procedure with different settings adjusted e.g., to different definitions, like frequency 12–14 Hz as defined in Rechtschaffen and Kales (1968) or slow (11–13 Hz) and fast (13–16 Hz) spindles separately.

Due to the lack of a precise definition of the minimum amplitude for spindles, one can either adapt a fixed threshold (e.g., Schimicek et al., 1994; Ventouras et al., 2005), usually optimized for a given recording (which causes obvious problems with generalization of the procedure to recordings from other labs/cohorts), or compute a threshold based upon the properties of the analyzed signal and in particular adapted to individual subject (e.g., Huupponen et al., 2000; Ray et al., 2010), which results in a more general procedure. We compute this threshold in relation to the RMS distribution. Exemplary distribution for one of the recordings is shown in **Figure 4**. To obtain the RMS distribution we filter the signal in the frequency band of sleep spindles (using 2nd order band-pass Butterworth filter with the cutoff frequencies set to 11 and 16 Hz). The RMS values were evaluated in successive, non-overlapping time windows with duration of 0.2 s. With this combination of bandwidth and window duration, one window includes more or less one period of oscillations of the filtered signal. Thus, in each window we can assume an approximate relation between amplitude and RMS as for a constant-amplitude sine wave. In such case peak-to-peak amplitude relates to the RMS as:

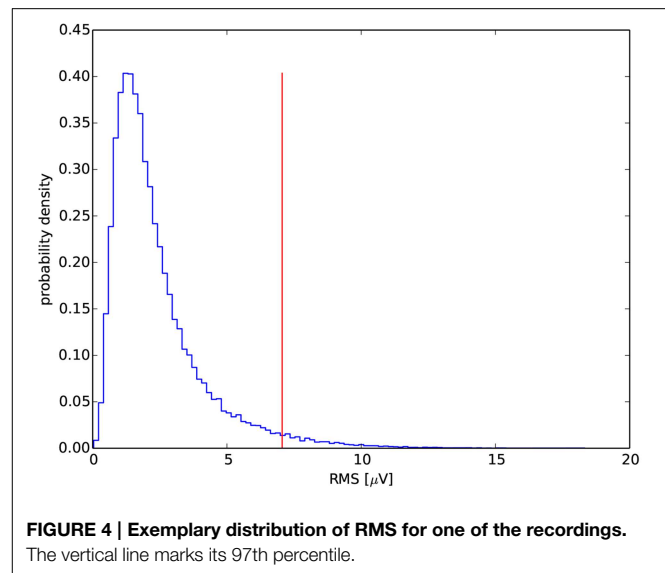
$$A = 2\sqrt{2}P_{\text{RMS}} \quad (11)$$

where P_{RMS} is the percentile of the mentioned RMS distribution, chosen to maximize resulting MCC.

3. Results

3.1. Performance of Sleep Spindles Detection in Individual Cases

As described in Section 2.4, the minimal amplitude of candidate waveform is a free parameter in the proposed detector of sleep



spindles. In order to have a complete picture of the detector performance on the current dataset, in **Figure 5A** we present the sensitivity, PPV and MCC for a range of RMS percentiles.

Figure 5B shows the distribution of the optimal, in the sense of maximizing MCC, percentiles for each of the recordings. The median of this distribution is the 97th percentile.

3.2. Cross-validation

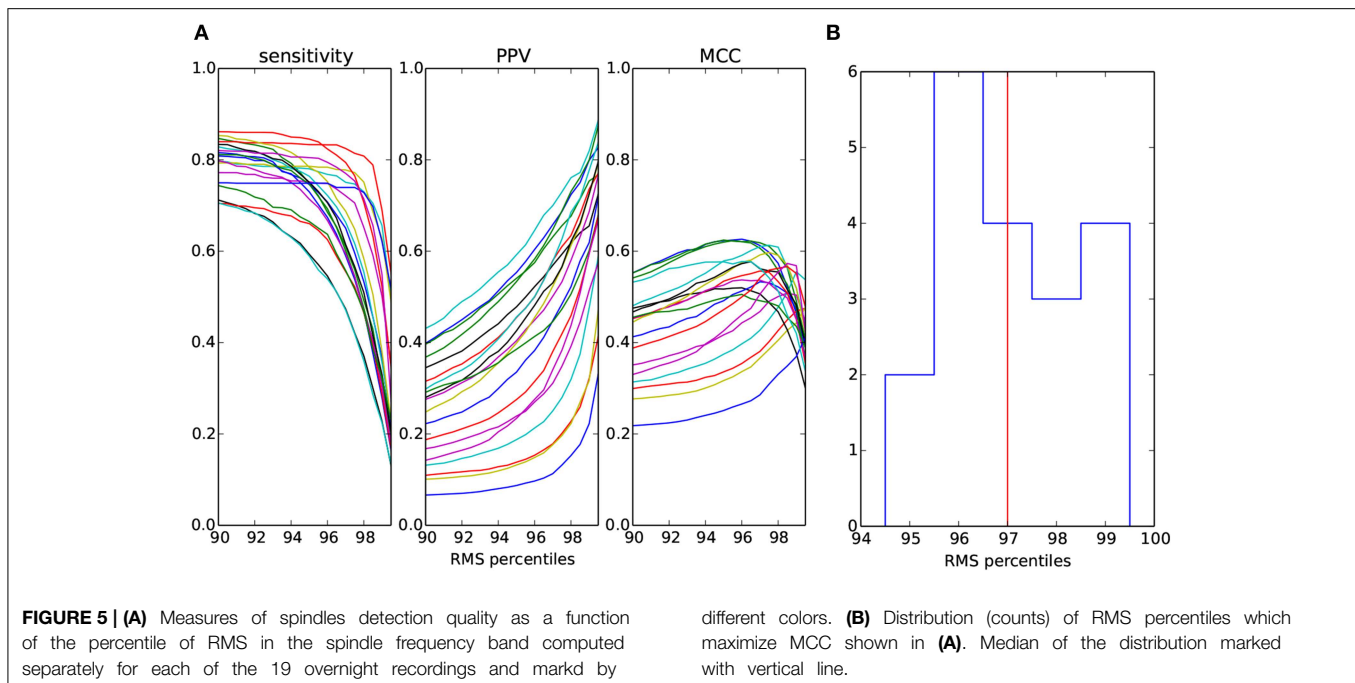
A common pitfall in the evaluation of the algorithms detecting sleep spindles is their explicit optimization for a particular dataset, often the same as the one used for presenting the performance of resulting algorithm. It is also a common problem in evaluation of detection algorithms, and the standard solution used in machine learning is called cross-validation.

For the evaluation of performance of the proposed method, we implement the following cross-validation procedure, related to the only parameter not taken directly from the definition of sleep spindles, which is the minimal amplitude expressed in terms of the percentile of RMS distribution in the frequency range of sleep spindles:

1. Randomly divide the available recordings in two disjoint subsets, further called the training set and the validation set.
2. Compute the optimal percentile for the training set.
3. Evaluate the performance on the validation set.
4. Repeat steps (1–3).

By averaging resulting performance measures over different random divisions of the available dataset we obtain an estimate of the average performance of the procedure on “unseen” data. This estimate tends to be a bit lower than the overall performance computed and estimated on the whole dataset at once.

We performed 100 iterations of the cross-validation procedure, each time randomly choosing 14 recordings for the training set used to compute the optimal RMS percentile. Then these 14 percentiles P_{RMS} , optimal for each of the recording separately, were averaged. The resulting average threshold



was applied to find the minimal spindle amplitudes for all the remaining 5 recordings. **Figure 6** shows the distribution of the resulting performance measures averaged over the validation sets. The summary statistics of performance are presented in **Table 1**.

3.3. EEG Profiles

Proposed approach offers precise detection of time centers and durations of sleep spindles and other transients. Apart from these, MP decomposition provides also an explicit and high resolution parameterization of their frequencies, amplitudes and phases. This opens a simple access to detailed information on the pattern of their occurrences across the whole analyzed recording, including:

- exact time occurrences of each detected structure with information about amplitude of each detected spindle.
- number of structures per epoch (in sleep analysis this is traditionally 20 or 30 s).
- percent of the epoch's time occupied by selected transients.

Although the last parameter has not been used for sleep spindles so far, all these reports are presented for demonstration in the three upper panels of **Figure 8**.

Sleep spindles are not the only EEG transients which can be effectively detected and parameterized by means of proposed approach. Another classic example of transient structures crucial for assessment of the sleep process are slow waves (Durka et al., 2005a). **Figure 7** presents example parameters allowing for selecting, from the same MP decomposition of the same signal, structures corresponding to slow waves: amplitude above 70 μ V, frequency 0.2–4 Hz, and time width above 0.5 s.

Figure 8 presents these profiles for sleep spindles and slow waves, computed in a fully automatic way without prior removal of artifacts. Examples of time-frequency definitions of structures in Svarog also include alpha, beta, theta and delta waves, and K-complexes (Malinowska et al., 2009). As explained in Section 4.1, all these profiles can be computed from the same MP decomposition, and reports for different settings of filters defining these structures, contrary to the underlying MP decomposition, are computed in seconds.

These profiles can be used for investigating several features of EEG, previously assessed by different specially constructed algorithms, or by visual inspection. For example:

- Report in the lower panel of **Figure 8**, showing the time fraction of each epoch occupied by slow waves, is crucial for sleep staging, since stages 3 and 4 are defined directly in terms of ranges of this parameter (Stage 3: 20–50%, Stage 4: above 50%). This correspondence was explored in Malinowska et al. (2009) for automatic construction of hypnograms, based directly upon the classical criteria from Rechtschaffen and Kales (1968).
- Profiles for these and other structures were used for assessment of the brain activity of patients in different states of disorders of consciousness (Malinowska et al., 2013).

4. Discussion

4.1. Computational Complexity of MP

As mentioned in Section 2.1, in each step of the MP algorithm we compute inner products of all the functions from the dictionary with the signal (or the residuum left from previous iterations). Implemented directly, this would typically result in millions of inner products, each computed on thousands of samples.

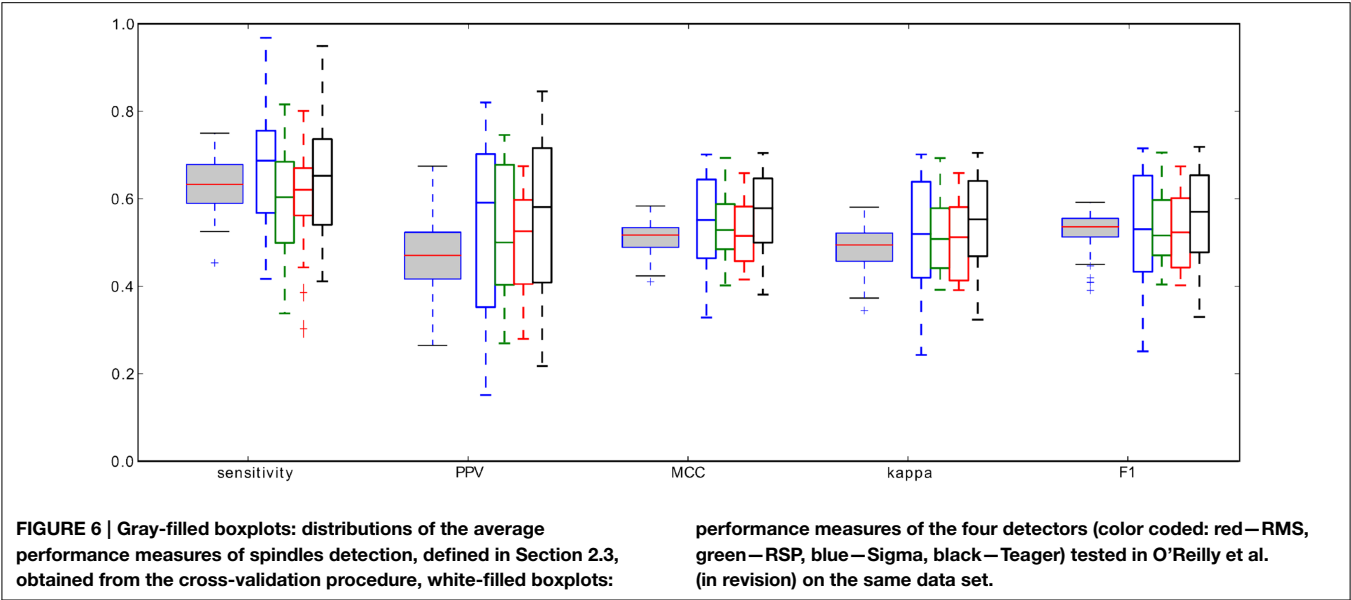


TABLE 1 | Summary of cross-validation performance statistics.

Measure	Median	First quartile	Third quartile	Mean	Standard deviation
Sensitivity	0.63	0.59	0.68	0.63	0.06
PPV	0.47	0.42	0.52	0.47	0.08
MCC	0.52	0.49	0.53	0.51	0.04
Cohen kappa	0.49	0.46	0.52	0.49	0.05
F ₁ -score	0.54	0.51	0.56	0.53	0.04

Such massive computations impose a significant burden even for modern computers. Fortunately, it is possible to decrease it significantly with mathematical and programming tricks. The former, implemented in the current version of the MP algorithm used for computations in this article and available together with Svarog from <http://braintech.pl/svarog>, are described in Mallat and Zhang (1993) and Kuś et al. (2013). However, this user-friendly software is still a research system, not aimed at commercial applications. Since the speed of computations was not the major goal here, not all the optimizations were explored yet. Also, as discussed in Section 2.1.2, we used a relatively dense dictionary, increasing significantly the computational burden: with 50 iterations per epoch, decomposition of one overnight recording took about 48 CPU-hours. Since the MP5 algorithm is single threaded, we were able to run 11 concurrent instances on a 12-core computer, thus decomposing in average one overnight recording every 4h approximately. While this may still look like a lot of computing time, let us recall that:

1. MP decomposition is performed only once per each analyzed signal, and as such needs not to be interactive. Using one such general decomposition, we can investigate any structures potentially present in the signal (Section 4.3) in a comfortably interactive mode. Results from one channel of an overnight recording like the one presented in **Figure 8** are computed in seconds.

2. There is still room for significant speed improvements, in the optimization of code (e.g., multithreading or using GPUs) as well as in the adjustments of the decomposition parameters to a particular problem. As an example of the latter we may quote an online procedure for detection of epileptic seizures in commercial EEG software by Persyst (<http://persyst.com>, patent US 6735467), based on a previous version of our MP algorithm (Durka et al., 2001).

4.2. Performance of Detection

Reported performance of sleep spindle detectors depends both on the properties of the detector and on the quality of experts scores. Therefore, the quantitative comparison of detectors is possible only on the same database of EEG recordings and scorings, otherwise the comparison is rather qualitative. It is especially so if the parameters of the detector are tuned to maximize the performance for a given dataset. Another problem in comparison between the results reported in literature is that various authors define the correct detection in different ways via the “window based” type of comparison—mainly in respect to the criteria defining the overlap between detectors and experts scores. We used “signal-sample-based” assessment of performance, since we find it much less ambiguous. In general, the values obtained in “signal-sample-based” type of comparison are more conservative than those obtained in “window based” comparison, as was demonstrated in O’Reilly et al. (in revision). Unfortunately,

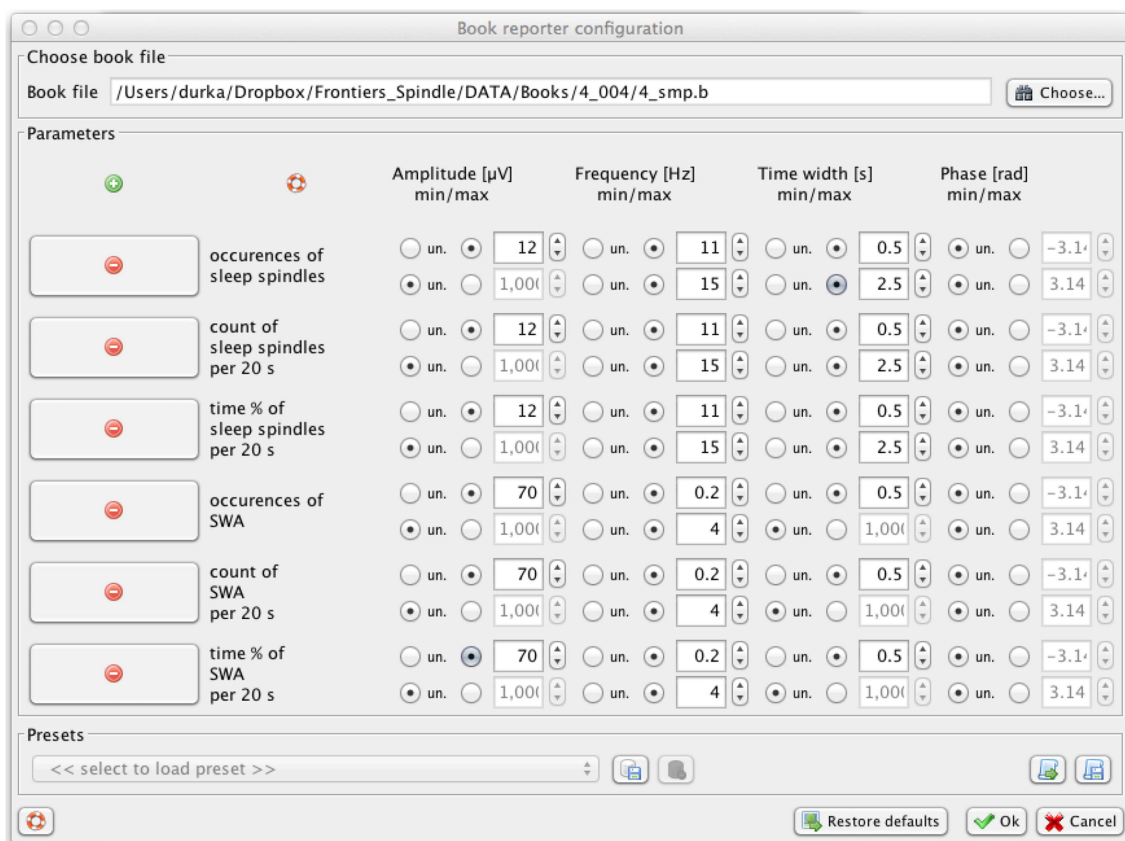


FIGURE 7 | Svarog window for setting the parameters of filters defining the structures chosen from MP decomposition for the profile presented in Figure 8. This functionality operates on the results of a previously computed MP decomposition (Figure 2).

“window based” comparison is the most common and for a long time was the only one considered for assessing the performance of spindles detection presented in literature. To give a general background we cite below some of the results.

For example, one of the first automated detection method with fixed amplitude threshold (Schimicek et al., 1994) showed sensitivity of 89.7% and a specificity of 93.5%. Other sleep spindles detection method using artificial neural networks (Ventouras et al., 2005) presented the sensitivity of the network ranges from 79.2 to 87.5% and specificity from 88.4 to 97%, with the false detection rate ($FDR = \frac{FP}{FP+TP}$) ranging from 2.1 to 21.5%. The methods where variability of sleep spindles amplitude across subjects have been taken into account for detection (e.g., Bódzis et al., 2009) reported sensitivity of 92.9 and 58.4% false detection rate. Another work (Huupponen et al., 2007)³ testing four different detection methods reported optimal sensitivity of 70% for a false detection rate of 32%. Ray et al. (2010) reported a sensitivity of 98.96% for a specificity of 88.49%, with a corresponding 37.2% false

³One should be careful on reading of this paper since the authors call false-positive rate what is usually referred to as false detection rate. False positive rate is generally considered as $FPR = \frac{FP}{FP+TN}$ and therefore, as for specificity, due to huge counts of *TN* relative to other counts, is of little interest for characterizing sleep spindle detectors.

detection rate in detection of sleep spindles in stage II with the minimal amplitude adjusted individually and 3 s scoring windows.

A more direct comparison of the detector presented in this work can be made with the six automatic detectors, known from publications, reimplemented and tested in Warby et al. (2014) (cf. Figure 4). The authors presented “precision-recall” plot obtained with “window based” comparison of the detectors⁴. Our detector would be placed at point (recall = 0.63, precision = 0.47) in that space, which is close to the middle of the automated group consensus curve. Also the *F*₁-score is close to the maximum performance for the auto group consensus. Such result would indicate that the proposed detector is well balanced and close to optimal among the automated detectors, but we have to keep in mind that we compare results for different datasets.

The most meaningful and direct comparison can be made with the four detectors tested in O’Reilly et al. (in revision), since they were tested on exactly the same data set, with same expert scoring, and using the same “signal-sample-based” type of comparison. For the ease of comparison, in Figure 6, we rearranged the original results presented in O’Reilly et al.

⁴Precision is another name for PPV and recall for sensitivity.

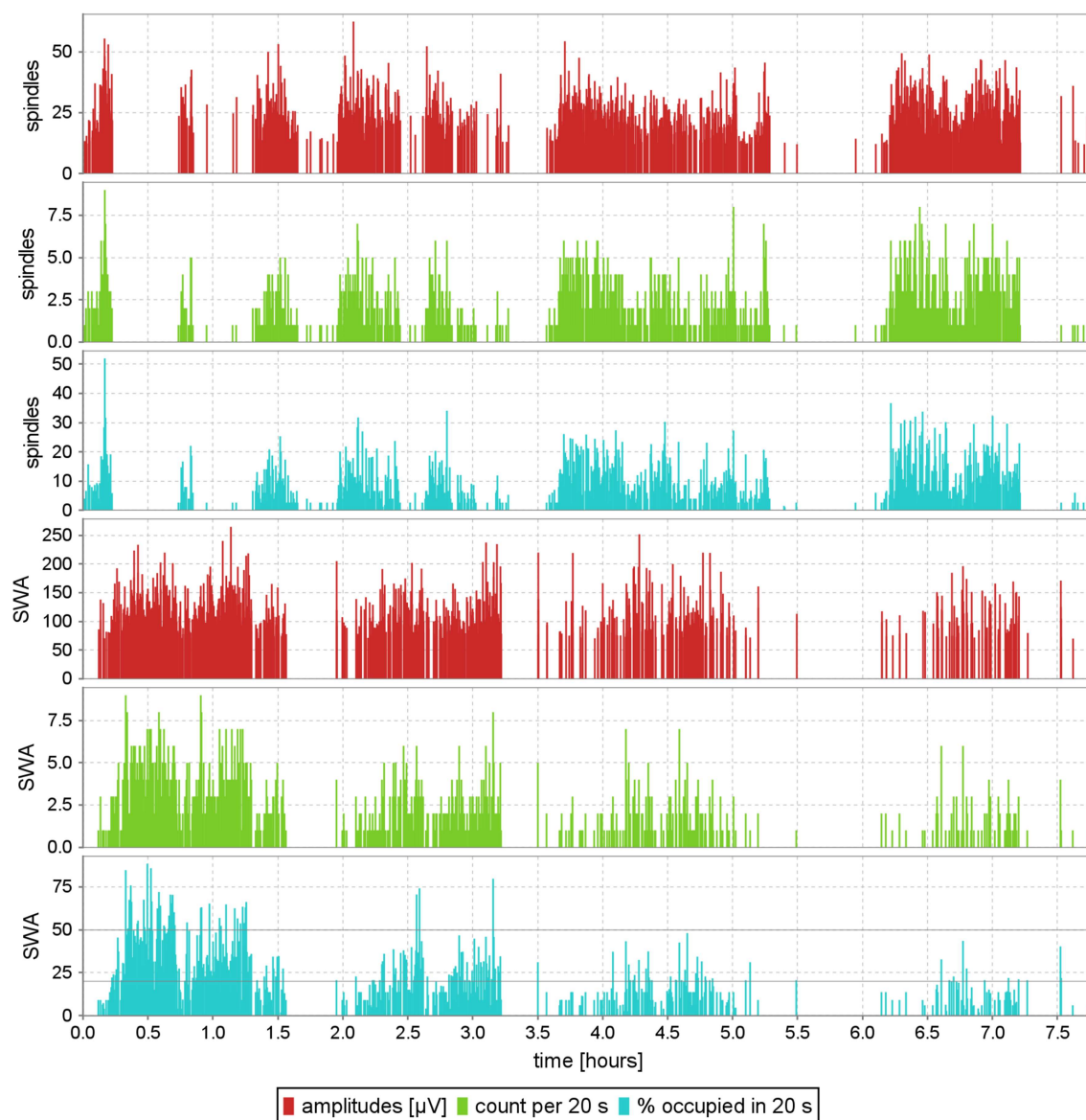


FIGURE 8 | Exemplary EEG profile of sleep, computed automatically without prior removal of artifacts, for the structures defined in

Figure 7; subject #4 of MASS-C1/SS2. Each of the red vertical lines is positioned at the exact time center of a spindle or slow wave, and its height represents the structure's amplitude. Green lines are positioned in 20-s intervals, and their height measures the number of structures detected in the

corresponding epochs. Cyan lines, like the green ones, relate to the properties of 20-s epochs, and give the percent of each epoch's time, occupied by the given structures. This parameter is especially relevant for slow waves, because (Rechtschaffen and Kales, 1968) defined sleep stages 3 and 4 explicitly by 20–50% and 50–100% ranges of this parameter (Malinowska et al., 2009).

(in revision). These detectors were: RMS—based on methodology proposed in Schimicek et al. (1994), RSP—relative spindle power detector based on Devuyst et al. (2011), Sigma—based on the sigma index proposed by Huupponen et al. (2007), and Teager—based on Teager energy operator, as in Ahmed et al. (2009). Comparison of all four classifiers tested by O'Reilly et al. as well as the MP-based classifier presented in this work, shown in **Figure 6**, have the same range of performance measures, if one takes into account the spread of the distribution of the measures, which in fact is quite broad. In our opinion, this fact points to the

limitations of consistency of expert's scorings which were used as the “gold standard,” or to the existence of some characteristics of the recording which affects the decisions of expert, but which are not included in the currently used definition of sleep spindles.

4.3. Universal Parametrization

In the context of a universal parameterization of EEG transients (Durka, 2005) it is also worth mentioning that proposed framework has a potential to solve a variety of important

problems in EEG analysis. Apart from the above examples, it was already shown to significantly improve the quality of EEG inverse solutions if used as a preprocessing and automatic detection of sleep spindles (Durka et al., 2005b), and sensitivity of estimates used in pharmaco EEG (Durka et al., 2002).

We believe that the availability of the free software and exemplary description of a framework for detection of sleep spindles paves the way to novel and creative applications of this high-resolution parametrization, to a large extent compatible with the tradition of visual analysis.

4.4. Data Sharing

Complete software package (with source code) used in this study for computing MP decompositions and generating **Figure 8**, as well as scripts for reading the results of MP decomposition in Matlab and Python (Section 2.1.4), are freely available from <http://braintech.pl/svarog>. Source code of the Svarog interface (in Java) and mp5 program for MP decomposition (in C) is available from <http://git.braintech.pl>.

Polysomnograms and human scoring of sleep spindles used in this study come from MASS database and can be downloaded from <http://ceams-carsm.ca/en/mass>. Access to polysomnographic recordings requires further accreditation from an authorized Ethics Research Board.

References

- Ahmed, B., Redissi, A., and Tafreshi, R. (2009). An automatic sleep spindle detector based on wavelets and the teager energy operator. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2009, 2596–2599. doi: 10.1109/IEMBS.2009.5335331
- Bódizs, R., Körmendi, J., Rigó, P., and Lázár, A. S. (2009). The individual adjustment method of sleep spindle analysis: methodological improvements and roots in the fingerprint paradigm. *J. Neurosci. Methods* 178, 205–213. doi: 10.1016/j.jneumeth.2008.11.006
- Devuyst, S., Dutoit, T., Stenuit, P., and Kerkhofs, M. (2011). Automatic sleep spindles detection—overview and development of a standard proposal assessment method. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2011, 1713–1716. doi: 10.1109/IEMBS.2011.6090491
- Durka, P., Szelenberger, W., Blinowska, K., Androsiuk, W., and Myszka, M. (2002). Adaptive time-frequency parametrization in pharmaco EEG. *J. Neurosci. Methods* 117, 65–71. doi: 10.1016/S0165-0270(02)00075-4
- Durka, P. J. (2005). On the methodological unification in electroencephalography. *Biomed. Eng. Online* 4:15. doi: 10.1186/1475-925X-4-15
- Durka, P. J., and Blinowska, K. J. (1995). Analysis of EEG transients by means of matching pursuit. *Ann. Biomed. Eng.* 23, 608–611. doi: 10.1007/BF02584459
- Durka, P. J., and Blinowska, K. J. (1997). “In pursuit of time-frequency representation of brain signals,” in *Time Frequency and Wavelets in Biomedical Signal Processing*, IEEE Press Series in Biomedical Engineering, ed M. Akay (New York, NY: IEEE press), 389–406.
- Durka, P. J., Ircha, D., and Blinowska, K. J. (2001). Stochastic time-frequency dictionaries for matching pursuit. *IEEE Tran. Signal Process.* 49, 507–510. doi: 10.1109/78.905866
- Durka, P. J., Malinowska, U., Szelenberger, W., Wakarow, A., and Blinowska, K. (2005a). High resolution parametric description of slow wave sleep. *J. Neurosci. Methods* 147, 15–21. doi: 10.1016/j.jneumeth.2005.02.010
- Durka, P. J., Matysiak, A., Montes, E. M., Sosa, P. V., and Blinowska, K. J. (2005b). Multichannel matching pursuit and EEG inverse solutions. *J. Neurosci. Methods* 148, 49–59. doi: 10.1016/j.jneumeth.2005.04.001

Author Contributions

PD has proposed and designed major steps of MP parameterization of EEG transients and detection of spindles, supervised and contributed to the development of the software, designed the current study and wrote most of the text. PR has written the interactive plugin for detection of structures and display of reports from MP decompositions, fixed the Svarog interface to MP and bugs found during preparation of this study, and consulted mathematical aspects of MP. UM contributed to tests of the software, data analysis and interpretation, drafting of the work and reviewing the manuscript. MZ contributed to tests of the software, tested several detection schemes and performed large part of data analysis and comparisons with visual detections. COR performed MP decompositions on MASS database, performed analyzes, and contributed in writing and reviewing the manuscript. JZ adjusted details of the detection algorithm, supervised the comparison with visual detection, performed cross-validation analyzes, and contributed in writing and reviewing the manuscript.

Funding

This work was partially supported from the Polish funds for science.

- Huupponen, E., Gómez-Herrero, G., Saastamoinen, A., Värrä, 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., Várri, A., Himanen, S. L., Hasan, J., Lehtokangas, M., and Saarinen. (2000). Optimization of sigma amplitude threshold in sleep spindle detection. *J. Sleep Res.* 9, 327–334. doi: 10.1046/j.1365-2869.2000.00220.x
- Ibert, C., Ancoli-Israel, S., Chesson, A., and Quan, S., (eds.). (2007). *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, 1st Edn.* Westchester: American Academy of Sleep Medicine.
- Kuś, R., Różański, P., and Durka, P. (2013). Multivariate matching pursuit in optimal Gabor dictionaries: theory and software with interface for EEG/MEG via Svarog. *Biomed. Eng. Online* 12:94. doi: 10.1186/1475-925X-12-94
- Malinowska, U., Chatelle, C., Bruno, M.-A., Noirhomme, Q., Laureys, S., and Durka, P. J. (2013). Electroencephalographic profiles for differentiation of disorders of consciousness. *Biomed. Eng. Online* 12:109. doi: 10.1186/1475-925X-12-109
- Malinowska, U., Klekowicz, H., Wakarow, A., Niemcewicz, S., and Durka, P. (2009). Fully parametric sleep staging compatible with the classic criteria. *Neuroinformatics* 7, 245–253. doi: 10.1007/s12021-009-9059-9
- Mallat, S. G., and Zhang, Z. (1993). Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.* 41, 3397–3415. doi: 10.1109/78.258082
- 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
- Ray, L. B., Fogel, S. M., Smith, C. T., and Peters, K. R. (2010). Validating an automated sleep spindle detection algorithm using an individualized approach. *J. Sleep Res.* 19, 374–978. doi: 10.1111/j.1365-2869.2009.00802.x
- Rechtschaffen, A., and Kales, A., (eds.). (1968). *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages in Human subjects.*

- Washington, DC: National Institutes of Health Publications, US Government Printing Office.
- Schimicek, P., Zeitlhofer, J., Anderer, P., and Saletu, B. (1994). Automatic sleep-spindle detection procedure: aspects of reliability and validity. *Clin. Electroencephalogr.* 25, 26–29.
- 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
- Warby, S. C., Wendt, S. L., Welinder, P., Munk, E. G. S., Carrillo, O., Sorensen, H. B. D., 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
- Żygierewicz, J., Blinowska, K. J., Durka, P. J., Szelenberger, W., Niemcewicz, S., and Androsiuk, W. (1999). High resolution study of sleep spindles. *Clin. Neurophysiol.* 110, 2136–2147.

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.

Copyright © 2015 Durka, Malinowska, Zieleniewska, O'Reilly, Różański and Żygierewicz. 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.



Automated detection of sleep spindles in the scalp EEG and estimation of their intracranial current sources: comments on techniques and on related experimental and clinical studies

Periklis Y. Ktonas^{1*} and Errikos-Chaim Ventouras²

¹ Sleep Study Unit, 1st Psychiatric Clinic, Eginition Hospital, University of Athens Medical School, Athens, Greece

² Department of Biomedical Engineering, Technological Educational Institution of Athens, Athens, Greece

*Correspondence: pktonas@uh.edu

Edited by:

Christian O'Reilly, McGill University, Canada

Reviewed by:

Stuart Fogel, Western University, Canada

Christian O'Reilly, McGill University, Canada

Keywords: sleep spindle, automated detection, inverse problem, intracranial sources, experimental studies, clinical studies

INTRODUCTION

Sleep spindles are short bursts of sleep EEG activity in the range of 11–15 Hz, reflecting central nervous system integrity and considered to promote sleep continuity, learning and memory consolidation processes. This contribution comments on the automated detection of sleep spindles and their intracranial sources, as well as on experimental and clinical studies for the characterization of spindles and their sources, and the study of their functional significance. Supporting literature is provided wherever appropriate, although comprehensive review is out of the scope of this opinion paper.

AUTOMATED DETECTION OF SLEEP SPINDLES

DETECTION METHODS

Visual EEG analysis heuristics, such as counting the number of peaks of the EEG signal within a time window, or counting the number of successive EEG waves having a specific amplitude and period within that window, can be utilized for spindle detection, provided relatively high sampling rates beyond the Nyquist criterion are chosen, e.g., 250 Hz (Principe and Smith, 1986; Ktonas, 1996). However, appropriate EEG pre-filtering with wide-band (low-Q) bandpass filters (Shirakawa et al., 1987) may be necessary. Techniques which are based on human pattern recognition can present problems because there is no explicit definition for a sleep spindle.

Spindle morphology may vary between the so-called “fast” and “slow” spindles, across subjects, with age and health condition (Nicolas et al., 2001; Ktonas et al., 2009). Appropriate initialization procedures in the detection system, such as adaptively adjusting amplitude or frequency parameters per subject, may help (Ray et al., 2010). Expert system-based approaches, incorporating complex domain knowledge, might be able to address these problems.

Spindle detection can be based on spectral analysis implemented via the Fast Fourier transform (FFT). Such techniques, although simple to implement, exhibit problems of FFT-based spectral analysis: inability to detect short (“phasic”) EEG events, unless the time window of the analysis is short as well (which may result in problems of frequency resolution), and difficulty in distinguishing between diffuse “background” activity in the spindle frequency band and well-defined spindles. These problems can be addressed by using time-frequency analysis techniques (e.g., wavelets) as well as matching pursuit procedures (which can be viewed as a generalization of wavelet analysis), although questions still remain as to the “best” mother wavelet or number and kind of atoms to use.

The above methods rely on the a priori knowledge of some electrographic characteristics defining the sleep spindle. The artificial neural network (ANN)

approach for detection may not depend on any such explicit knowledge (e.g., Ventouras et al., 2005). However, the generalization capability of ANN-based methods, which cannot be evaluated analytically (as in an expert system-based method), is not “guaranteed” and it depends in a quite non-linear way on the structure of the ANN architecture and on the training data. Combinations of possibly more than one ANN systems as pre-processors allowing any “spindle-like” waveform to be further evaluated, followed by a knowledge-based system mimicking an expert (or a consensus of experts) for more elaborate analysis, appear to be promising approaches.

A successful detection system exhibits mostly true positive spindle detections (TPs) and very few false positive spindle detections (FPs). We define TP performance (TPP) as follows: (the number of TPs)/(the number of spindles detected by the visual scorers). If possible, the visually detected spindles should reflect the consensus of several scorers. Visual scoring is still the “gold standard” to compare automated detection systems to, despite the fact that experts often make mistakes, may be biased using ill-defined procedures, and may not be always consistent. We define FP performance (FPP) as follows: (the number of FPs)/(the sum of FPs and TPs). TPP and FPP figures should be provided for testing data, which should be separate from training data. Both training

and testing data should contain records of several subjects, of various ages and pathologies, as sleep spindle morphology may vary as a function of age and pathology. In addition, the data should contain EEG epochs exhibiting various kinds of recording artifacts, such as movement and muscle (EMG) activity, since automated detection systems should be capable of analyzing routine sleep EEG records obtained in an artifact-prone clinical environment. Therefore, artifact detection or rejection capabilities should be incorporated into detection systems.

Deciding on “optimum” TPP and FPP figures is not straightforward. Satisfactory TPP and FPP figures should relate to the use of the automated system. For example, if the system is to be used for automated sleep staging in routine sleep EEG analysis, it may not be necessary to detect each and every sleep spindle, but enough of them so that a 30-s EEG epoch can be accurately assessed as sleep stage 2 (Rechtschaffen and Kales, 1968). However, in order not to misinterpret as sleep stage 2 other sleep stages where no spindles are expected, a “relatively good” FPP (say, less than 20%) may be appropriate. In cases where not missing spindles is of paramount importance, as in sleep EEG records of patients with neurological or psychiatric disorders where there is a paucity of spindles (e.g., dementia, schizophrenia), increasing TPP and decreasing FPP may be necessary. This could apply, for example, to clinical studies on the effect of pharmacotherapy in schizophrenia, where effects on thalamic centers involved in sleep spindle generation are investigated (Ferrarelli and Tononi, 2011). Assuming that high TPP and low FPP figures might necessitate a complicated system structure, it should be of interest to develop systems exhibiting some kind of modularity, whereby TPP and FPP could be altered depending on the use.

EXPERIMENTAL AND CLINICAL STUDIES

A reliable detection system can contribute to the effective and accurate quantification of sleep spindle occurrence patterns, either through spindle counts or spindle density figures (i.e., spindle number/time window of observation). It can also aid in topographical studies of “slow” and “fast” spindles, which should be of interest

(Zeithofer et al., 1997), as well as contribute in tracking the propagation of sleep spindles across the scalp, for the study of sleep spindle dynamics (O’Reilly and Nielsen, 2014). In some cases, the spindle sequence pattern (e.g., how inter-spindle time intervals are distributed in time) might be of importance (Ktonas et al., 2000), especially if spindle generation mechanisms are being studied. There is evidence that sleep spindles are generated through the interaction of cortico-thalamo-cortical neuronal networks, and that the so-called Slow Wave Oscillation (SWO), a cortical EEG rhythm of frequency content less than 1 Hz, serves as a “pacemaker” for the thalamic reticular nucleus to generate spindles (Steriade and Amzica, 1998). Studying sequence patterns in inter-spindle time intervals can provide information about SWO intra-frequency dynamics which may relate to cortical processes of interest, such as learning and memory consolidation (Molle et al., 2011).

Systems should provide the capability of extracting specific electrographic parameters from the detected spindles, such as mean amplitude, intra-spindle frequency and spindle length, which may relate to EEG generating mechanisms possibly affected by an experimental procedure (e.g., sleep deprivation, pharmacotherapy) or a neurological/psychiatric disorder. Accordingly, any changes in spindle mean amplitude may relate to changes in cortical processes, while changes in intra-spindle frequency and spindle length may relate to changes in thalamic or thalamo-cortical processes (Steriade and Amzica, 1998). Given their electrographic shape, sleep spindles could be viewed as amplitude-modulated and frequency-modulated (AM/FM) signals. Therefore, methodology for the analysis of analytic signals (e.g., Hilbert transforms) as well as time-frequency analysis techniques provide the opportunity of extracting parameters related to the instantaneous envelope and instantaneous frequency of spindles, allowing the possibility to study pathological processes that might affect such parameters, as, for example, in schizophrenia, dementia and cognitive dysfunction (Ktonas et al., 2009; Ferrarelli and Tononi, 2011; Carvalho et al., 2014).

ESTIMATION OF INTRACRANIAL CURRENT SOURCES FOR SLEEP SPINDLES

ESTIMATION METHODS

The non-invasive estimation of intracranial current sources for sleep spindles can be achieved by solving the inverse bioelectromagnetic problem, based on scalp EEG or MEG (magnetoencephalography) measurements. The sources are usually modeled as current dipoles. In the equivalent current dipole (ECD) approach, the number, location, amplitude and orientation of dipoles are to be determined. A set of dipoles is selected which best conforms to an optimization criterion.

In the Distributed Source Model (DSM) approach no restrictions are imposed on the number of sources to be computed. Optimization techniques are adopted for solving this highly under-determined problem, incorporating mathematically and/or biophysically inspired restrictions, but without certainty that no distribution other than the selected one could be closer to the real underlying distribution. Low-Resolution Electromagnetic Tomography (LORETA) is a DSM method selecting the solution which minimizes the Laplacian of the depth-weighted sources. Based on the assumption that contiguous neuronal assemblies have correlated activity, LORETA provides solutions that might be “over-smoothed.” Since anatomically contiguous areas can be functionally distinct, concurrent activity in such contiguous areas must be dealt with attention when inspecting the results of LORETA. Other DSM methods, like dynamic SPM (dSPM) and standardized LORETA (sLORETA), compute statistical scores indicating locations where activity would occur with low error probability, therefore creating statistical parametric maps which can provide more focused loci of activity than LORETA. Taking into account the rather diffuse distribution of spindle cortical activity, DSM methods seem more appropriate for spindle source estimation than ECD methods, since ECD methods limit the number of sources that can be investigated and, in order to perform adequately, the number of sources must be inferred a priori (Michel et al., 2004).

EXPERIMENTAL AND CLINICAL STUDIES

Source estimation techniques can be used to elucidate plausible neural generation mechanisms for sleep spindles and, in particular, the electrogenesis of “slow” and “fast” spindles. LORETA based on EEG has provided indications that fast (slow) spindle source activity is located posteriorly (anteriorly) in the cortex (Durka et al., 2005; Ventouras et al., 2010). Studies based on MEG data (Manshanden et al., 2002; Urakami, 2008) have found that four source areas, located in parieto-central and fronto-central cortical regions, bilaterally, adequately explain most of the variation in spindles, although indications for considering both slow and fast spindle source activity as a single event were provided using MEG data (Gomenyuk et al., 2009). However, the inversion of simultaneous EEG and MEG recordings (Dehghani et al., 2010) has found that there are significant differences between sources derived from EEG and those derived from MEG.

Although there is some degree of similarity among the source areas detected by the various studies, there is a need for providing a comparative analysis of a comprehensive set of inversion methods applied to an extensive set of data because of the different principles on which the various methods operate. Along these lines, the concurrent recording of EEG and MEG should be pursued. Similarly, several studies have used concurrent EEG and fMRI recordings, investigating the fMRI-obtained brain activation during sleep spindles (Caporro et al., 2012). EEG/MEG modalities are generally restrained to cortical imaging. However, the generators of spindles are thought to be thalamic and, therefore, not accessible to EEG/MEG. Concurrent EEG and fMRI recordings can provide information on “spindle-coincident” activation in sub-cortical formations, such as the thalamus. Therefore, the limitations of the bioelectromagnetic inverse problem methodologies can be surpassed, providing indications for relations of “slow” and “fast” spindles to thalamic and cortical activity (Schabus et al., 2007). Consequently, such studies should be actively pursued and are expected to significantly elucidate spindle generation mechanisms.

Application of inversion techniques in patient populations should be encouraged,

as in investigating the cortex involvement in the asymmetry of spindles after hemispheric stroke (Urakami, 2009) and the generation of spindles in temporal lobe epilepsy (Del Felice et al., 2013). A topic that has not yet been addressed concerns the extraction of parameters related to the phenomenology of intracranial current sources. Accordingly, it might be of interest to compute measures of current source spread and intensity as a function of time (along the duration of a spindle). Such approaches could help in differentiating healthy controls from patient populations, and in differentiating among various patient populations as well.

SUMMARY

This contribution provided comments on methodological issues related to the automated identification and characterization of sleep spindles and their intracranial sources, and to the understanding of their functional significance. Specific guidelines were presented for the computer-based detection and analysis of spindles and their intracranial sources, as well as for related experimental and clinical studies.

AUTHOR CONTRIBUTIONS

Periklis Y. Ktonas and Errikos-Chaim Ventouras contributed to the conception and design of the work, as well as to drafting and critically revising it, and to providing final approval of the version to be published. They both agree to be accountable for all aspects of the work.

REFERENCES

- Caporro, M., Haneef, Z., Yeh, H. J., Lenartowicz, A., Buttinelli, C., Parvizi, J., et al. (2012). Functional MRI of sleep spindles and K-complexes. *Clin. Neurophysiol.* 123, 303–309. doi: 10.1016/j.clinph.2011.06.018
- Carvalho, D. Z., Gerhardt, G. J., Dellagustin, G., de Santa-Helena, E. L., Lemke, N., Segal, A. Z., et al. (2014). Loss of sleep spindle frequency deceleration in Obstructive Sleep Apnea. *Clin. Neurophysiol.* 125, 306–312. doi: 10.1016/j.clinph.2013.07.005
- Dehghani, N., Cash, S. S., Chen, C. C., Hagler Jr, D. J., Huang, M., Dale, A. M., et al. (2010). Divergent cortical generators of MEG and EEG during human sleep spindles suggested by distributed source modeling. *PLoS ONE* 5:7. doi: 10.1371/journal.pone.0011454
- Del Felice, A., Arcaro, C., Storti, S. F., Fiaschi, A., and Manganotti, P. (2013). Slow spindles' cortical generators overlap with the epileptogenic zone in temporal epileptic patients: an electrical source

- imaging study. *Clin. Neurophysiol.* 124, 2336–2344. doi: 10.1016/j.clinph.2013.06.002
- Durka, P. J., Matysiak, A., Martinez Montes, A., Valdes Sosa, P., and Blinowska, K. J. (2005). Multichannel matching pursuit and EEG inverse solutions. *J. Neurosci. Methods* 148, 49–59. doi: 10.1016/j.jneumeth.2005.04.001
- Ferrarelli, F., and Tononi, G. (2011). The thalamic reticular nucleus and schizophrenia. *Schizophr. Bull.* 37, 306–315. doi: 10.1093/schbul/sbq142
- Gomenyuk, V., Roth, T., Moran, J. E., Jefferson, C., Bowter, S. M., Tepley, N., et al. (2009). Cortical locations of maximal spindle activity: magnetoencephalography (MEG) study. *J. Sleep Res.* 18, 245–253. doi: 10.1111/j.1365-2869.2008.00717.x
- Ktonas, P. Y. (1996). “Computer-based recognition of EEG patterns,” in *Frontier Science in EEG: Continuous Waveform Analysis (EEG Suppl. 45)*, eds R. M. Dasheiff and D. J. Vincent (Amsterdam: Elsevier), 23–35.
- Ktonas, P. Y., Golemati, S., Xanthopoulos, P., Sakkalis, V., Ortigueira, M. D., Tsekou, H., et al. (2009). Time-frequency analysis methods to quantify the time-varying microstructure of sleep EEG spindles: possibility for dementia biomarkers? *J. Neurosci. Methods* 185, 133–142. doi: 10.1016/j.jneumeth.2009.09.001
- Ktonas, P. Y., Paparrigopoulos, T., Monoyiou, E. A., Bergiannaki, J. D., and Soldatos, C. R. (2000). Sleep spindle incidence dynamics: a pilot study based on a Markovian analysis. *Sleep* 23, 419–423.
- Manshanden, I., de Munck, J. C., Simon, N. R., and Lopes da Silva, F. H. (2002). Source localization of MEG sleep spindles and the relation to sources of alpha band rhythms. *Clin. Neurophysiol.* 113, 1937–1947. doi: 10.1016/S1388-2457(02)00304-8
- Michel, C. M., Murray, M. M., Lantz, G., Gonzalez, S., Spinelli, L., and Grave de Peralta, R. (2004). EEG source imaging. *Clin. Neurophysiol.* 115, 2115–2222. doi: 10.1016/j.clinph.2004.06.001
- 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
- Nicolas, A., Petit, D., Rompre, S., and Montplaisir, J. (2001). Sleep spindle characteristics in healthy subjects of different age groups. *Clin. Neurophysiol.* 112, 521–527. doi: 10.1016/S1388-2457(00)00556-3
- O'Reilly, C., and Nielsen, T. (2014). 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
- Principe, J. C., and Smith, J. R. (1986). SAMICOS—a sleep analyzing microcomputer system. *IEEE Trans. Biomed. Eng.* 33, 935–941. doi: 10.1109/TBME.1986.325664
- Ray, L. B., Fogel, S. M., Smith, C. T., and Peters, K. R. (2010). Validating an automated sleep spindle detection algorithm using an individualized approach. *J. Sleep Res.* 19, 374–378. doi: 10.1111/j.1365-2869.2009.00802.x
- Rechtschaffen, A., and Kales, A. (1968). *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*. Washington, DC: Public Health Service, U.S. Government Printing Office.

- Schabus, M., Dang-Vu, T. T., Albouy, G., Baletau, 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
- Shirakawa, S., Smith, J. R., and Azumi, K. (1987). Filter transient response to EEG waveforms. *Int. J. Biomed. Comput.* 20, 87–95. doi: 10.1016/0020-7101(87)90017-1
- Steriade, M., and Amzica, F. (1998). Coalescence of sleep rhythms and their chronology in corticothalamic networks. *Sleep Res. Online* 1, 1–10.
- Urakami, Y. (2008). Relationships between sleep spindles and activities of cerebral cortex as determined by simultaneous EEG and MEG recording. *J. Clin. Neurophysiol.* 25, 13–24. doi: 10.1097/WNP.0b013e318162a8a4
- Urakami, Y. (2009). Relationships between sleep spindles and activities of the cerebral cortex after hemispheric stroke as determined by simultaneous EEG and MEG recordings. *J. Clin. Neurophysiol.* 26, 248–256. doi: 10.1097/WNP.0b013e3181af209c
- Ventouras, E., Ktonas, P., Tsekou, H., Paparrigopoulos, T., Kalatzis, I., and Soldatos, C. R. (2010). Independent component analysis for source localization of EEG sleep spindle components. *Comput. Intell. Neurosci.* 2010:329436. doi: 10.1155/2010/329436
- 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
- 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: 08 September 2014; accepted: 24 November 2014; published online: 10 December 2014.

Citation: Ktonas PY and Ventouras E-C (2014) Automated detection of sleep spindles in the scalp EEG and estimation of their intracranial current sources: comments on techniques and on related experimental and clinical studies. *Front. Hum. Neurosci.* 8:998. doi: 10.3389/fnhum.2014.00998

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

Copyright © 2014 Ktonas and Ventouras. 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.

Sleep spindle and K-complex detection using tunable Q-factor wavelet transform and morphological component analysis

Tarek Lajnef¹, Sahbi Chaibi¹, Jean-Baptiste Eichenlaub², Perrine M. Ruby³, Pierre-Emmanuel Aguera³, Mounir Samet¹, Abdennaceur Kachouri^{1,4} and Karim Jerbi^{3,5*}

¹ LETI Lab, Sfax National Engineering School, University of Sfax, Sfax, Tunisia, ² Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA, ³ DYCOG Lab, Lyon Neuroscience Research Center, INSERM U1028, UMR 5292, University Lyon I, Lyon, France, ⁴ Electrical Engineering Department, Higher Institute of Industrial Systems of Gabes, University of Gabes, Gabes, Tunisia, ⁵ Psychology Department, University of Montreal, Montreal, QC, Canada

OPEN ACCESS

Edited by:

Christian O'Reilly,
McGill University, Canada

Reviewed by:

Marek Adamczyk,
Max Planck Institute of Psychiatry,
Germany
Ivan Selesnick,
New York University, USA
Magdalena Zieleniewska,
University of Warsaw, Poland

*Correspondence:

Karim Jerbi,
Département de Psychologie,
Université de Montréal, Pavillon
Marie-Victorin C. P. 6128, Succursale
Centre-ville, Montréal, QC H3C 3J7,
Canada
karim.jerbi@umontreal.ca

Received: 09 February 2015

Accepted: 06 July 2015

Published: 28 July 2015

Citation:

Lajnef T, Chaibi S, Eichenlaub J-B, Ruby PM, Aguera P-E, Samet M, Kachouri A and Jerbi K (2015) Sleep spindle and K-complex detection using tunable Q-factor wavelet transform and morphological component analysis. *Front. Hum. Neurosci.* 9:414. doi: 10.3389/fnhum.2015.00414

A novel framework for joint detection of sleep spindles and K-complex events, two hallmarks of sleep stage S2, is proposed. Sleep electroencephalography (EEG) signals are split into oscillatory (spindles) and transient (K-complex) components. This decomposition is conveniently achieved by applying morphological component analysis (MCA) to a sparse representation of EEG segments obtained by the recently introduced discrete tunable Q-factor wavelet transform (TQWT). Tuning the Q-factor provides a convenient and elegant tool to naturally decompose the signal into an oscillatory and a transient component. The actual detection step relies on thresholding (i) the transient component to reveal K-complexes and (ii) the time-frequency representation of the oscillatory component to identify sleep spindles. Optimal thresholds are derived from ROC-like curves (sensitivity vs. FDR) on training sets and the performance of the method is assessed on test data sets. We assessed the performance of our method using full-night sleep EEG data we collected from 14 participants. In comparison to visual scoring (Expert 1), the proposed method detected spindles with a sensitivity of 83.18% and false discovery rate (FDR) of 39%, while K-complexes were detected with a sensitivity of 81.57% and an FDR of 29.54%. Similar performances were obtained when using a second expert as benchmark. In addition, when the TQWT and MCA steps were excluded from the pipeline the detection sensitivities dropped down to 70% for spindles and to 76.97% for K-complexes, while the FDR rose up to 43.62 and 49.09%, respectively. Finally, we also evaluated the performance of the proposed method on a set of publicly available sleep EEG recordings. Overall, the results we obtained suggest that the TQWT-MCA method may be a valuable alternative to existing spindle and K-complex detection methods. Paths for improvements and further validations with large-scale standard open-access benchmarking data sets are discussed.

Keywords: sleep, spindles, K-complex, automatic detection, electroencephalography (EEG), tunable Q-factor wavelet transform (TQWT), morphological component analysis (MCA), neural oscillations

Introduction

We spend about one third of our lives sleeping. Luckily, and as might be expected of an efficient organism, the time we spend sleeping is not wasted idling. Sleep plays a functional role mediating a range of cognitive processes including learning and memory consolidation (Maquet, 2001; Walker and Stickgold, 2004; Diekelmann and Born, 2010; Fogel et al., 2012; Albouy et al., 2013; Rasch and Born, 2013; Stickgold and Walker, 2013; Alger et al., 2014; Vorster and Born, 2015), problem solving (Cai et al., 2009), sensory processing (Bastuji et al., 2002; Perrin et al., 2002; Ruby et al., 2013a; Kouider et al., 2014) and dreaming (Nielsen and Levin, 2007; Hobson, 2009; Nir and Tononi, 2010; Blagrove et al., 2013; Ruby et al., 2013b; Eichenlaub et al., 2014a,b). Sleep disorders, as well as the mere lack of sleep, can have serious effects on our health, both by deteriorating the proper function of sleep-related brain processes and indirectly by being a risk factor for conditions such as weight gain, hypertension and diabetes (Anderson, 2015). The utmost importance of a good night's sleep is therefore unquestionable. However, many questions related to the mechanisms and role of the numerous electrophysiological signatures of sleep are still outstanding. The standard approach to monitor sleep is the use of Polysomnography (PSG) which combines multiple physiological recordings including electroencephalogram (EEG), electromyogram (EMG), electrocardiogram (ECG), and electrooculogram (EOG). In addition to being a central diagnosis tool for a range of sleep disorders (such as narcolepsy, idiopathic hypersomnia and sleep apnea), PSG is a valuable tool for sleep research performed in healthy individuals. In particular, the analysis of sleep EEG signals helps us understand its neurophysiological basis and functional role. Macro and micro-structures are present in sleep signals at various temporal scales. Macro structure analysis often refers to sleep staging, i.e., the segmentation of brain signals into 20 s or 30 s-long periods that represent different sleep stages, each with distinct cerebral signatures. On the other hand, micro structure analyses of brain signals during sleep consists of detecting short-lived microscopic events often considered to be hallmarks of specific sleep stages and of sleep-related cognitive processes, as well as potential signs of sleep anomalies. K-complexes and sleep spindles are among the most prominent micro-events studied in sleep studies, not only for their importance in sleep stage scoring (as they predominantly occur during S2 sleep stage), but also for their importance in the diagnosis of sleep disorders and the exploration of the functional role of sleep.

According to the American Academy of Sleep Medicine (AASM) (Iber et al., 2007), Sleep spindles are defined as: "A train of distinct waves having a frequency of 11–16 Hz with a duration ≥ 0.5 s, usually maximal in amplitude over central brain regions." These waveforms, which are controlled by thalamo-cortical loops (e.g., Steriade, 2003, 2005; Barthó et al., 2014), are the subject of an active area of investigation that seeks to understand the mechanisms and functions of the sleeping brain. Numerous studies have shown that sleep spindles have an important role in memory consolidation during sleep (Schabus et al., 2004; Morin et al., 2008; Diekelmann et al., 2009; Diekelmann and

Born, 2010; Barakat et al., 2011; Fogel et al., 2014; Lafortune et al., 2014). Moreover, sleep spindle characteristics undergo age-related changes (e.g., Seeck-Hirschner et al., 2012; Martin et al., 2013). Other studies suggest that sleep spindles are clinically important given that alterations in their density (number per minute) may be a symptom of neurological disorders such as dementia (e.g., Ktonas et al., 2009, 2014; Latreille et al., 2015), schizophrenia (e.g., Ferrarelli et al., 2010; Ferrarelli and Tononi, 2011), depression (Riemann et al., 2001), stroke recovery, mental retardation, and sleep disorders (De Gennaro and Ferrara, 2003).

K-complexes are defined by the AASM as "A well delineated negative sharp wave immediately followed by a positive component with a total duration ≥ 0.5 s, typically maximal at frontal electrodes" (Iber et al., 2007). The precise role of K-complexes in sleep is still a matter of debate. Some studies consider them as an arousal response, since they are often followed by micro-awakenings (Halász, 2005). Others give K-complexes a sleep "protection" function (Jahnke et al., 2012). Single-unit recordings during human sleep suggest that K-complexes may represent isolated down-states (Cash et al., 2009).

The ability to reliably detect the occurrence of sleep spindles and K-complexes in EEG recordings is therefore of major importance in a wide range of sleep investigations, ranging from basic research to clinical applications. Visual annotation of sleep spindles and K-complexes is tedious, time consuming, subjective and prone to human errors. The inter-agreement between multiple scorers (for spindles and K-complex visual marking) reported in the literature is relatively low (Zygierevicz et al., 1999; Devuyst et al., 2010; Warby et al., 2014). Therefore, as in sleep staging (e.g., O'Reilly et al., 2014; Lajnef et al., 2015), automatic or semi-automatic identification procedures are of great utility for the detection of sleep spindles and K-complexes. Approaches based on band-pass filtering and thresholding have been proposed for both spindles and K-complex detection (e.g., Huupponen et al., 2000; Devuyst et al., 2010). Template-based filtering using matching pursuit methods has also been used proposed (e.g., Schönwald et al., 2006). Other filtering approaches based on continuous wavelet transforms (CWTs) have also been explored (Erdamar et al., 2012). Moreover, signal classification methods have been used to detect K-complexes or spindles, for instance, using artificial neural networks (ANN) (e.g., Günes et al., 2011), Support Vector Machines (SVMs) (e.g., Acir and Güzelis, 2004) and decision-trees (Duman et al., 2009). Interestingly, only a handful of studies have investigated the detection of K-complex and spindles simultaneously using a common methodological framework (Jobert et al., 1992; Koley and Dey, 2012; Jaleel et al., 2013; Camilleri et al., 2014; Parekh et al., 2015).

In this study we propose a framework for joint spindle and K-complex detection. The proposed method combines a recently introduced discrete wavelet transform (DWT) known as the Tunable Q-factor Wavelet Transform (TQWT) (Selesnick, 2011a) with Morphological Component Analysis (MCA). This combination provides a natural and efficient way to decompose the EEG signal into transient (K-complex) and oscillatory (spindle) components. The results we obtain with full-night sleep EEG recordings from 14 participants demonstrate the

utility and added-value of the proposed method. Our method also performed well when compared with a standard spindle detection method and when applied to a publicly available spindle and K-complex data set.

Materials and Methods

K-Complex and Sleep Spindle Detection Method Overview

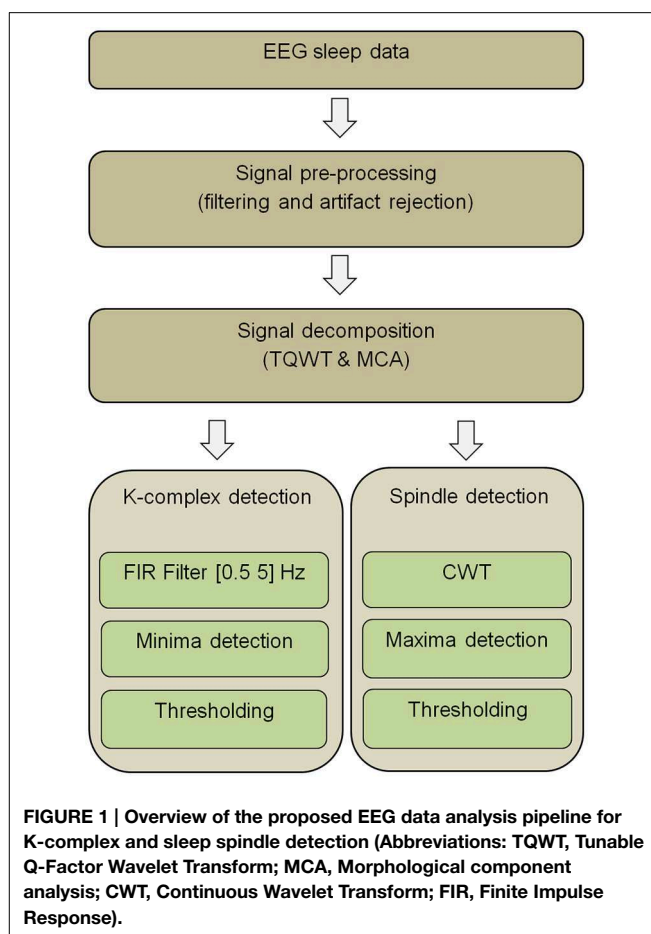
The main steps of the K-complex and spindle detection pipeline are presented in **Figure 1**. First, EEG segments are filtered so as to reduce the effect of potential artifacts. The filtered signals are then decomposed into oscillatory and transient components by combining a TQWT with MCA. Next, applying FIR filtering to the transient component unveils K-complex events, while applying a CWT to the oscillatory component unravels spindle events. The appropriate detection thresholds that need to be used in the final step are determined by plotting sensitivity against false discovery rate (FDR) for a range of potential thresholds [an approach akin to Receiver Operating Characteristic (ROC) curves] calculated from a subset of the data (training set). The ROC-like curves are obtained by repeatedly measuring sensitivity and FDR while varying the threshold parameters and using expert visual marking of K-complexes and spindles as ground truth. The steps that make up the proposed pipeline (**Figure 1**) are described in detail the next sections.

EEG Sleep Recordings Data Acquisition

The EEG data used in this study was collected from 14 healthy subjects aged 29.2 ± 8 years, all recorded at the DyCog Lab of the Lyon Neuroscience Research Center (CRNL, Lyon, France) with a sampling frequency of 1000 Hz. The data acquisition was part of a research program exploring cognition during sleep (Eichenlaub et al., 2012, 2014b; Ruby et al., 2013a,b). The EEG component of the polysomnography recordings across the 14 subjects were visually scored by an expert in successive windows of 30-s using the R and K guidelines (Rechtschaffen and Kales, 1968). The sleep staging step here gave us the possibility to choose to run our detection pipeline (a) exclusively on S2 sleep segments, or (b) on all sleep stages (as would be the case in the absence of sleep scoring). In other words, sleep staging is not a required pre-processing step for the detection method proposed here. Unless otherwise stated, all the analyses described were based on the standard EEG C3 channel.

Splitting the Data into Training and Test Sets

To evaluate the performance of the detection procedure, we divided the data base into a training set (used to derive optimal thresholds via ROC-like curves) and a test set used to compute the performance of the method. Thirty S2 segments and 15 non-S2 segments were randomly selected from the data of each of the 14 individuals (i.e., 630 sleep EEG data segments in total: 420 S2 segments and 210 non-S2 segments). This ensured a balanced representation of data from across all subjects. Note, that our emphasis on S2 stems from the fact that it is the sleep stage of primary interest for the detection of spindles and K-complexes.



As a general rule, we used equally sized training and test sets (210 segments for testing and 210 segments for training). The training and associated test sets consisted either of S2 segments only (scenario 1), or of a mixture of S2 and non-S2 segments (scenario 2). Note that in this second case, the test and training sets contained 105 S2 and 105 non-S2 segments. As we spend approximately half our sleep in stage 2 (Carskadon and Dement, 2011), this proportion was representative of using random sampling of sleep segments. In addition, for practical purposes, we also explored the effect of reducing the size of the training set to evaluate minimal training requirements (scenario 3).

Signal Preprocessing and Visual Annotation of Microstructures

The presence of various artifacts in sleep EEG adversely affects both visual and automatic detection of spindles and K-complexes. The EEG signals were therefore band-pass filtered with low and high cutoff frequencies at 0.2 and at 40 Hz, respectively. This was followed by visual inspection in search of potential remaining artifacts. In addition, visual annotation of K-complex and spindles on the EEG traces was independently performed by two experts and used as two alternative benchmarks. To facilitate this procedure, we designed a graphical user interface (GUI), which was used by our experts to visually explore the EEG data segments and identify K-complex

and spindles events. The results of the visual detection were saved to two separate text files containing segment number, start and end times/sample for each event. For example, the visual annotation by Expert 1 of the 420 segments of S2 sleep across all subjects led to the identification of 437 Spindles and 293 K-complexes (see details in **Table 1**).

EEG Signal Decomposition Using TQWT and MCA

K-complexes and spindles are microstructures that are morphologically different. One major difference is that K-complexes are transient while spindles are oscillatory. To exploit this distinction, we set out to combine the recently introduced TQWT discrete wavelet with MCA in order to conveniently decompose any given EEG segment into two signals; a K-complex channel and a spindle channel. The decomposition via TQWT and MCA is described below.

Tunable Q-factor Wavelet Transform (TQWT)

TQWT is a flexible fully DWT that was recently introduced by Selesnick (2011a,c), for which the Q-factor of the wavelet is easily tuned and adapted to the signal being investigated. In principle, a high Q-factor transform is suitable for oscillatory signals, whereas transient signals are modeled using low Q-factor wavelets. Like the dyadic DWT, TQWT consists of iteratively applying two-channel filter bank, where the low-pass output of each filter bank is the input to the next filter bank. A sub-band is then defined as the output signal of each high pass filter. Considering J the number of filter banks, there will be $J + 1$ sub-bands, i.e., J sub-bands coming from the high-pass filter output signal of each filter bank and the low-pass filter output signal of the final filter bank. At each level, the generation of low-pass sub-band $C^j[n]$ uses a low-pass filter $H_0^j(w)$ followed by low-pass (LP) scaling α , and similarly the generation of high-pass sub-band $d_j[n]$ uses a high-pass $H_1^j(w)$ and high-pass (HP) scaling β . $H_0^j(w)$ and $H_1^j(w)$ are defined as follows (Selesnick, 2011a):

$$H_0^{(j)}(w) = \begin{cases} \prod_{m=0}^{j-1} H_0\left(\frac{w}{\alpha^m}\right), & |w| \leq \alpha^j \pi \\ 0, & \alpha^j \pi < |w| < \pi \end{cases} \quad (1)$$

and

$$H_1^{(j)}(w) = \begin{cases} H_1\left(\frac{w}{\alpha^{j-1}}\right) \prod_{m=0}^{j-2} H_0\left(\frac{w}{\alpha^m}\right), & (1-\beta)\alpha^{j-1} \leq |w| \leq \alpha^{j-1}\pi \\ 0, & \text{for others } w \in [-\pi, \pi]. \end{cases} \quad (2)$$

TABLE 1 | Example of visual annotation results by Expert 1 based on 420 S2 segments.

	Spindles	K-complexes
Number of segments with	244	199
Number of segments without	176	221
Number of detected events	437	293

All main parameters were computed as described in the original study by Selesnick (2011a) and the user-manual of the TQWT toolbox (Selesnick, 2011b). Three key parameters that need to be set are the following:

- (i) **Q-factor:** In the context of the present TQWT implementation, the Q-factor is theoretically defined as $Q = (2 - \beta)/\beta$. As it reflects the oscillatory behavior of the wavelet, the Q-factor can be set to fit the nature of the signal to which it is applied. In other words the parameter Q can be used to tune the wavelet function to the signal it seeks to model. The signals of interest here are sleep spindles and K-complex events. So to tune the TQWT wavelet toward spindles, we selected a Q-value that corresponds to the minimum number of cycles in a spindle burst. As the latter occur predominantly within 11–16 Hz frequency range with a minimum duration of 0.5 s, we chose $Q = 5.5$. In contrast, to tune the wavelet toward the K-complex component (one cycle) we chose $Q = 1$, as this provided a wavelet that closely models the shape of a transient wave.
- (ii) **Maximum number of levels (J_{\max}):** The selection of J_{\max} depends on the length of the input signal (N) and the chosen filter scaling parameters (α and β) and is defined by the following equation: $J_{\max} = \log(\beta N/8) / \log(1/\alpha)$ (Selesnick, 2011a).
- (iii) **Redundancy parameter (r):** The redundancy parameter r controls the excessive ringing in order to localize the wavelet in time without affecting its shape. Here, it's defined as: $r = \beta/(1-\alpha)$. The specific value $r = 3$ has been previously recommended when processing biomedical signals (Selesnick, 2011a,b).

Morphological Component Analyze (MCA)

The goal of the MCA is to decompose a given signal x into two or more components on the basis of their sparse representation. In our case, MCA is used to decompose a given EEG signal x into an oscillatory component x_1 , and a transient signal x_2 , such that: $x = x_1 + x_2$, where $x, x_1, x_2 \in R^N$. Most importantly, this decomposition is carried out using the TQWT transform (described above) as the sparse representation of x (Selesnick, 2011c). According to the MCA implementation using basis pursuit de-noising with dual Q-factors described in Selesnick (2011b), the sparse wavelets coefficients w_1 and w_2 associated respectively with x_1 and x_2 can be estimated via the minimization of the following function:

$$\underset{w_1, w_2}{\operatorname{argmin}} \|x - \Phi_1^*(w_1) - \Phi_2^*(w_2)\|_2^2 + \sum_{j=1}^{J_1+1} \lambda_{1,j} \|w_{1,j}\|_1 + \sum_{j=1}^{J_2+1} \lambda_{2,j} \|w_{2,j}\|_1 \quad (3)$$

Where Φ_1 and Φ_2 are two matrices of TQWT parameters: (Q_1, r_1, J_1) and (Q_2, r_2, J_2) respectively, w_1 and w_2 are vectors which contain the concatenation of the wavelet transform sub-bands, and $\lambda_{1,j}$ and $\lambda_{2,j}$ are the regularization parameters associated respectively with the two types of wavelets (They are two vectors of lengths $J_1 + 1$ and $J_2 + 1$, respectively).

The sparse set of wavelet coefficients w_1 and w_2 are hence obtained, via the convergence of the objective function given by Equation (3). In the current study, the sparsity (few non-zero coefficients in w_1 and w_2 vectors) was achieved by setting the number of iterations for the convergence to 500. Next, the components x_1 and x_2 are estimated by: $x_1 = \Phi_1^* w_1$ and $x_2 = \Phi_2^* w_2$ (where Φ_1^* and Φ_2^* are the inverse TQWT matrices). Note that all parameters and variables described here were computed strictly as described in the original study by Selesnick (2011a) and user-manual of the TQWT toolbox (Selesnick, 2011b). **Figure 2** shows the results of the TQWT-MCA decomposition applied to an illustrative 30-s EEG segment that contains three spindles and one K-complex. Panels B and C show the decomposition into selected oscillatory and transient components. The next step is to apply a detection procedure to identify the individual spindles and K-complex events from both components. The detection step for each is described in the next sections.

Spindle Detection

The oscillatory component obtained from the EEG decomposition procedure described above is used to detect the occurrence of sleep spindles. Applying a simple threshold directly to this signal would not be appropriate since spindles do not have an established range of amplitudes. Instead, we decided to detect the spindles by filtering the oscillatory component using a CWT.

Continuous Wavelet Transform (CWT)

To optimize the selection of the wavelet function to use in the CWT analysis, we computed the cross-correlation between several wavelet functions (Teolis, 1996) and the spindle waveforms present in the training data set. Based on visual inspection of similarity with the spindle waveform, we chose to test the following wavelet functions: complex frequency B-spline wavelets (Fbsp), complex Morlet wavelets (Cmor), complex Shannon wavelets (Shan), and Gaussian wavelets (Gauss). **Figure 3** shows these individual wavelet functions as well as boxplots for the cross-correlation mean values obtained when using each one of them. Although the results were very close, Fbsp showed the highest maximal value (upper line of each box) and the highest median (red line) cross-correlation with the spindle waveforms. Therefore, we chose to use complex frequency B-spline wavelets which are defined as $bsp(t) = \sqrt{fb} \left[\text{sinc}^m \left(t \cdot \frac{fb}{m} \right) \cdot e^{i2\pi f_c t} \right]$, where m is an integer parameter ($m \geq 1$) that can be selected so as to ensure the best time-frequency resolution, fb is the bandwidth parameter and f_c is the wavelet center frequency. The CWT-based time-frequency maps computed throughout this study are based on this Fbsp wavelet function in the pre-defined frequency band of sleep spindles (i.e., 11–16 Hz).

Detection of Local Maxima and Thresholding

To detect the occurrence of sleep spindles from the time-frequency (T-F) map of the oscillatory component, we first search

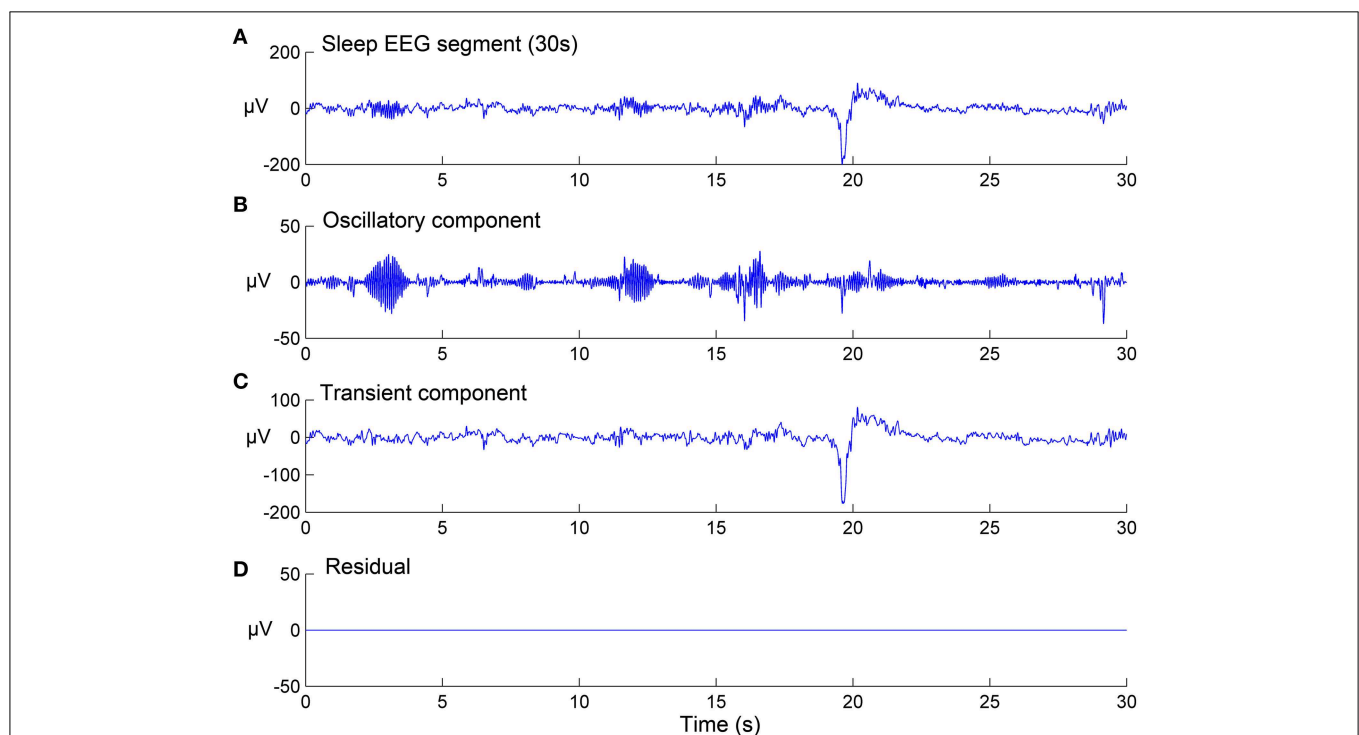


FIGURE 2 | Signal decomposition of a 30-s sleep EEG segment (A) into an oscillatory component (B) and a transient component (C) using TQWT-MCA method, with no residuals (D). See Section EEG Signal Decomposition using TQWT and MCA for method details.

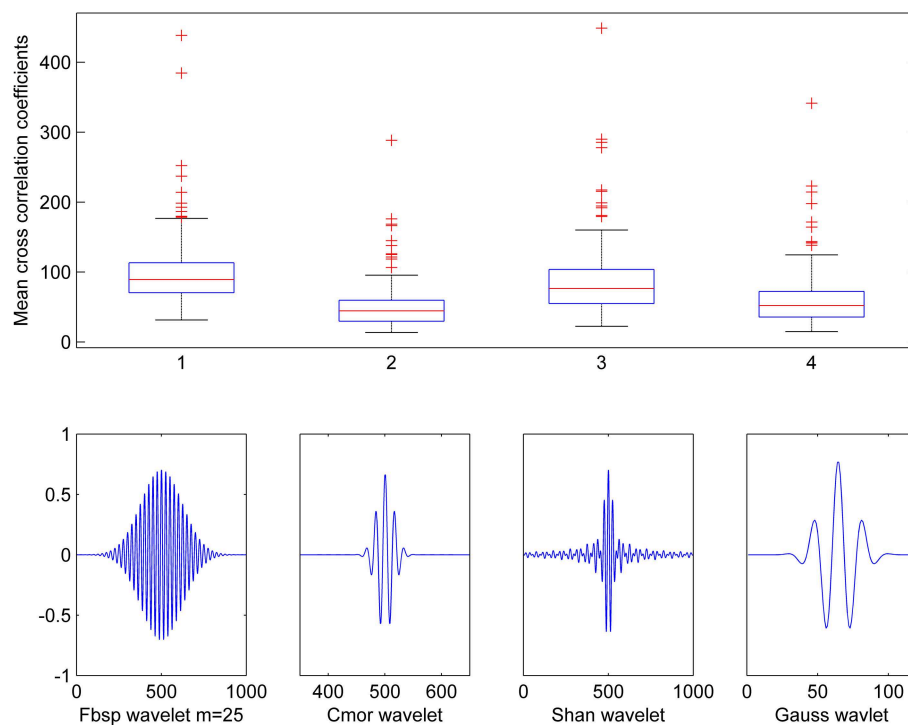


FIGURE 3 | Cross-correlation between various wavelet functions and spindles waveforms. Results show mean cross-correlation between 210 spindle waveforms and five distinct wavelet

functions: Frequency B-spline (Fbsp with $m = 25$), Complex Morlet wavelet (Cmor), Shannon wavelet (Shan), and Gaussian wavelet (Gauss).

for all local maxima by identifying T-F values that exceed those of all eight surrounding neighbors of any given value in the 2D time-frequency space (using a sliding window across the T-F two-dimensional space). Next, we apply a detection threshold to the obtained maxima in the T-F map. Selecting an optimal threshold is a critical step. We chose a procedure that determines the best threshold as the one that maximizes the difference between sensitivity (Sen) and FDR of spindle detection (Note that other options are of course possible and can easily be included in our framework). A practical way to achieve this goal is by using an ROC-like approach on a training data set. The concept is straight-forward: we compute the values of sensitivity and FDR of the detection method repeatedly as we gradually increase the threshold used in the last step. This procedure yields a curve that depicts how sensitivity and FDR co-vary as the detection threshold is changed. The optimal threshold is the one that maximizes the difference between sensitivity (ideally as high as possible) and FDR (ideally as low as possible). Note that the computation of FDR and sensitivity (see Section Performances Metrics) requires the use of some form of ground truth. Here we used expert visual marking as the benchmark. As our data was visually annotated (for K-complexes and spindles) by two experts, unless otherwise stated, we report all our results using, as ground truth, the annotation of each separately.

In summary, the optimal threshold derived from the “sensitivity vs. FDR” analysis is used when running the detection pipeline on the test data set. In order to evaluate the performance

of the method, we compute once again sensitivity and FDR, but now only on the results obtained with the test set. The interested reader can find more details on such training procedures for instance in the appendix of Chander (2007).

K-complex Detection

Unlike sleep spindles, K-complex waveform is distinguishable from EEG background activity by “a well delineated negative sharp wave.” Therefore, our rationale was that applying a negative amplitude threshold on the transient components (derived from the TQWT and MCA procedure) could be a promising way to detect such events. However, in order to reduce the effect of some high frequency waveforms which generate local minima with amplitudes close to those of the K-complex (Devuyst et al., 2010), we first apply a band-pass FIR filter [0.5–5 Hz] to the transient component produced by TQWT and MCA step. Next, K-complexes are detected from the list of all local minima in each segment using an optimal threshold value. Note that we constrained the interval between two successive detected minima to be at least 2 s long to reduce risks of false detections. An EEG structure composed of multiple successive local amplitude peaks (such as delta waves) could in theory lead to the detection of a succession of transients and thus lead to the identification of successive K-complexes. This is only acceptable if the successive events are separated by at least 2 s, as that is approximately the minimal interval expected between two real K-complexes. The method used to derive the best threshold value to use here for

K-complex identification is identical to the method described for threshold selection in the case of spindle detection: We use an ROC-like training procedure just as described in Section Detection of Local Maxima and Thresholding.

Performances Metrics

To compute the ROC-like curves used to derive detection thresholds (from the training set), and to evaluate the performance of our method (on the test set) we compute two basic metrics: the sensitivity (*Sen*) and *FDR* defined by Equations (4) and (5) respectively:

$$Sen = \frac{TP}{TP + FN} \quad (4)$$

$$FDR = \frac{FP}{FP + TP} \quad (5)$$

Where TP (true positive detections) are the events marked by the expert and correctly detected by our method, FN (false negative detections) are the events marked by the expert but not detected by the method and FP (false positive detections) represents the number of events detected by the method but which were not marked by the expert. Note that in detection contexts with strongly unbalanced occurrences of positive and negative cases, the ROC curve can provide an inadequate representation of the performance of a classifier (O'Reilly and Nielsen, 2013). This is the case here for the sleep EEG events we set out to detect because the continuous EEG segments consist predominantly of true negatives. This is why, instead of using standard ROC analysis, i.e., plotting sensitivity vs. false positive rate (or 1-specificity), we chose to depict sensitivity vs. *FDR*.

Expert Identification and Inter-annotator Agreement Metrics

Two annotators visually identified all K-complexes and spindle events in our database. Unless otherwise stated all automatic detection results are evaluated against the annotation of Expert 1 and 2, independently. When evaluating the minimal number of training segments needed for our method (Section Impact of the Amount of Available Training Data on the Performance) and when exploring the results on a subject by subject basis (Section Performance of the Method in Individual Subjects) we restricted the analysis to the segments where Expert 1 and Expert 2 fully agreed (consensus). Inter-annotator agreement was assessed using two metrics: (i) percent agreement (portion of events on which raters compared to total number of events) and (ii) Cohen's kappa coefficient κ , a statistical measure of inter-annotator agreement that takes into account the agreement occurring by chance (Cohen, 1960).

Results

The results of the proposed methodology are presented in the next sections as follows: First, we provide the results of the training step (ROC-based identification of optimal thresholds), followed by the performance of the method on test sleep data (S2 and non-S2). Next, we report also on the improvements achieved

by using the optional adjustment step where the expert reviews (accepts/rejects) the false positive outputs of the method. We then explore the practical utility of the method by monitoring its performance as a function of training set size. Unless otherwise stated, we report all our results using, as ground truth, the annotation of each one of the two experts separately. This provides further insights into the robustness of the method.

Detection of Optimal Threshold Values (Training Phase)

In the training phase, we used a subset of the data (training set) to derive "sensitivity vs. *FDR*" curves by evaluating sensitivity and false detection rates as we vary the detection threshold. Sensitivity and *FDR* were computed using 210 30-s EEG S2 data segments for threshold values that varied in steps of $10 \mu V^2$ for spindles and $2 \mu V$ for K-complexes (the unit reflects the fact that the thresholds are applied to time-frequency maps and voltages, respectively). The optimal threshold value, defined as the one that maximizes the difference between sensitivity and *FDR*, was determined from these curves and then used subsequently in the validation phase (i.e., using the test set). For spindle detection, this compromise in the training data was achieved by a threshold set to $290 \mu V^2$, yielding a sensitivity of 87.09% and an *FDR* of 45.68%. In the case of K-complex detection, a threshold value of $-70 \mu V$ provided the best compromise, with a sensitivity of 78.72% and an *FDR* of 23.44%. The above results were obtained when using Expert 1 as benchmark. The results were very similar when relying on the annotation by Expert 2 as benchmark: For spindle detection, this compromise in the training data was achieved by a threshold set to $300 \mu V^2$, yielding a sensitivity of 83.45% and an *FDR* of 27.68%. In the case of K-complex detection, a threshold value of $-70 \mu V$ provided the best compromise, with a sensitivity of 85.76% and an *FDR* of 32.22%. **Figure 4** shows an example that illustrates the results of the training step and how the optimal threshold levels are determined. The identified thresholds are then used when applying the detection pipeline to the test segments (see next section). Throughout the paper, the training strategy was applied using visual scoring either by Expert 1, Expert 2 or by only using the data segments for which both experts fully agreed (consensus). Unless otherwise stated, we report the results of each analysis by providing the results obtained against Expert 1 and Expert 2 independently.

Spindle and K-complex Detection Performance (Test Set)

To evaluate the performance of the pipeline and, in particular, assess the success of the threshold identification procedure, the spindle and K-complex specific thresholds identified in the training phase were then used to run the detection algorithm on previously unseen test segments. **Figure 5** illustrates the detection procedure on the same sample sleep segment shown presented in **Figure 2**. The global results obtained for all 210 test EEG S2-sleep segments are shown in **Table 2**. The full analysis (training and testing) was repeated twice, each time using a different scorer as ground truth to explore the robustness of the procedure. The results indicate that the method proposed here yields a

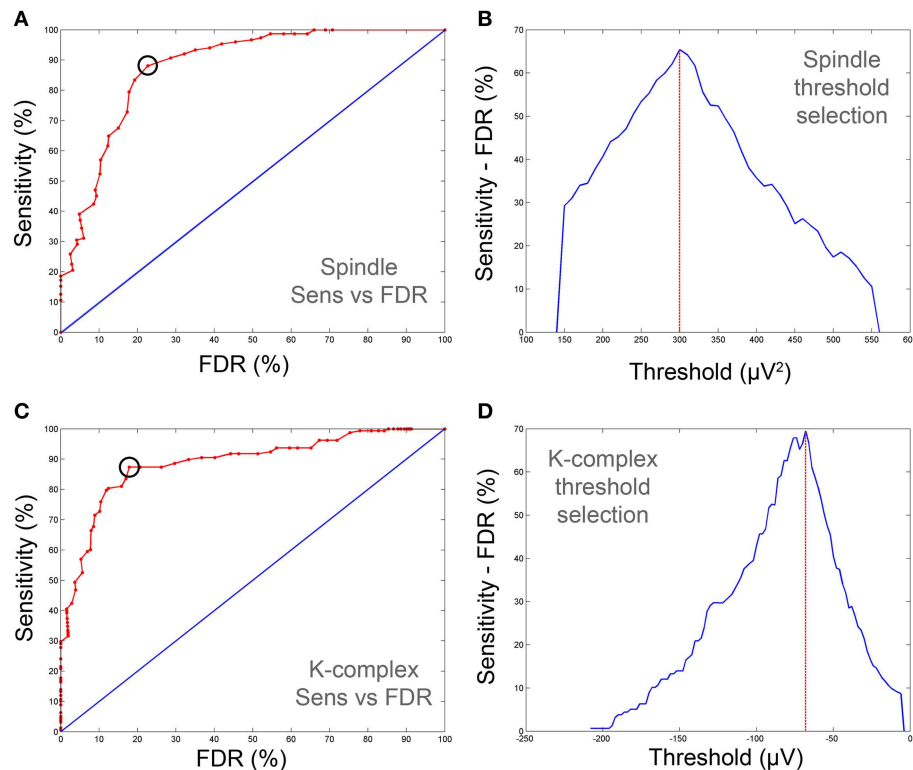


FIGURE 4 | Sensitivity-FDR plots used for the determination of optimal spindle and K-complex detection thresholds (on training data). (A,C) show ROC-like-curves of sensitivity vs. FDR for spindles and K-complex respectively. (B,D) depict the difference between sensitivity and FDR as the threshold is varied. The optimal thresholds were defined as

those corresponding to maximum difference (vertical red line). The corresponding cut-off point is also depicted on (A,C) with black circles. Note that the sensitivity and FDR in the illustrative examples presented here are computed against consensus scoring (i.e., agreement between both scorers).

reasonably high sensitivity both for spindles (scorer 1: 83.18%, scorer 2: 81.57%) and K-complex (scorer 1: 81.57%, scorer 2: 85.25%). The FDR values for spindles reached 39% (scorer 1) and 19.66% (scorer 2), while the FDR for K-complex detection was 29.54% and 32.82% for scorers 1 and 2, respectively. Note that the inter-rater overall agreement was 77.85% (Cohen's kappa 0.64) and 63.33% (Cohen's kappa 0.51) for spindle and K-complex identification respectively. **Table 2** also shows the method performance when applied exclusively to data segments for which both scorers agreed (100% inter-rater agreement, i.e., consensus scoring). In the case of spindle identification, this led to a sensitivity of 86.40% and an FDR of 29.22%.

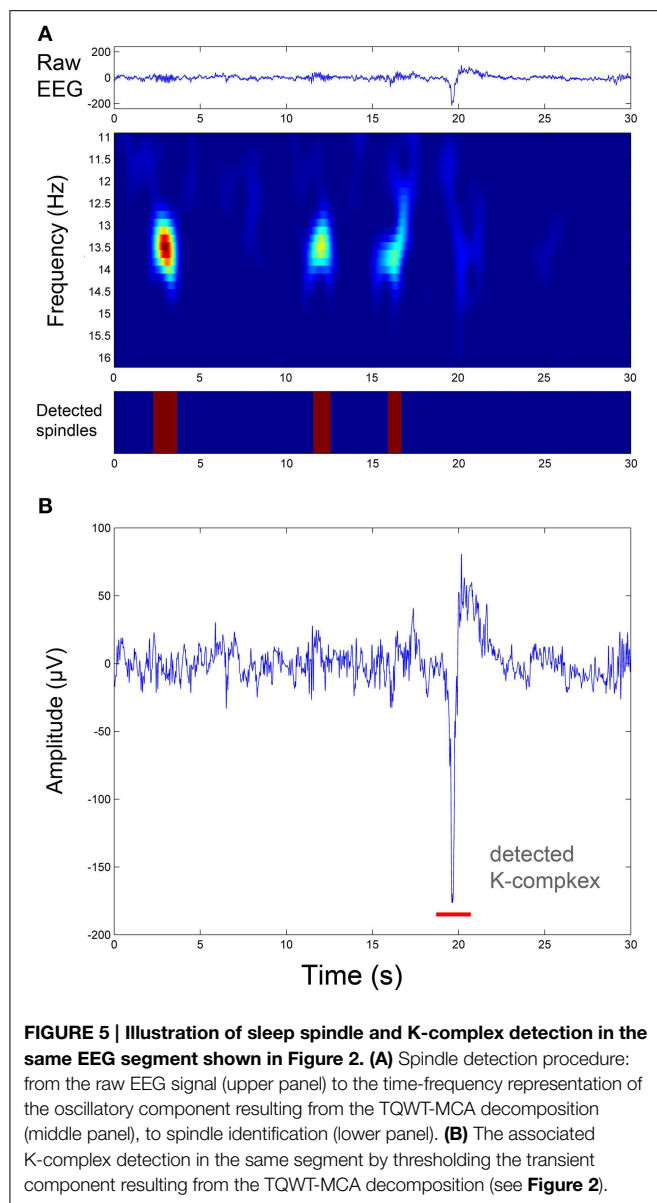
Performance Comparison with and without TQWT and MCA

How critical is the inclusion of the TQWT-MCA decomposition framework proposed here for the performance of the detection? To address this question we set out to evaluate the added-value of TQWT and MCA decomposition in the detection process. To this end, the entire pipeline was performed again on the same data set as above but this time with one notable difference: the TQWT and MCA steps were excluded from the method. In other words, instead of using oscillatory and transient components

(i.e., the output of TQWT-MCA), the detection process started directly from raw EEG signals for K-complex identification, and directly from its CWT transform for spindle detection. **Figure 6** compares the results obtained with and without the TQWT-MCA step. When using Expert 1 as ground truth, excluding the proposed decomposition led to a drop in sensitivity for spindle detection (from 83.18 down to 70%) and for K-complex detection (from 81.57 down to 76.97%). Deterioration was also observed in terms of increased false detections. The FDR values increased from 39 to 43.62% in spindles detection and rose from 29.54 to 49.09% for K-complex detection. The corresponding results obtained with Expert 2 as ground truth are comparable and are given in panels C and D of **Figure 6**. These findings quantify the specific added-value of the TQWT-MCA decomposition as a pre-processing step, as compared to direct detection on the raw EEG signal. In the discussion section, we further confirm these observations by comparing our method to another peak detection method previously published in the literature.

Scoring Adjustment Based on Expert Review of False Positives

We evaluated the potential performance enhancement that would be achieved by an additional (optional) step in which



the false positive detections of our algorithm were presented to the expert scorer for review. This allowed the scorer to decide to accept or reject events detected by the algorithm but that he had initially not marked. A dedicated GUI was developed for this score adjustment (SA) procedure. After this process was carried out a new file with the adjusted score was created and the whole detection pipeline was repeated (i.e., including the training and validation processes). The performance enhancement obtained with the SA procedure is shown in **Figure 6**. As expected, sensitivity increased and FDR decreased, for K-complexes and spindles. The most prominent improvements were a drop in spindle FDR from 39 to 21.33% and an increase in K-complex sensitivity from 81.57 to 87.27%, when using the annotations of Expert 1 (**Figures 6A,B**). Similar results were obtained when comparing against annotations by Expert 2 (**Figures 6C,D**). Note that this semi-automatic step

TABLE 2 | Method performance (Sensitivity and FDR) obtained by applying the pipeline to the validation data set (test segments) for spindles and K-complexes detection.

	Sensitivity (%)	FDR (%)
SCORER 1		
Spindle	83.18	39.00
K-complex	81.57	29.54
SCORER 2		
Spindle	83.10	19.66
K-complex	85.25	32.82
SCORER 1 and 2 (AGREEMENT)		
Spindle	86.40	29.22
K-complex	80.86	21.39

Results are shown for scorer 1, scorer 2 and also for the case where only data with full agreement between the two scorers were used.

is not considered part of the proposed methodology, as it requires visual marking of the whole data set. Nevertheless, this analysis quantifies the impact of the subjective scoring, and provides an estimate of the performance that the method could provide if the scorer provides a more consistent visual marking.

Stability of the Proposed Method with Regards to Sleep Stages

The results presented above were obtained with EEG segments that were recorded during S2, the sleep stage where K-complex and spindles are most frequent. However, as indicated above, our method does not require sleep staging as a preliminary pre-processing step. The method is in theory equally valid for EEG segments from all sleep stages. We therefore also examined the performance of the detection algorithm by using 420 EEG segments including data from all sleep stages. Half of the segments were S2 (i.e., 210 segments) and the other half were non-S2 sleep (i.e., 210 segments). The 210 non-S2 segments were composed of: 126 REM segments, 42 SWS segments and 42 S1 segments. Note that these proportions were chosen to be close to the natural distribution (frequency of occurrence) of the various sleep stages across a typical night's sleep (Carskadon and Dement, 2011). The motivation behind this selection was to create training and test sets with compositions as close as possible to what one would get from a random sampling of sleep EEG epochs, i.e., without access to sleep stage information. Using equal number of events across sleep stages (or running our analysis separately for each sleep stage) was not feasible with the data at hand given that some of the sleep stages, in particular S1 and REM, contain a very low number of spindles and K-complexes.

Globally speaking, the results of this analysis (see **Table 3**) show a slight increase in sensitivity but comes at the expense of an increase in FDR. This is most likely due to the fact that the thresholds are better tuned to the more numerous S2 events. Note that also in this analysis we see a reasonable agreement between the results obtained when using each of the two scorers as ground-truth.

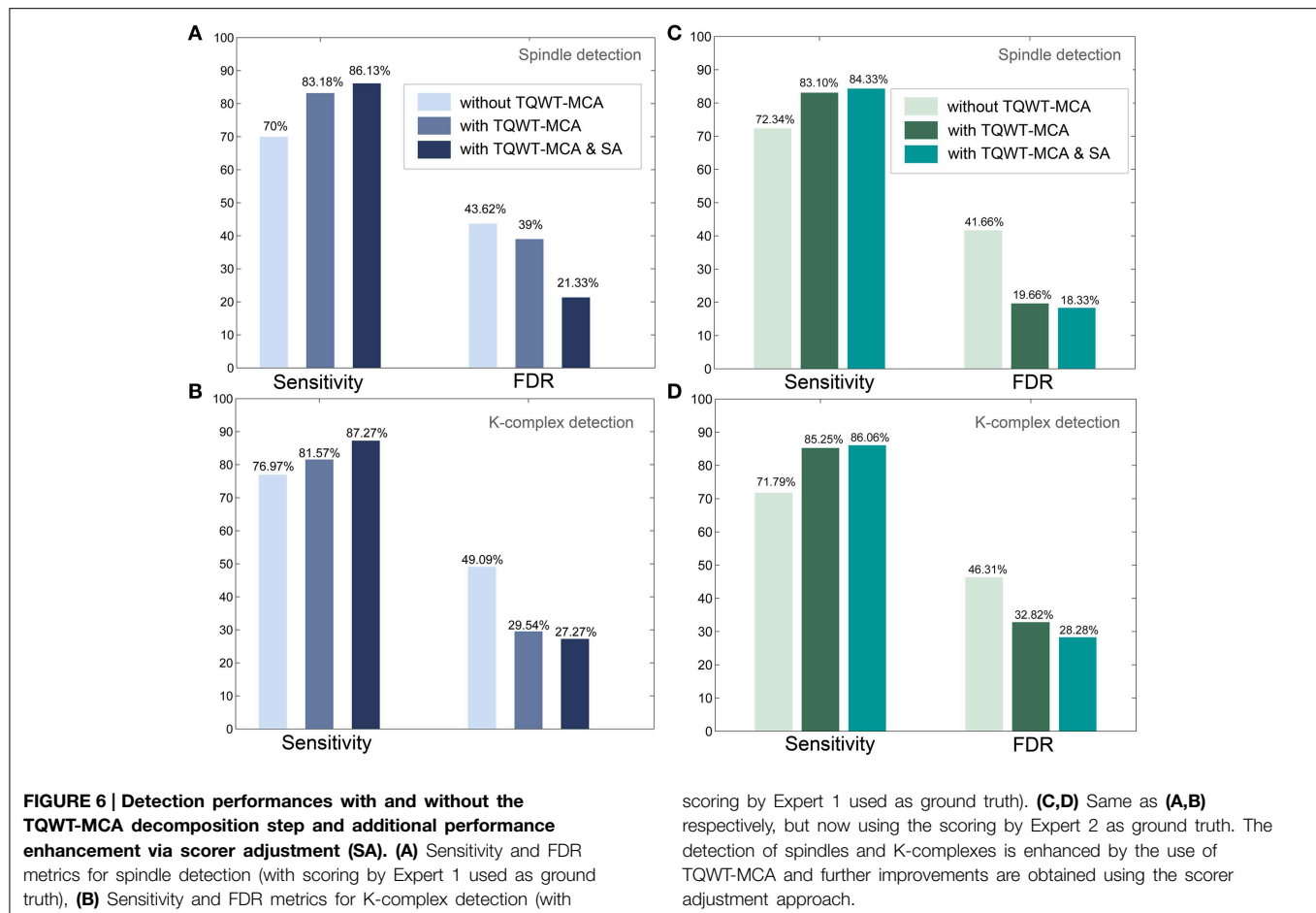


TABLE 3 | Method performance (Sensitivity and FDR) obtained by applying the pipeline to a validation data set (test segments) for spindle and K-complex detection that includes data from all sleep stages (S2 and non-S2 segments).

	Sensitivity (%)	FDR (%)
EXPERT 1		
Spindle	86.82	45.36
K-complex	80.23	37.27
EXPERT 2		
Spindle	85.05	32.19
K-complex	82.5	38.00

Results are shown for both scorers. The performances obtained if we restrict the detection to S2 segments are presented in **Table 2**.

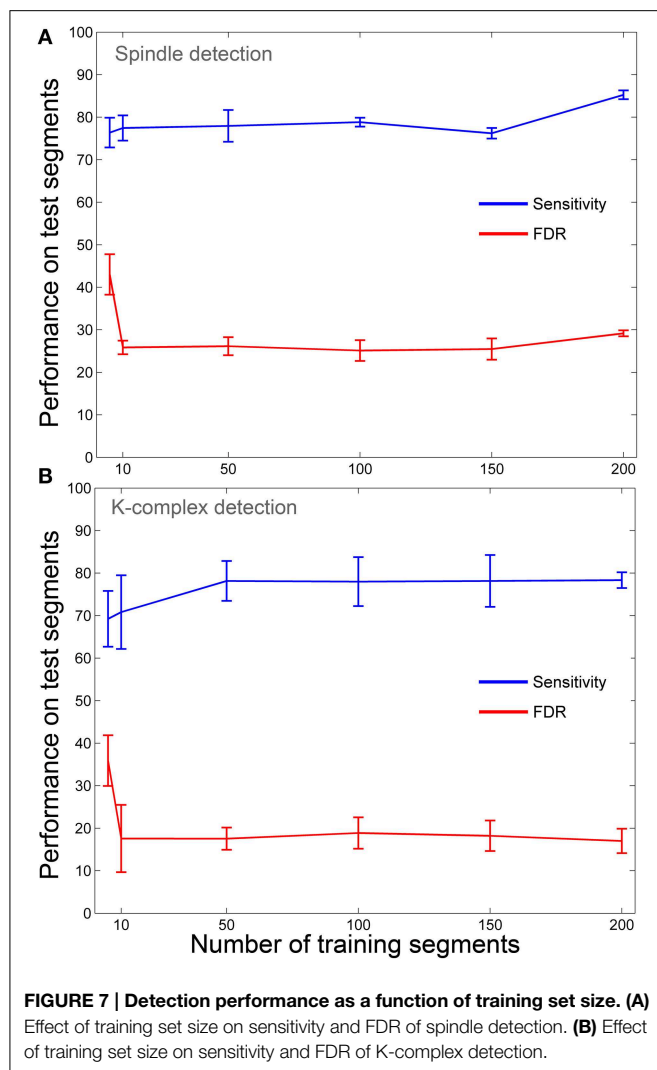
Impact of the Amount of Available Training Data on the Performance

The method proposed here is by definition a semi-automatic procedure since it has a built-in training step that uses visual marking of a subset of data to determine an optimal threshold that is to be used on the rest of the data. An important question is therefore: what is the minimal amount of visual scoring required by our method in order to achieve acceptable detection results? Obviously the method would be of little use, if half (or more)

of the K-complexes and spindles in the data need to be marked by an expert to ensure that it works. To address this question we launched the entire pipeline (training and testing) repeatedly, each time using an increasing number of training segments (starting from five segments up to 200 segments, the procedure was repeated five times at each size with random selection of segments). The aim was to see how quickly the sensitivity and FDR metrics stabilize. Here, we restricted the analysis to all segments where the annotations of both experts were in complete agreement (consensus). This was done to ensure robustness of the annotation and because of the lengthy computational cost associated with recalculating the whole analysis for annotations from each expert. The aim here was not to assess the effect of inter-expert variability, but rather to assess the dependency of our technique on the number of training samples. The results in **Figure 7** show that, luckily, the performance metrics reach a plateau already with a small number of training segments (below 20 segments for spindles and below 50 segments for K-complexes). This result indicates that the proposed method can be used with minimal visual marking.

Performance of the Method in Individual Subjects

The results presented so far were obtained by combining EEG sleep segments extracted from multiple subjects ($n = 14$).



But how robust is the proposed method for the detection of K-complexes and spindles in each individual subject? And, in particular, how good are the performances in single subjects when only a handful of events have been visually marked and thus available for training? To address this question we launched the entire detection pipeline in each subject individually using only 15 segments for training (On average these 15 30-s segments contained 18 ± 4.3 spindles and 11.9 ± 2.3 K-complex events). In addition, as in the previous analysis (Section Impact of the Amount of Available Training Data on the Performance) we restricted the analysis to all segments where the annotations of both experts were in agreement (consensus). The results listed in **Table 4** indicate a reasonably good performance in each individual. The means of the individual performances (achieved from only 18 spindles and 11.9 K-complexes on average) are in fact comparable to those achieved (see **Table 2**) when combining the data from all subjects and using half the data for training (210 segments, consisting of 141 K-complex and 217 spindle events). As a matter of fact, the mean sensitivity for spindle detection (i.e., 84.39%) which was obtained with very low number of training

TABLE 4 | Performance of the TQWT-MCA spindle and K-complex detection method in each subject with minimal training.

	Spindle			K-complex		
	Sens %	FDR %	Th (μV^2)	Sens (%)	FDR (%)	Th (μV)
S1	61.54	20.00	300	78.57	21.42	-68
S2	85.00	50.00	210	88.89	27.27	-72
S3	100.00	27.27	240	81.81	10.00	-80
S4	62.50	16.67	300	75.00	10.00	-82
S5	73.91	19.05	270	57.14	11.11	-94
S6	81.43	23.25	280	83.33	23.08	-74
S7	96.30	35.00	240	60.00	14.28	-78
S8	90.62	9.37	270	80.76	20.00	-94
S9	90.00	47.06	230	88.89	11.11	-66
S10	95.65	21.43	260	93.33	12.50	-62
S11	100.00	50.00	210	100.00	26.67	-70
S12	100.00	23.68	270	33.33	88.88	-78
S13	76.19	11.11	300	83.33	45.22	-98
S14	68.42	31.58	340	80.00	42.86	-54
Mean	84.39	27.53	265.7	77.45	26.04	-76.4

Sensitivity Sen (%), FDR (%), and the optimal threshold (Th) are reported for each individual but also as mean values across the whole population (bottom row). Only 15 annotated 30-s EEG segments were used for training in each subject (corresponding on average to 18 ± 4.3 spindles and 11.9 ± 2.3 K-complex training events).

samples is slightly higher than the value achieved with half of the whole data set when combining data across individuals. The results in **Table 4** confirm that individually determined thresholds provide good results and, because they were achieved with only 15 training segments, it also suggests that the proposed method does not require a lot of visual marking. Note however, that for practical reasons and for the sake of generalizability we recommend the use of a global detection threshold, just as we did in all previous sections.

Comparison with a Standard Detection Method

To gain insights into how our method compares to existing methods, we implemented a standard spindle detection method (Gais et al., 2002; Mölle et al., 2002) which has already been implemented or used as a standard method for comparison, in numerous publications (e.g., Gais et al., 2002; Mölle et al., 2002; Bergmann et al., 2012; Feld et al., 2013; Parekh et al., 2014, 2015; Warby et al., 2014). In brief, the procedure consists of the following steps: (1) filtering the EEG with a 12–15 Hz bandpass filter, (2) calculating the root mean square (RMS) of each 100 ms interval of the filtered signal, (3) counting the number of times the RMS power crossed a constant detection threshold T value for 0.5–3 s. In the original study, Mölle et al. (2002) set the threshold T to $10 \mu V$. To choose the best value for this parameter with regards to our data, we computed the performances we achieved using all T values between 5 and $12 \mu V$ (in 1 Hz steps) on the training test. The threshold that provided the best compromise between sensitivity and FDR was the one used when applying our method to the test data. **Table 5** compares the results obtained with this standard method (Möller et al., 2002) to those obtained

TABLE 5 | Comparison between the results (Sensitivity and FDR) achieved with our method to those obtained by applying a standard spindle detection technique (Möller et al., 2002), and to those achieved by a hybrid approach where we use the proposed TQWT + MCA analysis as a pre-processing step before running the standard RMS-based detection procedure.

	Filtering + RMS (Möller et al., 2002)				Current study	
	Standard		With TQWT		Expert 1	Expert2
	Expert 1	Expert 2	Expert 1	Expert 2		
Sensitivity	70.30	70.56	74.06	75.88	83.18	83.10
FDR	49.45	46.21	42.22	37.24	39	19.66

The performances of these three approaches are reported against Expert 1 and Expert 2 independently. The best results were achieved with the method proposed in this study.

with our method, but also to a hybrid approach where we use our TQWT + MCA analysis as a pre-processing before running the standard RMS-procedure proposed in Möller et al. (2002). The results in **Table 5** suggest that our method outperforms the RMS-based method on the same data set. In addition, we found that the performance of the RMS-based method (Möller et al., 2002) can be substantially improved if we first apply our TQWT-MCA processing to the data. Note that the thresholds T that yielded the best results with our data were 6 and 8 μV for the detection with and without TQWT-MCA, respectively.

Performance Evaluation on a Publicly Available Database

To investigate the performance of our method on sleep EEG data other than our own recordings, we detected spindles and K-complexes by applying our method to the DREAMS data set, a publicly available database of annotated sleep EEG. EEG recordings from two specific databases were used: The Sleep Spindle database and the K-complexes database, which have both been made available by University of MONS - TCTS Laboratory and Université Libre de Bruxelles—CHU de Charleroi Sleep Laboratory. The spindles data can be accessed online at: <http://www.tcts.fpms.ac.be/~devuyst/Databases/DatabaseSpindles/> while the K-complex data can be found at: <http://www.tcts.fpms.ac.be/~devuyst/Databases/DatabaseKcomplexes/>. The spindles and K-complexes databases consist respectively of 8 and 10 excerpts of 30 min of annotated central EEG channel extracted from whole-night PSG recordings. Here, we used recordings from the subjects that were recorded with identical sampling rate (200 Hz) and for which the visual annotation was complete. This meant that for spindle detection we used 6 participants out of 8 and for the K-complex detection we used the data from all 10 participants. We used the annotation by Expert 1 as benchmark since the annotations of Expert 2 are not available for all subjects. The straight-forward application of our method to these data, without any specific parameter adaptations, yielded a sensitivity of 71.77% and FDR of 30.54% for spindle detection, and a sensitivity of 83.31% and FDR of 36.31% for K-complex detection.

Discussion

The current study proposes a new method for joint detection of sleep spindles and K-complex events, two hallmarks of

NREM sleep stage 2, by conveniently splitting the EEG signal into oscillatory (spindles) and transient (K-complex) components. The decomposition is achieved by applying MCA on a sparse representation of EEG segments obtained by the recently introduced discrete TQWT (Selesnick, 2011a,b,c) with parameters specifically *tuned* to spindle and K-complex characteristics. The actual detection step relies on thresholding (a) the transient component in the search for K-complexes and (b) the time-frequency representation of the oscillatory component in search for sleep spindles. Optimal thresholds are extracted from ROC-like curves (sensitivity vs. FDR) in a training set, and the performance of the method is assessed on the test set.

Overall the method presented here provides a reasonable compromise between sensitivity and FDR with performances that were robust on several levels: First, the performances did not change much when the benchmarking ground-truth was switched from one scorer to another [Section Spindle and K-complex Detection Performance (Test Set)]. Second, the performance hardly changed whether only stage2 sleep EEG segments were used or if data from all sleep stages were examined (Section Stability of the Proposed Method with Regards to Sleep Stages). Third, and most importantly, our results show that the method does not require a large training set to derive optimal cut-off thresholds. By varying the number of segments used for training, we found that the performance in terms of sensitivity and FDR reaches a plateau within less than 20 training segments (Section Impact of the Amount of Available Training Data on the Performance, **Figure 7**). Finally, the latter observation was further confirmed by running the detection pipeline on individual subjects where the training (search for optimal threshold) was restricted to 15 segments (i.e., using on average 18 spindles and 12 K-complexes). This analysis revealed good sensitivity and relatively low FDR in each subject and also in terms of means over all individuals (Section Performance of the Method in Individual Subjects, **Table 4**).

The TQWT-MCA approach has been recently used to dissociate transient events with or without high frequency oscillations (HFOs) in intracranial EEG (Chaibi et al., 2014). The current study, is to our knowledge, the first to demonstrate the utility of the TQWT-MCA framework for the detection of sleep spindles and K-complexes.

Furthermore, the results we obtained by excluding the TQWT-MCA decomposition from the proposed framework, confirmed and quantified its contribution to the high

performances obtained (Section Performance Comparison with and without TQWT and MCA). Compared to the results obtained without the TQWT-MCA step, our method achieved an additional 13 point increase in percent sensitivity for spindles and a five point increase for K-complexes (**Figure 6**). Since the proposed decomposition is based on sparse representation of spindles and K-complexes, it reduces the effect of noise and artifacts in EEG signals, which may explain, at least in part, the improved performance of the subsequent CWT and FIR filtering.

In addition, we have shown that a simple visual marking adjustment step can lead to significant improvements, in particular by reducing FDR. In the scorer adjustment procedure the expert is presented with the false positive detections and is given the possibility to accept or reject detections that he had initially not indicated but that the algorithm identified as being positives. This SA procedure is not part of the recommended algorithm, rather a way to identify and quantify cases where the objective machine might actually outperform the subjective human scorer.

Parekh et al. (2014) propose a strategy to improve spindle detection by pre-processing the raw EEG signal using non-linear dual Basis Pursuit Denoising (BPD) which is also a way to separate the non-oscillatory transient components of the signal from the sustained rhythmic oscillations. The subsequent filtering of the oscillatory component enhances the spindles with regards to baseline, and thereby improves their detectability with standard spindle detectors. Using this technique with a readily available EEG spindle database provided a mean increase of 13.3% in the by-sample F1 score and 13.9% in the by-sample Matthews Correlation Coefficient score. A recent study by the same group also provides compelling evidence for the added value of using sparse optimization to detect spindles and K-complexes (Parekh et al., 2015). A direct comparison between these approaches and the methodology proposed here is not straightforward given the use of by-sample metrics in the Parekh et al. (2014, 2015) studies. Most importantly, the current method and those proposed by Parekh et al. (2014, 2015) provide converging evidence of improved spindle detection via time-frequency sparsity, and they collectively suggest that this framework is a promising path for enhanced performance of event detection in sleep EEG.

Overall, the results reported here (either by combining data across participants or by performing the detection algorithm separately for each individual) are comparable with the results of existing methods. However, we performed further analyses in order to gain additional insights into (a) how the performance of the pipeline proposed here compares to existing methodology (Section Comparison with a Standard Detection Method) and (b) how well it performs on other available data sets (Section Performance Evaluation on a Publicly Available Database). The results suggest that our method provides better detection than the RMS-based method and that the performance of the latter can be improved if we first apply the TQWT-MCA processing to the data before computing the RMS (**Table 5**). Furthermore, application of our method to the Devuyst et al. (2010, 2011) online database, yielded a sensitivity of 71.77% and FDR of 30.54% for spindle detection, and a sensitivity of 83.31% and

FDR of 36.31% for K-complex detection. The original papers associated with these databases do not directly report sensitivity and FDR, but these metrics can be inferred from the confusion matrices they provided for each expert. Using Expert 1 as ground truth (as we did here), they detected spindles with sensitivity of 68.40% and FDR of 62.04% (computed from confusion matrix in Devuyst et al., 2011). As for K-complexes, they were detected with sensitivity of 61% and FDR of 26.70% (computed from confusion matrix in Devuyst et al., 2010). Note, however, that the comparison between their findings and ours is limited by the fact that the recordings provided online does not allow us to explore the exact data sets used in Devuyst et al. (2010, 2011).

More generally, the comparison between existing methods for spindle and/or K-complex identification is not an easy endeavor. First of all, the different methods proposed are generally evaluated on different EEG data sets and with different scorers, often with substantial inter-rater variability (Wendt et al., 2015). Moreover, performance metrics also tend to differ across studies. Recent efforts seek to overcome such limitations by providing free access to high quality annotated sleep EEG data sets (O'Reilly et al., 2014). Such benchmark data carry the potential to significantly advance the field of automatic spindle and K-complex detection, as well as sleep staging. This was performed in a recent report by O'Reilly and Nielsen (2015) where the authors compared four automatic spindle detection algorithms: Teager detector (Ahmed et al., 2009), Sigma index (Huupponen et al., 2007), RSP (Devuyst et al., 2011), RMS (Möller et al., 2002). To this end, four data bases were used, two of which are open access: the DREAMS database (Devuyst et al., 2010, 2011) and the Montreal Archive of Sleep Studies (MASS) (O'Reilly et al., 2014). The results obtained and conclusions drawn from this important comparison highlight limitations and shortcomings of classical detection performance evaluations frameworks. In particular, the reported findings question the reliability of using expert scoring as gold standard. In addition, they highlight the necessity of using an exhaustive set of performance metrics: The authors recommend the use of sensitivity, precision and a more comprehensive statistic such as Matthew's correlation coefficient, F1-score, or Cohen's κ for adequate sleep spindle assessment. Comparison of our results with those presented in this comparative study is not straightforward because we use window-based performance metrics whereas the study by O'Reilly and Nielsen (2015) use a signal-sample metric, equivalent to the "by-sample" metric (Warby et al., 2014). This discrepancy is in itself problematic. Future studies should seek to evaluate detection performance using a unified set of evaluation metrics computed on large open-access benchmarking data bases. Such an assessment of the method proposed here would certainly help evaluate its strengths and limitations.

The current study is one of a few reports that have proposed a common methodological framework for the joint detection of K-complex and spindles (Jobert et al., 1992; Koley and Dey, 2012; Jaleel et al., 2013; Camilleri et al., 2014; Parekh et al., 2015). While Jobert et al. (1992) used matched filtering to detect sleep spindles and K-complex waveforms, Camilleri et al. (2014) used switching multiple models. The authors of the latter study

evaluated their method by computing sensitivity and specificity based on two expert manual scores and reported a sensitivity of 83.49 and 52.02% and a specificity of 78.89 and 90.55% for respectively spindles and K-complex detection. In addition, Koley and Dey (2012) used CWTs to detect a set of sleep EEG characteristic waveform, including spindles and K-complex. They reported a good accuracy of 92.6 and 93.9% but didn't mention any performance metrics that take false positive or false negative detection into account. Jaleel et al. (2013) proposed a pilot detection method based on a mimicking algorithm which imitates human visual scoring. However, no systematic evaluation of performance metrics was provided. The method proposed by Parekh et al. (2015) provides an elegant approach based on the decomposition of the EEG signals into three signal components (low-frequency, transient and non-oscillatory) and their results highlight the utility of sparse optimization in the improved detection of spindles and K-complexes.

Because of the naturally low number of K-complexes or spindles across some of the stages (S1 and REM in particular) it was impossible for us here to conduct our detection pipeline on each sleep stage individually. Instead, we evaluated the performance of our method by using either only S2 segments, or by pooling segments from all stages (S2 and non-S2 segments). Future studies with larger annotated sleep EEG databases will be needed to assess and compare the robustness of our method in each single sleep stage.

One way to increase the performance of our method could be to fine-tune parameters of the TQWT and of the MCA procedures on a subject by subject basis, so as to account for inter-individual differences in spindle and K-complex properties. To what extent the performance can be improved by modifying the tuning Q-factor (globally or for each individual) is not clear and could be the focus of further investigation. Future explorations may also benefit from exploring the use of alternative wavelets, such as the Morse wavelet (Lilly and Olhede, 2012) which has successfully been used in recent studies (Zerouali et al., 2013, 2014; O'Reilly et al., 2015).

Moreover, it is possible that the false positive detections in our pipeline include vertex waves mistakenly identified as K-complexes since the two events bare strong resemblances. Careful selection of the FIR filter parameters may help reduce this risk since vertex waves are shorter-lived events (<0.5 s).

A further path for performance improvement is to seek to identify spindles and K-complexes in multi-electrode data. The co-occurrence (and even delays) of the presence of these micro-structure across parietal, temporal and frontal brain areas would be very informative, and could even be used to increase detection performance. In addition, exploring the results obtained with the proposed method across all scalp-EEG channels could be helpful in assessing the distribution and propagation of K-complexes and spindles (O'Reilly and Nielsen, 2014a,b) and unraveling their underlying network dynamics (Zerouali et al., 2014). Note also

that the Q-factor of the TQWT can easily be tuned to incorporate differences in frequencies between, for instance, faster central spindles and slightly slower frontal spindles (e.g., Andrillon et al., 2011).

Another venue for future research would also be to attempt to incorporate into our framework recent findings of cross-frequency relationships among various electrophysiological signatures of sleep. In particular, high-frequency activity in the gamma-range, which has been shown to be involved in a variety of cognitive processes (e.g., Jerbi et al., 2009a,b; Jung et al., 2010; Dalal et al., 2011; Lachaux et al., 2012; Perrone-Bertolotti et al., 2012; Vidal et al., 2014), has also been shown to co-fluctuate with slower brain rhythms (Jensen and Colgin, 2007; Canolty and Knight, 2010; Soto and Jerbi, 2012). During sleep, gamma oscillations have been linked to spindles (e.g., Ayoub et al., 2012) and to slow wave sleep in intracranial EEG recordings (Dalal et al., 2010; Le Van Quyen et al., 2010; Valderrama et al., 2012) and in non-invasive EEG recordings (Piantoni et al., 2013). Whether including these cross-frequency relationships will enhance current detection tools remains to be seen.

Conclusion

The current study demonstrates the feasibility of identifying spindles and K-complex events in sleep EEG using a single methodological framework by literally *tuning into* the oscillatory characteristics of the target events via the TQWT. Because of the now well acknowledged challenges that face performance evaluation of automatic and semi-automatic procedures (O'Reilly et al., 2014), the next step would be to validate our method on a larger open-access benchmarking sleep database. This would allow us to perform fair and informative comparisons with other existing methods, and possibly to fine-tune the parameter selection for our method. From a broader perspective, the flexibility with which the TQWT and MCA decomposition (Selesnick and Bayram, 2009; Selesnick, 2011a,b,c) can be *tuned* to specific oscillatory or transient phenomena in the signal suggests that it could be a promising tool for the detection of other structures in sleep EEG signals beyond those included in this study, such as vertex wave, slow waves and apnea.

Acknowledgments

Tarek Lajnef was supported in part by travel funds from EDST doctoral program and the LETI Laboratory, Sfax, Tunisia. Jean-Baptiste Eichenlaub is supported by the Fyssen Foundation. This study was partly performed within the framework of the LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon, within the program ANR-11-IDEX-0007. This research was undertaken, in part, thanks to funding from the Canada Research Chairs program.

References

- Acir, N., and Güzelis, C. (2004). Automatic spike detection in EEG by a two-stage procedure based on support vector machines. *Comput. Biol. Med.* 34, 561–575. doi: 10.1016/j.combiomed.2003.08.003
- Ahmed, B., Redissi, A., and Tafreshi, R. (2009). An automatic sleep spindle detector based on wavelets and the teager energy operator. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2009, 2596–2599. doi: 10.1109/iembs.2009.5335331
- Albouy, G., King, B. R., Maquet, P., and Doyon, J. (2013). Hippocampus and striatum: dynamics and interaction during acquisition and sleep-related motor sequence memory consolidation. *Hippocampus* 23, 985–1004. doi: 10.1002/hipo.22183
- Alger, S. E., Chambers, A. M., Cunningham, T., and Payne, J. D. (2014). The role of sleep in human declarative memory consolidation. *Curr. Top. Behav. Neurosci.* 25, 269–306. doi: 10.1007/7854_2014_341
- Anderson, K. N. (2015). An update in sleep neurology: the latest bedtime stories. *J. Neurol.* 262, 487–491. doi: 10.1007/s00415-014-7602-6
- Andrillon, T., Nir, Y., Staba, R. J., Ferrarelli, F., Cirelli, C., Tononi, G., et al. (2011). Sleep spindles in humans: insights from intracranial EEG and unit recordings. *J. Neurosci.* 31, 17821–17834. doi: 10.1523/JNEUROSCI.2604-11.2011
- 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
- Barakat, M., Doyon, J., Debas, K., Vandewalle, G., Morin, A., Poirier, G., et al. (2011). Fast and slow spindle involvement in the consolidation of a new motor sequence. *Behav. Brain Res.* 217, 117–121. doi: 10.1016/j.bbr.2010.10.019
- Barthó, P., Slézia, A., Mátyás, F., Faradz-Zade, L., Ulbert, I., Harris, K. D., et al. (2014). Ongoing network state controls the length of sleep spindles via inhibitory activity. *Neuron* 82, 1367–1379. doi: 10.1016/j.neuron.2014.04.046
- Bastuji, H., Perrin, F., and Garcia-Larrea, L. (2002). Semantic analysis of auditory input during sleep: studies with event related potentials. *Int. J. Psychophysiol.* 46, 243–255. doi: 10.1016/S0167-8760(02)00116-2
- Bergmann, T. O., Mölle, M., Diedrichs, J., Born, J., and Siebner, H. R. (2012). Sleep spindle-related reactivation of category-specific cortical regions after learning face-scene associations. *Neuroimage* 59, 2733–2742. doi: 10.1016/j.neuroimage.2011.10.036
- Blagrove, M., Ruby, P., and Eichenlaub, J.-B. (2013). Dreams are made of memories, but maybe not for memory. *Behav. Brain Sci.* 36, 609–610. doi: 10.1017/S0140525X13001222
- Cai, D. J., Mednick, S. A., Harrison, E. M., Kanady, J. C., and Mednick, S. C. (2009). REM, not incubation, improves creativity by priming associative networks. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10130–10134. doi: 10.1073/pnas.0900271106
- Camilleri, T. A., Camilleri, K. P., and Fabri, S. G. (2014). Automatic detection of spindles and K-complexes in sleep EEG using switching multiple models. *Biomed. Signal Process. Control* 10, 117–127. doi: 10.1016/j.bspc.2014.01.010
- Canolty, R. T., and Knight, R. T. (2010). The functional role of cross-frequency coupling. *Trends Cogn. Sci.* 14, 506–515. doi: 10.1016/j.tics.2010.09.001
- Carskadon, M. A., and Dement, W. C. (2011). “Monitoring and staging human sleep,” in *Principles and Practice of Sleep Medicine, 5th Edn.*, eds M. H. Kryger, T. Roth, and W. C. Dement (St. Louis, MO: Elsevier Saunders), 16–26. doi: 10.1016/B978-1-4160-6645-3.00002-5
- Cash, S. S., Halgren, E., Dehghani, N., Rossetti, A. O., Thesen, T., Wang, C., et al. (2009). The human K-complex represents an isolated cortical down-state. *Science* 324, 1084–1087. doi: 10.1126/science.1169626
- Chaibi, S., Lajnef, T., Sakka, Z., Samet, M., and Kachouri, A. (2014). A reliable approach to distinguish between transient with and without HFOs using TQWT and MCA. *J. Neurosci. Methods* 232, 36–46. doi: 10.1016/j.jneumeth.2014.04.025
- Chander, R. E. E. G. (2007). *Algorithms to Detect High Frequency Oscillations in Human Intracerebral Electroencephalogram*. M.E. thesis. Montreal, QC: McGill University.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 37–46. doi: 10.1177/001316446002000104
- Dalal, S. S., Hamamé, C. M., Eichenlaub, J.-B., and Jerbi, K. (2010). Intrinsic coupling between gamma oscillations, neuronal discharges, and slow cortical oscillations during human slow-wave sleep. *J. Neurosci.* 30, 14285–14287. doi: 10.1523/JNEUROSCI.4275-10.2010
- Dalal, S. S., Vidal, J. R., Hamamé, C. M., Ossandón, T., Bertrand, O., Lachaux, J.-P., et al. (2011). Spanning the rich spectrum of the human brain: slow waves to gamma and beyond. *Brain Struct. Funct.* 216, 77–84. doi: 10.1007/s00429-011-0307-z
- De Gennaro, L., and Ferrara, M. (2003). Sleep spindles: an overview. *Sleep Med. Rev.* 7, 423–440. doi: 10.1053/smr.2002.0252
- Devuyst, S., Dutoit, T., Stenuit, P., and Kerkhofs, M. (2010). Automatic K-complexes detection in sleep EEG recordings using likelihood thresholds. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2010, 4658–4661. doi: 10.1109/iembs.2010.5626447
- Devuyst, S., Dutoit, T., Stenuit, P., and Kerkhofs, M. (2011). Automatic sleep spindles detection-overview and development of a standard proposal assessment method. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2011, 1713–1716. doi: 10.1109/iembs.2011.6090491
- Diekemann, S., and Born, J. (2010). The memory function of sleep. *Nat. Rev. Neurosci.* 11, 114–126. doi: 10.1038/nrn2762
- Diekemann, S., Wilhelm, I., and Born, J. (2009). The whats and whens of sleep-dependent memory consolidation. *Sleep Med. Rev.* 13, 309–321. doi: 10.1016/j.smrv.2008.08.002
- Duman, F., Erdamar, A., Eroglu, O., Telatar, Z., and Yetkin, S. (2009). Efficient sleep spindle detection algorithm with decision tree. *Expert Syst. Appl.* 36, 9980. doi: 10.1016/j.eswa.2009.01.061
- Eichenlaub, J.-B., Bertrand, O., Morlet, D., and Ruby, P. (2014b). Brain reactivity differentiates subjects with high and low dream recall frequencies during both sleep and wakefulness. *Cereb. Cortex* 24, 1206–1215. doi: 10.1093/cercor/bhs388
- Eichenlaub, J.-B., Nicolas, A., Daltrozzo, J., Redouté, J., Costes, N., and Ruby, P. (2014a). Resting brain activity varies with dream recall frequency between subjects. *Neuropsychopharmacology* 39, 1594–1602. doi: 10.1038/npp.2014.6
- Eichenlaub, J.-B., Ruby, P., and Morlet, D. (2012). What is the specificity of the response to the own first-name when presented as a novel in a passive oddball paradigm? An ERP study. *Brain Res.* 1447, 65–78. doi: 10.1016/j.brainres.2012.01.072
- Erdamar, A., Duman, F., and Yetkin, S. (2012). A wavelet and teager energy operator based method for automatic detection of K-Complex in sleep EEG. *Expert Syst. Appl.* 39, 1284–1290. doi: 10.1016/j.eswa.2011.07.138
- Feld, G. B., Wilhelm, I., Ma, Y., Groch, S., Binkowski, F., Mölle, M., et al. (2013). Slow wave sleep induced by GABA agonist tiagabine fails to benefit memory consolidation. *Sleep* 36, 1317–1326. doi: 10.5665/sleep.2954
- Ferrarelli, F., Peterson, M. J., Sarasso, S., Riedner, B. A., Murphy, M. J., Benca, R. M., et al. (2010). Thalamic dysfunction in schizophrenia suggested by whole-night deficits in slow and fast spindles. *Am. J. Psychiatry* 167, 39–48. doi: 10.1176/appi.ajp.2010.09121731
- Ferrarelli, F., and Tononi, G. (2011). The thalamic reticular nucleus and schizophrenia. *Schizophr. Bull.* 37, 306–315. doi: 10.1093/schbul/sbq142
- Fogel, S., Albouy, G., King, B., Vien, C., Karni, A., Benali, H., et al. (2014). “Motor memory consolidation depends upon reactivation driven by the action of sleep spindles,” in *Conference of the European Sleep Research Society (ESRS)*.
- Fogel, S., Martin, N., Lafortune, 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
- Gais, S., Mölle, M., Helms, K., and Born, J. (2002). Learning-dependent increases in sleep spindle density. *J. Neurosci.* 22, 6830–6834.
- Günes, S., Dursun, M., Polat, K., and Yosunkaya, S. (2011). Sleep spindles recognition system based on time and frequency domain features. *Expert Syst. Appl.* 38, 2455–2461. doi: 10.1016/j.eswa.2010.08.034
- Halász, P. (2005). K-complex, a reactive EEG graphoelement of NREM sleep: an oldchap in a new garment. *Sleep Med. Rev.* 9, 391–412. doi: 10.1016/j.smrv.2005.04.003
- Hobson, J. A. (2009). REM sleep and dreaming: towards a theory of protoconsciousness. *Nat. Rev. Neurosci.* 10, 803–813. doi: 10.1038/nrn2716
- 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., Värri, A., Himanen, S. L., Hasan, J., Lehtokangas, M., and Saarinen, J. (2000). Optimization of sigma amplitude threshold in sleep spindle detection. *J. Sleep Res.* 9, 327–334. doi: 10.1046/j.1365-2869.2000.00220.x

- Iber, C., Ancoli-Israel, S., Chesson, A., and Quan, S. F. (eds.). (2007). *The AASM manual for the scoring of sleep and associated events: rules, terminology, and technical specification*. 1st Edn. Westchester, IL: American Academy of Sleep Medicine.
- Jahnke, K., von Wegner, F., Morzelewski, A., Borisov, S., Maischein, M., Steinmetz, H., et al. (2012). To wake or not to wake? The two-sided nature of the human K-complex. *Neuroimage* 59, 1631–1638. doi: 10.1016/j.neuroimage.2011.09.013
- Jaleel, A., Tafreshi, R., Ahmed, B., and Boivin, D. B. (2013). “Pilot validation of a mimicking algorithm for detection of sleep spindles and K-complexes,” in *World Congress on Medical Physics and Biomedical Engineering May 26–31, 2012* (Berlin; Heidelberg: Springer Berlin Heidelberg), 562–565. Available online at: http://link.springer.com/10.1007/978-3-642-29305-4_147
- Jensen, O., and Colgin, L. L. (2007). Cross-frequency coupling between neuronal oscillations. *Trends Cogn. Sci.* 11, 267–269. doi: 10.1016/j.tics.2007.05.003
- Jerbi, K., Freyermuth, S., Minotti, L., Kahane, P., Berthoz, A., and Lachaux, J. (2009b). “Chapter 12 watching brain TV and playing brain ball,” in *International Review of Neurobiology* (Elsevier), 159–168. Available online at: <http://linkinghub.elsevier.com/retrieve/pii/S0074774209860121>
- Jerbi, K., Ossandón, T., Hamamé, C. M., Senova, S., Dalal, S. S., Jung, J., et al. (2009a). Task-related gamma-band dynamics from an intracerebral perspective: review and implications for surface EEG and MEG. *Hum. Brain Mapp.* 30, 1758–1771. doi: 10.1002/hbm.20750
- Jobert, M., Poiseau, E., Jähmig, P., Schulz, H., and Kubicki, S. (1992). Pattern recognition by matched filtering: an analysis of sleep spindle and K-complex density under the influence of lormetazepam and zopiclone. *Neuropsychobiology* 26, 100–107. doi: 10.1159/000118902
- Jung, J., Jerbi, K., Ossandon, T., Ryvlin, P., Isnard, J., Bertrand, O., et al. (2010). Brain responses to success and failure: direct recordings from human cerebral cortex. *Hum. Brain Mapp.* 31, 1217–1232. doi: 10.1002/hbm.20930
- Koley, B. L., and Dey, D. (2012). “Detection of characteristic waves of sleep EEG by continuous wavelet transform,” in (IEEE) *Computing and Communication Systems (NCCCS)* (Durgapur), 1–5. doi: 10.1109/ncccs.2012.6413025
- Kouider, S., Andrillon, T., Barbosa, L. S., Goupil, L., and Bekinschtein, T. A. (2014). Inducing task-relevant responses to speech in the sleeping brain. *Curr. Biol.* 24, 2208–2214. doi: 10.1016/j.cub.2014.08.016
- Ktonas, P. Y., Golemati, S., Xanthopoulos, P., Sakkalis, V., Ortigueira, M. D., Tsekou, H., et al. (2009). Time-frequency analysis methods to quantify the time-varying microstructure of sleep EEG spindles: possibility for dementia biomarkers? *J. Neurosci. Methods* 185, 133–142. doi: 10.1016/j.jneumeth.2009.09.001
- Ktonas, P. Y., Golemati, S., Xanthopoulos, P., Sakkalis, V., Ortigueira, M. D., Tsekou, H., et al. (2014). Automated detection of sleep spindles in the scalp EEG and estimation of their intracranial current sources: comments on techniques and on related experimental and clinical studies. *Front. Hum. Neurosci.* 8:998. doi: 10.3389/fnhum.2014.00998
- Lachaux, J. P., Axmacher, N., Mormann, F., Halgren, E., and Crone, N. E. (2012). High-frequency neural activity and human cognition: past, present and possible future of intracranial EEG research. *Prog. Neurobiol.* 98, 279–301. doi: 10.1016/j.pneurobio.2012.06.008
- Lafont, M., Gagnon, J.-F., Martin, N., Latreille, V., Dubé, J., Bouchard, M., et al. (2014). Sleep spindles and rapid eye movement sleep as predictors of next morning cognitive performance in healthy middle-aged and older participants. *J. Sleep Res.* 23, 159–167. doi: 10.1111/jsr.12108
- Lajnef, T., Chaibi, S., Ruby, P., Agüera, P.-E., Eichenlaub, J.-B., Samet, M., et al. (2015). Learning machines and sleeping brains: automatic sleep stage classification using decision-tree multi-class support vector machines. *J. Neurosci. Methods* 250, 94–105. doi: 10.1016/j.jneumeth.2015.01.022
- Latreille, V., Carrier, J., Lafont, M., Postuma, R. B., Bertrand, J.-A., Panisset, M., et al. (2015). Sleep spindles in Parkinson's disease may predict the development of dementia. *Neurobiol. Aging* 36, 1083–1090. doi: 10.1016/j.neurobiolaging.2014.09.009
- Le Van Quyen, M., Staba, R., Bragin, A., Dickson, C., Valderrama, M., Fried, I., et al. (2010). Large-scale microelectrode recordings of high-frequency gamma oscillations in human cortex during sleep. *J. Neurosci.* 30, 7770–7782. doi: 10.1523/JNEUROSCI.5049-09.2010
- Lilly, J. M., and Olhede, S. C. (2012). Generalized morse wavelets as a superfamily of analytic wavelets. *IEEE Trans. Signal Process.* 60, 6036–6041. doi: 10.1109/TSP.2012.2210890
- Maquet, P. (2001). The role of sleep in learning and memory. *Science* 294, 1048–1052. doi: 10.1126/science.1062856
- Martin, N., Lafont, 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
- Möller, 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.
- Morin, A., Doyon, J., Dostie, V., Barakat, M., Hadj Tahar, A., Korman, M., et al. (2008). Motor sequence learning increases sleep spindles and fast frequencies in post-training sleep. *Sleep* 31, 1149–1156.
- Nielsen, T., and Levin, R. (2007). Nightmares: a new neurocognitive model. *Sleep Med. Rev.* 11, 295–310. doi: 10.1016/j.smrv.2007.03.004
- Nir, Y., and Tononi, G. (2010). Dreaming and the brain: from phenomenology to neurophysiology. *Trends Cogn. Sci.* 14, 88–100. doi: 10.1016/j.tics.2009.12.001
- 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
- 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). doi: 10.1109/wosspa.2013.6602401
- 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., and Nielsen, T. (2015). Automatic sleep spindle detection: benchmarking with fine temporal resolution using open science tools. *Front. Hum. Neurosci.* 9:353. doi: 10.3389/fnhum.2015.00353
- Parekh, A., Selesnick, I. W., Rapoport, D. M., and Ayappa, I. (2014). *Sleep Spindle Detection Using Time-Frequency Sparsity*. Philadelphia, PA: IEEE. doi: 10.1109/spmb.2014.7002965
- Parekh, A., Selesnick, I. W., Rapoport, D. M., and Ayappa, I. (2015). Detection of K-complexes and sleep spindles (DETOKS) using sparse optimization. *J. Neurosci. Methods* 251, 37–46. doi: 10.1016/j.jneumeth.2015.04.006
- Perrin, F., Bastuji, H., and Garcia-Larrea, L. (2002). Detection of verbal discordances during sleep. *Neuroreport* 13, 1345–1349. doi: 10.1097/00001756-200207190-00026
- Perrone-Bertolotti, M., Kujala, J., Vidal, J. R., Hamame, C. M., Ossandon, T., Bertrand, O., et al. (2012). How silent is silent reading? Intracerebral evidence for top-down activation of temporal voice areas during reading. *J. Neurosci.* 32, 17554–17562. doi: 10.1523/JNEUROSCI.2982-12.2012
- Piantoni, G., Astill, R. G., Raymann, R. J. E. M., Vis, J. C., Coppens, J. E., and Van Someren, E. J. W. (2013). Modulation of gamma and spindle-range power by slow oscillations in scalp sleep EEG of children. *Int. J. Psychophysiol.* 89, 252–258. doi: 10.1016/j.ijpsycho.2013.01.017
- Rasch, B., and Born, J. (2013). About sleep's role in memory. *Physiol. Rev.* 93, 681–766. doi: 10.1152/physrev.00032.2012
- Rechtschaffen, A., and Kales, A. (1968). *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subject*. Washington, DC: US Government Printing Office, National Institute of Health Publication.
- Riemann, D., Berger, M., and Voderholzer, U. (2001). Sleep and depression, results from psychobiological studies: an overview. *Biol. Psychol.* 57, 67–103. doi: 10.1016/S0301-0511(01)00090-4
- Ruby, P., Blochet, C., Eichenlaub, J.-B., Bertrand, O., Morlet, D., and Bidet-Caulet, A. (2013a). Alpha reactivity to complex sounds differs during REM sleep and wakefulness. *PLoS ONE* 8:e79989. doi: 10.1371/journal.pone.0079989
- Ruby, P., Blochet, C., Eichenlaub, J.-B., Bertrand, O., Morlet, D., and Bidet-Caulet, A. (2013b). Alpha reactivity to first names differs in subjects with high and low dream recall frequency. *Front. Psychol.* 4:419. doi: 10.3389/fpsyg.2013.00419
- Schabus, M., Gruber, G., Parapatics, S., Sauter, C., Klösch, G., Anderer, P., et al. (2004). Sleep spindles and their significance for declarative memory consolidation. *Sleep* 27, 1479–1485.

- Schönwald, S. V., de Santa-Helena, E. L., Rossatto, R., Chaves, M. L., and Gerhardt, G. J. (2006). Benchmarking matching pursuit to find sleep spindles. *J. Neurosci. Methods* 156, 314–321. doi: 10.1016/j.jneumeth.2006.01.026
- Seeck-Hirschner, M., Baier, P. C., Weinhold, S. L., Dittmar, M., Heiermann, S., Aldenhoff, J. B., et al. (2012). Declarative memory performance is associated with the number of sleep spindles in elderly women. *Am. J. Geriatr. Psychiatry* 20, 782–788. doi: 10.1097/JGP.0b013e31823033da
- Selesnick, I. W. (2011a). Wavelet transform with tunable Q-factor. *IEEE Trans. Signal Process.* 59, 3560–3575. doi: 10.1109/TSP.2011.2143711
- Selesnick, I. W. (2011b). *TQWT Toolbox Guide*. Electrical and Computer Engineering, Polytechnic Institute of New York University. Available online at: <http://eeweb.poly.edu/iselesni/TQWT/index.html>
- Selesnick, I. W. (2011c). “Sparse signal representations using the tunable Q-factor wavelet transform,” in *Proc. SPIE 8138, Wavelets and Sparsity XIV, 81381U* (San Diego, CA). doi: 10.1117/12.894280
- Selesnick, I. W., and Bayram, İ. (2009). “Oscillatory plus transient signal decomposition using overcomplete rational-dilation wavelet transforms,” eds V. K. Goyal, M. Papadakis, and D. Van De Ville, 74460V. doi: 10.1117/12.826237
- Soto, J. L. P., and Jerbi, K. (2012). *Investigation of Cross-Frequency Phase-Amplitude Coupling in Visuomotor Networks Using Magnetoencephalography*. San Diego, CA: IEEE. doi: 10.1109/embc.2012.6346238
- Steriade, M. (2003). The corticothalamic system in sleep. *Front. Biosci.* 8, d878–d899. doi: 10.2741/1043
- Steriade, M. (2005). Sleep, epilepsy and thalamic reticular inhibitory neurons. *Trends Neurosci.* 28, 317–324. doi: 10.1016/j.tins.2005.03.007
- Stickgold, R., and Walker, M. P. (2013). Sleep-dependent memory triage: evolving generalization through selective processing. *Nat. Neurosci.* 16, 139–145. doi: 10.1038/nn.3303
- Teolis, A. (1996). *Computational Signal Processing with Wavelets*. Boston, MA: Birkhäuser Boston. doi: 10.1007/978-1-4612-4142-3
- Valderrama, M., Crépon, B., Botella-Soler, V., Martinerie, J., Hasboun, D., Alvarado-Rojas, C., et al. (2012). Human gamma oscillations during slow wave sleep. *PLoS ONE* 7:e33477. doi: 10.1371/journal.pone.0033477
- Vidal, J. R., Perrone-Bertolotti, M., Levy, J., De Palma, L., Minotti, L., Kahane, P., et al. (2014). Neural repetition suppression in ventral occipito-temporal cortex occurs during conscious and unconscious processing of frequent stimuli. *Neuroimage* 95, 129–135. doi: 10.1016/j.neuroimage.2014.03.049
- Vorster, A. P., and Born, J. (2015). Sleep and memory in mammals, birds and invertebrates. *Neurosci Biobehav. Rev.* 50, 103–119. doi: 10.1016/j.neubiorev.2014.09.020
- Walker, M. P., and Stickgold, R. (2004). Sleep-dependent learning and memory consolidation. *Neuron* 44, 121–133. doi: 10.1016/j.neuron.2004.08.031
- Warby, C. S., Wendt, S. L., Welinder, P., Munk, E. G. S., Carrillo, O., Sorensen, H. B. D., 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. D., Peppard, P. E., Jennum, P., Perona, P., et al. (2015). Inter-expert and intra-expert reliability in sleep spindle scoring. *Clin. Neurophysiol.* 126, 1548–1556. doi: 10.1016/j.clinph.2014.10.158
- Zerouali, Y., Herry, C. L., Jemel, B., and Lina, J. M. (2013). Localization of synchronous cortical neural sources. *IEEE Trans. Biomed. Eng.* 60, 770–780. doi: 10.1109/TBME.2011.2176938
- Zerouali, Y., Lina, J.-M., Sekerovic, Z., Godbout, J., Dube, J., Jolicoeur, P., et al. (2014). A time-frequency analysis of the dynamics of cortical networks of sleep spindles from MEG-EEG recordings. *Front. Neurosci.* 8:310. doi: 10.3389/fnins.2014.00310
- Zygierevicz, J., Blinowska, K. J., Durka, P. J., Szelenberger, W., Niemcewicz, S., and Androsiuk, W. (1999). High resolution study of sleep spindles. *Clin. Neurophysiol.* 110, 2136–2147. doi: 10.1016/s1388-2457(99)00175-3

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.

Copyright © 2015 Lajnef, Chaibi, Eichenlaub, Ruby, Aguera, Samet, Kachouri and Jerbi. 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.

Automatic sleep spindle detection: benchmarking with fine temporal resolution using open science tools

Christian O'Reilly^{1,2,3*} and Tore Nielsen^{2,3}

¹ MEG Laboratory, McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, QC, Canada, ² Dream and Nightmare Laboratory, Center for Advanced Research in Sleep Medicine, Hôpital du Sacré-Coeur de Montréal, Montreal, QC, Canada, ³ Département de Psychiatrie, Université de Montréal, Montreal, QC, Canada

OPEN ACCESS

Edited by:

Srikantan S. Nagarajan,
University of California, San Francisco,
USA

Reviewed by:

Christophe Phillips,
University of Liège, Belgium
Charmaine Demanuele,
Bernstein Center for Computational
Neuroscience Heidelberg-Mannheim,
Germany
Sabrina Lyngbye Wendt,
Zealand Pharma A/S, Denmark

*Correspondence:

Christian O'Reilly,
MEG Laboratory, McConnell Brain
Imaging Centre, Montreal Neurological
Institute, McGill University, 3801 Rue
University, Montreal,
QC H3A 2B4, Canada
christian.oreilly@mail.mcgill.ca

Received: 06 January 2015

Accepted: 01 June 2015

Published: 24 June 2015

Citation:

O'Reilly C and Nielsen T (2015)
Automatic sleep spindle detection:
benchmarking with fine temporal
resolution using open science tools.
Front. Hum. Neurosci. 9:353.
doi: 10.3389/fnhum.2015.00353

Sleep spindle properties index cognitive faculties such as memory consolidation and diseases such as major depression. For this reason, scoring sleep spindle properties in polysomnographic recordings has become an important activity in both research and clinical settings. The tediousness of this manual task has motivated efforts for its automation. Although some progress has been made, increasing the temporal accuracy of spindle scoring and improving the performance assessment methodology are two aspects needing more attention. In this paper, four open-access automated spindle detectors with fine temporal resolution are proposed and tested against expert scoring of two proprietary and two open-access databases. Results highlight several findings: (1) that expert scoring and polysomnographic databases are important confounders when comparing the performance of spindle detectors tested using different databases or scorings; (2) because spindles are sparse events, specificity estimates are potentially misleading for assessing automated detector performance; (3) reporting the performance of spindle detectors exclusively with sensitivity and specificity estimates, as is often seen in the literature, is insufficient; including sensitivity, precision and a more comprehensive statistic such as Matthew's correlation coefficient, F1-score, or Cohen's κ is necessary for adequate evaluation; (4) reporting statistics for some reasonable range of decision thresholds provides a much more complete and useful benchmarking; (5) performance differences between tested automated detectors were found to be similar to those between available expert scorings; (6) much more development is needed to effectively compare the performance of spindle detectors developed by different research teams. Finally, this work clarifies a long-standing but only seldomly posed question regarding whether expert scoring truly is a reliable gold standard for sleep spindle assessment.

Keywords: sleep spindles, automatic detection, temporal resolution, reliability, sensitivity, gold standard, assessment

Introduction

Sleep spindles are bursts of energy in the 11–16 Hz band with a characteristic waning and waxing oscillation pattern of about 0.5 to 2.0-s duration that arises periodically in electrical signals captured from, for example, implanted electrodes, electroencephalography, or magnetoencephalography. This transient waveform is a hallmark of stage 2 (N2) sleep and a biomarker of some diseases

(De Gennaro and Ferrara, 2003; Ferrarelli et al., 2010; Wamsley et al., 2012), cognitive faculties (Tamaki et al., 2008; Fogel and Smith, 2011; van der Helm et al., 2011), and even normal aging (Crowley et al., 2002). Thus, an effort to better characterize the properties of sleep spindles is becoming a priority topic for neuroscience and sleep medicine. A necessary step toward this goal is to establish a commonly accepted method for evaluating the performance of automated sleep spindle scoring systems. Some notable efforts have been made in this direction by Devuyst et al. (2011) who proposed a methodology and a publicly available database. However, as will be discussed later, this database is not sufficient in itself to robustly assess the performance of automated detectors and their assessment method does not respond to certain needs of the community studying sleep spindles. One limitation concerns the use of fine temporal resolution scoring for accurately describing the microstructural features of detected spindles.

The present paper contributes to the enterprise of improving automated tools for the scoring of polysomnographic (PSG) microevents like sleep spindles by describing four different, fine temporal resolution detectors. It also provides a thorough assessment of their performance and draws key conclusions about spindle detector performance assessment in general. In next section (Spindle Scoring Evaluation), we present methodological considerations on how to evaluate the performance of spindle scorers, whether human experts or automated detectors. The Methods section describes the algorithms for the four spindle detectors with modifications to increase their temporal resolution. The developed algorithms are made available in the public domain to help improve reproducibility of research, a challenging goal given the widespread use of in-house proprietary algorithms. This section also describes four polysomnographic databases used for our investigation. The Results section assesses the performance of the modified detectors using expert scoring as a gold standard. Results are discussed in the Discussion section and suggestions for future development and assessment of automated spindle detectors are proposed in the Conclusion section.

Spindle Scoring Evaluation

Two Different Applications, Two Different Sets of Requirements

There are two very different contexts within which to score spindles and two distinct sets of requirements for assessing their performance. The first context is to identify spindles as a preprocessing step for subsequent scoring of sleep stages. Indeed, according to both AASM (Iber et al., 2007) and Rechtschaffen and Kales (1968) guidelines, the presence of spindles is a key marker of sleep stage N2. In this context, knowing only if a spindle is present in some time window (e.g., the 30-s page used to score a stage) is sufficient. The second context for scoring spindles is to study their properties in relation to other phenomena such as disease symptoms or cognitive faculties. In this context, sleep stages are generally scored manually before automatic spindle detection is attempted; such stage scoring thus constitutes

useful *a priori* information for spindle detection. Here, more precise evaluation of spindle characteristics [frequency, root mean square (RMS), amplitude, etc.] are typically of central interest.

Also in this context, timing attributes of sleep spindles, such as their onset, offset, and duration, are of considerable interest and might even be critical in precisely computing more complex characteristics such as variation of the intra-spindle instantaneous frequency or spatial propagation patterns (e.g., O'Reilly and Nielsen, 2014a,b). However, these characteristics are often overlooked when spindle scoring is undertaken for sleep-staging purposes. For example, in the DREAMS database (Devuyst et al., 2011), one of the experts scored all sleep spindles except two as having exactly a 1-s duration. Although acceptable for sleep stage scoring, such detection is suboptimal for a finer characterization of spindle attributes. It also highlights a weakness of human scorers in comparison to automated systems: experts may interpret or apply scoring criteria differently depending on the application to which they think the spindles will be put.

Fine Temporal Assessment of Spindle Scoring

Although the assessment method proposed in Devuyst et al. (2011) might be adequate when spindles are detected for sleep stage scoring, they do not assess sleep spindles with a fine temporal resolution. From this paper, we can only infer that a 1-s scoring window was used for choosing between true positive (TP), false positive (FP), true negative (TN), and false negative (FN) cases as this was not explicitly stated in the methods. A high temporal resolution alternative to this approach would be to consider spindle scoring at a signal-sampling scale (i.e., for a $f_s = 256$ Hz sampling rate, 256 TP, FP, TN, or FN outcomes are counted per second of recorded signal). As shown in Figure 1, this *signal-sample-based* approach (equivalent to the “by-sample” evaluation in Warby et al., 2014) allows for finer assessment and solves some ambiguities that occur when using a *window-based* approach (as in Devuyst et al., 2011). For example, it is not clear whether condition (e) in Figure 1 should be counted as TP, FP, or FN because the spindles detected by the two scorers are not synchronized. The degree of allowed asynchrony is directly related to the width of the decision window.

Confusion Matrix and Related Statistics

Figure 2 gives the standard confusion matrix used for assessing diagnostic tools. From this matrix, Equations (1)–(3) give the definitions of accuracy, sensitivity (a.k.a. TP rate, recall, hit rate), and specificity (a.k.a. TN rate). These statistics are often used for diagnostic applications in general and for spindle detector assessment in particular.

$$accuracy = \frac{TN + TP}{P + N} \quad (1)$$

$$sensitivity = \frac{TP}{TP + FN} \quad (2)$$

$$specificity = \frac{TN}{FP + TN} \quad (3)$$

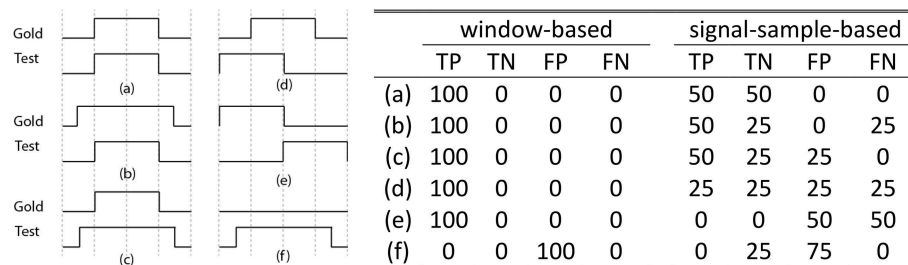


FIGURE 1 | The left panel shows six common situations [labeled as (a–f)] occurring when comparing the detection of a gold standard scorer (Gold) with another scorer (Test). The x-axis on these plots represents time. On the y-axis, a high (low) value indicates the presence (absence) of a spindle. For example, case (a) shows perfect agreement between the gold standard and the tested scorer. Resulting assessments

(TN, TP, FP, and FN, in percent) for the proposed signal-sample-based approach and for the window-based method used in Devuyst et al. (2011) are given in rightward panel. Note: The length of the scored signal is taken as being 1 s, such that there is only one decision taken for the window-based method, whereas there are f_s decisions for the signal-sample-based method.

		Gold standard outcome		Total
		Positive	Negative	
Test outcome	Positive	True positive (TP)	False positive (FP)	P'
	Negative	False negative (FN)	True negative (TN)	N'
Total		P	N	

FIGURE 2 | Confusion matrix used to assess the performance of diagnostic systems. Two scorings are necessary for this kind of assessment, one considered as giving the true outcome (gold standard) and one for which performance is established as a deviation from the true outcome (Test).

$$\text{positive predictive value} = \frac{TP}{FP + TP} \quad (4)$$

$$\text{negative predictive value} = \frac{TN}{FN + TN} \quad (5)$$

Equations (4) and (5) define two other, less frequently used, statistics: the Positive Predictive Value (PPV, a.k.a. precision) and the Negative Predictive Value (NPV). Furthermore, it is noteworthy that the False Discovery Rate (FDR) is linked to the PPV such that $FDR = 1 - PPV$. This is also true for specificity and the False Positive Rate (FPR, a.k.a. fall-out) which are related by $FPR = 1 - \text{specificity}$.

It should also be noted that accuracy is a measure of agreement between two scorings, and as such, it is independent of which scoring is used as gold standard and which is used as Test. Moreover, sensitivity and PPV are two sides of a coin; sensitivity becomes PPV when the gold standard scorer is interchanged with the Test scorer. This is also true for the relation between NPV and specificity. Thus, by listing the values of these five variables

TABLE 1 | Statistics for the comparison of spindle scorers from (Devuyst et al., 2011).

Gold	Test	Accuracy	Sensitivity	PPV	Specificity	NPV
V1	Au	95.18	66.13	30.17	96.65	99.02
V1	V2	95.47	56.40	33.81	96.73	98.40
V2	Au	95.90	63.01	58.34	98.03	98.13

Only time samples from stage 2 sleep are used to calculate these statistics.

(accuracy, sensitivity, specificity, PPV, NPV) the results of testing a scorer X against a scorer Y are completely known from the outcomes of inverse comparisons.

Devuyst et al. (2011) developed an automatic spindle detector (Au) and compared its performance with the scoring of two human experts (V1 and V2). The average performances of all three, assessed using our signal-sample-based method, are compared in Table 1. The reported statistics are all more conservative than when using the window-based method. For example, the sensitivity of the automated system (Au) is about 65% with a PPV between 30 and 60%, depending on the expert, as compared to a sensitivity of 70% and a PPV of 74% reported in Devuyst et al. (2011).

Note that the sleep spindle detection problem shows a large number of negative cases (N) with respect to the number of positive cases (P), e.g., according to the scoring of V1, the ratio between these two variables varies between 30 and 400, depending on the subject. As discussed further in O'Reilly and Nielsen (2013), in these unbalanced situations where $N \gg P$, specificity, NPV, and to a lesser extent, accuracy will always tend to be close to 1. Apparently very good specificity and sensitivity statistics alone may in fact be misleading as they can conceal a very low PPV. Thus, reported outcomes should concentrate on sensitivity and PPV (or equivalently, on the false detection rate) rather than on the typically reported sensitivity-specificity pairs. Furthermore, accuracy should be considered only as a statistic allowing comparison with other detectors and not as a statistic that is sufficient for claiming good performance in its own right.

Also as noted in O'Reilly and Nielsen (2013), these basic statistics are best supplemented with more robust statistics such as Cohen's κ (Cohen, 1960), Matthew's coefficient of correlation (MCC) (Matthews, 1975), or the F-measure—especially in the case of unbalanced datasets. Since none of these measures has yet been established as the standard for spindle scoring, we report results for all three of them.

Cohen's κ coefficient is defined by:

$$\kappa = \frac{\text{accuracy} - P_e}{1 - P_e} \quad (6)$$

where P_e is the probability of random agreement (given the bias of both scorers) defined such that:

$$P_e = \frac{P'P + N'N}{(P + N)^2} \quad (7)$$

MCC is defined by:

$$\text{MCC} = \frac{\text{TP} * \text{TN} - \text{FP} * \text{FN}}{\sqrt{P' * P * N' * N}} \quad (8)$$

The F-measure is defined by:

$$F_{\beta_C} = (1 + \beta_C^2) \frac{\text{PPV} * \text{sensitivity}}{\text{PPV} * \beta_C^2 + \text{sensitivity}} \quad (9)$$

which is a weighted harmonic mean of PPV and sensitivity with the factor β_C allows one to put more emphasis on either sensitivity or PPV (Chinchor, 1992). A special case of this measure is the F1-score which weights sensitivity and precision equally. In this case, Equation (9) reduces simply to:

$$F_1 = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \quad (10)$$

Decision Thresholds

Generally, at least at some internal level, automated classifiers produce a decision outcome X on a continuous scale, e.g., an estimated probability that a given sample is a positive. In such cases, deciding whether a tested sample should be considered as a positive or a negative applies a decision threshold λ_d such that the sample is considered as positive if $X \geq \lambda_d$ and as negative if $X < \lambda_d$. This implies that the statistics (1)–(5) are highly dependent on the value used for λ_d , making the comparison between classifiers difficult if based only on *threshold-dependent* statistics evaluated with some specific decision threshold. To obtain a more complete assessment, it is therefore preferable to evaluate the behavior of these statistics as a function of the decision threshold.

Threshold-Independent Analysis

In the context of signal detection, evaluating the performance at a specific decision threshold can be problematic. Indeed, if a first classifier obtains both sensitivity and specificity scores of 0.8 whereas a second classifier obtains scores of 0.75 and 0.85 for the two statistics, it is not clear which classifier should be selected as the best. In such a situation, the choice ultimately depends on the

costs associated with FPs and FNs, costs that are often unknown or subject to change over time or situations. Moreover, from these statistics alone it is impossible to know if there is a threshold λ_d such that one classifier will rank higher than the other on both measures simultaneously.

Receiver Operating Characteristic (ROC) Curve

ROC curves (see Fawcett, 2006 and Wojtek and David, 2009, for comprehensive overviews) have been proposed precisely to answer this question. They allow assessing classifiers under various operating conditions, i.e., using different values of λ_d .

The ROC curve is a parametric curve in the sensitivity-specificity space parameterized using the decision threshold. That is, every specific λ_d threshold is associated with a (sensitivity, specificity) point on the ROC curve, a random classifier forming a straight diagonal line from coordinates (0, 0) to (1, 1). ROC curves are increasingly used in detection problems including the assessment of spindle detectors.

Dealing with Asymmetry: the PR Curve

Using a measure complementary to the ROC curve such as the Precision-Recall (PR) curve¹ might also prove useful given the significant asymmetry between the number of negative and positive cases encountered in the spindle detection problem (Davis and Goadrich, 2006; O'Reilly and Nielsen, 2013). In this unbalanced situation, the specificity tends toward very high values for any threshold selected in practical applications because choosing thresholds associated with lower specificity would imply unacceptably low PPV. This results in only a small useful portion of the ROC curve which, therefore, benefits from being complemented with information about the behavior of the PPV statistic. This can be achieved by providing PR curves, which are parametric curves that link the TP rate to PPV, using the decision threshold as parameter. Compared to the ROC curve, the PR curve therefore eschews reliance on specificity and depends upon PPV, a more meaningful statistic for asymmetrical problems.

Correlations among Spindle Features

Detectors should also be compared for their ability to extract spindles bearing similar properties. This is probably the most important feature for detectors that are used either for characterizing sleep spindles or for investigating relationships between sleep spindle features (e.g., oscillation frequency, amplitude) and subject characteristics (e.g., age, gender, neuropsychological test scores). To evaluate this aspect of a detector, the average values of spindle features are computed within the spindle sets extracted with respect to both the gold standard and the tested classifier. This is performed separately for every recording condition (recording nights, recording channels). Then correlations between these values are computed across recording conditions using the Spearman's rank correlation coefficient. Such computation is performed for a range of threshold values to evaluate the behavior and the reliability of the detector against threshold variation but also to better assess the optimal operating threshold.

¹Also referred to as Positive Tradeoff (PT) curve in (O'Reilly and Nielsen, 2013).

High correlations should be obtained if spindles extracted by the gold standard (e.g., an expert) and a tested classifier are to be considered as assessing the same phenomenon. Indeed, if automated classifiers were to detect many more spindles than an expert (i.e., produce many FPs) but correlations between experts and automated detectors for spindle characteristics were high, we could draw two conclusions. First, that both scorings could be used to obtain similar outcomes and, second, that a higher number of spindles detected by the automated systems would probably not be an indication of FPs from the detector but rather of FNs from the expert.

In this paper, five spindle characteristics are investigated: duration, root-mean-square (RMS) amplitude, frequency slope, mean frequency, and density. Duration is defined as the length of the time window during which a *detection function* is above the decision threshold, as will be discussed more thoroughly when presenting the detectors. The window spanning the duration of the whole sleep spindle is used for RMS computation.

Technical details related to the computation of the frequency slope are described elsewhere (O'Reilly and Nielsen, 2014b). In short, it is calculated as the slope of the linear relationship between the time and the instantaneous average frequency of a spindle oscillation. It assesses the tendency of a spindle oscillating frequency not to be stable in time but to vary more or less linearly. Density is the number of detected spindles per minute. Mean frequency is computed as the average frequency of the fast Fourier transform (FFT) as described in Equation (11).

$$f_{\text{mean}} \stackrel{\text{def}}{=} \frac{\int_{10}^{16} f \cdot \text{FFT}(f) df}{\int_{10}^{16} \text{FFT}(f) df} \quad (11)$$

Methods

Databases

Four different PSG databases were used for our investigation. This diversity allowed us to assess the impact of heterogeneous databases on automated scoring and to evaluate the resilience of these detectors when used in different setups. To provide results that are easy to compare with those of other research teams, two of the databases used are open access: the DREAMS database (DDB) (Devuyst, 2013) and the Montreal Archive of Sleep Studies (MASS) (O'Reilly et al., 2014).

DDB contains eight 30 min-long EEG signals recorded on channel CZ-A1, except for two using channel C3-A1. Six recordings were sampled at 200 Hz, one at 100 Hz and one at 50 Hz. Subjects were 4 men and 4 women of about 45 years of age [standard deviation (SD): 8 years] with several different pathologies (dysomnia, restless legs syndrome, insomnia, apnoea/hypopnoea syndrome). Spindles were manually annotated by two experts (V1 and V2; V2 only annotated 6 nights). The authors of this database did not specify which scoring rules experts used for scoring spindles.

As of now, the MASS contains one cohort (C1) of 200 complete-night recordings sampled at 256 Hz and split into five subsets. The second subset (C1/SS2) contains 19 nights from young healthy subjects. For this subset, sleep spindles are scored

by two experts (V4 and V5) on N2 epochs and on channel C3 with linked-ear reference. A complete description can be found in O'Reilly et al. (2014). It should be noted that relatively low inter-rater agreement is expected between these two scorers since V4 used traditional AASM scoring rules whereas V5 used an approach similar to (Ray et al., 2010). In this case, both broad-band EEG signals (0.35–35 Hz band) and sigma filtered signals (11–17 Hz band) were used in scoring to facilitate the identification of short duration, small amplitude or obscured (e.g., by delta waves or K-complexes) spindles. Also, no minimal spindle duration was used by V5 and four nights (out of the 19) were not scored due to recordings that were judged to reflect poor quality sleep (e.g., alpha intrusions during N2) or intermittent signal quality/artifact (Fogel, personal communication).

The third database (NDB) is taken from an experiment described in detail in Nielsen et al. (2010). Only the subset of subjects not suffering from nightmares and only the two last recording nights (of a total of three consecutive nights) were used. The NDB subject sample contains 14 men [24.7 ± 5.9 (SD) years old] and 14 women [24.6 ± 6.2 (SD) years old]. Subjects were fitted with 4 referential EEG channels from the international 10–20 electrode placement system (C3, C4, O1, O2); 4 EOG channels; 4 EMG channels; 1 cardiac channel for bipolar ECG; and 1 respiration channel for nasal thermistry. Tracings were scored by trained polysomnographers applying standard criteria and using Harmonie v6.0b software. Sleep spindles were visually scored on either C3 or C4 by an expert (V3) using R&K scoring rules.

The fourth database (SDB) contains 19 complete nights from 10 young and healthy subjects (9 were recorded for two consecutive nights). Subjects were fitted with a complete 10–20 EEG electrode grid; 2 EOG channels; 3 EMG channels; 1 cardiac channel for bipolar ECG. Signals were recorded at 256 Hz using a Grass Model 15 amplifier. A linked-ear reference was used for EEG recording. Tracings were scored by trained polysomnographers applying standard criteria and using Harmonie v6.0b software. Sleep spindles were visually scored on Fz, Cz, and Pz by one of the experts (V4) who also scored the MASS spindles. In this case, spindles were scored when a burst of activity in the 12–16 Hz band was observed for 0.5–2.0 s duration.

In the following, only EEG signals from stage N2 sleep were considered. **Table 2** lists the characteristics of these four databases.

Automatic Spindle Detection with Fine Resolution

Many automatic detectors have been developed to address the tedious task of identifying sleep spindles manually (Schimicek et al., 1994; Acir and Güzeleş, 2004; Ventouras et al., 2005; Schonwald et al., 2006; Huupponen et al., 2007; Ahmed et al., 2009; Duman et al., 2009; Devuyst et al., 2011; Babadi et al., 2012). However, no implementation of these detectors has been released to the public domain—see however, other papers of this special issue which propose such open-source detectors (Durka et al., 2015; O'Reilly et al., 2015; Tsanas and Clifford, 2015)—, making it very difficult to reproduce reported results based only on the description of algorithms (Ince et al., 2012).

TABLE 2 | Specifications of the databases used in our assessment.

	DDB	NDB	MASS (SS2)	SDB
Access	Open	Closed	Open	Closed
Sampling rate	200 Hz in 6 cases; 100 Hz in 1 case; 50 Hz in 1 case	256 Hz	256 Hz	256 Hz
Number of subjects	4 men and 4 women	14 men and 14 women	8 men and 11 women	4 men and 6 women
Number of recordings per subject	1	2	1	2 for 9 subjects; 1 for 1 subject
Age	45.9 ± 8.0	24.7 ± 6.1	23.6 ± 3.7	22.4 ± 2.5
Health	Several different pathologies; see text	Healthy	Healthy	Healthy
Epoch duration	30 s	20 s	20 s	20 s
Stage scoring rules	R&K	R&K	R&K	R&K
Number of epochs per recording	360 ± 0	1483 ± 158.7	1447 ± 126.5	1480 ± 113.4
Number of N2 epochs per recording	215 ± 34.5	582 ± 109.4	689 ± 112.3	833 ± 86.2
Number of scorers	2 (V2 scored only 6 recordings)	1	2 (V5 scored only 15 recordings)	1
Spindle scoring rules	Unknown	R&K	ASSM for V4; see text for V5	0.5–2.0 s duration; 12–16 Hz band
Time-resolved scoring	Yes for V1; no for V2	Yes	Yes	Yes
Scored derivation	CZ-A1 in 6 cases; C3-A1 in 2 cases	C3 or C4 with computed ear-linked reference	C3 with resistor ear-linked reference	Fz, Cz, and Pz with resistor ear-linked reference

When using the notation $X \pm Y$, X is the mean and Y is the standard deviation.

Moreover, the algorithms of these detectors generally have a coarse temporal resolution of $\pm \frac{W_l}{2}$ where W_l is the length of an analysis window typically varying between 200 and 1000 ms. For a better characterization of spindles using fine temporal resolution, we target $\pm \frac{1}{2f_s}$. For comparative purposes, we here implement four fine resolution versions of originally coarse resolution detectors described in the literature; these detectors are based on RMS amplitude, sigma index, relative power, and the Teager energy operator. The implemented detectors are part of the Spynkle Python package, a publicly available spindle detection and analysis software toolbox (O'Reilly, 2013c).

All of the implemented detectors share the same basic structure. They first compute a *detection function* f_d , i.e., a function whose amplitude varies with the probability of spindle presence. Spindles are detected when f_d exceeds some *effective decision threshold* λ_d for a continuous duration between l_{min} and l_{max} . We qualify this threshold as effective to distinguish it from the *common threshold* λ_c (fixed value) from which λ_d is computed (i.e., it can be adaptive or not, depending on the detector). For the investigation reported in this paper l_{min} was set to 0.5 s—a suggested minimal sleep spindle duration (Iber et al., 2007)—and l_{max} to 2.0 s to avoid spurious detection of unrealistically long spindles. This upper bound is large enough to capture relevant events considering that spindle duration is generally shorter than 2.0 s; e.g., Silber et al. (2007) reported a 0.5–1.2 s range in young adults. The decision threshold can be either static or vary as a function of the EEG signal assessed for the whole night, the current NREM-REM cycle, or the current stage of the current NREM-REM cycle. For this paper, we used sleep cycles defined

as in Aeschbach and Borbely (1993) but other definitions are available as well (e.g., Feinberg and Floyd, 1979; Schulz et al., 1980). We also provide for the possibility of allowing portions of f_d to go below λ_d within the time window spanned by a spindle (i.e., it is a supplementary exception that takes precedence over the l_{min} criterion) as long as these portions are less than t_{gap} seconds long². **Figure 3** shows the pseudo-code of this general architecture.

To illustrate this detection process, **Figure 4** shows a raw signal from the second subject of DDB and its 11–16 Hz band-passed filtered version as well as the detection function and effective detection thresholds for our four detectors. Detected spindles are indicated by shaded regions.

In the following section, we describe how to obtain the detection function f_d as well as the effective thresholds λ_d for each of our four detectors (see also the Supplementary Materials for related pseudo-codes).

RMS Amplitude Detector

This algorithm is based on a methodology adopted by many researchers in the domain (e.g., Molle et al., 2002; Clemens et al., 2005; Schabus et al., 2007) and initially proposed by

²The t_{gap} parameter is included as a property of the Python classes implementing the spindle detectors. Thus, its value can be easily changed if needed. t_{gap} values used for the present investigation are reported below for reproducibility purposes, but the impact of this parameter has not been thoroughly tested yet (i.e., it has been used in an informal, trial-and-error, manual optimization) since testing it systematically would add a factor that would render our analyses prohibitively complex and computationally intensive. It is therefore likely that the t_{gap} values used in this study are suboptimal.

```

computePerCycle : Boolean; true to compute thresholds separately for all cycles
computePerStage: Boolean; true to compute thresholds separately for all stages
detectStages : List of stages where to detect spindles

detectedSpindles  $\leftarrow$  []
IF computePerCycle THEN
  cycles: List of sleep cycles
  FOR cycle IN cycles DO
    IF computePerStage THEN
      FOR stage IN detectStages DO
        signal  $\leftarrow$  recording of stage during cycle
        detectedSpindles.append(detectSpindle(signal))
      ELSE
        signal  $\leftarrow$  recording for all detectStages during cycle
        detectedSpindles.append(detectSpindle(signal))
      END IF
    END IF
  ELSE
    IF computePerStage THEN
      FOR stage IN detectStages DO
        signal  $\leftarrow$  recording of stage during the whole night
        detectedSpindles.append(detectSpindle(signal))
      ELSE
        signal  $\leftarrow$  recording for all detectStages during whole night
        detectedSpindles.append(detectSpindle(signal))
      END IF
    END IF
  END IF

FUNCTION detectSpindle(signal)
  Compute  $f_d(\text{signal})$ 
  spindles  $\leftarrow$  continuous segments where  $f_d(\text{signal}) > \lambda_d$  with possible gaps smaller
    than  $t_{gap}$ 
  RETURN elements in spindles with duration within  $[l_{min}, l_{max}]$ 
END FUNCTION

```

FIGURE 3 | Pseudo-code for the general architecture of the proposed detectors. At the end of this algorithm, detected spindles are contained in the *detectedSpindles* list.

Schimicek et al. (1994). Raw EEG signals from each channel are band-pass filtered, rejecting activity outside the spindle band. In our case, we used a 1000th order forward-backward finite impulse response filter with a Hanning window with cut-off frequencies at 11 and 16 Hz. The detection function f_{d_RMS} is defined as the RMS amplitude of the filtered signal computed within a window of length W_l repeating itself through the entire recording. The value for the effective threshold (λ_{d_RMS}) is computed as the λ_{c_RMS} percentile—the 95th percentile is generally used in the literature—of the distribution of the f_{d_RMS}

function. Since the signal amplitude may vary between and within recording nights, this effective threshold is computed separately for every sleep stage of every NREM-REM cycle of a recording.

To increase the time resolution of this method from $\pm \frac{W_l}{2}$ to $\pm \frac{1}{2f_s}$, the window used to compute the RMS can slide by one sample (maximally overlapped) instead of W_l samples (contiguous) at a time. Using a matrix-based programming language (e.g., Matlab, Python with NumPy), this can be performed efficiently even in night-long signals.

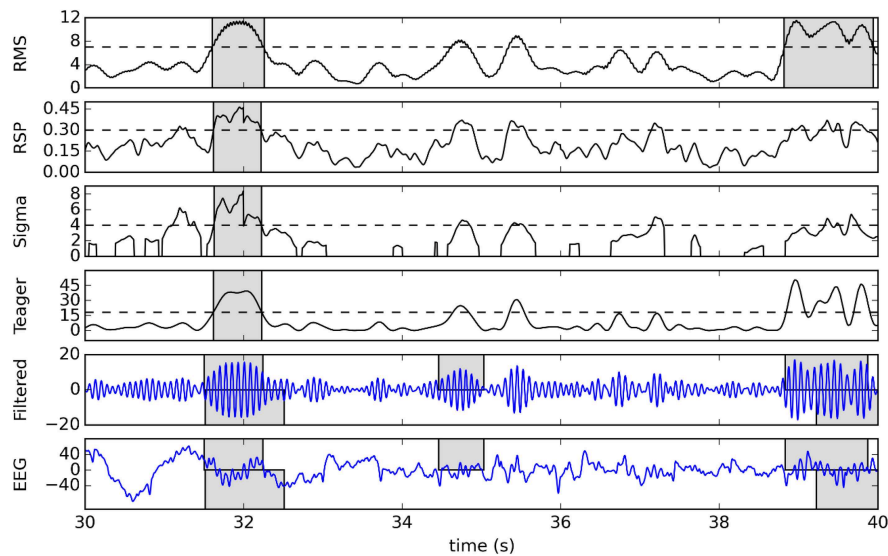


FIGURE 4 | Example of detection using a 10-s sample from the second subject of DDB. The original signal and its 11–16 Hz band-passed version are plotted in the two bottom graphs, with gray boxes showing

expert scoring (top rectangles for V1, bottom for V2). The four plots in the upper portion of the figure show corresponding detection functions (solid lines), effective thresholds (dashed lines), and detected spindles (gray boxes).

For this paper, a 200-ms averaging window and a $t_{gap} = 0$ were used.

Sigma Index Detector

This detector is based on the sigma index (Huupponen et al., 2007). To obtain good time accuracy with an acceptable computational load, we use a time-frequency representation known as the S-transform (ST) (Stockwell et al., 1996) instead of using Fast Fourier Transform (FFT) on contiguous or overlapping windows. The ST is equivalent to a short-time Fourier transform (i.e., a Fourier transform computed over small time periods using a sliding window) with a Gaussian window function whose width varies inversely with the signal frequency. Formally, this transform is expressed as:

$$ST(t, f) \stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} h(\tau) \frac{|f|}{\sqrt{2\pi}} e^{-\frac{(t-\tau)^2 f^2}{2}} e^{-i2\pi f \tau} d\tau \quad (12)$$

with t and f being transform time and frequency and $h(t)$ being the signal to be transformed. For simplicity, we used the discrete version of this transform but a fast version (i.e., similar to what the FFT is to the discrete FT) could also be used if efficiency is an important consideration (Brown et al., 2010).

To minimize processing time, the ST is computed only on the 4–40 Hz band. Since this operation cannot be performed on the whole night at once because of random-access memory limitations and heavy computational overhead³, the ST is applied on windows of 4.2 s. Windows are overlapped over 0.2 s and only the 0.1–4.1 range is used to remove artifacts at the temporal borders of the computed transform. Once the $ST(t, f)$ array is

obtained from the EEG signal, we determine the value $\max(t) = \max_{f_{spin}}(ST(t, f_{spin}))$ where $f_{spin} = [11, 16]$ Hz is the frequency range for spindle detection. In other words, $\max(t)$ is the maximal energy along the frequency axis at a given time t , in the sigma band. We then determine the detection function as the sigma index.

$$f_{d_SIGMA}(t) = \begin{cases} 0 & \text{if } \max(ST(t, f_{\alpha})) > \max(t) \\ f_{\alpha} & \text{else} \end{cases} \quad (13)$$

with $m_l(t) = \text{mean}(ST(t, f_l))$, $m_h(t) = \text{mean}(ST(t, f_h))$, f_l is the 4–10 Hz band, f_h is the 20–40 Hz band, and f_{α} is 7.5–10 Hz band. That is, for each time t , the sigma index is the maximal energy in the spindle band normalized by the average between the energy values in the f_l and f_h bands to control for wide band artifacts such as those caused by muscular activity. Moreover, this index is completed by an alpha rejection step which states that the value of the sigma index is canceled out if the maximal energy in the alpha band f_{α} is larger than the maximal energy in the sigma band.

Although computed using different signal processing algorithms, the sigma index used here follows the definition proposed in Huupponen et al. (2007). These authors suggest applying a threshold $f_{d_SIGMA}(t) > \lambda_{d_SIGMA}$ with $\lambda_{d_SIGMA} = \lambda_{c_SIGMA} = 4.5$. Note that there is no difference between the effective and the common threshold in this case, the effective threshold being taken as a fixed value. We further used $t_{gap} = 0.1$.

Relative Spindle Power Detector

Following the ideas proposed in Devuyst et al. (2011), we implemented a detection function based on the relative spindle power (RSP).

³This is true for the discrete ST. However, since the fast ST can be computed in-place (i.e., without additional attribution of memory), it should be computable on the whole night at once.

TABLE 3 | Definition for effective thresholds λ_d ; tested variation ranges, optimal values according to our investigations, and previously suggested values in the literature for common thresholds λ_c .

Detector	Threshold definition for λ_d	Tested range for λ_c	Optimal value for λ_c	Suggested in literature for λ_c
RSP	λ_{c_RSP}	[0.1, 0.5]	0.3	0.22
RMS	$\text{percentile}(\lambda_c, f_{d_RMS})$	[0.7, 0.995]	0.92	0.95
Sigma	λ_{c_SIGMA}	[1.0, 8.0]	4.0	4.5
Teager	$\lambda_{c_TEAGER} * \overline{f_{d_TEAGER}}$	[0.5, 8.0]	3.0	0.6

All effective decision thresholds are applied directly to the corresponding detection functions f_d . The $\text{percentile}(p, s)$ function computes the percentile p of the distribution of a signal s . $\overline{f_{d_TEAGER}}$ stands for the average value of f_{d_TEAGER} .

$$f_{d_RSP}(t) = \frac{\int_{11}^{16} ST(t, f) df}{\int_{0.5}^{40} ST(t, f) df}. \quad (14)$$

That is, it represents the instantaneous ratio of the power of the EEG signal in the 11–16 Hz band divided by its power in the 0.5–40 Hz band. Power computation is performed using the S-transform as described in the previous section.

The implementation details for this detector are exactly the same as for the detector based on the sigma index, except that $f_{d_SIGMA}(t)$ is changed to $f_{d_RSP}(t)$ and an adequate threshold is applied ($\lambda_{d_RSP} = \lambda_{c_RSP} = 0.22$ was proposed in Devuyst et al., 2011). We further used $t_{gap} = 0$.

Teager Detector

Based on Ahmed et al. (2009) and Duman et al. (2009), we used the Teager energy operator as another detection function. This operator is defined as:

$$f_{d_TEAGER} = h^2(n) - h(n-1)h(n+1) \quad (15)$$

where $h(n)$ is the digital signal (e.g., the EEG time series in our case) which is transformed into the detection function f_{d_TEAGER} by the right-hand side of the equation and n is the (discrete) time variable. Duman et al. (2009) propose a decision threshold at $\lambda_{c_TEAGER} = 60\%$ of the average amplitude (i.e., $\lambda_{d_TEAGER} = \lambda_{c_TEAGER} * \overline{f_{d_TEAGER}}$ where $\overline{f_{d_TEAGER}}$ is the mean value of f_{d_TEAGER}). We further used $t_{gap} = 0$.

Scripting

For transparency and better reproducibility of these results, Python scripts used to generate the results presented are provided in the *examples* repository of the Spynle package version 0.4.0 available at https://bitbucket.org/christian_oreilly/spynle.

Artifacts

No artifact rejection was performed prior to spindle detection. Some detection functions were designed to reject artifacts, e.g., the sigma-index which is designed to reject alpha band activity and muscular artifact. We wanted to test these detectors in the worst conditions to determine their resilience even in the presence of artifacts.

Results

Five analyses performed in this study are described in detail in the next sections. The first compares the detectors against each

expert scorer using threshold-dependent statistics computed for a range of decision threshold values (see **Table 3** for actual ranges). The second analysis is similar but compares correlations between pairs of detectors/experts for average values of spindle characteristics. The third analysis presents ROC and PR curves for the different detectors using expert scoring as a gold standard. The fourth analysis assesses threshold-dependent statistics for detectors operating with common thresholds judged to be optimal according to our investigations (see **Table 3** for corresponding values). These thresholds are subjective choices made by visual inspection following a thorough assessment and motivated by the fact that they balance performance estimates (i.e., attempt to maximize the MCC, F1 and the Cohen κ ; see **Figure 5**) across the expert scorings⁴.

A final section presents comparative processing times for the four proposed detectors.

Comparative Performances for Threshold-Dependent Statistics

Figure 5 shows results obtained for threshold-dependent statistics using large ranges of decision thresholds for testing against each expert scoring. Whereas simpler statistics generally monotonically increase (specificity and PPV) or decrease (sensitivity) with respect to the decision threshold, more complete statistics (e.g., Cohen κ , F1, and MCC) are low for extreme thresholds and maximal for intermediate values, better capturing the tradeoff between low FPs and FNs.

Reliability of Spindle Characteristics

Results from previous sections show the extent of the agreement between automated detectors and experts. However, for investigating relationships between sleep spindle properties and subject characteristics it is important to know to what extent the latter relationships are affected by these partial agreements. In other words, we want to verify if these correlations can be reliably assessed regardless of the specific expert or detector used to score spindles. To assess this, the median values of some sleep spindle characteristics (RMS amplitude, density, duration, oscillation mean frequency, instantaneous slope of

⁴We consider these thresholds to be a good tradeoff for most uses. However, depending on the application, one might want to give more weight to sensitivity or to precision. For a specific application, one can choose the operating point that will result in expected performances using **Figures 5–7**.

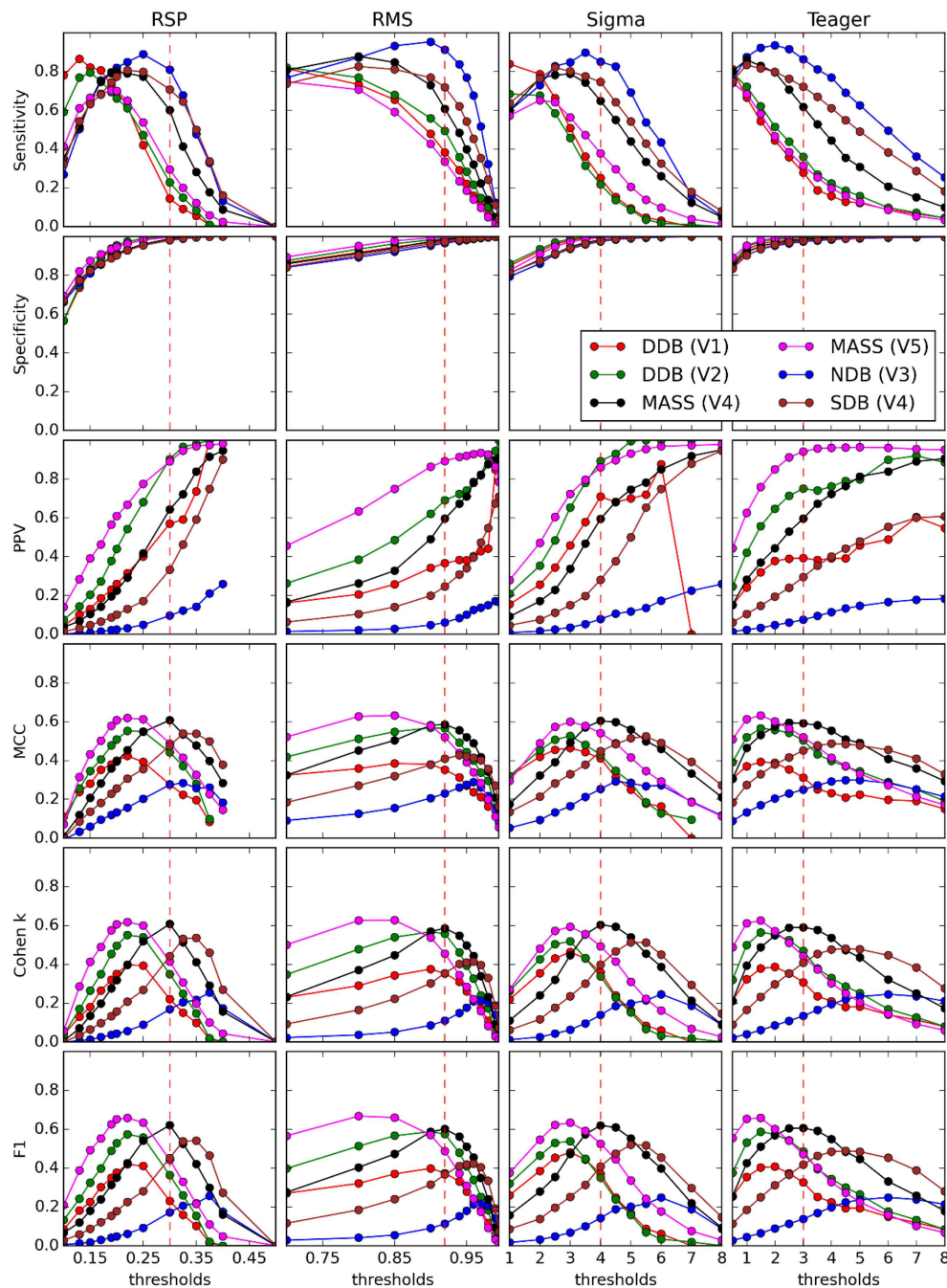


FIGURE 5 | Performance of four detectors (columns) compared against five experts (V1–V5) scoring four databases (DDB, NDB, MASS, SDB) (see legend on graphs) for six threshold-dependent statistics (rows). Vertical dashed red lines show optimal thresholds, as reported in Table 3.

intra-spindle frequency) are computed for each scored channel of each recorded night. These sets of median values are then compared between pairs of detectors/experts using Spearman correlations. Such computation is performed again for a large range of detection thresholds as shown in Figure 6. In this figure, correlations for V2's estimates of duration are not reported because this expert did not score spindle duration (i.e., every

spindle was noted as having a 1-s duration, except for two spindles of 0.49 and 0.5 s).

Figure 6 shows how spindle characteristics correlate between experts and automatic detectors but do not allow evaluation of whether there is any offset between the different scorings. Presence of such offsets can be assessed in Figure 7 which shows actual spindle characteristic values.

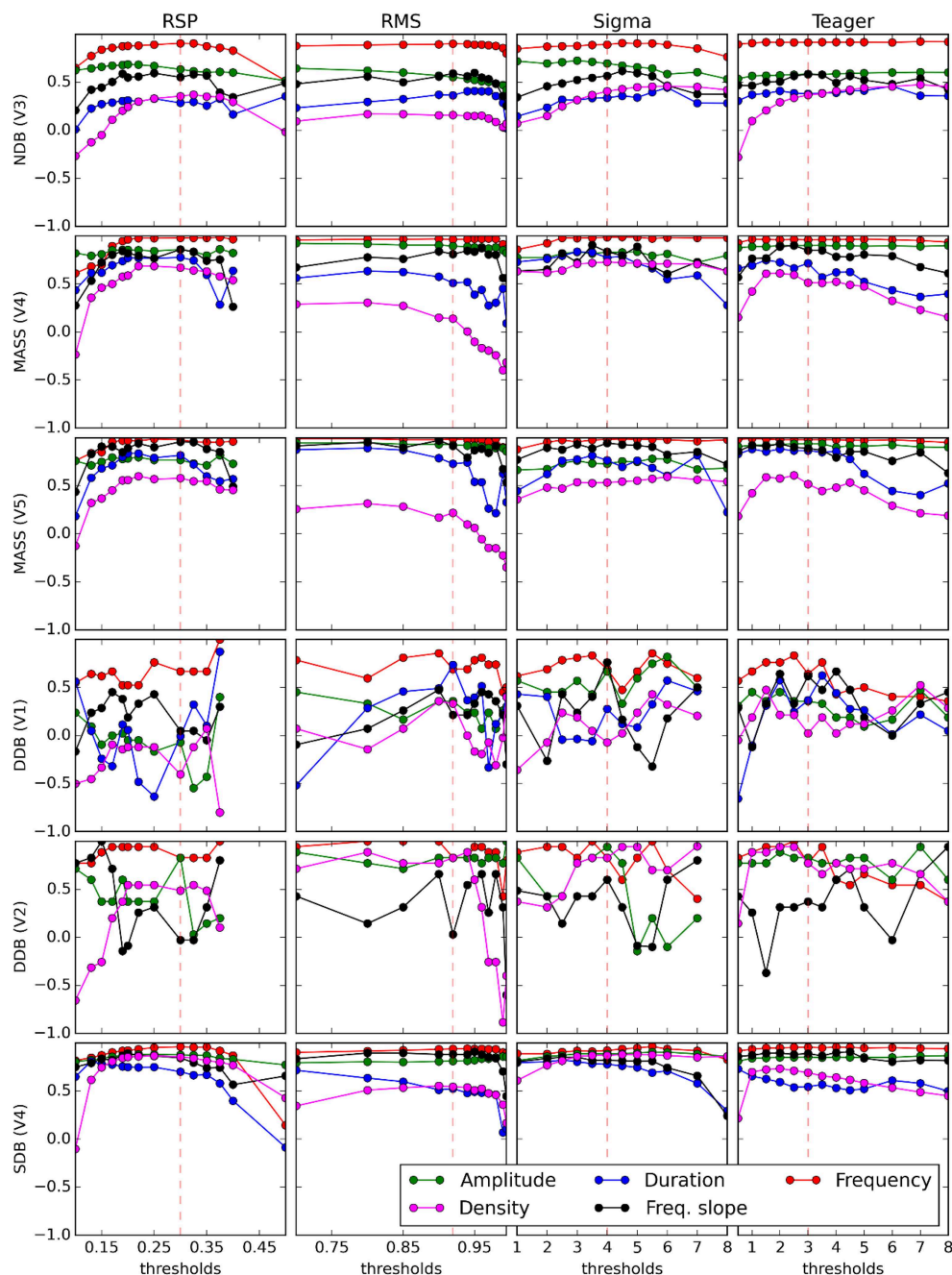


FIGURE 6 | Curves display variation of the Spearman coefficient of correlation between the median value of spindle features (see legend on graphs) computed from an expert scorer (rows) and an

automated detector (columns). Decision thresholds are varied in graph abscises. Vertical dashed red lines show optimal thresholds, as reported in Table 3.

ROC and PR Curves

Figures 8 and 9 show the ROC and PR curves, respectively, for each of the four classifiers. Given the asymmetry of the spindle detection problem, the portion of the ROC curve with specificity less than 0.8 is of no interest since this portion corresponds to useless operating conditions with PPV below 0.2 (this can be observed by comparing specificity and PPV graphs in Figure 5).

Thus, ROC graphs have been truncated to focus on the most informative parts.

As can be seen, PR and ROC curves do not increase monotonically, as is generally expected for such curves. This is a consequence of setting an upper limit on spindle duration. Indeed, with such a limit, using lower thresholds causes an increase in sensitivity up to a certain limit, after which

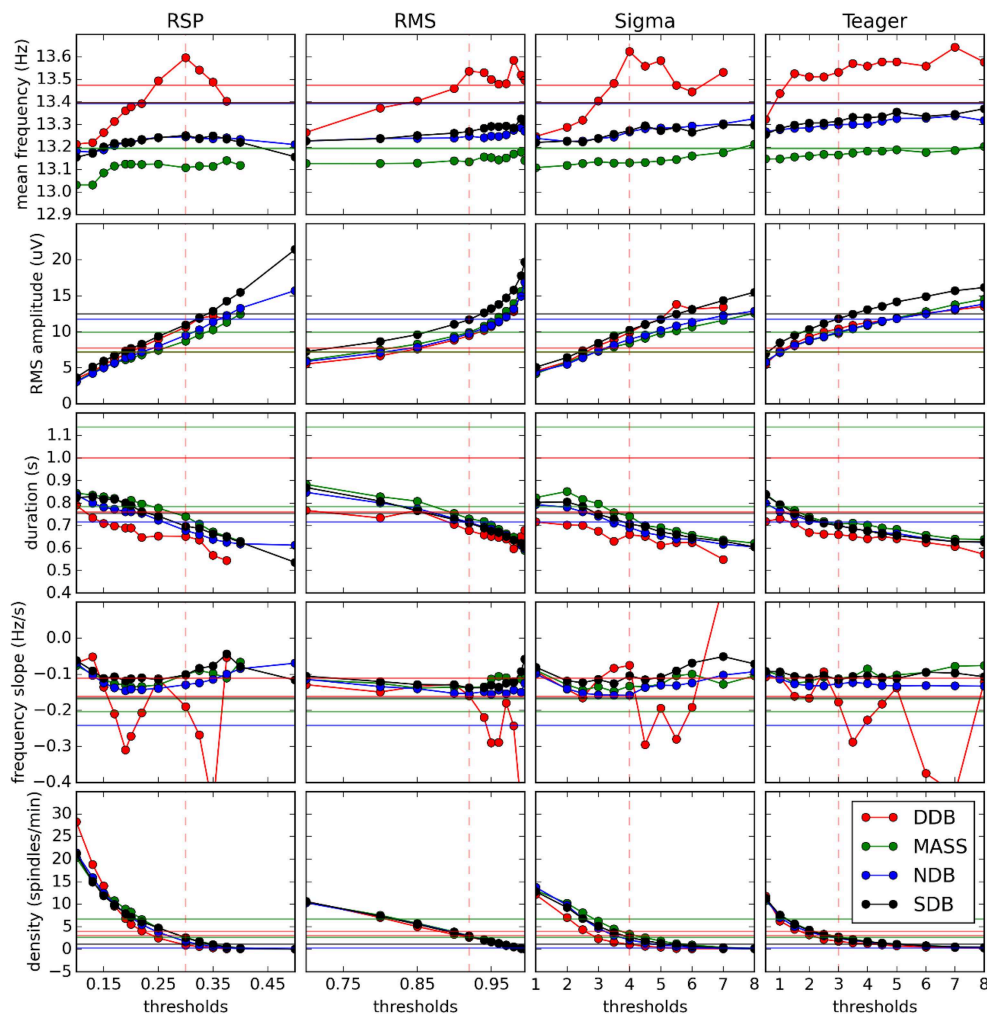


FIGURE 7 | Variation of spindle characteristics as a function of decision threshold for each of the four automatic detectors. Straight horizontal lines show values for expert scoring. Vertical dashed red lines show optimal thresholds as reported in **Table 3**.

excessively long spindles occur and are rejected, lowering the specificity.

Threshold-Dependent Statistics at Optimal Decision Threshold

Figure 10 shows performances that can be expected when comparing each expert scoring to the different detectors using optimal decision thresholds as specified in **Table 3**. Accordingly, these plots would change for a different choice of threshold. Each box represents the distribution of the median value of a given statistic (e.g., specificity) across recording conditions (recording nights, EEG derivations) for a specific expert's scoring [e.g., DDB (V1)] and a specific detector (e.g., RMS).

Processing Time

Computations were performed on Intel Core i7-3970X processors @ 3500 GHZ, using 32 GB of RAM memory (DDR3 @ 800 Hz), running a 64-bit Windows 7 operating

system. Since this system has 12 cores and spindle detectors run in single threads, the detection of spindles for all nights, with all 4 detectors, at all threshold values—i.e., detection of spindles for 4488 whole-nights and 408 30-min long signals—was automated and run in 11 parallel detection processes using BlockWork (O'Reilly, 2013b) and EEG Analyzer (O'Reilly, 2013a).

Aside from detection performances, processing time required by the detectors is sometimes an important practical constraint. For example, our assessment would have taken about half a CPU-years if spindle detection for a whole-night of EEG signal took 1 h to complete. Fortunately, the proposed detectors are substantially faster. **Figure 11** compares the average processing time for each detector, with durations assessed on the MASS nights. Most of the computation time required for spindle detection is associated with three distinct tasks: loading the signals in memory (blue), detecting the spindle (green), and saving the annotations on hard drive (red). As would be expected, only the event detection is significantly affected by the choice of detector. There is about one

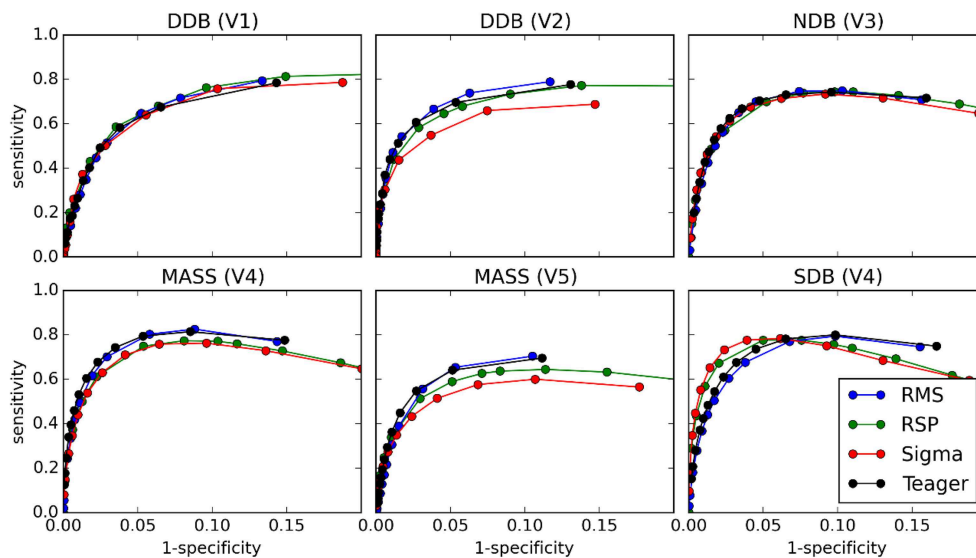


FIGURE 8 | ROC curves for comparisons between the four classifiers (tests) and scoring by experts (gold standard).

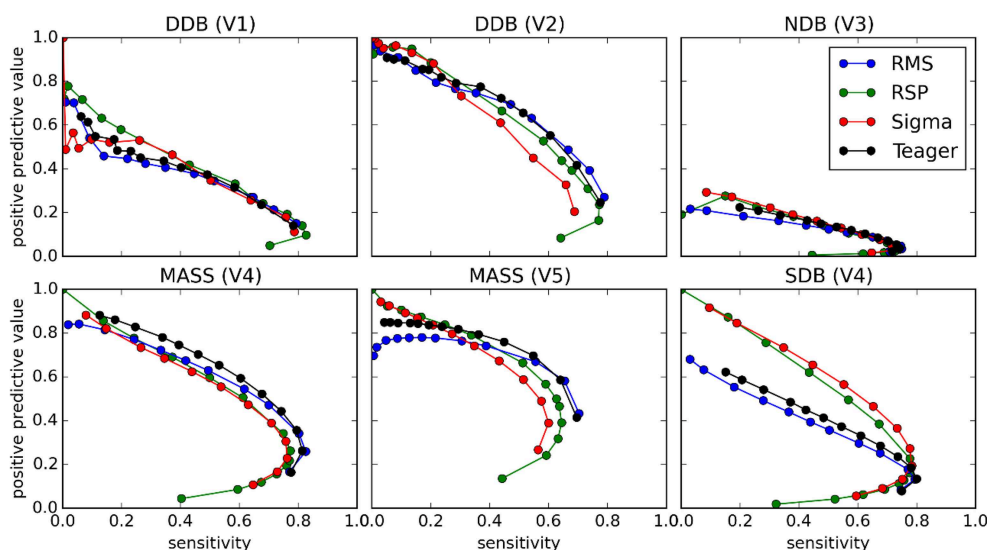


FIGURE 9 | PR curves for comparisons between the four classifiers (tests) and scoring by experts (gold standard).

order of magnitude between the processing time requirements for event detection of the fastest (Teager; 32 s) and slowest (RSP; 402 s) detectors.

Discussion

Comparative Performance Assessment for Spindle Detectors

As discussed in the Spindle Scoring Evaluation section, the most interesting characteristics for threshold-dependent evaluation of sleep spindle detectors are sensitivity and PPV (precision) as well as more complete statistics such as Cohen's κ , F1-score,

and MCC. Specificity is of low interest since the relative scarcity of spindles in sleep EEG forces it to take high values for any reasonable PPV. This is exemplified in **Figures 5** and **10**. In fact, specificity values can be considered misleading in that they give the false impression that a detector has good performance even if it is not necessarily the case. In light of this, it appears prudent to report PPV or FDR instead of specificity as a measure of a detector's ability to reject FPs.

It is, however, obvious from **Figure 9** that the impact of the choice of an expert/database combination has even more influence on PPV than the choice of a detector. This highlights the fact that PPV is directly related to how conservative the

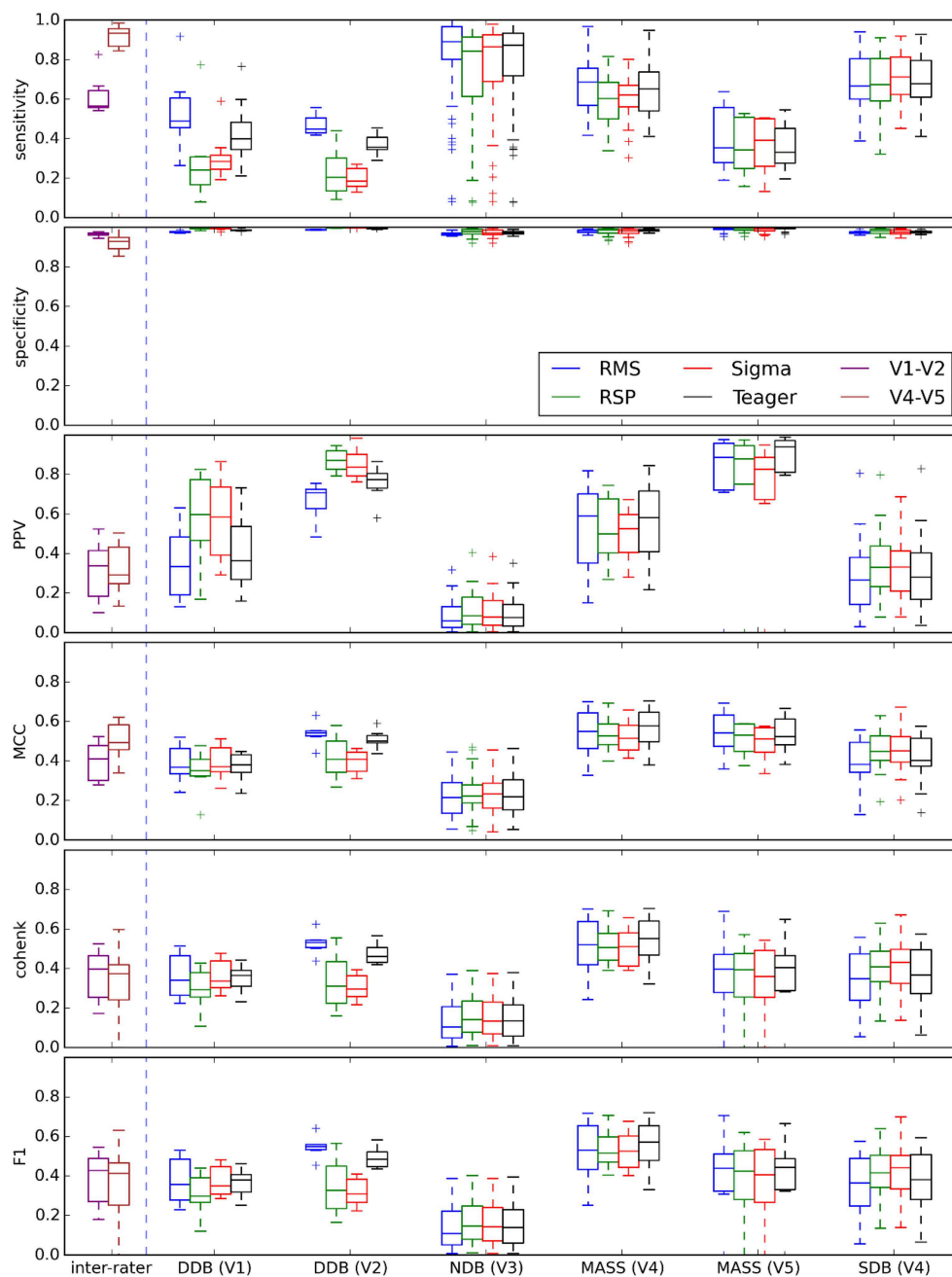
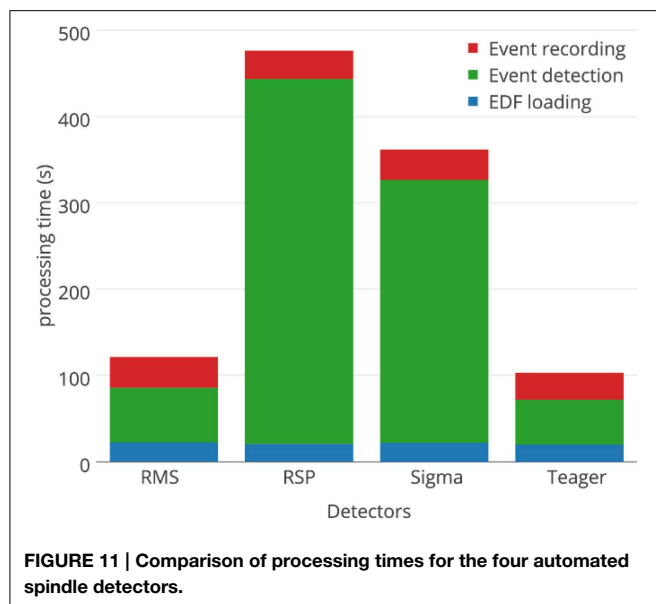


FIGURE 10 | Box plots summarizing the distribution of threshold-dependent statistics for expert/detector comparisons using decision thresholds judged optimal as reported in Table 3. Results for each detector are color coded (see the

legend). Note that the first two boxes (in purple and brown) at the left of each vertical dashed line show the distributions of statistics comparing experts V1 and V2 (DDB) and V4 and V5 (MASS).

expert is when detecting spindles (i.e., the extent to which an expert systematically scores fewer spindles per night than do other experts; see also spread of the optima for MCC, Cohen's κ , and F1-score in **Figure 5** which depicts the same phenomenon). It suggests that PPV is more indicative of the relative importance of FNs from the expert part than FPs from the detector part. In this context, it appears ill-advised to compare spindle detectors

for which assessments were performed on different databases or different expert scorings. Indeed, the expert scoring and database are two important confounding factors that can completely mask true differences in detector performance. The importance of these confounders on PPV is particularly obvious, but is clearly also true of the other performance statistics (sensitivity, MCC, F1-score, Cohen κ) as can be seen in **Figure 10**. Fortunately,



open-access databases that can be used for comparative purposes are starting to become available. We hope that the present results will incite researchers to propose additional open-access databases or to contribute to existing ones.

Choosing the best decision threshold is rather difficult and almost impossible to do objectively using sensitivity and PPV curves (Figure 5), ROC curves (Figure 8), or PR curves (Figure 9). Such a choice requires estimation of the costs associated with both FP and FN errors. Since these costs are difficult to evaluate and can vary depending on context, MCC, F1-score, and Cohen's κ provide attractive alternatives. These three statistics give very similar assessments with clearly identifiable maxima close to FN/FP tradeoffs that are generally adopted in the literature. Since correlation coefficients are well understood by the general scientific community whereas use of Cohen's κ is restricted more to the field of psychology, MCC might be a good choice of statistics to report. Furthermore, MCC lends itself readily to parametrical statistical analysis since it is related to the χ^2 distribution (Baldi et al., 2000). The F-measure, on the other hand, has the advantage of explicitly specifying weights on the relative importance of sensitivity versus PPV, whereas the tradeoff is implicit in MCC and Cohen's κ . Similarly, the F1-score implicitly considers these two statistics as being of equal importance, something that might not be true in general. Regardless, no consensus has yet emerged concerning which of these three statistics is best to report, but reporting all three might be preferable when assessing a detector on an open-access database so as to maximize the possibility of comparing detector performances across studies. In any case, at least one such statistic should be reported to provide a more comprehensive view of the detector's performance.

Another important conclusion is that there is an inherent difficulty deciding which automated detector performs best relative to expert scoring using statistics computed at only one specific threshold. Shifts in the decision threshold can

produce very different results. Thus, reporting the value of threshold-dependent statistics over some reasonable range of decision thresholds is desirable.

It should also be noted that, because databases and experts constitute two important sources of variability, one should exercise caution in comparing results from studies presenting algorithms that use general classification rules based on heuristics (e.g., the detectors proposed here) with those from studies using detectors that are trained on a database of pre-scored spindles (e.g., Acir and Güzeleş, 2004) unless the training and the testing subsets in the latter are taken from different databases and scored by different experts. Indeed, the maximally attainable performances for heuristic and trained systems are quite different. In the former case, the best performances that can be expected when comparing a detector with different experts are limited by the relatively low average agreement between experts (inter-expert reliability). In the latter case, if scoring from the same expert is used both for training and testing, the maximal performance that the automated detector can attain is only limited by intra-expert reliability.

Impact of Scorers on Averaged Spindle Characteristics

As can be seen in Figure 6, the inter-scorer reliability of spindle characteristics can be loosely ranked, from most to least reliable, as follows: frequency, amplitude, frequency slope, duration, and density. This ordering does not seem to be affected much by the choice of detector. It seems, however, that all curves can be displaced up or down by differences in the quality of the database and the expert scoring. Also, it is perhaps concerning to see that spindle density—the most frequently used spindle characteristic in sleep research—is in fact the least reliably evaluated characteristic. This is not surprising though since density is the only characteristic considered here that is not computed by averaging its value across spindles (i.e., the density is defined directly at the subject level as a count whereas the other characteristics are defined at the level of individual spindles and their value at the level of the subject is obtained by averaging across a large number of spindles). Including, for example, 10% more or fewer events in the averaging process may not cause a large difference for stable characteristics. However, this would cause a rather large error ($\pm 10\%$) for density.

Figure 7 also shows that at optimal thresholds there is generally good agreement between the characteristics of spindles labeled by experts and by detectors, with no large offsets between these two kinds of scorings. In this figure, we see that the frequency slope cannot be reliably evaluated on the DDB. This is likely due to the short duration of the recordings (30 min instead of whole nights) which does not allow for the detection of enough spindles to stabilize computation of the median value. This is most visible for the frequency slope because this measure is harder to estimate reliably on individual spindles than are other properties such as average frequency. Except for this specific case and the results for frequency, detectors tend to agree closely across databases, contrary to the experts. This is consistent with the hypothesis that different experts work with different detection thresholds.

Choice of an Open-Access Database

Results obtained with DDB have a restricted utility because of severe limitations on the features of this database. For example, the DDB is relatively small, containing only 4 h of recording (8 sequences of 30 min) on one channel. This results in unreliable assessment as can be seen in **Figure 6**. In contrast, the portion of MASS that was scored for spindles is much larger; about 150 h of recording (19 nights of about 8 h). Another limitation of DDB is in its recording parameters. For example, the EEG of one subject is sampled at 50 Hz, which theoretically allows assessment of frequencies up to 25 Hz without aliasing; however, in practice imperfect filtering produces aliasing even at lower frequencies. **Figure 10** also shows generally similar agreement between expert scoring on MASS and on DDB, even if low agreement was expected for MASS given the fact that it was scored by two different teams using two different approaches. Thus, using only the DDB does not appear to be sufficient to provide a robust assessment of spindle detectors and a more complete database such as MASS is preferable for such a purpose.

On the other hand, DDB has the advantage of presenting signals for clinical cases. These can serve as examples or for case studies. Also, DDB has a high value in open-science for fast validation, teaching, and tutorials since it is directly downloadable on the Internet, something not possible for ethical reasons with MASS.

ROC and PR Curves

Results from ROC and PR curves are not conclusive. Detector rankings according to these curves vary from expert to expert. They may, therefore, not constitute the most appropriate tools for assessing spindle detectors. In a related vein, because of the asymmetry of the spindle detection problem, most of the ROC curve is associated with uninteresting operating conditions. Computing the area under this curve hence produces an aggregated measure that is obtained from mostly useless conditions. Therefore, the area under the ROC curve (AUC) does not appear appropriate for assessing the performance of spindle detectors.

Choosing the Best Detector

Even with the thorough assessment proposed here, we cannot with good confidence determine the best classifier. Our ability to do so is limited by the lack of a highly reliable gold standard.

Moreover, the required characteristics of a detector may change depending on the desired application. Here is a short list of some of the most important qualities/features that vary with different applications: (1) requirement or not of sleep stage scoring; (2) rejection or not of artifacts; (3) temporal precision of spindle detection; (4) simplicity of the algorithm; (5) efficiency of the code (e.g., code execution time); (6) overall classification performance; (7) reliability of detected spindle characteristics; (8) capacity for extracting spindles that are correlated with other dependent variables (e.g., neurophysiological and neuropsychological variables).

In general, RMS and Teager detectors are good picks for applications requiring simple deployment and rapid processing. The Sigma detector, however, seems more reliable for estimating

spindle characteristics when compared against expert scoring. Further, we found that 0.3, 0.92, 4.0, and 3.0 are appropriate values for decision thresholds used with the RSP, RMS, Sigma, and Teager detectors, respectively. These thresholds are close to previously proposed values for Sigma (4.5 vs. 4.0) and RMS (0.95 vs. 0.92). The threshold for the RSP detector is also not too discrepant from previously proposed values (0.22 vs. 0.30). However, for the Teager detector, the previously proposed threshold is five times lower than the one found here (0.6 vs. 3.0). The reason for such a discrepancy between our results and those of Duman et al. (2009) is presently unknown. This detector seems, however, particularly sensible to characteristics of the database. Thus, finding a better approach to adapt the effective decision threshold to the characteristics of individual subjects might help to stabilize the performance of this detector.

Note also that, except for DDB, our assessment was made on young healthy subjects. This is important because sleep spindle properties (e.g., density, frequency, morphology, spatial distribution) vary with age and brain and sleep disorders, such as sleep apnea. These associations have practical implications for using these detectors on clinical datasets. For example, recordings taken from the elderly might need lower decision thresholds to accommodate less pronounced spindles in this population. Precision of the detector would evidently suffer from such an accommodation. Thus, a thorough assessment of the behavior of these detectors is advisable before using them with populations known to have smaller amplitude spindles, more artifacts, or smaller signal-to-noise ratios (SNR).

The Problem of the Gold Standard

As previously mentioned, our results suggest that a significant proportion of the FPs traditionally attributed to automatic detectors might rather be due to FNs from experts. This raises the question of the adequacy of expert scoring as a gold standard for evaluating spindle detectors. The general reliability of expert scoring can indeed be questioned considering that, in our findings, expert scoring has more influence on automatic detection than does the choice of automated detector. This is further supported by the fact that experts V1–V2 and V4–V5 agree more closely with one another than they do with most of the automated detectors (see **Figure 10**). These results are in line with reports of a relatively low reliability for expert scoring. For example, F1-scores of $72 \pm 7\%$ (Cohen κ : 0.52 ± 0.07) for intra-rater agreement and $61 \pm 6\%$ (Cohen κ : 0.52 ± 0.07) for inter-rater agreement have been reported by Wendt et al. (2014).

Our results suggest that there is ample room for improvement of automatic spindle detectors. However, the extent of this improvement is unclear because of low reliability of the gold standard currently available for spindle identification. Without a robust gold standard, results will continue to be limited by average inter-rater agreement. Consensus from a large number of crowd-sourced scoring judges is a possible alternative to expert scoring as a gold standard (Warby et al., 2014), but it remains unproven that common agreement of a large number of low-qualification scorers will provide better detection of atypical, unusual or non-obvious spindles than will experts. Low-qualification scorers will in all likelihood show high reliabilities

only on large amplitude spindles with large signal-to-noise ratios. Similarly, it is unclear if consensus scoring of a few experts would, in the long run, be retained as a practical solution. This would require substantial resources and runs counter to the tremendous efforts invested in automation of sleep spindle scoring designed to reduce the burden of manual processing to begin with. It might prove to be a sound approach for scoring only subsets of recordings that can be used for training classifiers to detect the entire database (e.g., O'Reilly et al., 2015). Alternatively, manual validation of automatically scored recordings could prove to be quicker for experts than would be manual scoring of the tracings, and thus would provide a reasonable compromise.

Although automated spindle detectors have been in use for several decades, their development and assessment still require substantial work. As they mature, expert scoring will need to be abandoned in favor of criteria based on construct validation results that reflect the growing capacities of computerized automation and statistical assessment. This task could be facilitated by incorporating correlations between detected spindle characteristics and psychological, physiological and demographic dependent variables. We would expect that spindle features obtained from random detectors would correlate only poorly with such variables, whereas spindle features obtained from detectors tapping genuine neurophysiological phenomena would correlate robustly.

Limitations

Consensus scoring was not pursued for this study but it clearly warrants consideration in future work. For this analysis, double scoring was only available for two databases. The first (DDB) produced rather unreliable results while the second (MASS) produced low inter-expert agreement. Higher inter-rater agreement in MASS could have been pursued by allowing both experts to consult and align their scorings. We would argue, however, that this is not representative of scoring used in the field. We chose instead to ask experts from two different centers to score these recordings as they would in their research. These low agreements are more representative of the variability in expert scoring that we observe between studies published by different centers than is an artificially increased agreement of experts aligning their scoring through consultation.

It is also noteworthy that no artifact rejection was performed prior to spindle detection. Thus, our results show the relative resilience of these detectors to the presence of artifacts. However, in clinical settings where many artifacts are expected, signals should be adequately preprocessed (i.e., cleaned of artifacts) to ensure robust detection. This is especially true for consistent artifacts that might affect the computation of detection thresholds, e.g., the presence of many high-amplitude arousals or flat segments. Fortunately, the use of percentile statistics in the definition of thresholds should render these thresholds relatively robust compared to thresholds based on, e.g., averages and standard deviations, as long as artifacts introduce only a non-significant amount of activity to the top percentiles of the amplitude distribution.

Conclusion

As we demonstrate in the present paper, assessing the performance of automated spindle detectors is a complex enterprise. The superiority of a new detector can no longer be supported merely by reporting that threshold-dependent variables such as sensitivity and precision are superior to those of previously published detectors. These basic statistics should be supplemented—at a minimum—by more complete statistics such as MCC. However, because there exist no commonly agreed upon testing conditions (i.e., standard databases, relative positive and negative error costs, etc.) and since these conditions may change with different usage contexts, better estimates of external validity (and, thus, a more general validation) can be obtained by reporting the values of these statistics across a range of decision thresholds. The most useful results are obtained by also providing access to the detector source code such that other research teams may test the detector's performance under different conditions. If authors are not willing to share source code, sharing of at least an executable copy with documentation should be considered.

Aside from the dynamics of spindle detectors themselves, other important topics in detector assessment concern the methodological environment of the evaluation. One key topic is the availability of a validated gold standard against which automatic scoring may be evaluated. Expert scoring has been used *de facto* as a trustworthy gold standard, but this assumption is challenged by our results. Although for the present experts will most certainly keep their gold standard status in spindle detection, the definition of a more reliable and commonly agreed upon standard is urgently needed if progress in the domain is to continue.

A second matter needing attention is the availability of EEG databases. As shown in our results, outcomes from different databases can be quite different depending on the database representativeness (i.e., characteristics of the subject sample), size (i.e., are there enough records to obtain stable averages?), and reliability (appropriate sampling frequency, recording equipment, etc.). The availability of shared databases is critical for the development of new algorithms and the benchmarking of various systems on the same set of biological recordings. Pooling of multiple scorings from experts of different research teams could also help in capturing inter-expert variability when developing classifiers that require training.

Unfortunately, the implementation (both executables and source code) of existing sleep spindle detectors described in the literature are not widely available, making their reproducibility, standardization, and benchmarking difficult to attain. In an effort to stimulate progress in this regard, we provide open source spindle detectors for use by the other researchers working in this area (see the Spynle package, O'Reilly, 2013c) along with a comprehensive assessment of their performance.

Funding

Funding provided by Canadian Institutes of Health Research (MOP-115125) and Natural Sciences and Engineering Research

Council of Canada (312277) to Tore Nielsen and a postdoctoral fellowship to Christian O'Reilly.

Acknowledgments

The authors thank Tyna Paquette and Sonia Frenette for technical assistance, Elizaveta Solomonova for scoring the NDB, Julie Carrier for sharing the data of SDB, Stuart Fogel for

providing a second set of expert scoring for MASS, and Simon Warby for valuable comments on this manuscript.

Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fnhum.2015.00353>

References

- Acir, N., and Güzelış, 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.tb00065.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* (Minneapolis, MN), 2596–2599.
- 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
- Baldi, P., Brunak, S., Chauvin, Y., Andersen, C. A., and Nielsen, H. (2000). Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics* 16, 412–424. doi: 10.1093/bioinformatics/16.5.412
- Brown, R. A., Lauzon, M. L., and Frayne, R. (2010). A general description of linear time-frequency transforms and formulation of a fast, invertible transform that samples the continuous s-transform spectrum nonredundantly. *IEEE Trans. Signal Process.* 58, 281–290.
- Chinchor, N. (1992). "MUC-4 evaluation metrics," in *Proceedings of the 4th Conference on Message Understanding* (McLean, VA), 22–29.
- 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
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 37–46. doi: 10.1177/001316446002000104
- Crowley, K., Trinder, J., Kim, Y., Carrington, M., and Colrain, I. M. (2002). The effects of normal aging on sleep spindle and K-complex production. *Clin. Neurophysiol.* 113, 1615–1622. doi: 10.1016/S1388-2457(02)00237-7
- Davis, J., and Goadrich, M. (2006). "The relationship between Precision-Recall and ROC curves," in *Proceedings of the 23rd International Conference on Machine Learning* (Pittsburgh, PA: ACM).
- De Gennaro, L., and Ferrara, M. (2003). Sleep spindles: an overview. *Sleep Med. Rev.* 7, 423–440. doi: 10.1053/smr.2002.0252
- Devuyst, S. (2013). *The DREAMS K-complexes Database*. Available online at: <http://www.tcts.fpm.ac.be/~devuyst/Databases/DatabaseKcomplexes/>
- Devuyst, S., Dutoit, T., Stenuit, P., and Kerkhofs, 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* (Boston, MA), 1713–1716.
- Duman, F., Erdamar, A., Eroglu, O., Telatar, Z., and Yetkin, S. (2009). Efficient sleep spindle detection algorithm with decision tree. *Expert Syst. Appl.* 36, 9980–9985. doi: 10.1016/j.eswa.2009.01.061
- Durka, P. J., Malinowska, U., Zieleniewska, M., O'Reilly, C., Rózański, P. T., and Żygierewicz, J. (2015). Spindles in Svarog: framework and software for parametrization of EEG transients. *Front. Hum. Neurosci.* 9:258. doi: 10.3389/fnhum.2015.00258
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognit. Lett.* 27, 861–874. doi: 10.1016/j.patrec.2005.10.010
- Feinberg, I., and Floyd, T. C. (1979). Systematic trends across the night in human sleep cycles. *Psychophysiology* 16, 283–291. doi: 10.1111/j.1469-8986.1979.tb02991.x
- Ferrarelli, F., Peterson, M. J., Sarasso, S., Riedner, B. A., Murphy, M. J., Benca, R. M., et al. (2010). Thalamic dysfunction in schizophrenia suggested by whole-night deficits in slow and fast spindles. *Am. J. Psychiatry* 167, 1339–1348. doi: 10.1176/appi.ajp.2010.09121731
- Fogel, S. M., and Smith, C. T. (2011). The function of the sleep spindle: a physiological index of intelligence and a mechanism for sleep-dependent memory consolidation. *Neurosci. Biobehav. Rev.* 35, 1154–1165. doi: 10.1016/j.neubiorev.2010.12.003
- 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
- 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.
- Ince, D. C., Hatton, L., and Graham-Cumming, J. (2012). The case for open computer programs. *Nature* 482, 485–488. doi: 10.1038/nature10836
- Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta* 405, 442–451. doi: 10.1016/0005-2795(75)90109-9
- 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.
- Nielsen, T. A., Paquette, T., Solomonova, E., Lara-Carrasco, J., Popova, A., and Levrier, K. (2010). REM sleep characteristics of nightmare sufferers before and after REM sleep deprivation. *Sleep Med.* 11, 172–179. doi: 10.1016/j.sleep.2008.12.018
- O'Reilly, C. (2013a). *EEG Analyzer*. Available online at: https://bitbucket.org/christian_oreilly/eeg-analyzer
- O'Reilly, C. (2013b). *PyBlockWork*. Available online at: https://bitbucket.org/christian_oreilly/pyblockwork
- O'Reilly, C. (2013c). *Spyndle*. Available online at: https://bitbucket.org/christian_oreilly/spyndle
- 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
- 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).
- 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

- Ray, L. B., Fogel, S. M., Smith, C. T., and Peters, K. R. (2010). Validating an automated sleep spindle detection algorithm using an individualized approach. *J. Sleep Res.* 19, 374–378. doi: 10.1111/j.1365-2869.2009.00802.x
- 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., Baletau, 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., Zeitlhofer, J., Anderer, P., and Saletu, B. (1994). Automatic sleep-spindle detection procedure: aspects of reliability and validity. *Clin. Electroencephalogr.* 25, 26–29. doi: 10.1177/155005949402500108
- Schonwald, S. V., de Santa-Helena, E. L., Rossatto, R., Chaves, M. L., and Gerhardt, G. J. (2006). Benchmarking matching pursuit to find sleep spindles. *J. Neurosci. Methods* 156, 314–321. doi: 10.1016/j.jneumeth.2006.01.026
- Schulz, H., Dirlich, G., Balteskonis, S., and Zulley, J. (1980). The REM-NREM sleep cycle: renewal process or periodically driven process? *Sleep* 2, 319–328.
- Silber, M. H., Ancoli-Israel, S., Bonnet, M. H., Chokroverty, S., Grigg-Damberger, M. M., Hirshkowitz, M., et al. (2007). The visual scoring of sleep in adults. *J. Clin. Sleep Med.* 3, 121–131.
- Stockwell, R. G., Mansinha, L., and Lowe, R. P. (1996). Localization of the complex spectrum: the S transform. *IEEE Trans. Signal Process.* 44, 998–1001.
- Tamaki, M., Matsuoka, T., Nittono, H., and Hori, T. (2008). Fast sleep spindle (13–15 Hz) activity correlates with sleep-dependent improvement in visuomotor performance. *Sleep* 31, 204–211.
- Tsanas, A., and Clifford, G. D. (2015). Stage-independent, single lead EEG sleep spindle detection using the continuous wavelet transform and local weighted smoothing. *Front. Hum. Neurosci.* 9:181. doi: 10.3389/fnhum.2015.00181
- van der Helm, E., Gujar, N., Nishida, M., and Walker, M. P. (2011). Sleep-dependent facilitation of episodic memory details. *PLoS ONE* 6:e27421. doi: 10.1371/journal.pone.0027421
- 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
- Wamsley, E. J., Tucker, M. A., Shinn, A. K., Ono, K. E., McKinley, S. K., Ely, A. V., et al. (2012). Reduced sleep spindles and spindle coherence in schizophrenia: mechanisms of impaired memory consolidation? *Biol. Psychiatry* 71, 154–161. doi: 10.1016/j.biopsych.2011.08.008
- 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].
- Wojtek, J. K., and David, J. H. (2009). *ROC Curves for Continuous Data*. Boca Raton: Chapman & Hall/CRC.

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.

Copyright © 2015 O'Reilly and Nielsen. 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.



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:

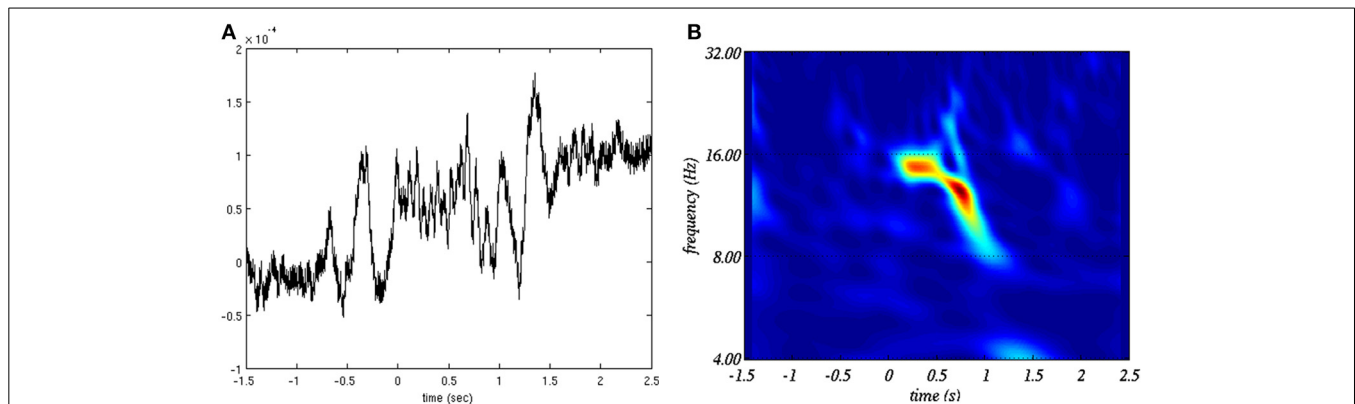


FIGURE 1 | Example of sleep spindle (A) in time domain and (B) in the time-frequency plane.

$$\tilde{x}_n^{\text{sigma}} = \frac{2a_\sigma}{a_\alpha + a_\beta}$$

$$= \frac{2 \max \left\{ |W|_{[10.5 - 16 \text{ Hz}; t_n^{\max}]} \right\}}{\text{mean} \left\{ |W|_{[4 - 10 \text{ Hz}; t_n^{\max}]} \right\} + \text{mean} \left\{ |W|_{[20 - 40 \text{ Hz}; t_n^{\max}]} \right\}} \quad (6)$$

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 :

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

where $f_n^{\max} = f_i$ with $i = \underset{1 \leq i \leq N_f}{\text{argmax}} \left| w_{i,j}^{(m)} \right|$ 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

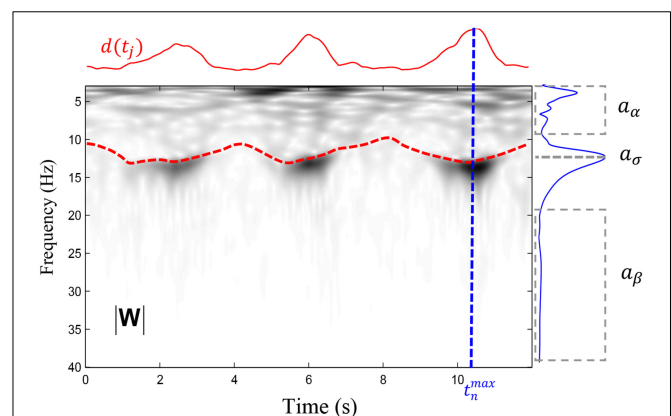


FIGURE 2 | Illustration of the variables entering in the computation of $\tilde{x}_n^{\text{sigma}}$ as defined in (6). The time-frequency plane ($|W|$), in gray levels, is represented by the modulus of wavelet coefficients computed for frequencies between 4 and 40 Hz with 0.1 Hz resolution. The red 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 instantaneous 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 $a_\alpha = \text{mean} \left\{ |W|_{[4 - 10 \text{ Hz}; t_n^{\max}]} \right\}$ and $a_\beta = \text{mean} \left\{ |W|_{[20 - 40 \text{ Hz}; t_n^{\max}]} \right\}$. We further take the maximal amplitude in the spindle band to obtain $a_\sigma = \max \left\{ |W|_{[10.5 - 16 \text{ Hz}; t_n^{\max}]} \right\}$.

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:

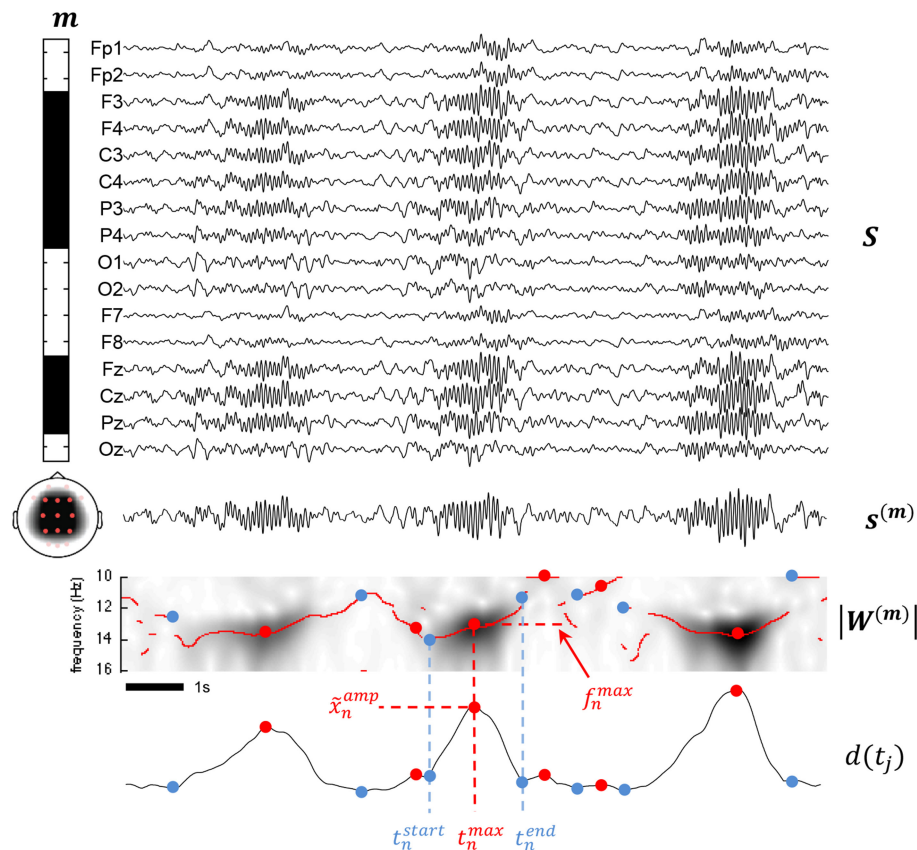


FIGURE 3 | This figure summarizes various aspects of the proposed methodology. The montage defined by the m vector is illustrated by the leftward band displaying binary weights including frontal, parietal, and central channels and excluding the other channels from the mixture. Below the band, a scalp map shows the topological coverage of this montage (red dots for included electrodes, pink for excluded). At the right of the montage band, all available channels are stack in a matrix S . The mixture signal $s^{(m)}$ is obtained by matrix multiplication of m and S

as described in (1). Applying the CWT to $s^{(m)}$ and taking the modulus, we obtain a time-frequency map $|W^{(m)}|$ as shown. Using (4) on $|W^{(m)}|$, we compute the $d(t_j)$ detection function used to determine the times t_n^{start} , t_n^{max} , and t_n^{end} . \tilde{x}_n^{amp} is obtained as $d(t_n^{max})$ and f_n^{max} as f_i with

$$i = \underset{1 \leq i \leq N_f}{\operatorname{argmax}} |W_{i,j}^{(m)}|.$$

$$t_j = t_n^{max}$$

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

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^{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, $\tilde{x}_n^{expert} = 1$ 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 $\tilde{x}_n^{sensitive} = [\tilde{x}_n^{amp} \ \tilde{x}_n^{sigma}]$ is not associated with a spindle. For this assessment, a sample of the null-hypothesis, i.e., non-spindle events, is built from $\tilde{x}_n^{sensitive}$ of all events with $\tilde{x}_n^{stage} = 5$. 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

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

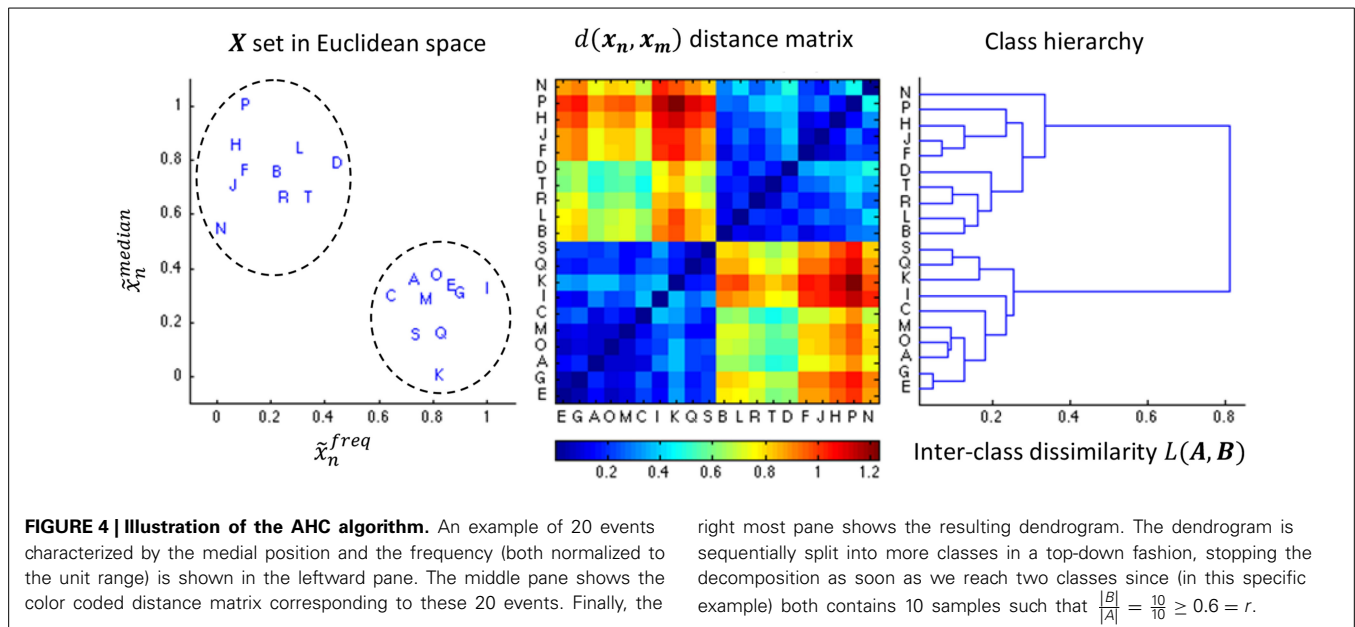
For assessing performances, we used a terminology borrowed from confusion matrices. Four classification outcomes can be encountered in the dual-class problem considered here: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). If we consider a variable $x_n^{selected}$ 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:

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

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

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

Meaure	Formula
Sensitivity (Se)	$Se = \frac{N_{TP}}{N_{FN} + N_{TP}}$
Specificity (Sp)	$Sp = \frac{N_{TN}}{N_{FP} + N_{TN}}$
False positive rate (FPr)	$FPr = \frac{N_{FP}}{N_{FP} + N_{TN}} = 1 - Sp$
False Discovery Rate (FDr)	$FDr = \frac{N_{FP}}{N_{FP} + N_{TP}}$
False positive proportion (FPP)	$FPP = \frac{N_{FP}}{N_{FN} + N_{TP}}$
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	$F1 = \frac{2N_{TP}}{2N_{TP} + N_{FP} + N_{FN}}$
Cohen κ	$\kappa = \frac{X - (N_{TN} + N_{TP})N_e}{X - N_e^2}$ with $X = (N_{TP} + N_{FP})(N_{TP} + N_{FN}) + (N_{FP} + N_{TN})(N_{FN} + N_{TN})$



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

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

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 k Ω 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**.

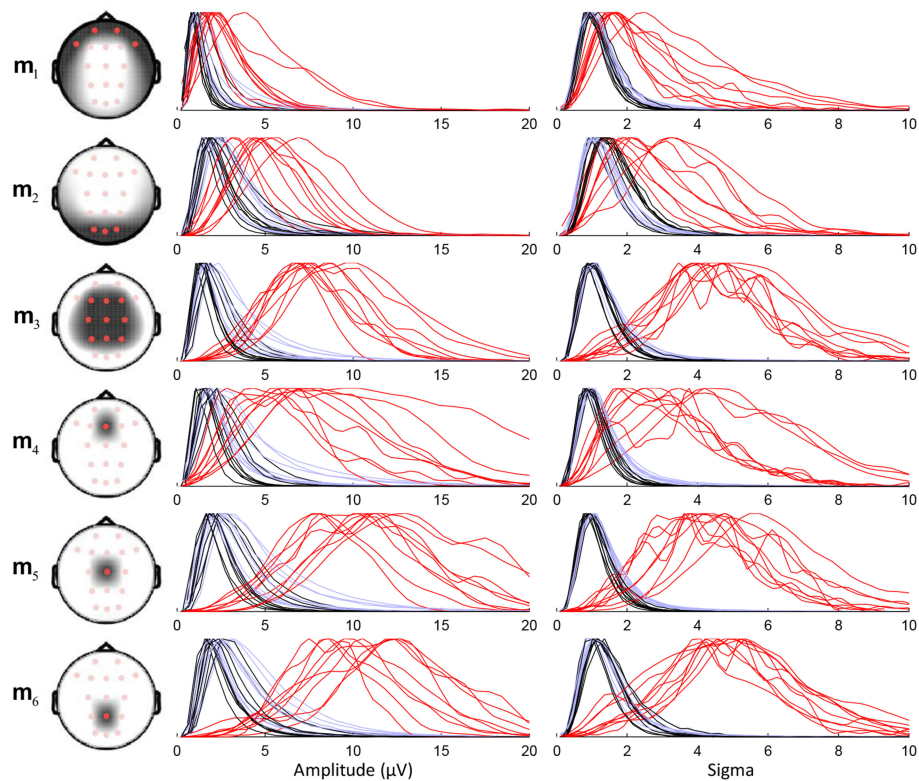


FIGURE 5 | Density plots with normalized maximal amplitude for the distribution of sensitive detection features (amplitude and sigma index) for six different montages. The head drawings show the topographical coverage of each montage. Feature distributions are plotted

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.

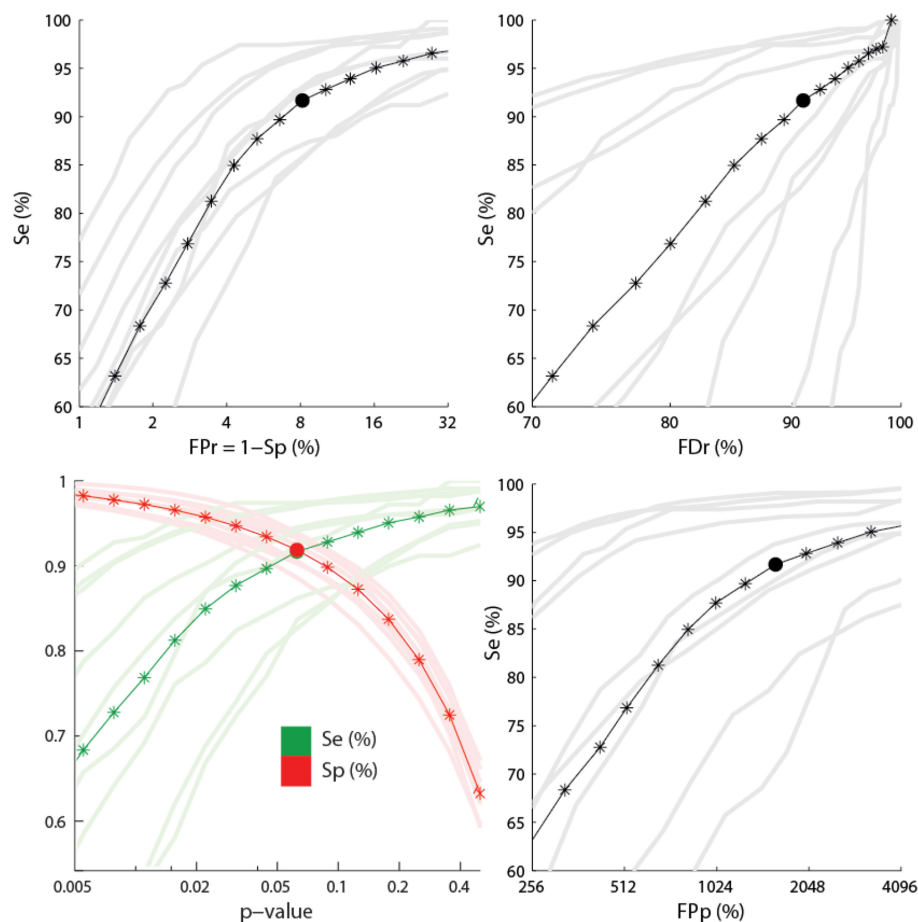


FIGURE 6 | Performance for the sensitive detection. The lower left panel shows the sensitivity (green) and the specificity (red) as function of the p -value used for feature thresholding. The three other panels show ROC-like curves using FPr, FDr, and FPP. Dark colors show average statistics whereas

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 $S_e = S_p$ and when $\alpha = 0.1$ for the sensitive and the specific phase.

Mesure	Sensitive phase		Specific phase
	$S_e = S_p$	$\alpha = 0.1$	$\alpha = 0.1$
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

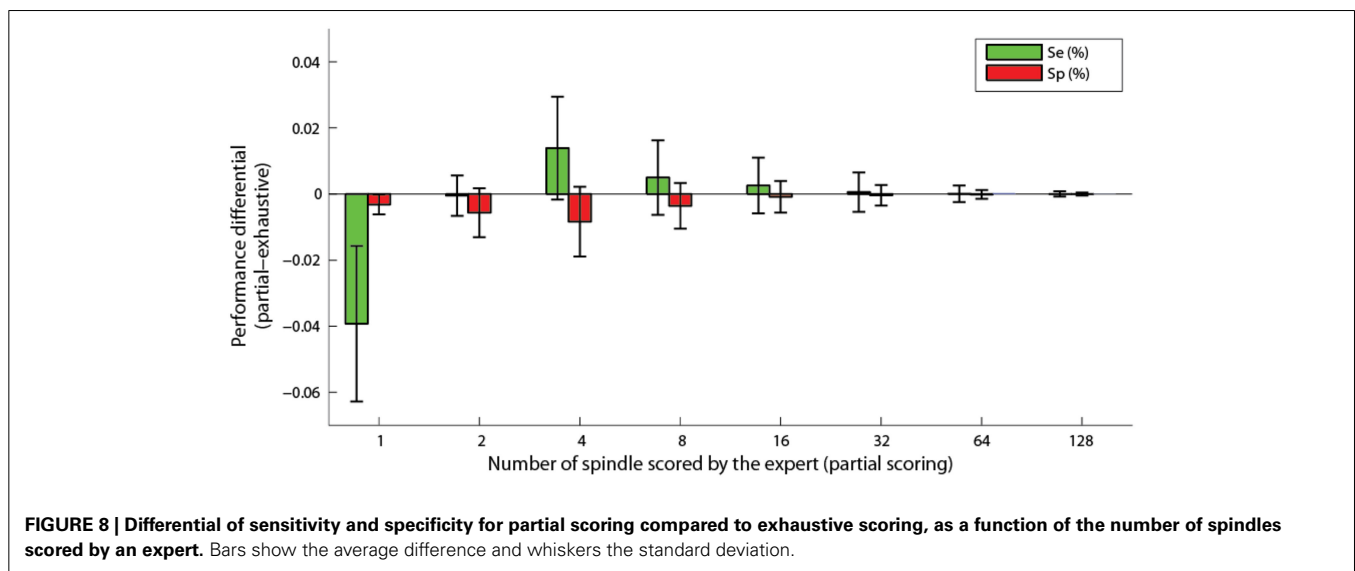
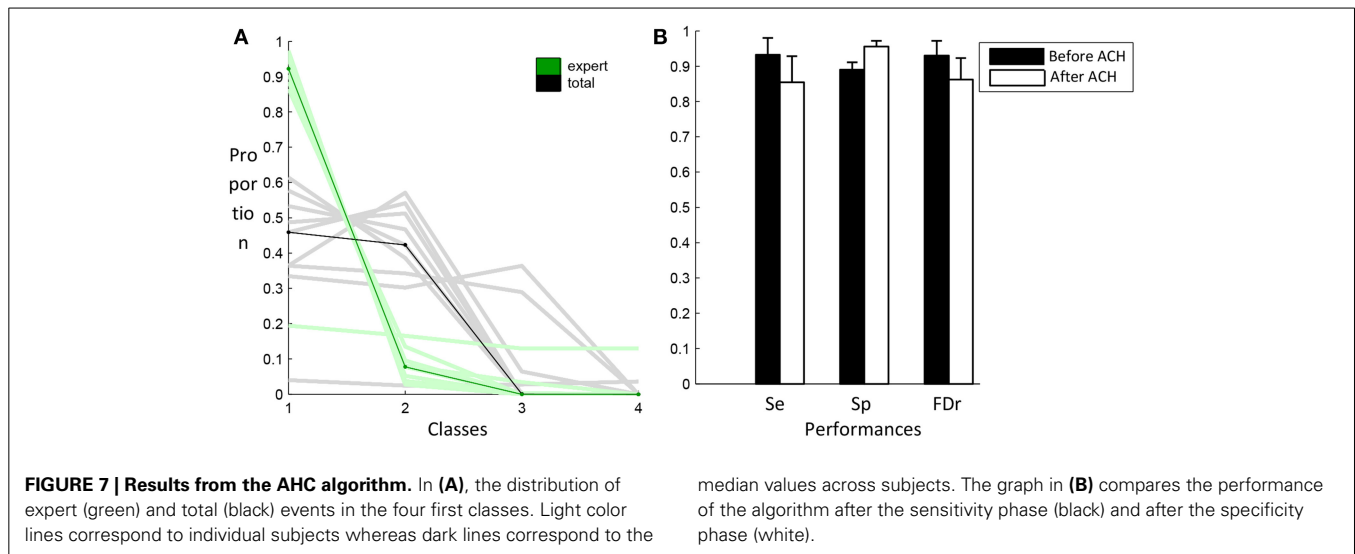
Frequency and medial position

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 four 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.

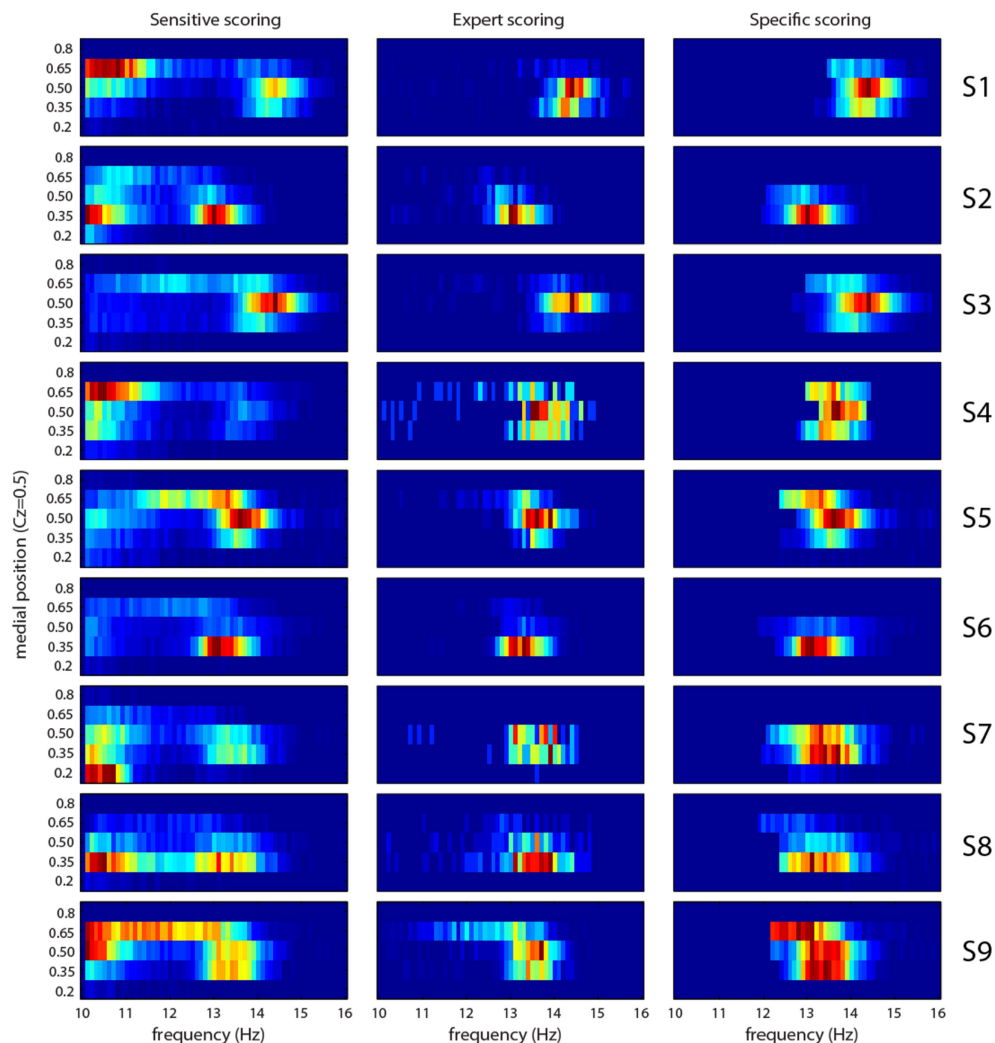


FIGURE 9 | Joint distribution of frequency and medial position (0.15 is occipital and 0.9 is prefrontal) features for each subject (S1–S9 rows) and for sensitive detection (first column), expert scoring

(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.

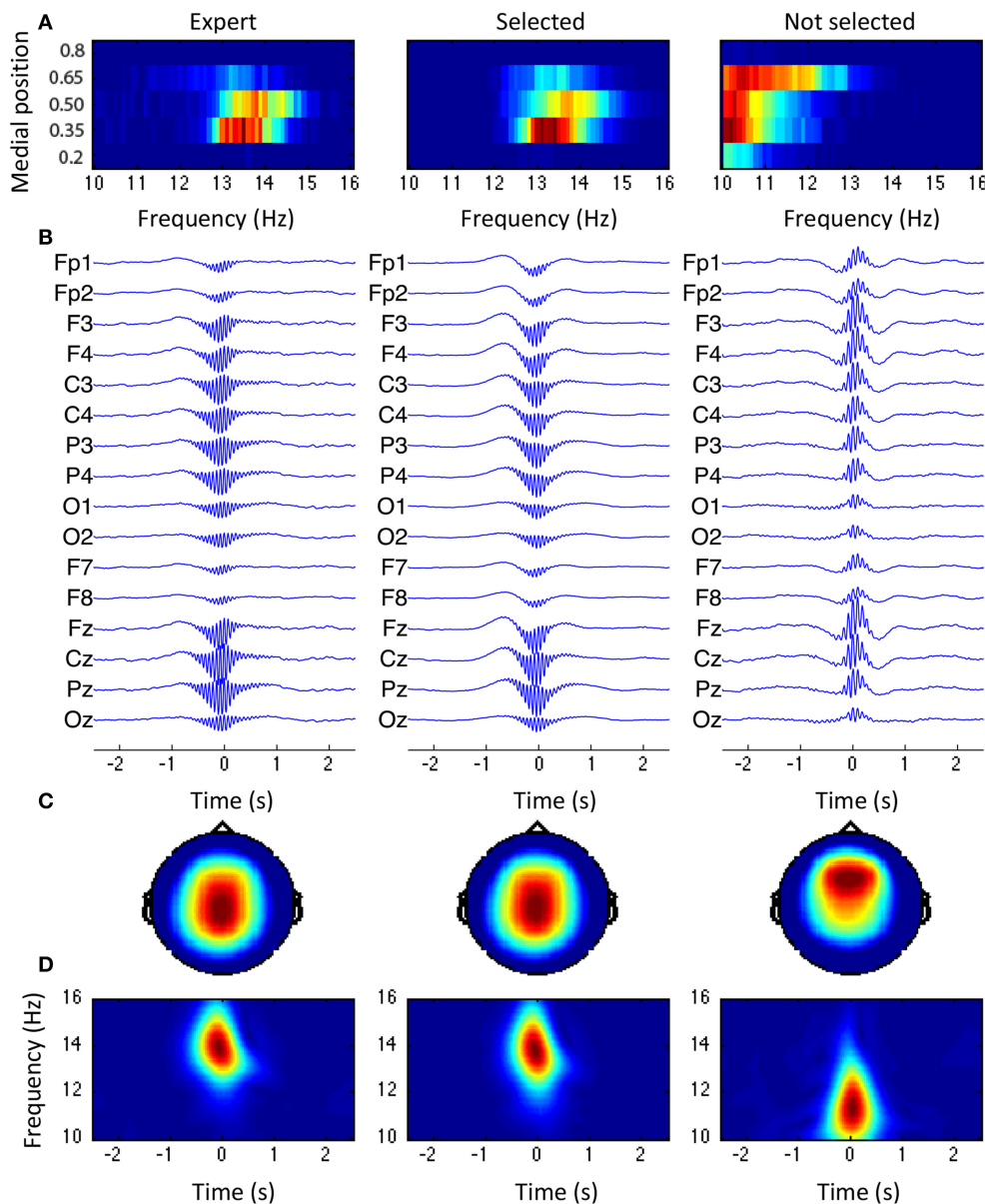


FIGURE 10 | Grand average across subjects. From left to right: spindle scored by the expert, events from the selected class, and events from the rejected classes. **(A)** Joint distribution of the frequency and the medial

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 expert. 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

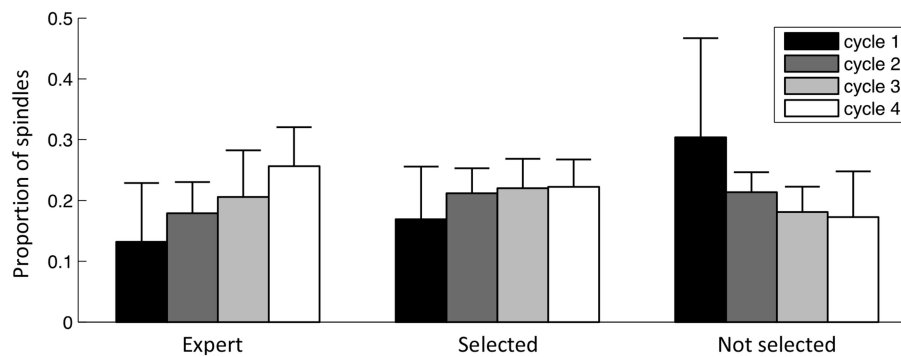


FIGURE 11 | Spindle distribution across sleep cycles. Bars indicate the mean value across subjects, whiskers indicate standard deviations. Values are normalized for subjects and classes (i.e., each of these three sets of bars sums to the unity).

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 SPINDLE 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

- Acir, N., and Güzelış, 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.tb00065.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., Miazhyńska, 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., Fabb, 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/smr.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., Escudé, B., Guillemain, 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
- Devuyst, S., Dutoit, T., Stenuit, P., and Kerkhofs, 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., Lafortune, 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
- Huopponen, 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
- Huopponen, 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., Lafortune, 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., Baletau, 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
- Schmicek, P., Zeitlhofer, 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?ptiID=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., Schmicek, 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.



Expert and crowd-sourced validation of an individualized sleep spindle detection method employing complex demodulation and individualized normalization

Laura B. Ray¹, Stéphane Sockeel², Melissa Soon^{1,3}, Arnaud Bore², Ayako Myhr¹, Bobby Stojanoski¹, Rhodri Cusack^{1,3}, Adrian M. Owen^{1,3}, Julien Doyon^{2,4} and Stuart M. Fogel^{1,2,3,4*}

OPEN ACCESS

Edited by:

Simon C. Warby,
Stanford University, USA

Reviewed by:

Ian M. Colrain,
SRI International, USA
Róbert Bódizs,
Sемmelweis University, Hungary
Roy Cox,
Beth Israel Deaconess Medical
Center/Harvard Medical School, USA

*Correspondence:

Stuart M. Fogel,
Brain and Mind Sleep Research
Laboratory, Brain and Mind Institute,
Western University, London,
ON N6A 5B7, Canada
sfogel@uwo.ca;
<http://www.bmisleeplab.uwo.ca>

Received: 28 January 2015

Accepted: 31 August 2015

Published: 24 September 2015

Citation:

Ray LB, Sockeel S, Soon M, Bore A, Myhr A, Stojanoski B, Cusack R, Owen AM, Doyon J and Fogel SM (2015) Expert and crowd-sourced validation of an individualized sleep spindle detection method employing complex demodulation and individualized normalization. *Front. Hum. Neurosci.* 9:507. doi: 10.3389/fnhum.2015.00507

¹ Brain and Mind Institute, Western University, London, ON, Canada, ² Functional Neuroimaging Unit, Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal, Montreal, QC, Canada, ³ Department of Psychology, Western University, London, ON, Canada, ⁴ Department of Psychology, University of Montreal, Montreal, QC, Canada

A spindle detection method was developed that: (1) extracts the signal of interest (i.e., spindle-related phasic changes in sigma) relative to ongoing “background” sigma activity using complex demodulation, (2) accounts for variations of spindle characteristics across the night, scalp derivations and between individuals, and (3) employs a minimum number of sometimes arbitrary, user-defined parameters. Complex demodulation was used to extract instantaneous power in the spindle band. To account for intra- and inter-individual differences, the signal was z-score transformed using a 60 s sliding window, per channel, over the course of the recording. Spindle events were detected with a z-score threshold corresponding to a low probability (e.g., 99th percentile). Spindle characteristics, such as amplitude, duration and oscillatory frequency, were derived for each individual spindle following detection, which permits spindles to be subsequently and flexibly categorized as slow or fast spindles from a single detection pass. Spindles were automatically detected in 15 young healthy subjects. Two experts manually identified spindles from C3 during Stage 2 sleep, from each recording; one employing conventional guidelines, and the other, identifying spindles with the aid of a sigma (11–16 Hz) filtered channel. These spindles were then compared between raters and to the automated detection to identify the presence of true positives, true negatives, false positives and false negatives. This method of automated spindle detection resolves or avoids many of the limitations that complicate automated spindle detection, and performs well compared to a group of non-experts, and importantly, has good external validity with respect to the extant literature in terms of the characteristics of automatically detected spindles.

Keywords: sleep, spindle, EEG, detection, automated, crowdsourcing

Introduction

Sleep spindles are brief (typically <1 s, up to 3 s) discrete phasic bursts of sigma (~ 11 –16 Hz) activity, with a waxing and waning amplitude envelope, which characterize non-rapid eye movement (NREM) sleep. Sleep spindles have garnered much interest in terms of their physiological mechanisms and cerebral correlates (Steriade, 2006; Schabus et al., 2007; Bonjean et al., 2011), putative function for sleep maintenance (Nicolas et al., 2001; Dang-Vu et al., 2010; Schabus et al., 2012), most recently in terms of their function for memory consolidation during sleep (Gais et al., 2002; Schabus et al., 2004; Fogel and Smith, 2006, 2011; Nishida and Walker, 2007; Bergmann et al., 2011), relationship to cognitive abilities (Smith et al., 2004; Bódizs et al., 2005, 2008; Fogel and Smith, 2006, 2011; Schabus et al., 2006; Fogel et al., 2007; Peters et al., 2007; Geiger et al., 2011; Ujma et al., 2014) and clinical relevance (Gibbs and Gibbs, 1962; Bixler and Rhodes, 1968; Shibagaki et al., 1982; Limoges et al., 2005; Steriade, 2005; Ferrarelli et al., 2007). Until recently, the study of the sleep spindle has been hindered by the labor-intensive task of visually identifying thousands of individual spindle events over the course of several hours of sleep and the resulting difficulty in obtaining expertly scored, publically available data sets for benchmarking. The investigation of sleep spindles has invigorated the proliferation of a variety of automated spindle detection methods (Broughton et al., 1978; Campbell et al., 1980; Zeitlhofer et al., 1997; Crowley et al., 2002; Mölle et al., 2002; Bódizs et al., 2009; Ray et al., 2010; Martin et al., 2012; Wamsley et al., 2012). However, the task of accurately detecting spindles has proven to be a significant methodological challenge. These challenges include, but are not limited to, the onerous task of analyzing lengthy, high temporal resolution recordings, and the high variability in signal-to-noise ratio over the course of the night, between derivations and individuals. Resolving these issues is complicated by the wide variety of methods being employed and incomplete or inconsistent validation procedures for these methods. This is further compounded by the absence of a “base truth” or appropriate and publically available “gold standard” to compare detection methods. Finally, validating automated detection methods by comparing their performance to human scorers may be insufficient as this assumes that: (1) human scorers are superior at detecting spindle events, and (2) automated detectors only perform correctly when functioning according to the narrow definition for visual identification of spindles. The absence of established method(s) could lead to erroneous scientific results or produce findings that are difficult to interpret and replicate.

Most commonly employed methods of spindle detection can be broadly classified into several categories based on the way that the signal of interest is extracted. These categories include: (1) methods that employ counting the number of peaks in a defined period of time, (2) band-pass filtering and root mean squared (RMS) transformations, (3) Fourier-based, and (4) wavelet-based techniques. In the following paragraphs, we compare and contrast some of the most commonly employed methods used to extract spindle-related activity for the purposes of automated

detection, highlighting some of the strengths and challenges of each.

Techniques that employ counting the number of peaks or zero crossings in a given time period (Principe and Smith, 1982; Schimicek et al., 1994; Zeitlhofer et al., 1997; Ray et al., 2010) may be advantageous to characterize spindle events once detected, however as a means of extracting spindle-related activity for the purposes of detection, these methods are susceptible to artifacts and can be contaminated by other naturally occurring EEG activity in other frequency bands of non-interest. As a result, the effectiveness of these techniques depend on how the EEG is preprocessed, thus making signal extraction relative to noise a challenge, nonetheless they are suitable for the extraction of the signal of interest. Similarly, band-pass filtering the signal to the sigma band and further RMS transformation (Clemens et al., 2007) does extract the signal of interest, and transforms the signal into all positive values, however, the oscillatory nature of the signal remains intact. This aids in characterizing spindle events, however, detection of the onset, peak and offset directly from an RMS transformed signal is no more straightforward than identifying events in the raw EEG signal, and thus the vastness of irregularities in the shape of the spindle, or changes in the frequency content and amplitude of each spindle over time, complicate detection and accurate identification of each spindle event. Moreover, deviation from the ideal frequency response of a band-pass filter (i.e., size of the transition band and related ripple effects) is a function of the window type and filter order. This is a potential challenge for slow spindles, whereby the adjacent frequencies, such as alpha activity (due to cortical arousals), may lead to false positives. In addition, given that when the sigma band is further divided into smaller and adjacent ~ 1.5 –2 Hz bands for slow (e.g., 11–13.5 Hz) and fast (e.g., 13.5–16 Hz) spindles, overlap between slow and fast spindle activity could lead to difficulty discriminating between spindle types. These issues could be overcome by employing filters with a sufficiently high filter order, and also, if spindles are first detected using the whole spindle bandwidth (e.g., 11–16 Hz) and each spindle is subsequently classified as slow or fast based on its peak (or mean) frequency following detection, spindles can be categorized orthogonally. These issues apply equally to other methods employing filters (including the current method).

Techniques that employ filtering and Fast Fourier Transform (FFT) techniques (Uchida et al., 1994; Huupponen et al., 2006) can be advantageous, however, the frequency resolution of FFT is determined by the sampling rate, window size and overlap. In addition, while FFT is well suited to handle signals that are linear and stationary, EEG is a dynamic, complex and noisy signal that originates from a combination of cortical and subcortical generators, whose relative contribution to scalp-recorded oscillations, in various mixed frequencies, changes dynamically over time. Thus, like many other biological signals, the EEG is a non-stationary and non-linear signal. Frequency extraction using Fourier-based methods can yield dramatically different results (Klonowski, 2007) as the signal evolves over time (i.e., time-domain information is lost). In relation to this caveat, Fourier-based methods are not necessarily optimal for extracting very brief phasic events, or to discriminate the activity of a phasic

event from the ongoing EEG. Thus, the ability of FFT to extract spindle-related activity is limited by selecting an appropriate window type, size and resulting frequency resolution, and may involve trial-and-error to select a multitude of appropriate model parameters, thus, care must be taken when utilizing FFT and similar techniques to extract spindle-related activity from the ongoing EEG.

By contrast, wavelet-based decomposition and other bandpass filtering techniques (Huuipponen et al., 2006; Wamsley et al., 2012) have the advantage of representing the signal in both time and frequency domains, and thus can be advantageous with respect to FFT, particularly for detecting brief events. However, wavelet-based approaches are computationally intensive and require a-priori assumptions about the signal of interest (e.g., spindles) in order to select the ideal “mother wavelet” (e.g., Meyer, Mortlet, or Mexican hat). Determining the wavelet type may involve many trial-and-error decisions in order to be optimized. As compared to other approaches, Wavelet-based techniques have been found to perform well as compared to FFT and RMS-based methods (Warby et al., 2014), however, they have been found to be susceptible to filter distortions (Ktonas et al., 2009), which could be problematic for brief events such as spindles.

The proposed method employed complex demodulation (CD; Walter, 1968) to extract the instantaneous power in a precise frequency band, and is desirable in that it does not make assumptions about the linearity or stationarity of the signal, and thus is well suited to detect events, such as sleep spindles in the EEG. CD has been shown to be an effective and flexible method to analyze real signals such as EEG, with less distortion (due to lowpass filtering) than Hilbert transformations, Wavelet decomposition, and matching pursuit (Ktonas et al., 2009). CD performs well compared to band-pass filtering, phase-locked loop demodulation, peak amplitude and zero-crossing detection (Ktonas and Papp, 1980). CD transforms the signal of interest in such a way that detection is straightforward (*n.b.*, yields a time series in the same temporal resolution as the original, with only positive data point values by demodulating the signal, in μV^2) and does not require any other a-priori decisions for signal extraction, other than the determination of the frequency band of interest, which for spindles is typically defined around 11–16 Hz (although it is important to note that there is considerable variability in the definition of the spindle band in the extant literature).

Over-and-above the challenges involved in signal extraction, considerable differences exist in terms of sleep spindles between individuals and over the course of the entire night, as well as within each NREM period (Silverstein and Levy, 1976; Werth et al., 1997; De Gennaro et al., 2000, 2005; Himanen et al., 2002; Ray et al., 2010). A commonly used approach to individualize detection amplitude thresholds is to use a detection threshold that is, for example, at the 95th percentile of the entire recording (Gais et al., 2002; Barakat et al., 2011; Nir et al., 2011; Cox et al., 2014). While this aids in overcoming the inter-individual differences in sleep spindles, it does not account for either the significant changes in spindle-related activity relative to the overall “background” sigma activity that evolves over the course

of individual NREM periods, or over the course of a whole night. In addition, spindles vary from one electrode site to another, and thus one amplitude threshold per subject may not be ideal for all derivations. Here, instead of adapting the detection threshold to the signal, or using multiple individualized thresholds, we have employed a sliding window that spans several epochs of NREM sleep (60 s; a period long enough to contain at least one spindle), and transforms each data point of the CD EEG into z-scores, based on the mean and standard deviation calculated from the centered 60-s window. The use of a sliding window allows for a single, fixed amplitude threshold, accounting for the changes in sigma activity that occur within each NREM cycle (Himanen et al., 2002), over the course of the entire night and across scalp derivations.

Finally, one of the major challenges of automated spindle detection, is the large number of aforementioned user-defined parameters, including but not limited to: (1) filter type, (2) window function (and related parameters, type, length, overlap, etc.), and (3) wavelet choice. Other user-defined parameters are often necessary to define attributes of the spindle including, but not limited to: (1) amplitude threshold, (2) frequency band, (3) minimum duration, (4) maximum duration, and (5) inter-spindle interval. Depending on the particular method, there can be a veritable infinite number of combinations of parameters to decide upon, prior to detection. While the current method is by no means parameter-free, an effort has been made to minimize the number of parameters and arbitrary decisions that are essential for maximum effectiveness and flexibility.

We automatically detected spindles on recordings obtained from the Montreal Archive of Sleep Studies (MASS; www.ceams-carsm.ca/en/MASS), an openly available database of overnight sleep recordings. Here, we compared automatically detected spindles to expert manual scoring using either conventional AASM guidelines, or with the visual aid of a sigma (11–16 Hz) band-pass filtered channel. We also compared expert manual scoring to the scoring of a group of non-experts using the aid of the sigma-filtered channel. And finally, we compared the automated detection to the non-experts to assess the utility of crowd-sourcing techniques to serve as an efficient means to develop a gold standard basis for comparison.

Here, we present a method for sleep spindle detection, inspired by algorithms first introduced in an analog system by Campbell et al. (Campbell et al., 1980; Hao et al., 1992; Ktonas et al., 2009). The method in the current study: (1) extracts the signal of interest (i.e., spindle-related phasic changes in sigma) relative to ongoing “background” sigma activity using CD; (2) accounts for intra-individual characteristics of sleep spindles (e.g., changes over the course of the night, and differences at various scalp locations) and the inter-individual differences in spindle characteristics; (3) utilizes as few, potentially arbitrary, user-defined parameters as possible (e.g., to avoid a multitude of signal extraction/model parameters, amplitude thresholds, minimum/maximum cut-offs, etc.), (4) compares the performance of three different visual detection approaches to one another and each visual detection method to the automated detection, and finally, (5) validates this method by comparing to established characteristics of spindles: (i) during

Stage 2 sleep (NREM2) and slow wave sleep (SWS), (ii) at frontal and parietal derivations, (iii) for fast and slow spindle types, and (iv) across consecutive NREM sleep cycles. The current method provides an alternative approach intended to address (or circumvent) the major above-mentioned challenges for accurate, automated spindle detection using a relatively straightforward approach.

Methods

Participants and EEG Data Set

PSG recordings (including sleep stage scoring annotations) were obtained from the publically available (upon request) Montreal Archive of Sleep Studies (MASS; www.ceams-carsm.ca/en/MASS) from the SS2 database (O'Reilly et al., 2014) and included recordings from 19 subjects (11 female) with a mean age of 23.6 years. Overnight PSG data were acquired on a Grass Model 12 amplifier using Harmonie acquisition software (V5.4, Natus Medical Inc., San Carlos, USA) from 21 EEG channels (Fp1, Fpz, Fp2, F7, F3, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2, A1, A2). EEG was recorded at 256 samples/s using -6 dB filters, 0.4 s time constant, low cutoff filter at 0.3 Hz, and computed linked reference from A1 to A2. Sleep stages were scored according to Rechtschaffen and Kales (1968) in 20 s epochs (Table 1). PSG records and sleep stage annotations were converted from EDF+ to EEGLab format using in-house file conversion software written for Matlab (R2014a, Mathworks, Matick, MA, USA).

All subjects had a Beck Depression Score <13 and did not report any history of mental disorders. Subjects did not take antidepressant medications and were not currently (or within the last 10 years) diagnosed with major mental illness or personality disorder. Upon visual inspection of the data, four subjects were excluded from analyses, two for excessive alpha intrusion (01-02-0004, 01-02-0016), one for frequent EEG arousals indicative of a sleep disorder (01-02-0008) and one due to intermittent poor quality EEG for one of the channels of interest (01-02-0015). Ethical approval to use the MASS SS2 PSG and sleep scoring annotations was obtained by the local Ethics Review Board at Western University, London, Ontario, Canada.

Expert Manual Spindle Scoring

Two experts from different sites (Expert 1: London, Ontario; Expert 2: Montreal, Quebec) manually scored spindles from C3 in NREM2 displayed in 20-s epochs, for the entire recording in all

15 subjects included in the study. These annotations are available from the MASS database. The visual identification method employed by each expert differed with the exception that: Expert 1 visually identified and manually marked the beginning and end of each spindle from a duplicate C3 channel, filtered to the sigma band (11–16 Hz), and did not use any explicit minimum duration criteria. This visualization technique is used to help identify spindles that would otherwise be obscured by slow wave activity (e.g., by k-complexes, delta waves), and to identify spindles that have a short duration and small amplitude. This method allows the Expert scorer to visualize activity in a way that is closer to how many spindle detection algorithms “see” the EEG, with the intention that this may improve the accuracy of manual detection and make for a more valid comparison to automated detection methods. Otherwise, spindle scoring conformed to AASM guidelines (Iber, 2007). On the other hand, Expert 2 adhered to AASM guidelines, did not score using the duplicate, filtered channel, and scored spindles greater than 0.5 s in duration. The spindle duration, amplitude and frequency of each spindle event were calculated in the same way as the automated detection (see Section Automated Spindle Detection, below).

Non-expert Manual Spindle Scoring

Sleep spindles were also manually identified by a group of non-expert scorers using Amazon's web-based crowd sourcing platform (Amazon Mechanical Turk: <https://www.mturk.com/mturk/>) in order to collect spindle scoring from a large sample of non-experts (see Supplementary Figures 1–5). Two recordings were not included (01-02-0018 and 01-02-0019) in the non-expert scoring data set described above (see Section Participants and EEG Data Set), as a result of changes to Amazon's terms and conditions mid-way through data collection. This policy change restricted use of the Mechanical Turk payment service to residents of the United States, preventing data collection to be completed. The remaining data from the 13 EEG recordings (199,860 s of data from NREM2) sleep were divided into segments of about 2000 s. This was done in order to provide small, manageable amounts of data to be manually scored by the non-experts, for which they were compensated for their time. There was no limit on how many segments each individual non-expert could score from the dataset, but the same non-expert was permitted to score the same segment only once. A total of 406 unique non-experts contributed to the manual spindle scoring by marking at least one 2000 s segment. On average, 18.4 ($SD = 1.2$, range 15–20) non-experts scored each ~2000 s segment of data. Similar to the method used by Expert 1, the interface itself (Supplementary Figure 1) displayed EEG in 20 s epochs for the sigma (11–16 Hz) filtered C3 channel. This was done in order to simplify the task of identifying spindles for non-experts, to reduce ambiguity and to simplify and minimize the amount of training required (Supplementary Figure 2). One advantage of using the sigma-filtered channel was that non-experts were not required to learn anything about EEG and very little about sleep spindles *per se* (Supplementary Figure 3). Rather, they were trained by exemplars on a de-noised signal, making event identification more straightforward than spindles embedded in ongoing EEG in NREM2. Non-experts were required to become

TABLE 1 | Sleep architecture results (M ± SD).

Sleep stage	Duration (minutes)	Duration (% TST)
Wake	60.93 ± 44.21	
NREM1	39.27 ± 21.19	6.3 ± 3.40
NREM2	360.03 ± 46.37	55.7 ± 6.17
SWS	114.27 ± 41.17	17.4 ± 5.43
REM	133.03 ± 23.02	20.6 ± 3.74
Total	646.60 ± 55.98	

familiarized with a simple set of 3 tools in order to use the web-based interface (Supplementary Figure 4). These tools allowed them to navigate from one epoch to another (Supplementary Figure 5, #1), highlight spindles (Supplementary Figure 5, #2) and to indicate when there were no spindles present on the epoch (Supplementary Figure 5, #3).

Automated Spindle Detection

EEG processing was carried out using EEGLab (V13) and Matlab (R2014a) (**Figure 1**) on the same data set (see Section Participants and EEG Data Set) using the same EEG channel (C3) as the expert and non-expert scorers. Thus, the validation between automated detection and visual raters is limited to NREM2 sleep from a single central (C3) derivation. Spindles were also detected from additional channels at frontal (F3) and parietal (P3) sites in both NREM2 and SWS across the first four NREM cycles to further explore the characteristics of the automatically detected spindles, in order to provide additional validation of known topographic distribution (Werth et al., 1997; Zeitlhofer et al., 1997), temporal patterns (Werth et al., 1997; De Gennaro et al., 2000) and the characteristics (Bódizs et al., 2009) of spindles. Prior to detection, the EEG was low-pass filtered at 35 Hz. Movement artifact was detected from the EMG channel (highpass filtered at 10 Hz) when the second order derivative of the signal exceeded $20 \mu\text{V/ms}$. The EEG was marked as “bad data” $\pm 3\text{ s}$ about the detected movement.

CD was employed on the normally filtered (0.3–35 Hz) EEG, to extract the instantaneous power (in μV^2) about the frequency of interest (13.5 Hz), while eliminating all other frequencies

outside the spectrum of interest (11–16 Hz). CD is carried out in two principle steps on the original data ($X(t)$), that is taken to be the signal of interest, plus everything else ($Z(t)$). Amplitude (A) and phase (P) vary with respect to the carrier frequency (ω), defined mathematically as:

$$X(t) = A(t) \cos(\omega t + P(t)) + Z(t)$$

In the first step of the CD, the frequency spectrum of interest, about a carrier frequency (in this case, 13.5 Hz), is shifted left by the demodulating frequency, toward the origin (i.e., zero frequency) by multiplying $X(t)$ by $\exp\{-i\omega t\}$ according to the method originally described by Walter (1968):

$$Y(t) = X(t) \exp\{-i\omega t\}$$

This can also be written as its analytical analog, as follows, which reveals 3 terms (a, b, c):

$$Y(t) = \frac{A(t)}{2} \exp\{iP(t)\} \quad (\text{a})$$

$$+ \frac{A(t)}{2} \exp\{-i(2\omega t + P(t))\} \quad (\text{b})$$

$$+ Z(t) \exp\{-i\omega t\} \quad (\text{c})$$

The result $Y(t)$ contains the shifted component at 0 Hz (term a), and a second component that varies at twice the shifted carrier frequency 2ω (term b), plus all other frequency components (term c). In the second step, the signal is low pass filtered (infinite impulse response, 4th order butterworth filter, using “filtfilt” from

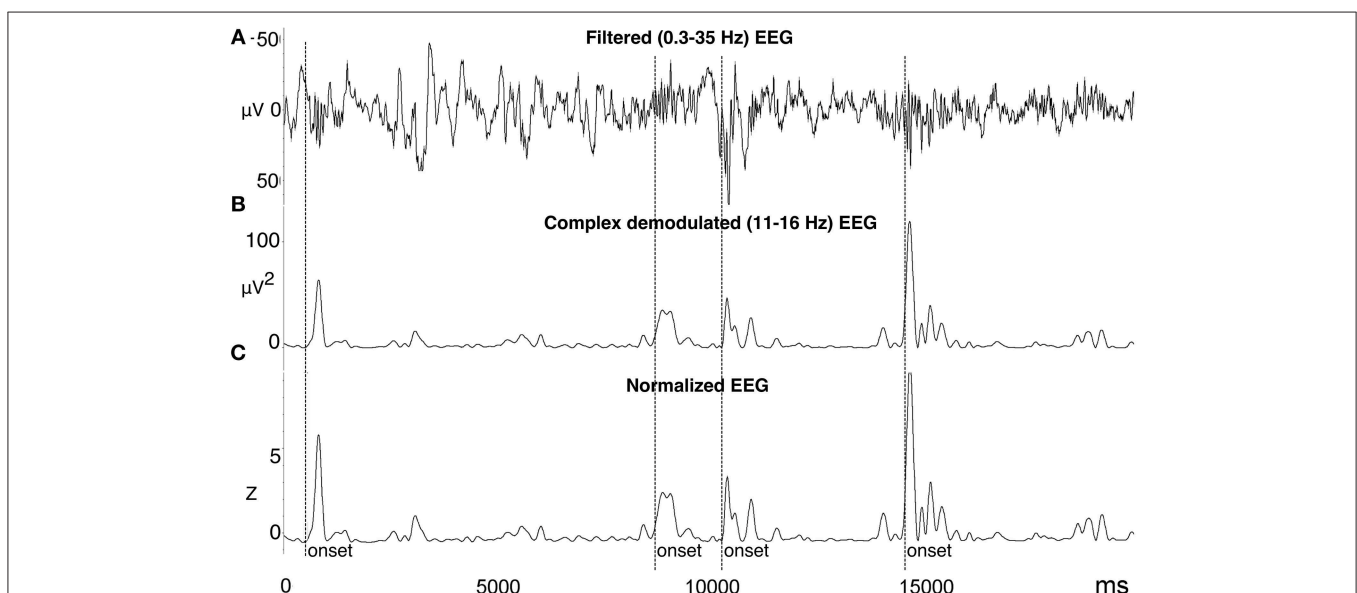


FIGURE 1 | Automated spindle detection method processing steps. (A) Step 1, the EEG was filtered using a high pass 0.3 Hz filter, low pass 35 Hz filter, and bad data and artifact was identified. **(B)** Step 2, the EEG was transformed using complex demodulation (CD), producing a new time series of instantaneous magnitude (μV^2) in the frequency band of interest (e.g., 11–16 Hz). **(C)** Step 3, the CD time series was normalized to Z-scores calculated from a 60-s sliding window about each data point. Spindle onsets were detected when $Z > 2.33$ (i.e., 99th percentile). To more accurately measure the entire length of the spindle, the onset was adjusted to be the first point at which $Z = 0.5$ prior to the amplitude threshold Z, and the offset as the first point at which $Z = 0.5$ after the amplitude threshold Z. Figure reproduced from Fogel et al. (2014b).

Matlab, to avoid phase shifts) so that the first term is preserved, and the frequency content of the complex signal outside the frequency band of interest may be considered negligible (Ktonas et al., 2009). Filtering removes the unwanted 2nd (b) and 3rd (c) terms and smoothes the resulting signal (with a length of $2T - 1$, where $T = 2\pi/\omega$ is the demodulation period), to retain the demodulated and smoothed amplitude time series, where prime indicates smoothed:

$$Y'(t) = \frac{1}{2}A'(t) \exp\{iP'(t)\}$$

Following the CD transformation, the present method transforms the data from each channel, by normalizing the signal using a z-score transformation derived from a centered 60-s sliding window. This is similar to other methods that employ an individualized amplitude threshold (Gais et al., 2002; Barakat et al., 2011; Nir et al., 2011; Cox et al., 2014), calculated from a percentile score of the whole recording (e.g., 95%), except that instead of adapting the detection threshold on a per-subject basis, here, the signal is transformed so that a single threshold can be applied to all subjects, at all scalp derivations, across the entire recording that accounts for the variation of spindle-related activity to ongoing sigma over time.

To detect spindle events, an amplitude threshold corresponding to the 99th percentile ($Z = 2.33$) was used. Events occurring during “bad data” and outside NREM sleep were subsequently removed. Finally, the onset and offset of the spindle event is determined to be when the amplitude approaches zero, in this case, $Z = 0.5$ and the duration (offset-onset, in seconds) encompassing the whole spindle event can then be calculated. Spindle event markers (onset and offset) were then moved to the EEG prior to demodulation, filtered from 11 to 16 Hz so that the mean frequency (peak-to-peak mean distance, in Hz) and peak amplitude (max peak-to-peak value, in μV) could be calculated in the same units as the original EEG signal. For the purposes of further characterizing the automatically detected spindles at frontal (F3) and parietal (P3) sites in NREM2 and SWS (see Section Characteristics of Automatically Detected Spindles), each individual spindle event was categorized and binned into either slow (11–13.5 Hz) or fast (13.5–16 Hz) spindles based on the mean frequency of each spindle event. Further, to investigate the changes in spindle characteristics over the course of the night, spindles were binned into the first four NREM cycles. NREM cycles were defined as periods of consolidated NREM sleep comprising at least 15 consecutive minutes (forty-five 20 s epochs) of NREM sleep separated by consolidated REM sleep comprising at least 2 consecutive minutes (six 20 s epochs) of REM sleep.

Inter-rater Reliability

The inter-rater agreement between methods (either between visual scoring methods, or automated detection vs. Experts, or compared to non-experts) was tested using a method adapted from Ray et al. (2010). Three second epochs were used to identify the presence or absence of spindles to count true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). This was done so that TN could be easily quantified

in some meaningful way. Consensus between non-experts was simply calculated as the proportion of non-expert scorers that identified a spindle at the same point in time. For non-expert comparisons, this was carried out at 10 different levels of consensus among raters, ranging from 0.1 to 0.9 (Supplementary Figure 6). Statistics were calculated for the level of consensus where the mean F1 score (the harmonic mean of recall and precision, a composite score that represents a single measure of inter-rater agreement) was maximal.

More precisely, in the case where there was an overlap between spindles scored by one scorer and the other (expert, automatic or non-expert), the 3-s epoch was counted as TP, otherwise it was counted as FN. In the case where the other scorer scored a spindle, and there was no overlap with an event, the 3-s epoch where the “spindle” occurred was counted as FP. In the case where there was no spindle scored from either scorer, this 3-s epoch was counted as TN. Each comparison could only be made once.

Spindles are sparsely distributed throughout the total duration of NREM2. This leads to a disproportionate number of TN results, which can inflate sensitivity. The 3 s windows were used to judge inter-rater agreement in order to minimize this, however, it does not completely eliminate the issue. Thus, the recall ($TP/(TP + FN)$) and precision ($TP/(TP + FP)$) were used in addition to the conventional measures of agreement that can be biased by a high proportion of TN (e.g., specificity, negative predictive value (NPV) and false positive rate). Despite the employment of a relatively large 3 s window to judge the inter-rater agreement, there were still a disproportionate number of TN judgments (Table 2). Thus, the F1 scores and the phi (Φ) coefficient (another balanced single measure that is appropriate when classes are of different sizes, where 1 represents perfect agreement and -1 represents complete disagreement between judges) were also reported. The statistical significance of Φ can also be determined. Importantly, the F1 score and phi coefficient are advantageous in that they are unbiased by the direction of the comparison between judges.

Results

Inter-rater Agreement for Visual Identification of Spindles

Expert 1 vs. Expert 2

Overall, Expert 1 had a high mean proportion of correctly identified events relative to the total number of events identified by Expert 2 (i.e., precision = 0.85, ± 0.21), but Expert 2 had a low mean proportion of spindles that were correctly identified relative to the total number of events scored by Expert 1 (i.e., recall = 0.40, ± 0.14). There was a very high proportion of periods without spindles that were correctly identified by Expert 2 as compared to Expert 1 (i.e., specificity = 0.97, ± 0.04) and a high proportion of 3 s periods of EEG without spindles identified by Expert 2 (NPV = 0.80, ± 0.07), with a false positive rate of only 0.03, ± 0.04 . When recall and precision are both maximal (i.e., equal to 1), this represents perfect performance, and when recall and precision are plotted against one another (Figure 2A), data

TABLE 2 | Group mean percent and marginal totals (\pm SD) of true positive, false positive, true negative and false positive epochs comparing expert vs. expert, expert vs. non-expert and expert vs. automatically detected spindles.

	Positive	Negative	Total
EXPERT 1 vs. EXPERT 2 DETECTIONS			
True	11.48%	69.22%	3882 \pm 451.83
False	1.97%	17.34%	928 \pm 194.45
Total	647 \pm 232.70	4164 \pm 413.58	4811 \pm 323.14
EXPERT 1 vs. NON-EXPERT DETECTIONS			
True	25.36%	62.58%	3552.93 \pm 325.84
False	8.27%	3.79%	487.23 \pm 178.64
Total	1358.54 \pm 284.08	2681.62 \pm 220.40	4040.16 \pm 252.24
EXPERT 2 vs. NON-EXPERT DETECTIONS			
True	9.76%	78.89%	3568.15 \pm 232.16
False	7.67%	3.68%	456.77 \pm 149.07
Total	701.31 \pm 172.87	3323.61 \pm 208.37	4024.92 \pm 190.62
EXPERT 1 vs. AUTOMATED DETECTIONS			
True	19.94%	63.66%	4024.73 \pm 380.99
False	7.48%	8.91%	789.14 \pm 230.50
Total	1320.27 \pm 231.83	3493.60 \pm 379.66	4813.87 \pm 305.74
EXPERT 2 vs. AUTOMATED DETECTIONS			
True	10.10%	68.77%	3795.13 \pm 376.78
False	17.83%	3.30%	1016.74 \pm 176.37
Total	1343.67 \pm 242.89	3468.20 \pm 310.25	4811.87 \pm 276.57
NON-EXPERT vs. AUTOMATED DETECTIONS			
True	53.69%	13.58%	1142.68 \pm 326.11
False	26.08%	6.65%	555.92 \pm 196.02
Total	1354.92 \pm 211.88	343.68 \pm 310.25	1698.60 \pm 261.07

points crowd the upper-right hand corner. However, as shown in **Figure 2A**, data points were dispersed along the left hand side of the plot, which resulted in low F1 scores (**Figure 2B**; mean $F1 = 0.54, \pm 0.17$), and a low and non-statistically significant phi coefficient ($\Phi = 0.49, \pm 0.18, p > 0.05$).

Expert 1 vs. Non-expert Consensus

Overall, and consistent with a previous report (Warby et al., 2014), Expert 1 and the consensus of non-experts performed with very high agreement (**Figures 2A,B**). The non-expert detection of spindles had both a high proportion of spindles that were correctly identified relative to the total number of expert events (i.e., recall = 0.87, ± 0.08) and a high proportion of correctly identified events relative to the total number of spindles detected by the group of non-experts (i.e., precision = 0.75, ± 0.13). There was also a very high proportion of actual periods without spindles that were correctly identified by non-experts (i.e., specificity = 0.88, ± 0.07) and a high proportion of correctly identified 3 s periods of EEG without spindles identified by non-experts (NPV = 0.94, ± 0.05), with a false positive rate of only 0.12, ± 0.07 . Finally, the F1 scores were high ($F1 = 0.81 \pm 0.07$, **Figure 2B**) with points crowding the upper-right hand corner of the recall-precision plot (**Figure 2A**), and the phi coefficients [mean $\Phi = 0.72, \pm 0.07, \chi^2_{(1)} = 6.82, p < 0.001$] were high, and

statistically significant, suggesting excellent overall agreement between Expert 1 and the consensus of non-experts.

Expert 2 vs. Non-expert Consensus

In contrast to the comparison to Expert 1, the non-experts correctly identified fewer spindles relative to the total number of Expert 2 events (i.e., recall = 0.73, ± 0.20) and a lower proportion of correctly identified events relative to the total number of spindles detected by the group of non-experts (i.e., precision = 0.56, ± 0.18) with agreement also being more variable across recordings (**Figure 2A**). There was a very high proportion of actual periods without spindles that were correctly identified by non-experts (i.e., specificity = 0.91, ± 0.05) and a high proportion of correctly identified 3 s periods of EEG without spindles identified by non-experts (NPV = 0.96, ± 0.05), with a false positive rate of only 0.09 ± 0.05 . However, when considering measures unbiased by TN events, the F1 scores were on average lower (mean $F1 = 0.63 \pm 0.16$) although the phi coefficient did reach statistical significance [mean $\Phi = 0.57, \pm 0.19, \chi^2_{(1)} = 4.27, p = 0.039$].

Characteristics of Visually Identified Spindles

The most apparent differences in the characteristics of spindles identified by the various visual scoring approaches were for spindle duration and amplitude. In general, Expert 1 and non-experts identified spindles with very similar distributions of durations (Cohen's $d = 0.14$) ranging from about 0.2–3 s in length (**Figure 3A**), whereas Expert 2 identified spindles in a more restricted range between about 0.5 and 2 s in length (**Figure 3A**), whose distribution overlapped less with Expert 1 (Cohen's $d = 0.85$) and the consensus of the non-experts (Cohen's $d = 0.63$). A similar pattern was observed for amplitude whereby Expert 1 tended to score more spindles with smaller amplitudes (**Figure 4A**) than Expert 2 (Cohen's $d = 0.63$), with the distribution of non-expert spindle amplitudes overlapping to a greater extent with Expert 1 (Cohen's $d = 0.2$) than Expert 2 (Cohen's $d = 0.37$), respectively (**Figure 4A**). By contrast, there was considerable overlap between visual scoring approaches for mean frequency (**Figure 5A**) between Experts 1 and 2 (Cohen's $d = 0.08$), Expert 1 and non-experts (Cohen's $d = 0.16$) and between Expert 2 and non-experts (Cohen's $d = 0.23$). In terms of mean frequency, however, from inspection of **Figure 5A**, it appears that non-experts tended to identify more spindles with a slower frequency than either Expert 1 or 2, perhaps due to mistakenly identifying brief arousals (i.e., alpha activity) as spindles.

Expert 1 vs. Expert 2

Spindles scored by Expert 1 and Expert 2 (**Table 3**) differed significantly in terms of spindle duration [$t_{(14)} = 13.42, p < 0.001$], amplitude [$t_{(14)} = 2.76, p = 0.015$], total number [$t_{(14)} = 5.26, p < 0.001$], but not mean frequency ($p > 0.7$). Despite these differences, the characteristics of the spindles identified by the two experts were linearly related to one another; suggesting that the experts systematically (and consistently) identified spindles differently across recordings on average, for duration [**Figure 3B**, $r_{(13)} = 0.69, p = 0.004$], amplitude [**Figure 4B**, $r_{(13)} = 0.96$,

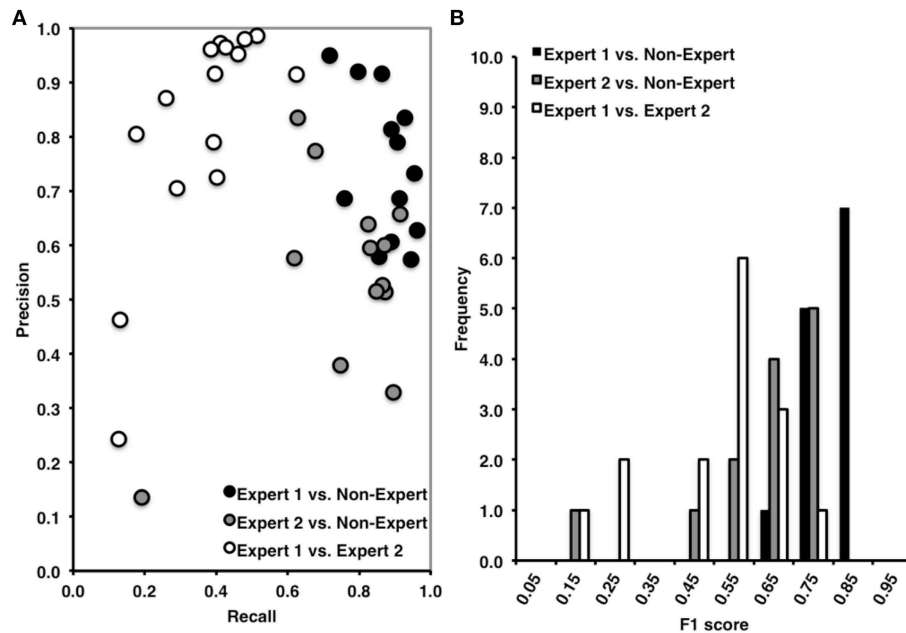


FIGURE 2 | (A) High precision and recall across recordings when comparing Expert 1 to non-expert spindle scoring (black), low recall and variable precision across recordings when comparing Expert 1 to Expert 2 (open), and intermediate precision and recall between Expert 2 and non-experts (gray). **(B)** Inter-rater agreement was consistently high across subjects for Expert 1 vs. non-expert detections, ranging from 0.60 to 0.90 (Mean $F1 = 0.81, \pm 0.07$), low and variable agreement between Expert 1 and Expert 2 ranging from 0.10 to 0.80 (Mean $F1 = 0.54, \pm 0.17$), and intermediate and variable agreement between Expert 2 and non-experts ranging from 0.10 to 0.80 (mean $F1 = 0.63, \pm 0.16$). F1 score = harmonic mean of recall and precision.

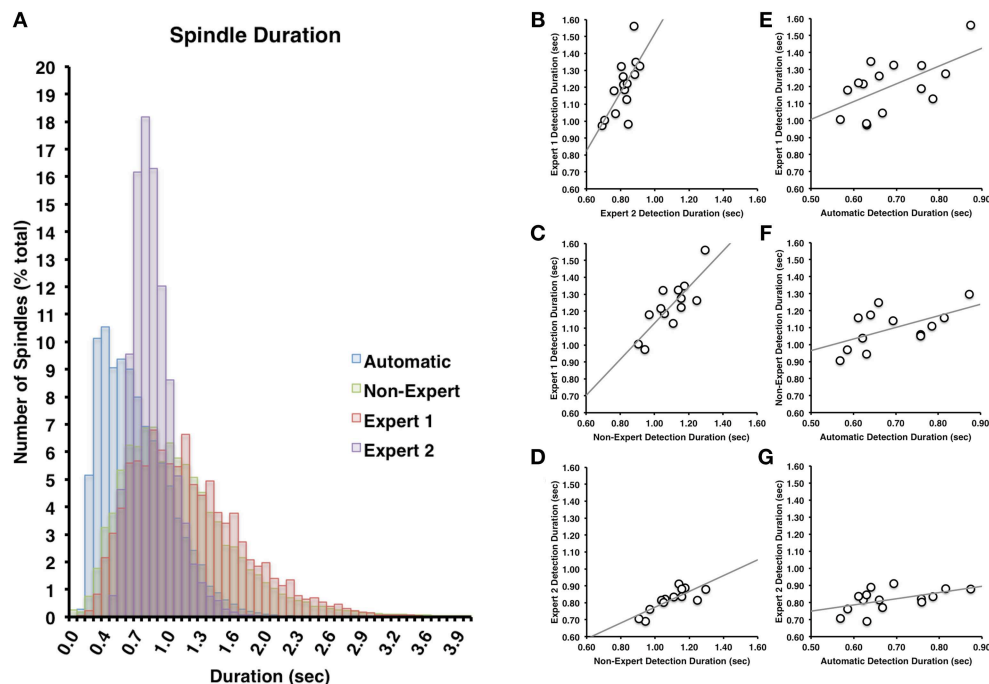


FIGURE 3 | (A) There was a great deal of overlap between Expert 1 and non-experts in terms of spindle duration (Cohen's $d = 0.14$), but less overlap with Expert 2 (Cohen's $d = 0.85$) or between Expert 2 and non-experts (Cohen's $d = 0.63$). Spindle duration of automatically detected spindles were generally shorter in duration than Expert 1 (Cohen's $d = 1.12$), Expert 2 (Cohen's $d = 0.57$), or non-experts (Cohen's $d = 0.91$). Spindle duration among visual identification methods **(B–D)** and between automatic and visual detection **(E–G)** were all highly inter-correlated (all $p < 0.05$).

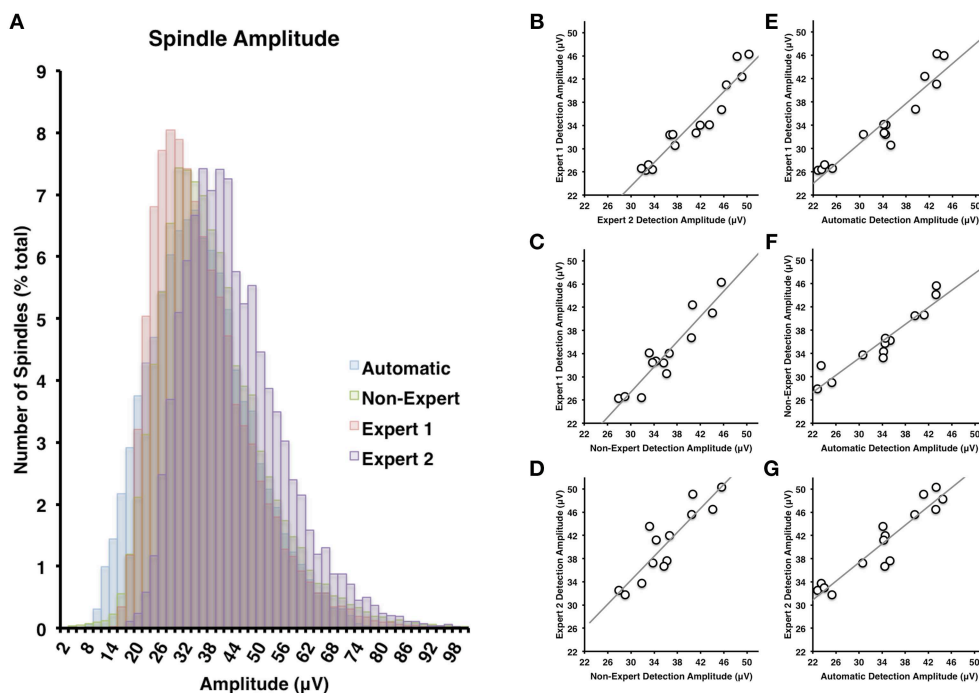


FIGURE 4 | (A) There was the greatest deal of overlap between non-experts and the automatically detected spindles in terms of spindle amplitude (Cohen's $d = 0.24$), but less overlap between the non-experts and Expert 1 (Cohen's $d = 0.85$) or Expert 2 (Cohen's $d = 0.63$). Spindle amplitudes of automatically detected spindles were generally smaller than Expert 2 (Cohen's $d = 0.68$) and overlapped the most with Expert 1 (Cohen's $d = 0.05$). Spindle amplitude among visual identification methods (**B–D**) and between automatic and visual detection (**E–G**) were all highly inter-correlated (all $p < 0.05$).

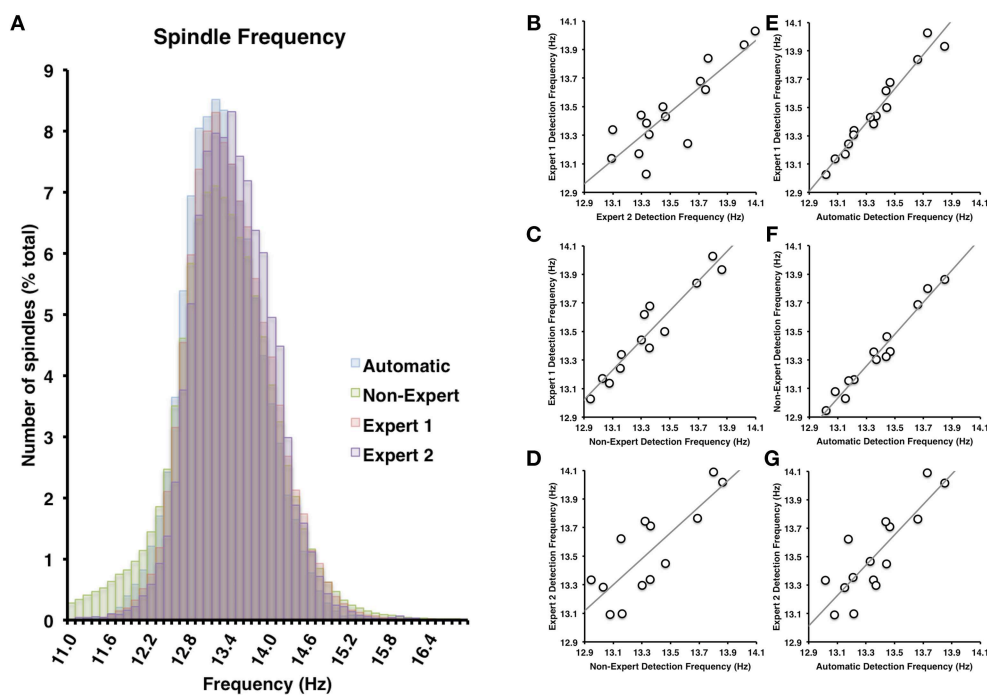


FIGURE 5 | (A) There was nearly complete overlap between the four scoring methods employed (all Cohen's $d < 0.25$). Spindle frequency among visual identification methods (**B–D**) and between automatic and visual detection (**E–G**) were all very highly inter-correlated (all $p < 0.05$).

TABLE 3 | Group mean (\pm standard deviation) spindle characteristics for automatically and manually detected spindles by experts and a group of non-experts.

Detection Method	Duration (s)	Frequency (Hz)	Amplitude (μ V)	Number
Expert 1	1.20 \pm 0.16	13.47 \pm 0.30	34.34 \pm 6.84	1422 \pm 410.62
Expert 2	0.82 \pm 0.06*,#	13.51 \pm 0.30	40.60 \pm 6.44*	772.73 \pm 386.35*,#
Non-experts	1.10 \pm 0.11	13.35 \pm 0.28	36.12 \pm 5.17	1140 \pm 339.53
Automatic	0.69 \pm 0.09*,+,#	13.37 \pm 0.24	34.06 \pm 7.52+	1438 \pm 240.14*,#

*Indicates significant difference from Expert 1, + indicates significant difference from Expert 2, and # indicates significant difference from Non-experts, $p < 0.05$, two-tailed t-test. Mean values for number reported for non-experts.

$p < 0.001$] mean frequency [Figure 5B, $r_{(13)} = 0.86$, $p < 0.001$] and number [$r_{(13)} = 0.81$, $p < 0.001$]. Taken together, this suggests that Experts 1 and 2 identified spindles with different characteristics, and did so systematically across recordings.

Expert 1 vs. Non-expert Consensus

By contrast, there were no significant differences (Table 3) in the characteristics of the spindles identified by Expert 1 as compared to the consensus of the non-experts in terms of spindle duration ($p > 0.05$), mean frequency ($p > 0.2$), amplitude ($p > 0.4$) or total number identified ($p > 0.06$). Given these similarities, it is not surprising that there was also a very high correlation for duration [Figure 3C, $r_{(13)} = 0.82$, $p < 0.001$], amplitude [Figure 4C, $r_{(13)} = 0.93$, $p < 0.001$] mean frequency [Figure 5C, $r_{(13)} = 0.96$, $p < 0.001$] and number [$r_{(13)} = 0.71$, $p = 0.003$] across subjects. Thus, suggesting that Expert 1 and non-experts identified spindles with similar characteristics and did so consistently across recordings.

Expert 2 vs. Non-expert Consensus

By contrast, the characteristics of the spindles identified by Expert 2 differed significantly from non-experts (Table 3) in terms of spindle duration [$t_{(14)} = 11.47$, $p < 0.001$] and total number [$t_{(14)} = 2.83$, $p = 0.013$], but not frequency ($p > 0.1$) and amplitude ($p > 0.05$). Despite the differences in spindle characteristics between Expert 2 and non-experts, there was a significant linear relationship for the spindle characteristics between Expert 2 and non-experts for duration [Figure 3D, $r_{(13)} = 0.80$, $p < 0.001$], amplitude [Figure 4D, $r_{(13)} = 0.89$, $p < 0.001$], mean frequency [Figure 5D, $r_{(13)} = 0.81$, $p < 0.001$] and total number [$r_{(13)} = 0.83$, $p < 0.001$]. Thus, similar to the comparison between Expert 1 and Expert 2, in general, Expert 2 identified spindles with different characteristics than non-experts and did so in a consistent manner across recordings.

Automated Detection vs. Visual Scoring

Automated Detection vs. Expert 1

The automated detection method had both a high proportion of spindles that were correctly identified relative to the total number of events identified by Expert 1 (i.e., recall = 0.69, ± 0.11) and a high and balanced proportion (with respect to recall) of correctly identified events relative to the total number of automatically detected events (i.e., precision = 0.73, ± 0.15) (Figure 6A).

As expected, there was a high proportion of actual periods without spindles that were correctly identified (i.e., specificity = 0.89, ± 0.05) and a high proportion of correctly identified 3 s periods of EEG without spindles (NPV = 0.88, ± 0.08), with a false positive rate of only 0.11, ± 0.05 . Overall, we observed high agreement between the automated and manual detection by Expert 1 [F1 = 0.71, ± 0.06 and $\Phi = 0.60$, ± 0.06 , $\chi^2_{(1)} = 5.31$, $p = 0.021$; Figure 6B].

Automated Detection vs. Expert 2

By contrast, while the automated detection identified a high number of spindles relative to the total number of events identified by Expert 2 (recall = 0.75, ± 0.23) there was a low number of correctly identified events relative to the number of automatically detected events (precision = 0.36, ± 0.17) (Figure 6A). Specificity (0.79, ± 0.04) and negative predictive value (0.95, ± 0.04) were also high, with a low false positive rate (0.21, ± 0.04), however these metrics are likely inflated by the high number of TN. When taken into consideration, the F1 scores (F1 = 0.49, ± 0.04) and phi coefficient ($\Phi = 0.42$, ± 0.20 , $p > 0.05$) were low and non-statistically significant. Thus, suggesting that the automated detection method also detected the majority of spindles identified by Expert 2, but made additional detections that Expert 2 did not.

Automated Detection vs. Non-expert Consensus

Similar to Expert 1, the automated detection method performed comparatively as well or better as compared to the consensus of non-experts (Figure 6A), as indicated by high recall = 0.80, ± 0.11 , precision = 0.67, ± 0.10 , specificity = 0.85, ± 0.08 , negative predictive value = 0.92, ± 0.03 and a low false positive rate = 0.15, ± 0.08 . The F1 scores (Figure 6B) were also consistently high F1 = 0.73, ± 0.04 , as was the phi coefficient [$\Phi = 0.62$, ± 0.07 , $\chi^2_{(1)} = 5.00$, $p = 0.025$]. In summary, the automated detection method performed well as compared to Expert 1 and the consensus of non-experts, but with less agreement and consistency as compared to Expert 2.

Characteristics of Automatically Detected Spindles

Characteristics of Automatically Detected Spindles vs. Visually Detected Spindles

The automated detection method identified spindles that were smaller both in terms of duration (Table 3 and Figure 3A) as compared to Expert 1 [$t_{(14)} = 19.45$, $p < 0.001$], Expert 2 [$t_{(14)} = 5.41$, $p < 0.001$] and non-experts [$t_{(14)} = 17.25$, $p < 0.001$], supporting the notion that even with the use of a highly filtered channel to simplify and aid in the visual identification of sleep spindles, automated methods are able to identify and measure smaller spindles. Expert 2 identified spindles that were also larger in terms of amplitude [$t_{(14)} = 2.73$, $p = 0.016$], whereas Expert 1 ($p > 0.9$) and non-experts ($p > 0.4$) identified spindles of the same amplitude as the automated detection (Table 3). Spindle frequency did not differ from visual scoring (all $p > 0.1$). Spindle duration (Figures 3E–G), amplitude (Figures 4E–G) and frequency (Figures 5E–G) for automatically detected spindles

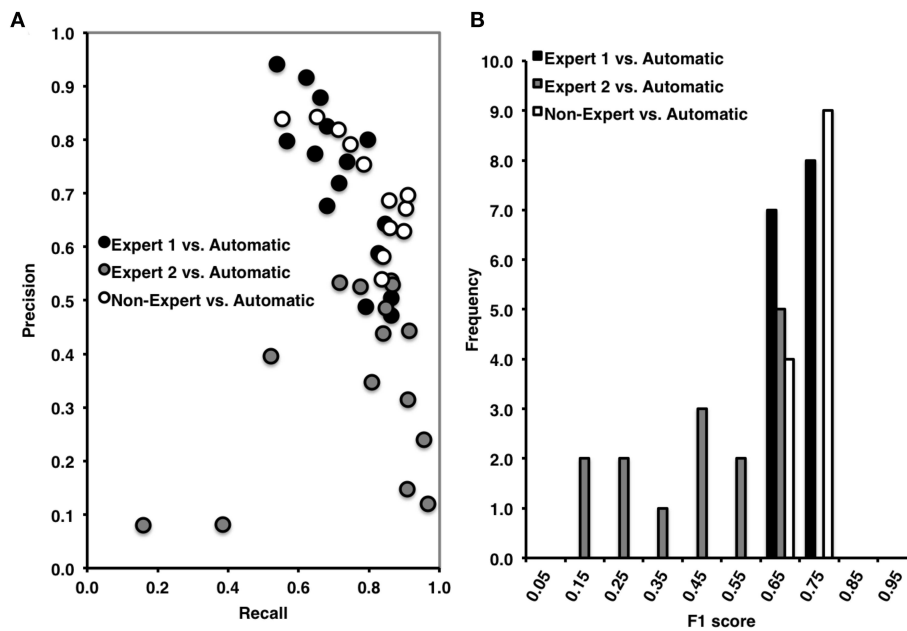


FIGURE 6 | (A) High precision and recall across recordings when comparing automated to Expert 1 spindle scoring (black) and to non-experts (open), but low precision and high, but variable recall when comparing Expert 2 to automatic spindle detection (gray). **(B)** Inter-rater agreement was consistently high across recordings scored by Expert 1 vs. automatic detections, ranging from 0.60 to 0.80 (Mean F1 = 0.71, ± 0.06) and in non-experts vs. automatic detection, ranging from 0.60 to 0.80 (F1 = 0.73, ± 0.04), but was low and variable between Expert 2 and the automatic detection, ranging from 0.10 to 0.70 (Mean F1 = 0.49, ± 0.04). F1 score = harmonic mean of recall and precision.

were significantly correlated with the spindles identified by visual scoring (all $p < 0.05$).

Distribution of Spindle Frequencies during NREM2 and SWS at Frontal and Parietal Regions

Consistent with previous reports (Zeitlhofer et al., 1997) **Figure 7** reveals that a greater number of faster frequency spindles predominated parietal regions whereas a greater number of slower frequency spindles predominated frontal regions in both NREM2 (**Figure 7A**, Cohen's $d = 0.43$) and SWS (**Figure 7B**, Cohen's $d = 0.78$). This dissociation was supported by significant spindle type (fast, slow) \times site (frontal, parietal) ANOVAs on automatically detected spindle density in NREM2 and SWS, which revealed that fast spindles predominated parietal regions as compared to slow spindles at frontal regions in both NREM2 [$F_{(1,14)} = 149.62$, $p < 0.001$], and SWS [$F_{(1,14)} = 194.19$, **Table 4**].

Spindle Density

Spindle characteristics over the course of NREM cycles and across frontal and parietal regions followed well-established patterns (**Figure 8**). A cycle (NREM cycle 1–4) \times spindle type (fast, slow) \times site (frontal, parietal) ANOVA for spindle density revealed a significant three-way interaction [$F_{(3,42)} = 3.98$, $p = 0.014$]. This was driven by a higher density of slow spindles ($3.38, \pm 0.62$) than fast spindles ($1.16, \pm 0.54$) at F3 as compared to a higher density of fast spindles ($3.31, \pm 0.91$) than slow spindles ($1.52, \pm 0.69$) at P3 [$F_{(1,14)} = 149.62$, $p < 0.001$]. Spindle density also differed across NREM cycles in a U-shaped

pattern (Himanen et al., 2002), but more so for fast spindles than slow spindles, as indicated by a significant type by NREM cycle interaction [$F_{(3,42)} = 4.74$, $p = 0.006$].

Spindle Duration

A similar pattern of results was observed for spindle duration, however the cycle by spindle type by site three-way interaction was not significant ($p > 0.4$). Slow sleep spindles ($0.63, \pm 0.01$) were longer in duration than fast spindles ($0.47, \pm 0.06$) at F3, but not at P3 (slow = $0.61, \pm 0.19$, fast = $0.66, \pm 0.07$), as supported by a significant spindle type by site interaction [$F_{(1,14)} = 38.91$, $p < 0.001$]. Spindle duration also varied over the course of the night as a function of: (1) spindle type, whereby slow spindles flowed an inverted U-shaped pattern more so than fast spindles [$F_{(3,42)} = 5.31$, $p = 0.003$], and (2) site, whereby spindles at P3 regions followed an inverted U-shaped more so than at F3 [$F_{(3,42)} = 6.27$, $p = 0.001$].

Spindle Amplitude

In terms of spindle amplitude, there was a significant cycle by spindle type by site three-way interaction [$F_{(3,42)} = 3.01$, $p = 0.041$], whereby fast spindles increased over the course of NREM cycles at frontal regions and decreased over the course of NREM cycles at parietal regions. However, there were no other significant interactions or main effects, thereby suggesting that spindle amplitude was relatively stable over the course of the night at frontal and parietal regions for both slow and fast spindles.

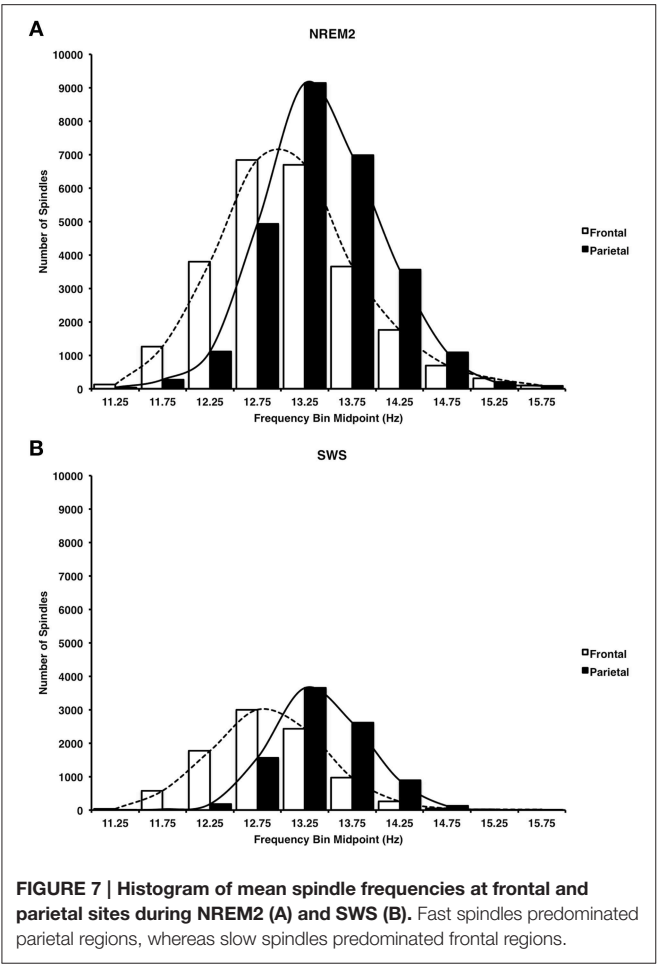


FIGURE 7 | Histogram of mean spindle frequencies at frontal and parietal sites during NREM2 (A) and SWS (B). Fast spindles predominated parietal regions, whereas slow spindles predominated frontal regions.

TABLE 4 | Group mean (\pm standard deviation) of fast and slow spindle density during NREM2 and SWS at frontal and parietal regions.

	NREM2		SWS	
	F3	P3	F3	P3
Fast	1.16 \pm 0.14	3.31 \pm 0.24	0.78 \pm 0.12	3.66 \pm 0.30
Slow	3.38 \pm 0.16	1.52 \pm 0.18	4.47 \pm 0.22	1.49 \pm 0.22

Spindle Frequency

By contrast, spindle frequency was very stable over the course of the night as a function of site ($p > 0.9$) and spindle type ($p > 0.4$), and there was no cycle by spindle type by site three-way interaction. However, fast spindles were faster at F3 (14.12, \pm 0.11) than fast spindles at P3 (13.76, \pm 0.15) whereas slow spindles did not differ at F3 (12.80, \pm 0.14) and P3 (12.75, \pm 0.10) as supported by a significant site by spindle type interaction [$F_{(1, 14)} = 11.85, p = 0.004$].

Discussion

In summary, the strengths of this automated detection method are: (1) CD was used to extract the signal of interest; a

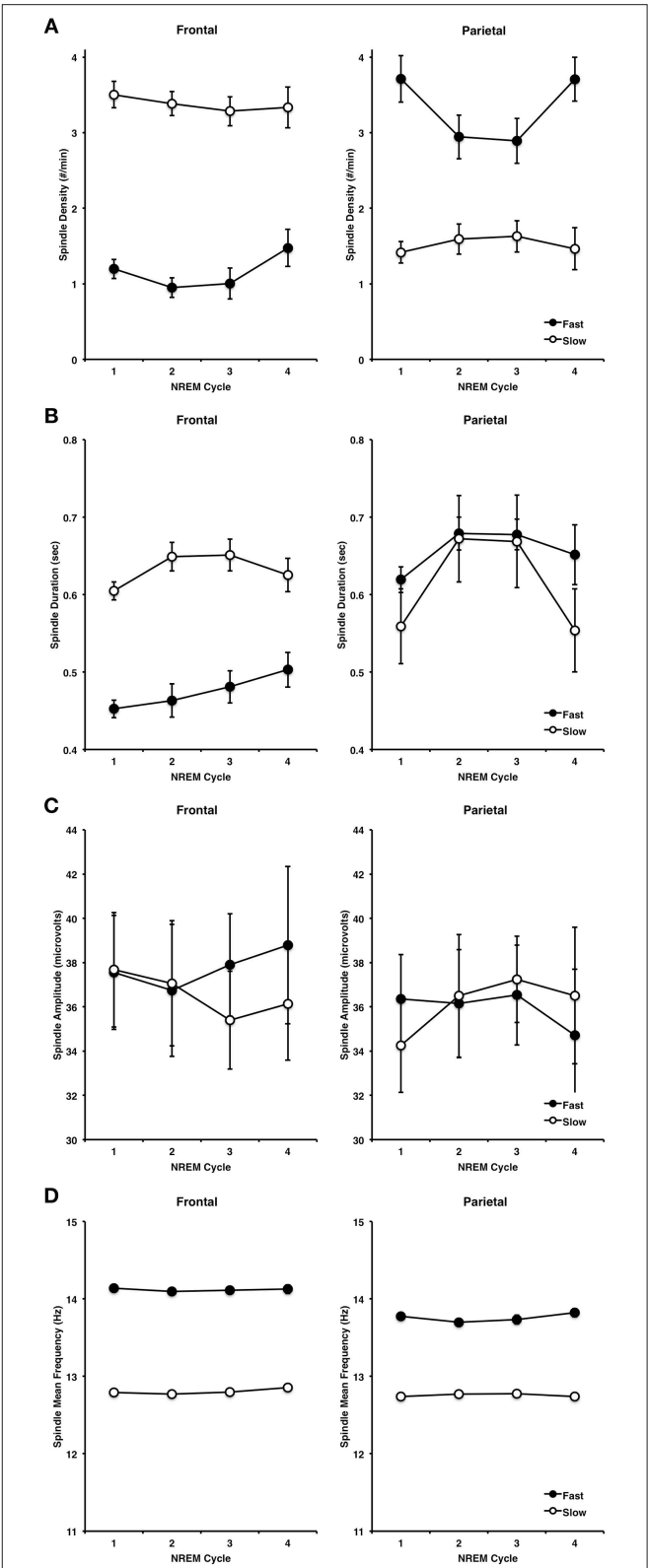


FIGURE 8 | Spindle characteristics over the course of the first four NREM periods, at frontal and parietal regions for fast and slow spindle types, including density (A), duration (B), amplitude (C) and frequency (D).

method that is appropriate for brief events in a well-defined frequency range for non-linear, non-stationary signals such as EEG, and transforms the signal to a waveform that makes event detection straightforward; (2) a sliding window was used to calculate the M and SD for the z-score normalization to account for intra-individual changes in the ratio of spindle-related sigma to the changes in ongoing sigma over time, and standardizes the amplitude of the signal across scalp locations and individuals; and (3) this method permits the effective use of a single, intuitive, user-defined amplitude parameter, with very few other parameters to extract the signal of interest, that are relatively intuitive (although sometimes non-trivial) to decide upon (e.g., spindle frequency bandwidth and normalization window duration). The validation was conducted on a freely available database of EEG, independently scored by two experts that employed two different methods to visually identify spindles, using spindle annotations that are available to other researchers for comparison. Thus, future direct comparisons to other detection methods are possible. Improving the reliability and validity of automated spindle detection will enable researchers to investigate the neural and functional correlates of spindles with greater confidence and reproducibility.

The results of the comparison between experts, highlights the difficulty in comparing automated detection methods to human visual scoring. Here, one expert (Expert 2) used conventional guidelines (e.g., AASM), while the other expert (Expert 1) utilized the aid of a sigma filtered channel to help identify spindles that are either difficult to discriminate from the ongoing EEG (i.e., spindles obscured by slow activity, or are small, or have unusual morphology in the normally filtered signal, e.g., 0.3–35 Hz). There were considerable differences between Expert 1 and Expert 2 in terms of low inter-rater agreement and in the characteristics of the spindles that were identified. Expert 1 also had a much higher level of agreement and identified spindles with similar characteristics as compared to the consensus of non-experts (who also used a sigma filtered channel to identify spindles) and the automated detection, than did Expert 2. By having human scorers view the EEG in a way that is closer to how the algorithm “sees” the EEG, this may have putatively improved agreement between automated and visual scoring and also may have minimized the differences in the characteristics of the spindles that were identified between automated and visual scoring. These results suggest that the use of the additional filtered channel allowed Expert 1 and non-experts to identify spindles that were difficult to visually identify, whereas Expert 2 identified far fewer spindles in general, that differed in their characteristics. This highlights the caveats of validating spindle detection methods against expert scoring as they can vary considerably from one individual to another (Warby et al., 2014), and also depending on adherence to established guidelines. To compare the automated detection to a potentially less idiosyncratic detection, here, we also compared the automated detection to a group of non-experts, to assess the utility of crowd-sourcing techniques. Overall, non-experts performed with a very high level of agreement as compared to the automated detection method. Thus, suggesting that manual scoring using web-based crowd sourcing tools could serve to generate a valid gold standard, and could even replace automated

detection of spindles, if the goal is to perform as close to the ideal performance of an expert scorer as possible. That said, automated detection methods do have their advantages over humans in terms of cost effectiveness and speed. They are also superior at precisely calculating the beginning and ending of individual spindles, can be tuned to perform better, can be used to identify spindles on multiple channels, and can perform well in the face of large amplitude, slow oscillations that visually obscure spindles. This is particularly advantageous in slow wave sleep where manual spindle detection is more challenging.

Importantly, in addition to comparing this method to experts and non-expert visual scoring methods, we investigated the characteristics of the spindles that were automatically detected to determine whether these spindles conform to known patterns from the extant literature. In summary, a greater density of fast spindles were observed at parietal than frontal regions, whereas a greater density of slow spindles were observed at frontal regions than parietal regions (Bódizs et al., 2009), and the change in spindle density followed the previously reported U-shaped pattern (for spindle power) over the course of the night (Himanen et al., 2002). Moreover, slow spindles were longer in duration than fast spindles at frontal regions and longer than both slow and fast spindles at parietal regions (Bódizs et al., 2009), whereas amplitude and frequency were relatively stable over the course of the night. Thus, many of the characteristics of the automatically detected spindles were consistent with known characteristics. Ultimately, given that scalp-recorded spindles are generated by the oscillatory firing of thalamocortical neurons (Steriade, 2006), future validation work comparing scalp-detected spindles to intracranial (e.g., unit activity) (Frauscher et al., 2015) may permit automated detection of spindles recorded from the scalp to be validated and identified more precisely.

The current method does not use any explicit minimum duration criteria for spindle detection and, due to the inherently straightforward approach used to extract the signal of interest (i.e., CD) minimizes - but does not eliminate—the number of parameters that require trial-and-error adjustment to optimize detection. Many existing definitions are based on minimum duration criteria (e.g., 0.5 s) derived from spindles large enough to be observed in the raw, mixed-frequency EEG (~ 0.5–35 Hz) by the naked eye alone. By excluding spindles <0.5 s in duration, this could possibly exaggerate inter-individual and group differences or lead to a systematic bias in the detection of large spindles. Of note, as can be seen in **Figure 8B**, the vast majority of fast frontal spindles that were automatically detected were < 0.5 s in duration. Automated techniques that require a minimum spindle duration to be decided a-priori, could benefit from loosening this criteria to determine the functional significance of short-duration spindles. This could be particularly problematic for elderly and psychiatric populations that have smaller spindles. Despite having no minimum duration criteria, the current method did not detect virtually any spindles shorter than 0.2 s. This likely contributed to the difference in spindle duration between expert and automated spindle detections, however, it is also likely that the ability to manually and

precisely score spindle duration is dependent on several factors, most notably the manual dexterity of the scorer, visual display settings, temporal resolution, and precision of the marking tool. In addition, the automated detection is able to detect and precisely measure (i.e., to the data point) very short duration events. This interpretation is supported by the fact that spindle duration was significantly and linearly related between visual and automated methods, thus suggesting that while visual and automated methods differ overall, there is a linear relationship between them, and thus, the difference may be due to the human scorer marking events systematically longer than the automated detection method (n.b., compare range of values for x-axis vs. y-axis in **Figure 3E**).

Both sigma power and spindles vary over the course of a night of sleep, and within individual NREM periods (De Gennaro and Ferrara, 2003; De Gennaro et al., 2005). In addition, spindles are relatively stable from night-to-night within an individual, but there are considerable inter-individual differences. Thus, it is crucial for automated spindle detection methods to account for these dynamic changes for accurate detection. Previous methods have accounted for inter-individual differences by adjusting the detection threshold that can be set to, e.g., the 95th or 99th percentile of the entire recording (Gais et al., 2002; Barakat et al., 2011; Nir et al., 2011; Cox et al., 2014). However, in order to account for variations, not only per individual and per derivation, here we employed the use of a sliding window to normalize the signal to the 99th percentile to adaptively detect spindles as the size of spindles change over the course of the night relative to the “background” non-spindle-related sigma activity. While intuitively, this may improve spindle detection, it is possible that for extremely intense periods of spindle activity (e.g., when the mean sigma activity is extremely elevated), that some smaller spindles may go undetected, and by contrast, in periods with very little spindle activity (e.g., when mean sigma activity is extremely low), that some very small spindles may be detected, or even lead to false detections. We feel however that this is unlikely as the window size employed is sufficiently long (60 s, equivalent to in this case, 3 epochs of consecutive NREM sleep), and thus would be unlikely to have very sustained periods of either high or low spindle activity great enough to systematically introduce a high number of false positives or false negatives. That said, additional work may be required to either optimize the size of the sliding window, or refine the method to automatically adapt the size of the sliding window.

Based on previously reported oscillatory frequency, topographic (Zeitlhofer et al., 1997) and functional activation differences between slow and fast spindles (Schabus et al., 2007), sleep spindles can be categorized as either slow or fast. Many current detection methods detect slow and fast spindles in two separate detection passes (Ray et al., 2010). For example, the frequency limits are set to detect slow spindles (e.g., 11–13.5 Hz), and then in a separate run on the same data, frequency limits are set to detect fast spindles (e.g., 13.5–16 Hz). This approach will invariably result in (perhaps the majority

of) the same spindle events to be detected twice, due to the overlap of the frequency extraction of the two adjacent bands. The current method detects spindles in the full band (e.g., 11–16 Hz) in one pass and categorizes spindles *post-hoc* as either slow or fast, based on each individual spindle events’ mean oscillatory frequency. This approach is advantageous such that the categorization of slow and fast spindles is orthogonal (i.e., so that the same spindle is not identified as both slow and fast).

The present investigation used a sample of young healthy subjects to validate the automated detection of spindles. In order to assess how this method preforms in populations where spindles are generally less frequent and smaller, such as elderly subjects (Martin et al., 2012), in clinical populations (Limoges et al., 2005; Steriade, 2005; Ferrarelli et al., 2007) or in noisy recordings, such as simultaneous EEG-fMRI, formal validation would also be required. However, preliminary validation results show automated detection using the present method, had a high inter-rater reliability with an established method in young and older subjects ($r = 0.98$) (Fogel et al., 2014b) and in EEG recorded simultaneously with fMRI (Fogel et al., 2014a).

The main advantage of this method is the employment of CD in conjunction with the normalization of the signal over time to account for inter- and intra-individual differences in spindles. An effort was made to minimize the number of parameters that require trial-and-error or arbitrary decisions, and the detection method has been validated against two experts employing different approaches (from a freely available repository) and a group of non-experts. In conclusion, the present method resolves or avoids many of the limitations of automated spindle detection, and performs well compared to a group of non-experts, and importantly, has good external validity with respect to the extant literature in terms of the characteristics of automatically detected spindles.

Acknowledgments

Many thanks to Sonia Frenette for her contribution to the MASS database, and to Dr. Christian O’Reilly who was instrumental to the development of this valuable repository and research tool. The authors would also like to acknowledge Dr. Julie Carrier for her input and guidance. This research was supported by a Canada Excellence Research Chair (CERC) grant (CERC-215063) to author AO, a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery grant (NSERC 418293DG-2012) to author RC and a NSERC postdoctoral fellowship (PDF-377124-2009) to author SF.

Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fnhum.2015.00507>

References

- Barakat, M., Doyon, J., Debas, K., Vandewalle, G., Morin, A., Poirier, G., et al. (2011). Fast and slow spindle involvement in the consolidation of a new motor sequence. *Behav. Brain Res.* 217, 117–121. doi: 10.1016/j.bbr.2010.10.019
- Bergmann, T. O., Mölle, M., Diedrichs, J., Born, J., and Siebner, H. R. (2011). Sleep spindle-related reactivation of category-specific cortical regions after learning face-scene associations. *Neuroimage* 59, 2733–2742. doi: 10.1016/j.neuroimage.2011.10.036
- Bixler, E. O., and Rhodes, J. M. (1968). Spindle activity during sleep in cultural-familial mild retardates. *Psychophysiology* 5, 212.
- Bódizs, R., Kis, T., Lázár, A. S., Havrán, L., Rigó, P., Clemens, Z., et al. (2005). Prediction of general mental ability based on neural oscillation measures of sleep. *J. Sleep Res.* 14, 285–292. doi: 10.1111/j.1365-2869.2005.00472.x
- Bódizs, R., Körmendi, J., Rigó, P., and Lázár, A. S. (2009). The individual adjustment method of sleep spindle analysis: methodological improvements and roots in the fingerprint paradigm. *J. Neurosci. Methods* 178, 205–213. doi: 10.1016/j.jneumeth.2008.11.006
- Bódizs, R., Lázár, A. S., and Rigó, P. (2008). Correlation of visuospatial memory ability with right parietal EEG spindling during sleep. *Acta Physiol. Hung.* 95, 297–306. doi: 10.1556/APhysiol.95.2008.3.5
- Bonjean, M., Baker, T., Lemieux, M., Timofeev, I., Sejnowski, T., and Bazhenov, M. (2011). Corticothalamic feedback controls sleep spindle duration *in vivo*. *J. Neurosci.* 31, 9124–9134. doi: 10.1523/JNEUROSCI.0077-11.2011
- Broughton, R., Healey, T., Maru, J., Green, D., and Pagurek, B. (1978). A phase locked loop device for automatic detection of sleep spindles and stage 2. *Electroencephalogr. Clin. Neurophysiol.* 44, 677–680. doi: 10.1016/0013-4694(78)90134-7
- 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., Mölle, M., Eross, L., Barsi, P., Halász, P., and Born, J. (2007). Temporal coupling of parahippocampal ripples, sleep spindles and slow oscillations in humans. *Brain* 130(Pt 11), 2868–2878. doi: 10.1093/brain/awm146
- Cox, R., Hofman, W. F., de Boer, M., and Talamini, L. M. (2014). Local sleep spindle modulations in relation to specific memory cues. *Neuroimage* 99, 103–110. doi: 10.1016/j.neuroimage.2014.05.028
- Crowley, K., Trinder, J., Kim, Y., Carrington, M., and Colrain, I. M. (2002). The effects of normal aging on sleep spindle and k-complex production. *Clin. Neurophysiol.* 113, 1615–1622. doi: 10.1016/S1388-2457(02)00237-7
- 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/smr.2002.0252
- De Gennaro, L., Ferrara, M., and Bertini, M. (2000). Topographical distribution of spindles: variations between and within nrem sleep cycles. *Sleep Res. Online* 3, 155–160. doi: 10.1046/j.1365-2869.2000.00193.x
- De Gennaro, L., Ferrara, M., Vecchio, F., Curcio, G., and Bertini, M. (2005). An electroencephalographic fingerprint of human sleep. *Neuroimage* 26, 114–122. doi: 10.1016/j.neuroimage.2005.01.020
- Ferrarelli, F., Huber, R., Peterson, M. J., Massimini, M., Murphy, M., Riedner, B. A., et al. (2007). Reduced sleep spindle activity in schizophrenia patients. *Am. J. Psychiatry* 164, 483–492. doi: 10.1176/appi.ajp.164.3.483
- Fogel, S., Albouy, G., King, B., Vien, C., Karni, A., Benali, H., et al. (2014b). Motor memory consolidation depends upon reactivation driven by the action of sleep spindles. *J. Sleep Res.* 23, 47.
- Fogel, S. M., Albouy, G., Vien, C., Popovicci, R., King, B. R., Hoge, R., et al. (2014a). fMRI and sleep correlates of the age-related impairment in motor memory consolidation. *Hum. Brain Mapp.* 35, 3625–3645. doi: 10.1002/hbm.22426
- Fogel, S. M., Nader, R., Cote, K. A., and Smith, C. T. (2007). Sleep spindles and learning potential. *Behav. Neurosci.* 121, 1–10. doi: 10.1037/0735-7044.121.1.1
- Fogel, S. M., and Smith, C. T. (2006). Learning-dependent changes in sleep spindles and stage 2 sleep. *J. Sleep Res.* 15, 250–255. doi: 10.1111/j.1365-2869.2006.00522.x
- Fogel, S. M., and Smith, C. T. (2011). The function of the sleep spindle: a physiological index of intelligence and a mechanism for sleep-dependent memory consolidation. *Neurosci. Biobehav. Rev.* 35, 1154–1165. doi: 10.1016/j.neubiorev.2010.12.003
- Frauscher, B., von Ellenrieder, N., Dubeau, F., and Gotman, J. (2015). Scalp spindles are associated with widespread intracranial activity with unexpectedly low synchrony. *Neuroimage* 105, 1–12. doi: 10.1016/j.neuroimage.2014.10.048
- Gais, S., Mölle, M., Helms, K., and Born, J. (2002). Learning-dependent increases in sleep spindle density. *J. Neurosci.* 22, 6830–6834.
- Geiger, A., Huber, R., Kurth, S., Ringli, M., Jenni, O. G., and Achermann, P. (2011). The sleep EEG as a marker of intellectual ability in school age children. *Sleep* 34, 181–189.
- Gibbs, E. L., and Gibbs, F. A. (1962). Extreme spindles: correlation of electroencephalographic sleep pattern with mental retardation. *Science* 138, 1106–1107. doi: 10.1126/science.138.3545.1106
- Hao, Y. -L., Ueda, Y., and Ishii, N. (1992). Improved procedure of complex demodulation and an application to frequency analysis of sleep spindles in EEG. *Med. Biol. Eng. Comput.* 30, 406–412. doi: 10.1007/BF02446168
- Himanan, S. L., Virkkala, J., Huhtala, H., and Hasan, J. (2002). Spindle frequencies in sleep EEG show u-shape within first four NREM sleep episodes. *J. Sleep Res.* 11, 35–42. doi: 10.1046/j.1365-2869.2002.00273.x
- Huupponen, E., De Clercq, W., Gómez-Herrero, G., Saastamoinen, A., Egiastian, K., Värrä, A., et al. (2006). Determination of dominant simulated spindle frequency with different methods. *J. Neurosci. Methods* 156, 275–283. doi: 10.1016/j.jneumeth.2006.01.013
- Iber, C. (2007). *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. Westchester, IL: American Academy of Sleep Medicine.
- Klonowski, W. (2007). From conformons to human brains: an informal overview of nonlinear dynamics and its applications in biomedicine. *Nonlinear Biomed. Phys.* 1:5. doi: 10.1186/1753-4631-1-5
- Ktonas, P. Y., Golemati, S., Xanthopoulos, P., Sakalis, V., Ortigueira, M. D., Tsekou, H., et al. (2009). Time-frequency analysis methods to quantify the time-varying microstructure of sleep EEG spindles: possibility for dementia biomarkers? *J. Neurosci. Methods* 185, 133–142. doi: 10.1016/j.jneumeth.2009.09.001
- Ktonas, P. Y., and Papp, N. (1980). Instantaneous envelope and phase extraction from real signals: theory, implementation, and an application to EEG analysis. *Signal Processing* 2, 373–385. doi: 10.1016/0165-1684(80)90079-1
- Limoges, E., Mottron, L., Bolduc, C., Berthiaume, C., and Godbout, R. (2005). Atypical sleep architecture and the autism phenotype. *Brain* 128, 1049–1061. doi: 10.1093/brain/awh425
- Martin, N., Lafortune, M., Godbout, J., Barakat, M., Robillard, R., Poirier, G., et al. (2012). Topography of age-related changes in sleep spindles. *Neurobiol. Aging* 34, 468–476. doi: 10.1016/j.neurobiolaging.2012.05.020
- Möller, 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.
- Nicolas, A., Petit, D., Rompré, S., and Montplaisir, J. (2001). Sleep spindle characteristics in healthy subjects of different age groups. *Clin. Neurophysiol.* 112, 521–527. doi: 10.1016/S1388-2457(00)00556-3
- Nir, Y., Staba, R. J., Andrillon, T., Vyazovskiy, V. V., Cirelli, C., Fried, I., et al. (2011). Regional slow waves and spindles in human sleep. *Neuron* 70, 153–169. doi: 10.1016/j.neuron.2011.02.043
- Nishida, M., and Walker, M. P. (2007). Daytime naps, motor memory consolidation and regionally specific sleep spindles. *PLoS ONE* 2:e341. doi: 10.1371/journal.pone.0000341
- 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
- Peters, K. R., Smith, V., and Smith, C. T. (2007). Changes in sleep architecture following motor learning depend on initial skill level. *J. Cogn. Neurosci.* 19, 817–829. doi: 10.1162/jocn.2007.19.5.817
- Principe, J. C., and Smith, J. R. (1982). Sleep spindle characteristics as a function of age. *Sleep* 5, 73.
- Ray, L. B., Fogel, S. M., Smith, C. T., and Peters, K. R. (2010). Validating an automated sleep spindle detection algorithm using an individualized approach. *J. Sleep Res.* 19, 374–378. doi: 10.1111/j.1365-2869.2009.00802.x
- Rechtschaffen, A., and Kales, A. (1968). *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*. Los

- Angeles, CA: Brain Information Service/ Brain Research Institute, University of California.
- Schabus, M., Dang-Vu, T. T., Albouy, G., Baletau, 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
- Schabus, M., Dang-Vu, T. T., Heib, D. P. J., Boly, M., Desseilles, M., Vandewalle, G., et al. (2012). The fate of incoming stimuli during NREM sleep is determined by spindles and the phase of the slow oscillation. *Front. Neurol.* 3:40. doi: 10.3389/fneur.2012.00040
- Schabus, M., Gruber, G., Parapatics, S., Sauter, C., Klösch, G., Anderer, P., et al. (2004). Sleep spindles and their significance for declarative memory consolidation. *Sleep* 27, 1479–1485.
- Schabus, M., Hödlmoser, K., Gruber, G., Sauter, C., Anderer, P., Klösch, G., et al. (2006). Sleep spindle-related activity in the human EEG and its relation to general cognitive and learning abilities. *Eur. J. Neurosci.* 23, 1738–1746. doi: 10.1111/j.1460-9568.2006.04694.x
- Schimicek, P., Zeitlhofer, J., Anderer, P., and Saletu, B. (1994). Automatic sleep-spindle detection procedure: aspects of reliability and validity. *Clin. Electroencephalogr.* 25, 26–29. doi: 10.1177/155005949402500108
- Shibagaki, M., Kiyono, S., and Watanabe, K. (1982). Spindle evolution in normal and mentally retarded children: a review. *Sleep* 5, 47–57.
- Silverstein, L. D., and Levy, C. M. (1976). The stability of the sigma sleep spindle. *Electroencephalogr. Clin. Neurophysiol.* 40, 666–670. doi: 10.1016/0013-4694(76)90142-5
- Smith, C. T., Nixon, M. R., and Nader, R. S. (2004). Posttraining increases in REM sleep intensity implicate REM sleep in memory processing and provide a biological marker of learning potential. *Learn. Mem.* 11, 714–719. doi: 10.1101/lm.74904
- Steriade, M. (2005). Sleep, epilepsy and thalamic reticular inhibitory neurons. *Trends Neurosci.* 28, 317–324. doi: 10.1016/j.tins.2005.03.007
- Steriade, M. (2006). Grouping of brain rhythms in corticothalamic systems. *Neuroscience* 137, 1087–1106. doi: 10.1016/j.neuroscience.2005.10.029
- Uchida, S., Atsumi, Y., and Kojima, T. (1994). Dynamic relationships between sleep spindles and delta waves during a NREM period. *Brain Res. Bull.* 33, 351–355. doi: 10.1016/0361-9230(94)90205-4
- Ujma, P. P., Konrad, B. N., Genzel, L., Bleifuss, A., Simor, P., Pótári, A., et al. (2014). Sleep spindles and intelligence: evidence for a sexual dimorphism. *J. Neurosci.* 34, 16358–16368. doi: 10.1523/JNEUROSCI.1857-14.2014
- Walter, D. O. (1968). The method of complex demodulation. *Electroencephalogr. Clin. Neurophysiol. Suppl.* 27, 53–57.
- Wamsley, E. J., Tucker, M. A., Shinn, A. K., Ono, K. E., McKinley, S. K., Ely, A. V., et al. (2012). Reduced sleep spindles and spindle coherence in schizophrenia: mechanisms of impaired memory consolidation? *Biol. Psychiatry* 71, 154–161. doi: 10.1016/j.biopsych.2011.08.008
- Warby, S. C., Wendt, S. L., Welinder, P., Munk, E. G. S., Carrillo, O., Sorensen, H. B. D., 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
- Werth, E., Achermann, P., Dijk, D. J., and Borbély, 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
- 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.

Copyright © 2015 Ray, Sockeel, Soon, Bore, Myhr, Stojanoski, Cusack, Owen, Doyon and Fogel. 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.

Stage-independent, single lead EEG sleep spindle detection using the continuous wavelet transform and local weighted smoothing

Athanasios Tsanas^{1,2,3*} and Gari D. Clifford^{3,4,5}

¹ Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Oxford, UK, ² Wolfson Centre for Mathematical Biology, Mathematical Institute, University of Oxford, Oxford, UK, ³ Nuffield Department of Medicine, Sleep and Circadian Neuroscience Institute, University of Oxford, UK, ⁴ Department of Biomedical Informatics, Emory University, Atlanta, GA, USA, ⁵ Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, USA

OPEN ACCESS

Edited by:

Christian O'Reilly,
McGill University, Canada

Reviewed by:

E. J. W. VanSomeren,
Netherlands Institute for
Neuroscience, Netherlands
Marek Adamczyk,
Max Planck Institute of Psychiatry,
Germany

*Correspondence:

Athanasios Tsanas,
Mathematical Institute, University of
Oxford, Andrew Wiles Building,
Woodstock Road, Oxford OX2 6GG,
UK
tsanas@maths.ox.ac.uk;
tsanasthanasis@gmail.com

Received: 15 November 2014

Accepted: 17 March 2015

Published: 08 April 2015

Citation:

Tsanas A and Clifford GD (2015)
Stage-independent, single lead EEG
sleep spindle detection using the
continuous wavelet transform and
local weighted smoothing.
Front. Hum. Neurosci. 9:181.
doi: 10.3389/fnhum.2015.00181

Sleep spindles are critical in characterizing sleep and have been associated with cognitive function and pathophysiological assessment. Typically, their detection relies on the subjective and time-consuming visual examination of electroencephalogram (EEG) signal(s) by experts, and has led to large inter-rater variability as a result of poor definition of sleep spindle characteristics. Hitherto, many algorithmic spindle detectors inherently make signal stationarity assumptions (e.g., Fourier transform-based approaches) which are inappropriate for EEG signals, and frequently rely on additional information which may not be readily available in many practical settings (e.g., more than one EEG channels, or prior hypnogram assessment). This study proposes a novel signal processing methodology relying solely on a single EEG channel, and provides objective, accurate means toward probabilistically assessing the presence of sleep spindles in EEG signals. We use the intuitively appealing continuous wavelet transform (CWT) with a Morlet basis function, identifying regions of interest where the power of the CWT coefficients corresponding to the frequencies of spindles (11–16 Hz) is large. The potential for assessing the signal segment as a spindle is refined using local weighted smoothing techniques. We evaluate our findings on two databases: the MASS database comprising 19 healthy controls and the DREAMS sleep spindle database comprising eight participants diagnosed with various sleep pathologies. We demonstrate that we can replicate the experts' sleep spindles assessment accurately in both databases (MASS database: sensitivity: 84%, specificity: 90%, false discovery rate 83%, DREAMS database: sensitivity: 76%, specificity: 92%, false discovery rate: 67%), outperforming six competing automatic sleep spindle detection algorithms in terms of correctly replicating the experts' assessment of detected spindles.

Keywords: decision support tool, hypnogram, signal processing algorithms, sleep spindle, sleep structure assessment

Introduction

Sleep spindles are characteristic oscillatory patterns of brain activity which can be visually detected in human electroencephalography (EEG) signals. These transient patterns are typically portrayed as nearly sinusoidal waxing and waning waveforms with a characteristic frequency profile of 11–16 Hz [formerly this range was narrowed between 12 and 14 Hz in the Rechtschaffen and Kales criteria (Rechtschaffen and Kales, 1968), and different research labs might use slightly different frequency ranges] (Iber et al., 2007; Kryger et al., 2010). Interestingly, although sleep spindles exhibit substantially varying characteristics (amplitude, duration, density) in the population, they are fairly stable for individuals (Werth et al., 1997). Spindles are generated in the thalamus, and contemporary evidence suggests they can be classified into slow spindles (11–13 Hz) and fast spindles (13–16 Hz), which are believed to regulate different activation patterns (DeGennaro and Ferrara, 2003).

The presence of sleep spindles is one of the hallmarks for determining stage 2 (S2) in the *hypnogram*, which provides an overall representation of sleep structure successively assigning short signal segments (known as *epochs*, usually of 30 s duration) to one of five sleep stages (Iber et al., 2007). They have been associated with various higher cognitive processes in particular memory (Tamminen et al., 2010), but also learning performance (Schmidt et al., 2006) and skill performance (Astill et al., 2015). Moreover, there is a growing body of research literature highlighting their potential as biomarkers: a number of studies have reported clinically significant differences in spindle characteristics for a range of neurological disorders (Ferrarelli et al., 2007; Wamsley et al., 2012; Christensen et al., 2014).

The gold standard for the determination of sleep spindles has traditionally been achieved through visual inspection of the EEG by sleep physiology experts. Despite the best attempts of experts to standardize protocols, expert-based assessments rely on expensive human resources, depend on the rater's experience and level of expertise, are laborious and prone to errors due to fatigue, and by nature cannot scale to handle very large datasets. As with all cases where the gold standard is set by *subjective* assessments of trained experts, there can always be an argument that an automated algorithmic process could provide an alternative, often sufficiently accurate, robust, scalable, replicable, cost-effective, and objective mode to achieve the aim; indicative studies highlighting these concepts include Grove and Meehl (1996), Seshadrinathan et al. (2010), and Tsanas (2012) amongst many others. At the very least, the development of algorithmic tools can facilitate and expedite the work of trained experts particularly due to the sheer amount of the growing availability of massive datasets.

There are several approaches that have been proposed to tackle the problem of automatic sleep spindle detection. The majority of the proposed algorithms rely on a time-frequency analysis. In all cases, a major hurdle is the determination of appropriate thresholds, which may need to be optimized for each individual. Unfortunately, it is difficult to define universally applicable thresholds due to the large variability in spindle characteristics

amongst individuals (Werth et al., 1997). Frequently, the setting of these thresholds for many algorithms require prior hypnogram assessment, and subsequent focusing only on stage 2 sleep (Möller et al., 2002; Wamsley et al., 2012) or Non Rapid Eye Movement (NREM) sleep (Ferrarelli et al., 2007; Martin et al., 2013). However, we argue that all these approaches are quite restrictive, particularly because in practice we want to completely automate the EEG signal processing task without requiring prior hypnogram assessment by experts. Detecting spindles might be the end goal in one application, but could also be used to guide automated sleep staging assessment. Another generic approach for many algorithms is attempting to determine the presence of spindles by successively searching over pre-defined short windowed EEG segments [typically 1 s, e.g., see Huupponen et al. (2007), although some approaches rely on the detection of spindles in the more traditional 30-s epochs used in hypnogram assessment]. A major limitation with this approach is that one needs to specify a small signal segment to assess whether a spindle occurred within that segment and loosely approximate the spindle onset and offset.

Recently Wendt et al. (2012) introduced a fusion approach to detect spindles applying their sleep detection algorithm on two EEG channels (central and occipital). However, spindles are known to occur locally (Kryger et al., 2010) and hence there is no guarantee that both the central and occipital deflections will identify the spindle; furthermore, this complicates the practical task of spindle assessment by imposing the requirement that additional recordings are available (ideally a single channel would be sufficient for detecting spindles locally). It should be noted that localized sleep can occur, and therefore a single channel cannot reveal the overall sleep structure for the entire brain. In practice we want to focus on specific brain areas, detecting spindles *locally*, e.g., at the central regions where the spindle density is maximal (Kryger et al., 2010); some interesting recent work has focused on spindle propagation (O'Reilly and Nielsen, 2014b).

One of the simplest algorithmic approaches for detecting spindles is to band-pass the EEG signal and assess the presence of spindles by setting an appropriate (relative) threshold on the amplitude of the band-passed version of the signal (Schimicek et al., 1994), which is both sensible and remains topical to this day at least as a benchmark. Similarly, the ubiquitous Fourier Transform (FT) has been investigated in this application (Huupponen et al., 2007). However, there are inherent limitations of the FT in that it implicitly assumes a periodic signal, and also that it requires a sufficiently adequate number of samples for the spectrum estimation; in practice this sets a minimum requirement of about 1 s signal segment (Pardey et al., 1996). In turn, this means that with FT it is fundamentally impossible to correctly determine the spindle onset and offsets accurately as highlighted previously. Wavelet analysis is particularly suitable for analyzing non-stationary signals (such as the EEG), thus overcoming certain shortcomings of the traditional spectral analysis with the FT, and hence has justifiably attracted interest recently in the spindle detection domain (Sitnikova et al., 2009; Wamsley et al., 2012).

This study extends the methodology of recent approaches using the Continuous Wavelet Transform (CWT) with Morlet

basis functions (Sitnikova et al., 2009; Wamsley et al., 2012). The Morlet wavelet has been widely used in many practical applications because it has the desirable property that it minimizes the product of the wavelet's time and frequency spreads; hence it optimizes the time-frequency resolution (Addison, 2002). The main novelty of this work lies in the processing of the relative normalized power of the CWT coefficients to determine spindle candidates. Whereas previous studies computed the moving average of the power of the CWT coefficients to detect spindles directly, we first rank the CWT coefficients in terms of their normalized power at each time instant. Then, we compute the instantaneous ratio of the CWT coefficients falling within the scale spindle range (corresponding to the standard 11–16 Hz frequency range) over the top 10 ranked CWT coefficients. This ratio denotes the “instantaneous strength” of detecting a spindle, which is subsequently processed with weighted moving average methods to detect spindles. The proposed algorithm overcomes several shortcomings of competing algorithms: (a) it does not require processing successive small (e.g., 1 s) signal segments which blur the determination of true onset and offset of spindles (instead the algorithm works directly the entire signal), (b) it does not require prior hypnogram assessment, (c) it uses a single EEG lead. Moreover, using the proposed algorithm we can determine the frequency variation contour as a function of time within each spindle: these features may have clinical relevance, a fact which is often overlooked by contemporary competing approaches (for example, FT-based approaches cannot readily provide this information).

Materials and Methods

This section summarizes the dataset used in this study, summarizes some of the previously published algorithms against which the new sleep spindle detection algorithm developed in this study is benchmarked, and outlines the evaluation criteria for assessing the performance of the algorithms.

Data

We used two publicly available databases in this study.

The first database was collected during the DREAMS project (Devuyst et al., 2011), which aimed to provide a platform to assist assessment of automatic detection algorithms. The sleep spindles database contains recordings from eight participants with diverse sleep pathologies (dysomnia, restless legs syndrome, insomnia, apnoea/hypopnoea syndrome). Two EOG channels (P8-A1, P18-A1), three EEG channels (CZ-A1 or C3-A1, FP1-A1, and O1-A1) and one submental EMG channel were recorded, using a sampling frequency of 200 Hz (six signals), 100 Hz (one signal), or 50 Hz (one signal). A segment of 30 min of a central EEG channel (C3-A1 or Cz-A1) was extracted from each whole-night recording, and two experts have independently annotated the presence of sleep spindles. The second expert has only annotated six out of the eight recordings, and has not provided the exact duration of the assessed spindles (hence, it was all assigned to be 1 s in duration). Although the hypnograms (according to standard Rechtschaffen and Kales criteria) were

available, these were not used in the assessment of the spindles by the experts. The dataset along with additional information is publicly available from: <http://www.tcts.fpms.ac.be/~devuyst/Databases/DatabaseSpindles/>.

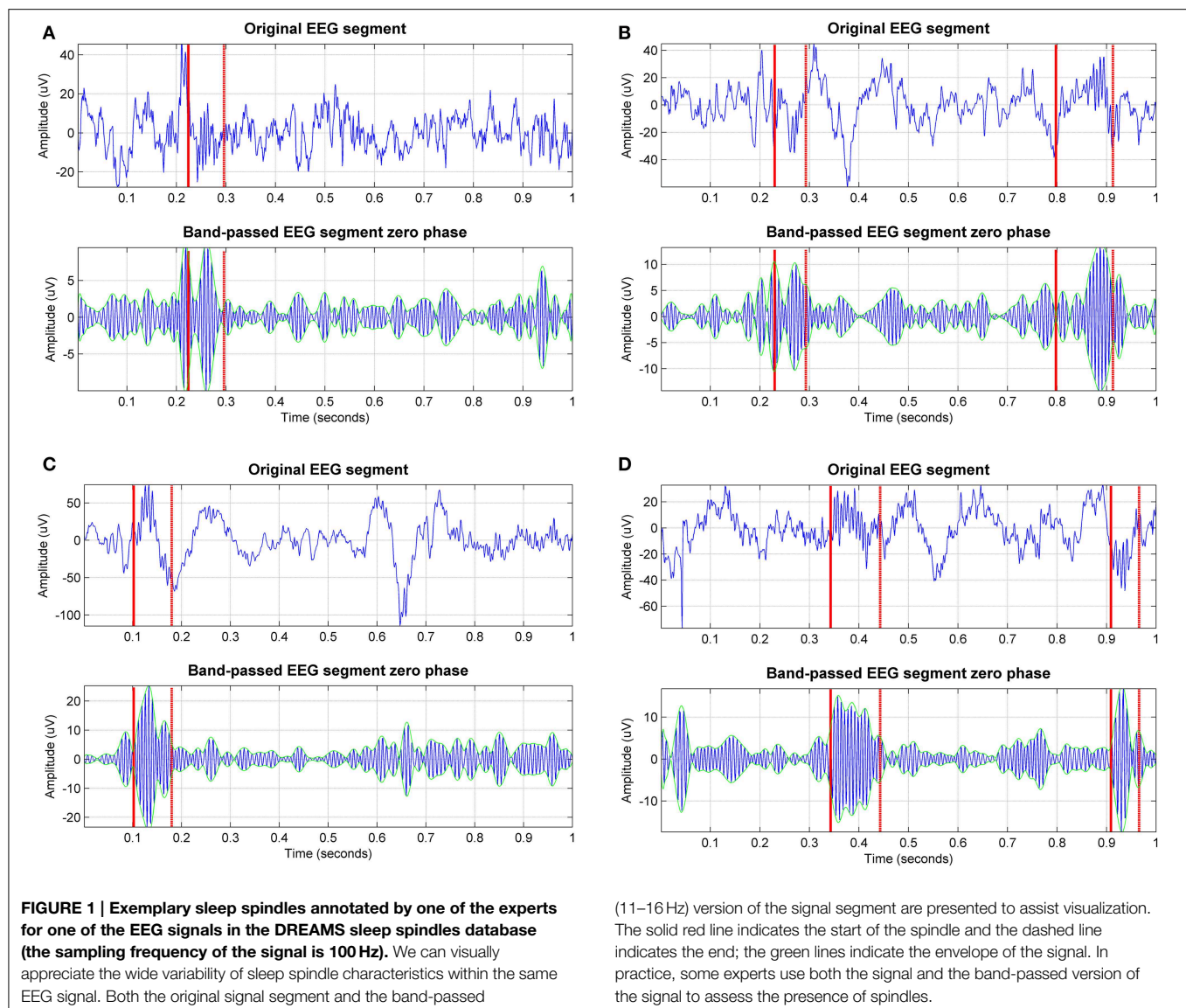
The second database was collected as part of a large project looking into sleep, the Montreal Archive of Sleep Studies (MASS) (O'Reilly et al., 2014a). It contains overnight PSG recordings from 19 healthy controls: specifically, electroencephalography (EEG) montage of 19 channels, 4 electro-oculography (EOG), electromyography (EMG), electrocardiography (ECG), and respiratory signals. The EEG signals were sampled at 256 Hz. The database was annotated independently by two experts for sleep spindles. The second expert has only annotated 15 out of the 19 signals for sleep spindles. Hypnograms (according to standard Rechtschaffen and Kales criteria) were also made available. For further details see O'Reilly et al. (2014a). The dataset became available to the authors of this study *after* the development of the algorithms and the original submission of the manuscript; we deliberately decided not to further fine-tune the original algorithms developed using the DREAMS data to guide the sleep spindle estimation process, in order not to bias the presented findings in any way. The dataset can be accessed from: <http://www.ceams-carsm.ca/en/MASS>.

In all cases, the EEG signals were resampled at 100 Hz.

Methods

Before delving into the details of the sleep spindle detection algorithms, it is useful to revisit the definition of spindles, and visualize some examples annotated by experts in order to motivate the subsequent algorithmic development. According to the latest recommendation of the AASM Manual for the scoring of sleep, a spindle is defined as “a train of distinct waves with frequency 11–16 Hz (most commonly 12–14 Hz) with a duration ≥ 0.5 s, usually maximal in amplitude in the central derivations.” (Iber et al., 2007). The spindle frequency range is nowadays generally accepted to be 11–16 Hz, but the range over which researchers focus may vary slightly depending on the research lab, e.g., 10.5–16 Hz (Huupponen et al., 2007), or 12–15 Hz (Ferrarelli et al., 2007); the standard reference book “Principles of Sleep Medicine” quotes the range 10–15 Hz (Kryger et al., 2010). We note there is no formal recommendation for the use of amplitude thresholds to detect a spindle, although many researchers have explicitly used amplitude criteria in their algorithmic implementations (Devuyst et al., 2011; Wamsley et al., 2012). Also, many researchers have relaxed the requirement of the minimum spindle duration, e.g., 0.4 s (Wamsley et al., 2012) or even as low 0.3 s instead (Warby et al., 2014). In practice, most spindles are typically around 0.5–1.5 s (very occasionally might be over 2 s), and typically most researchers impose a maximum length constraint (typically 3 s, e.g., Warby et al., 2014) in their algorithmic approaches.

Sleep textbooks often depict sleep spindles as waxing and waning, nearly sinusoidal waveforms; however, in practice spindle waveforms are markedly noisy, exhibiting diverse characteristics. **Figure 1** illustrates some spindles detected by experts for the same signal in the DREAMS sleep spindle database (Devuyst et al., 2011). It is striking that all these transient waveforms



(stemming from the same EEG recording and being only a few seconds or minutes apart) display such widely varying features (for example compare the peak-to-peak amplitudes). Nevertheless, all these illustrative examples are considered true spindles according to at least one of the two experts and set the ground truth against which all automated sleep spindle detection algorithms are benchmarked. For each signal we also present its band-passed version at the spindle frequency range. Following visual inspection of these plots, we can postulate that amplitude may be a misleading criterion to assess automatically the presence of spindles; on the other hand, the presence of the spindle appears to be more consistent when also observing the band-pass version of the signals. This exploratory step may assist in the motivation and understanding of the sleep detection algorithms which are presented in the following sections.

Contemporary Sleep Spindle Detection Algorithms

For simplicity and to conform to the terminology of Warby et al. (2014) we will denote with a_x each of the sleep spindle detection algorithms used in this study, where the subscript indicates the corresponding algorithm. In this section we summarize the six spindle detection algorithms used in Warby et al. (2014) (denoted here with a_1 – a_6), and in the following section we will introduce the new algorithmic approaches. These algorithms (occasionally with slight modifications) have been widely used in a number of studies, and therefore can be considered indicative of the most popular contemporary approaches to automatically detect sleep spindles. We used the Matlab implementations provided by Warby et al. (2014) for a_1 – a_6 and the description of the algorithms below follows their algorithmic modifications; hence the described algorithms differ slightly in comparison to the original algorithms. Our own algorithms were also implemented

in Matlab, and are made freely available on Physionet (www.physionet.org) and the first author's website.

Algorithm a₁, Bódizs' average amplitude spectrum

The first algorithm, a₁, is due to Bódizs et al. (2009), and attempts to tackle the problem of intra-subject variability in terms of EEG characteristics by incorporating subject-specific information (hence building upon the findings of Werth et al. (1997) that the variability of the spindle characteristics is low for each individual). The algorithm detects spindles in customized frequency ranges (identifying slow and fast spindles) using the average amplitude spectrum of NREM sleep using epochs of 4 s. The decision to evaluate the presence of a spindle is based on the amplitude threshold in each of the two band-pass regions for slow spindles or fast spindles. The implementation by Warby et al. (2014) used here requires both a central and an occipital EEG channel.

Algorithm a₂, Ferrarelli's band pass and signal envelope algorithm

The second algorithm, a₂, was proposed by Ferrarelli et al. (2007) and with slight modifications has been used in some recent studies, e.g., Astill et al. (2015). The algorithm applies a band-pass filter (11–15 Hz) to the NREM data (epochs), and the envelope of the resulting signal is subsequently used. An amplitude threshold (threshold₁) is then set relative to the mean signal amplitude (because different channels exhibit different amplitude profiles). A spindle is marked by first detecting a local maximum in the envelope of the filtered signal above threshold₁, and its duration is determined by identifying the preceding and following instances when this amplitude falls below a lower threshold (threshold₂), i.e., detecting the nearest troughs below threshold₂ (local minima). The slightly different versions of this type of algorithm set threshold₁ and threshold₂ slightly differently than the original algorithm, but the essential main idea remains the same.

Algorithm a₃, Mölle's band pass RMS overlapping moving window

The third algorithm, a₃, was described by Mölle et al. (2002). This algorithm is also band-pass filtering the NREM data at the spindle frequency range (12–15 Hz), and subsequently computes the Root Mean Squared (RMS) value of the filtered data over a short-frame overlapping (50%) moving window of 100 ms. Then, spindles are determined only on the data from sleep stage 2 depending on whether the RMS value exceeds an amplitude threshold (set at 1.5 times the standard deviation of the band-pass filtered signal) and the duration is within the acceptable spindle limits (0.3–3 s).

Algorithm a₄, Martin's band pass RMS percentile moving window

The fourth algorithm, a₄, by Martin et al. (2013) is conceptually very similar to a₃. It differs from a₃ in terms of the spindle frequency range used (11–15 Hz) for the band-pass filter, the use of a non-overlapping time window (25 ms) to compute the RMS values, and the threshold for detecting the spindle which is set to be the 95th percentile of the RMS signal.

Algorithm a₅, Wamsley's CWT moving average

The fifth algorithm, a₅, was developed by Wamsley et al. (2012). Contrary to the algorithms described so far, this algorithm is based on the CWT, which has some desirable properties for analyzing EEG signals as discussed previously. The algorithm relies on prior hypnogram assessment and attempts to detect spindles during stage 2. The signal is transformed into the wavelet domain using the complex Morlet wavelet basis function. The Morlet scales corresponding approximately to the pseudo-frequencies of interest (10–16 Hz) were used, and the moving average of the coefficients using a 100 ms sliding window was computed; when it exceeded a threshold for a minimum of 0.3 s a spindle was registered. The threshold was set using only the amplitude of epochs assessed as stage 2 by experts.

Algorithm a₆, Wendt's two-channel band pass and signal envelope combination

The sixth algorithm, a₆, was developed by Wendt et al. (2012). This algorithm is conceptually similar to a₂, the main difference is that the boundaries for the spindle detection are determined using local extrema of the signal envelope and its rate of change, whereas a₂ relied on local minima. A further difference is that both a central and an occipital EEG channels are used in the band 11–16 Hz, and the spindle detection is a result of the combination of the two different sets of envelopes.

Recently, Warby et al. (2014) applied the six algorithms described so far in a large private database with sleep spindles from 110 healthy controls, and reported that the best algorithm in terms of accurately detecting spindles and minimizing false detections was a₅, closely followed by a₄. We note that all six algorithms described so far (a₁–a₆) rely on prior hypnogram assessment, which was provided given that the sleep stages assessed by experts was available for this database. We note that this fact effectively places competing algorithms which do not have access to direct algorithmic performance comparisons. The following new algorithms (a₇–a₈) do not rely on prior sleep staging information, but we aim to demonstrate that the new algorithms are nevertheless very competitive.

Novel Sleep Spindle Detection Algorithms

We have already highlighted the intuitively appealing features of the CWT for analyzing EEG signals due to its time-frequency localization properties, and the fact that it does not make assumptions regarding signal periodicity. Exploring the data by visual inspection of the true spindles (see **Figure 1**) seems to indicate that amplitude-based characteristics may be misleading (this is also implicit in the AASM criteria where no amplitude recommendation is made when assessing spindles); hence the primary focus of the developed algorithms is on the frequency content of the signal. Strictly speaking, we work directly with the CWT scales which correspond to the (pseudo)frequencies of interest (11–16 Hz). We defined 131 Morlet scales with a resolution of 0.1 in the range 2–15 (corresponding pseudo-frequencies: 5.4–40.6 Hz), which led to 24 scales lying within the spindle scale range. There is a non-linear mapping of the scales to their corresponding pseudo-frequencies, which is a function of the wavelet

basis function and the sampling frequency of the signal. For the Morlet wavelet with a signal sampling frequency of 100 Hz, the scales of interest (*spindle scale range*) are 5.1–7.4. We used a lower threshold of pseudo-frequency at 5.4 Hz above which we try to assess the probability of having a spindle so as to avoid challenging settings of spindles occurring on the background of large-amplitude slow oscillations (the delta frequency range, 1–4 Hz). Conceptually the starting basis of the proposed algorithms is similar to the study by Wamsley et al. (2012) (algorithm a₅), who subsequently thresholded the CWT coefficients at the spindle frequency range using a moving average of 100 ms sliding window. What distinguishes the algorithms proposed in this study compared to previous algorithms using the CWT is the different processing of the extracted Morlet CWT coefficients and the fact that we do not rely on expert-based hypnogram (in particular determining sleep stage 2) assessment.

Figure 2 presents a high-level flowchart of the two new algorithms introduced in this study. All sleep spindle detection algorithms developed in the research literature have some free parameters (typically these are some thresholds, e.g., on amplitude values). Similarly, the proposed algorithms in this study rely on a number of free parameters which need to be optimized: the chosen values were determined by testing on random subsamples of the training data so that regions of relative stability were found; exhaustive searches over the parameter space were not possible due to the size of the data set. We deliberately decided not to pursue rigorous optimization of these parameter values, in order to avoid overfitting the characteristics in the DREAMS database (effectively this would be training and testing on the same data). It is likely that the parameter values chosen could benefit from further refinement to optimize the outputs of the proposed algorithms, but a larger database would be needed.

Algorithm a₇, CWT instantaneous probabilistic estimate with moving averaging

The algorithm a₇, uses the following steps after the computation of the CWT coefficients:

- Computes the normalized percentage power of the CWT coefficients (henceforth referred to as *normalized coefficients*).
- Sorts the normalized coefficients in descending order at each time instant and works on the top 10 Morlet CWT scales corresponding to the top normalized coefficients (thus resulting in a matrix of size *number of signal samples* × 10).
- Computes instantaneous probabilistic estimate of spindle occurrence at the spindle scale range using the following algorithmic expression:

$$P(s_i) = \frac{1}{L} \cdot \sum_{i=1}^T (1./\langle \mathcal{M}_i \rangle)$$

where $P(s_i)$ denotes the probability of having a spindle at a given sample i , T is the cardinality of the top 10 scales corresponding to the sorted top 10 CWT normalized coefficients at instant i coinciding with the spindle scale range (i.e., for each sample i , we find how many of the top 10 sorted scales corresponding to

the normalized coefficients match the scales in the spindle scale range), $\langle \mathcal{M}_i \rangle$ contains the positions of the detected scales intersecting with the spindle scale range in the 10-element vector and the operator “./” denotes element-wise division. The value $P(s_i)$ effectively expresses the confidence that the sample i is part of a spindle (the higher the value, the more likely this sample may be part a spindle). The underlying concept is that if a sufficiently large number of successive samples (corresponding to some minimum time duration to be defined) have large probabilities denoting spindles, then that sequence will be denoted as a spindle. Effectively, we determine how many of the top 10 sorted scales matched the spindle scale range, and weigh these scales based on where they feature in the list with the instantaneous top 10 scales. If none of the sorted top 10 scales overlapped with the spindle scale range then $P(s_i)$ is zero. L denotes a normalization constant factor which was computed as $L = \sum_{i=1}^T (1./\langle 1 \dots 10 \rangle)$.

- Now, we need to smooth the instantaneous $P(s_i)$ estimates based on their K neighbors $\{P(s_{i-K/2}) \dots P(s_{i+K/2})\}$ to determine whether some EEG segments (regions) of arbitrary length within some duration boundaries (here 0.5–1.5 s) correspond to a spindle. Essentially, we have scales corresponding to the spindle scale frequency and we want to smooth neighboring regions to decide whether these are above the minimum duration threshold (in practice we very rarely have *all* consecutive samples in a spindle exhibiting large proportion of the scales belonging to the spindle scale range). Conceptually, this is similar to the concept that Wamsley et al. (2012) used, smoothing the data using a moving average of 0.1 s. Similarly, we used a moving average filter of 0.1 s to obtain the $P_{smooth}(s_i)$.
- It is possible that certain $P(s_i) < P_{smooth}(s_i)$ and we want to encourage relative large values to maximize the probability of detecting true spindles; hence we applied a final check: $P_{final}(s_i) = \max_{\forall i} (P_{smooth}(s_i), P(s_i))$.
- The candidate spindle instances (as a first pass) were detected at those samples when $P_{final}(s_i) > 0.3$ (for as many successive samples as the threshold remains valid). We remark this threshold (and all free parameters in this spindle detection algorithm such as number of top scales to investigate and K) were not rigorously optimized to avoid over-fitting the database used in this study. Instead we have attempted to determine “good” parameter values, which may be refined if presented with additional databases which will assist in properly optimizing the values of the free parameters.
- Finally, we need to group together regions which contain series of samples with high probabilities of denoting spindles. This was achieved using flags to denote if successive regions containing candidate spindles would group in terms of their proximity, average probabilistic estimate of having a spindle in a region defined between samples (i_1, \dots, i_2) $\{P(s_{i_1}) \dots P(s_{i_2})\}$, and the duration of the candidate spindle. Specifically, we grouped successive candidate spindles in the following cases:

- The duration between successive spindles was less than 0.3 s, and both successive spindles exhibited average

Input: Single-lead EEG (typically central EEG), sampling frequency

Output: Two-dimensional matrix with onset and offset of spindles (in terms of samples of the original EEG signal presented to the spindle detection algorithm)

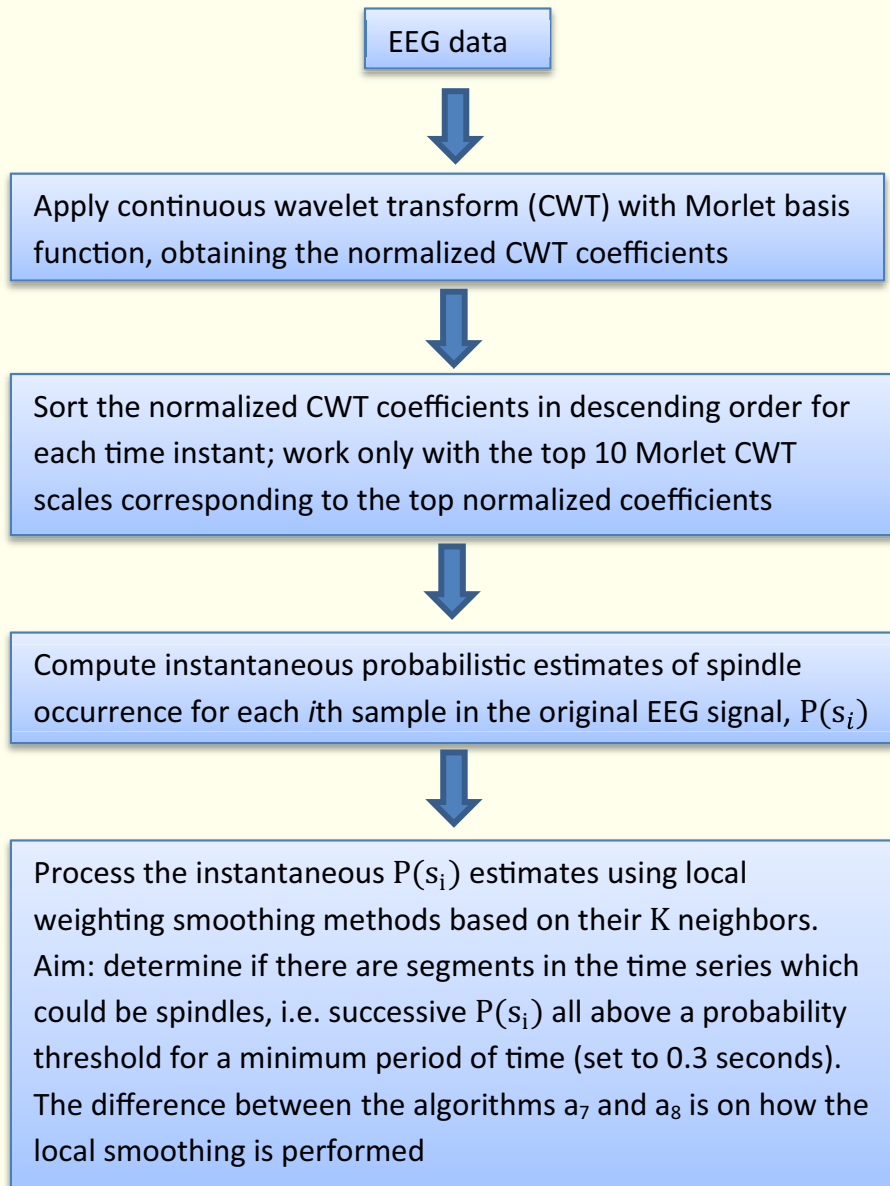


FIGURE 2 | Flowchart of the proposed algorithms in this study.

probabilistic strength above a threshold, i.e., both spindles appeared to be very likely true spindles: $\left(\frac{1}{i_2 - i_1} \cdot \sum_{i=i_1}^{i_2} P(s_i)\right) > 0.7$, and the duration of both

successive spindles was at least 0.1 s (case: “strong” spindles).

- (ii) The duration between successive spindles was less than 0.3 s and both successive spindles exhibited average

probabilistic strength: $\left(\frac{1}{i_2 - i_1} \cdot \sum_{i=i_1}^{i_2} P(s_i)\right) > 0.6$ and both were at least 0.3 s long (case: “long spindles”).

Algorithm a₈, CWT instantaneous probabilistic estimate with distance and amplitude weighted averaging

The algorithm a₈, is very similar to a₇. The difference lies in how we process the instantaneous probability spindle estimates $P(s_i)$ to affect neighboring $P(s_j)$ values. That is, the first steps (a)–(c) are identical, and step (d) processes the computed $P(s_i)$ using the exponential weighted moving average concept (instead of moving average). The underlying idea is that we want to update $P(s_i)$ values depending on their neighboring $P(s_j)$ values as a weighted function of their distances and a weighted function of their magnitude (which is weighted exponentially to promote EEG regions where instantaneous $P(s_i)$ estimates are large). Specifically, step (d) now becomes:

- (d) We used smoothing over 0.2 s, linearly scaling the effect of samples $P(s_j)$ on $P(s_i)$ as a function of their distance from $P(s_i)$, i.e., $\{w_t\}_{t=-10, t \neq 0} = \frac{1}{|t|} \cdot P(s_{i+t})$. In order to augment the effect of large $P(s_i)$ values (which denote great confidence that the sample i is part of a spindle) we exponentiated these values. Overall, conceptually it is similar to using an exponential weighted moving average approach. Algorithmically this is expressed as:

$$P_{smooth}(s_i) = \left[P(s_i) + \frac{1}{\sum_{t=-10, t \neq 0}^{10} w_t} \cdot \sum_{t=-10, t \neq 0}^{10} (\exp(P(s_{i+t})) - 1) \cdot w_t \right]$$

where the notation $[\cdot]$ denotes that the value is upper bounded to be 1, and the notation “ \cdot ” denotes element wise multiplication. The subsequent steps (e)–(g) are identical to a₇ to detect a spindle. We remark that a₈ is by design heavily weighting regions where there is a possibility of observing a spindle, but these regions will likely contain many cases which are not likely to be spindles.

Evaluation of Sleep Spindle Detection Algorithms

Both the DREAMS sleep spindles database and the MASS database have been annotated by two experts. Given the large inter-rater variability (e.g., for the DREAMS database the first rater has marked 289 spindles whereas the second rater has marked 409 spindles), there are two approaches to determine the ground truth. One approach is to only consider cases where both experts agree, an approach used previously for the DREAMS database by other researchers (Devuyst et al., 2011; Nonclercq et al., 2013). However, this biases the results, because one might argue that cases where both experts agree may denote “easily detectable” spindles; hence in this study we used all assessments by both experts, removing one of the double entries (in those cases where both experts agreed, in the DREAMS database we removed the assessment by the second expert because only

the first expert had also provided the duration of the assessed spindle).

Each of the sleep spindle algorithms used in this study results in estimates summarized in the format $N \times 2$, where N denotes the number of detected spindles for each EEG signal: the first column contains the estimated onset, and the second column the spindle duration. This facilitates direct comparison with the ground truth which is in the same format. In order to assess the performance and fairly compare all algorithms, we used the following commonly used metrics:

- True Positive Rate (TPR) (%), also known as *sensitivity*: $TPR = TP/(TP + FN)$ (is the proportion of spindles assessed by experts correctly identified by an algorithm, ideally we want this to be 100%).
- True Negative Rate (TNR) (%), also known as *specificity*: $TNR = TN/(TN + FP)$ (is the proportion of non-spindles assessed by experts correctly identified by an algorithm, ideally we want this to be 100%).

Specificity is also the complement of the False Positive Rate (FPR), defined as $FPR = FP/(FP + TN)$: $\text{specificity} = 100 - FPR$.

- False Discovery Rate (FDR): $FDR = FP/(TP + FP)$.
- Cohen’s kappa coefficient, where: $k = \frac{TP + TN - \Pr(e)}{1 - \Pr(e)}$, with $\Pr(e) = \frac{TP + FN}{N} \cdot \frac{TP + FP}{N} + \left(1 - \frac{TP + FN}{N}\right) \cdot \left(1 - \frac{TP + FP}{N}\right)$, and $N = TP + FP + TN + FN$

Cohen’s kappa coefficient was originally developed to assess inter-rater agreement, and some researchers suggest it takes into account agreement between raters which could be attributed to chance. Effectively, this implies that when raters are uncertain they guess about their decision, which some researchers have suggested is unlikely in many practical settings. Some of the problems and limitations of Cohen’s kappa have been discussed by Gwet (2008); we cautiously include it in this study because some research papers published in the sleep spindle detection literature have used it. We also used and put greater emphasis on the weighted kappa in this study because spindles are rare events in the EEG signal and we wanted to weigh accordingly for spindles correctly detected and spindles missed by the spindle detection algorithms (that is, we set the weight for TP and FN to be 10 times compared to the weight assigned to FP and TN).

- Absolute difference in the onset timings between the ground truth and the estimated onset.

where True Positive (TP) denotes agreement between the algorithm and the ground truth about the detection of a spindle, False Negative (FN) denotes a true spindle as assessed by the experts which was missed by the algorithm, False Positive (FP) when the algorithm detected a spindle that was not assessed as a spindle by the experts, and True Negative (TN) was defined as in Devuyst et al. (2011): $TN = \text{signal duration in seconds} - FP - TP - FN$. We assess a true positive when the absolute difference between the onset of the ground truth and the estimated spindle onset by the

algorithm is less than 0.5 s. Other studies have used different, less stringent definitions to assess whether an algorithm has matched the expert's assessment in correctly detecting a spindle. Some studies assess whether a spindle was detected within a sliding pre-specified time-interval (epoch), e.g., Duman et al. (2009), however this does not assess directly the accuracy in determining the spindle onset. Other studies, e.g., Nonclercq et al. (2013), consider that an algorithm has correctly detected a spindle if there was *any* overlap between the duration of the estimated spindle and the true spindle duration. However, this may positively bias sleep detection algorithms which provide spindle estimates with large durations.

Results

Evaluation of the Spindle Detection Algorithms on the DREAMS Sleep Spindles Database

Tables 1–3 summarize the performance of the sleep spindle detection algorithms used in this study for each of the eight signals. Ideally, a good algorithm exhibits large sensitivity and specificity, and low false discovery rate.

We observe relatively large deviations in the performance of the sleep spindle detection algorithms across the eight signals. Overall, the new algorithm a_7 exhibits large sensitivity and specificity. The more complicated new algorithm a_8 can accurately

TABLE 1 | Sensitivity (%) of the spindle detection algorithms across the eight EEG signals (higher values indicate better performance).

	Signal ₁	Signal ₂	Signal ₃	Signal ₄	Signal ₅	Signal ₆	Signal ₇	Signal ₈	Mean \pm std
a_1	70.6	56.6	53.3	40.6	45.6	78.6	27.8	75	56.0 \pm 17.8
a_2	14	3.90	11.1	9.40	20.4	29.1	16.7	10.4	14.4 \pm 7.7
a_3	86.7	68.8	84.4	42.2	95.1	91.5	77.8	75	77.7 \pm 16.8
a_4	46.7	63.6	77.3	32.8	63.1	68.4	61.1	50	57.9 \pm 14.0
a_5	12.5	49.4	84.4	31.3	15.5	64.1	55.6	47.9	45.1 \pm 24.4
a_6	79.3	85.7	77.8	45.3	81.6	81.2	72.2	83.3	75.8 \pm 12.9
a_7	84.4	80.5	73.3	65.6	70.9	66.7	88.9	77.1	75.9 \pm 8.3
a_8	89	80.5	82.2	68.8	96.1	88	77.8	83.3	83.2 \pm 8.2

The best performing algorithm for each case appears in bold.

TABLE 2 | Specificity (%) of the spindle detection algorithms across the eight EEG signals (higher values indicate better performance).

	Signal ₁	Signal ₂	Signal ₃	Signal ₄	Signal ₅	Signal ₆	Signal ₇	Signal ₈	Mean \pm std
a_1	85	79.8	83.9	82.9	82.6	85.2	80.7	79.1	82.4 \pm 2.3
a_2	99.6	100	99.6	98.8	99.2	99.4	98.9	99.1	99.3 \pm 0.4
a_3	91.1	97.6	75.1	92.7	88.5	77.8	89.1	39	81.4 \pm 18.7
a_4	98.5	98.3	97	96.5	98.2	98.6	95.6	94.3	97.1 \pm 1.6
a_5	99.8	99.2	96.1	96.3	99.6	98.8	97.1	96.1	97.9 \pm 1.6
a_6	86.6	67	87	87.1	91.1	92.5	82	79.5	84.1 \pm 8.1
a_7	94.6	93.4	94.5	87.3	95.5	97.3	94.1	78.1	91.8 \pm 6.3
a_8	78.6	76.3	77.8	68.1	80.9	86.6	75.5	55.7	74.9 \pm 9.4

The best performing algorithm for each case appears in bold.

TABLE 3 | False discovery rate (%) of the spindle detection algorithms across the eight EEG signals (lower values indicate better performance).

	Signal ₁	Signal ₂	Signal ₃	Signal ₄	Signal ₅	Signal ₆	Signal ₇	Signal ₈	Mean \pm std
a_1	72.3	89	92.2	91.9	86.3	73	98.6	91.1	86.8 \pm 9.4
a_2	26.9	0	58.3	77.8	38.2	22.7	87	75	48.2 \pm 31.0
a_3	56	44.2	92	82.4	66.7	77.7	93.3	96.7	76.1 \pm 19.0
a_4	28.4	38	60.5	74.4	32.3	23.1	87.6	80.6	53.1 \pm 25.7
a_5	19	25.5	64.2	76.5	27.3	21.9	83.9	74.7	49.1 \pm 28.1
a_6	67.6	89.6	86.7	88.5	64.3	57.2	96.1	90	80.0 \pm 14.6
a_7	44.1	64.6	74.4	84	51.3	37.1	86.9	91.2	66.7 \pm 20.7
a_8	74.6	86.8	91.3	92.6	76.6	68.6	96.9	95.1	85.3 \pm 10.6

The best performing algorithm for each case appears in bold.

detect more spindles than the competing approaches including a_7 (large sensitivity), at the cost of decreased specificity and increased false discovery rate. We have also evaluated the absolute difference in the onset timings between the ground truth and the estimated onset: this was fairly consistent amongst the algorithms with a mean absolute difference in onset timings ranging between 0.15 and 0.2 s and the standard deviation ranging between 0.11 and 0.15 s. Overall, all algorithms performed similarly with respect to correctly detecting onset spindle timing. We have emphasized that Cohen's kappa suffers from certain limitations (Gwet, 2008) and we use it here cautiously simply to facilitate comparisons with other studies in the research literature. Specifically the (unweighted) Cohen kappa was (mean \pm standard deviation): $a_1 = 0.15 \pm 0.12$, $a_2 = 0.19 \pm 0.11$, $a_3 = 0.29 \pm 0.22$, $a_4 = 0.46 \pm 0.20$, $a_5 = 0.37 \pm 0.19$, $a_6 = 0.25 \pm 0.18$, $a_7 = 0.40 \pm 0.20$, $a_8 = 0.18 \pm 0.14$.

Evaluation of the Spindle Detection Algorithms on the MASS Database

We have also evaluated the performance of all eight algorithms in terms of correctly detecting the sleep spindles in the MASS database. The results are summarized in **Table 4**. Interestingly, the findings in terms of sensitivity, specificity, and FDR are similar across the two databases used in this study. The algorithm a_7 outperforms the competing approaches in terms of sensitivity whilst being very competitive in terms of specificity. As indicated previously, we prefer the weighted Cohen kappa (see **Table 4**) penalizing more severely missed true spindles compared to false positives. Nevertheless, to facilitate direct comparisons with the research literature the unweighted Cohen kappa for the algorithms is also reported (mean \pm standard deviation): $a_1 = 0.20 \pm 0.11$, $a_2 = 0.22 \pm 0.04$, $a_3 = 0.28 \pm 0.24$, $a_4 = 0.51 \pm 0.13$, $a_5 = 0.38 \pm 0.18$, $a_6 = 0.37 \pm 0.18$, $a_7 = 0.24 \pm 0.12$, $a_8 = 0.16 \pm 0.09$.

Algorithmic Comparisons with Results Reported in the Research Literature

Many researchers have indicated that it is not easy to directly compare the performance of different algorithms across studies because of the different criteria used to detect spindles and assess the performance of the automated algorithms (Devuyst et al., 2011; Nonclercq et al., 2013). **Table 4** attempts to summarize many of these published findings in the research literature as an indicative reference, but we emphasize these results should be cautiously interpreted when comparing algorithms unless they have been tested on the same database using identical criteria to assess performance. **Table 5** summarizes the four performance metrics in this study (sensitivity, specificity, FDR, weighted Cohen's kappa) in terms of percentile scores, thus providing a good overview of the overall performance of each algorithm (including their behavior at extremes).

Discussion

This study revisited the problem of accurate and automatic detection of sleep spindles using a single EEG channel. We reviewed some indicative and widely used signal processing approaches

toward this aim, and highlighted some of the underlying problems. Two new signal processing approaches which are based on the CWT with Morlet basis were proposed and demonstrated to be very competitive against some commonly used algorithms found in the research literature. Interestingly, there was no universally best algorithm for all signals, although a_3 , a_6 , and a_7 appear to display relatively large sensitivity and specificity scores. We found that the new algorithm a_7 led to a range of 65.6–88.9% sensitivity scores and a range of 78.1–97.3% specificity scores for the DREAMS database, which compare favorably against competing approaches. The new algorithm a_8 exhibits higher sensitivity and lower specificity in the DREAMS database, on average, hence it might be more suitable primarily in cases where a human expert will post-process the estimates to determine whether the detected spindles correspond to true spindles. We re-iterate that the DREAMS sleep spindles database used in this study suffers from large inter-rater variability: the first rater has marked 289 spindles whereas the second rater has marked 409 spindles. Hence, the inter-rater agreement is lower than the agreement between raters reported in other studies (Huupponen et al., 2007), which may suggest automatic detection of spindles in this dataset may be challenging.

The original manuscript submission did not include the MASS database and hence the development of the spindle detection algorithm relied only on the DREAMS data. We have deliberately refrained from any additional fine-tuning of a_7 and a_8 to optimize performance in the MASS data, which might have potentially improved our reported results on the MASS database. It is reassuring that the proposed algorithms work very well on the MASS data, in particular a_7 . It is also encouraging to see that the results of sensitivity, specificity, FDR and weighted Cohen's kappa are similar across the two databases (see **Table 4**) for all algorithms: this inspires confidence regarding the objective merits of each algorithm, and may be a good indicator of the performance of the sleep spindle detection algorithms in new, unseen datasets. It is possible that other studies relying on a single database to develop and test their spindle detection algorithms might have over-trained on that particular dataset, so we find the reported findings on the MASS database (truly out-of-sample) to be particularly compelling. **Table 5** provides an overall summary of performance of the sleep spindle algorithms on both databases, including extremes (i.e., the algorithms at their worst and at their best) by reporting percentile values. We note that a_7 in particular is very competitive across the entire range of the distribution of performances, particularly for the MASS database (and interestingly, exhibiting good performance even for the 5th and 25th percentiles, i.e., it is fairly stable across individuals compared to many of the competing algorithms).

For reference purposes we have summarized the findings of multiple sleep spindle studies in the research literature in **Table 4**. However, direct comparison of findings across studies in this application is not straightforward for a number of reasons: (a) many studies rely solely on data stemming from healthy controls which are arguably easier to analyze than data from pathological cohorts (or process EEG artifact-free data, whereas the DREAMS sleep spindle database used here contains data from various sleep disorders), (b) the criteria for identifying sleep

TABLE 4 | Summary of automated spindle detection results in the research literature and in this study.

Study	Spindle assessment				Participants and data collected	Database	Algorithm requires hypnogram	Spindle detector TP evaluation
	Sensitivity (%)	Specificity (%)	FDR (%)	Weighted Cohen kappa				
Schonwald et al., 2006	81.2	81.2	N/R	N/R	9 healthy adults, extracted 24 segments from each subject using 20 s epochs, removed epochs with artifacts	Private (N = 9)	Yes	Second-by-second analysis
Huupponen et al., 2007	70.0	98.6	32	N/R	12 healthy adults, entire night recordings	Private (N = 12)	Yes	The absolute difference between the detected spindle onset and the spindle onset determined by the experts was less than 0.5 s.
Causa et al., 2010	88.2	89.7	11.9	N/R	56 healthy children overnight recordings, 27 recordings used for training, 10 recordings for validation, and 19 for testing performance	Private (N = 56)	No	At least 75% spindle duration overlap between detected and expert assessed spindle
Warby et al. (2014) applying a ₁	74	81	89	N/R	110 healthy adults, (4 min of artifact-free stage 2 sleep from 100 subjects and ~38 min of stage 2 sleep from 10 subjects)	Private (N = 110)	Yes	At least 20% spindle duration overlap between detected and expert assessed spindle
Warby et al. (2014) applying a ₂	17	99	48	N/R	See above entry	Private (N = 110)	Yes	See above entry
Warby et al. (2014) applying a ₃	71	81	89	N/R	See above entry	Private (N = 110)	Yes	See above entry
Warby et al. (2014) applying a ₄	43	98	58	N/R	See above entry	Private (N = 110)	Yes	See above entry
Warby et al. (2014) applying a ₅	33	99	44	N/R	See above entry	Private (N = 110)	Yes	See above entry
Warby et al. (2014) applying a ₆	57	96	70	N/R	See above entry	Private (N = 110)	Yes	See above entry
Devuyst et al., 2011	70.2	98.6	N/R	N/R	8 diagnosed with various sleep disorders (30 min segments), two raters for all signals; one rater only for two signals. Use only six signals and only cases where raters agree	DREAMS sleep spindle database (publicly available) (N = 6)	No	N/R
Nonclercq et al., 2013	75.1	96.7	N/R	N/R	See above entry	DREAMS (N = 6)	No	There is overlap between the duration of the detected spindle and the spindle duration assessed by experts
Present study a ₁	56.0	82.4	86.8	0.37	8 from various sleep disorders (30 min segments), two raters for all signals; one rater only for two signals. Use all eight signals including “difficult” cases where raters do not agree	DREAMS (N = 8)	Yes	The absolute difference between the detected spindle onset and the spindle onset determined by the experts was less than 0.5 s
Present study a ₂	14.4	99.3	48.2	0.17	See above entry	DREAMS (N = 8)	Yes	See above entry
Present study a ₃	77.7	81.4	76.1	0.55	See above entry	DREAMS (N = 8)	Yes	See above entry

(Continued)

TABLE 4 | Continued

Study	Spindle assessment				Participants and data collected	Database	Algorithm requires hypnogram	Spindle detector TP evaluation
	Sensitivity (%)	Specificity (%)	FDR (%)	Weighted Cohen kappa				
Present study a ₄	57.9	97.1	53.1	0.59	See above entry	DREAMS (N = 8)	Yes	See above entry
Present study a ₅	45.1	97.9	49.1	0.47	See above entry	DREAMS (N = 8)	Yes	See above entry
Present study a ₆	75.8	84.1	80.0	0.55	See above entry	DREAMS (N = 8)	Yes	See above entry
Present study a ₇	75.9	91.8	66.7	0.66	See above entry	DREAMS (N = 8)	No	See above entry
Present study a ₈	83.2	74.9	85.3	0.50	See above entry	DREAMS (N = 8)	No	See above entry
Present study a ₁	65.5	85.1	82.7	0.46	19 overnight PSG from healthy controls; two raters for 15 signals, one rater for four signals	MASS database S2 (publicly available) (N = 19)	Yes	See above entry
Present study a ₂	16.5	99.2	49.5	0.20	See above entry	MASS (N = 19)	Yes	See above entry
Present study a ₃	73.5	78.2	75.3	0.46	See above entry	MASS (N = 19)	Yes	See above entry
Present study a ₄	66.2	97.5	48.1	0.64	See above entry	MASS (N = 19)	Yes	See above entry
Present study a ₅	41.3	98.8	45.3	0.43	See above entry	MASS (N = 19)	Yes	See above entry
Present study a ₆	73.0	90.5	69.1	0.60	See above entry	MASS (N = 19)	Yes	See above entry
Present study a ₇	83.8	90.2	82.6	0.64	See above entry	MASS (N = 19)	No	See above entry
Present study a ₈	77.2	76.9	86.5	0.46	See above entry	MASS (N = 19)	No	See above entry

Sensitivity (%) = $TP/(TP + FN)$, Specificity (%) = $TN/(TN + FP)$, False Discovery Rate (FDR) (%) = $FP/(TP + FP)$. TP stands for true positive, TN for true negative, FP for false positive, and FN for false negative. The last column briefly explains the method used to assess how the automatic sleep spindle detector was deemed to succeed in detecting the spindle as registered by the experts. See Section Evaluation of Sleep Spindle Detection Algorithms for more details.

spindles are inconsistent, (c) different research teams use slightly different definitions of spindles, (d) in some cases researchers have only reported the detection accuracy but have not provided details about the number of erroneous detections, therefore making comparison against some conservative approaches (algorithms which aim to minimize the number of falsely reported spindles) unfair. For all these reasons, probably the most efficient and appropriate scientific approach is to apply multiple sleep spindle detection algorithms across multiple datasets and directly compare their performance. Causa et al. (2010) have reported better sensitivity (88.2%) and specificity scores (89.7%) compared to results in other studies (including the current study). However, that study focused only on healthy children, and those findings might not be generalizable to studies focusing on other cohorts (healthy adults, and adults diagnosed with a sleep-related disorder). Two prior studies have focused on the DREAMS sleep spindle database which facilitate comparison of findings: Devuyst et al. (2011) reported sensitivity score 70.2% and specificity score 98.6%. Likewise, Nonclercq et al. (2013) reported sensitivity scores ranging between 65.8 and 82.8% and specificity scores ranging between 96.7 and 98.7% for the first six signals in the

database. However, we note that in both studies the authors used as ground truth only those cases where the experts agreed on the first six signals, which potentially biases the results (spindles detected by either one of the raters are probably borderline and more difficult to assess, but on the other hand are probably also more interesting). Similarly, the MASS database is a new publicly available database and we anticipate future studies will benchmark algorithms against this database.

Ideally, a sleep spindle detection algorithm should correctly detect all true spindles without indicating the presence of additional (erroneous) spindles (an artifact or other class of event erroneously considered to be spindle). In practice, there is a tradeoff compromising between maximizing the detection of true spindles (true positive rate) and minimizing the false assessment of EEG segments as spindles. Essentially this is the case with the closely related algorithms a₇ and a₈ proposed in this study. The algorithm a₈ can typically correctly detect more spindles than a₇ at the cost of increasing the number of falsely detected spindles (increased false discovery rate). We note that a₆ and a₃ are similarly more prone compared to competing algorithms to decide that spindles have occurred in the EEG signal: this causes their

TABLE 5 | Summary of statistics (percentiles) of the performance metrics of the spindle detection algorithms for the DREAMS and MASS databases.

	Sensitivity (%)					Specificity (%)					FDR (%)					Weighted Cohen kappa				
	5	25	50	75	95	5	25	50	75	95	5	25	50	75	95	5	25	50	75	95
a ₁	27.8	43.1	54.9	72.8	78.6	79.1	80.3	82.8	84.5	85.2	72.3	79.6	90	92.1	98.6	0.06	0.27	0.36	0.51	0.64
	54.1	60.8	65.3	69.3	80.84	82.1	83.7	85.3	86.4	88.4	66.3	76.3	80.6	90.9	97.7	0.22	0.43	0.49	0.54	0.63
a ₂	3.9	9.9	12.6	18.6	29.1	98.8	99.0	99.3	99.6	100	0	24.8	48.3	76.4	87.0	0.05	0.13	0.15	0.23	0.32
	10.9	13.0	14.6	17.5	30.1	98.9	98.9	99.2	99.4	99.6	33.8	41.8	43.9	64.2	67.0	0.12	0.15	0.18	0.20	0.39
a ₃	42.2	71.9	81.1	89.1	95.1	39	76.5	88.8	91.9	97.6	44.2	61.3	80	92.7	96.7	0.08	0.43	0.58	0.75	0.81
	34.5	58.1	81.7	88.8	91.6	39.4	64.2	83.9	93.4	97.0	35.8	58.7	76.8	96	98.7	0	0.10	0.62	0.75	0.82
a ₄	32.8	48.4	62.1	66.0	77.3	94.3	96.1	97.6	98.4	98.6	23.1	30.4	49.2	77.5	87.6	0.36	0.49	0.61	0.69	0.77
	41.2	56.2	64.8	77.6	96.2	95.7	97.2	97.6	98.2	98.7	23.5	33.2	43.7	64.4	88.5	0.40	0.58	0.68	0.73	0.82
a ₅	12.5	23.4	48.7	59.9	84.4	96.1	96.2	97.9	99.4	99.8	19.0	23.7	45.7	75.6	83.9	0.13	0.26	0.54	0.61	0.81
	3.5	24.7	39.6	48.8	91.4	97.1	98.5	98.9	99.5	99.7	20.6	29.8	39.7	58.8	82.0	0.040	0.35	0.43	0.54	0.78
a ₆	45.3	75.0	80.3	82.5	85.7	67.0	80.8	86.8	89.1	92.5	57.2	65.9	87.6	89.8	96.1	0.32	0.40	0.55	0.70	0.75
	52.4	69.8	72.7	76.0	92.79	76.7	85.9	92.8	95.2	97.4	45.7	55.6	66.1	80.7	97.0	0.23	0.60	0.65	0.69	0.74
a ₇	65.6	68.8	75.2	82.5	88.9	78.1	90.4	94.3	95.1	97.3	37.1	47.7	69.5	85.5	91.2	0.46	0.60	0.69	0.72	0.80
	64.7	80.1	86.3	89.6	92.9	83.6	88.1	90.1	94.1	95.9	51.3	81.1	85.7	90.6	92.3	0.49	0.60	0.64	0.70	0.74
a ₈	68.8	79.2	82.8	88.5	96.1	55.7	71.8	77.1	79.8	86.6	68.6	75.6	89.1	93.9	96.9	0.26	0.29	0.50	0.70	0.74
	65.1	72.7	79.3	82.2	87	67.6	72.6	76.2	81.1	86.5	71.0	83.0	86.8	92.2	97.6	0.24	0.36	0.49	0.58	0.63

The first row for each algorithm a₁–a₈ corresponds to the (5,25,50,75,95) percentiles in the DREAMS database, and the second row to the percentiles in the MASS database.

true positive rate to be generally higher at the cost of additional false positives. O'Reilly and Nielsen (2014b) envisage that “*most probably, manual [sleep spindle] scoring will progress toward semi-automation benefitting from further advances in signal processing*” an assertion we find plausible. In that sense, if sleep spindle assessment is performed semi-automatically (prior assessment by an algorithm and subsequent checking by an expert) it is beneficial to correctly detect as many spindles as possible, even at the cost of erroneously recording spindles (i.e., increasing sensitivity at the cost of an increased false positive rate). There is probably no universal solution to this problem, and the sensitivity trade-off might need to be a free parameter of sleep spindle algorithms which could be appropriately adjusted by the operator of the algorithm.

We remark that some of the sleep spindle detection algorithms used in this study require more than a single-EEG channel to detect spindles. For example, a₁ and a₆ require the use of an additional EEG channel, and a₁–a₅ need to be presented with the hypnogram assessment (moreover the algorithm a₅ explicitly requires stage 2 assessments). We emphasize again that the proposed algorithms in this study (a₇ and a₈) have minimal requirements in terms of the input data in order to detect spindles: a single EEG channel is sufficient. Therefore, we argue that these new algorithms may be more readily deployable on databases which have not been scored by experts prior to sleep spindle estimation (no sleep staging requirement). Nevertheless, future studies could further explore whether the use of

additional EEG channels and/or hypnogram might increase the sleep spindle detection accuracy.

A critical aspect for comparing algorithms in this application is the definition of TP, TN, FP, FN. In some studies it is not explicitly clear how authors deemed that the automated sleep spindle detector has matched the assessment of an expert in correctly identifying a sleep spindle. There is no clear consensus in the research literature currently; the last column in **Table 4** summarizes some of the different approaches that have been used. We agree with Causa et al. (2010) who criticize other studies that the criteria used for algorithmic assessment are not made explicit, and would encourage other researchers to meticulously report the methodology followed to mark their assessments; ideally this methodology should be standardized to facilitate direct comparisons of algorithmic concepts.

Inspection of the results revealed that different sleep spindle detection algorithms have the potential to detect different spindles under different conditions. This would suggest that exploring some data fusion approaches might have good potential in this application. Data fusion in conceptually related applications (combining the outputs of multiple signal processing algorithms which estimate some property of the signal) has shown great promise (Mitchell, 2012; Tsanas et al., 2014; Zhu et al., 2014). In fact, simple combination approaches of the first six sleep spindle detection algorithms used in this study have been previously explored by Warby et al. (2014) but the authors did not report any significant improvement over the single best algorithm;

future studies could further explore some principled data fusion frameworks in this application.

Acknowledgments

This study was supported by the Wellcome Trust through a Centre Grant No. 098461/Z/12/Z, “The University of Oxford Sleep and Circadian Neuroscience Institute (SCNi).” The first database used in this study (DREAMS sleep spindles database) is publicly available from the University of MONS—TCTS Laboratory (Stéphanie Devuyst, Thierry Dutoit) and Université Libre de Bruxelles—CHU de Charleroi Sleep Laboratory

(Myriam Kerkhofs). The second database used in this study (MASS database) was a collaborative effort led by Christian O’Reilly and colleagues at the University of Montreal (the Montreal Archive of Sleep Studies). We want to thank both research teams for making their datasets publicly available. AT is particularly grateful to Christian O’Reilly for all his help during this project.

Supplementary Material

Documented Matlab source code is available from the first author’s website, and from www.physionet.org.

References

- Addison, P. (2002). *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance*. Bristol: CRC press.
- Astill, R. G., Piantoni, G., Raymann, R. J. E. M., Vis, J. C., Coppens, J. E., Walker, M. P. et al. (2015). Sleep spindle and slow wave frequency reflect motor skill performance in primary school-age children. *Front. Hum. Neurosci.* 8:910. doi: 10.3389/fnhum.2014.00910
- Bódizs, R., Körmendi, J., Rigó, P., and Lázár, A. S. (2009). The individual adjustment method of sleep spindle analysis: methodological improvements and roots in the fingerprint paradigm. *J. Neurosci. Methods* 178, 205–213. doi: 10.1016/j.jneumeth.2008.11.006
- Causa, L., Held, C. M., Causa, J., Estevez, P. A., Perez, C. A., Chamorro, R., et al. (2010). Automated sleep spindle detection in healthy children polysomnograms. *IEEE Trans. Biomed. Eng.* 57, 2135–2146. doi: 10.1109/TBME.2010.2052924
- Christensen, J. A. E., Kempfner, J., Zoetmulder, M., Leonthin, H. L., Arvastsonm, L., Christensen, S., et al. (2014). Decreased sleep spindle density in patients with idiopathic REM sleep behavior disorder and patients with Parkinson’s disease. *Clin. Neurophysiol.* 125, 512–519. doi: 10.1016/j.clinph.2013.08.013
- DeGennaro, L., and Ferrara, M. (2003). Sleep spindles: an overview. *Sleep Med. Rev.* 7, 423–440. doi: 10.1053/smr.2002.0252
- Devuyst, S., Dutoit, T., Stenuit, P., and Kerkhofs, M. (2011). Automatic sleep spindles detection—overview and development of a standard proposal assessment method. *Proc. IEEE Eng. Med. Biol. Soc.* 2011, 1713–1716. doi: 10.1109/IEMBS.2011.6090491
- Duman, F., Erdamar, A., Eroglu, O., Telatar, Z., and Yetkin, S. (2009). Efficient sleep spindle detection algorithm with decision tree. *Exp. Syst. Appl.* 36, 9980–9985. doi: 10.1016/j.eswa.2009.01.061
- Ferrarelli, F., Huber, R., Peterson, M. J., Massimini, M., Murphy, M., Riedner, B. A., et al. (2007). Reduced sleep spindle activity in schizophrenia patients. *Am. J. Psychiatry* 164, A62. doi: 10.1176/appi.ajp.164.3.483
- Grove, W. M., and Mehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: the clinical–statistical controversy. *Psychol. Public Policy Law* 2, 293–323. doi: 10.1037/1076-8971.2.2.293
- Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *Br. J. Math. Stat. Psychol.* 61, 29–48. doi: 10.1348/000711006X126600
- 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
- Iber, C., Ancoli-Israel, S., Chesson, A. L., 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.
- Kryger, M. H., Roth, T., and Dement, W. C. (2010). *Principles and Practice of Sleep Medicine, 5th Edn*. Elsevier Saunders.
- Martin, N., Lafortune, 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
- Mitchell, H. B. (2012). *Data Fusion: Concepts and Ideas, 2nd Edn*. Berlin: Springer.
- Möller, 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.
- Nonclercq, A., Urbain, C., Verheulpen, D., Decaestecker, C., Van Bogaert, P., and Peigneux, P. (2013). Sleep spindle detection through amplitude-frequency normal modeling. *J. Neurosci. Methods* 214, 192–203. doi: 10.1016/j.jneumeth.2013.01.015
- O’Reilly, C., Gosselin, N., Carrier, J., and Nielsen, T. (2014a). 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. (2014b). 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
- Pardey, J., Roberts, S., and Tarassenko, L. (1996). A review of parametric modeling techniques for EEG analysis. *Med. Eng. Phys.* 18, 2–11. doi: 10.1016/1350-4533(95)00024-0
- Rechtschaffen, A., and Kales, A. (1968). *A Manual of Standardized Terminology, Techniques and Scoring System For Sleep Stages of Human Subjects*. Washington, DC: US Dept. of Health, Education, and Welfare; National Institutes of Health.
- Schmicek, P., Zeithofer, J., Anderer, P., and Saletu, B. (1994). Automatic sleep-spindle detection procedure: aspects of reliability and validity. *Clin. Electroencephalogr.* 25, 26–29. doi: 10.1177/155005949402500108
- Schmidt, C., Peigneux, P., Muto, V., Schenkel, M., Knoblauch, V., Münch, M., et al. (2006). Encoding difficulty promotes postlearning changes in sleep spindle activity during napping. *J. Neurosci.* 26, 8976–8982. doi: 10.1523/JNEUROSCI.2464-06.2006
- Schonwald, S. V., de Santa-Helena, E. L., Rossatto, R., Chaves, M. L., and Gerhardt, G. J. (2006). Benchmarking matching pursuit to find sleep spindles. *J. Neurosci. Methods* 156, 314–321. doi: 10.1016/j.jneumeth.2006.01.026
- Seshadrinathan, K., Soundararajan, R., Bovik, A. C., and Cormack, L. K. (2010). Study of subjective and objective quality assessment of video. *IEEE Trans. Image Processing* 19, 1427–1441. doi: 10.1109/TIP.2010.2042111
- Sitnikova, E., Hramov, A. E., Koronovsky, A. A., and van Luijtelaar, G. (2009). Sleep spindles and spike-wave discharges in EEG: their generic features, similarities and distinctions disclosed with Fourier transform and continuous wavelet analysis. *J. Neurosci. Methods* 180, 304–316. doi: 10.1016/j.jneumeth.2009.04.006
- Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., and Gareth Gaskell, M. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *J. Neurosci.* 30, 14356–14360. doi: 10.1523/JNEUROSCI.3028-10.2010
- Tsanas, A. (2012). *Accurate Telemonitoring of Parkinson’s Disease Symptom Severity Using Nonlinear Speech Signal Processing and Statistical Machine Learning*. Ph.D. thesis, University of Oxford, UK.
- Tsanas, A., Zañartu, M., Little, M. A., Fox, C., Ramig, L. O., and Clifford, G. D. (2014). Robust fundamental frequency estimation in sustained vowels: detailed

- algorithmic comparisons and information fusion with adaptive Kalman filtering. *J. Acoust. Soc. Am.* 135, 2885–2901. doi: 10.1121/1.4870484
- Wamsley, E. J., Tucker, M. A., Shinn, A. K., Ono, K. E., McKinley, S. K., Ely, A. V., et al. (2012). Reduced sleep spindles and spindle coherence in schizophrenia: mechanisms of impaired memory consolidation? *Biol. Psychiatry* 71, 154–161. doi: 10.1016/j.biopsych.2011.08.008
- Warby, S. C., Wendt, S. L., Welinder, P., Munk, E. G. S., Carrillo, O., Sorensen, H. B. D., 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., Christensen, J. A. E., Kempfner, J., Leonthin, H. L., Jennum, P., and Sorensen, H. B. D. (2012). Validation of a novel automatic sleep spindle detector with high performance during sleep in middle aged subjects. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2012, 4250–4253. doi: 10.1109/EMBC.2012.6346905
- 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
- Zhu, T., Johnson, A. E. W., Behar, J., and Clifford, G. D. (2014). Crowd-sourced annotation of ECG signals using contextual information. *Ann. Biomed. Eng.* 42, 871–884. doi: 10.1007/s10439-013-0964-6
- 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.

Copyright © 2015 Tsanas and Clifford. 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.

Corrigendum: A comparison of two sleep spindle detection methods based on all night averages: individually adjusted vs. fixed frequencies

Péter P. Ujma¹, Ferenc Gombos², Lisa Genzel³, Boris N. Konrad⁴, Péter Simor^{5,6}, Axel Steiger⁴, Martin Dresler^{4,7*} and Róbert Bódizs^{1,2}

¹ Institute of Behavioral Science, Semmelweis University, Budapest, Hungary, ² Department of General Psychology, Pázmány Péter Catholic University, Budapest, Hungary, ³ Centre for Cognitive and Neural Systems, University of Edinburgh, Edinburgh, UK, ⁴ Max Planck Institute of Psychiatry, Munich, Germany, ⁵ Department of Cognitive Sciences, Budapest University of Technology and Economics, Budapest, Hungary, ⁶ Nyíró Gyula Hospital, National Institute of Psychiatry and Addictions, Budapest, Hungary, ⁷ Donders Institute, Radboud University Medical Centre, Nijmegen, Netherlands

Keywords: sleep spindles, EEG, individual adjustment method, IAM, algorithm

A Corrigendum on

A comparison of two sleep spindle detection methods based on all night averages: individually adjusted vs. fixed frequencies

by Ujma, P. P., Gombos, F., Genzel, L., Konrad, B. N., Simor, P., Steiger, A., et al. (2015). *Front. Hum. Neurosci.* 9:52. doi: 10.3389/fnhum.2015.00052

OPEN ACCESS

Edited and reviewed by:

Simon C. Warby,
Stanford University, USA

*Correspondence:

Martin Dresler,
dresler@mpipsykl.mpg.de

Received: 01 June 2015

Accepted: 06 July 2015

Published: 21 July 2015

Citation:

Ujma PP, Gombos F, Genzel L, Konrad BN, Simor P, Steiger A, Dresler M and Bódizs R (2015) Corrigendum: A comparison of two sleep spindle detection methods based on all night averages: individually adjusted vs. fixed frequencies. *Front. Hum. Neurosci.* 9:415. doi: 10.3389/fnhum.2015.00415

The description of the Individual Adjustment Method (IAM) algorithm for sleep spindle analyses (Ujma et al., 2015) contained an error, which we hereby rectify. On page 5, line 7–8, instead of $f(x) = e^{-(x - xm)/(w/2)}$, the equation should read as follows:

$$f(x) = e^{-(((x - xm)/(w/2))^2)}$$

References

Ujma, P. P., Gombos, F., Genzel, L., Konrad, B. N., Simor, P., Steiger, A., et al. (2015). A comparison of two sleep spindle detection methods based on all night averages: individually adjusted vs. fixed frequencies. *Front. Hum. Neurosci.* 9:52. doi: 10.3389/fnhum.2015.00052

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.

Copyright © 2015 Ujma, Gombos, Genzel, Konrad, Simor, Steiger, Dresler and Bódizs. 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.



A comparison of two sleep spindle detection methods based on all night averages: individually adjusted vs. fixed frequencies

Péter Przemysław Ujma¹, Ferenc Gombos², Lisa Genzel³, Boris Nikolai Konrad⁴, Péter Simor^{5,6}, Axel Steiger², Martin Dresler^{4,7*} and Róbert Bódizs^{1,2}

¹ Institute of Behavioral Science, Semmelweis University, Budapest, Hungary

² Department of General Psychology, Pázmány Péter Catholic University, Budapest, Hungary

³ Centre for Cognitive and Neural Systems, University of Edinburgh, Edinburgh, UK

⁴ Department of Clinical Research, Max Planck Institute of Psychiatry, Munich, Germany

⁵ Department of Cognitive Sciences, Budapest University of Technology and Economics, Budapest, Hungary

⁶ Nyró Gyula Hospital, National Institute of Psychiatry and Addictions, Budapest, Hungary

⁷ Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Centre, Nijmegen, Netherlands

Edited by:

Simon C. Warby, Stanford University, USA

Reviewed by:

Thien Thanh Dang-Vu, Concordia University, Canada

Suzana Schonwald, Hospital de Clínicas de Porto Alegre, Brazil
Dennis Angel Dean, Brigham and Women's Hospital and Harvard Medical School, USA

*Correspondence:

Martin Dresler, Department of Clinical Research, Max Planck Institute of Psychiatry, Kraepelinstraße 2-10, 80804 München, Germany
e-mail: dresler@mpipsykl.mpg.de

Sleep spindles are frequently studied for their relationship with state and trait cognitive variables, and they are thought to play an important role in sleep-related memory consolidation. Due to their frequent occurrence in NREM sleep, the detection of sleep spindles is only feasible using automatic algorithms, of which a large number is available. We compared subject averages of the spindle parameters computed by a fixed frequency (FixF) (11–13 Hz for slow spindles, 13–15 Hz for fast spindles) automatic detection algorithm and the individual adjustment method (IAM), which uses individual frequency bands for sleep spindle detection. Fast spindle duration and amplitude are strongly correlated in the two algorithms, but there is little overlap in fast spindle density and slow spindle parameters in general. The agreement between fixed and manually determined sleep spindle frequencies is limited, especially in case of slow spindles. This is the most likely reason for the poor agreement between the two detection methods in case of slow spindle parameters. Our results suggest that while various algorithms may reliably detect fast spindles, a more sophisticated algorithm primed to individual spindle frequencies is necessary for the detection of slow spindles as well as individual variations in the number of spindles in general.

Keywords: EEG, sleep spindles, sigma waves, automatic detections, fixed frequency method, IAM, comparison

INTRODUCTION

Sleep spindles are oscillations emerging from interacting thalamocortical, corticothalamic, and reticular networks in NREM sleep (Steriade and Deschenes, 1984; Amzica and Steriade, 2000; Steriade, 2000; Fogel and Smith, 2011), which are thought to play an important role in sleep-related brain plasticity (Genzel et al., 2014). Due to their trait-like nature and relationship to plasticity, sleep spindles are frequently studied as candidate indexes of individual variations in cognitive performance. Sleep spindles are remarkably individual features: sleep spindle parameters are characterized by high intra-individual stability and inter-individual variability (De Gennaro et al., 2005), a strong genetic background (De Gennaro et al., 2008), and a correlation with anatomical properties of the brain (Piantoni et al., 2013; Saletin et al., 2013).

Due to their high prevalence and specific signal properties automatic detection methods have proven to be viable and preferable alternatives to visual detection. Some of the earliest studies (Broughton et al., 1978; Campbell et al., 1980) used phase-locked loop devices for automatic sleep spindle detection and already reported an adequate agreement with visual detection.

An early combined software-hardware system (Ferri et al., 1989) also reliably replicated visual spindle detection results. Software solutions for automatic spindle detection were introduced somewhat later (Schimicek et al., 1994) and reported relatively high (approx. 70%) specificity for 90% sensitivity, while an improved method (Devuyst et al., 2006) could increase this to almost 76% in a clinical sample. More recently, sophisticated automatic sleep spindle detection methods using artificial neural networks (Acir and Güzelis, 2004; Ventouras et al., 2005) and decision trees (Duman et al., 2009) reached even higher performance, with correct classification frequently exceeding 90%.

Automatic sleep spindle recognition was further refined by adapting algorithms that take into account the inter-individual differences in sleep spindle activity, which vastly exceed intra-individual variation (De Gennaro et al., 2005) and emerge—among others—as a function of age and sex (Driver et al., 1996; Carrier et al., 2001; Huupponen et al., 2002; Genzel et al., 2012). Sleep spindle detection methods have been developed to operate with individually adjusted amplitude limits (Huupponen et al., 2000, 2007; Ray et al., 2010). A novel algorithm (Bódizs et al.,

2009; Ujma et al., 2014) based on the electrophysiological fingerprint theory of human sleep (De Gennaro et al., 2005, 2008) is the Individual Adjustment Method (IAM), which takes into account inter-individual variations not only in the amplitude, but also in the frequency of sleep spindles. In the IAM, sleep spindles are therefore not only detected based on individual amplitude thresholds, but also within the exact frequency bands where they are present in a given individual. A similarly adaptive detection method (based on a probabilistic model) is reported in Nonclercq et al. (2013).

A comparison of four different spindle detection methods (Huupponen et al., 2007) reported acceptable, but not overwhelming concordance. A recent study (Warby et al., 2014) investigated the agreement in spindle detection between expert human raters, non-experts recruited in an internet crowdsourcing effort, and automatic detection algorithms. Concordance was strongest among human experts, followed by non-experts operation in a crowdsourcing scheme, and weakest among automatic algorithms.

While the progress in automatic sleep spindle detection methods is impressive, there are numerous concerns which must be addressed in this field. A practical criticism may arise from the fact that automatic sleep spindle detections are frequently validated against visual detections: however, agreement in the visual scoring of spindles is not perfect (Campbell et al., 1980; Warby et al., 2014), the visual detection of spindles is often considered as a consensus from several raters which may bias results (Ray et al., 2010), and—despite stronger agreement among human raters than algorithms (Warby et al., 2014)—the use of human expert opinion as an absolute gold standard is philosophically questionable in itself (Bódizs et al., 2009).

Further criticism must be given to the fact of the use of standard signal detection terminology (such as sensitivity and specificity) in case of sleep spindle detection algorithms. Sleep spindles are frequent phenomena, but even so the vast majority of a sleep EEG recording does not consist of sleep spindles. Therefore, correct negative classifications are by far the most common result produced by any sleep spindle detector, which might drastically inflate specificity. The ratio of correct hits and false detections—including misses and false positives—would be a much more conservative, but also more informative measure of detection performance.

Sleep spindles are not only biological signals, but important markers of individual traits (De Gennaro et al., 2005, 2008) as well as powerful correlates of human cognition (among others: Bódizs et al., 2005; Schabus et al., 2006; Fogel et al., 2007; Ujma et al., 2014). Therefore, an alternative option in order to assess detection algorithms would be to investigate how much they can reproduce trait-like individual averages (instead of comparing individual spindle detections).

To our knowledge, it has never been investigated how strongly spindle measures of different detection methods are correlated if not individual spindle detections, but subject averages are considered. This can evidently not be predicted from the signal detection characteristics of the comparison of individual spindle detections of various methods—albeit the literature usually reports moderate agreement between the individual spindle detections

of different algorithms, it is unknown whether the different spindle samples obtained by different methods approximate the same individual averages. Therefore, the aim of our study was to reveal the correlation between individual sleep spindle parameters calculated with two different detection methods.

MATERIALS AND METHODS

SUBJECTS

We examined polysomnographic data of 161 healthy volunteers (88 males, 73 females, age between 17 years and 69 years, mean age 29.4 years, StD 10.7 years) recorded on the second night spent in a sleep laboratory. All procedures were approved by the responsible institution's ethical board and subjects gave informed consent. A semi-structured interview excluded any history of neurologic or psychiatric disease, but six subjects suffered from frequent nightmares. Subjects were free of drugs and prescription medication (except for contraceptives, all data self-reported). Alcohol and excessive caffeine consumption (over two cups of coffee before noon) was not allowed. Eight subjects were smokers, while the rest were non-smokers (self-reported). This dataset used for analysis was the same as in Ujma et al. (2014), except for the inclusion of one female subject who was excluded from the previous study due to her unavailable IQ score.

SLEEP RECORDINGS

All subjects spent two nights in a sleep laboratory and polysomnographic data from the 2nd night was used for analysis. Since the study was performed in cooperation between multiple sleep laboratories, recordings were performed in four slightly different designs.

For 31 subjects, recordings were performed with 18 EEG electrodes using a Flat Style SLEEP La Mont Headbox device with a HBX32-SLP preamplifier (La Mont Medical Inc. USA), with a sampling rate of 249 Hz, hardware prefiltering 0.5–70 Hz and a precision of 12 bit.

For 16 subjects signals were collected, prefiltered (0.33–1500 Hz, 40 dB/decade anti-aliasing hardware input filter), amplified and digitized with 4096 Hz/channel sampling rate (synchronous) and 12 bit resolution by using the 32 channel EEG/polysystem (Brain-Quick BQ 132S, Micromed, Italy). A further 40 dB/decade anti-aliasing digital filter was applied by digital signal processing which low-pass filtered the data at 450 Hz. Finally, the digitized and filtered EEG was undersampled at 1024 Hz.

For 114 subjects, recordings were performed with a Comlab 32 Digital Sleep Lab device (Schwarzer, Germany) with a sampling rate of 250 Hz, hardware prefiltering 0.53–70 Hz and a precision of 8 bit. In 94 of these subjects, 22 EEG electrode sites were used, while in the others 20 subjects 10 EEG electrodes were used. Common recording sites in all subjects which were used in the analysis were Fp1, Fp2, F3, F4, Fz, F7, F8, C3, C4, Cz, P3, P4, T3, T4, T5, T6, O1, and O2, all referred to the mathematically linked mastoids. For the 20 subjects with only 10 electrodes, data from Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1 and O2 was available and for the other electrodes these subjects were treated as missing data.

In order to correct for potentially different baseline amplitudes depending on the recording device (Vasko et al., 1997),

the analog-digital conversion and filtering characteristics of all recording devices were measured and sleep spindle amplitudes were corrected for the measured differences as follows (Ujma et al., 2014). We determined the amplitude reduction rate of each recording system by calculating the proportion between digital (measured) and analog (generated) amplitudes of sinusoid signals at typical sleep spindle frequencies (10, 11, 12, 13, 14, and 15 Hz) for both inducing (40 and 355 μ V amplitude) signals. Machine-specific amplitude reduction rates were given as the mean amplitude rate between digital and analog values at the two amplitudes and six measured frequencies. Sleep spindle amplitudes were corrected by dividing their calculated values by the amplitude reduction rate of the recording system. Given the individual- and derivation-specific adjustment inherent to both the Fixed frequency method (FixF) and the IAM, sleep spindle densities and durations are amplitude-insensitive measures. Thus, there is no need for the compensation of the different recording systems in these values.

Sleep recordings of the second nights were scored according to standard criteria (Iber et al., 2007) on a 20 s basis and artifacts were removed by visual inspection on a 4 s basis. Sleep spindle analysis was performed on artifact-free segments of NREM sleep.

ANALYSES

Fixed frequency method of sleep spindle analysis

For the FixF method we determined the 11–13 Hz range as a slow spindle frequency band and the 13–15 Hz window as a fast spindle frequency band. These frequencies were selected to ensure consistency with previous studies (Schabus et al., 2006, 2007, 2008; Chatburn et al., 2013), which used a similar approach for the separation of slow and fast spindles.

Sleep spindles were automatically detected within artifact-free NREM sleep periods on every EEG derivation. For slow spindle detection, data were bandpass-filtered between 11 Hz and 13 Hz. The root mean squares of the filtered signals were determined for 0.25 s length time windows. Next a threshold was calculated at the 95th percentile of the root mean square values for every EEG derivation. A spindle was identified when at least two consecutive root mean square time points exceeded the threshold, and the duration criterion (≥ 0.5 s) was met. Four spindle characteristics were calculated; these were density (number of spindles/min); amplitude (peak-to-peak difference in voltage, expressed in μ V); duration (s), and frequency (number of cycles/s, in Hz). The same procedure was followed for detecting fast spindles, using a band pass filter of 13–15 Hz (Schabus et al., 2007; Gruber et al., 2013).

Sleep spindle analysis according to the IAM

The second sleep spindle detection algorithm was the IAM (Bódizs et al., 2009). This sleep spindle detection method takes into account both inter-individual variations and intra-individual consistency in sleep spindle frequency (De Gennaro et al., 2005, 2008), analyzing sleep spindles at the individual peak frequency for all subjects.

The IAM procedure (Bódizs et al., 2009) consisted of several steps as described below (illustrated on **Figure 1**).

- i. Average amplitude spectra. Non-overlapping 4 s artifact-free NREM sleep EEG segments are Hanning-tapered (50%), then zero-padded to 16 s. Average amplitude spectra of all-night NREM sleep EEG derivations is computed between 9 Hz and 16 Hz by using an FFT routine (frequency resolution: 0.0625 Hz).
- ii. Individually adjusted frequency limits of slow and fast sleep spindles. Determination of the individual slow and fast sleep spindle frequencies is based on second order derivatives of the 9–16 Hz amplitude spectra. In order to avoid small fluctuations in convex and concave segments average amplitude spectra of 0.0625 Hz resolution (i) is downsampled (decimated) by a factor of 4 (resulting in a resolution of 0.25 Hz) before calculating the derivation-specific second-order derivatives in this frequency range. Derivation-specific second order derivatives of the amplitude spectra are then averaged over all EEG derivations resulting in a whole-scalp second order derivative for each subject. Individual-specific frequency limits of sleep spindles are defined as pairs of zero crossing points encompassing a negative peak in the whole-scalp second order derivatives. These zero-crossing points are rounded to the closest bins within the high-resolution (0.0625 Hz) amplitude spectra obtained in step i. Two pairs of individual-specific frequency limits and corresponding ranges are defined (one for slow and one for fast spindles). In cases of uncertainty (lack of zero crossing points indicating slow spindles or partial overlap between slow and fast sleep spindles in some cases), frequencies with predominance of power in averaged frontal (Fp1, Fp2, F3, F4, Fz, F7, F8) over averaged centro-parietal (C3, C4, Cz, P3, P4) amplitude spectra were considered as slow spindle frequencies ($N = 18$). There was no case of uncertainty related to the individual-specific frequency boundaries of fast sleep spindles.
- iii. Individual-specific spindle middle frequencies. Slow spindle middle frequency of a given subject was quantified as the arithmetic mean of the individual-specific lower and upper limits for slow spindling as obtained above (ii). In case of fast sleep spindling the arithmetic mean of the lower and the upper frequency limits of fast sleep spindles were considered.
- iv. Individual- and derivation-specific amplitude criteria for sleep spindles. Spindles are defined as those EEG segments contributing to the peak region of the average amplitude spectrum. Hence we intended to obtain an amplitude criterion corresponding to the line determined by the γ -values (μ V) pertaining to the individually adjusted pairs of frequency limits (ii) in the average amplitude spectra (i).
 - a. The number of high resolution (0.0625 Hz) frequency bins (i) falling in the individual-specific slow- and fast sleep spindle frequency ranges (ii) is determined.
 - b. The amplitude spectral values (i) at the individually adjusted frequency limits for slow and fast sleep spindles (ii) are determined. This is performed in a derivation-specific manner.
 - c. Number of bins for slow and fast sleep spindling (iv/a) are multiplied by the arithmetic mean of the pairs of derivation-specific amplitude spectral values for slow and

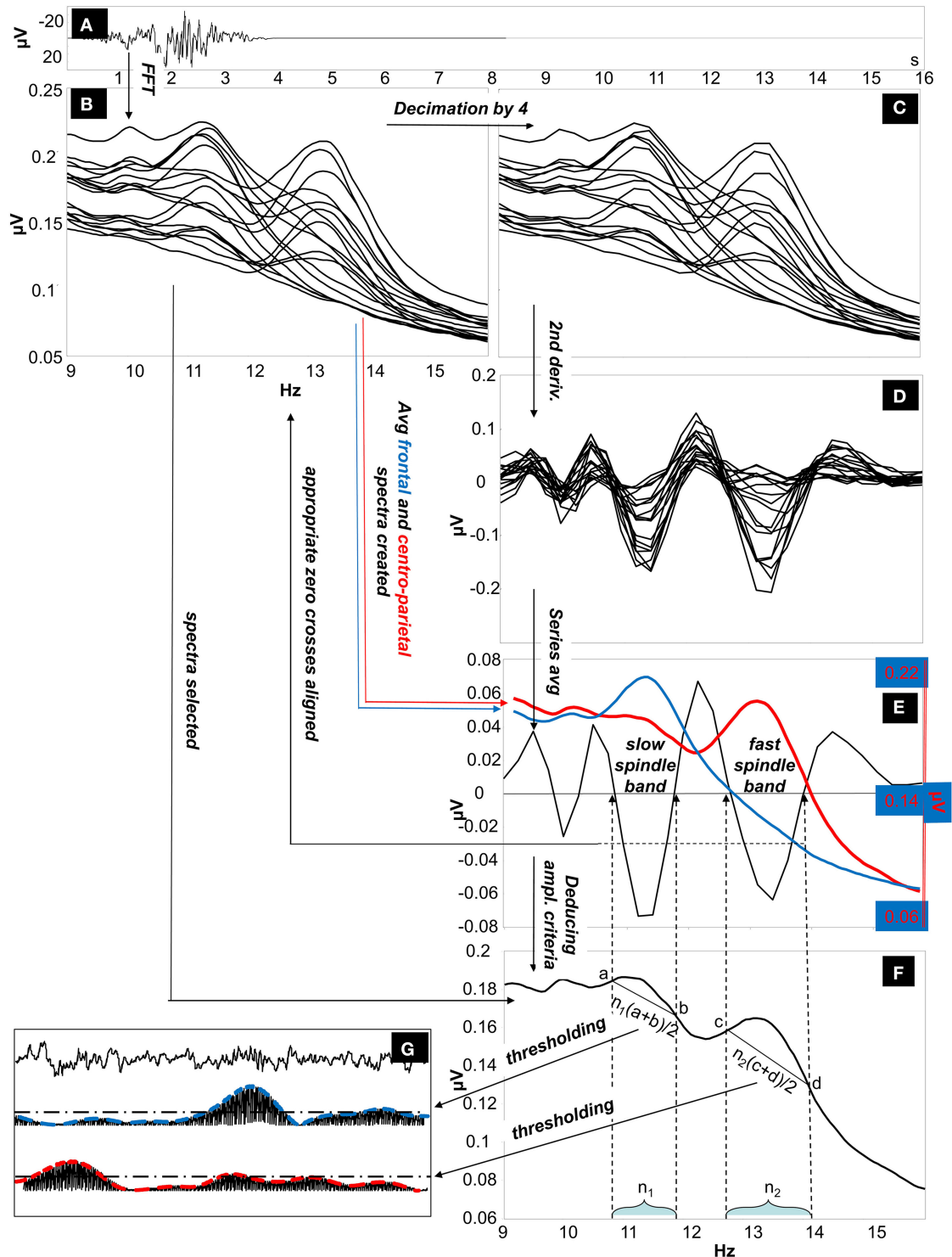


FIGURE 1 | The Individual Adjustment Method (IAM) of sleep spindle analysis. (A) Four-second EEG epoch Hanning-tapered and zero padded to 16 s. (B) Fast Fourier Transformation (FFT) is used to calculate 9–16 Hz average amplitude spectra of all night NREM sleep EEG from Hanning-tapered and zero-padded segments (derivations: Fp1, Fp2, F3, F4, Fz, F7, F8, T3, T4, T5, T6, C3, C4, Cz, P3, P4, O1, O2 referred to the mathematically-linked mastoids). (C) Amplitude spectra are decimated (down-sampled) by a factor of 4. (D) Second order derivatives of the decimated amplitude spectra. (E) Calculating the

whole-scalp second order derivatives by averaging all series. The resulting average series is overplotted with the averaged frontal (Fp1, Fp2, F3, F4, Fz, F7, F8) and centro-parietal (C3, C4, Cz, P3, P4) amplitude spectra (the left-side Y axis is for average second-order derivatives, while the second Y axis on the right is for average amplitude spectra). Appropriate zero-crossing points encompassing individual-specific slow and fast sleep spindle bands are selected on the 9–16 Hz frequency scale. (F) Derivation-specific amplitude criteria are calculated. (G) Thresholding of the envelopes of the slow and fast-spindle filtered signal.

fast sleep spindle frequency limits (iv/b), respectively. Outcomes are individual- and derivation specific amplitude criteria for slow and fast sleep spindle detections.

- v. Envelopes of sleep spindling. EEG data is band-pass filtered for the slow and fast spindle frequency ranges by using an FFT-based Gaussian filter with 16 s windows: $f(x) = e^{-(x-x_m)/(w/2)}$, where x varies between zero and the Nyquist frequency according to the spectral resolution, x_m is the middle frequency of the spindle range (iii), and w is the width of the spindle range (ii) (ii and iii). Filtered signal is rectified and smoothed by a moving average weighted with a Hanning window of 0.1 s length and multiplied with $\pi/2$ (the latter is the inverse of the mean of a rectified sine wave).
- vi. Detection and characterization of sleep spindles. If envelopes of this band-pass filtered and rectified data (v) exceed the individual and derivation-specific threshold as defined above (iv) for at least 0.5 s, a sleep spindle is detected. Sleep spindles detected this way are analyzed and average sleep spindle density (number of spindles per minute), sleep spindle duration (s), as well as median and maximum amplitude (expressed as all-night means of intra-spindle envelopes in μV at the middle of the detected spindles and at the maxima of the spindles, respectively) is calculated for the subject.

STATISTICS

FixF and IAM spindle parameters were compared using paired-sample t -tests ($\alpha = 0.05$). The Benjamini-Hochberg method of false detection rate correction was performed in order to correct for multiple comparisons.

We computed Pearson's point-moment correlation coefficients between comparable sleep spindle measures (that is, sleep spindle parameters computed from the same electrode) produced by IAM, and the FixF method.

RESULTS

IAM FREQUENCY BANDS

For the IAM method, individual slow spindle lower frequency limits ranged from 8.98 Hz to 12.95 Hz (mean: 10.96 Hz), while higher frequency limits ranged from 10.14 Hz to 13.7 Hz (mean: 11.9 Hz). Slow spindle middle frequencies ranged from 9.59 to 13.28 Hz (mean: 11.43 Hz). Fast spindle lower frequency limits ranged from 11.82 Hz to 14.77 Hz (mean: 13.06 Hz), while higher frequency limits ranged from 13.04 Hz to 16.03 Hz (mean: 14.36 Hz). Fast spindle middle frequencies ranged from 12.49 Hz to 15.38 Hz (mean: 13.71 Hz).

Individual slow spindle frequency bands were on average 0.94 Hz wide (range: 0.34–2.2 Hz). Individual fast spindle frequency bands were on average 1.3 Hz wide (range: 0.84–1.89 Hz).

Figure 2 shows the distribution of individual sleep spindle frequencies.

FixF vs. IAM SPINDLE PARAMETERS

IAM provides an approximately twice higher sleep spindle density than the FixF method in case of both slow and fast spindles as well as 1.5–2 times longer sleep spindle durations. Standard deviations of the individual averages of the FixF parameters are much smaller

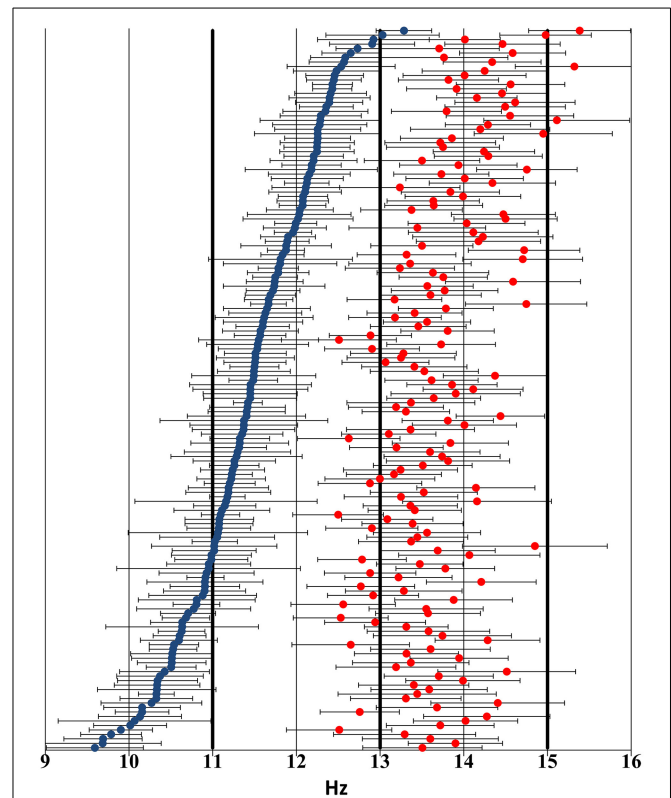


FIGURE 2 | Distribution of individual sleep spindle frequency bands across the 9–16 Hz frequency domain based on visual inspection of the sleep EEG spectrum and the zero crossings of the second-order derivatives thereof. Blue markers indicate slow spindle middle frequencies, while red markers show fast spindle middle frequencies. Negative and positive error bars illustrate lower and higher individual frequency limits, respectively. Thick lines highlight the 11 Hz, 13 Hz, and 15 Hz thresholds used in the FixF method. Subjects have been ordered by slow spindle middle frequency to ensure better visibility.

than in case of IAM parameters, even proportionally to the lower mean values.

Sleep spindle parameters are shown in **Table 1**. It must be noted that while FixF and IAM amplitude measures are displayed and compared, they are not expected to be on the same scale due to the narrower frequency band of IAM and the fact that in the FixF method amplitude was expressed as the mean maximum peak-to-peak voltage difference within a spindle, while in IAM amplitude was defined as the mean maximum of intra-spindle envelopes of the individually band-passed EEG.

The difference between comparable FixF and IAM spindle parameters is significant in all cases at $p < 0.0001$, and all comparisons remain significant after correction for multiple comparisons.

CORRELATIONS BETWEEN FixF AND IAM SPINDLE PARAMETERS

Despite the differences in the results, individual spindle parameters obtained with the FixF and IAM methods are strongly correlated in case of the amplitude and duration of fast spindles. These correlations are always over 0.5 for amplitude and over 0.4 for duration and they are highest (>0.8 for amplitude, >0.7 for

Table 1 | Sleep spindle parameters calculated by IAM and the fixed frequency method (FixF).

		Mean		StD		t-value			Mean		StD		t-value
		FixF	IAM	FixF	IAM				FixF	IAM	FixF	IAM	
C3							Fz						
Slow spindles	Density	3.469	6.830	0.163	1.428	−29.663	Slow spindles	Density	3.586	6.876	0.143	1.245	−31.273
	Duration	0.673	1.413	0.045	0.467	−19.995		Duration	0.668	1.435	0.032	0.462	−19.755
	Amplitude	42.053	3.548	12.045	1.848	40.093		Amplitude	57.322	4.902	15.897	2.507	38.814
Fast spindles	Density	3.657	7.176	0.202	0.921	−47.383	Fast spindles	Density	3.614	6.571	0.201	1.007	−34.319
	Duration	0.699	1.074	0.049	0.141	−31.956		Duration	0.693	1.435	0.047	0.462	−19.067
	Amplitude	43.171	5.471	13.632	1.533	34.871		Amplitude	46.909	5.588	16.803	1.732	29.150
C4							O1						
Slow spindles	Density	3.487	6.878	0.143	1.430	−29.947	Slow spindles	Density	3.215	6.737	0.126	1.947	−22.905
	Duration	0.666	1.411	0.042	0.462	−20.348		Duration	0.662	1.365	0.069	0.476	−18.540
	Amplitude	41.946	3.638	11.001	1.831	43.583		Amplitude	28.171	2.460	10.093	1.406	32.013
Fast spindles	Density	3.662	6.878	0.190	1.430	−28.299	Fast spindles	Density	3.554	7.062	0.217	1.104	−39.561
	Duration	0.697	1.411	0.045	0.462	−19.516		Duration	0.703	1.073	0.052	0.146	−30.259
	Amplitude	44.092	5.542	13.640	1.536	35.637		Amplitude	30.627	4.062	11.602	1.395	28.844
Cz							O2						
Slow spindles	Density	3.401	6.692	0.156	1.526	−25.571	Slow spindles	Density	3.219	6.728	0.114	1.944	−22.866
	Duration	0.674	1.381	0.053	0.465	−17.976		Duration	0.656	1.366	0.064	0.479	−18.640
	Amplitude	49.867	4.211	13.568	2.094	39.630		Amplitude	28.348	2.479	10.497	1.377	31.006
Fast spindles	Density	3.688	6.692	0.182	1.526	−23.302	Fast spindles	Density	3.544	7.051	0.218	1.109	−39.346
	Duration	0.700	1.381	0.045	0.465	−17.349		Duration	0.699	1.066	0.051	0.142	−30.824
	Amplitude	58.700	7.324	18.546	2.076	32.806		Amplitude	29.904	3.975	10.755	1.319	30.364
F3							P3						
Slow spindles	Density	3.653	6.920	0.158	1.193	−34.449	Slow spindles	Density	3.275	6.743	0.154	1.741	−25.187
	Duration	0.673	1.459	0.032	0.459	−21.674		Duration	0.694	1.376	0.069	0.471	−18.183
	Amplitude	52.046	4.518	14.978	2.341	39.780		Amplitude	37.668	3.050	10.762	1.687	40.321
Fast spindles	Density	3.588	6.323	0.205	0.982	−34.597	Fast spindles	Density	3.619	7.506	0.206	0.932	−51.670
	Duration	0.691	1.014	0.044	0.117	−32.780		Duration	0.718	1.110	0.051	0.149	−31.496
	Amplitude	39.676	4.846	13.952	1.525	31.488		Amplitude	44.219	5.773	14.429	1.670	33.585
F4							P4						
Slow spindles	Density	3.661	6.966	0.149	1.182	−35.205	Slow spindles	Density	3.286	6.761	0.147	1.754	−25.046
	Duration	0.669	1.456	0.029	0.456	−21.834		Duration	0.683	1.371	0.068	0.473	−18.297
	Amplitude	53.057	4.585	15.579	2.316	39.050		Amplitude	36.159	2.992	10.716	1.616	38.834
Fast spindles	Density	3.607	6.357	0.204	0.996	−34.328	Fast spindles	Density	3.613	7.468	0.195	0.961	−49.894
	Duration	0.687	1.456	0.043	0.456	−21.289		Duration	0.716	1.104	0.048	0.150	−31.314
	Amplitude	40.907	4.945	16.694	1.541	27.218		Amplitude	42.701	5.532	14.211	1.646	32.967
Slow spindles	Density	3.661	6.953	0.164	1.318	−29.526	Slow spindles	Density	3.529	6.927	0.144	1.521	−26.503
	Duration	0.669	1.420	0.031	0.455	−19.595		Duration	0.645	1.388	0.031	0.470	−18.779
	Amplitude	35.259	3.253	9.820	1.617	38.323		Amplitude	26.048	2.312	8.012	1.201	34.914
Fast spindles	Density	3.481	5.561	0.186	1.101	−22.198	Fast spindles	Density	3.529	6.210	0.156	1.142	−27.718
	Duration	0.660	0.964	0.035	0.102	−33.518		Duration	0.654	0.993	0.039	0.113	−33.920
	Amplitude	23.331	2.998	7.859	0.875	30.642		Amplitude	19.469	2.529	6.010	0.702	33.362
F8							T4						
Slow spindles	Density	3.670	6.979	0.160	1.334	−29.352	Slow spindles	Density	3.525	6.929	0.145	1.546	−26.119
	Duration	0.665	1.418	0.028	0.456	−19.639		Duration	0.639	1.379	0.028	0.466	−18.911
	Amplitude	35.654	3.303	9.349	1.626	40.625		Amplitude	25.325	2.348	6.821	1.198	39.535

(Continued)

Table 1 | Continued

		Mean		StD		t-value			Mean		StD		t-value
		FixF	IAM	FixF	IAM				FixF	IAM	FixF	IAM	
F7							T3						
Fast spindles	Density	3.479	5.534	0.207	1.099	−21.900	Fast spindles	Density	3.498	6.031	0.184	1.228	−24.304
	Duration	0.655	0.960	0.035	0.101	−34.075		Duration	0.647	0.985	0.037	0.117	−32.600
	Amplitude	23.663	3.039	7.774	0.885	31.412		Amplitude	19.517	2.601	6.174	0.785	32.388
Fp1							T5						
Slow spindles	Density	3.632	7.043	0.154	1.228	−34.960	Slow spindles	Density	3.333	6.753	0.139	1.808	−22.473
	Duration	0.674	1.448	0.029	0.455	−21.537		Duration	0.646	1.354	0.048	0.480	−17.495
	Amplitude	44.540	3.755	18.255	1.943	28.190		Amplitude	25.427	2.276	7.801	1.293	34.888
Fast spindles	Density	3.462	5.500	0.199	1.061	−23.961	Fast spindles	Density	3.544	6.849	0.188	1.058	−36.667
	Duration	0.669	0.969	0.036	0.099	−36.057		Duration	0.690	1.045	0.049	0.139	−28.762
	Amplitude	28.654	3.325	16.333	1.010	19.640		Amplitude	25.257	3.279	8.719	1.074	29.812
Fp2							T6						
Slow spindles	Density	3.644	7.064	0.152	1.238	−34.793	Slow spindles	Density	3.347	6.785	0.125	1.852	−22.075
	Duration	0.672	1.445	0.028	0.454	−21.579		Duration	0.638	1.348	0.042	0.475	−17.763
	Amplitude	44.180	3.783	17.303	1.927	29.443		Amplitude	24.813	2.241	8.184	1.213	32.509
Fast spindles	Density	3.477	5.569	0.204	1.070	−24.360	Fast spindles	Density	3.541	6.750	0.185	1.074	−35.098
	Duration	0.665	0.965	0.035	0.098	−36.612		Duration	0.680	1.033	0.048	0.132	−29.933
	Amplitude	28.325	3.345	13.607	1.031	23.228		Amplitude	23.782	3.108	7.491	0.876	32.665

Density, duration and amplitude means, standard deviations (StD) and comparison t-values are shown.

duration) in derivations where fast spindles are most prominent (central and parietal electrodes) as well as in occipital derivations. There is, surprisingly, a negative correlation between fast spindle density calculated by the IAM and the FixF method.

There is only a weak concordance between FixF and IAM slow spindle parameters. There is no significant FixF-IAM correlation in case of slow spindle density and duration, and only a modest correlation in case of slow spindle amplitude ($r < 0.5$ except for F3).

Table 2 presents the Pearson correlation coefficients depicting the linear relationship between corresponding IAM and FixF spindle parameters on all electrodes.

Given that 1) our sample consisted of several datasets recorded on various EEG devices and 2) the FixF ranges we analyzed—while based on previous literature—did not correspond well to the frequency ranges computed by IAM, we re-analyzed our sample divided in subsamples as well as with different FixF ranges set with slow spindles between 10 Hz and 12.5 Hz and fast spindles between 12.5 Hz and 15 Hz. In both re-analyses, we attempted to replicate our most prominent results, and investigated fast spindle parameters on P4 and slow spindle parameters on F3. F3 was selected over Fz because of the higher availability of this electrode in the sample.

Results are similar across subsamples: that is, fast spindle density is negatively correlated; slow spindle density and duration are not correlated, slow spindle amplitude is moderately and positively correlated while fast spindle duration

and amplitude are strongly and positively correlated. FixF-IAM correlations for slow spindles on F3 are as follows for density ($r_{\text{Budapest1}} = 0.427$, $p = 0.016$; $r_{\text{Budapest2}} = -0.032$, $p = 0.908$; $r_{\text{Munich}} = 0.129$, $p = 0.086$), duration ($r_{\text{Budapest1}} = 0.086$, $p = 0.647$; $r_{\text{Budapest2}} = -0.143$, $p = 0.597$; $r_{\text{Munich}} = -0.072$, $p = 0.448$) and amplitude ($r_{\text{Budapest1}} = 0.353$, $p = 0.052$; $r_{\text{Budapest2}} = 0.498$, $p = 0.049$; $r_{\text{Munich}} = 0.519$, $p < 0.001$). FixF-IAM correlations for fast spindles on P4 are as follows for density ($r_{\text{Budapest1}} = -0.28$, $p = 0.127$; $r_{\text{Budapest2}} = -0.282$, $p = 0.291$; $r_{\text{Munich}} = -0.359$, $p < 0.001$), duration ($r_{\text{Budapest1}} = 0.844$, $p < 0.001$; $r_{\text{Budapest2}} = 0.661$, $p = 0.005$; $r_{\text{Munich}} = 0.805$, $p < 0.001$) and amplitude ($r_{\text{Budapest1}} = 0.75$, $p < 0.001$; $r_{\text{Budapest2}} = 0.798$, $p < 0.001$; $r_{\text{Munich}} = 0.861$, $p < 0.001$).

Application of the new frequency bands also did not change the pattern of consistency of our methods significantly. With the 10–12.5 Hz FixF windows, FixF-IAM correlations for slow spindles on F3 are the following: $r_{\text{density}} = 0.083$, $p = 0.292$; $r_{\text{duration}} = -0.069$, $p = 0.39$; $r_{\text{amplitude}} = 0.419$, $p < 0.001$. With the 12.5–15 Hz FixF windows, FixF-IAM correlations for fast spindles on P4 are the following: $r_{\text{density}} = -0.149$, $p = 0.06$; $r_{\text{duration}} = 0.802$, $p < 0.001$; $r_{\text{amplitude}} = 0.66$, $p < 0.001$.

DISCUSSION

While previous studies compared sleep spindle detections between various manual and automatic methods (Huupponen et al., 2007; Warby et al., 2014), to our knowledge no previous study compared individual averages of sleep spindle parameters

Table 2 | Correlation coefficients and *p*-values between compatible sleep spindle parameters calculated by IAM and the fixed frequency method.

C3		Slow			Fast			Fz		Slow			Fast		
		Density	Duration	Amplitude	Density	Duration	Amplitude			Density	Duration	Amplitude	Density	Duration	Amplitude
r		−0.072	0.095	0.379*	−0.292*	0.802*	0.788*	r		0.248*	−0.087	0.402*	−0.341*	0.737*	0.782*
p		0.364	0.230	0.000	0.000	0.000	0.000	p		0.003	0.306	0.000	0.000	0.000	0.000
C4		Slow			Fast			O1		Slow			Fast		
		Density	Duration	Amplitude	Density	Duration	Amplitude			Density	Duration	Amplitude	Density	Duration	Amplitude
r		−0.090	0.049	0.297*	−0.341*	0.785*	0.814*	r		−0.155	0.052	0.415*	−0.257*	0.793*	0.827*
p		0.255	0.541	0.000	0.000	0.000	0.000	p		0.049	0.515	0.000	0.001	0.000	0.000
Cz		Slow			Fast			O2		Slow			Fast		
		Density	Duration	Amplitude	Density	Duration	Amplitude			Density	Duration	Amplitude	Density	Duration	Amplitude
r		−0.067	0.056	0.320*	−0.367*	0.792*	0.842*	r		−0.019	0.041	0.317*	−0.282*	0.775*	0.804*
p		0.430	0.511	0.000	0.000	0.000	0.000	p		0.816	0.610	0.000	0.000	0.000	0.000
F3		Slow			Fast			P3		Slow			Fast		
		Density	Duration	Amplitude	Density	Duration	Amplitude			Density	Duration	Amplitude	Density	Duration	Amplitude
r		0.159	−0.050	0.623*	−0.296*	0.665*	0.720*	r		−0.146	0.105	0.304*	−0.347*	0.811*	0.849*
p		0.044	0.529	0.000	0.000	0.000	0.000	p		0.064	0.183	0.000	0.000	0.000	0.000
F4		Slow			Fast			P4		Slow			Fast		
		Density	Duration	Amplitude	Density	Duration	Amplitude			Density	Duration	Amplitude	Density	Duration	Amplitude
r		0.162	−0.078	0.365*	−0.235*	0.654*	0.754*	r		−0.144	0.116	0.260*	−0.392*	0.826*	0.855*
p		0.041	0.328	0.000	0.003	0.000	0.000	p		0.069	0.143	0.001	0.000	0.000	0.000
F7		Slow			Fast			T3		Slow			Fast		
		Density	Duration	Amplitude	Density	Duration	Amplitude			Density	Duration	Amplitude	Density	Duration	Amplitude
r		0.095	−0.086	0.429*	−0.121	0.459*	0.637*	r		0.048	−0.078	0.321*	−0.235*	0.644*	0.724*
p		0.259	0.310	0.000	0.151	0.000	0.000	p		0.567	0.359	0.000	0.005	0.000	0.000
F8		Slow			Fast			T4		Slow			Fast		
		Density	Duration	Amplitude	Density	Duration	Amplitude			Density	Duration	Amplitude	Density	Duration	Amplitude
		0.071	−0.074	0.421*	−0.040	0.419*	0.670*	r		−0.051	−0.059	0.331*	−0.037	0.547*	0.565*
		0.404	0.383	0.000	0.637	0.000	0.000	p		0.548	0.485	0.000	0.658	0.000	0.000
Fp1		Slow			Fast			T5		Slow			Fast		
		Density	Duration	Amplitude	Density	Duration	Amplitude			Density	Duration	Amplitude	Density	Duration	Amplitude
r		0.010	0.104	0.277*	0.052	0.349*	0.502*	r		−0.114	0.082	0.440*	−0.352*	0.767*	0.798*
p		0.901	0.191	0.000	0.512	0.000	0.000	p		0.177	0.329	0.000	0.000	0.000	0.000
Fp2		Slow			Fast			T6		Slow			Fast		
		Density	Duration	Amplitude	Density	Duration	Amplitude			Density	Duration	Amplitude	Density	Duration	Amplitude
r		0.237*	−0.158*	0.297*	−0.025	0.399*	0.526*	r		0.097	0.052	0.238*	−0.351*	0.703*	0.695*
p		0.002	0.046	0.000	0.753	0.000	0.000	p		0.251	0.542	0.004	0.000	0.000	0.000

Significant correlations (after multiple comparison correction) are marked with an asterisk.

calculated by various methods. Moreover, comparisons of individual detections were usually performed on many spindles from a small number of subjects. We investigated the convergent validity of two well-known algorithms by correlating all-night averages of individual sleep spindle parameters in a large database of subjects. In this approach, the agreement

between individual detections is admittedly less important than agreement between individual averages. Overall, our results highlight both similarities and differences in the two sleep spindle detection methods we compared, and they do not provide overwhelming evidence for the convergence of the two methods.

IAM is tuned to individual spindle frequencies as well as individual and derivation-specific amplitude limits, making it inherently more sensitive as evidenced by higher spindle density and longer duration. FixF, on the other hands, focuses on the upper 5% of the amplitude distribution of filtered EEG signals. While FixF appears to detect “the tips of the icebergs” with this approach, the fast spindles detected by FixF are able to realistically approximate the same fast spindle durations and amplitudes as the IAM. Concordance is much weaker, however, in case of slow spindle amplitude, completely absent in case of slow spindle density and duration, while a very surprising negative correlation between fast spindle densities were found. To explain these findings, some empirical tendencies must be considered.

First, while the 13–15 Hz FixF window for fast spindles was similar to the empirically determined individual frequencies of IAM fast spindles, this was not the case for the 11–13 Hz slow spindle window. Fast spindle middle frequencies were below 13 Hz in only 11.24% of all cases and over 15 Hz in 1.87% of cases, while slow spindle middle frequencies were below 11 Hz in 27.5% of all cases and over 13 Hz in 1.25% of all cases. This poor demarcation of slow spindles in the FixF method might explain why FixF slow spindle parameters correlate more strongly with IAM fast spindle parameters than IAM slow spindle parameters (FixF slow vs. IAM fast correlations on Cz: $r_{\text{density}} = -0.092$ $p = 0.275$; $r_{\text{duration}} = 0.547$ $p < 0.001$; $r_{\text{amplitude}} = 0.603$ $p < 0.001$, with similar tendencies on all electrodes, see **Table 2** for correlations with IAM slow spindle parameters). This finding, together with poor agreement on density measures suggests that some FixF slow spindles may actually be classified as fast spindles by the IAM procedure and vice versa, explaining the confusion in both density measures and slow spindle parameters in general. This phenomenon is exemplified by some dissimilar findings in the field. That is, both slow and fast sleep spindle measures correlate with cognitive abilities in cases when the FixF method is used (Schabus et al., 2006, 2008), while in case of IAM fast spindles are much more stable correlates of cognitive performance (Bódizs et al., 2005, 2008; Ujma et al., 2014). It must be noted that sleep spindles are not stationary sinusoidal processes: they are known to shift frequencies (chirp). Negative spindle chirps (decreasing frequencies) have been reported in humans (Andrillon et al., 2011; Schonwald et al., 2011), while increasing spindle frequencies were reported in rats (Sitnikova et al., 2014). These frequency shifts are not large enough to eclipse the difference between slow and fast spindles (Andrillon et al., 2011) but spectral chirps arising in spindles close to the 13 Hz boundary might be large enough to make them “jump” it and be detected in the opposite category.

Second, the average width of the individual fast spindle frequency band was 1.3 Hz, while in case of slow spindles it was only 0.94 Hz. That is, our results show that individual fast spindle frequency bands rarely fell outside the 13–15 Hz range and they were generally closer to the 2 Hz window of the FixF method than slow spindle frequency bands. The fact the re-analysis with FixF bands resembling the empirically determined individual frequency bands of IAM (10–12.5/12.5–15 Hz, compare with IAM frequency bands on **Figure 2**) did not significantly improve concordance between the two methods suggests that the differences in individual spindle bandwidth may be even more important for

the lack of concordance between the two methods than the mere whereabouts of the frequency limits. This is in line with previous results from an adaptive, probabilistic model (Nonclercq et al., 2013) which reported a similar robustness to the input frequency range.

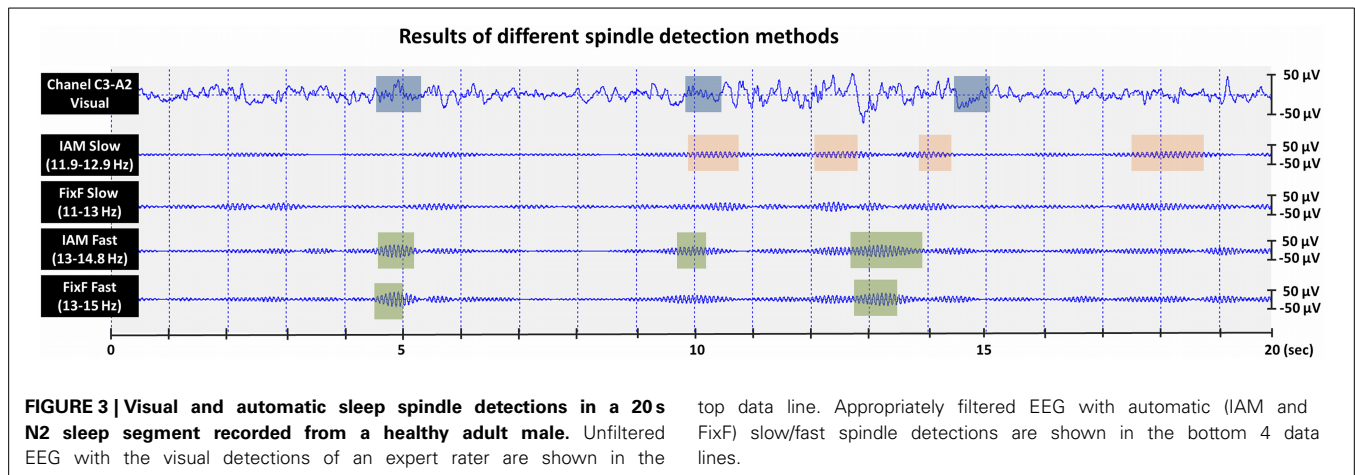
Based on the above findings we hypothesize that the lack of consistency between FixF and IAM slow spindle parameters is caused by the above factors: IAM slow spindles are determined at a lower and narrower frequency, with a larger distance from fast spindle frequencies in the same subject. The same phenomenon might be speculated to explain the negative correlation between IAM and FixF fast spindle density: in subjects with higher numbers of fast spindles (by IAM definitions) around the 13 Hz cutoff point cross-contamination with slow spindles may have been elevated in the FixF method.

There is little consistency in the sleep spindle detection methods used in previous research literature concerning the relationship between spindles and human cognition. Not all studies about the relationship between sleep spindle parameters and individual differences in psychometric variables separated slow and fast spindles: many analyzed sleep spindles in general or spectral power from a broader frequency band (Clemens et al., 2005; Fogel and Smith, 2006; Fogel et al., 2007; Tucker and Fishbein, 2009; Lustenberger et al., 2012; Gruber et al., 2013). Most studies which specifically analyzed slow and fast spindles and their correlation with psychometric variables used a *post-hoc* classification of spindles based on their central frequency, usually with 13 Hz as the split point (Schabus et al., 2006, 2008; Chatburn et al., 2013). Other studies used a slightly different *ad-hoc* division of sigma power into slow (11.5–12.5 Hz) and fast (13.5–14.5 Hz) sigma bands (Bang et al., 2014). Only a handful of studies relied on individually determined spindle frequencies, either by using the IAM method (Bódizs et al., 2005, 2008; Ujma et al., 2014) by computing individual relative sigma power defined as power ± 2 Hz around a single maximal spectral peak relative to the background EEG (Gottselig et al., 2002; Geiger et al., 2011) or by using an adaptive, probabilistic method (Nonclercq et al., 2013).

In sum, our results show that in case of fast spindles, duration and amplitude can be estimated reliably with both fixed and individual frequency methods. Much less consistency can be reached in case of slow spindles, and fixed cutoff frequencies may also lead to a poor separation of slow and fast spindles. Our results suggest that the cutoff frequencies and bandwidths for slow and fast spindles must be selected carefully and individually determined frequency bands should be considered.

It is notable that the concordance between the two methods is generally highest on typical spindle locations (frontal electrodes for slow spindles and centro-parietal electrodes for fast spindles). Concordance is usually lowest on temporal leads, but remains relatively high in occipital leads, in line with the relatively high spindle amplitude on these electrodes reported in the same dataset (Ujma et al., 2014). Lead-specific findings suggest that the lack of concordance between different spindle detection algorithms is especially problematic when non-prominent (e.g., temporal) leads are investigated.

There are limitations of our study that must be mentioned. First, the technical standards of the American Academy of Sleep



Medicine (2007) are not met in several subsamples of our study. That is, the analog-to-digital conversion rate is low (8 bits) in the largest subsample ($N = 114$), while the sampling rates are close to the minimally required values (249 and 250 Hz) in two subsamples. Second, while the study compared methodically well-established methods with previous practical applications in science, it must be acknowledged that IAM and the FixF algorithm operate with different philosophical underpinnings, they are designed to detect different features: the FixF method considers the background-relative amplitude of the filtered signal as the key feature of a spindle event, while IAM looks for an amplitude threshold based on the inflection points of the individual EEG spectrum (IAM and FixF detections, together with visual detections are illustrated on **Figure 3**). Therefore, a perfect agreement between their results cannot be expected, and in the absence of a “gold standard” the inherent superiority of any method cannot be ascertained.

REFERENCES

- Acir, N., and Güzelış, 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
- Amzica, F., and Steriade, M. (2000). Integration of low-frequency sleep oscillations in corticothalamic networks. *Acta Neurobiol. Exp. (Wars)* 60, 229–245.
- Andrillon, T., Nir, Y., Staba, R. J., Ferrarelli, F., Cirelli, C., Tononi, G., et al. (2011). Sleep spindles in humans: insights from intracranial EEG and unit recordings. *J. Neurosci.* 31, 17821–17834. doi: 10.1523/JNEUROSCI.2604-11.2011
- Bang, J. W., Khalilzadeh, O., Hämmäläinen, M., Watanabe, T., and Sasaki, Y. (2014). Location specific sleep spindle activity in the early visual areas and perceptual learning. *Vision Res.* 99, 162–171. doi: 10.1016/j.visres.2013.12.014
- Bódizs, R., Kis, T., Lazar, A. S., Havran, L., Rigo, P., Clemens, Z., et al. (2005). Prediction of general mental ability based on neural oscillation measures of sleep. *J. Sleep Res.* 14, 285–292. doi: 10.1111/j.1365-2869.2005.00472.x
- Bódizs, R., Körmendi, J., Rigó, P., and Lázár, A. S. (2009). The individual adjustment method of sleep spindle analysis: methodological improvements and roots in the fingerprint paradigm. *J. Neurosci. Methods* 178, 205–213. doi: 10.1016/j.jneumeth.2008.11.006
- Bódizs, R., Lázár, A., and Rigo, P. (2008). Correlation of visuospatial memory ability with right parietal EEG spindling during sleep. *Acta Physiol. Hung.* 95, 297–306. doi: 10.1556/APhysiol.95.2008.3.5
- Broughton, R., Healey, T., Maru, J., Green, D., and Pagurek, B. (1978). A phase locked loop device for automatic detection of sleep spindles and stage 2. *Electroencephalogr. Clin. Neurophysiol.* 44, 677–680. doi: 10.1016/0013-4694(78)90134-7
- 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
- Carrier, J., Land, S., Buysse, D. J., Kupfer, D. J., and Monk, T. H. (2001). The effects of age and gender on sleep EEG power spectral density in the middle years of life (ages 20–60 years old). *Psychophysiology* 38, 232–242. doi: 10.1111/1469-8986.3820232
- Chatburn, A., Coussens, S., Lushington, K., Kennedy, D., Baumert, M., and Kohler, M. (2013). Sleep spindle activity and cognitive performance in healthy children. *Sleep* 36, 237–243. doi: 10.5665/sleep.2380
- 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
- De Gennaro, L., Ferrara, M., Vecchio, F., Curcio, G., and Bertini, M. (2005). An electroencephalographic fingerprint of human sleep. *Neuroimage* 26, 114–122. doi: 10.1016/j.neuroimage.2005.01.020
- De Gennaro, L., Marzano, C., Fratello, F., Moroni, F., Pellicciari, M. C., Ferlazzo, F., et al. (2008). The electroencephalographic fingerprint of sleep is genetically determined: a twin study. *Ann. Neurol.* 64, 455–460. doi: 10.1002/ana.21434
- Devuyst, S., Dutoit, T., Didier, J. F., Meers, F., Stanus, E., Stenuit, P., et al. (2006). Automatic sleep spindle detection in patients with sleep disorders. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 1, 3883–3886. doi: 10.1109/IEMBS.2006.259298
- Driver, H. S., Dijk, D. J., Werth, E., Biedermann, K., and Borbely, A. A. (1996). Sleep and the sleep electroencephalogram across the menstrual cycle in young healthy women. *J. Clin. Endocrinol. Metab.* 81, 728–735.
- Duman, F., Erdamar, A., Eroglu, O., Telatar, Z., and Yetkin, S. (2009). Efficient sleep spindle detection algorithm with decision tree. *Expert Syst. Appl.* 36, 9980–9985. doi: 10.1016/j.eswa.2009.01.061
- Ferri, R., Dequae, P., Colognola, R. M., Musumeci, S. A., and Bergonzi, P. (1989). A simple electronic and computer system for automatic spindle detection. *Neurophysiol. Clin.* 19, 171–177. doi: 10.1016/S0987-7053(89)80057-7
- Fogel, S. M., Nader, R., Cote, K. A., and Smith, C. T. (2007). Sleep spindles and learning potential. *Behav. Neurosci.* 121, 1–10. doi: 10.1037/0735-7044.121.1.1
- Fogel, S. M., and Smith, C. T. (2006). Learning-dependent changes in sleep spindles and Stage 2 sleep. *J. Sleep Res.* 15, 250–255. doi: 10.1111/j.1365-2869.2006.00522.x
- Fogel, S. M., and Smith, C. T. (2011). The function of the sleep spindle: a physiological index of intelligence and a mechanism for sleep-dependent memory consolidation. *Neurosci. Biobehav. Rev.* 35, 1154–1165. doi: 10.1016/j.neubiorev.2010.12.003
- Geiger, A., Huber, R., Kurth, S., Ringli, M., Jenni, O. G., and Achermann, P. (2011). The sleep EEG as a marker of intellectual ability in school age children. *Sleep* 34, 181–189.
- Genzel, L., Kiefer, T., Renner, L., Wehrle, R., Kluge, M., Grozinger, M., et al. (2012). Sex and modulatory menstrual cycle effects on sleep related memory consolidation. *Psychoneuroendocrinology* 37, 987–998. doi: 10.1016/j.psyneuen.2011.11.006

- Genzel, L., Kroes, M. C., Dresler, M., and Battaglia, F. P. (2014). Light sleep versus slow wave sleep in memory consolidation: a question of global versus local processes? *Trends Neurosci.* 37, 10–19. doi: 10.1016/j.tins.2013.10.002
- Gottselig, J. M., Bassetti, C. L., and Achermann, P. (2002). Power and coherence of sleep spindle frequency activity following hemispheric stroke. *Brain* 125, 373–383. doi: 10.1093/brain/awf021
- Gruber, R., Wise, M. S., Frenette, S., Knäuper, B., Boom, A., Fontil, L., et al. (2013). The association between sleep spindles and IQ in healthy school-age children. *Int. J. Psychophysiol.* 89, 229–240. doi: 10.1016/j.ijpsycho.2013.03.018
- Huupponen, E., Gómez-Herrero, G., Saastamoinen, A., Väri, 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., Himanen, S. L., Varri, A., Hasan, J., Lehtokangas, M., and Saarinen, J. (2002). A study on gender and age differences in sleep spindles. *Neuropsychobiology* 45, 99–105. doi: 10.1159/000048684
- Huupponen, E., Varri, A., Himanen, S. L., Hasan, J., Lehtokangas, M., and Saarinen, J. (2000). Optimization of sigma amplitude threshold in sleep spindle detection. *J. Sleep Res.* 9, 327–334. doi: 10.1046/j.1365-2869.2000.00220.x
- Iber, C., Ancoli-Israel, S., Chesson, A., and Quan, S. (2007). *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specification*. Westchester, IL: American Academy of Sleep Medicine.
- Lustenberger, C., Maric, A., Durr, R., Achermann, P., and Huber, R. (2012). Triangular relationship between sleep spindle activity, general cognitive ability and the efficiency of declarative learning. *PLoS ONE* 7:e49561. doi: 10.1371/journal.pone.0049561
- Nonclercq, A., Urbain, C., Verheulpen, D., Decaestecker, C., Van Bogaert, P., and Peigneux, P. (2013). Sleep spindle detection through amplitude-frequency normal modelling. *J. Neurosci. Methods* 214, 192–203. doi: 10.1016/j.jneumeth.2013.01.015
- Piantoni, G., Poil, S. S., Linkenkaer-Hansen, K., Verweij, I. M., Ramautar, J. R., Van Someren, E. J., et al. (2013). Individual differences in white matter diffusion affect sleep oscillations. *J. Neurosci.* 33, 227–233. doi: 10.1523/JNEUROSCI.2030-12.2013
- Ray, L. B., Fogel, S. M., Smith, C. T., and Peters, K. R. (2010). Validating an automated sleep spindle detection algorithm using an individualized approach. *J. Sleep Res.* 19, 374–378. doi: 10.1111/j.1365-2869.2009.00802.x
- Saletin, J. M., Van Der Helm, E., and Walker, M. P. (2013). Structural brain correlates of human sleep oscillations. *Neuroimage* 83, 658–668. doi: 10.1016/j.neuroimage.2013.06.021
- Schabus, M., Dang-Vu, T. T., Albouy, G., Baletau, 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.* 104, 13164–13169. doi: 10.1073/pnas.0703084104
- Schabus, M., Hodlmoser, K., Gruber, G., Sauter, C., Anderer, P., Klossch, G., et al. (2006). Sleep spindle-related activity in the human EEG and its relation to general cognitive and learning abilities. *Eur. J. Neurosci.* 23, 1738–1746. doi: 10.1111/j.1460-9568.2006.04694.x
- Schabus, M., Hoedlmoser, K., Pecherstorfer, T., Anderer, P., Gruber, G., Parapatics, S., et al. (2008). Interindividual sleep spindle differences and their relation to learning-related enhancements. *Brain Res.* 1191, 127–135. doi: 10.1016/j.brainres.2007.10.106
- Schmicek, P., Zeitlhofer, J., Anderer, P., and Saletu, B. (1994). Automatic sleep-spindle detection procedure: aspects of reliability and validity. *Clin. Electroencephalogr.* 25, 26–29. doi: 10.1177/155005949402500108
- Schonwald, S. V., Carvalho, D. Z., Dellagustin, G., De Santa-Helena, E. L., and Gerhardt, G. J. (2011). Quantifying chirp in sleep spindles. *J. Neurosci. Methods* 197, 158–164. doi: 10.1016/j.jneumeth.2011.01.025
- Sitnikova, E., Hramov, A. E., Grubov, V., and Koronovsky, A. A. (2014). Time-frequency characteristics and dynamics of sleep spindles in WAG/Rij rats with absence epilepsy. *Brain Res.* 16, 290–299. doi: 10.1016/j.brainres.2013.11.001
- Steriade, M. (2000). Corticothalamic resonance, states of vigilance and mentation. *Neuroscience* 101, 243–276. doi: 10.1016/S0306-4522(00)00353-5
- Steriade, M., and Deschenes, M. (1984). The thalamus as a neuronal oscillator. *Brain Res.* 320, 1–63. doi: 10.1016/0165-0173(84)90017-1
- Tucker, M. A., and Fishbein, W. (2009). The impact of sleep duration and subject intelligence on declarative and motor memory performance: how much is enough? *J. Sleep Res.* 18, 304–312. doi: 10.1111/j.1365-2869.2009.00740.x
- Ujma, P. P., Konrad, B. N., Genzel, L., Bleifuss, A., Simor, P., Pótari, A., et al. (2014). Sleep spindles and intelligence: evidence for a sexual dimorphism. *J. Neurosci.* 34, 16358–16368. doi: 10.1523/JNEUROSCI.1857-14.2014
- Vasko, R. C., Brunner, D. P., Monahan, J. P., Doman, J., Boston, J. R., El-Jaroudi, A., et al. (1997). Power spectral analysis of EEG in a multiple-bedroom, multiple-polygraph sleep laboratory. *Int. J. Med. Inform.* 46, 175–184. doi: 10.1016/S1386-5056(97)00064-6
- 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
- Warby, S. C., Wendt, S. L., Welinder, P., Munk, E. G. S., Carrillo, O., Sorensen, H. B. D., 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

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: 02 November 2014; accepted: 19 January 2015; published online: 17 February 2015.

Citation: Ujma PP, Gombos F, Genzel L, Konrad BN, Simor P, Steiger A, Dresler M and Bódizs R (2015) A comparison of two sleep spindle detection methods based on all night averages: individually adjusted vs. fixed frequencies. *Front. Hum. Neurosci.* 9:52. doi: 10.3389/fnhum.2015.00052

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

Copyright © 2015 Ujma, Gombos, Genzel, Konrad, Simor, Steiger, Dresler and Bódizs. 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.

Using a quadratic parameter sinusoid model to characterize the structure of EEG sleep spindles

Abdul J. Palliyali, Mohammad N. Ahmed and Beena Ahmed*

Electrical and Computer Engineering Program, Texas A&M University at Qatar, Doha, Qatar

OPEN ACCESS

Edited by:

Christian O'Reilly,
McGill University - Montreal
Neurological Institute, Canada

Reviewed by:

Petros Xanthopoulos,
University of Central Florida, USA
Christian O'Reilly,
McGill University - Montreal
Neurological Institute, Canada
Errikos-Chaim Michael Ventouras,
Technological Educational Institution
of Athens, Greece

*Correspondence:

Beena Ahmed,
Electrical and Computer Engineering
Program, Texas A&M University at
Qatar, PO Box 23874, Doha, Qatar
beena.ahmed@qatar.tamu.edu

Received: 14 November 2014

Accepted: 28 March 2015

Published: 05 May 2015

Citation:

Palliyali AJ, Ahmed MN and Ahmed B
(2015) Using a quadratic parameter
sinusoid model to characterize the
structure of EEG sleep spindles.
Front. Hum. Neurosci. 9:206.
doi: 10.3389/fnhum.2015.00206

Sleep spindles are essentially non-stationary signals that display time and frequency-varying characteristics within their envelope, which makes it difficult to accurately identify its instantaneous frequency and amplitude. To allow a better parameterization of the structure of spindle, we propose modeling spindles using a Quadratic Parameter Sinusoid (QPS). The QPS is well suited to model spindle activity as it utilizes a quadratic representation to capture the inherent duration and frequency variations within spindles. The effectiveness of our proposed model and estimation technique was quantitatively evaluated in parameter determination experiments using simulated spindle-like signals and real spindles in the presence of background EEG. We used the QPS parameters to predict the energy and frequency of spindles with a mean accuracy of 92.34 and 97.73% respectively. We also show that the QPS parameters provide a quantification of the amplitude and frequency variations occurring within sleep spindles that can be observed visually and related to their characteristic “waxing and waning” shape. We analyze the variations in the parameters values to present how they can be used to understand the inter- and intra-participant variations in spindle structure. Finally, we present a comparison of the QPS parameters of spindles and non-spindles, which shows a substantial difference in parameter values between the two classes.

Keywords: sleep spindles, sleep spindles model, sleep spindle structure, sleep stages, sleep spindle morphology

Introduction

Spindles are rhythmic transients present in the electroencephalogram (EEG) characteristic of stage two sleep. Though varying definitions of spindles exist in literature, the American Academy of Sleep Medicine (AASM) has standardized them by describing spindles as “oscillatory bursts on EEG, of 11–16 Hz sinusoidal waves, with a duration of 0.5–2 s and waxing and waning envelope” (Rechtschaffen and Kales, 1968; Iber et al., 2007).

Sleep spindles are used to aid sleep staging (Rechtschaffen and Kales, 1968; Iber et al., 2007). Recent research has shown that they play a role in memory formation and sleep “stability” (Wei et al., 1999; Fogela and Smith, 2011). They have also been found to have an association with various pathological phenomenon such as depression, epilepsy, Parkinson, Alzheimer, and schizophrenia, further raising their significance (Bódizs et al., 2009; Wamsley et al., 2012; Tezer et al., 2014). For example in Fogela and Smith (2011) the authors propose that spindles can be used as possible physiological markers of intellectual ability; spindle properties were found to be highly correlated with tests of intelligence such as IQ tests. The authors also discuss the role of spindles in the consolidation of declarative memory by aiding the interaction between the hippocampus and the thalamus.

Similarly, in Bódizs et al. (2005), the authors showed that the grouping and density of fast spindles correlated positively with mental ability measured from standard Raven Progressive Matrices test. Authors in Tezer et al. (2014) reported a significant decrease in the power and density of spindles before epileptic seizures especially in extra temporal lobe epilepsies. Participants with schizophrenia were also found to have drastically reduced density, number and coherence of sleep spindles (Wamsley et al., 2012).

These analyses require accurate labeling of sleep spindles in EEG recordings, which is time-consuming and error-prone when done manually. Automated spindle detection is thus gathering increasing attention from the research community. As spindles are of sinusoidal nature, characterized by progressively increasing, then gradually decreasing amplitude, most spindle detectors utilize features best suited for sinusoidal functions such as Filter banks, Fast Fourier Transforms, Wavelets, and Matching pursuit (Schönwald et al., 2006; Huupponen et al., 2007; Bódizs et al., 2009). The accuracy of these features however decreases when the frequency content of the background EEG overlaps the spindle range causing an increase in the number of false positives. Automatic sleep spindle detection is also hindered due to fluctuations in the frequency patterns and large inter-individual variability (Campbell et al., 1980; Kunz et al., 2000). However, a more significant issue in the development of accurate sleep spindle detectors is the proper training or tuning of these detectors. The broad AASM definition for sleep spindles leaves the manual marking of spindles in EEG data open to some interpretation, leading to low inter-expert agreement for spindle scoring (Kunz et al., 2000). A study by Wendt et al. found an average intra-expert agreement of $72 \pm 7\%$ ($\kappa: 0.66 \pm 0.07$) and an average inter-expert agreement of $61 \pm 6\%$ ($\kappa: 0.52 \pm 0.07$) (Wendt et al., 2015). Thus, the accuracy of sleep spindle detectors when trained and tested using data scored from a single scorer can fall significantly when tested against data scored by other experts. This also makes it difficult to develop validated assessment criteria for automatic sleep spindle detectors to compare the performance of proposed detectors.

A number of mathematical models have been proposed to better characterize the structure of sleep spindles, thus enabling a better understanding of their structure and facilitating further analysis (Olbrich and Achermann, 2005, 2008; Xanthopoulos et al., 2006; Ktonas et al., 2007; Perumalsamy et al., 2009; Nonclercq et al., 2013). In Olbrich and Achermann (2005), the authors fitted autoregressive (AR) models onto EEG data and used it to analyze oscillatory patterns including spindles. The authors further expanded their work in Olbrich and Achermann (2008) to study the temporal organization of spindles. Though, spindles were detected by studying damping constants of the AR model, no physical characteristics of the spindle were modeled. A similar approach was later proposed in Perumalsamy et al. (2009) where oscillations in EEG including spindles were detected using AR models through surrogate data testing. In Nonclercq et al. (2013), the authors modeled the amplitude and frequency of spindles using bivariate normal distributions. The work, motivated by the widely varying values of spindle properties, used tolerance intervals of normal models to detect spindles. However, it was limited

to the detection of spindles and did not model intra-spindle variations of these properties.

Spindle models as above have been adequate for applications such as the detection of spindles. However, they fail to incorporate details such as the intra-spindle variation of frequencies or “skewness” of the envelope. These details more than often vary with abnormalities or other factors, requiring a model that parameterizes these variations. As spindles have strong amplitude and frequency modulations, non-stationary sinusoidal analysis where the amplitude and frequency are allowed to evolve within the analysis frame are required. In this context, Ktonas et al. (2007) modeled spindles as amplitude and frequency modulated sinusoids. The model consisted of six parameters that captured the time varying microstructure of spindles. The authors also compared various time-frequency analysis methods for parameter estimation in Xanthopoulos et al. (2006) and concluded that complex demodulation provided the best results. They report promising preliminary results with simulated spindles and some selected spindles from three healthy controls and three dementia participants (Ktonas et al., 2009), but do not present detailed validation studies with the model parameters. Furthermore, the sinusoidal form approximation imposed by the model means non-sinusoidal variations in the spindle envelope and instantaneous frequency, as shown in **Figure 1**, cannot be tracked completely as also discussed by the authors in Ktonas et al. (2009).

In this paper, we extend the work done on spindle modeling using amplitude and frequency modulation with a new Quadratic Parameter Sinusoid (QPS) model to improve the representation of the intra-spindle amplitude and frequency variations without increasing complexity. The model utilizes a quadratic representation to modulate the specific amplitude and frequency variations within spindles. The QPS model was originally used to model non-stationary speech and music (Marques and Almeida, 1989). Non-stationary speech frames were approximated as a sum of time varying frequency and amplitude sinusoids and spectrally analyzed using Short Time Fourier Transforms. The QPS model is well suited to model spindle activity due to its ability to accurately model instantaneous frequency, phase and amplitude in non-stationary signals without the need to assume local stationarity.

The rest of the paper is structured as follows. In the Materials and Methods section we define the QPS model, explain the methodology utilized to estimate the model parameters and experiments conducted to validate the QPS model. We then summarize the results obtained from parameter estimation on simulated spindles with additive white noise and delta EEG as well as real spindles, followed by a discussion of the results and conclusions.

Materials and Methods

Quadratic Parameter Sinusoid

Sleep spindles have a waxing and waning sinusoidal form which enables them to be represented as a modulated sinusoidal whose instantaneous frequency and amplitude continuously varies with time. A sleep spindle $s(t)$ can thus be represented as

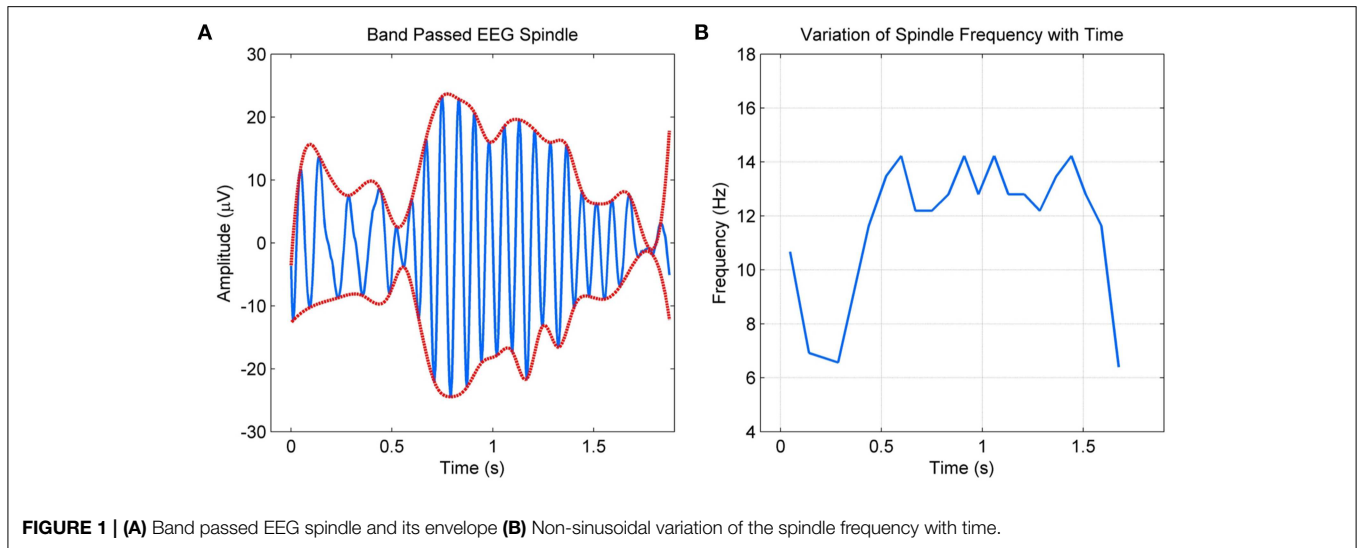


FIGURE 1 | (A) Band passed EEG spindle and its envelope **(B)** Non-sinusoidal variation of the spindle frequency with time.

$$s(t) = e^{A(t)} \cos P(t) \quad (1)$$

where $A(t)$ represents the instantaneous logarithmic amplitude and $P(t)$ the instantaneous phase. The instantaneous frequency $F(t)$ can be obtained from the time derivative of $P(t)/2\pi$. Due to the non-stationary nature of EEG, both $A(t)$ and $F(t)$ will be time-varying, making their determination non-trivial. For each spindle, as shown in Ito and Yano (1989), both $A(t)$ and $P(t)$ can be approximated using Taylor's polynomials around a center time t_c . $P(t)$ is given by

$$P(t) = \sum_{n=0}^{\infty} p_n(t - t_c)^n / n! \quad (2)$$

where,

$$p_n = \frac{d^{(n)}P(t_c)}{dt^{(n)}} \quad (3)$$

For frequency to be time-varying, there must be at least one non-zero p_n for $n \geq 2$ in $P(t)$. Hence, the minimum possible approximation of $P(t)$ would be as a quadratic function if the higher order terms are assumed to be negligible. $A(t)$ can similarly be represented as a quadratic function. This allows the sleep spindle to be defined as a Quadratic-Parameter Sinusoid (QPS) that is given by

$$s(t) = e^{(a+bt+ct^2)} \cos(d+et+ft^2) \quad (4)$$

where a, b, c, d, e and f are the parameters of the quadratic functions $A(t)$ and $P(t)$ from (1) respectively. As (4) gives only the real part of the QPS, the general form of $s(t)$ is given by

$$s(t) = e^{(a+bt+ct^2)} e^{i(d+et+ft^2)} \quad (5)$$

Figure 2A compares a spindle obtained from an EEG recording to a QPS model generated spindle in **Figure 2C**. The model

was applied to the band passed version of the spindle as shown in **Figure 2B**. The figure shows considerable similarities between the waxing and waning envelope of the spindle and model.

The 6 parameters (a–f) of the QPS function determine characteristics such as frequency, change in frequency, amplitude, variation in amplitude and the envelope shape of the signal $s(t)$. The parameters a, b , and c largely determine the amplitude and the shape of the envelope of the QPS and hence, that of the spindle. a is the approximate instantaneous log-amplitude at time, $t = 0$, at which the QPS is centered; b the rate of change of amplitude; c the Gaussian parameter which determines the shape and duration of the curve (Abe and Smith, 2005). In symmetrical spindles, b would be zero, with increasing/decreasing values shifting the time at which the spindle reaches maximum amplitude. Negative values of c cause the signal to decay, giving the spindle its rising and waning shape. **Figures 3A–C** illustrates the variations in amplitude of $s(t)$ caused by increasing values of b and decreasing values of c .

The remaining three parameters d, e , and f influence the frequency characteristics and phase of the signal. d represents the initial phase at $t = 0$. The initial frequency of the signal is given by e , whereas f represents the frequency rate change (Ito and Yano, 2007). In the absence of drastic variations, parameter e determines the dominant spindle frequency and f causes a linear variation in this frequency within the spindle duration. **Figures 3D–F** show the variation that occurs in spindle frequency with increasing e and f .

The highly nonlinear structure of the QPS signal makes parameter estimation of the QPS for a real spindle non-trivial. The problem is further compounded due to the presence of background noise in the EEG. We used non-linear least square (NLLS) estimation using the “Levenberg-Marquardt” technique to obtain the parameters for the QPS model due to its relative simplicity and dependability.

NLLS estimation algorithms are iterative numerical methods that attempt to converge toward optimal parameter values by successively minimizing a sum of squares cost function.

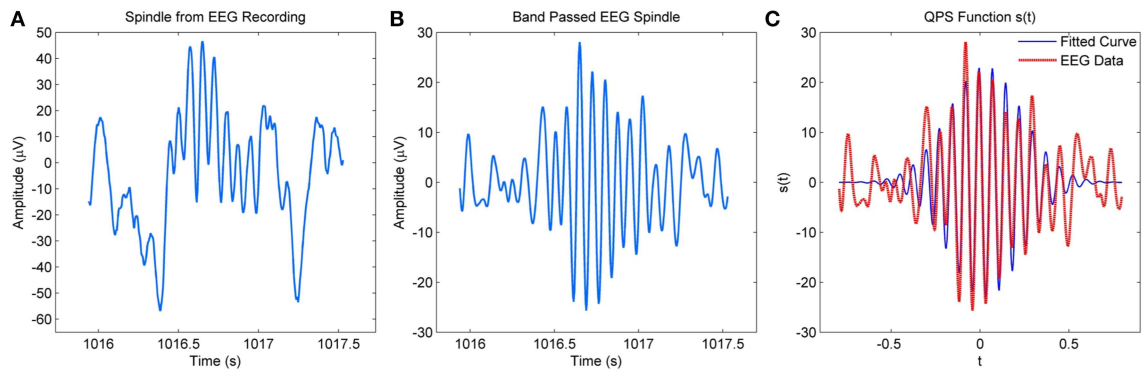


FIGURE 2 | (A) Raw spindle from MASS-C1/SS2 EEG recording **(B)** Band-passed version of the EEG spindle **(C)** QPS spindle generated using the parameters of the band-passed version of the spindle.

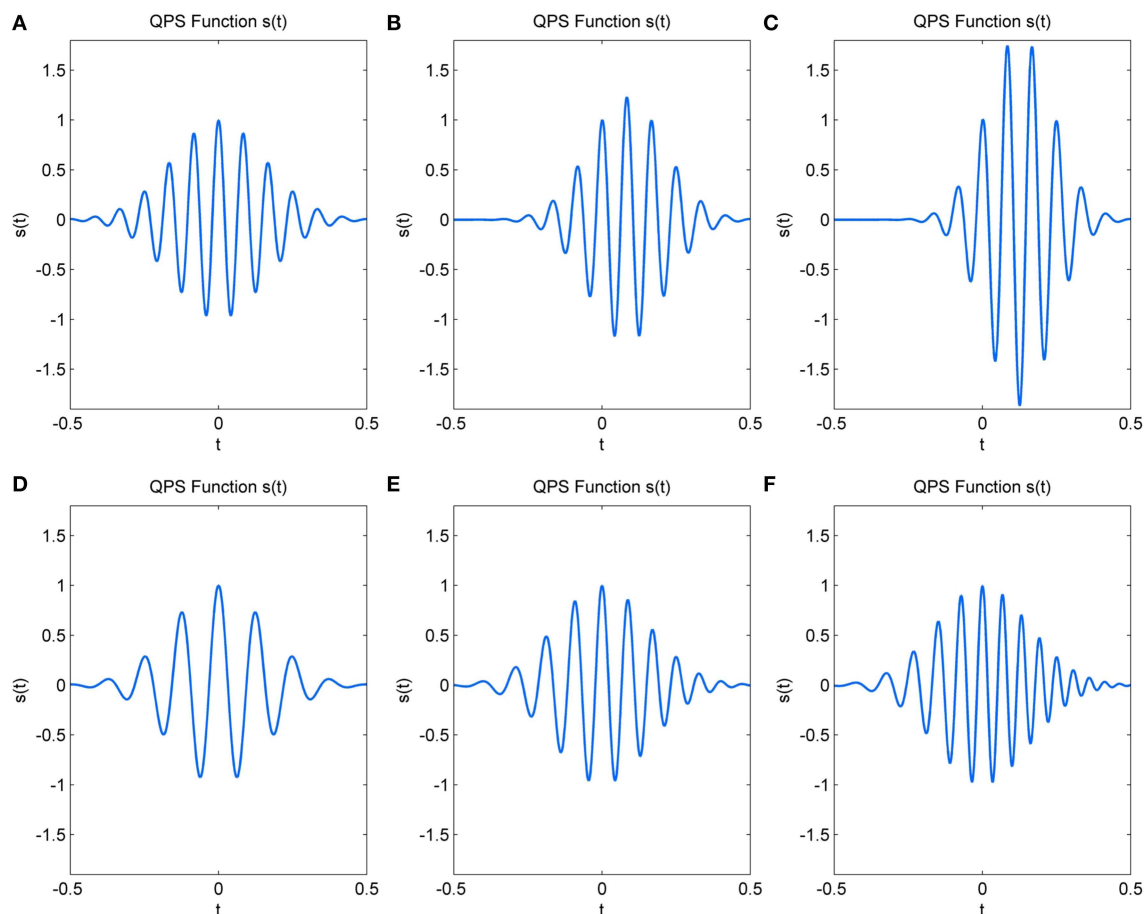


FIGURE 3 | Change in simulated QPS spindle with (A) $b = 0, c = -20$ (B) $b = 5, c = -30$ (C) $b = 10, c = -40$ (D) $e = 50, f = 0$ (E) $e = 70, f = 20$ (F) $e = 90, f = 40$.

The “Levenberg-Marquardt” technique utilized in this work is a standard NLLS implementation that adaptively varies the parameter update between Gradient descent and Gauss-Newton methods using a damping factor. If an iteration results in a large reduction of the cost, the damping factor is decreased bringing

the algorithm closer to the Gauss-Newton approach. On the other hand, if an iteration produces negligible cost reduction, the damping factor is increased to mimic a more Gradient-descent strategy. Like all NLLS algorithms, the algorithm can converge to local minima and is heavily dependent on the initial conditions.

In our work, convergence was ensured by initializing the parameters to spindle-like values and applying constraints consistent with the AASM spindle definition.

Experimental Validation Methodology

Our proposed QPS spindle model was validated using two datasets. The first dataset consisted of a group of simulated spindles with known parameter values. The second dataset consisted of real spindles from the MASS (Montreal Archive of Sleep Studies) database (O'Reilly et al., 2014). This database includes about 1700 h of PSG recording sampled at 256 Hz (O'Reilly et al., 2014). EEG recordings, annotated by two expert scorers, V4 and V5, were retrieved from the 19 participants of MASS-C1/SS2 database. The participants in this subset comprised of 11 women and 8 men with a mean age of 24.3 and 23.2 years respectively and an age range of 18–33 years (O'Reilly et al., 2014). The two expert scorers, V4 and V5 had an average Cohen's Kappa of 0.389 across all participants (O'Reilly and Nielsen, in revision). "It should be noted that relatively low inter-rater agreement is expected between these two scorers since V4 used traditional AASM scoring rules whereas V5 used an approach similar to Ray et al. (2010), O'Reilly and Nielsen (in revision)." Recordings from 4 participants (01-02-0004, 01-02-0008, 01-02-0015, and 01-02-0016) were not scored by V5 as they "were judged reflecting poor quality sleep (e.g., alpha intrusion during N2) or intermittent signal quality/artifact" (O'Reilly and Nielsen, in revision). Hence these recordings were discarded; spindles in the second dataset were thus isolated from the EEG recordings of 15 participants using the annotations of two expert scorers, V4 and V5, with 500 to 1000 spindles per participant. Prior approval for the study was obtained from the TAMU Institutional Review Board.

The accuracy of parameter estimation by the NLLS estimation algorithm was first validated on a simulated spindle dataset, as artificial spindles provided known reference values allowing errors to be quantified. Next, the robustness of the NLLS to varying levels of additive noise was quantified using "Goodness of Fit" measures. White Gaussian noise with wide-ranging SNR values and EEG segments consisting of strong delta components (representative of background EEG) were added to a number of simulated spindles. The QPS parameters of the resultant noisy signal were obtained using NLLS and compared with the parameter values of the original QPS prior to addition of noise.

The performance of the model with real EEG data in MASS-C1/SS2 was then evaluated by determining the error in estimated spindle frequency and energy for spindles marked by both the expert scorers individually and the common spindles marked by both scorers. Trends in the distribution of parameter values across all participants were also analyzed to obtain a better understanding of how spindle structure varied across the participants and how spindles marked by two scorers affect the distribution of these parameter values. The impact of each QPS parameter on the overall spindle shape was also studied by tracking variations in parameter values over the spindle value range. Finally, the ability of QPS parameter values to differentiate between spindles and non-spindle EEG activity was analyzed by comparing parameter values for sample non-spindle and spindle EEG regions in MASS-C1/SS2 database. The results from

each of the above validation experiments are detailed in the next section.

Results

Validation of QPS Model on Simulated Spindles Accuracy of Parameter Estimation

The parameters of a simulated spindle with added white Gaussian noise at an SNR of 10 dB were estimated using the NLLS algorithm. Both the true and estimated parameter values are given in **Table 1**, with the estimated parameters matching the true values within a narrow confidence interval. **Figure 4** illustrates the estimated signal (shown in red) superimposed on the noisy signal (shown in blue).

NLLS Performance in the Presence of White Gaussian Noise

We computed the following goodness of fit (GOF) measures on five simulated spindles with spindle like parameter values and varying SNR values:

1. Sum of Squared Errors (SSE)

$$SSE = \sum_{i=1}^n (s_i - \hat{s}_i)^2 \quad (6)$$

TABLE 1 | True and estimated parameters for a simulated spindle.

Parameter	True value	Estimated value	Confidence bounds
<i>a</i>	0	0.003995	(−0.01843, 0.02642)
<i>b</i>	0	0.01055	(−0.1505, 0.1716)
<i>c</i>	−20	−19.35	(−20.35, −18.35)
<i>d</i>	0	−0.01817	(−0.0406, 0.004255)
<i>e</i>	75	75.03	(74.87, 75.19)
<i>f</i>	0	0.6464	(−0.3566, 1.649)

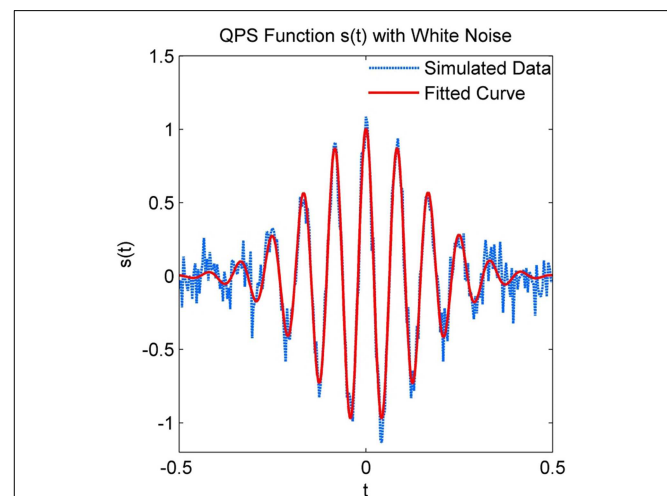


FIGURE 4 | Simulated QPS spindle with white Gaussian noise and the predicted QPS spindle using estimated parameters.

where, s_i is the i^{th} sample of the original signal, \hat{s}_i the i^{th} sample of the estimated signal and n is the number of samples, in our case n is 256. An ideal fit will result in an $SSE = 0$.

2. Rsquare

$$Rsquare = 1 - \frac{SSE}{SST} \quad (7)$$

where, $SST = \sum_{i=1}^n (s_i - \bar{s})^2$ is the total sum of squares about the mean \bar{s} . Rsquare measures the proportion of variance accounted for by the model and should ideally be 1.

3. Degree of Freedom adjusted Rsquare (Adjusted Rsquare)

$$AdjRsquare = 1 - \frac{SSE(n-1)}{SST(v)} \quad (8)$$

where, $v = n - m$; v is the residual degree of freedom and m the number of coefficients, In our case, m is 6 and v equals 250.

4. Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{MSE} = \sqrt{SSE/v} \quad (9)$$

Figures 5A–D plot the four GOF measures for a range of SNRs in the five simulated spindles. As seen, all four GOF measures approach their ideal values with increasing SNR. To determine the impact of the initial parameter values used in the NLLS algorithm on the final converged values, we also executed the NLLS algorithm using a range of different initial conditions for the

same spindle. **Figures 6A–D** show that the parameter estimates still converge at all SNRs despite variation in initial conditions indicating the robustness of NLLS algorithm. As expected, both **Figures 5, 6** show that parameters estimated with the NLLS converged to their true values with higher SNR.

NLLS Performance in the Presence of Delta Noise

We also evaluated the performance of NLLS in estimating QPS model parameters in the presence of strong delta components, since real EEG spindles have these components. The QPS model shown in blue in **Figure 7A** was simulated using the true parameter values from **Table 2**. A random EEG segment with delta components was then retrieved from the raw EEG recording of MASS-C1/SS2 participant 1, amplified by a factor of 2 and then added to the simulated QPS model from **Figure 7A**. The resulting signal is shown in red in **Figure 7A**. The NLLS algorithm was then used to estimate the parameters of the resulting signal with strong delta components. The QPS model generated using these estimated parameters is shown in red in **Figure 7B** superimposed on the original simulated noise-free spindle in blue. As seen from **Figure 7B**, there is no marked difference between the simulated QPS model and predicted QPS model in the presence of delta noise.

The accuracy of the QPS model parameters in the presence of delta noise was evaluated by adding 190 raw EEG segments

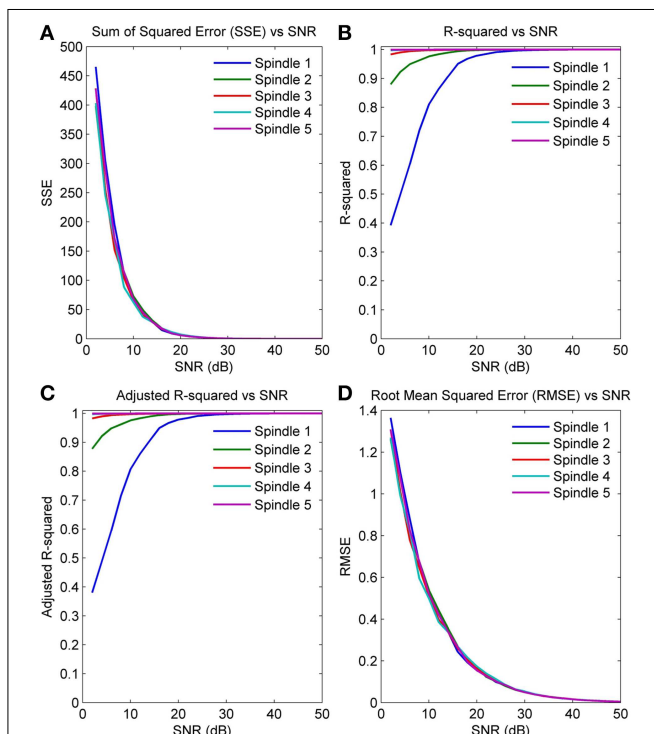


FIGURE 5 | GOF measures calculated over a range of SNR values for five simulated spindles (A) Sum of Squared Error (B) R-Squared Error (C) Adjusted R-Squared Error (D) Root Mean Squared Error.

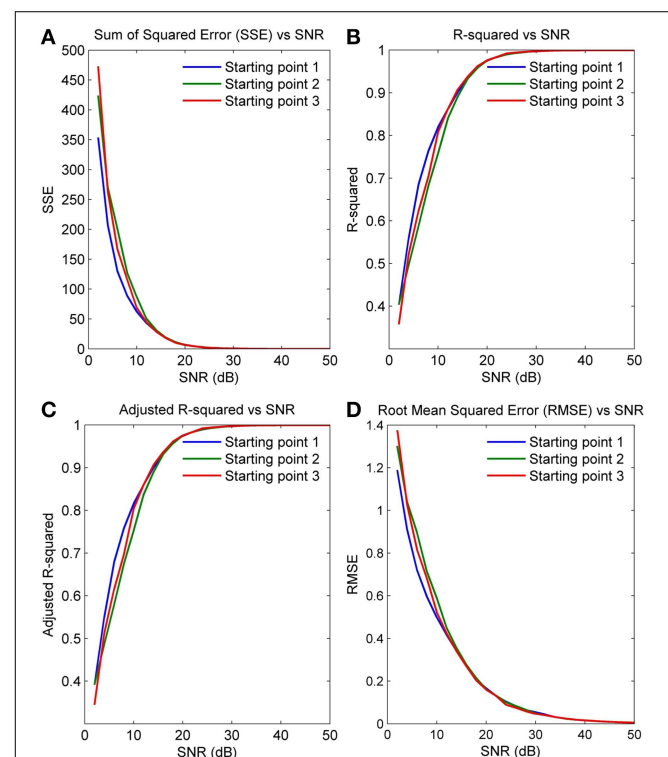


FIGURE 6 | GOF measures calculated over a range of SNR values for a single spindle but with different initial values for NLLS (A) Sum of Squared Error (B) R-Squared Error (C) Adjusted R-Squared Error (D) Root Mean Squared Error.

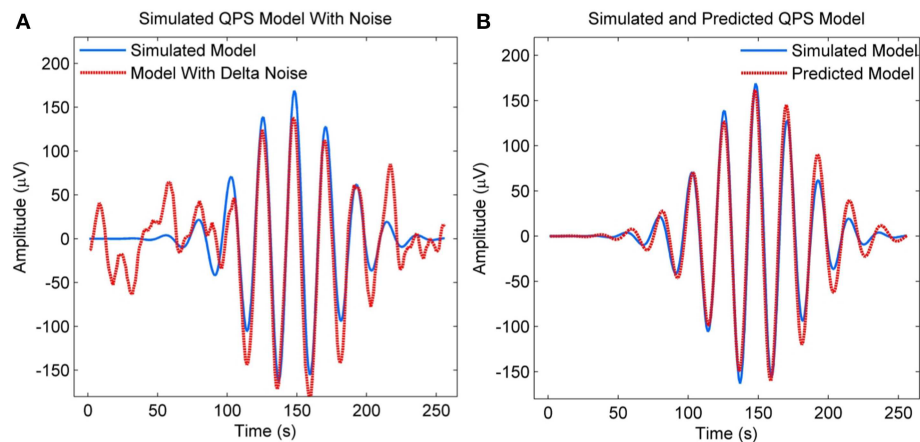


FIGURE 7 | (A) Simulated QPS spindle along with the spindle with added delta components **(B)** Simulated QPS spindle along with the predicted QPS spindle from the noisy spindle with added delta components.

TABLE 2 | True and estimated parameter values in the presence of delta noise.

Parameter	True value	Mean estimated value	Minimum value	Maximum value
<i>a</i>	5	4.990	4.722	5.290
<i>b</i>	4	4.122	1.369	8.025
<i>c</i>	-30	-29.998	-47.685	-14.420
<i>d</i>	1	0.903	-5.363	7.342
<i>e</i>	70	70.098	67.138	77.264
<i>f</i>	5	4.220	-28.004	20.876

with delta components to the simulated model from **Figure 7A**. The parameter values of the QPS model for these noisy signals were then estimated using NLLS and compared to its actual value. **Table 2** shows the mean, minimum and maximum estimated parameter values and the range of estimated values as computed by NLLS in the presence of delta noise. The percentage difference between the true parameter value and the mean estimated parameter value is less than 3% for all parameters except for parameters *d* and *f*, where the percentage difference is 9.7 and 15.6% respectively.

The boxplot in **Figure 8** shows the distribution of estimated parameter values in the presence of delta noise. The true parameter value is indicated using a blue square. The distribution of parameter values is shown using a red box, with the whiskers encompassing $\pm 2.7\sigma$ of the data set. As seen from **Figure 8** and **Table 2**, the greatest variation is seen in the values of parameters *c* and *f*, indicating a lower accuracy in estimating these models parameters in the presence of delta noise.

Validation of QPS Model on Real Spindles Accuracy of Energy and Frequency Estimation

The QPS model was tested on real spindles by estimating model parameters for spindles in the EEG data of 15 participants obtained from the MASS-C1/SS2 database. Since actual

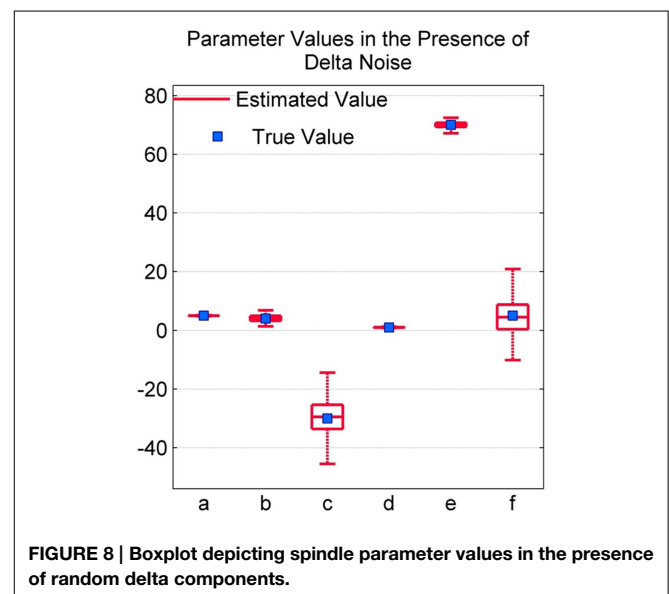
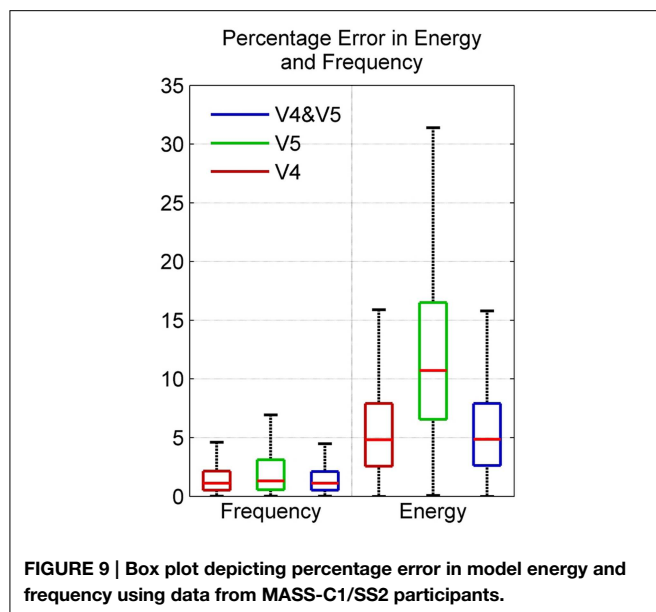


FIGURE 8 | Boxplot depicting spindle parameter values in the presence of random delta components.

spindle parameter values were not known, the QPS model was validated by computing the energy and frequency of the QPS generated spindle and comparing it to the spindle energy and frequency. Energy was calculated by computing the area within the envelope. As the envelope of the QPS generated spindle is given by parameters *a*, *b*, and *c*, comparing the energy of the generated spindle allowed us to validate the accuracy of these model parameters. The frequency of the QPS generated spindle was obtained from the parameter estimate ($e/2\pi$) and the spindle frequency obtained from the most dominant peak of the frequency spectrum.

The boxplot in **Figure 9** shows the distribution of energy and frequency error using scorers V4, V5, and both V4 and V5; here the whiskers correspond to $\pm 2.7\sigma$. Assuming normal distribution, the frequency error of $\sim 99.3\%$ of the data set is $\leq 4.6\%$ for scorer V4 and $\leq 6.9\%$ for V5. The energy error is $\leq 15.9\%$ for V4



and $\leq 31.4\%$ for V5. The low variation in frequency error among the scorers as seen in the figure is due to the tighter constraints on frequency values in spindle-marking rule; whereas, the ambiguous definition of spindle amplitude leads to a higher variation in energy error in the two scorers. The subject-specific scoring criteria used by V5 which was based upon each subject's mean peak spindle amplitude meant that their marked spindles fell in a different and narrower amplitude range to the range of spindles marked by V4; this resulted in a low average inter-scorer agreement between V4 and V5 and higher error rate for the QPS model for V5 marked spindles. The relatively low percentage frequency error of the QPS model suggests that it accurately captures the frequency content of spindle. On the contrary, there is a relatively higher energy error as our model attenuates faster than what occurs in actual spindles as seen in **Figure 2**.

Table 3 shows the mean percentage error in energy and frequency of spindles as scored by scorers V4 and V5 and the mean percentage error in energy and frequency of spindles marked in common by both these scorers. As seen, the overall average mean error in energy and frequency for all participants is the lowest for spindles marked by both the scorers (last row of **Table 3**). Furthermore, the same observation holds true for the mean error in energy and frequency for most of the individual participants. Reliable spindle scoring is typically achieved by using only spindles marked by multiple scorers. The lower error rate for commonly marked spindles indicate that the QPS model provides an accurate representation of “reliably” marked spindles.

Detailed Validation on MASS-C1/SS2 Database

Figures 10A–F show the distribution of parameter values for all participants using scorers V4, V5, and both V4 and V5. Here, the whiskers correspond $\pm 2.7\sigma$ of the data set. As seen, parameters d and e have an identical distribution for both scorers, V4 and V5. This indicates that there is greater agreement among the scorers in the frequency and phase content of the signal. Furthermore,

TABLE 3 | Mean percentage error in energy and frequency of spindles.

MASS-C1/SS2 Participant #	Mean energy error			Mean frequency error		
	V4	V5	V4 and V5	V4	V5	V4 and V5
1	4.204	10.474	4.045	1.379	2.764	1.389
2	5.383	10.606	5.079	1.631	2.198	1.610
3	4.059	4.059	4.114	1.602	1.602	1.730
5	5.150	11.524	5.032	1.866	3.090	1.916
6	4.849	11.866	4.685	1.880	4.668	2.085
7	8.567	15.262	7.907	3.023	3.292	2.197
9	5.510	14.340	5.492	2.249	3.618	1.543
10	6.127	15.683	6.242	2.080	5.092	2.022
11	4.509	11.640	4.798	1.569	2.258	1.522
12	8.209	13.598	7.689	2.138	4.087	2.069
13	5.030	12.380	4.850	2.000	1.902	1.299
14	5.380	13.420	5.469	1.639	2.367	1.561
17	3.568	12.662	3.579	1.640	2.886	2.022
18	8.520	11.478	7.020	2.813	3.356	2.549
19	5.144	10.021	5.362	1.677	2.423	1.795
Average	5.614	11.934	5.424	1.946	3.040	1.821

for all parameters, the spindles marked in common by both V4 and V5 show a distribution pattern similar to that of V4. The lower agreement among the scorers regarding spindle amplitude is due to the different scoring criteria used by the scorers. V4 used standard AASM scoring rules while V5 used a subject-specific amplitude threshold to score spindles (Ray et al., 2010).

The error bar in **Figures 11A–F**, with whiskers representing $\pm 2\sigma$ show the distribution of parameter values across all the 15 participants using spindles that have been marked in common by scorers V4 and V5. As seen, the mean values of parameters a and b fall within the narrow range of (2.5, 3) and (0, 1). Additionally, the mean values for parameter e fall within the spindle characteristic frequency range of 11–16 Hz for all participants. The low variance of parameters a and b for scorers, as given in **Table 4**, is in line with spindle amplitude and shape scoring criteria. The table also indicates that parameters c , d , e , and f have the most variation. The variance in e is representative of the 11–16 Hz spindle frequency range. Parameters a , b , and e give spindles the characteristic waxing and waning shape as defined in the AASM guidelines whereas c , d , and f are more likely to account for the intra- and inter-participant variability in the spindle structure. d is more likely to account for the intra-participant variability whereas c and f could impact the inter-participant variability in the spindle structure.

Effect of QPS Model Parameters on Spindle Shape

To evaluate the effect of variation of each QPS parameter on the shape of a marked spindle from the MASS-C1/SS2 database, we linearly increased the value of each of the six parameters of the fitted QPS spindle while keeping the value of the other five parameters constant. We thus regenerated new QPS spindles with

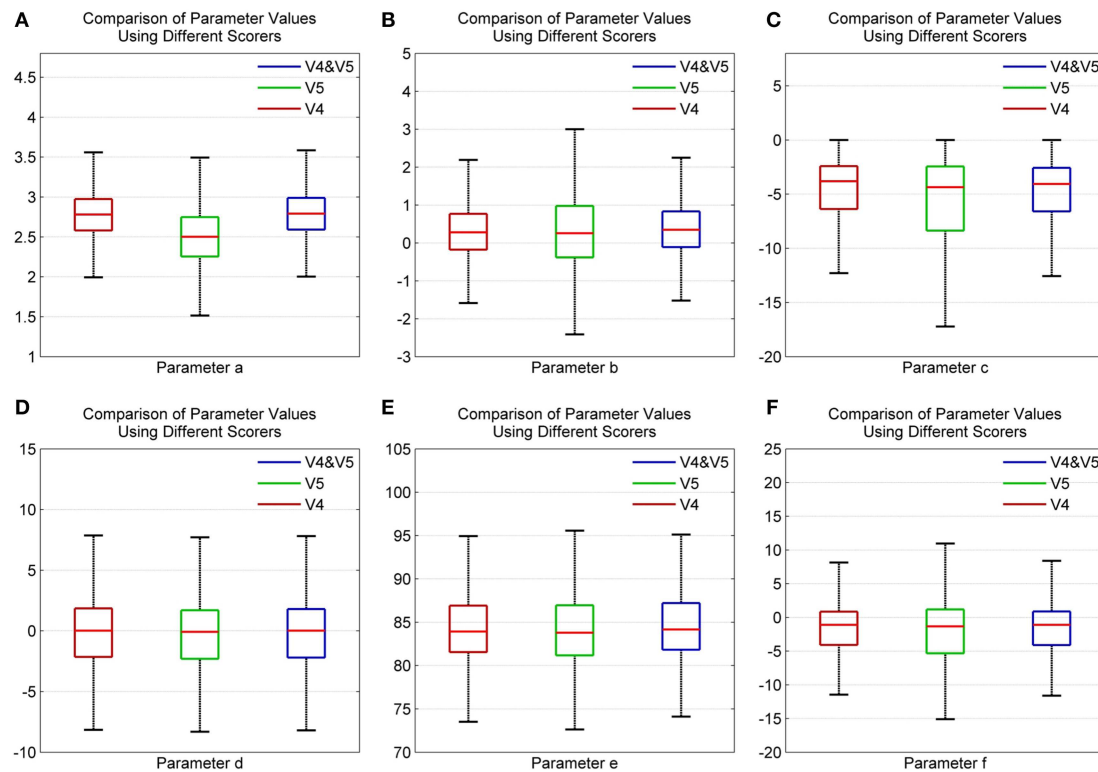


FIGURE 10 | Box plot depicting the comparison of spindle parameter values using different scorers for parameters (A) a (B) b (C) c (D) d (E) e and (F) f using data from MASS-C1/SS2 participants.

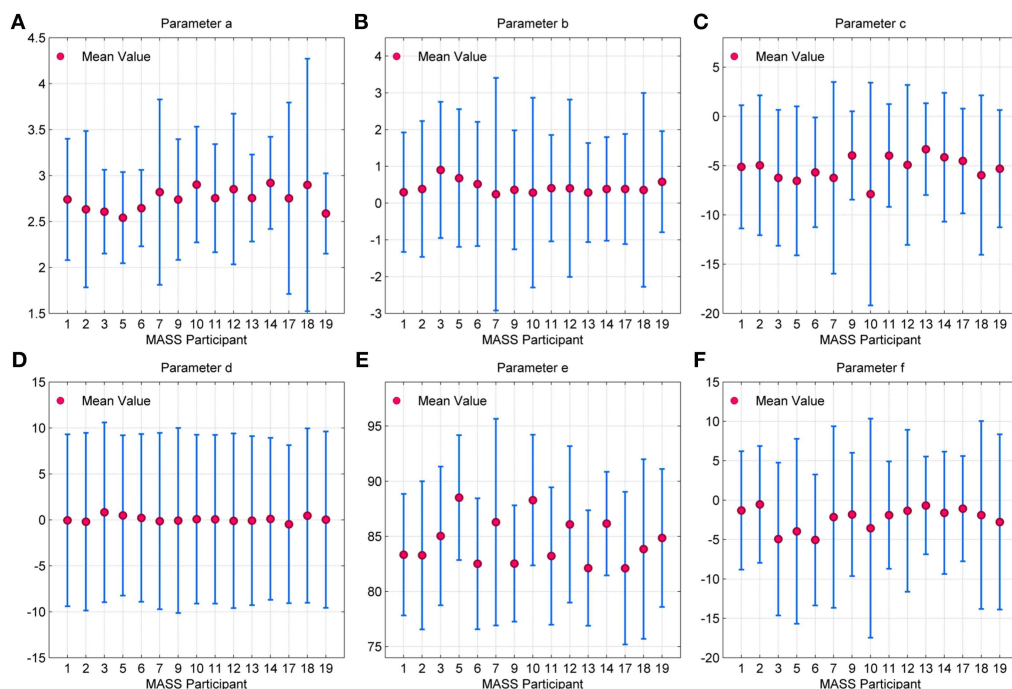


FIGURE 11 | Error bar depicting the distribution of values for parameters (A) a (B) b (C) c (D) d (E) e and (F) f for each MASS-C1/SS2 participant. Here, the whiskers represent 2σ .

TABLE 4 | Variance of parameter values for all participants.

Parameters	V4	V5	V4 and V5
<i>a</i>	0.195	0.575	0.168
<i>b</i>	1.115	3.946	1.087
<i>c</i>	14.252	44.246	15.265
<i>d</i>	21.727	20.761	22.027
<i>e</i>	16.790	20.640	14.928
<i>f</i>	25.677	57.356	24.456

five constant parameters and a linearly increasing sixth parameter. Instead of choosing an arbitrary constant parameter value, the mean value of the other five parameters of spindles from participant 1 of MASS-C1/SS2 (01-02-0001) scored by scorer V4 was used.

Parameter *a*

Figure 12A presents the generated QPS spindle for different values of parameter *a*. The plots demonstrate that increasing the value of *a* increases the peak to peak of the generated spindle but does not impact the sinusoidal content of the signal. As *a* increases from *a* = 1.89 to 2.76, the peak to peak value increases from 13.1 to 31.4, thus signifying that the amplitude of generated QPS spindle has a strong positive correlation to the value of *a*.

Parameter *b*

Figure 12B shows generated QPS spindles with varying values of parameter *b*. It can be observed that parameter *b* values of −1.27, −0.63, and 0.686 changes the peak to peak value of the generated spindles to 28.6, 27.1 and 27.2 respectively, but the relative change in peak to peak value is not as pronounced as the variation caused by change in *a*. **Figure 12B** further illustrates that *b* produces asymmetry in the spindle, with the spindle shifting along the time axis.

Parameter *c*

Figure 12C presents the generated QPS spindles for different values of parameter *c*. The plots indicate that parameter *c* controls the rate of decay while producing minute variations in the peak to peak value of generated QPS spindles. The decay rate decreases with increasing value of *c*. For instance, **Figure 12C** shows that the fastest decay rate occurs with the lowest value of *c* (*c* = −14.2), but as *c* approaches 0 (*c* = −2.77), the generated spindle loses its characteristic “spindle-like” shape. Only a small number of spindles in the MASS-C1/SS2 database had values of *c* approaching 0 (0.5% of all spindles), indicating a low proportion of cases showing a large fitting error.

Parameter *d*

Generated QPS spindles with varying values of parameter *d* are shown in **Figure 12D**. A dashed black line has been added in the individual plots of **Figure 12D** to indicate the value of *t* at which the spindle attains the maximum peak amplitude value. These plots demonstrate that the variation in parameter *d* induces a phase shift in the generated spindle. Since parameters *a*, *b*, *c*, *e*, and *f* are fixed, all three spindles shown here have the same

amplitude and frequency with only the position of the maximal value shifting due to *d* (phase shift).

Parameter *e*

The value of parameter *e* and the corresponding QPS model generated spindle can be seen in **Figure 12E**. The figures indicate that increasing the value of parameter *e* increases the frequency of the generated spindle without affecting its amplitude, thus corroborating that parameter *e* corresponds to the angular frequency of the spindle [see the Accuracy of Energy and Frequency Estimation section]. Discarding outliers, we found all values of *e* to fall within the characteristic spindle frequency range of 11–16 Hz.

Parameter *f*

Figure 12F shows the value of parameter *f* and the corresponding QPS spindle. The initial frequency was fixed at a constant value of 13.2 Hz. As seen here, parameter *f* values of −11.2, −3.62, and 4.4 changes the model frequency to 13.5, 14, and 14 Hz respectively. The figures indicate that increasing the value of parameter *f* induces minor variations in the frequency of generated spindles, thus signifying that the intra-spindle variation in the frequency of the QPS spindle is correlated to the change in *f*.

Variation of Parameter Values in an Overnight Recording

Figure 13 provides the variation in the QPS parameter values for the spindles marked by scorer V4 in the overnight recording of MASS-C1/SS2 participant 3 (01-02-0003). As expected from the results in **Table 4**, the least variation over the night's spindles can be seen in parameters *a* and *b*, whereas the most variation is in *c*, *d*, *e*, and *f*. Interestingly all parameters show a cyclic rise and fall over the course of the night. **Figure 13B** shows a decrease in the variation of parameter *b* and increase in its minima during the middle of the recording. **Figure 13D** also shows a decrease in the variation of parameter *d*, however this occurs later in the recording and is accompanied by a visible dip in the maxima values instead. Parameters *a* and *b* on the other hand show an increase in the peak-to-peak values during the middle of the recording. **Figure 13** gives an example of how the QPS parameter values of spindles in an overnight recording can be tracked to better understand the natural physiological variations that can occur during the night.

Comparison of QPS Spindle and Non-spindle Parameters

In our final experiment, the NLLS algorithm was applied to random non-spindle EEG regions. These were obtained by randomly selecting 500 segments of unmarked EEG data that were 1 second in duration using the two scorers, V4 and V5. The data included all the 15 MASS-C1/SS2 participants and were classified into two groups. The first group contained non-spindles from only sleep stage two (Group 1), whereas the second group contained non-spindles from all sleep stages (Group 2). Special focus was paid to stage two data (Group 1) as spindles are typically observed in EEG during sleep stage two. The resulting set of parameter values given by NLLS were then compared to those obtained from QPS spindles.

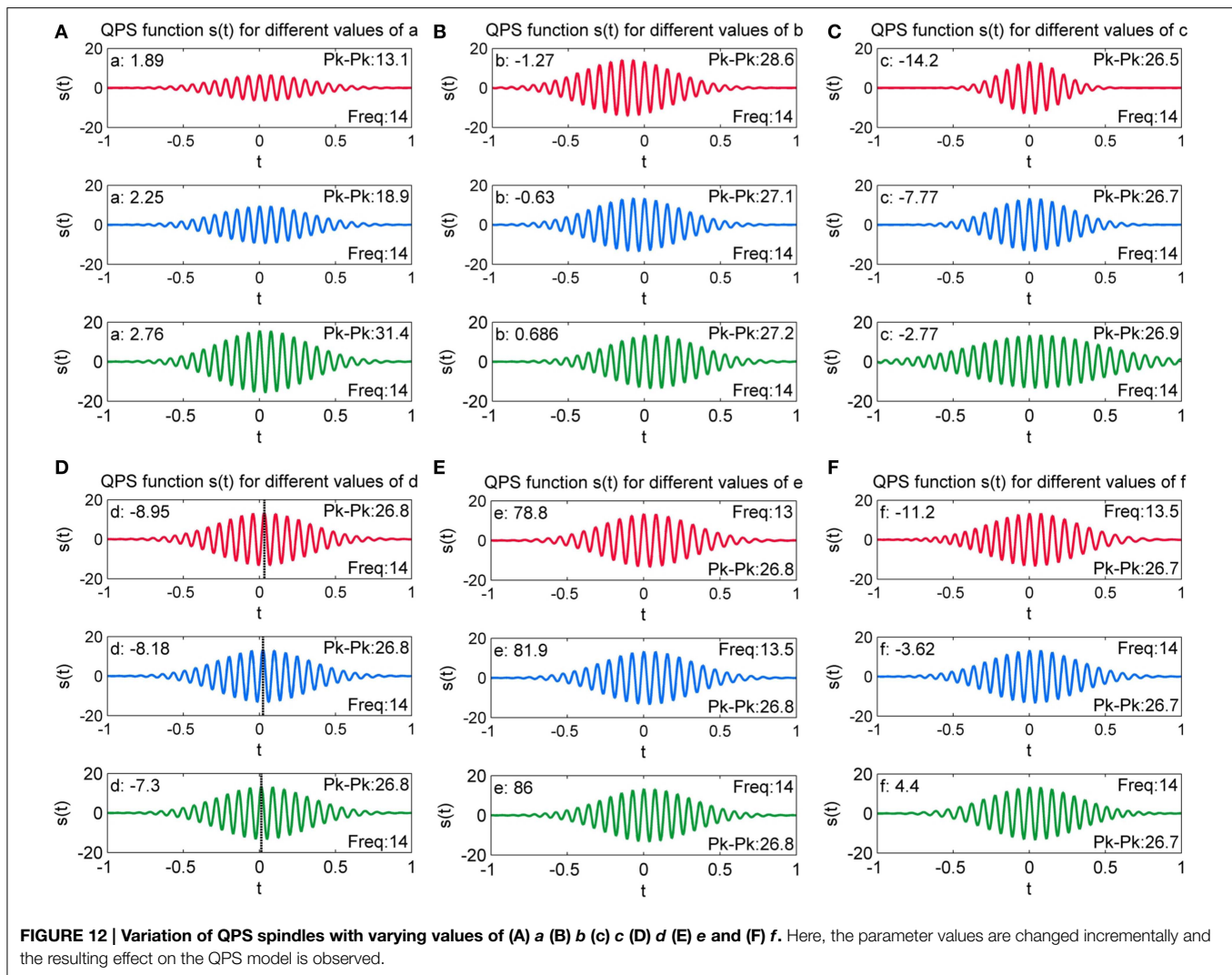


Table 5 shows the results from a two-sided non-parametric t -test comparing parameter values from QPS spindles and non-spindles using the mean spindle parameter values as obtained in the Detailed Validation on MASS-C1/SS2 Database section. Here, $h = 0$ indicates that the null hypothesis (parameter values from spindles and non-spindles come from distributions with equal means) cannot be rejected at a significance level of 1%. The p -value for each parameter is also shown in Table 5. As seen, parameters a and c were significant at the 0.01 level for both scorers and the two groups. With Group 2 non-spindles, b , e , and f were also significantly different from spindles for scorer V5 but not for V4.

Table 6 shows the results from a two-sided non-parametric t -test using different sets of initial conditions for spindles and non-spindles. Given the wide range of possible non-spindle waveforms, the NLLS was initiated with all parameters = 0 for non-spindles, whereas the NLLS was initialized with the mean spindle parameter values for spindles. As seen in Table 6, all parameters show significant difference for both the scorers and the two groups; with the only exception being Group 1 non-spindles for parameter c of scorer V4.

The dependency of the NLLS on the initial conditions limits the parameters of QPS function from accurately differentiating between spindles and non-spindles, as seen from the results in Tables 5, 6. We expected a significant difference in parameter e values for spindles and non-spindles. However when the initial value of parameter e for both non-spindles and spindles was set at 90, the value corresponding to the spindle frequency, parameter e values for non-spindles converged to a local minimum close to that value; significant differences in parameter e values could thus not be observed in the results from scorer V4. The difference in parameters a and c for all the groups using both scorers indicates significant difference in the amplitude variations of QPS spindles and non-spindles. Using different initial conditions for non-spindles resulted in significant differences for all parameter values from those of spindles.

Discussion and Conclusions

In this paper we proposed a new method to model the instantaneous frequency and amplitude variations occurring within sleep

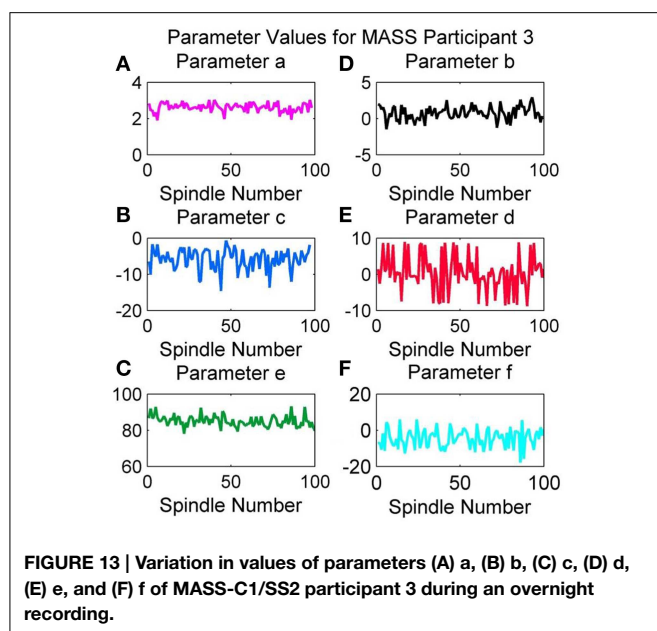


FIGURE 13 | Variation in values of parameters (A) a, (B) b, (C) c, (D) d, (E) e, and (F) f of MASS-C1/SS2 participant 3 during an overnight recording.

TABLE 5 | Results of two sided *t*-test comparing parameters obtained from QPS spindles and non-spindles using same initial conditions.

Parameters	V4 (Group 2)		V5 (Group 2)		V4 (Group 1)		V5 (Group 1)	
	<i>h</i>	<i>p</i>	<i>h</i>	<i>p</i>	<i>h</i>	<i>p</i>	<i>h</i>	<i>p</i>
<i>a</i>	1	0	1	0	1	0	1	0
<i>b</i>	0	0.112	0	0.052	0	0.263	1	0
<i>c</i>	1	0	1	0	1	0	1	0
<i>d</i>	0	0.051	0	0.429	0	0.559	0	0.379
<i>e</i>	0	0.809	1	0.003	0	0.550	1	0
<i>f</i>	0	0.063	1	0	0	0.011	1	0

spindles. Our proposed QPS model is able to account for the non-stationarity observed in sleep spindles within the analysis window by accurately approximating the frequency and logarithmic amplitude of the signal using quadratic functions of time. Our results illustrate that QPS successfully models the various intra-spindle characteristics within its six parameters. Parameter estimation using standard NLLS methods resulted in good convergence and was robust in the presence noise, both of which are vital given the presence of background EEG. The relative error in frequency estimates was less than 5% when compared to the dominant peak in the spindle frequency spectrum for a majority of the participants.

The reversibility between the determined parameters and signal waveform is also an important characteristic of the QPS modeling. As seen in **Figures 4, 7**, it is possible to regenerate a cleaner version using the QPS parameters. Unlike other techniques, the QPS model also provides the instantaneous phase, which is indispensable in signal reconstruction. The results in the Validation of QPS Model on Simulated Spindles section show that it possible to use the QPS to regenerate cleaner versions of spindles in EEG with large artifacts and background noise. The noise component

TABLE 6 | Two sample *t*-test result comparing parameters obtained from QPS spindles and non-spindles using different initial conditions.

Parameters	V4 (Group 2)		V5 (Group 2)		V4 (Group 1)		V5 (Group 1)	
	<i>h</i>	<i>p</i>	<i>h</i>	<i>p</i>	<i>h</i>	<i>p</i>	<i>h</i>	<i>p</i>
<i>a</i>	1	0	1	0	1	0	1	0
<i>b</i>	1	0	1	0	1	0	1	0
<i>c</i>	1	0.003	1	0	0	0.331	1	0
<i>d</i>	1	0	1	0	1	0	1	0
<i>e</i>	1	0	1	0	1	0	1	0
<i>f</i>	1	0	1	0	1	0	1	0

identified in the spindles could then be used to de-noise adjacent areas of the sleep EEG.

Characterizing sleep spindles using the QPS parameters could help restrict the inconsistency in scoring due to the differing subjective interpretation of scorers, which will in turn assist in the proper training and tuning of accurate sleep spindle detectors. As seen in **Table 4**, parameters *c*, *d*, and *f* had the most variation. The broad AASM definition for sleep spindles currently leaves the manual marking of spindles in EEG data open to some interpretation, leading to low inter-expert agreement for spindle scoring. Thus, the accuracy of sleep spindle detectors when trained and tested using data scored from a single scorer can fall significantly when tested using data scored by other experts. Spindle scoring reliability is typically reduced by having multiple scorers detect spindles manually and accepting only commonly marked spindles. The frequency (1.8%) and energy (5.4%) error estimates for the QPS model were lowest for the spindles marked by both scorers (**Table 3**), indicating that it provides a more accurate representation of the “reliably” marked spindles. Providing guidance on an acceptable range for all QPS parameters in the spindle scoring criteria using spindles marked by multiple scorers can help reduce scoring inconsistencies.

Accurately characterizing the structure of sleep spindles could enable researchers develop a better understanding of the relationship between sleep spindles and various physiological phenomena such as sleep “stability,” memory formation and other pathological problems, e.g., depression, epilepsy, Parkinson, Alzheimer and schizophrenia (Wei et al., 1999; Bódizs et al., 2005; Fogela and Smith, 2011; Wamsley et al., 2012; Tezer et al., 2014). The relationship of spindle amplitude and frequency, from parameter *a* and *e* with these phenomena have been researched. However, their impact on the rate of decay of the spindle envelope (*c*), the phase shift (*d*) and frequency variation (*f*) have not been studied to date. The QPS parameters offer quantitative representations of spindle structure that can be interpreted visually, as presented in the Effect of QPS Model Parameters on Spindle Shape section. Variations in these parameters can be analyzed to determine if they are disorder, scorer or participant specific.

Additional potential uses of the QPS model include the generation of a wide range of simulated spindles to help accurately train automatic detectors as well as manual scorers. The simulated QPS spindles can also be utilized to provide a reference to define more precise scoring rules, normalize real spindles from

multiple participants and also compare real spindles against to track naturally or pathologically occurring variations.

The similarity in distribution patterns and limited range of the QPS parameter values (**Figure 11**) indicate that there is potential in their use in an automatic spindle scoring algorithm. The results in the Comparison of QPS Spindle and Non-Spindle Parameters section however show that NLLS estimation are highly dependent on the initial conditions used. The parameter values showed significant difference between the two groups when different initial conditions were used for spindles and non-spindles. However when the same initial conditions were used for both groups, surprisingly only the amplitude based parameters a and c were significant and not the frequency based e . These results indicated that the QPS model can only be used for spindle detection if preceded by a priori parameter estimation to obtain the initial conditions to be used in the NLLS for each epoch or an alternative QPS parameter estimation technique, e.g., an analytical method, that does not depend on initial conditions is utilized instead of the NLLS algorithm.

In this study, parameter estimation was performed using the NLLS algorithm. Results obtained with NLLS need to be compared with other parameter estimation techniques. Furthermore, as discussed above NLLS results can depend on the chosen initial conditions. Like other recursive methods, the NLLS

algorithm can be computationally expensive. Future work will include identifying more robust algorithms for parameter estimation including analytical methods, thus overcoming the burden of initial conditions and ensuring global convergence. Simplified as well as expanded versions of the QPS model with more parameters also need to be explored as they may enhance the characterization of spindle structure.

We also intend to use the QPS parameters to develop an accurate sleep spindle detection algorithm, taking into account the limitations stated above and test it on spindles from the MASS-C1/SS2 database. The accuracy of the automated detector will be compared to existing spindle detector available through the Spynldle toolbox. Finally, as mentioned earlier, the QPS model opens up the potential to examine in detail the impact of sleep abnormalities and disorders as well as other physiological processes on sleep spindles.

Acknowledgments

The work was supported by NPRP grant # [5-1327-2-568] from the Qatar National Research Fund which is a member of Qatar Foundation. The statements made herein are solely the responsibility of the authors.

References

- Abe, M., and Smith, J. O. (2005). "AM/FM rate estimation for time-varying sinusoidal modeling," in *IEEE International Conference on Acoustics, Speech, and Signal Processing* (Philadelphia, PA), 201–204.
- Bódizs, R., Kis, T., Lázár, A. S., Havrán, L., Rigó, P., Clemens, Z., et al. (2005). Prediction of general mental ability based on neural oscillation measures of sleep. *J. Sleep Res.* 14, 285–292. doi: 10.1111/j.1365-2869.2005.00472.x
- Bódizs, R., Körmendi, J., Rigó, P., and Lázár, A. (2009). The individual adjustment method of sleep spindle analysis: methodological improvements and roots in the fingerprint paradigm. *J. Neurosci. Methods* 178, 205–213. doi: 10.1016/j.jneumeth.2008.11.006
- Campbell, K., Kumar, A., and Hofman, W. (1980). Human and automatic validation of a phase-locked loop spindle detection. *Electroencephalogr. Clin. Neurophysiol.* 45, 602–605.
- Fogela, S. M., and Smith, C. T. (2011). The function of the sleep spindle: a physiological index of intelligence and a mechanism for sleep-dependent memory consolidation. *Neurosci. Biobehav. Rev.* 35, 1154–1165. doi: 10.1016/j.neubiorev.2010.12.003
- Huupponen, E., Gómez-Herrero, G., Saastamoinen, A., Värrä, A., Hasan, J., and Himanen, S. (2007). Development and comparison of four sleep spindle detection methods. *Artif. Intell. Med.* 40, 157–170. doi: 10.1016/j.artmed.2007.04.003
- Iber, C., Ancoli-Israel, S., Chesson, A., and Quan, S. (2007). *The AASM Manual for the Scoring of Sleep and Associated Events: rules, Terminology and Technical Specifications*. Westchester: American Academy of Sleep Medicine.
- Ito, M., and Yano, M. (1989). Sinusoidal modeling for nonstationary voiced speech based on a local vector transform. *J. Acoust. Soc. Am.* 121, 763–765.
- Ito, M., and Yano, M. (2007). Sinusoidal modeling for nonstationary voiced speech based on a local vector transform. *J. Acoust. Soc. Am.* 121, 1717–1727. doi: 10.1121/1.2431581
- Ktonas, P., Golemati, S., Xanthopoulos, P., Sakkalis, V., Ortigueira, M., Tsekou, H., et al. (2007). "Potential dementia biomarkers based on the time-varying microstructure of sleep EEG spindles," in *29th Annual International Conference of the IEEE EMBS* (Lyon: Cité Internationale).
- Ktonas, P. Y., Golemati, S., Xanthopoulos, P., Sakkalis, V., Ortigueira, M., Tsekou, H., et al. (2009). Time-frequency analysis methods to quantify the time-varying microstructure of sleep EEG spindles: Possibility for dementia biomarkers? *J. Neurosci. Methods* 185, 133–142. doi: 10.1016/j.jneumeth.2009.09.001
- Kunz, D., Danker-Hopfe, H., Gruber, G., Lorenzo, J. L., Himanen, S. L., Kemp, B., et al. (2000). Interrater reliability between eight European sleep-labs in healthy subjects of all age-groups. *J. Sleep Res.* 9, 106. doi: 10.1515/bmte.2000.45.s1.433
- Marques, J. S., and L. B., Almeida (1989). Frequency-varying sinusoidal modeling of speech. *IEEE Trans. Acoust. Speech Signal Process.* 5, 763–765. doi: 10.1109/29.17571
- Nonclercq, A., Urbain, C., Verheulpen, D., Decaestecker, C., Bogaert, P. V., and Peigneux, P. (2013). Sleep spindle detection through amplitude-frequency normal modelling. *J. Neurosci. Methods* 214, 192–203. doi: 10.1016/j.jneumeth.2013.01.015
- 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
- Olbrich, E., and Achermann, P. (2008). Analysis of the temporal organization of sleep spindles in the human sleep EEG using a phenomenological modeling approach. *J. Biol. Phys.* 34, 341–349. doi: 10.1007/s10867-008-9078-z
- 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
- Perumalsamy, V., Sankaranarayanan, S., and Rajamony, S. (2009). Sleep spindles detection from human sleep EEG signals using autoregressive (AR) model: a surrogate data approach. *J. Biomed. Sci. Eng.* 2, 294–303. doi: 10.4236/jbise.2009.25044
- Ray, L. B., Fogel, S. M., Smith, C. T., and Peters, K. R. (2010). Validating an automated sleep spindle detection algorithm using an individualized approach. *J. Sleep Res.* 19, 374–378. doi: 10.1111/j.1365-2869.2009.00802.x

- Rechtschaffen, A., and Kales, A. (1968). *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*. Washington, DC: Public Health Service, U.S. Government Printing Office.
- Schönwald, S., Santa-Helena, E. D., Rossatto, R., Chaves, M., and Gerhardt, G. (2006). Benchmarking matching pursuit to find sleep spindles. *J. Neurosci. Methods* 156, 314–321. doi: 10.1016/j.jneumeth.2006.01.026
- Tezer, I., Rémi, J., Erbil, N., Noachtar, S., and Saygi, S. (2014). A reduction of sleep spindles heralds seizures in focal epilepsy. *Clin. Neurophysiol.* 125, 2207–2211. doi: 10.1016/j.clinph.2014.03.001
- Wamsley, E. J., Tucker, M. A., Shinn, A. K., Ono, K. E., McKinley, S. K., Ely, A. V., et al. (2012). Reduced sleep spindles and spindle coherence in schizophrenia: mechanisms of impaired memory consolidation? *Biol. Psychiatry* 41, 154–161. doi: 10.1016/j.biopsych.2011.08.008
- Wei, H. G., Riel, E., Czeisler, C. A., and Dijk, D. J. (1999). Attenuated amplitude of circadian and sleep-dependent modulation of electroencephalographic sleep spindle characteristics in elderly human subjects. *Neurosci. Lett.* 260, 29–32. doi: 10.1016/S0304-3940(98)00851-9
- Wendt, S. L., Welinder, P., Sorensen, H. B., Peppard, P. E., Jennum, P., Perona, P., et al. (2015). Inter-expert and intra-expert reliability in sleep spindle scoring. *Clin. Neurophysiol.* doi: 10.1016/j.clinph.2014.10.158. [Epub ahead of print].
- Xanthopoulos, P., Golemati, S., Sakalis, V., Ktonas, P. Y., Zervakis, M., and Soldatos, C. R. (2006). “Modeling the time-varying microstructure of simulated sleep EEG spindles using time-frequency analysis methods,” in *28th IEEE EMBS Annual International Conference* (New York, NY).

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.

Copyright © 2015 Palliyali, Ahmed and Ahmed. 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.



Sleep spindle and slow wave frequency reflect motor skill performance in primary school-age children

Rebecca G. Astill^{1,2}, Giovanni Piantoni^{1,3}, Roy J. E. M. Raymann¹, Jose C. Vis^{1,4}, Joris E. Coppens⁵, Matthew P. Walker⁶, Robert Stickgold⁷, Ysbrand D. Van Der Werf^{8,9} and Eus J. W. Van Someren^{1,10*}

¹ Department of Sleep and Cognition, Netherlands Institute for Neuroscience, Royal Netherlands Academy of Arts and Sciences, Amsterdam, Netherlands

² Department of Clinical Neurophysiology, Amsterdam Sleep Centre, Slotervaartziekenhuis, Amsterdam, Netherlands

³ Department of Neurology, Massachusetts General Hospital, Boston, MA, USA

⁴ Sleepvision, Berg en Dal, Netherlands

⁵ Department of Technology and Software Development, Netherlands Institute for Neuroscience, Royal Netherlands Academy of Arts and Sciences, Amsterdam, Netherlands

⁶ Sleep and Neuroimaging Laboratory, Department of Psychology, University of California, Berkeley, CA, USA

⁷ Department of Psychiatry, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, USA

⁸ Department of Emotion and Cognition, Netherlands Institute for Neuroscience, Royal Netherlands Academy of Arts and Sciences, Amsterdam, Netherlands

⁹ Department of Anatomy and Neurosciences, VU University and Medical Center, Amsterdam, Netherlands

¹⁰ Departments of Integrative Neurophysiology and Medical Psychology, Center for Neurogenomics and Cognitive Research (CNCR), Neuroscience Campus Amsterdam, VU University and Medical Center, Amsterdam, Netherlands

Edited by:

Simon C. Warby, Stanford University, USA

Reviewed by:

Julie Carrier, Université de Montréal, Canada

Reut Gruber, McGill University, Canada

*Correspondence:

Eus J. W. Van Someren,
Department of Sleep and Cognition,
Netherlands Institute for
Neuroscience, Royal Netherlands
Academy of Arts and Sciences,
Meibergdreef 47, 1105 BA
Amsterdam, Netherlands
e-mail: e.van.someren@
nin.knaw.nl

Background and Aim: The role of sleep in the enhancement of motor skills has been studied extensively in adults. We aimed to determine involvement of sleep and characteristics of spindles and slow waves in a motor skill in children.

Hypothesis: We hypothesized sleep-dependence of skill enhancement and an association of interindividual differences in skill and sleep characteristics.

Methods: 30 children (19 females, 10.7 ± 0.8 years of age; mean \pm SD) performed finger sequence tapping tasks in a repeated-measures design spanning 4 days including 1 polysomnography (PSG) night. Initial and delayed performance were assessed over 12 h of wake; 12 h with sleep; and 24 h with wake and sleep. For the 12 h with sleep, children were assigned to one of three conditions: modulation of slow waves and spindles was attempted using acoustic perturbation, and compared to yoked and no-sound control conditions.

Analyses: Mixed effect regression models evaluated the association of sleep, its macrostructure and spindles and slow wave parameters with initial and delayed speed and accuracy.

Results and Conclusions: Children enhance their accuracy only over an interval with sleep. Unlike previously reported in adults, children enhance their speed independent of sleep, a capacity that may to be lost in adulthood. Individual differences in the dominant frequency of spindles and slow waves were predictive for performance: children performed better if they had less slow spindles, more fast spindles and faster slow waves. On the other hand, overnight enhancement of accuracy was most pronounced in children with more slow spindles and slower slow waves, i.e., the ones with an initial lower performance. Associations of spindle and slow wave characteristics with initial performance may confound interpretation of their involvement in overnight enhancement. Slower frequencies of characteristic sleep events may mark slower learning and immaturity of networks involved in motor skills.

Keywords: children, learning, motor skill, memory, sleep, spindles, slow waves, frequency

INTRODUCTION

The importance of sleep for learning and memory processes has been established firmly. A large number of studies in adults have shown that sleep contributes to efficient consolidation of

both declarative memory—the memory for facts and events—and procedural memory—the memory for skills and procedures (Maquet, 2001; Walker and Stickgold, 2004; Stickgold and Walker, 2005; Diekelmann et al., 2009; Rasch and Born, 2013;

Landmann et al., 2014). Sleep does more than merely prevent forgetting by providing a time-period without interference: for certain motor skills, sleep can even enhance performance without further training. In adults, a contribution of sleep may have been demonstrated most robustly for the consolidation and enhancement of newly learned visuomotor skills, especially of a finger-sequence tapping task (Walker et al., 2002; Morin et al., 2008; Van Der Werf et al., 2009b; Barakat et al., 2011, 2013; Albouy et al., 2013a). This task requires participants to tap a particular sequence with their fingers as fast and accurately as possible. It has been consistently shown that performance on this task saturates to a certain individual level, without further improvement unless participants try again after a period of sleep. Only if participants sleep within a certain time window after their first saturating training session, does their subsequent performance improve by about 10–20% without further training (Walker et al., 2002; Van Der Werf et al., 2009b).

What are the neuronal processes underlying this performance enhancement by sleep? Numerous studies, mostly in adults, have investigated the specific aspects of sleep-electroencephalography (EEG) that could provide clues to neuronal processes involved. These investigations have addressed qualitative aspects of the sleep-EEG macrostructure, including sleep stages, as well as quantitative aspects of the sleep-EEG, notably its power spectrum and the microstructural discrete events of sleep spindles and slow waves. Investigations of qualitative aspects of the sleep-EEG aspects of sleep revealed that overnight skill enhancement is associated with the amount of stage 2 sleep, especially in the later part of the night (Walker et al., 2002). This finding immediately points to the involvement of a specific microstructural aspect of the sleep-EEG, because stage 2 sleep is characterized by the appearance of sleep spindles (Rechtschaffen and Kales, 1968). These transient bursts of about 12–15 Hz activity reflect thalamo-cortical oscillations (Steriade, 2006). Indeed, sleep spindles have repeatedly been linked to procedural memory consolidation and enhancement (for a review see Fogel and Smith, 2011).

Along a continuum of dominant frequencies, spindles have been divided into slower and faster spindles (Feld and Born, 2012). Slow spindles dominate over frontal EEG derivations and are thought to involve the superior frontal gyrus, while fast spindles show up stronger in central and parietal EEG derivations and are thought to involve the precuneus, hippocampus, medial frontal cortex, and sensorimotor areas (Schabus et al., 2007; Dehghani et al., 2011). Relevant to the present study, the topographic representation of sleep spindles change with age (Tanguay et al., 1975; Shinomiya et al., 1999). Frontal spindles are more prominent in younger children while older children show more centroparietal spindles (Shinomiya et al., 1999).

Slow spindles are more pronounced during slow wave sleep. The slow waves of sleep represent alternating periods of hyperpolarization (down-states) and depolarization (up-states) of neurons in the cerebral cortex. Spindles are especially likely to occur at the transition to the down-state of a slow oscillation. Fast spindles occurring during slow wave sleep are more likely to occur at the transition from the down-state to the up-state

(Möller and Born, 2011). Fast spindles are most prominent during stage 2 sleep (Feld and Born, 2012). In their original study, Walker et al. (2002) showed that overnight skill enhancement is associated with the amount of stage 2 sleep, especially in the later part of the night where slow wave activity (SWA) hardly occurs. In accordance with this initial observation, fast spindles have commonly been associated with overnight enhancement of a visuomotor skill (Nishida and Walker, 2007; Tamaki et al., 2008; Barakat et al., 2011), with the overnight restoration of episodic learning ability (Mander et al., 2011) and with the overnight integration of new information in existing knowledge (Tamminen et al., 2010, 2013). Nevertheless, at least one study suggests that slow spindles rather than fast spindles are important in overnight cognitive processing (Holz et al., 2012).

In addition to spindles, slow waves have also been associated with sleep-dependent performance enhancement, possibly correlated with the role of spindles (Holz et al., 2012). The overnight enhancement of an implicit visuomotor skill is associated with the increase in slow wave power the pre-sleep training elicits in subsequent sleep (Huber et al., 2004; Määttä et al., 2010). Relevant to the present study, Kurth et al. (2012) showed in children that the maturation of simple motor skills, complex motor skills, visuomotor skills, language skills and cognitive control skills is predicted by the topographical distribution of SWA.

In contrast to adults, far less is known however about the role of sleep and associated oscillations in memory consolidation across childhood. Some studies have reported a sleep-dependent consolidation of declarative memory (Fischer et al., 2007; Backhaus et al., 2008; Wilhelm et al., 2008), but no overnight enhancement of skills (Fischer et al., 2007; Wilhelm et al., 2008). However, closer inspection of the data obtained in the finger-tapping task and mirror tracing skill tasks has indicated that children's performance is significantly improved, both across offline periods of sleep and wakefulness (Wilhelm et al., 2008; Prehn-Kristensen et al., 2009). Moreover, 9- and 12-year old children showed less susceptibility to daytime interference of a newly acquired motor memory than 17 year olds (Dorfberger et al., 2007). This supports the interpretation that children have the capacity for memory consolidation over periods of both sleep and wakefulness, the latter being diminished or even lost with the development into adulthood.

With respect to the involvement of sleep specific sleep oscillations in performance enhancement in children, Kurdziel et al. (2013) found that a daytime nap in 4 year old children enhanced recall on a hippocampal-dependent visuospatial task resembling the card-deck "Memory" game. Moreover, sleep spindle density during the intervening nap was positively correlated with the memory performance benefit ($r = 0.65$). However, these memory associations may have been secondary to a negative correlation of spindle density with initial baseline memory performance ($r = -0.67$), thereby offering more improvement opportunity in children with lower baseline ability. Of note, a negative correlation of spindle density with baseline performance was also reported in 4–8 year old children (Chatburn et al., 2013).

Building on these prior findings, the first aim of the present study was to address the hypothesis that motor skill enhancement

is dependent on sleep in school-aged children, as it has been reported to be in adults. The second aim was to determine whether both baseline motor skill performance and offline enhancements were significantly predicted by specific aspects of the sleep-EEG. In particular, we focused on the role of fast and slow sleep spindles and slow waves of sleep. Thirdly, to attain support for the hypotheses beyond observational correlations between sleep and memory in children, we implemented an experimental manipulation aimed at changing spindles and slow waves, thus exploring causality. Pharmacological manipulation of spindle density affects sleep-dependent performance enhancement of sequence finger tapping (Rasch et al., 2009) but may not easily be approved of by medical ethics committees for application in children, and may induce other systematic effects. We therefore aimed to manipulate spindles and slow waves only during slow wave sleep, using a validated selective acoustic interference of sleep at the first occurrence of slow waves (Van Der Werf et al., 2009a). This method selectively and effectively suppresses slow waves (Van Der Werf et al., 2009a) and therefore their co-occurrence with spindles, allowing for a better discrimination of the role of sleep spindles vs. slow waves, and sleep spindles that occur in stage 2 vs. those that occur in slow wave sleep. Moreover, since fast spindles are more prominent during stage 2 sleep and slow spindles occur more pronounced during slow wave sleep, selective suppression of slow waves further offers the ability to more clearly disambiguate the role of fast vs. slow spindles in memory processing.

METHODS

PARTICIPANTS

Participants were recruited through a national competition designed to promote an interest in science amongst primary schools in the Netherlands. The two final school classes of the winning school were invited to take part in the current study. For ethical reasons, all children for which informed consent was obtained participated in the experiment, including children with diagnosed psychiatric or neurological illnesses. By allowing them to participate, their condition remained concealed to their peers. Their data were however excluded from analysis. The data of two participants were excluded because of a diagnosis with Pervasive Developmental Disorder—Not Otherwise Specified (PDD-NOS). Useful data were obtained from 30 participants, 19 females (10.7 ± 0.8 years; mean \pm SD). No apparent sleep disorders were present as indicated by Dutch translations of the abbreviated Child's Sleep Habits Questionnaire (CSHQ, cutoff score 41; Owens et al., 2000b) and Sleep Disturbance Scale for Children (SDSC, cutoff score 39; Bruni et al., 1996) filled out by the parents and the Sleep Self Report (Owens et al., 2000a) filled out by the children. The local medical ethics committee approved of the procedures and written informed consent was obtained from the parents.

PROCEDURAL TASK

The current study used a paradigm frequently employed to examine sleep-dependent procedural performance enhancement in adults: the finger-tapping task (Karni et al., 1995; Walker et al., 2002). The task consists of two sessions: an initial learning

acquisition session, followed by an offline time period of either wake or sleep, after which there was a delayed recall test session to investigate the development of offline performance changes, relative to the end of the initial acquisition session. In the current version, each learning session consisted of 12 trials of 23-s duration, separated by 20-s breaks. The delayed recall session consisted of six additional trials, again separated by 20-s breaks. During a trial, participants were asked to continuously tap a five-digit sequence on a computer keyboard (e.g., 4-1-3-2-4) as fast and as accurately as possible with their non-dominant hand. Four parallel versions of the task were used and these were counterbalanced across participants and across the four experimental conditions: 41324, 32413, 14231 and 23142.

Key-presses were recorded using E-prime (Psychology Software Tools Inc., Pittsburgh, USA) and processed to derive two main variables of interest for each trial: (1) speed, i.e., the number of correct sequences per 23-s trial; and (2) accuracy, i.e., the percentage of key taps that resulted in correct sequences, relative to all key taps.

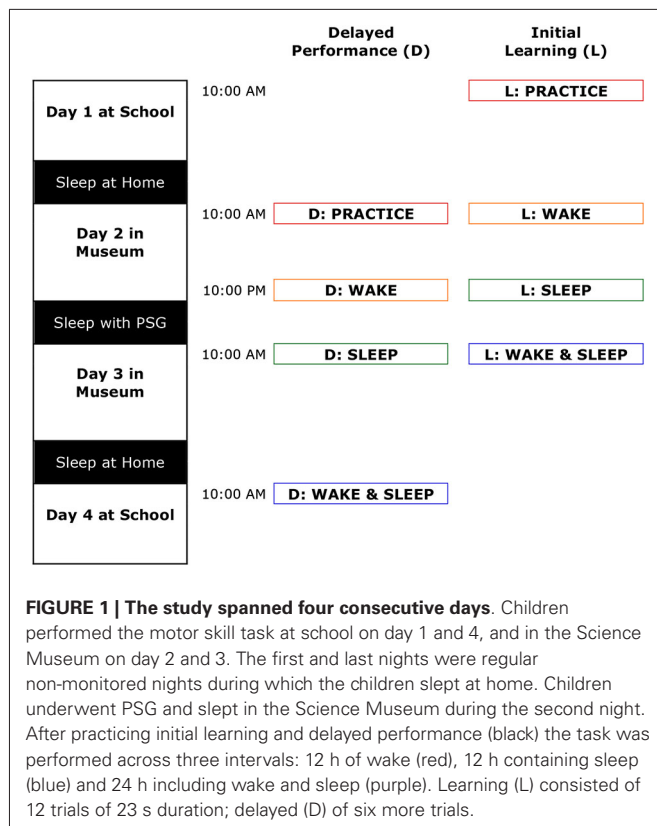
EXPERIMENTAL DESIGN

Using a repeated-measures design, participants performed finger-tapping learning and recall sessions three times, preceded by an additional initial acquisition learning (L) and recall (R) practice sessions to get familiar with the task. Assessments spanned four consecutive weekdays with morning sessions at 10:00 AM and the evening session at 10:00 PM. As indicated in **Figure 1**, after the initial learning and recall practice sessions, performance changes were assessed in a fixed order over the following intervals: (1) 12 h containing wake (the *Wake* interval); (2) 12 h including sleep (the *Sleep* interval); and (3) 24 h including both wake and sleep (the *Wake & Sleep* interval). In the 12-h *Sleep* interval, participants stayed in individual bedrooms in a purposefully built sleep-lab in the Science Museum "Nemo" (Amsterdam, Netherlands) for polysomnography (PSG) recordings. Every three children were supervised by at least one sleep technician. The nights in-between the learning and recall training sessions and the *Wake & Sleep* interval were spent at home, during which the children slept in their own bed as per usual.

POLYSOMNOGRAPHY (PSG)

During the 12-h *Sleep* interval, participants were fit with eight Au electrodes: two for electroencephalography (EEG) on frontopolar (FPz) and central (Cz) positions according to the 10–20 system, two for electrooculography (EOG) placed diagonally across the eyes, two for electromyography (EMG) attached submentally, a ground electrode positioned on the forehead and a reference electrode (A1) fit on the left mastoid. Polysomnography was performed using the Embla A10 system (Flaga hf, Reykjavik, Iceland). Data were recorded online, and transferred onto a personal computer. The Embla A10 system initially samples the data at 2000 Hz and subsequently down-samples it digitally to 200 Hz. Filtering was limited to the Embla's integrated highpass DC filter at 1 Hz (-3 dB at 0.3 Hz) and 50 Hz notch filter (1 Hz bandwidth).

During the night that the children spent in the sleep-lab, they were randomly assigned to one of three acoustic manipulation



conditions. All children wore in-ear headphones. The first condition has been described previously (Van Der Werf et al., 2009a) and aimed at suppressing slow wave sleep. In brief, we developed a custom analysis plug-in for the Somnologica 2 software (Flaga, Reykjavik, Iceland) that performed online calculation of the relative contribution of the SWA band (0.4–4 Hz) to the frequency spectrum as a measure of the depth of sleep. When the contribution of SWA exceeded a threshold level, the headphone emitted a beeping noise that continued to increase in amplitude in six discrete steps until it reached a maximum. The sound continued until the level of SWA dropped below the threshold. To avoid erroneous inclusion of slow EOG signals in the 0.4–4 Hz EEG band, the sound was not emitted when the signals from the two EOG leads were negatively correlated, reflecting conjugated eye movements; a positive correlation reflects leakage of SWA into the EOG leads. Using this system, we have successfully achieved slow wave sleep suppression in elderly volunteers (Van Der Werf et al., 2009a).

The second acoustic manipulation condition concerned a yoked control group, who received the same auditory stimuli, but unrelated to their own slow wave sleep. They received a copy of the auditory stimuli that were given in a closed-loop way to their sleeping neighbor. Finally, the third, placebo, condition consisted of merely wearing the in-ear headphones without providing any acoustic stimulation.

Children were blinded to the condition they were assigned to and were told that tones would be played in the night, but that they might not become aware of them.

EEG ANALYSIS

Macrosleep

Electroencephalography was scored visually, blinded to the condition, in 30-s epochs using Somnologica software (Flaga hf, Reykjavik, Iceland) according to standard sleep scoring criteria (Rechtschaffen and Kales, 1968) with the adaptation of viewing EEG at 100 μ V/cm instead of the recommended 50 μ V/cm, to account for the very large amplitude of sleep EEG oscillations in children (Piantoni et al., 2013a). Macrosleep variables quantified were Time In Bed (TIB), Total Sleep Time (TST), Sleep Onset Latency, Latency to the First REM epoch, Wake after Sleep Onset, Sleep Efficiency and the Percentages of Stage 1, 2, SWS and REM sleep relative to TST.

Preprocessing for quantitative EEG analysis

The visual scoring included a rating of presence of artifacts. Epochs of 30 s that contained even the slightest artifact, including an arousal, were omitted from quantitative EEG analyses.

Spindles

Automated spindle detection was performed using a previously reported algorithm (Ferrarelli et al., 2007) implemented in Matlab (The MathWorks Inc, Natick, USA). Artifact-free EEG in stages S2, S3, and S4 across the entire night was bandpass-filtered between 9 and 15 Hz using an infinite impulse response filter (Figures 4A,B). We then computed the time-course of the amplitude by taking the envelope of the filtered signal (Figure 4B). For each channel and participant, the mean of the envelope over the artifact-free stages S2, S3, and S4 was used to calculate the upper threshold: all amplitude fluctuations of the filtered signal surpassing 4.5-fold the average amplitude value calculated above were considered putative spindles (Figure 4C). The beginning and end of each spindle was defined by a lower threshold, set at 25% of the upper threshold value (Figure 4C). A minimal duration of 450 ms was used to avoid the detection of brief events. Visual inspection of the performance of the automated algorithm indicated the need of slight adaptations in the parameter settings as compared to the settings used in Ferrarelli et al. (2007), in particular we used a lower threshold for spindle detection and we applied an additional smoothing window. Spindle outcome variables were: duration, maximal amplitude, duration \times maximal amplitude, and density (the number of spindles per valid epoch of sleep) of slow (frequency <12 Hz) and fast (frequency \geq 12 Hz) spindles.

Slow waves

Automated slow wave detection was performed using an algorithm based on previously published methods (Massimini et al., 2004; Riedner et al., 2007) implemented in Matlab (The MathWorks Inc, Natick, USA). Artifact-free EEG classified as S2, S3 and S4 was high-pass filtered at 0.16 Hz (transition band width = 0.02 Hz) and low-pass filtered at 4 Hz (transition band width = 0.6 Hz), using a least-square zero phase-shift 200th order FIR filter. In the filtered signal, slow waves were defined by the appearance of a particular order of occurrences: a down-going zero crossing, a negative peak, an upgoing zero crossing,

a positive peak, and a final downgoing zero crossing. A slow wave was counted if the duration between the downgoing and upgoing zero crossing (the negative half wave) was between 0.3 and 1 s. No amplitude criteria were set. Slow wave outcome variables were the durations and peak amplitudes of the negative and positive half-wave and total wave (using downward and upward zero-crossings, see e.g., Heib et al., 2013); the steepness of the rising slope of the negative half-wave (see Piantoni et al., 2013b); and the density (the number of slow waves per epoch of NREM stage 2 and SWS sleep; see Piantoni et al., 2013b).

STATISTICAL ANALYSIS

The four paragraphs below describe the analysis plan, respectively addressing: the effect of sleep on performance; the association of sleep variables with performance baseline and overnight enhancement; the effect of acoustic perturbation on sleep outcome variables; and the effect of acoustic perturbation on performance outcome variables.

Effect of sleep on performance

In order to maximally exploit the variance information of speed and accuracy data of individual trials, they were not averaged, but rather analyzed using mixed models (MLwiN, Centre for Multilevel Modeling, Institute of Education, London, UK). Mixed models take an interdependence of data points into account; allowing trials to be nested within sessions, which are subsequently nested within participants. Maximal use of information was attained by including trials at the level of performance saturation (see Figure 2: the last six trials of the learning sessions and all six trials of the recall sessions).

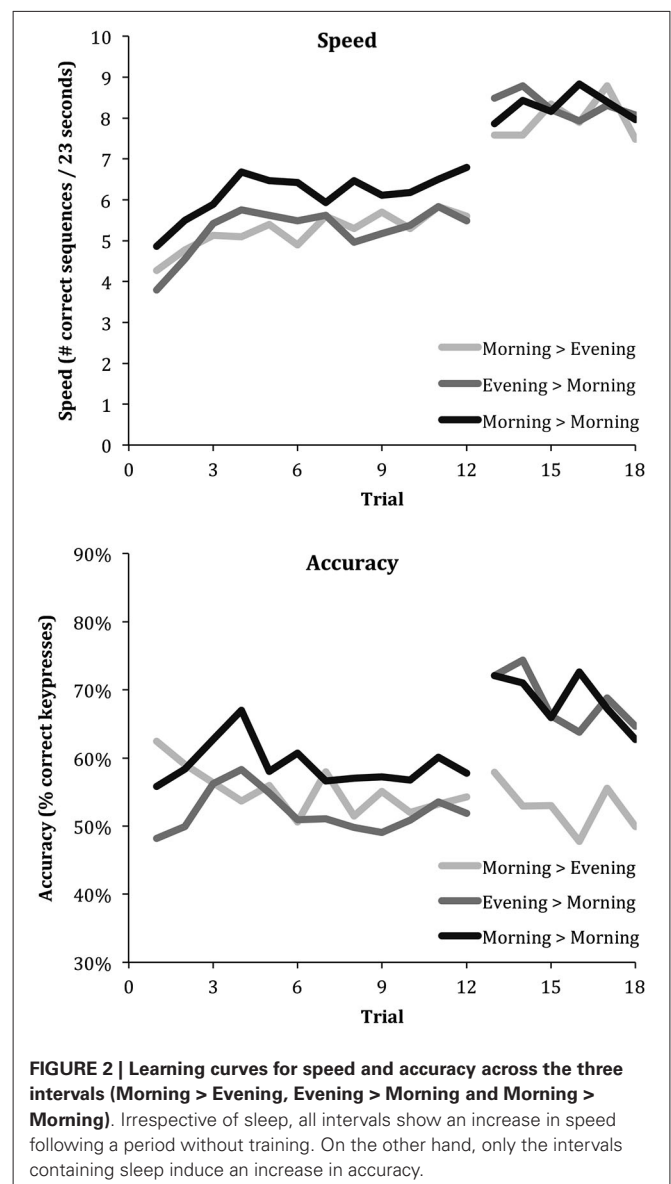
In order to evaluate the effect of sleep on initial (baseline) performance and performance enhancement, the dependent variables “speed” and “accuracy” assessed over all sessions were analyzed using the regression equation:

$$Y_{ijkl} = \beta_{0ijkl} + \beta_1 * \text{Recall}_{jkl} + \beta_2 * \text{Slept}_{jkl} + \beta_3 * \text{Recall} * \text{Slept}_{jkl}$$

where: Y is the dependent variable (either “speed” or “accuracy”), measured on trial i of the initial learning vs. delayed part j of session k of child l ; β_0 is the model intercept; “Recall” is a binary (dummy) variable that indicates whether the trial was a recall (1) or initial learning (0) trial; “Slept” is a binary (dummy) variable that indicates whether the present session was (1) or was not (0) preceded by a previous session followed by a period of sleep; “Recall*Slept” is a binary (dummy) variable that indicates the interaction between “Recall” and “Slept”. This interaction represents the sleep-dependent effect on recall. The variable is 1 for recall trials in sessions that are separated from the previous session by a period including sleep and 0 for all learning trials and recall trials in sessions that are separated from the previous session by a period of wakefulness only.

ASSOCIATION OF SLEEP VARIABLES WITH PERFORMANCE BASELINE AND OVERNIGHT ENHANCEMENT

In order to evaluate the effect of sleep variables assessed during the third night on baseline performance and performance



enhancement across that night, the dependent variables “speed” and “accuracy” were analyzed using the regression equation:

$$Y_{ijkl} = \beta_{0ijk} + \beta_1 * \text{Recall}_{jk} + \beta_2 * \text{Sleepvariable}_{jk} + \beta_3 * \text{Recall} * \text{Sleepvariable}_{jk}$$

where: Y is the dependent variable (either “speed” or “accuracy”), measured on trial i of the initial learning vs. delayed part j of child k ; β_0 is the model intercept; “Delayed” is a binary (dummy) variable that indicates whether the trial was a delayed (1) or initial learning (0) trial; “Sleepvariable” is the sleep variable of interest in the current analysis and indicates the nonspecific (i.e., sleep-unspecific) association of the sleep variable with performance; “Delayed*Sleepvariable” represents the interaction between “Delayed” and “Sleepvariable”. This interaction represents the sleep variable-dependent change in performance from the initial learning session to the delayed session.

Effect of acoustic perturbation on sleep outcome variables

Kruskal-Wallis tests (SPSS 12.0.1 for Windows, Chicago, USA) were applied to evaluate differences in macrosleep and quantitative EEG variables between acoustic perturbation conditions. The more robust Kruskal-Wallis tests were preferred over ANOVAs because variance estimates, although not precise due to the small and unequal sample sizes of the three groups, seemed to differ for some variables.

Effect of acoustic perturbation on performance outcome variables

In order to evaluate the effect of sleep perturbation, during the third night, on baseline performance and performance enhancement across that night, the dependent variables "speed" and "accuracy" were analyzed using the regression equation:

$$Y_{ijkl} = \beta_{0ijk} + \beta_1 * \text{Delayed}_{jk} + \beta_2 * \text{Slow Wave Triggered Sound}_{jk} + \beta_3 * \text{YokedSound}_{jk} + \beta_4 * \text{Delayed} * \text{Slow Wave Triggered Sound}_{jk} + \beta_5 * \text{Delayed} * \text{YokedSound}_{jk}$$

where: Y is the dependent variable (either "speed" or "accuracy"), measured on trial i of the initial learning vs. delayed part j of child k ; β_0 is the model intercept; "Delayed" is a binary (dummy) variable that indicates whether the trial was a delayed (1) or initial learning (0) trial; "SlowWaveTriggered-Sound" and "YokedSound" are two dummy binary (dummy) variables that code whether (1) or not (0) the child was assigned to the stimulation condition; both are zero for the control condition; "Delayed*SlowWaveTriggeredSound" and "Delayed*YokedSound" represent the interactions of "Delayed" with the conditions. These interactions represent the condition-dependent change in performance from the initial learning session to the delayed session.

For all mixed effect models, the significance of the regression coefficient estimates of interest was evaluated using the Wald test, that calculates a z -value as the ratio of the coefficient estimate over its standard error (Twisk, 2003). Effects with $P < 0.05$ were regarded significant.

RESULTS

In three children, one of the learning sessions was missed, twice because of equipment malfunctioning, once because the subject did not feel well temporarily. The corresponding delayed trials were omitted accordingly. In two participants one consistently noisy sleep-EEG channel (Cz) was omitted from analyses. Completely artifact-free data used for quantitative EEG analysis accounted for 65.1% ($\pm 1.3\%$; SEM) of the total PSG data acquired. The percentage of epochs containing even the slightest artifact slowly increased during the sleep period from 23% in the first hour of the night to 43% in the last hour of the night.

EFFECT OF ACOUSTIC PERTURBATION ON SLEEP AND PERFORMANCE OUTCOME VARIABLES

Counter to the impact in adults (Van Der Werf et al., 2009a), Kruskal-Wallis tests on acoustic perturbation confirmed no significant differences in either macrosleep outcome variables or NREM oscillations of the sleep recordings of children included in the closed loop slow wave suppression group ($n = 9$), the yoked

control group ($n = 10$) and the no-noise group ($n = 11$): TIB ($P = 0.759$), TST ($P = 0.847$), Sleep Onset Latency ($P = 0.758$), Latency to the First REM epoch ($P = 0.458$), Sleep Efficiency ($P = 0.742$) and the percentages of Wakefulness ($P = 0.192$) Stage 1 ($P = 0.599$), 2 ($P = 0.659$), SWS ($P = 0.493$) and REM sleep (0.373), spindle variables (FPz: $0.194 < \text{all } P < 0.706$, Cz: $0.257 < \text{all } P < 0.913$) or slow wave outcome variables (FPz: $0.135 < \text{all } P < 0.966$, Cz: $0.662 < \text{all } P < 0.981$). The analyses confirm that children slept through the acoustic perturbation without any measurable effect on their macrosleep or quantitative sleep variables. Mixed effect models confirmed that the overnight change in motor skill speed and accuracy were not affected by either the Slow Wave-Triggered or Yoked Sound ($0.505 < \text{all } P < 0.975$). Due to the lack of effect of acoustic stimulation, further results aggregate the data of all children, irrespective of condition.

EFFECT OF SLEEP ON PERFORMANCE

Figure 2 shows the trial-by-trial average speed and accuracy for the *Wake*, *Sleep* and *Wake & Sleep* conditions. Mixed effect models evaluated how speed and accuracy were affected at delay ("Delayed" effect), by sleep between the present and previous session ("Slept" effect), and by a sleep-dependent effect specific to delay ("Slept*" "Delayed" interaction), i.e., showing only in the previously trained sequences but not in the subsequent newly trained sequences. According to the output generated by mixed effect model estimation, all estimated effects are shown as average \pm standard error of the mean.

The analysis showed a very significant "Delayed" effect on speed, which increased on average from the six final training trials to the six delayed trials by 2.617 ± 0.421 correct sequences (48% of the initial performance that was 5.459, $Z = 6.216$, $P = 5E^{-10}$). Overall speed, i.e., aggregated over both delayed trials and initial learning trials, did not depend on whether children had slept in between the present and prior session ("Slept" effect: 0.339 ± 0.338 correct sequences, $Z = 0.947$, $P = 0.34$). There was no "Delayed*Slept" effect on speed, indicating that the performance increase occurred independently of whether children had slept in between the initial learning and delayed session; neither was there a sleep-dependent delay-specific effect on speed (-0.096 ± 0.570 , $Z = -0.168$, $P = 0.87$). Thus, children showed strong speed improvements both after a period of sleep and after a period of wakefulness, selectively for the previously learned sequences, without affecting performance on the subsequent newly trained sequences.

In contrast, there was a highly significant sleep-dependent effect on accuracy, which increased by $12.4 \pm 4.6\%$ (26% of the initial accuracy that was 47.6%, $Z = 2.696$, $P = 0.007$) specifically for the delayed trials, without any sleep-unspecific delayed effect ($-1.6 \pm 3.5\%$, $Z = -0.457$, $P = 0.65$) or non-delay-specific effect of sleep ($-3.0 \pm 4.6\%$, $Z = -0.652$, $P = 0.51$). Thus, children showed a strong reduction in error rates only after a period of sleep and only for the previously learned sequences, without affecting performance on the subsequent newly trained sequences, meaning that sleep did not affect performance on subsequent newly trained sequences.

Figure 3 shows an integrated view of the changes in speed and accuracy from initial learning to retesting of the same sequence for

each of the three intervals (*Sleep, Wake, Wake & Sleep*) as vectors. It illustrates how speed increases independent of whether or not the interval contained sleep (rightward change), while accuracy increases only if the interval contained sleep (upward change).

ASSOCIATION OF MACROSLEEP VARIABLES WITH PERFORMANCE BASELINE AND OVERNIGHT ENHANCEMENT

The overnight increase in accuracy was more pronounced in children with a higher percentage of SWS ($0.85 \pm 0.43\%$ per % more SWS, $Z = 1.977$, $P < 0.05$). Given that the range of SWS percentages found in the group of children was 23% to 46%, this finding suggests that the increase in accuracy may differ up to 20% ($0.85 \times 23\%$: for every % more SWS a child shows, it has a 0.85% higher accuracy, and there is a difference of 23% between the child with the lowest and highest percentage slow wave sleep).

SPINDLE CHARACTERISTICS AND THEIR ASSOCIATION WITH PERFORMANCE BASELINE AND OVERNIGHT ENHANCEMENT

Given the frequency distribution of spindles at FPz and Cz (Figure 5), the cut-off to discriminate fast and slow spindles was set at 12 Hz. Spindles were more prevalent and of a faster frequency at Cz. Table 1 summarizes the spindle characteristics. Mixed effect models evaluated the association of spindle characteristics with both the overall level and the overnight change in performance. Significant effects were found only for the density of slow and fast spindles.

With respect to overall performance, i.e., not specific for overnight enhancement and including all trials, children with a higher density of slow spindles at either Cz or FPz have lower overall speed (-5.45 ± 1.63 correct sequences/spindle per sleep epoch, $Z = -3.342$, $P < 0.001$) and accuracy ($-27.5 \pm 12.4\%$ /spindle per epoch, $Z = -2.218$, $P < 0.03$). In contrast, children with a higher density of fast spindles have a higher overall speed (4.46 ± 1.52

correct sequences/spindle per sleep epoch, $Z = 2.919$, $P < 0.004$) and, if anything, a non-significant higher accuracy ($15 \pm 11.5\%$ per spindle per epoch more, $Z = 1.304$, $P = 0.19$).

With respect to the overnight enhancement of performance, children with a higher density of slow spindles show a stronger overnight increase in accuracy ($16.1 \pm 6.8\%$ more increase/spindle per epoch, $Z = 2.368$, $P = 0.02$), but not speed ($P = 0.45$). In contrast, individual differences in fast spindle density did not show an association with overnight change in either speed ($P = 0.61$) or accuracy ($p = 0.39$).

Because slow spindles occurred more frequently at FPz and fast spindles more at Cz, we performed ancillary analyses to investigate whether the findings reflected differential effects of FPz vs. Cz spindles instead of slow vs. fast spindles. Neither the overall density of FPz spindles, nor the overall density of Cz spindles, were associated with either overall speed or accuracy or their overnight enhancement ($0.16 < P < 0.76$). To further explore the relevance of spindle frequency, we ran ancillary analyses on the predictive value of the mean frequencies at FPz and at Cz for motor skill speed and accuracy. Children show a higher overall speed if they have a higher mean frequency of their spindles measures either at FPz (3.99 ± 1.95 correct sequences/Hz, $Z = 2.046$, $P = 0.04$) or at Cz (3.62 ± 1.60 correct sequences/Hz, $Z = 2.268$, $P = 0.02$). The mean spindle frequencies were not associated with overall accuracy ($P = 0.48$ and $P = 0.10$ for FPz and Cz respectively), nor with overnight enhancement of speed or accuracy ($0.38 < P < 0.91$).

In summary, overall performance is best in children with a high density of fast spindles and a low density of slow spindles. Children with a high density of slow spindles profit most from sleep to attain a higher accuracy.

SLOW WAVE CHARACTERISTICS AND THEIR ASSOCIATION WITH PERFORMANCE BASELINE AND OVERNIGHT ENHANCEMENT

Table 1 summarizes the characteristics of slow waves detected in S2, S3 and S4. Because of the frequency-specific associations of spindles with overall performance and sleep-dependent enhancement, it was of particular interest to investigate whether a similar frequency-specific effect of slow waves was present, i.e., whether their duration (inverse of frequency) mattered for performance. Indeed, significant effects were found only for slow wave duration.

With respect to overall performance (i.e., including all trials, not specific for overnight enhancement), children with a longer average duration of their slow waves had a lower overall speed, no matter whether the slow wave duration was derived from FPz (-0.102 ± 0.041 less correct sequences/ms longer duration, $Z = -2.457$, $P = 0.014$) or Cz (-0.084 ± 0.043 less correct sequences/ms longer duration, $Z = -1.960$, $P < 0.050$). Likewise, children with a longer average duration of their slow waves had a lower overall accuracy, significantly so for slow wave duration derived at FPz (-0.77 ± 0.30 lower % accuracy per milliseconds longer duration, $Z = -2.567$, $P = 0.010$) and almost significant for slow wave duration derived at Cz (-0.57 ± 0.30 lower % accuracy per milliseconds longer duration, $Z = -1.900$, $P = 0.057$). Given that the range of individual differences in the average duration of slow waves (FPz: 763–828; Cz: 752–822) covers up to 70 ms, the findings suggest

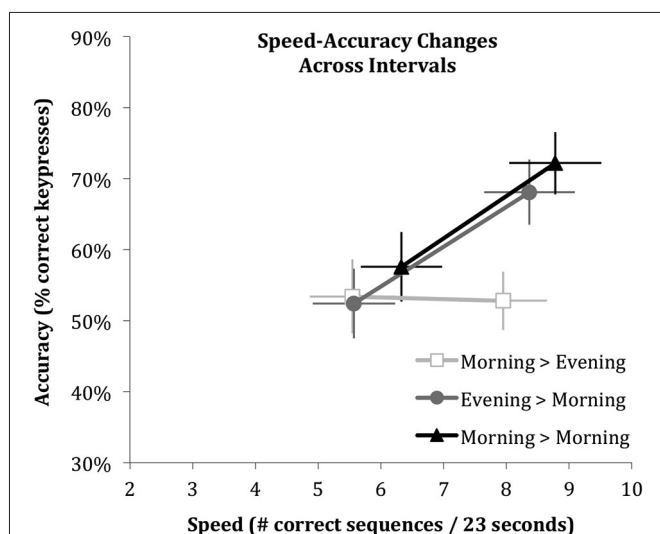


FIGURE 3 | Performance changes across the three intervals shown as vectors of speed and accuracy. Error bars indicate standard errors of the mean derived in mixed effect analyses. Note that speed increases across all intervals, whereas accuracy improves only across the intervals that include sleep.

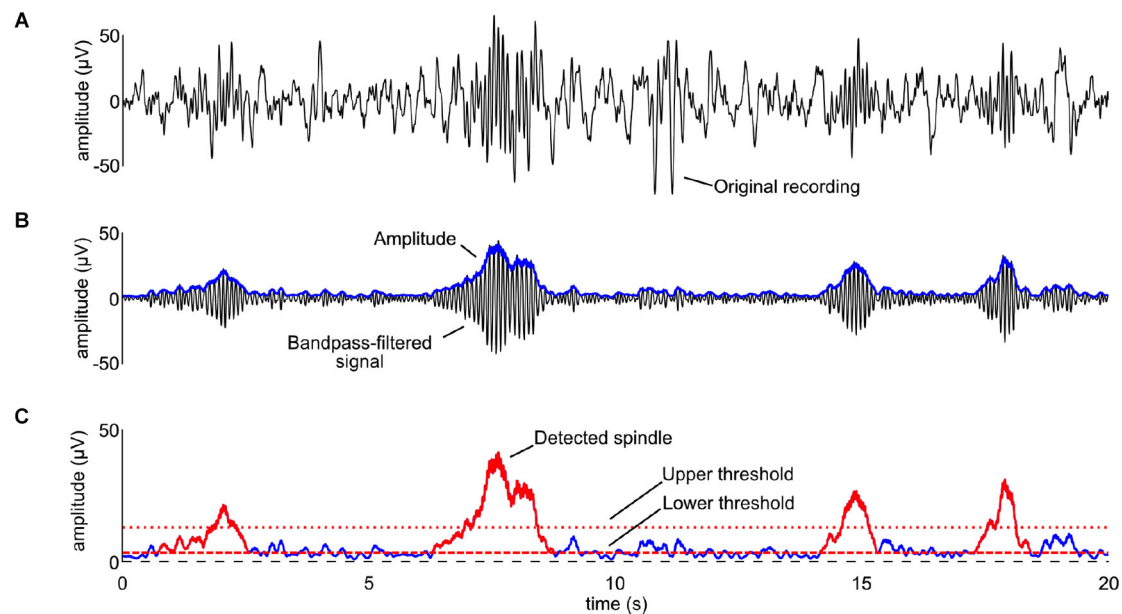


FIGURE 4 | The spindle detection procedure described in detail. (A) The original recording for one participant in stage S2. **(B)** The signal was bandpass-filtered between 9 and 15 Hz (black line) and the time-course of its amplitude was computed by rectifying the signal, applying a low-pass filter at 4 Hz (Nir et al., 2007), and multiplying by $\sqrt{2}$ (blue line). **(C)** An upper threshold equal to the 4.5 times the mean of the amplitude in

stages S2, S3, and S4 was used for the detection of the spindles (dotted red line). A lower threshold was used to define the beginning and end of each spindle (dotted dashed line). Detected spindles are shown as red traces superimposed on the time-course of the amplitude. Note that the x-axis is the same for all the panels, while the y-axis in the bottom panel is twice as large as that of panels (A) and (B).

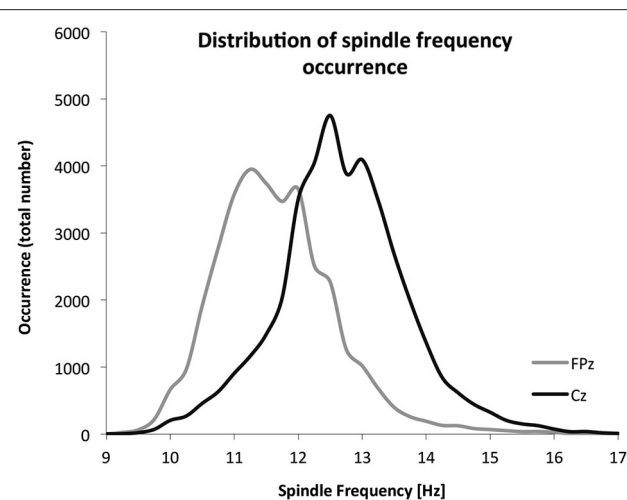


FIGURE 5 | The frequency distribution of all detected spindles. A total of 37,177 spindles were detected on FPz (30 children), and a total of 39,951 spindles were detected on Cz (28 children).

slow wave duration-associated individual differences in speed of up to about six correct sequences and in accuracy of up to about 50%.

With respect to the overnight enhancement of performance, children with a longer average duration of their slow waves showed a stronger overnight increase in accuracy, significantly so

for slow wave duration at Cz ($0.36 \pm 0.16\%$ stronger increase in accuracy per milliseconds longer duration, $Z = 2.25$, $P = 0.024$) and almost significant for slow wave duration derived at FPz ($0.32 \pm 0.17\%$ stronger increase in accuracy per milliseconds longer duration, $Z = 1.882$, $P = 0.060$). Slow wave duration was not associated with overnight changes in speed (FPz: $P = 0.66$; Cz: $P = 0.42$). Given the range of individual differences in the average duration of slow waves mentioned above, the findings suggest slow wave duration-associated individual differences in the overnight increase in accuracy of up to about 25%.

In summary, overall performance is best in children with a faster slow waves. Children with slower slow waves profit most from sleep to attain a higher accuracy.

ASSOCIATION BETWEEN INDIVIDUAL DIFFERENCES IN FAST AND SLOW SPINDLE DENSITY WITH AVERAGE SLOW WAVE DURATION

Given the findings overall performance is best in children with faster slow waves, a high density of fast spindles and a low density of slow spindles, *post hoc* correlations were calculated over the individual's pairs of these slow wave and parameters. The average duration of slow waves measured at FPz was negatively correlated with the density of fast spindles ($r = -0.40$, $p = 0.03$) and almost significantly positively correlated with the density of slow spindles ($r = 0.37$, $p = 0.05$). The average duration of slow waves measured at Cz showed no significant correlation with the density of either fast spindles ($r = -0.05$, $p = 0.80$) or slow spindles ($r = 0.10$, $p = 0.61$). In summary, there is a significant association between the dominant

Table 1 | Sleep variables averaged over all children.

	Mean \pm SEM	Mean \pm SEM
Macrostructure characteristics		
Time in Bed (TIB) (min)	464.7 \pm 1.76	
Total Sleep Time (TST) (min)	432.0 \pm 4.62	
Sleep Onset Latency (min)	19.3 \pm 2.34	
First REM Latency (min)	94.3 \pm 6.92	
Wake After Sleep Onset (min)	9.4 \pm 2.10	
Sleep Efficiency %	92.9 \pm 0.88	
% Stage 1 (of TST)	3.4 \pm 0.63	
% Stage 2 (of TST)	39.0 \pm 1.32	
% Slow Wave Sleep (of TST)	34.0 \pm 1.06	
% REM (of TST)	23.7 \pm 0.82	
Sleep spindle characteristics		
	FPz	Cz
Duration (ms)	1100.96 \pm 24.70	1206.72 \pm 17.99
Amplitude (μ V)	17.84 \pm 0.66	27.74 \pm 0.80
Duration*Amplitude (μ Vs)	20.37 \pm 1.00	34.52 \pm 1.29
Frequency (Hz)	11.55 \pm 0.06	12.49 \pm 0.08
Density (# / 30 s epoch)	2.24 \pm 0.07	2.48 \pm 0.06
Slow wave characteristics		
Duration negative		
half wave (ms)	457.44 \pm 1.32	451.52 \pm 1.35
Duration positive		
half wave (ms)	333.36 \pm 1.66	327.56 \pm 1.69
Total duration (ms)	790.80 \pm 2.76	780.14 \pm 2.98
Amplitude negative		
half wave (μ V)	-55.63 \pm 1.92	-76.87 \pm 2.42
Amplitude positive		
half wave (μ V)	58.62 \pm 2.36	76.00 \pm 2.59
Peak-to-peak amplitude (μ V)	126.96 \pm 4.53	162.33 \pm 5.48
Up-slope negative		
half wave (μ V/ms)	0.31 \pm 0.01	0.43 \pm 0.01
Density (# / 30 s epoch)	23.72 \pm 0.19	22.59 \pm 0.19

frontal frequency of two characteristic sleep microstructural events with relevance for motor skill performance: the average duration of a slow waves measured and the density of fast spindles.

DISCUSSION

The present study set out to investigate the following questions. We hypothesized that motor skill enhancement is dependent on sleep in school-aged children. We moreover hypothesized that initial motor skill performance, and its enhancement after an interval without training, depend on the parameters that quantify the sleep-EEG macrostructure and microstructural properties of spindles and slow waves. Finally, to complement associational findings, we aimed to evaluate whether the hypotheses would be supported by an intervention aimed at manipulation of spindles and slow waves.

Similar to findings in adults (Walker et al., 2002; Van Der Werf et al., 2009b), the current report demonstrated children express offline enhancements in motor skill accuracy only if this interval includes a period of sleep. However, unlike previously reported in adults, children enhance their speed no matter whether the interval includes a period of sleep. In contrast to previous reports with similar results (Fischer et al., 2007; Wilhelm et al., 2008; Prehn-Kristensen et al., 2009), we do not interpret these results

to indicate that children fail to show a speed enhancement over a period of sleep. Children do in fact show an enhancement of speed over a period of sleep, but as well over a period without sleep. Our interpretation is rather that children, like adults, do have the ability to enhance motor speed over a period of sleep, but the offline improvement can also be achieved across the different brain state of wakefulness (and thus perhaps by a different brain-state mechanisms). A speculative suggestion from our findings, that could be addressed in long-term follow-up studies on the development from childhood to adulthood, is that the capacity to improve performance without the necessity of sleep may be lost in adulthood. This suggestion is in line with recent findings indicating that procedural memory stabilizes during waking much faster in children than in adults (Ashtamker and Karni, 2013; Adi-Japha et al., 2014). Although children enhance their motor speed over periods of sleep and wake alike, sleep is required for an increase in accuracy (Figure 3).

An important new finding of the present study concerns the question of whether initial motor skill performance, or its enhancement after an interval without training, depend on specific aspects of the sleep-EEG microstructure. The results consistently indicate that individual differences in the dominant frequency of thalamo-cortical oscillations marks differences in both initial performance and sleep-dependent skill enhancement. Children with lower dominant frequencies of spindles and slow waves performed worse, as consistently indicated by the findings that children performed better if they had less slow spindles, more fast spindles and faster slow waves. The negative association between overall performance and the density of slow spindles is in line with a recent study by Kurdziel et al. (2013) who found, in 4-year children, that spindle density during a nap correlated negatively ($r = -0.67$) with baseline performance on a hippocampal-dependent visuospatial task resembling the card-deck "Memory" game. The hippocampus has also been implicated in sleep-dependent consolidation of motor sequence learning (Albouy et al., 2013b,c,d).

On average, characteristic oscillations in the EEG are slower in children than in adults and indeed also the peak frequency of sleep spindles increases as children mature (De Gennaro and Ferrara, 2003; Jenni and Carskadon, 2004; Tarokh and Carskadon, 2010). Our findings therefore suggest that dominant physiological frequencies of the characteristic sleep events may reflect trait-like markers of maturity within neuronal networks involved in cognition, including that associated with offline motor skill enhancement. It appears timely to consider large-scale multivariate follow-up studies to disentangle individual traits from developmental aspects, as well as common vs. differential involvement of spindle characteristics in motor skills, explicit memory and intellectual abilities (Geiger et al., 2011, 2012; Chatburn et al., 2013; Gruber et al., 2013; Hoedlmoser et al., 2014).

With respect to the overnight increase in performance, there appears to be a discrepancy at first sight between findings based on the density of slow and fast spindles vs. the findings based on the mean frequency of spindles. A stronger overnight increase in accuracy was associated with a higher density of slow spindles but not with a lower mean frequency of spindles. We interpret this

finding as support for distinct types of spindles, as suggested by a bimodal distribution (**Figure 5**). The mean frequency depends on the number of both slow and fast spindles, and can be low irrespective of overall density. Overnight accuracy enhancement appears specifically associated with the abundance of slow spindles. The finding that the density of slow spindles, rather than fast spindles as in adults, is associated with the overnight increase in accuracy is interesting, since in children and adolescents, there is a slower frequency peak in the spindle-related sigma power (Jenni et al., 2005; Kurth et al., 2010). Thus, it may be that this leftward shift in the dominant spindle frequency curve, relative to adults, is involved in this differential association, and could still reflect similar overlapping consolidation mechanisms. Indeed, sleep spindle frequency in human adults has been associated with structural gray matter properties of the hippocampus. Moreover, surface EEG recorded spindles in human adults are associated with coinciding hippocampal activation. Should similar spindle-hippocampal associations be identified in child, this may provide one potential neural pathway through which spindle-related motor skill improvements are transacted in child, especially since the hippocampus is importantly involved in explicit motor skill learning (Walker et al., 2005; Steele and Penhune, 2010; Saletin et al., 2013).

Heib et al. (2013) showed a positive correlation between individual differences in the duration of the positive half-wave of the slow oscillation and their overnight changes in memory for word pairs. They speculated that a prolonged depolarizing up-state extends the time window for neuronal replay and thus enhances overnight memory improvement. No increase in the duration of slow oscillations in response to learning was found in this study, nor in a previous similar study (Möller et al., 2009). These studies did not investigate whether individuals with longer positive half-waves might have had lower initial, pre-sleep, performance, and thus more room for overnight improvement similar to the current findings in children. Our present findings suggest that it may be important to investigate whether associations of sleep parameters with overnight improvements are secondary to associations of the same sleep parameters with initial performance. In the present study, the use of mixed effect multiple regression models allowed for a separation of these different associations.

Interestingly, the enhancement of accuracy over a period of sleep and of speed over a period of either sleep or wakefulness, is of a greater magnitude than has previously been reported in adults. The overnight improvement of speed, irrespective of sleep, was about 45%, which is more than twice the sleep-dependent speed improvement reported in the original study in adults (Walker et al., 2002). The overnight improvement in accuracy was 49%. Whereas no sleep-dependent change in accuracy reported in the original study in adults (Walker et al., 2002), later studies found accuracy improvements of up to 48% (Kuriyama et al., 2004). A parsimonious explanation of the findings is that participants that show an initial low performance, as is the case in the present study in children, have more headroom for improvement. This interpretation is supported by the fact that the strongest sleep-dependent increase in accuracy occurred in those that initially performed worst, i.e., those with lower dominant frequencies of spindles and slow waves. A recent study in

4-year old children also observed an inverse association between initial performance and sleep-dependent improvement (Kurdziel et al., 2013). As was the case for slow spindles (typical of young children) in our present study, they observed that sleep spindle density was negatively correlated with baseline performance and positively correlated with the change in memory performance across the nap period. In that study, children with a higher sleep spindle density initially performed worse and benefitted more from sleep for subsequent performance. Importantly, if associations of spindle and slow wave characteristics with initial performance are not accounted for, they may confound interpretation of their involvement in overnight enhancement.

The current study result need to be appreciated within the context of several inherent limitations. First, the sleep of children was so resistant to acoustic manipulation that we did not succeed in our aim to take the level of evidence for a role of spindles and slow waves in overnight a step further, from observational data to experimental intervention. The present findings confirm previous findings (Busby et al., 1994) suggesting that children have a much more powerful thalamic gate to shut off environmental monitoring during sleep.

A second limitation is that during the night of polysomnographic recording the children performed the task later in the evening than their habitual bedtime and slept relatively short. With respect to the late assessment, **Figure 3** shows no systematically worse performance. The speed during both learning and recall in the evening did not differ from the speed during learning and recall in the morning, and the accuracy during learning in the evening did not differ from the accuracy during learning in the morning. These considerations support the interpretation that the lack of accuracy improvement in the morning-to-evening condition is specifically due to a lack of sleep. With respect to sleep duration, a recent systematic review on normal sleep patterns in children concluded that 11-year olds on average sleep 9 a night (Galland et al., 2012). Sleep duration was somewhat restricted in the present protocol due to the task assessment protocol with strict 12 h and 24 h intervals, so that the evening task assessment started at 10:00 PM. This resulted in a late bedtime as compared to their habitual bedtime (8:46 PM \pm 00:21 min). Sleep duration may moreover have been somewhat restricted due to the excitement of the children about participating in a study that included sleeping a night in a Science Museum. The distribution of sleep stage durations in the present study was however very similar to those reported in previous studies on sleep in children (Fischer et al., 2007; Backhaus et al., 2008; Wilhelm et al., 2008). Ideally, a replication study would assess whether the reported associations hold if children are recorded at home according to their habitual sleep schedule.

A third limitation is that sleep was recorded in a non-shielded environment, which may have induced a larger number of epochs containing artifacts than would be expected in the environment of a well-controlled sleep-laboratory. A further limitation is that no extensive clinical evaluation on sleep disturbances was performed.

Finally, it should be noted that performing a motor skill task prior to bedtime may in itself alter the distribution of sleep spindles. Studies in humans and animals have consistently

shown spindle activity to increase following training on several tasks, including the motor sequence tapping task used in the present study (Nishida and Walker, 2007; Barakat et al., 2011). Barakat et al. (2011) studied how sleep was affected by pre-sleep training on the same finger-tapping task that was used in the present study. They found that, compared to training on a control task, the motor sequence tapping task increased the density of fast spindles, while the density of slow spindles did not change. Subjects with the strongest training-elicited increase in fast spindle density showed the strongest sleep-dependent speed enhancement. Slow spindle density was not related to the sleep-dependent enhancement. Accuracy was not investigated. The association may be specific to the type of motor skill, because data presented by Tamaki et al. (2008; **Table 1**) suggest a decrease rather than increase in the number of fast spindles after training a mirror tracing skill. Moreover, although we cannot exclude the possibility that the motor skill task performance prior to bedtime increased spindle activity, it should be noticed that the functional relevance of such increase may be limited to the cortical area that are most prominently activated by the task, an area below the C4 electrode (Nishida and Walker, 2007).

In summary, the present findings indicate that even without sleep, children have the ability to increase the speed of their motor skills without training, a capacity that seems to be lost in adulthood. Moreover, whereas the majority of previous studies focused on sleep-dependent consolidation and enhancement, the present findings underscore the importance of investigating the associations of slower vs. faster oscillating spindles and slow waves with initial performance (Bódizs et al., 2005; Schabus et al., 2008), and the necessity to investigate how overnight improvements may be limited by high initial performance and enhanced by low initial performance. Overall, the present findings suggest that slower frequency oscillations of the characteristic sleep events may mark a less mature neuronal networks involved in motor skills and slower learning curves. This finding can be seen as a warning for a likely confound: if associations of spindle and slow wave characteristics with initial performance are not accounted for, they may confound interpretation of their selective involvement in overnight enhancement.

ACKNOWLEDGMENTS

Data were obtained during “The Great Sleep Experiment” event, organized and sponsored by Netherlands Organization of Scientific Research and the Netherlands Institute for Neuroscience, and sponsored by Beter Bed, Medcare, IBM, Amstel Botel, Cambridge Neurotechnology, J&J Pharmaceutical Research and Development, Nederlandse Vereniging voor Slaap-Waak Onderzoek, Onderzoeksschool Neurowetenschappen Amsterdam, Philips, SEIN Zwolle, Gelre ziekenhuizen Zutphen, OLVG Amsterdam, Sint Lucas Andreas Ziekenhuis Amsterdam, Spaarne Ziekenhuis Hoofddorp and Zaans Medisch Centrum Zaandam. This work was further supported by grants from the Netherlands Organization of Scientific Research (NWO): VICI Innovation (Grant 453-07-001), and National Initiative Brain and Cognition Research Program “Innovative learning materials and methods” (Grant 056-32-013).

We would like to thank all the participating children and their teachers, all 22 volunteering lab technicians and the research staff who made this event possible. Finally, we would like to thank Prof. R. Huber for his kind permission to use and adapt his automated spindle detection scripts, and Prof. M. Massimini for the slow wave script.

REFERENCES

- Adi-Japha, E., Badir, R., Dorfberger, S., and Karni, A. (2014). A matter of time: rapid motor memory stabilization in childhood. *Dev. Sci.* 17, 424–433. doi: 10.1111/desc.12132
- Albouy, G., Fogel, S., Pottiez, H., Nguyen, V. A., Ray, L., Lungu, O., et al. (2013a). Daytime sleep enhances consolidation of the spatial but not motoric representation of motor sequence memory. *PLoS One* 8:e52805. doi: 10.1371/journal.pone.0052805
- Albouy, G., King, B. R., Maquet, P., and Doyon, J. (2013b). Hippocampus and striatum: dynamics and interaction during acquisition and sleep-related motor sequence memory consolidation. *Hippocampus* 23, 985–1004. doi: 10.1002/hipo.22183
- Albouy, G., Sterpenich, V., Vandewalle, G., Darsaud, A., Gais, S., Rauchs, G., et al. (2013c). Interaction between hippocampal and striatal systems predicts subsequent consolidation of motor sequence memory. *PLoS One* 8:e59490. doi: 10.1371/journal.pone.0059490
- Albouy, G., Vandewalle, G., Sterpenich, V., Rauchs, G., Desseilles, M., Balteau, E., et al. (2013d). Sleep stabilizes visuomotor adaptation memory: a functional magnetic resonance imaging study. *J. Sleep Res.* 22, 144–154. doi: 10.1111/j.1365-2869.2012.01059.x
- Ashtamker, L., and Karni, A. (2013). Motor memory in childhood: early expression of consolidation phase gains. *Neurobiol. Learn. Mem.* 106, 26–30. doi: 10.1016/j.nlm.2013.07.003
- Backhaus, J., Hoeckesfeld, R., Born, J., Hohagen, F., and Junghanns, K. (2008). Immediate as well as delayed post learning sleep but not wakefulness enhances declarative memory consolidation in children. *Neurobiol. Learn. Mem.* 89, 76–80. doi: 10.1016/j.nlm.2007.08.010
- Barakat, M., Carrier, J., Debas, K., Lungu, O., Fogel, S., Vandewalle, G., et al. (2013). Sleep spindles predict neural and behavioral changes in motor sequence consolidation. *Hum. Brain Mapp.* 34, 2918–2928. doi: 10.1002/hbm.22116
- Barakat, M., Doyon, J., Debas, K., Vandewalle, G., Morin, A., Poirier, G., et al. (2011). Fast and slow spindle involvement in the consolidation of a new motor sequence. *Behav. Brain Res.* 217, 117–121. doi: 10.1016/j.bbr.2010.10.019
- Bódizs, R., Kis, T., Lázár, A. S., Havrán, L., Rigó, P., Clemens, Z., et al. (2005). Prediction of general mental ability based on neural oscillation measures of sleep. *J. Sleep Res.* 14, 285–292. doi: 10.1111/j.1365-2869.2005.00472.x
- Bruni, O., Ottaviano, S., Guidetti, V., Romoli, M., Innocenzi, M., Cortesi, F., et al. (1996). The sleep disturbance scale for children (SDSC). Construction and validation of an instrument to evaluate sleep disturbances in childhood and adolescence. *J. Sleep Res.* 5, 251–261. doi: 10.1111/j.1365-2869.1996.00251.x
- Busby, K. A., Mercier, L., and Pivik, R. T. (1994). Ontogenetic variations in auditory arousal threshold during sleep. *Psychophysiology* 31, 182–188. doi: 10.1111/j.1469-8986.1994.tb01038.x
- Chatburn, A., Coussens, S., Lushington, K., Kennedy, D., Baumert, M., and Kohler, M. (2013). Sleep spindle activity and cognitive performance in healthy children. *Sleep* 36, 237–243. doi: 10.5665/sleep.2380
- De Gennaro, L., and Ferrara, M. (2003). Sleep spindles: an overview. *Sleep Med. Rev.* 7, 423–440. doi: 10.1053/smr.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
- Diekmann, S., Wilhelm, I., and Born, J. (2009). The whats and whens of sleep-dependent memory consolidation. *Sleep Med. Rev.* 13, 309–321. doi: 10.1016/j.smr.2008.08.002
- Dorfberger, S., Adi-Japha, E., and Karni, A. (2007). Reduced susceptibility to interference in the consolidation of motor memory before adolescence. *PLoS One* 2:e240. doi: 10.1371/journal.pone.0000240
- Feld, G. B., and Born, J. (2012). “Sleep EEG rhythms and system consolidation of memory,” in *Sleep and Brain Activity*, ed M. G. Frank (San Diego: Academic Press), 187–226.

- Ferrarelli, F., Huber, R., Peterson, M. J., Massimini, M., Murphy, M., Riedner, B. A., et al. (2007). Reduced sleep spindle activity in schizophrenia patients. *Am. J. Psychiatry* 164, 483–492. doi: 10.1176/appi.ajp.164.3.483
- Fischer, S., Wilhelm, I., and Born, J. (2007). Developmental differences in sleep's role for implicit off-line learning: comparing children with adults. *J. Cogn. Neurosci.* 19, 214–227. doi: 10.1162/jocn.2007.19.2.214
- Fogel, S. M., and Smith, C. T. (2011). The function of the sleep spindle: a physiological index of intelligence and a mechanism for sleep-dependent memory consolidation. *Neurosci. Biobehav. Rev.* 35, 1154–1165. doi: 10.1016/j.neubiorev.2010.12.003
- Galland, B. C., Taylor, B. J., Elder, D. E., and Herbison, P. (2012). Normal sleep patterns in infants and children: a systematic review of observational studies. *Sleep Med. Rev.* 16, 213–222. doi: 10.1016/j.smrv.2011.06.001
- Geiger, A., Huber, R., Kurth, S., Ringli, M., Achermann, P., and Jenni, O. G. (2012). Sleep electroencephalography topography and children's intellectual ability. *Neuroreport* 23, 93–97. doi: 10.1097/wnr.0b013e32834e7e8f
- Geiger, A., Huber, R., Kurth, S., Ringli, M., Jenni, O. G., and Achermann, P. (2011). The sleep EEG as a marker of intellectual ability in school age children. *Sleep* 34, 181–189.
- Gruber, R., Wise, M. S., Frenette, S., Knäuper, B., Boom, A., Fontil, L., et al. (2013). The association between sleep spindles and IQ in healthy school-age children. *Int. J. Psychophysiol.* 89, 229–240. doi: 10.1016/j.jipsycho.2013.03.018
- Heib, D. P., Hoedlmoser, K., Anderer, P., Zeithofer, J., Gruber, G., Klimesch, W., et al. (2013). Slow oscillation amplitudes and up-state lengths relate to memory improvement. *PLoS One* 8:e82049. doi: 10.1371/journal.pone.0082049
- Hoedlmoser, K., Heib, D. P., Roell, J., Peigneux, P., Sadeh, A., Gruber, G., et al. (2014). Slow sleep spindle activity, declarative memory and general cognitive abilities in children. *Sleep* 37, 1501–1512. doi: 10.5665/sleep.4000
- Holz, J., Piosczyk, H., Feige, B., Spiegelhalter, K., Baglioni, C., Riemann, D., et al. (2012). EEG sigma and slow-wave activity during NREM sleep correlate with overnight declarative and procedural memory consolidation. *J. Sleep Res.* 21, 612–619. doi: 10.1111/j.1365-2869.2012.01017.x
- Huber, R., Ghilardi, M. F., Massimini, M., and Tononi, G. (2004). Local sleep and learning. *Nature* 430, 78–81. doi: 10.1038/nature02663
- Jenni, O. G., Achermann, P., and Carskadon, M. A. (2005). Homeostatic sleep regulation in adolescents. *Sleep* 28, 1446–1454.
- Jenni, O. G., and Carskadon, M. A. (2004). Spectral analysis of the sleep electroencephalogram during adolescence. *Sleep* 27, 774–783.
- Karni, A., Meyer, G., Jezard, P., Adams, M. M., Turner, R., and Ungerleider, L. G. (1995). Functional MRI evidence for adult motor cortex plasticity during motor skill learning. *Nature* 377, 155–158. doi: 10.1038/377155a0
- Kurziel, L., Duclos, K., and Spencer, R. M. (2013). Sleep spindles in midday naps enhance learning in preschool children. *Proc. Natl. Acad. Sci. U S A* 110, 17267–17272. doi: 10.1073/pnas.1306418110
- Kuriyama, K., Stickgold, R., and Walker, M. P. (2004). Sleep-dependent learning and motor-skill complexity. *Learn. Mem.* 11, 705–713. doi: 10.1101/lm.76304
- Kurth, S., Ringli, M., Geiger, A., LeBourgeois, M., Jenni, O. G., and Huber, R. (2010). Mapping of cortical activity in the first two decades of life: a high-density sleep electroencephalogram study. *J. Neurosci.* 30, 13211–13219. doi: 10.1523/jneurosci.2532-10.2010
- Kurth, S., Ringli, M., Lebourgeois, M. K., Geiger, A., Buchmann, A., Jenni, O. G., et al. (2012). Mapping the electrophysiological marker of sleep depth reveals skill maturation in children and adolescents. *Neuroimage* 63, 959–965. doi: 10.1016/j.neuroimage.2012.03.053
- Landmann, N., Kuhn, M., Piosczyk, H., Feige, B., Baglioni, C., Spiegelhalter, K., et al. (2014). The reorganisation of memory during sleep. *Sleep Med. Rev.* doi: 10.1016/j.smrv.2014.03.005. [Epub ahead of print].
- Määttä, S., Landsness, E., Sarasso, S., Ferrarelli, F., Ferreri, F., Ghilardi, M. F., et al. (2010). The effects of morning training on night sleep: a behavioral and EEG study. *Brain Res. Bull.* 82, 118–123. doi: 10.1016/j.brainresbull.2010.01.006
- Mander, B. A., Santhanam, S., Saletin, J. M., and Walker, M. P. (2011). Wake deterioration and sleep restoration of human learning. *Curr. Biol.* 21, R183–R184. doi: 10.1016/j.cub.2011.01.019
- Maquet, P. (2001). The role of sleep in learning and memory. *Science* 294, 1048–1052. doi: 10.1126/science.1062856
- Massimini, M., Huber, R., Ferrarelli, F., Hill, S., and Tononi, G. (2004). The sleep slow oscillation as a traveling wave. *J. Neurosci.* 24, 6862–6870. doi: 10.1523/jneurosci.1318-04.2004
- Möller, M., and Born, J. (2011). Slow oscillations orchestrating fast oscillations and memory consolidation. *Prog. Brain Res.* 193, 93–110. doi: 10.1016/b978-0-444-53839-0.00007-7
- Möller, M., Eschenko, O., Gais, S., Sara, S. J., and Born, J. (2009). The influence of learning on sleep slow oscillations and associated spindles and ripples in humans and rats. *Eur. J. Neurosci.* 29, 1071–1081. doi: 10.1111/j.1460-9568.2009.06654.x
- Morin, A., Doyon, J., Dostie, V., Barakat, M., Hadj Tahar, A., Korman, M., et al. (2008). Motor sequence learning increases sleep spindles and fast frequencies in post-training sleep. *Sleep* 31, 1149–1156.
- Nir, Y., Fisch, L., Mukamel, R., Gelbard-Sagiv, H., Arieli, A., Fried, I., et al. (2007). Coupling between neuronal firing rate, gamma LFP and BOLD fMRI is related to interneuronal correlations. *Curr. Biol.* 17, 1275–1285. doi: 10.1016/j.cub.2007.06.066
- Nishida, M., and Walker, M. P. (2007). Daytime naps, motor memory consolidation and regionally specific sleep spindles. *PLoS One* 2:e341. doi: 10.1371/journal.pone.0000341
- Owens, J. A., Maxim, R., Nobile, C., McGuinn, M., and Msall, M. (2000a). Parental and self-report of sleep in children with attention-deficit/hyperactivity disorder. *Arch. Pediatr. Adolesc. Med.* 154, 549–555. doi: 10.1001/archpedi.154.6.549
- Owens, J. A., Spirito, A., and McGuinn, M. (2000b). The children's sleep habits questionnaire (CSHQ): psychometric properties of a survey instrument for school-aged children. *Sleep* 23, 1043–1051.
- Piantoni, G., Astill, R. G., Raymann, R. J. E. M., Vis, J. C., Coppens, J. E., and Van Someren, E. J. W. (2013a). Modulation of gamma and spindle-range power by slow oscillations in scalp sleep EEG of children. *Int. J. Psychophysiol.* 89, 252–258. doi: 10.1016/j.jipsycho.2013.01.017
- Piantoni, G., Poil, S. S., Linkenkaer-Hansen, K., Verweij, I. M., Ramautar, J. R., Van Someren, E. J., et al. (2013b). Individual differences in white matter diffusion affect sleep oscillations. *J. Neurosci.* 33, 227–233. doi: 10.1523/jneurosci.2030-12.2013
- Prehn-Kristensen, A., Göder, R., Chirobeja, S., Bressmann, I., Ferstl, R., and Baving, L. (2009). Sleep in children enhances preferentially emotional declarative but not procedural memories. *J. Exp. Child Psychol.* 104, 132–139. doi: 10.1016/j.jecp.2009.01.005
- Rasch, B., and Born, J. (2013). About sleep's role in memory. *Physiol. Rev.* 93, 681–766. doi: 10.1152/physrev.00032.2012
- Rasch, B., Pommer, J., Diekmann, S., and Born, J. (2009). Pharmacological REM sleep suppression paradoxically improves rather than impairs skill memory. *Nat. Neurosci.* 12, 396–397. doi: 10.1038/nn.2206
- Rechtschaffen, A., and Kales, A. (1968). *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*. Bethesda: United States Department of Health, Education and Welfare.
- Riedner, B. A., Vyazovskiy, V. V., Huber, R., Massimini, M., Esser, S., Murphy, M., et al. (2007). Sleep homeostasis and cortical synchronization: III. A high-density EEG study of sleep slow waves in humans. *Sleep* 30, 1643–1657.
- Saletin, J. M., van der Helm, E., and Walker, M. P. (2013). Structural brain correlates of human sleep oscillations. *Neuroimage* 83, 658–668. doi: 10.1016/j.neuroimage.2013.06.021
- 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
- Schabus, M., Hoedlmoser, K., Pecherstorfer, T., Anderer, P., Gruber, G., Parapatics, S., et al. (2008). Interindividual sleep spindle differences and their relation to learning-related enhancements. *Brain Res.* 1191, 127–135. doi: 10.1016/j.brainres.2007.10.106
- Shinomiya, S., Nagata, K., Takahashi, K., and Masumura, T. (1999). Development of sleep spindles in young children and adolescents. *Clin. Electroencephalogr.* 30, 39–43. doi: 10.1177/155005949903000203
- Steele, C. J., and Penhune, V. B. (2010). Specific increases within global decreases: a functional magnetic resonance imaging investigation of five days of motor sequence learning. *J. Neurosci.* 30, 8332–8341. doi: 10.1523/jneurosci.5569-09.2010

- Steriade, M. (2006). Grouping of brain rhythms in corticthalamic systems. *Neuroscience* 137, 1087–1106. doi: 10.1016/j.neuroscience.2005.10.029
- Stickgold, R., and Walker, M. P. (2005). Memory consolidation and reconsolidation: what is the role of sleep? *Trends Neurosci.* 28, 408–415. doi: 10.1016/j.tins.2005.06.004
- Tamaki, M., Matsuoka, T., Nittono, H., and Hori, T. (2008). Fast sleep spindle (13–15 Hz) activity correlates with sleep-dependent improvement in visuomotor performance. *Sleep* 31, 204–211.
- Tamminen, J., Lambon Ralph, M. A., and Lewis, P. A. (2013). The role of sleep spindles and slow-wave activity in integrating new information in semantic memory. *J. Neurosci.* 33, 15376–15381. doi: 10.1523/jneurosci.5093-12.2013
- Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., and Gaskell, M. G. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *J. Neurosci.* 30, 14356–14360. doi: 10.1523/jneurosci.3028-10.2010
- Tanguay, P. E., Ornitz, E. M., Kaplan, A., and Bozzo, E. S. (1975). Evolution of sleep spindles in childhood. *Electroencephalogr. Clin. Neurophysiol.* 38, 175–181. doi: 10.1016/0013-4694(75)90227-8
- Tarokh, L., and Carskadon, M. A. (2010). Developmental changes in the human sleep EEG during early adolescence. *Sleep* 33, 801–809.
- Twisk, J. W. R. (2003). *Applied Longitudinal Data Analysis for Epidemiology*. Cambridge: Cambridge University Press.
- Van Der Werf, Y. D., Altena, E., Schoonheim, M. M., Sanz-Arigita, E., Vis, J. C., De Rijke, W., et al. (2009a). Sleep benefits subsequent hippocampal functioning. *Nat. Neurosci.* 12, 122–123. doi: 10.1038/nn.2253
- Van Der Werf, Y. D., Van Der Helm, E., Schoonheim, M. M., Ridderikhoff, A., and Van Someren, E. J. W. (2009b). Learning by observation requires an early sleep window. *Proc. Natl. Acad. Sci. U S A* 106, 18926–18930. doi: 10.1073/pnas.0901320106
- Walker, M. P., Brakefield, T., Morgan, A., Hobson, J. A., and Stickgold, R. (2002). Practice with sleep makes perfect: sleep-dependent motor skill learning. *Neuron* 35, 205–211. doi: 10.1016/s0896-6273(02)00746-8
- Walker, M. P., and Stickgold, R. (2004). Sleep-dependent learning and memory consolidation. *Neuron* 44, 121–133. doi: 10.1016/j.neuron.2004.08.031
- Walker, M. P., Stickgold, R., Alsop, D., Gaab, N., and Schlaug, G. (2005). Sleep-dependent motor memory plasticity in the human brain. *Neuroscience* 133, 911–917. doi: 10.1016/j.neuroscience.2005.04.007
- Wilhelm, I., Diekelmann, S., and Born, J. (2008). Sleep in children improves memory performance on declarative but not procedural tasks. *Learn. Mem.* 15, 373–377. doi: 10.1101/lm.803708

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: 27 July 2014; accepted: 23 October 2014; published online: 11 November 2014.

Citation: Astill RG, Piantoni G, Raymann RJEM, Vis JC, Coppens JE, Walker MP, Stickgold R, Van Der Werf YD and Van Someren EJW (2014) Sleep spindle and slow wave frequency reflect motor skill performance in primary school-age children. *Front. Hum. Neurosci.* 8:910. doi: 10.3389/fnhum.2014.00910

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

Copyright © 2014 Astill, Piantoni, Raymann, Vis, Coppens, Walker, Stickgold, Van Der Werf and Van Someren. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and 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.



Sleep spindling and fluid intelligence across adolescent development: sex matters

Róbert Bódizs^{1,2*}, Ferenc Gombos², Péter P. Ujma¹ and Ilona Kovács²

¹ Institute of Behavioural Sciences, Semmelweis University, Budapest, Hungary

² Department of General Psychology, Pázmány Péter Catholic University, Budapest, Hungary

Edited by:

Christian O'Reilly, McGill University, Canada

Reviewed by:

Roger Godbout, Université de Montréal, Canada

Rebecca Nader, Trent University, Canada

*Correspondence:

Róbert Bódizs, Institute of Behavioural Sciences, Semmelweis University, Nagyvárad tér 4, Budapest H-1089, Hungary
e-mail: bodizs.robert@med.semmelweis-univ.hu

Evidence supports the intricate relationship between sleep electroencephalogram (EEG) spindling and cognitive abilities in children and adults. Although sleep EEG changes during adolescence index fundamental brain reorganization, a detailed analysis of sleep spindling and the spindle-intelligence relationship was not yet provided for adolescents. Therefore, adolescent development of sleep spindle oscillations were studied in a home polysomnographic study focusing on the effects of chronological age and developmentally acquired overall mental efficiency (fluid IQ) with sex as a potential modulating factor. Subjects were 24 healthy adolescents (12 males) with an age range of 15–22 years (mean: 18 years) and fluid IQ of 91–126 (mean: 104.12, Raven Progressive Matrices Test). Slow spindles (SSs) and fast spindles (FSs) were analyzed in 21 EEG derivations by using the individual adjustment method (IAM). A significant age-dependent increase in average FS density ($r = 0.57$; $p = 0.005$) was found. Moreover, fluid IQ correlated with FS density ($r = 0.43$; $p = 0.04$) and amplitude ($r = 0.41$; $p = 0.049$). The latter effects were entirely driven by particularly reliable FS-IQ correlations in females [$r = 0.80$ ($p = 0.002$) and $r = 0.67$ ($p = 0.012$), for density and amplitude, respectively]. Region-specific analyses revealed that these correlations peak in the fronto-central regions. The control of the age-dependence of FS measures and IQ scores did not considerably reduce the spindle-IQ correlations with respect to FS density. The only positive spindle-index of fluid IQ in males turned out to be the frequency of FSs ($r = 0.60$, $p = 0.04$). Increases in FS density during adolescence may index reshaped structural connectivity related to white matter maturation in the late developing human brain. The continued development over this age range of cognitive functions is indexed by specific measures of sleep spindling unraveling gender differences in adolescent brain maturation and perhaps cognitive strategy.

Keywords: sleep spindling, EEG, adolescence, gender, IQ, Raven Progressive Matrices Test, sigma waves

INTRODUCTION

Adolescence is a critical period in the maturation of the neural architecture and in the related development of cognitive functions. This period is characterized by the late maturation of association areas involved in top-down control of thoughts and action (Casey et al., 2005). New findings in developmental psychology and neuroscience reveal that a fundamental reorganization of the brain takes place in adolescence (Konrad et al., 2013). The major reorganization of cortical networks during adolescence is indexed by the changing patterns of synchronous, oscillatory activity (Uhlhaas et al., 2009). Moreover, evidence suggests profound changes in the organization and function of cortical networks during transition from adolescence to adulthood (Uhlhaas and Singer, 2011). These changes may have substantial implications for the understanding of cognitive functions and

cognitive development (Uhlhaas et al., 2009; Uhlhaas and Singer, 2011).

Intellectual ability is closely related to cortical development in children and adolescents. The level of intelligence is associated with the trajectory of cortical development, primarily in frontal regions implicated in the maturation of intelligent activity: vigorous cortical thinning by early adolescence is a positive index of IQ (Shaw et al., 2006; Gogtay and Thompson, 2010). Furthermore, results emphasize the possibility that an individual's intellectual capacity relative to their peers can decrease or increase in the teenage years. Decreases and increases were found to depend on structural and functional changes of specific brain regions (Ramsden et al., 2011).

Striking sex differences in the functional architecture (Ingallhalikar et al., 2014) and developmental trajectory (Simmonds et al., 2014) of the brain of children, adolescents and young adults were established recently. The above cited studies suggest that males and females are characterized by modularity and cross-modularity of the neural architecture as

Abbreviations: BMI, Body mass index; FS, Fast spindle; IAM, Individual Adjustment Method (of sleep spindle analysis); RPMT, Raven Progressive Matrices Test; S2, Stage 2 sleep; SS, Slow spindle; SWS, slow wave sleep.

well as linear and non-linear white matter growth, respectively. In addition, gender roles were also shown to have a modulatory effect on regional brain volumes of children and adolescents (Belfi et al., 2014). Striking sex differences in the neural correlates of intelligence were reported in terms of waking electroencephalogram (EEG; Neubauer et al., 2002; Jausovec and Jausovec, 2005) and brain anatomy (Gur et al., 1999; Haier et al., 2005): neural connectivity measures and white matter structures are reliable neurobiological correlates of intelligence in women but not in men. Consequently, sex and gender are of primary interest when investigating the brain-derived factors related with adolescent neurocognitive development. Neural connectivity and white matter-related indices expressing cross-modular brain organization are candidate neurobiological markers of cognitive efficiency in females, but not in males.

Sleep EEG is considered to be the clearest window through which to view adolescent brain development (Colrain and Baker, 2011a). The sleep EEG changes during adolescence were considered as indexes of fundamental brain reorganization (Feinberg and Campbell, 2010). As Colrain and Baker (2011a) acknowledged, EEG power reflects the sum of inhibitory and excitatory postsynaptic potentials in thousands of neural columns sampled by an individual electrode, and the curve describing changes in delta EEG over the lifespan is remarkably similar to those based on postmortem anatomic synaptic density measures and cerebral metabolic rate (Feinberg and Campbell, 2010). Most of the known sleep architectural or quantitative EEG measures strongly and reliably depend on the chronological age of the adolescent subjects. Reports on developmental changes in human sleep most frequently emphasize age-related increases in Stage 2 (S2) sleep percentage, and decreases in slow wave sleep (SWS) percentage. The above changes are reflected in age-related decreases in quantitative EEG measures of sleep EEG delta and theta waves during both NREM and REM sleep (Ringli and Huber, 2011; Colrain and Baker, 2011b; Feinberg and Campbell, 2013).

Sleep spindles are groups of rhythmic neuronal oscillations in the frequency range of sigma waves (11–16 Hz), constituting the hallmarks and major defining features of NREM sleep (De Gennaro and Ferrara, 2003; Lüthi, 2014). Hypotheses on the preferential involvement of sleep spindles in sleep-related neural plasticity (Timofeev et al., 2002), offline information processing (Fogel and Smith, 2011), and sleep protection (Dang-Vu et al., 2010) have been put forward. Individual profiles in sleep EEG spindling reflect the microstructural properties of white matter tracts as measured by diffusion weighted magnetic resonance imaging, with high levels of spindling being related to high axial diffusivity in white matter structures (Piantoni et al., 2013). Moreover, sleep spindles were shown to constitute a physiological index of overall mental efficiency or intelligence (Fogel and Smith, 2011). Several studies emphasized the differences between the frontally and centro-parietally dominant slow (~11–13 Hz) and fast (~13–16 Hz) spindles (SSs and FSs), respectively (De Gennaro and Ferrara, 2003). Apart from frequency and topography other differences in specific features characterizing spindle types are seen in the hemodynamic activities indexing neural activation patterns associated with SSs and

FSs (Schabus et al., 2007). Moreover, increasing evidence supports the thesis on the specificity of the cognitive correlates of SSs and FSs: SSs were shown to correlate with visual perceptual learning (Bang et al., 2014), while FSs with more complex abilities and processes, like fluid intelligence (Bódizs et al., 2005), visuospatial memory (Bódizs et al., 2008), learning ability (Lustenberger et al., 2012) and word-location associations (Cox et al., 2014).

In spite of the hints on the potential significance of sleep EEG spindle measures in unraveling the details of the neurodevelopmental processes of adolescence (Tarokh et al., 2011), there are only a few controversial reports focusing specifically on this issue. Although it was claimed that sleep spindle activity changes with maturation until the age of 16 years in terms of length and density (Scholle et al., 2007) there is only scarce data on late adolescence or on the transition from adolescence to adulthood. In contrast to Scholle et al. (2007), Shinomiya et al. (1999) reported a decrease in the power of slow sleep spindling until the age of 13 years, but little change in the power of fast centroparietal spindles between 4 and 24 years. These controversies might result from an inappropriate methodological approach of the individual-specific and developmentally changing frequencies of SSs and FSs. Given the finding on the relationship between individual level of sleep spindling and white matter integrity (Piantoni et al., 2013) as well as the continuing white matter development during late adolescence (Peters et al., 2012) the issue of adolescent development in sleep spindling is of utmost importance. The potential significance of a detailed analysis of sleep spindling during adolescence is further supported by the correlations of specific sleep spindle measures with late developing, higher order intellectual performances of preadolescent (Geiger et al., 2011, 2012; Chatburn et al., 2013; Gruber et al., 2013) and adult human volunteers (Bódizs et al., 2005, 2008; Schabus et al., 2006, 2008; Lustenberger et al., 2012). According to our knowledge, no data on the sleep spindle-intellectual ability relationship in adolescents was published in the literature. Thus, the potential relevance of the above mentioned sleep spindle-related EEG indexes in revealing the individual patterns of cognitive development remained largely neglected in previous reports.

In summary adolescence is a critical period of brain maturation and cognitive development, presumably characterized by increasing sexual dimorphism and gender-divergence. In spite of the fact that sleep EEG was acknowledged as a prominent route in discerning the neurodevelopmental processes of adolescence and individual-specific measures of sleep spindling were shown to reflect complex cognitive processes and faculties in both children and adults, no prior study explicitly addressed the neurocognitive developmental aspects of sleep spindle oscillation in adolescents. The corroboration of the above cited evidence for a positive association of fast sleep spindling with complex, human-specific cognitive performances and faculties in both children and adults, with the unequivocal growth of white matter structures in the adolescent brain, and with the relationship between white matter integrity and sleep spindling lead us to hypothesize that fast sleep spindling correlates positively with chronological age (H1). By completing the above considerations with the claim

suggesting that white matter is the major determining neural substrate of thinking in women, but not in men we further hypothesize that fast sleep spindling predicts overall mental efficiency as measured by intelligence tests primarily in females (H2).

Hypotheses were tested in a home polysomnographic study focusing on the effects of chronological age and developmentally acquired overall mental efficiency (fluid IQ) with sex as a potential modulating factor.

MATERIALS AND METHODS

SUBJECTS

Subjects ($N = 24$, 12 males) were adolescents of Hungarian nationality recruited by a convenience sampling procedure. Age range was 15–22 years, while mean age was 18 years (SD: 2.3 years). The whole examined age range was subdivided into four subgroups (groups of 15–16, 17–18, 19–20 and 21–22 years old subjects). Six participants were included in each subgroup: 3 females and 3 males. Thus, subjects were evenly distributed over the age range. Mean height of the subjects was 173.04 cm (range: 160–198, SD: 10.57). Subjects' weight averaged 63.83 kg (range: 47–92, SD: 11.92), while their body mass index (BMI) was between the normal limits (mean: 21.19, range: 17.68–27.01, SD: 2.6).

Subjects were interviewed on their health status by the authors of the study. Exclusion criteria for the participants were self-reported sleep problems or diagnoses of psychiatric, neurological or other medical disorders. Subjects were requested to not to drink alcohol containing beverages, to not to take drugs other than caffeine before noon and to not to take naps during the study.

The research protocol was approved by the Ethical Committee of the Pázmány Péter Catholic University Budapest. Adult participants or the parents of the underage participants signed informed consent for the participation in the study according to the Declaration of Helsinki.

PROCEDURES

Fluid intelligence was tested by using the Raven Progressive Matrices Test (RPMT), which is based on items assessing the abilities in the field of non-verbal reasoning (Raven et al., 1976). Scores of the RPMT were shown to be among the most reliable measures of the general factor of mental abilities (Gray and Thompson, 2004). Raw RPMT scores were transformed to IQ by using the Hungarian standards (Raven et al., 2004). As a consequence the term IQ reflects fluid instead of crystallized intelligence throughout our paper. Subjects' sleep was recorded at their homes by using ambulatory home polysomnography. Sleep recordings on two consecutive weekend nights were performed according to the subjects' sleeping habits. We used a portable SD LTM 32BS Headbox together with a BRAIN QUICK System PLUS software (Micromed, Italy) for polysomnographic data recording. We recorded EEG according to the 10–20 system (Jasper, 1958) at 21 recording sites (Fp1, Fp2, Fpz, F3, F4, F7, F8, Fz, C3, C4, Cz, P3, P4, Pz, T3, T4, T5, T6, O1, O2, Oz) referred to the mathematically linked mastoids. Bipolar EOG, ECG and submental as well as tibialis EMG were also recorded. Electroencephalogram

and polygraphic data were high-pass filtered at 0.15 Hz and low-pass filtered at 250 Hz (both 40 dB/decade). Data were collected with an analog to digital conversion rate of 4096 Hz/channel (synchronous, 22 bit). A further 40 dB/decade anti-aliasing digital filter was applied by digital signal processing (firmware) which low pass filtered the data at 463.3 Hz before the decimation by a factor of 4, resulting in a sampling rate of 1024 Hz.

Sleep recordings of the second nights were visually scored according to standard criteria (Rechtschaffen and Kales, 1968) in 20 s epochs. The following definitions were used for sleep architecture evaluation: time in bed (as the time from lights out to final awakening), total sleep time (defined as the amount of sleep from sleep onset to final awakening), wake time after sleep onset (WASO, excluding wakefulness after the final awakening), sleep efficiency (calculated as the percent of sleep time without WASO divided by the time in bed), sleep latency (defined as the period between lights off and the first appearance of S2 sleep), non-rapid eye movement (NREM), Stage 1 (S1), S2, SWS (defined as the amount of time spent in Stages 3 and 4), rapid eye movement sleep (REM), REM latency (defined as the period between sleep onset and the first epoch scored as REM), number of sleep cycles (number of REM periods separated from each other by more than 15 min), average REM period duration (duration of REM sleep divided by the number of REM periods) and average sleep cycle duration in minutes (sleep time from the sleep onset to the end of the last REM period divided by the number of sleep cycles).

The 4 s epochs containing artifactual sleep EEG (movement, sweating or technical artifacts) were manually removed before further automatic sleep EEG analyses. One male subject was excluded from the below listed quantitative EEG analyses (but not from the above mentioned sleep architectural one) because of technical artifacts interfering with deliberate and reliable signal processing approaches.

The Individual Adjustment Method (IAM) of sleep spindle analysis (Bódizs et al., 2009) was used to unravel the potential peculiarities of NREM sleep (stages 2–4) EEG spindling. In short the principle of sleep spindle detection is the idea that individual spindles are those groups of waves which last at least 0.5 s and contribute to one or two of the major peaks in the 9–16 Hz average amplitude spectra of NREM sleep EEG. Individual-specific spectral peaks were formalized by calculating the zero crossing points of their second order derivatives. The lower frequency peak corresponds to SSs while the higher frequency peak to FSs. As a result, features like mean density (spindles/min), duration (s) and amplitude (μV) of SSs and FSs can be determined in an individual- and derivation-specific manner. The dominant individual-specific frequency (Hz) of SSs and FSs is inherently derivation-independent in the IAM procedure. Based on the derivation-specific data on density, duration and amplitude we created averages for five regions: all derivations (region-independent), frontal derivations (Fp1, Fp2, Fpz, F3, F4, F7, F8, Fz), centro-parietal derivations (C3, C4, Cz, P3, P4, Pz), temporal derivations (T3, T4, T5, T6) and occipital derivations (O1, O2, Oz). Region-specific averages were used for descriptive purposes while the region-independent average values were starting points of inferential statistics.

Additional analyses were based on Fast Fourier Transform-based measurement of binwise spectral power in the 8–16 Hz range of all-night average NREM sleep (stages 2–4) EEG covering alpha and sigma waves. In line with the relevant guidelines, spectral power was log-transformed before the statistical analyses (Pivik et al., 1993; Jobert et al., 2013). This transformation is required in order to normalize the distribution of power values. Besides log-transformation, z-scores of the 8–16 Hz spectra were also analyzed. This latter transformation is justified by the findings supporting the striking trait-like reliability (De Gennaro et al., 2005) and the marked sensitivity (Bódizs et al., 2012) of this sleep EEG scores expressing discrete frequency points of the individual shapes of the sleep EEG spectra. Both log-transformed power (10th base) and z-transformed normalization ($x-m/SD$) were used in separate statistical models. Our aim was to compare the results based on the more sophisticated IAM of sleep spindle analysis with the relatively simple spectral analysis. While IAM is sensitive to sleep spindle features at the individual frequencies, spectral power mapping is able to provide evidence for the importance of sleep spindle activity occurring at specific frequencies.

STATISTICS

Descriptive statistics on IQ, as well as on sleep architecture and regional sleep spindling are provided. As for inferential statistics, we followed a top-down approach by using consecutive tests progressing from global to gender-specific and local effects. The average (region-independent) sleep spindle variables (frequencies, densities, durations and amplitudes of individual specific SSs and FSs) were correlated with the output variables (age and IQ) by using the Pearson product-moment procedure. In case of the emergence of a significant region-independent correlation the next step was to analyze sexual dimorphism of the relationship (by comparing correlations for females and males using the Fisher r -to- z transformation), as well as to depict the potential region-specificities of the significant global effects by subjecting the derivation-specific sleep spindle vs. output variable correlations to the procedure of descriptive data analysis (Abt, 1987) adapted to quantified neurophysiology with mapping (Abt, 1990; Duffy et al., 1990). This procedure tests the global null hypothesis (“all individual null hypotheses in the respective region are true”) at level $\alpha = 0.05$, against the alternative that at least one of the null hypotheses is wrong. According to Abt (1987) and Duffy et al. (1990) local, uncorrected significances at the level of $\alpha = 0.05$ (descriptive significances) define the R ger’s areas (R ger, 1978). If N is the number of electrodes in the R ger’s area, the investigator is required to choose a minimal number of unspecified null hypotheses (M), less than N , to be nominally rejected at a new, more conservative α level. Typically the value M/N is 1/2 or 1/3. The corresponding new α levels for these values are $\alpha/2 = 0.025$ and $\alpha/3 = 0.017$, respectively. We will use an M/N value of 1/2 and a corresponding new α of 0.025 in our analyses. If any M values (half of the correlation coefficients if $M/N = 1/2$) within the R ger’s area individually reach the new α level of significance the overall null hypothesis is rejected for the R ger’s area at the 0.05 level. This means that for at least one EEG derivation in the R ger’s area

the relationship is significant, allowing the investigator to make global confirmatory statement with controlled uncertainty. In order to obtain a better localization of regions with significant correlations between sleep spindling and IQ the correlations were represented by significance probability maps (Hassainia et al., 1994). Finally, we tested the age-independence of the relationship between sleep spindling and IQ by recalculating the significant spindle-IQ correlations with the effects of age partialled out.

Binwise NREM sleep EEG spectral data between 8 and 16 Hz was correlated with age and with IQ in females and males by using the same methodology as described above.

RESULTS

FLUID INTELLIGENCE

Raven Progressive Matrices Test-derived IQ-scores of the sample resulted in a group average of 104.12 (range: 91–126, SD: 10.82). Neither age ($r = 0.30$; $p = 0.15$), nor weight ($r = 0.13$; $p = 0.51$), height ($r = 0.14$; $p = 0.50$) nor BMI ($r = 0.06$; $p = 0.77$) correlated significantly with IQ. Males and females did not differ in their general mental abilities ($t = 0.31$; $p = 0.75$).

SLEEP ARCHITECTURE AND SLEEP SPINDLING

Details on sleep architecture of our sample are depicted in Table 1. In short subjects had a normal sleep structure with 4–8 sleep cycles, an average total sleep time of 8.23 h, a sleep efficiency of 94.84%, over 59% of S2, 12% of SWS and 25% of REM sleep (Table 1).

Slow spindle densities, durations and amplitudes prevail in the frontal regions. In contrast densities, durations and amplitudes of FSs peak in the centroparietal area (Table 2).

Table 1 | Descriptive statistics of sleep architectural variables*.

	Mean	Min	Max	SD
Time in bed (min)	521.65	399.00	639.33	59.29
Total sleep time (min)	494.33	368.33	617.00	54.60
Sleep efficiency (%)	94.84	85.25	99.09	3.36
Wake time (min)	27.31	4.33	85.00	19.03
Relative wake time (%)	5.15	0.90	14.74	3.36
WASO (min)	19.50	1.00	81.66	19.02
Sleep latency (min)	10.72	2.00	38.00	10.09
NREM time (min)	365.86	302.00	447.00	38.33
Relative NREM time (%)	74.16	66.28	81.99	4.00
S1 time (min)	10.68	3.00	33.66	6.36
Relative S1 time (%)	2.16	0.62	6.28	1.23
S2 time (min)	294.34	208.33	386.00	49.70
Relative S2 time (%)	59.59	43.61	75.83	7.93
SWS time (min)	60.83	3.00	162.33	37.15
Relative SWS time (%)	12.40	0.56	33.98	7.70
REM time (min)	128.47	66.33	170.00	27.35
Relative REM time (%)	25.83	18.00	33.71	4.00
REM latency (min)	79.02	44.66	150.00	28.08
Number of sleep cycles	5.08	4.00	8.00	0.97
Average REM period time (min)	25.63	15.66	34.75	5.23
Average sleep cycle time (min)	99.35	74.83	133.91	15.32

*S1—Stage 1 sleep; S2—Stage 2 sleep; SWS—slow wave sleep; WASO—wake after sleep onset.

Table 2 | Descriptive statistics on sleep spindling*.

	Mean	Min	Max	SD
SS frequency (Hz)	11.28	9.88	13.10	0.83
Average SS density (spindles \times min $^{-1}$)	6.83	4.45	9.29	1.46
Frontal SS density (spindles \times min $^{-1}$)	7.14	5.17	9.18	1.14
Centroparietal SS density (spindles \times min $^{-1}$)	6.68	4.06	9.35	1.56
Temporal SS density (spindles \times min $^{-1}$)	6.73	4.00	9.60	1.70
Occipital SS density (spindles \times min $^{-1}$)	6.46	2.85	9.58	2.03
Average SS duration (s)	1.41	0.86	2.58	0.54
Frontal SS duration (s)	1.48	0.89	2.63	0.53
Centroparietal SS duration (s)	1.39	0.85	2.56	0.54
Temporal SS duration (s)	1.38	0.83	2.53	0.54
Occipital SS duration (s)	1.33	0.76	2.55	0.55
Average SS amplitude (μ V)	3.48	1.30	6.91	1.65
Frontal SS amplitude (μ V)	4.42	1.50	9.86	2.25
Centroparietal SS amplitude (μ V)	3.64	1.39	7.21	1.72
Temporal SS amplitude (μ V)	2.26	0.84	4.24	1.02
Occipital SS amplitude (μ V)	2.34	0.80	4.89	1.10
FS frequency (Hz)	13.29	12.55	14.24	0.47
Average FS density (spindles \times min $^{-1}$)	7.27	5.50	8.29	0.75
Frontal FS density (spindles \times min $^{-1}$)	6.59	5.07	7.73	0.75
Centroparietal FS density (spindles \times min $^{-1}$)	8.11	6.58	9.74	0.75
Temporal FS density (spindles \times min $^{-1}$)	7.20	5.23	8.57	0.92
Occipital FS density (spindles \times min $^{-1}$)	7.45	4.86	9.01	0.93
Average FS duration (s)	1.10	0.89	1.29	0.09
Frontal FS duration (s)	1.02	0.86	1.16	0.07
Centroparietal FS duration (s)	1.21	0.97	1.45	0.10
Temporal FS duration (s)	1.07	0.88	1.29	0.09
Occipital FS duration (s)	1.15	0.86	1.40	0.11
Average FS amplitude (μ V)	5.10	3.45	7.08	1.04
Frontal FS amplitude (μ V)	4.88	2.95	7.02	1.08
Centroparietal FS amplitude (μ V)	7.16	4.64	10.38	1.60
Temporal FS amplitude (μ V)	3.26	2.42	4.37	0.50
Occipital FS amplitude (μ V)	4.01	2.08	5.74	1.10

*SS—slow spindle, FS—fast spindle.

SLEEP SPINDLING AND AGE

Average FS density correlated positively with chronological age ($r = 0.57$; $p = 0.005$; **Figure 1**). No other sleep spindle measures were significantly related with the age of our subjects. There was no significant difference between the age vs. FS density correlations of females and males [$r = 0.62$ and $r = 0.52$, respectively; $p = 0.76$ (two-sided)].

The region-specific analysis revealed a significant age-related increase in FS density measured at 16 of 21 derivations (F3, F4, Fz, C3, C4, Cz, T3, T4, T5, T6, P3, P4, Pz, O1, O2, Oz) defining a significant R ger's area (16/16 p values < 0.025) consisting of frontal, centroparietal, temporal and occipital regions, but not of frontopolar-orbitofrontal (Fp1, Fp2, Fpz, F7, F8) ones (**Figure 2**).

SLEEP SPINDLING AND IQ

Intelligence quotient was shown to be significantly and positively related to average FS density ($r = 0.43$; $p = 0.04$) and amplitude ($r = 0.41$; $p = 0.049$). While females were characterized by significant FS density vs. IQ, as well as FS amplitude vs. IQ correlations [$r = 0.80$ ($p = 0.002$) and $r = 0.67$

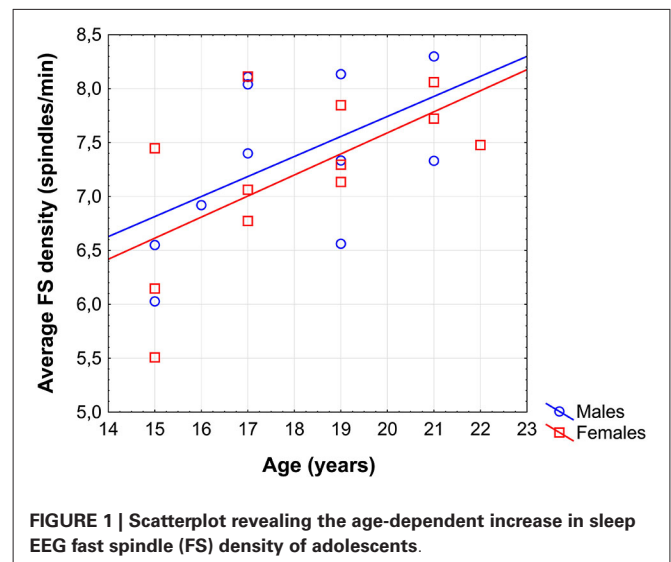


FIGURE 1 | Scatterplot revealing the age-dependent increase in sleep EEG fast spindle (FS) density of adolescents.

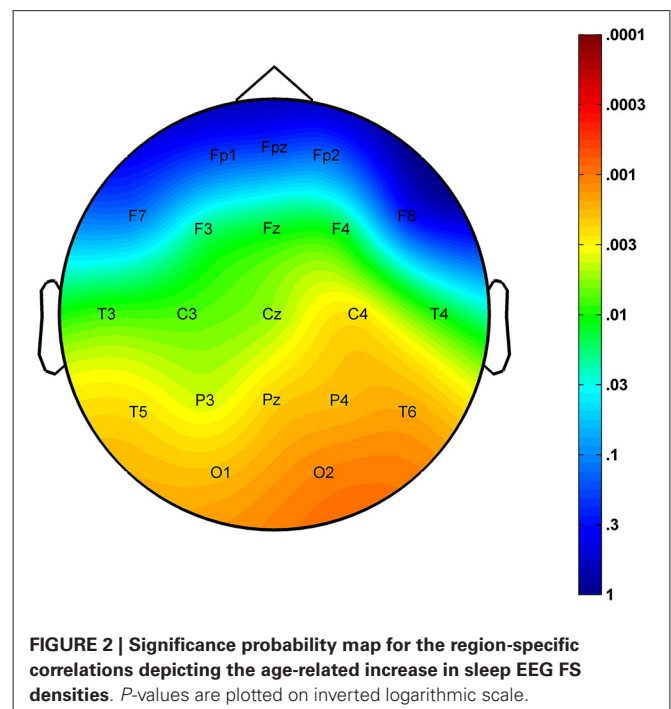
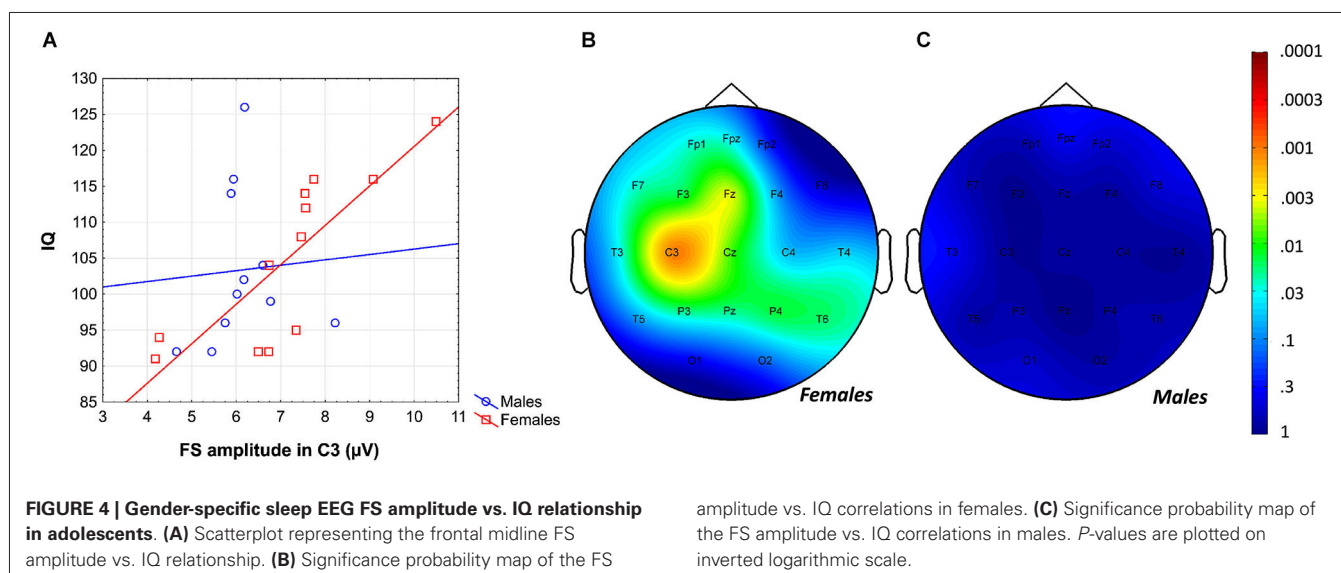
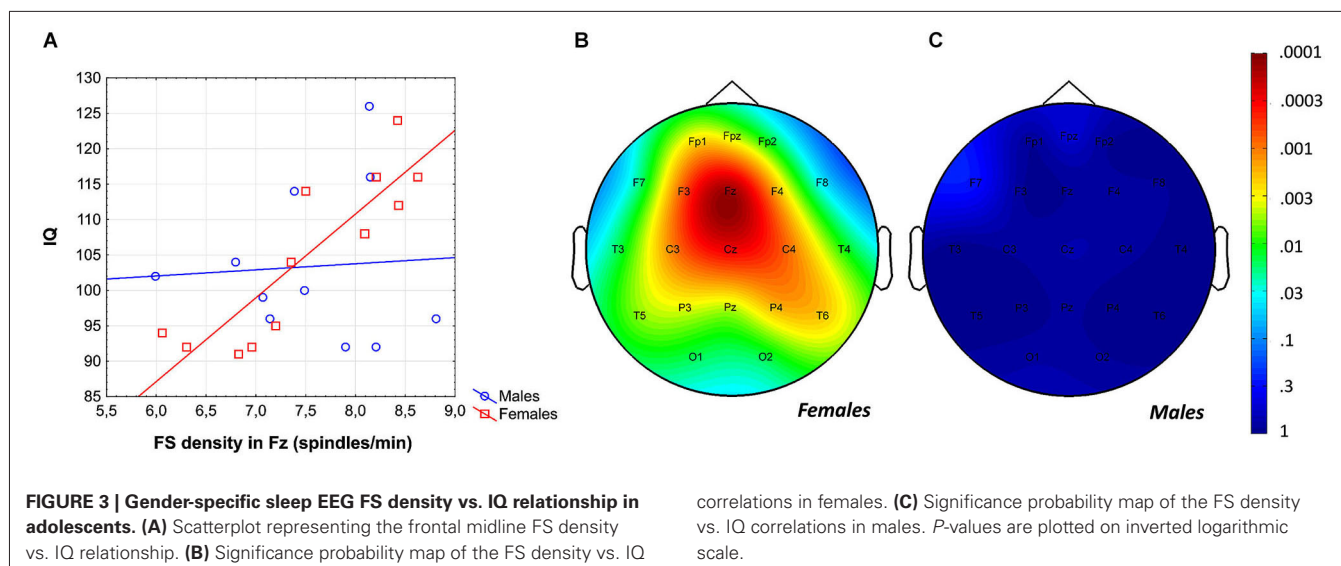


FIGURE 2 | Significance probability map for the region-specific correlations depicting the age-related increase in sleep EEG FS densities. P -values are plotted on inverted logarithmic scale.

($p = 0.012$), respectively, males were not [$r = 0.00$ ($p = 0.99$) for both measures]. Differences between the correlation coefficients depicting the linear relationship between FS density vs. IQ of females and males was significant ($p = 0.017$, one-sided). However, the female-male difference in FS amplitude vs. IQ correlation proved to be a tendency only ($p = 0.055$, one-sided). One-sided statistics were used because of our explicit hypothesis on female predominance in the spindle vs. IQ correlations.

The region-specific analysis of the FS density vs. IQ correlation of females revealed significant correlations in 21 out of 21 derivations, 19 of which were significant at the level of 0.025



(Figure 3). Thus, findings fulfill the criteria for rejecting the global null hypothesis. Maximal significances were revealed over the frontal midline region ($r = 0.90$; $p = 0.0001$ at derivation Fz).

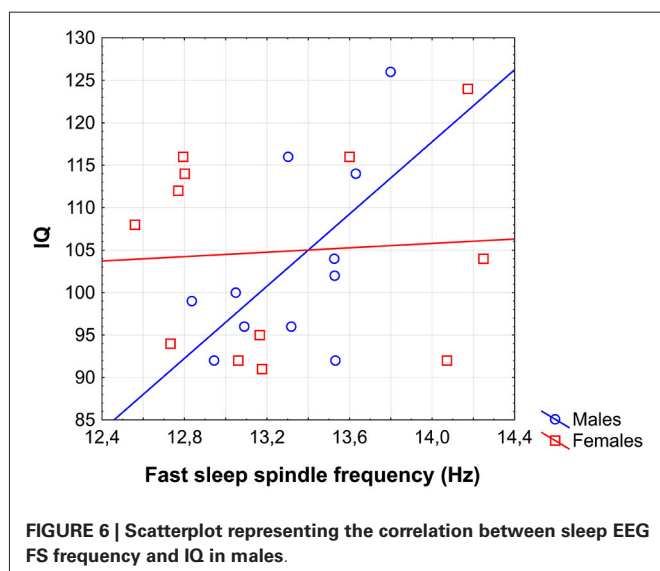
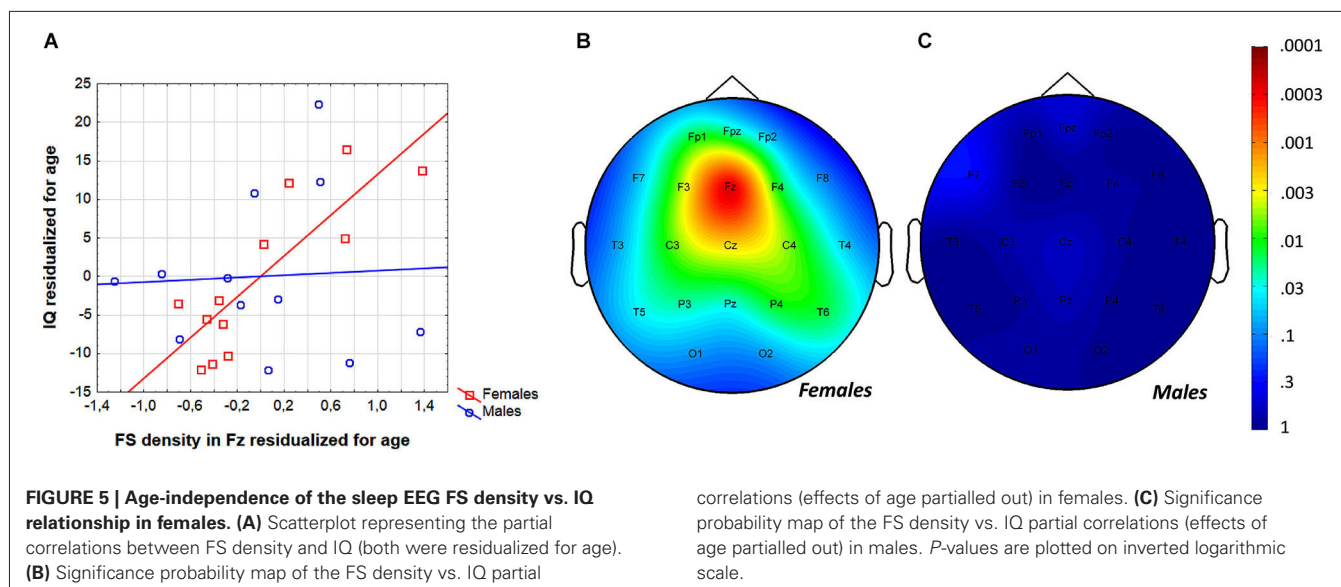
Likewise, the region-specific analysis of the FS amplitude vs. IQ correlation of females revealed significant correlations in 12 out of 21 derivations (Fp1, Fpz, F3, F7, Fz, C3, Cz, P3, P4, Pz, T3, T6), 8 of which were significant at the level of 0.025 (Figure 4). Again, based on these findings the global null hypothesis can be rejected. Maximal significances were revealed over the left central region ($r = 0.82$; $p = 0.001$ at derivation C3).

AGE-CORRECTED RELATIONSHIPS BETWEEN SLEEP SPINDLING AND IQ

In order to test whether individual levels of fast sleep spindling age-independently predict general mental ability in adolescent

females, partial correlations were calculated and entered in the procedure of descriptive data analysis and significance probability mapping (Figure 5). We found 13 significant correlations (out of 21) between FS density and IQ with the effects of age partialled out. The Rüger's area consisted of a wide region including frontopolar-prefrontal, central, parietal and posterior temporal locations (Fp1, Fpz, F3, F4, Fz, C3, C4, Cz, T5, T6, P3, P4, Pz) with p values less than 0.025 at 11 derivations. Thus, the area includes significant FS density vs. IQ partial correlation (with the effects of age held constant) in adolescent females. Maximal correlation emerged at the frontal midline derivation Fz ($r = 0.90$; $p = 0.0002$).

The same analyses were run with FS amplitudes. Eight out of 21 partial correlations were significant in adolescent females, depicting a scattered parasagittal area (F7, Fz, C3, Cz, T6, P3, P4, Pz) with four p values being less than 0.025. Thus, the null



hypothesis cannot be unambiguously rejected for this R ger's area.

ARE THERE ANY SLEEP SPINDLE CORRELATES OF IQ IN MALES?

In previous analyses we progressed from global to sex-specific and local effects. This approach could hinder the recognition of some weaker, male-specific correlations between sleep spindles and IQ. In order to reveal any male-specific sleep spindle correlates of IQ in adolescents the correlations between all sleep spindle variables and IQ were checked for the male subgroup only. Analysis revealed a significant correlation of FS frequency with IQ in males ($r = 0.60$; $p = 0.04$; **Figure 6**). Partialling out the effects of age even slightly increased the strength of this relationship ($r = 0.65$; $p = 0.04$). No other correlation between sleep spindle measures and IQ in males proved to be significant.

EEG SIGMA POWER

In females, neither log-transformed EEG powers nor z-scores revealed significant associations with IQ after the R ger area correction, with or without control for the effects of age.

In males, however, a positive association between log-transformed EEG power on F3, C3 and C4 between 13.75 and 15 Hz ($r_{\max} = 0.70$; $p = 0.014$ on F3 at 14 Hz) is significant after R ger correction, while there is a tendency (with significant correlations not surviving R ger correction) for a negative correlation between IQ and log-transformed power between 12.75 and 13 Hz on T5 and Pz (**Figure 7A**). Using EEG power z-scores, a significant negative correlation between IQ and power is present between 12 and 13.25 Hz on C3, C4, P3, P4, Pz, T3, T4, T5, T6, O1 and O2 ($r_{\max} = -0.78$; $p = 0.001$ on T5 at 12.75 Hz; **Figure 7B**). Similar results were obtained if age-controlled correlations were used. In this case, no R ger-significant effects are evident in females, while there is a significant negative correlation between IQ and power z-scores between 12 and 13.5 Hz (on C3, P3, P4, Pz, T3, T4, T5, T6, O1, O2, and Oz) in males. The positive correlation between IQ and log power is present between 13.75 and 15 Hz (on F3, C3, and C4) in males, but does not reach significance after correcting with the R ger area method.

DISCUSSION

We performed a home polysomnographic study in order to unravel the developmental peculiarities of sleep spindling during adolescence as well as to test the predicted sexual dimorphism in the sleep spindle-IQ relationship during the period of the late maturation of the frontal lobes. Advantages of our study are the familiar, thus relatively non-disturbing sleeping environments and settings. Moreover, sleep was timed according to the preferred sleeping times of our subjects during two consecutive weekend nights. These circumstances are reflected in relatively long total sleep times (**Table 1**), at least when compared to laboratory based average values (Ohayon et al., 2004). Since longer sleep times lead

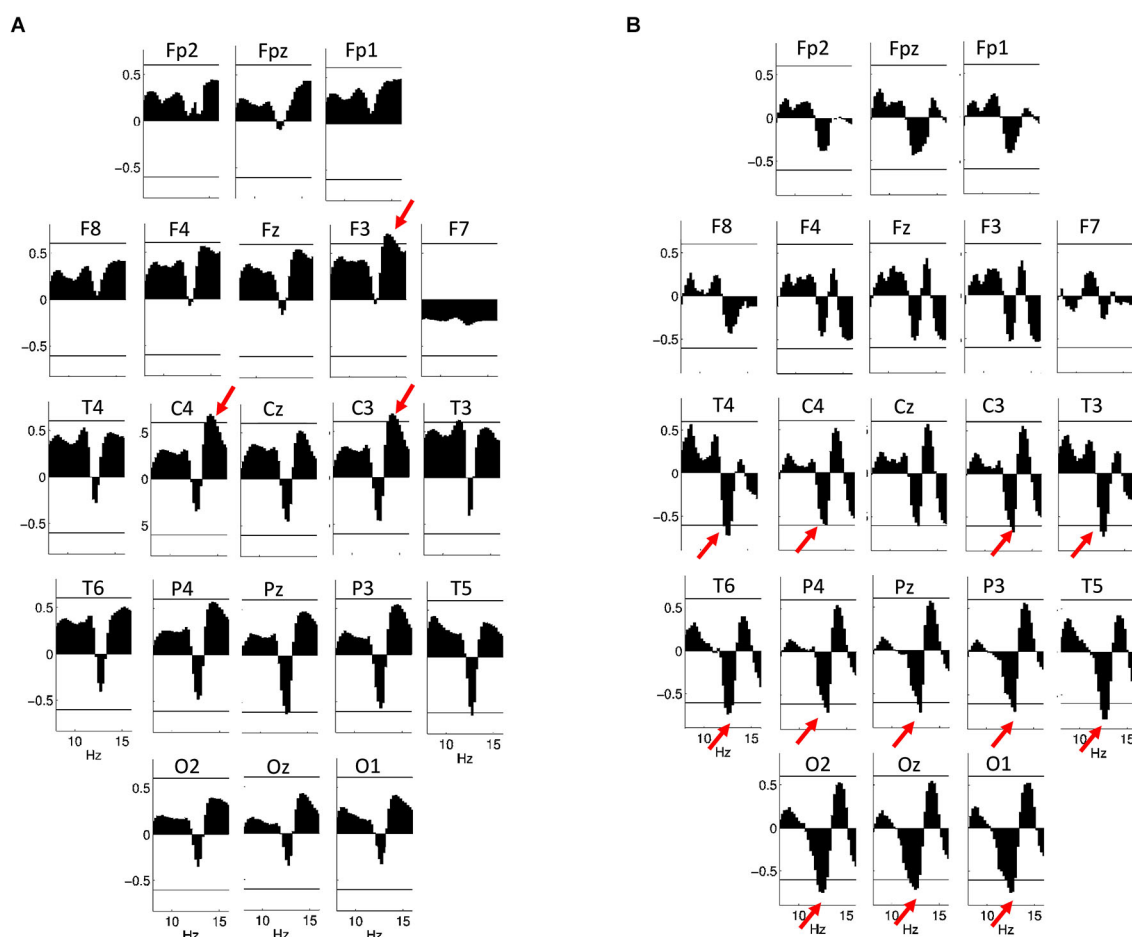


FIGURE 7 | Correlations between NREM sleep EEG spectral power of 8–16 Hz frequency and IQ in males. Graphs are indicating region-specific correlations as revealed at different scalp locations. Horizontal lines denote critical values for $p < 0.05$. **(A)** Binwise spectral data were log-transformed (10th base) before implementing correlation analyses. Positive correlations of NREM sleep EEG 13.75–15 Hz spectral power at derivations F3, C3 and C4 with IQ (red arrows) are significant

after controlling for multiple testing according to the procedure of descriptive data analysis. **(B)** Binwise spectral data were z-transformed before implementing correlation analyses. Negative correlations of NREM sleep EEG 12–13.25 Hz spectral power at derivations C3, C4, P3, P4, Pz, T3, T4, T5, T6, O1, O2 and Oz with IQ (red arrows) are significant after controlling for multiple testing according to the procedure of descriptive data analysis.

to increases in S2 and REM sleep, the relative times spent in these two sleep stages were higher than usual while relative SWS times were lower. Given the fact that sleep spindles are most expressed in S2 sleep (De Gennaro and Ferrara, 2003) the above circumstances are not likely to mask the neurocognitive developmental aspects of sleep spindles.

Recent reports revealed the relationship between individual levels of sleep spindling and white matter integrity (Piantoni et al., 2013). Moreover, white matter continues to develop during late adolescence (Peters et al., 2012) resulting in continuously increasing integration and decreasing segregation of structural connectivity with age (Hagmann et al., 2010). We have shown that the prevalence (density) of centroparietally dominant FS of adolescents increases with age in both sexes, suggesting that “network refinement mediated by white matter maturation” (Hagmann et al., 2010) might be indexed by specific measures of sleep spindling (i.e., FS density). Thus, our current finding on

the age-dependent increase in FS density in adolescents coheres with the above mentioned neuroimaging data (Hagmann et al., 2010; Peters et al., 2012; Piantoni et al., 2013) and strengthens/expands the reliability of the hypothesis suggesting that fundamental reorganization of cortical networks during adolescence is indexed by the changing patterns of synchronous, oscillatory activity (Uhlhaas et al., 2009; Konrad et al., 2013). Therefore, it is reasonable to assume that beside sleep EEG delta and theta activity indexing adolescent brain maturation (Feinberg and Campbell, 2013), sleep spindling is another neurophysiological marker with potential neurodevelopmental relevance. Given the widely accepted hypothesis on the thalamo-cortical origin of sleep EEG spindle oscillations (De Gennaro and Ferrara, 2003; Lüthi, 2014) the fundamental reorganization of the adolescent brain probably involves the developmental enhancement of the functionality of cortico-thalamic networks. As for the additional neurodevelopmental aspects of sleep spindling, it is worth noting,

that the age-dependent increase of FS density during adolescence is the mirror image of the age vs. FS relationship of adult subjects, as the latter is characterized by a decline in spindling with increasing ages (Bódis et al., 2009). Thus, the increasing FS density during adolescence suggests an inverted U-like relationship between age and fast sleep spindling during the human lifespan with maximal spindling emerging during the periods of maximal cognitive efficacy.

There are several previous studies investigating the relationship between cognitive abilities and sleep EEG spindling. Most of these studies are based on data from adult volunteers (Bódis et al., 2005, 2008; Schabus et al., 2006, 2008; Lustenberger et al., 2012), some of them on investigations on preadolescent children (Geiger et al., 2011, 2012; Chatburn et al., 2013; Gruber et al., 2013), while none of them specifically addressed the period of late maturation of frontal lobes and related higher order cognitive functions. Here we aim to fill this gap by analyzing the period of adolescence and the transition from adolescence to adulthood from the perspective of sleep EEG spindle oscillation. Our present results on sleep spindle-IQ correlation and its predominantly frontal topography echoes previous findings (Bódis et al., 2005; Fogel and Smith, 2011), further strengthens the primary role of frontal regions in intelligence (Gray and Thompson, 2004; Shaw et al., 2006), but also completes the picture with the issue of sex-specificity: FS density and amplitude was strongly and positively related with IQ in females only. Sleep spindles were shown to reflect the structural properties of white matter tracts (Piantoni et al., 2013). Thus, the female-specificity of the FS-IQ relationship reported here is reminiscent of earlier reports suggesting that anatomical measures of white matter structures are markers of cognitive ability in women, but not men (Gur et al., 1999; Haier et al., 2005). As white matter structures in fact serve efficient large-scale neural connectivity, the evidence indicating that EEG connectivity measures of the wakeful resting state are predictive of intelligence exclusively in women (Neubauer et al., 2002; Jausovec and Jausovec, 2005) might pertain in the same pattern of sexual dimorphism. In contrast with females, males were not characterized by a tight relationship of FS density or amplitude with IQ. Males, however, in contrast to females, were characterized by a positive FS frequency vs. IQ correlation. This was supported by spectral power data, which suggested a pattern of negative correlation between IQ and sigma power around 13 Hz as well as a positive correlation with higher sleep spindle frequencies around 14 Hz. Together, these results suggest that in adolescent males the tuning of sleep spindles to a higher, adult-like FS frequency is a more stable correlate of IQ than either amplitude or duration at the given individual frequency. While sleep spindle frequency has been shown to be a correlate of cognitive ability (Geiger et al., 2011; Bódis et al., 2012), our results do not rule out the possibility that this correlation between IQ and spindle frequency is due to the effect of a maturation process which has already taken place in females of the same age.

Female sleep spindling frequency was shown to be influenced by the phase of the menstrual cycle (Ishizuka et al., 1994). As we did not control our subjects for the menstrual cycle phase effects this could hinder the depiction of the FS frequency-IQ

relationship in females. Although, Tarokh et al. (2011) hypothesized that the increase of sleep spindle frequency during adolescence reflects the myelination of neural circuitry, there is no supporting evidence for this statement. However, we consider the above detailed sexually dimorphic correlations as further evidences for the fractionation of the general factor of intelligence into components (Conway and Kovács, 2013). Females, in contrast to males, rely on large-scale integration of neural circuitry during solving the complex non-verbal reasoning tasks of the RPMT. We hypothesize that this difference might emerge from different cognitive strategies of females and males. Indeed, there is evidence for certain sexual dimorphisms in cognitive strategies (Waller and Lin, 2012). Moreover, the report on the relationship between white matter structure and sleep spindling (Piantoni et al., 2013) together with our present finding on the relationship of individual level in FSs with IQ in females, but not in males serve as indirect evidences for the claim that women and men think with their white and gray matter, respectively (Zaidi, 2010).

Apart from the above mentioned difference between females and males other factors could contribute to the findings on the sexual dimorphism of the sleep spindle-IQ relationship of the present report. Among these factors the differences in the timing and the course of maturational processes (De Bellis et al., 2001) has to be mentioned.

There are several limitations of our study among which the relatively low number of subjects and the lack of longitudinal data must be mentioned. A higher number of subjects as well as a follow-up of our volunteers could provide a further refinement of our findings on the developmental aspects of sleep spindling and its relationship with general mental abilities in adolescents. Moreover, we did not monitor respiratory parameters and leg movements during sleep. Although sleep apnea and periodic leg movements during sleep are rare phenomena during adolescence we cannot completely rule out the possibility of the presence of these syndromes in some of our subjects.

To sum up our main empirical findings and conclusions we emphasize the following statements: (1) FS density is increasing during adolescent development; (2) FS density is an age-independent positive correlate of fluid intelligence in female adolescents. This latter effect is maximal over the frontal area; (3) FS frequency is a positive, age-independent index of fluid intelligence in male adolescents; (4) efficient network reorganization in the adolescent brain is indexed by specific, individually adjusted sleep spindle measures.

AUTHOR CONTRIBUTIONS

Róbert Bódis contributed to: the conception and design of the study, the visual scoring of sleep records, the quantitative analysis of sleep spindling, statistical analysis (descriptive, and sleep spindle-related inferential), interpretation of data for the work, designing and creating the figures of the manuscript, writing and critical review of the manuscript. Ferenc Gombos contributed to: the conception and design of the study, the quantitative analysis of sleep spindling and power spectra, interpretation of data for the work, designing and creating the figures of the manuscript,

writing and critical review of the manuscript. Péter P. Ujma contributed to: statistical analysis of the spectral power measures, interpretation of data for the work, designing and creating the figures for the manuscript, writing and critical review of the manuscript. Ilona Kovács contributed to: the conception and design of the study, interpretation of data for the work, designing the figures of the manuscript, writing and critical review of the manuscript. All authors (Róbert Bódizs, Ferenc Gombos, Péter P. Ujma, Ilona Kovács) approved the final version of the manuscript to be published and agreed for all aspects of the work regarding accuracy and integrity.

ACKNOWLEDGMENTS

Research supported by the Hungarian Scientific Research Fund (OTKA-NK104481 to Ilona Kovács).

REFERENCES

- Abt, K. (1987). Descriptive data analysis: a concept between confirmatory and exploratory data analysis. *Methods Inf. Med.* 26, 77–88.
- Abt, K. (1990). Statistical aspects of neurophysiologic topography. *J. Clin. Neurophysiol.* 7, 519–534. doi: 10.1097/00004691-199010000-00007
- Bang, J. W., Khalilzadeh, O., Hämläinen, M., Watanabe, T., and Sasaki, Y. (2014). Location specific sleep spindle activity in the early visual areas and perceptual learning. *Vision Res.* 99, 162–171. doi: 10.1016/j.visres.2013.12.014
- Belfi, A. M., Conrad, A. L., Dawson, J., and Nopoulos, P. (2014). Masculinity/femininity predicts brain volumes in normal healthy children. *Dev. Neuropsychol.* 39, 25–36. doi: 10.1080/87565641.2013.839681
- Bódizs, R., Gombos, F., and Kovács, I. (2012). Sleep EEG fingerprints reveal accelerated thalamocortical oscillatory dynamics in Williams syndrome. *Res. Dev. Disabil.* 33, 153–164. doi: 10.1016/j.ridd.2011.09.004
- Bódizs, R., Kis, T., Lázár, A. S., Rigó, P., Havrán, L., Clemens, Z., et al. (2005). Prediction of general mental ability based on neural oscillation measures of sleep. *J. Sleep Res.* 14, 285–292. doi: 10.1111/j.1365-2869.2005.00472.x
- Bódizs, R., Körmendi, J., Rigó, P., and Lázár, A. S. (2009). The individual adjustment method of sleep spindle analysis: methodological improvements and roots in the fingerprint paradigm. *J. Neurosci. Methods* 178, 205–213. doi: 10.1016/j.jneumeth.2008.11.006
- Bódizs, R., Lázár, A. S., and Rigó, P. (2008). Correlation of visuospatial memory ability with right parietal EEG spindling during sleep. *Acta Physiol. Hung.* 95, 297–306. doi: 10.1556/APhysiol.95.2008.3.5
- Casey, B. J., Tottenham, N., Liston, C., and Durston, S. (2005). Imaging the developing brain: what have we learned about cognitive development? *Trends Cogn. Sci.* 9, 104–110. doi: 10.1016/j.tics.2005.01.011
- Chatburn, A., Coussens, S., Lushington, K., Kennedy, D., Baumert, M., and Kohler, M. (2013). Sleep spindle activity and cognitive performance in healthy children. *Sleep* 36, 237–243. doi: 10.5665/sleep.2380
- Colrain, I. M., and Baker, F. C. (2011a). Sleep EEG, the clearest window through which to view adolescent brain development. *Sleep* 34, 1287–1288. doi: 10.5665/SLEEP.1260
- Colrain, I. M., and Baker, F. C. (2011b). Changes in sleep as a function of adolescent development. *Neuropsychol. Rev.* 21, 5–21. doi: 10.1007/s11065-010-9155-5
- Conway, A. R. A., and Kovács, K. (2013). Individual differences in intelligence and working memory: a review of latent variable models. *Psychol. Learn. Motiv.* 58, 233–270. doi: 10.1016/B978-0-12-407237-4.00007-4
- Cox, R., Hofman, W. F., de Boer, M., and Talamini, L. M. (2014). Local sleep spindle modulations in relation to specific memory cues. *Neuroimage* 99, 103–110. doi: 10.1016/j.neuroimage.2014.05.028
- 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 Bellis, M. D., Kehavan, M. S., Beers, S. R., Hall, J., Frustaci, K., Masalehdan, A., et al. (2001). Sex differences in brain maturation during childhood and adolescence. *Cereb. Cortex* 11, 552–557. doi: 10.1093/cercor/11.6.552
- De Gennaro, L., and Ferrara, M. (2003). Sleep spindles: an overview. *Sleep Med. Rev.* 7, 423–440. doi: 10.1053/smr.2002.0252
- De Gennaro, L., Ferrara, M., Vecchio, F., Curcio, G., and Bertini, M. (2005). An electroencephalographic fingerprint of human sleep. *Neuroimage* 26, 114–122. doi: 10.1016/j.neuroimage.2005.01.020
- Duffy, F. H., Jones, K., Bartels, P., Albert, M., McNulty, G. B., and Als, H. (1990). Quantified neurophysiology with mapping: statistical inference, exploratory and confirmatory data analysis. *Brain Topogr.* 3, 3–12. doi: 10.1007/bf01128856
- Feinberg, I., and Campbell, I. G. (2010). Sleep EEG changes during adolescence: an index of a fundamental brain reorganization. *Brain Cogn.* 72, 56–65. doi: 10.1016/j.bandc.2009.09.008
- Feinberg, I., and Campbell, I. G. (2013). Longitudinal sleep EEG trajectories indicate complex patterns of adolescent brain maturation. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 304, R296–R303. doi: 10.1152/ajpregu.00422.2012
- Fogel, S. M., and Smith, C. T. (2011). The function of the sleep spindle: a physiological index of intelligence and a mechanism for sleep-dependent memory consolidation. *Neurosci. Biobehav. Rev.* 35, 1154–1165. doi: 10.1016/j.neubiorev.2010.12.003
- Geiger, A., Huber, R., Kurth, S., Ringli, M., Achermann, P., and Jenni, O. G. (2012). Sleep electroencephalography topography and children's intellectual ability. *Neuroreport* 23, 93–97. doi: 10.1097/WNR.0b013e32834e7e8f
- Geiger, A., Huber, R., Kurth, S., Ringli, M., Jenni, O. G., and Achermann, P. (2011). The sleep EEG as a marker of intellectual ability in school age children. *Sleep* 34, 181–189.
- Gogtay, N., and Thompson, P. M. (2010). Mapping gray matter development: implications for typical development and vulnerability to psychopathology. *Brain Cogn.* 72, 6–15. doi: 10.1016/j.bandc.2009.08.009
- Gray, J. R., and Thompson, P. M. (2004). Neurobiology of intelligence: science and ethics. *Nat. Rev. Neurosci.* 5, 471–482. doi: 10.1038/nrn1405
- Gruber, R., Wise, M. S., Frenette, S., Knäuper, B., Boom, A., Fontil, L., et al. (2013). The association between sleep spindles and IQ in healthy school-age children. *Int. J. Psychophysiol.* 89, 229–240. doi: 10.1016/j.ijpsycho.2013.03.018
- Gur, R. C., Turetsky, B. I., Matsui, M., Yan, M., Bilker, W., Huggett, P., et al. (1999). Sex differences in brain gray and white matter in healthy young adults: correlations with cognitive performance. *J. Neurosci.* 19, 4065–4072.
- Hagmann, P., Sporns, O., Madan, N., Cammoun, L., Pienaar, R., Wedeen, V. J., et al. (2010). White matter maturation reshapes structural connectivity in the late developing human brain. *Proc. Natl. Acad. Sci. U S A* 107, 19067–19072. doi: 10.1073/pnas.1009073107
- Haier, R. J., Jung, R. E., Yeo, R. A., Head, K., and Alkire, M. T. (2005). The neuroanatomy of general intelligence: sex matters. *Neuroimage* 25, 320–327. doi: 10.1016/j.neuroimage.2004.11.019
- Hassainia, F., Petit, D., and Montplaisir, J. (1994). Significance probability mapping: the final touch in t-statistic mapping. *Brain Topogr.* 7, 3–8. doi: 10.1007/bf01184832
- Ingallhalikar, M., Smith, A., Parker, D., Satterthwaite, T. D., Elliott, M. A., Ruparel, K., et al. (2014). Sex differences in the structural connectome of the human brain. *Proc. Natl. Acad. Sci. U S A* 111, 823–828. doi: 10.1073/pnas.1316909110
- Ishizuka, Y., Pollak, C. P., Shirakawa, S., Kakuma, T., Azumi, K., Usui, A., et al. (1994). Sleep spindle frequency changes during the menstrual cycle. *J. Sleep Res.* 3, 26–29. doi: 10.1111/j.1365-2869.1994.tb00100.x
- Jasper, H. H. (1958). Report of the committee on methods of clinical examination in electroencephalography. *Electroencephalogr. Clin. Neurophysiol.* 10, 370–375. doi: 10.1016/0013-4694(58)90053-1
- Jausovec, N., and Jausovec, K. (2005). Sex differences in brain activity related to general and emotional intelligence. *Brain Cogn.* 59, 277–286. doi: 10.1016/j.bandc.2005.08.001
- Jobert, M., Wilson, F. J., Roth, T., Ruigt, G. S. F., Anderer, P., Drinkenburg, W. H. I. M., et al. (2013). Guidelines for the recording and evaluation of pharmacological sleep studies in man: the International Pharmacology-EEG Society (IPEG). *Neuropsychobiology* 67, 127–167. doi: 10.1159/000343449
- Konrad, K., Firk, C., and Uhlhaas, P. J. (2013). Brain development during adolescence: neuroscientific insights into this developmental period. *Dtsch. Arztebl. Int.* 110, 425–431. doi: 10.3238/arztebl.2013.0425
- Lustenberger, C., Maric, A., Dürr, R., Achermann, P., and Huber, R. (2012). Triangular relationship between sleep spindle activity, general cognitive ability and the efficiency of declarative learning. *PLoS One* 7:e49561. doi: 10.1371/journal.pone.0049561
- Lüthi, A. (2014). Sleep spindles: where they come from, what they do. *Neuroscientist* 20, 243–256. doi: 10.1177/1073858413500854. [Epub ahead of print].

- Neubauer, A. C., Fink, A., and Schrausser, D. G. (2002). Intelligence and neural efficiency: the influence of task content and sex on the brain-IQ relationship. *Intelligence* 30, 515–536. doi: 10.1016/S0160-2896(02)00091-0
- Ohayon, M. M., Carskadon, M. A., Guilleminault, C., and Vitiello, M. V. (2004). Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. *Sleep* 27, 1255–1273.
- Peters, B. D., Szeszko, P. R., Radua, J., Ikuta, T., Gruner, P., DeRosse, P., et al. (2012). White matter development in adolescence: diffusion tensor imaging and meta-analytic results. *Schizophr. Bull.* 38, 1308–1317. doi: 10.1093/schbul/sbs054
- Piantoni, G., Poil, S. S., Linkenkaer-Hansen, K., Verweij, I. M., Ramautar, J. R., Van Someren, E. J., et al. (2013). Individual differences in white matter diffusion affect sleep oscillations. *J. Neurosci.* 33, 227–233. doi: 10.1523/JNEUROSCI.2030-12.2013
- Pivik, R. T., Broughton, R. J., Coppola, R., Davidson, R. J., Fox, N., and Nuwer, M. R. (1993). Guidelines for the recording and quantitative analysis of electroencephalographic activity in research context. *Psychophysiology* 30, 547–558. doi: 10.1111/j.1469-8986.1993.tb02081.x
- Ramsden, S., Richardson, F. M., Josse, G., Thomas, M. S., Ellis, C., Shakeshaft, C., et al. (2011). Verbal and non-verbal intelligence changes in the teenage brain. *Nature* 479, 113–116. doi: 10.1038/nature10514
- Raven, J. C., Court, J. H., and Raven, J. (1976). *Manual for Raven's Progressive Matrices*. London: H. K. Lewis.
- Raven, J. C., Raven, J., and Court, J. H. (2004). *Raven Haladó Progresszív Mátrixok Kézikönyv*. Budapest: Edge2000 Kft.
- Rechtschaffen, A., and Kales, A. (1968). *Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Participants*. Los Angeles: UCLA Brain Information Service/Brain Research Institute.
- Ringli, M., and Huber, R. (2011). Developmental aspects of sleep slow waves: linking sleep, brain maturation and behavior. *Prog. Brain Res.* 193, 63–82. doi: 10.1016/b978-0-444-53839-0.00005-3
- Rüger, B. (1978). Das maximale Signifikanzniveau des tests: "Lehne Ho ab, wennk untern gegebene Tests zur ablehnung führen". *Metrika* 25, 171–178. doi: 10.1007/bf02204362
- Schabus, M., Dang-Vu, T. T., Albouy, G., Baletau, 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
- Schabus, M., Hödlmoser, K., Gruber, G., Sauter, C., Anderer, P., Klösch, G., et al. (2006). Sleep spindle-related activity in the human EEG and its relation to general cognitive and learning abilities. *Eur. J. Neurosci.* 23, 1738–1746. doi: 10.1111/j.1460-9568.2006.04694.x
- Schabus, M., Hoedlmoser, K., Pecherstorfer, T., Anderer, P., Gruber, G., Parapatics, S., et al. (2008). Interindividual sleep spindle differences and their relation to learning-related enhancements. *Brain Res.* 1191, 127–135. doi: 10.1016/j.brainres.2007.10.106
- Scholle, S., Zwacka, G., and Scholle, H. C. (2007). Sleep spindle evolution from infancy to adolescence. *Clin. Neurophysiol.* 118, 1525–1531. doi: 10.1016/j.clinph.2007.03.007
- Shaw, P., Greenstein, D., Lerch, J., Clasen, L., Lenroot, R., Gogtay, N., et al. (2006). Intellectual ability and cortical development in children and adolescents. *Nature* 440, 676–679. doi: 10.1038/nature04513
- Shinomiya, S., Nagata, K., Takahashi, K., and Masumura, T. (1999). Development of sleep spindles in young children and adolescents. *Clin. Electroencephalogr.* 30, 39–43. doi: 10.1177/155005949903000203
- Simmonds, D. J., Hallquist, M. N., Asato, M., and Luna, B. (2014). Developmental stages and sex differences of white matter and behavioral development through adolescence: a longitudinal diffusion tensor imaging (DTI) study. *Neuroimage* 92, 356–368. doi: 10.1016/j.neuroimage.2013.12.044
- Tarokh, L., Van Reen, E., LeBourgeois, M., Seifer, R., and Carskadon, M. A. (2011). Sleep EEG provides evidence that cortical changes persist into late adolescence. *Sleep* 34, 1385–1393. doi: 10.5665/SLEEP.1284
- Timofeev, I., Grenier, F., Bazhenov, M., Houweling, A. R., Sejnowski, T. J., and Steriade, M. (2002). Short- and medium-term plasticity associated with augmenting responses in cortical slabs and spindles in intact cortex of cats in vivo. *J. Physiol.* 542, 583–598. doi: 10.1111/jphysiol.2001.013479
- Uhlhaas, P. J., Roux, F., Singer, W., Haenschel, C., Sireteanu, R., and Rodriguez, E. (2009). The development of neural synchrony reflects late maturation and restructuring of functional networks in humans. *Proc. Natl. Acad. Sci. U S A* 106, 9866–9871. doi: 10.1073/pnas.0900390106
- Uhlhaas, P. J., and Singer, W. (2011). The development of neural synchrony and large-scale cortical networks during adolescence: relevance for the pathophysiology of schizophrenia and neurodevelopmental hypothesis. *Schizophr. Bull.* 37, 514–523. doi: 10.1093/schbul/sbr034
- Waller, H. R., and Lin, A. L. (2012). Sexual dimorphisms in spatial memory, facial recognition and reaction time. *J. Stud. Res.* 1, 41–48.
- Zaidi, Z. F. (2010). Gender differences in human brain: a review. *Open Anat. J.* 2, 37–55. doi: 10.2174/1877609401002010037

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: 15 September 2014; accepted: 08 November 2014; published online: 28 November 2014.

Citation: Bódizs R, Gombos F, Ujma PP and Kovács I (2014) Sleep spindling and fluid intelligence across adolescent development: sex matters. *Front. Hum. Neurosci.* 8:952. doi: 10.3389/fnhum.2014.00952

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

Copyright © 2014 Bódizs, Gombos, Ujma and Kovács. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and 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.

Sleep spindle alterations in patients with Parkinson's disease

Julie A. E. Christensen^{1,2,3*}, Miki Nikolic², Simon C. Warby⁴, Henriette Koch^{1,2,3}, Marielle Zoetmulder^{2,5}, Rune Frandsen², Keivan K. Moghadam⁶, Helge B. D. Sorensen¹, Emmanuel Mignot³ and Poul J. Jennum^{2,7}

¹ Biomedical Engineering, Department of Electrical Engineering, Technical University of Denmark, Kongens Lyngby, Denmark, ² Danish Center for Sleep Medicine, Department of Clinical Neurophysiology, Glostrup University Hospital, Glostrup, Denmark, ³ Stanford Center for Sleep Sciences and Medicine, Psychiatry and Behavioral Sciences, Stanford University, Palo Alto, CA, USA, ⁴ Center for Advanced Research in Sleep Medicine, Sacré-Coeur Hospital of Montréal, University of Montréal, Montréal, QC, Canada, ⁵ Department of Neurology, Bispebjerg Hospital, Copenhagen, Denmark, ⁶ Department of Biomedical and Neuromotor Sciences (DIBINEM), University of Bologna, Bologna, Italy, ⁷ Center for Healthy Ageing, University of Copenhagen, Copenhagen, Denmark

OPEN ACCESS

Edited by:

Christian O'Reilly,
McGill University, Canada

Reviewed by:

Ki-Young Jung,
Korea University
Medical Center, South Korea
Géraldine Rauchs,
Institut National de la Santé et de la
Recherche Médicale, France
Veronique Latreille,
Hopital du Sacré-Coeur
de Montréal, Canada

*Correspondence:

Julie A. E. Christensen,
Technical University of Denmark,
Orsted's Plads, Building 349, DK-2800
Kongens Lyngby, Denmark
julie.a.e.christensen@gmail.com

Received: 16 December 2014

Accepted: 11 April 2015

Published: 01 May 2015

Citation:

Christensen JAE, Nikolic M, Warby SC, Koch H, Zoetmulder M, Frandsen R, Moghadam KK, Sorensen HBD, Mignot E and Jennum PJ (2015) Sleep spindle alterations in patients with Parkinson's disease. *Front. Hum. Neurosci.* 9:233. doi: 10.3389/fnhum.2015.00233

The aim of this study was to identify changes of sleep spindles (SS) in the EEG of patients with Parkinson's disease (PD). Five sleep experts manually identified SS at a central scalp location (C3-A2) in 15 PD and 15 age- and sex-matched control subjects. Each SS was given a confidence score, and by using a group consensus rule, 901 SS were identified and characterized by their (1) duration, (2) oscillation frequency, (3) maximum peak-to-peak amplitude, (4) percent-to-peak amplitude, and (5) density. Between-group comparisons were made for all SS characteristics computed, and significant changes for PD patients vs. control subjects were found for duration, oscillation frequency, maximum peak-to-peak amplitude and density. Specifically, SS density was lower, duration was longer, oscillation frequency slower and maximum peak-to-peak amplitude higher in patients vs. controls. We also computed inter-expert reliability in SS scoring and found a significantly lower reliability in scoring definite SS in patients when compared to controls. How neurodegeneration in PD could influence SS characteristics is discussed. We also note that the SS morphological changes observed here may affect automatic detection of SS in patients with PD or other neurodegenerative disorders (NDDs).

Keywords: Parkinson's disease, sleep spindle morphology, EEG, neurodegeneration, biomarker

Introduction

Parkinson's disease (PD) is a neurodegenerative disorder (NDD) characterized primarily by motor symptoms, including bradykinesia, rigidity, postural instability, and tremor. Although the disease process in PD is not restricted to a specific brain area, these symptoms are mostly caused by the loss of dopaminergic neurons in the substantia nigra pars compacta resulting in a reduction or depletion of dopamine (Galvin et al., 2001). Lewy body aggregations of alpha-synuclein in the brain are a central feature of PD pathology (Galvin et al., 2001). These inclusions typically start in caudal areas of the brain and progress anteriorly (Braak et al., 2003), and may take place years prior to involvement of the substantia nigra and associated development of motor symptoms.

Abbreviations: AASM, American Academy of Sleep Medicine; EEG, electroencephalography; iRBD, idiopathic REM sleep behavior disorder; MSA, Multiple System Atrophy; NDD, Neurodegenerative disorders; PD, Parkinson's disease; PSG, polysomnographic; REM, Rapid eye movements; SS, Sleep spindles.

Specifically, Braak et al.'s PD staging is based on Lewy-body distribution, which rise from the dorsal motor nucleus of the vague nerve in the medulla and in the olfactory bulb (stage 1) emerging through the subceruleus-ceruleus complex and the magnocellularis reticular nucleus (stage 2), the substantia nigra, the pedunculopontine nucleus and the amygdala (stage 3), the temporal mesocortex (stage 4), and finally reaching the neocortex (stage 5 and 6). Stage 1 and 2 were considered as pre-Parkinsonian states, stage 3 and 4 as Parkinsonian states and 5 and 6 as late-Parkinsonian states (Braak et al., 2003).

In addition to the motor manifestations that define PD, non-motor symptoms such as sleep problems, depression, dementia and attention deficit (Chaudhuri et al., 2011, 2006), autonomic symptoms as abnormal heart rate variability (Sorensen et al., 2012, 2013) and gastrointestinal symptoms such as nausea and constipation (Garcia-Ruiz et al., 2014) are all well known in patients with PD. Stating the presence of at least two of the four motor symptoms resting tremor, bradykinesia, rigidity, and postural imbalance typically makes the clinical diagnosis of PD, although it has been indicated that the pathological changes in the striatal dopaminergic system develop several years before the clinical appearance of PD. Further development of the pathology may result in Lewy Body Dementia.

Twenty years ago, it was discovered that idiopathic rapid eye movement (REM) sleep behavior disorder (iRBD) is closely related to Parkinsonism (Schenck et al., 1996, 2013a; Salawu et al., 2010). Indeed, the presence of iRBD, even without the presence of motor or cognitive complaints, confers a significant risk of conversion into synucleinopathies including PD (Iranzo, 2011; Schenck et al., 2013b). The diagnosis of RBD requires complaints or an anamnesis describing dream enactment behaviors as well as a manifestation of REM sleep without atonia (RWA) as measured by polysomnography (PSG) (Stevens and Comella, 2013; American Academy of Sleep Medicine, 2014). The idiopathic form of RBD (iRBD) is diagnosed when no concurrent neurological disease is found, and International classification of Sleep Disorders criteria for RBD are met (Stevens and Comella, 2013; American Academy of Sleep Medicine, 2014). Specifically, measures of RWA (Postuma et al., 2010; Kempfner et al., 2013), slow wave characteristics (Latreille et al., 2011), sleep stability and differences in electroencephalographic (EEG) or electrooculographic micro- and macro-sleep patterns have been investigated in patients with iRBD and/or PD (Christensen et al., 2012, 2013, 2014b).

Reduced sleep spindle (SS) density and activity have been identified in patients with PD and iRBD (Puca et al., 1973; Myslobodsky et al., 1982; Emser et al., 1988; Comella et al., 1993; Christensen et al., 2014a; Latreille et al., 2015). SS are generated by a complex interaction involving thalamic, limbic, and cortical areas. A di-synaptic circuit between thalamic reticular neurons and thalamocortical relay cells, both located in the thalamus, can spontaneously generate spindle-like oscillations, which are conveyed to the cortex by the axons of the thalamocortical relay cells. These cells receive feedback from cortical pyramidal cells as well as input from pre-thalamic fibers originating from the brainstem and posterior hypothalamus (Steriade et al., 1993; Steriade and Timofeev, 2003). As such the thalamus holds a

primary role in generating and controlling SS. SS have been reported to have a gating role with regard to the flow of thalamic sensory input, and thus may have a sleep-preserving role (De Gennaro and Ferrara, 2003). Also, several studies have reported SS to have an important role in memory consolidation, synaptic plasticity and cognition (Steriade and Timofeev, 2003; Schabus et al., 2006; Fogel and Smith, 2011; Fogel et al., 2012; Latreille et al., 2015). The formation of SS begins in the infant brain (De Gennaro and Ferrara, 2003), but SS characteristics such as density and amplitude change with age (Nicolas et al., 2001; De Gennaro and Ferrara, 2003), suggesting that SS play an important role in normal cognitive functioning.

Although a reduction in SS density is not specific to PD, SS and other EEG features may be potential useful as biomarkers of disease progression or therapeutic efficacy in PD and other NDDs (Nguyen et al., 2010; Leiser et al., 2011; Micanovic and Pal, 2014). However, the identification of SS is a difficult task; studies assessing inter-scorer variance in normal sleep have shown significant variance in SS identification, both between human experts and between automated SS detectors (Warby et al., 2014; Wendt et al., 2014). SS identification and characterization in pathological sleep is not well studied, but previous evidence suggests that SS may have different characteristics in PD patients (Latreille et al., 2015), and therefore may interfere with traditional sleep staging in patients (Comella et al., 1993; Jensen et al., 2010; Christensen et al., 2014b; Koch et al., 2014).

In this study, we aimed to identify changes in SS density and specific morphological characteristics of SS in patients with PD. Since five sleep experts identified SS independently, we were also able to assess inter-expert variation of SS identification in EEG of patients and controls. By identifying specific changes in SS characteristics, we aimed to better understand the mechanism and to what extent the neurodegenerative progress influences SS characteristics, also identifying specific spindle features that may be useful as prognostic biomarkers of disease. A secondary aim was to help guide the specialized development of automatic SS detectors to be used on EEG from patients with NDDs.

Materials and Methods

Subjects and Recordings

Polysomnographic (PSG) EEG data from 15 patients with PD and 15 sex- and age-matched control subjects with no history of movement disorder, dream-enacting behavior or other previously diagnosed sleep disorders were included in this study. The subjects were all recruited from the Danish Center for Sleep Medicine (DCSM) in the Department of Clinical Neurophysiology, Glostrup University Hospital in Denmark. All patients were evaluated by a movement specialist with a comprehensive medical and medication history and a PSG analyzed according to the American Academy of Sleep Medicine (AASM) standard (Iber et al., 2007). The diagnostic certainty for PD at Danish neurological departments has been reported to be 82% (Wermuth et al., 2012). None of the PD patients had dementia at inclusion, but one of the patients with PD later developed Multiple System Atrophy (MSA), indicated as the Parkinsonian type (MSA-P) as the patient had predominating

PD-like symptoms. Subjects were excluded from the study if they were taking medications known to effect sleep (antidepressants, antipsychotics, hypnotics). However, dopaminergic treatments were permitted despite their potential effect on vigilance and SS characteristics (Puca et al., 1973; Micalef et al., 2009). In addition to ethical concerns regarding discontinuing dopaminergic treatment in these subjects, we wanted to avoid deleterious discontinuation effects on the PSG, as well as unpleasant and negative motor effects that could interfere with the study. The quality of each PSG recording was individually examined, and recordings with disconnections or significant amounts of signal artifact were not included. Demographic data and PSG variables for the two groups are seen in **Table 1**.

Manual Labeling of Sleep Spindles

For each subject, eight blocks of five consecutive epochs of non-REM sleep stage 2 (N2) of 30-s duration were selected randomly from the PSG recording in between lights off and lights on. The blocks were randomly chosen and ranked by use of Matlab's *randsample*-function. One-by-one and in the prioritized order, the blocks were visually checked for major movements or other contaminating artifacts. The first eight artifact-free blocks were chosen as the ones to be scored for SS. A total of five independent sleep experts identified SS in these blocks, where only the C3-A2 EEG derivation was visible. The signals were filtered with a notch filter at 50 Hz and a band-pass filter with cutoff frequencies at 0.3 Hz and 35 Hz, as indicated by AASM standards (Iber et al., 2007). All analyzed signals had a sampling frequency of 256 Hz. The experts assigned a confidence score to each identified spindle, to indicate the amount of confidence in the identification (as described previously in Warby et al., 2014). In this way, each SS was given a confidence weighting of 1 for “definitely SS,” 0.75 for “probably a SS” and 0.5 for “maybe a SS.”

The scoring procedure was performed in a Matlab-based software program “EEG viewer” developed by MN at DCSM. The program mimics a standard sleep scoring program in a clinical

setting, and includes the standard features so the experts have the same opportunities to view and navigate the PSG data as they are used to when analyzing sleep in the clinic. The program ensures that if an epoch to be scored does not have any marked SS, the expert is required to click a box saying “no spindles in current epoch.” This ensures that the total of 40 epochs of N2 sleep per subject was analyzed by each expert. The experts were blinded for which group the subjects belong to.

The final SS identifications used for morphology measures were defined using the group consensus rule described in Warby et al. (2014). Spindle identifications from five different experts with weighted confidence scores for each SS were averaged at each sample point and aggregated into a single consensus. Sample points that had an average score of higher than the group consensus threshold $T_{gc} = 0.25$ were included in the final group consensus, and the morphology measures were computed on these group consensus SS. It was decided to use $T_{gc} = 0.25$ as this was found to be the best in Warby et al. (2014).

Spindle Characteristics and between Group Comparisons

The morphology of the identified SS was characterized by their (1) duration, (2) oscillation frequency, (3) maximum peak-to-peak amplitude, (4) percent-to-peak amplitude, and (5) SS density per minute; all of which are well-evaluated elsewhere (Warby et al., 2014). The morphology measures were all computed using Matlab 2013b. Before any of the measures were computed, the central EEG signal was filtered forward and reverse with (1) a notching filter with the notch at 50 Hz and a bandwidth of 50/35 Hz (at -3 dB) and (2) a 4th order Butterworth band-pass filter with cut off frequencies (-3 dB) at 0.3 Hz and 35 Hz.

For each SS the duration was computed in seconds as

$$dur = \frac{\# \text{ samples}}{f_s},$$

where $f_s = 256$ Hz is the sampling frequency and $\# \text{ samples}$ defines the number of samples. The samples were consecutive and obeyed the consensus rule. The oscillation frequency was defined in Hz and was for each SS estimated as

$$f_{osc} = \frac{K}{2 \cdot dur},$$

where K defines the total number of extrema points detected using Matlab's *findpeaks*-function applied on a 5-point moving average smoothed version of the SS signal and with a minimum peak-to-peak distance of 11 samples. The maximum points were found by applying the *findpeaks*-function directly, and the minima points were found by applying the function on the flipped signal, and the total number of extrema points was set as the sum of the two. These settings were chosen, as they were considered best for estimating the f_{osc} when visually investigating numerous randomly selected examples of SS. The maximum peak-to-peak amplitude was for each SS estimated as

$$A_{p2p} = \max(|A_e(k+1) - A_e(k)|), k = 1, 2, \dots, K-1,$$

TABLE 1 | Demographic and PSG data for the two groups studied.

Characteristics	PD patients	Controls	P
Total counts (Male/Female)	15 (7/8)	15 (7/8)	–
Age (Years)	62.7 ± 5.8	62.9 ± 5.9	0.90
BMI (kg/m ²)	25.3 ± 3.5	22.1 ± 2.5	0.02
Disease duration (years)	6.7 ± 4.5	NA	–
Hoehn and Yahr stage	2.0 ± 1.2	NA	–
UPDRS part III “on”	20.9 ± 7.0	NA	–
ACE	90.2 ± 4.8	NA	–
Levodopa equivalent dosage (mg)	621.1 ± 301.5	NA	–
Levodopa use [n (%)]	10 (67)	NA	–
Dopamine agonist use [n (%)]	14 (93)	NA	–
Sleep efficiency (%)	79.7 ± 14.1	87.1 ± 8.4	0.09
Time in bed (min)	448.1 ± 82.0	499.6 ± 63.7	0.07
LM index (number/hour)	31.8 ± 34.8	30.4 ± 35.3	0.91

BMI, Body Mass Index; UPDRS, Unified Parkinson's disease rating scale; ACE, Addenbrooke's cognitive examination; LM, Leg movements.

where A_e is a vector holding the amplitude values for each of the K detected extrema points. To investigate the influence on SS from K-complexes or delta waves, the maximum peak-to-peak amplitude was estimated twice for each SS; once without any further frequency filtering of the data, and once where the data was forward and reverse filtered with a 10th order high-pass filter with cut off frequency (-3 dB) at 4 Hz to remove low frequency, high amplitude waves that may interfere with the peak-to-peak calculation. The percent-to-peak amplitude gives a simple measure between 0 and 1 of the symmetry of the spindle and it was computed for each SS as

$$Sym = \frac{\# \text{ samples before point of } A_{p2p}}{\# \text{ samples}},$$

where the point of A_{p2p} is defined as the point between the maxima and minima delineating A_{p2p} . Finally, the density was computed for each subject as the number of SS per minute of investigated data, described as

$$Density = \frac{2 \cdot \# \text{ SS}}{\# \text{ epochs reviewed}}.$$

The morphology measures were computed for the SS identifications for each expert, as well as for the spindles included in the group consensus. For the SS included in the group consensus, a minimum duration threshold $dur_{th} = 0.2$ s was used, and resulted in the exclusion of only three spindles. This threshold is less than the minimum duration stated by the AASM scoring (0.5 s). However, others have shown that apparent spindles <0.5 s are clearly recognizable by sleep experts, and have similar characteristics to spindles >0.5 s (Warby et al., 2014). We used a minimum duration threshold of 0.2 s because we wanted to determine whether PD patients and controls have specific differences in these shorter spindles. When computing the measures for the SS identifications for each expert, all the SS were included, regardless of their confidence score and duration. Two-sided Wilcoxon rank sum tests with a significance level of $\alpha = 0.05$ were used for each of the measures to test for significant differences between the two groups.

Inter-Expert Reliability When Scoring SS

Inter-expert reliability measures were computed for each of the 10 available expert-pairs. True positives (TP) define the number of samples where both experts have marked SS, true negatives (TN) define the number of samples where both experts have not marked SS, false positives (FP) define the number of samples where the reference-expert has not marked SS, and the other expert has and false negatives (FN) define the number of samples where the reference-expert has marked SS, but the other expert has not. For each comparison, the reliability measures were indicated as the F_1 -score and the Cohen's Kappa coefficient (κ). The F_1 -score is the harmonic mean of precision (P) and recall (R) and reaches its best value at 1 (perfect agreement) and the worst

at 0 (no agreement). It is computed as

$$F_1\text{-score} = \frac{2 \cdot R \cdot P}{R + P}, \text{ where}$$

$$R = \frac{TP}{TP + FN} \text{ and } P = \frac{TP}{TP + FP}.$$

The κ is often used to measure inter-annotator reliability as it takes the agreement occurring by chance into account. It reached its best value at 1 (perfect agreement) and worst at -1 (no agreement). It reaches 0 when accuracy is equal to what is expected by chance. It is computed as

$$\kappa = \frac{\frac{TP + TN}{N} - Pr}{1 - Pr}, \text{ where}$$

$$Pr = \frac{TP + FN}{N} \cdot \frac{TP + FP}{N} + \left(1 - \frac{TP + FN}{N}\right) \cdot \left(1 - \frac{TP + FP}{N}\right),$$

where $N = TP + TN + FP + FN$ defines the total number of samples reviewed. The relative strength of agreement associated with κ can be described by the labels "poor" ($\kappa < 0.00$), "slight" ($0.00 \leq \kappa \leq 0.20$), "fair" ($0.21 \leq \kappa \leq 0.40$), "moderate" ($0.41 \leq \kappa \leq 0.60$), "substantial" ($0.61 \leq \kappa \leq 0.80$) and "almost perfect" ($0.81 \leq \kappa \leq 1.00$) (Landis and Koch, 1977). The F_1 -score and κ are symmetric regarding false detections and will therefore both yield the same regardless of which expert were used as the reference.

Results

For the SS included in the group consensus, it was found that patients with PD show SS that are significantly different from controls in terms of duration, oscillation frequency and max peak-to-peak amplitude. Additionally, patients with PD have significantly different SS density compared to controls. Specifically, it was found that patients with PD have decreased SS density ($-38.17\%/ -0.71$ SS/min), and that their SS are longer ($+11.69\%/ +0.09$ s), have a lower frequency ($-2.27\%/ -0.29$ Hz) and higher max peak-to-peak amplitude ($+19.61\%/ 9.45$ μ V) compared to controls (Table 2). No significant differences were identified for the symmetry measure. The maximum peak-to-peak amplitude estimated after removal of frequencies below 4 Hz was still significantly different between groups. Of note, patients with PD still showed a higher max peak-to-peak amplitude ($+20.95\%/ 9.49$ μ V) compared to controls. The five SS morphology measures are illustrated in Figure 1. From left to right, the eight first ID numbers in both groups are females ranging from the youngest to the oldest. The last seven IDs in both groups are males, also ranging from the youngest to the oldest. One of the patients later developed MSA and is illustrated with black.

The patients had significantly fewer spindles than the controls (p -value < 0.05). Ten patients and only four controls had less than 10 SS in the 40 epochs of N2 sleep that were assessed; four

TABLE 2 | Mean (μ) and standard deviation (σ) for the spindle characteristics found for each of the experts' identifications as well as for the spindles in the group consensus.

Spindle characteristic	Expert 1 (947 SS)		Expert 2 (752 SS)		Expert 3 (952 SS)		Expert 4 (282 SS)		Expert 5 (2135 SS)		Group consensus (901 SS)		P
	PD	C	PD	C	PD	C	PD	C	PD	C	PD	C	
Duration [sec, $\mu \pm \sigma$]	0.93 \pm 0.44	0.84 \pm 0.41	0.66 \pm 0.29	0.67 \pm 0.29	0.74 \pm 0.29	0.68 \pm 0.27	0.88 \pm 0.20	0.77 \pm 0.24	1.19 \pm 0.52	1.12 \pm 0.51	0.86 \pm 0.35	0.77 \pm 0.36	<0.002 ^{a,c,d,e,GC}
Frequency [Hz, $\mu \pm \sigma$]	12.38 \pm 1.27	12.69 \pm 1.27	12.92 \pm 1.24	13.07 \pm 1.11	12.45 \pm 1.22	12.62 \pm 1.34	12.73 \pm 1.14	13.13 \pm 1.00	11.69 \pm 1.24	12.03 \pm 1.28	12.51 \pm 1.21	12.80 \pm 1.23	<0.02 ^{a,b,d,e,GC}
Max peak-to-peak amplitude [μ V, $\mu \pm \sigma$]	57.37 \pm 17.23	48.26 \pm 15.37	56.96 \pm 18.14	46.88 \pm 15.96	57.75 \pm 17.15	48.52 \pm 15.47	64.60 \pm 16.68	49.95 \pm 14.04	51.44 \pm 18.14	45.02 \pm 15.73	57.64 \pm 17.34	48.19 \pm 15.55	<0.001 ^{a,b,c,d,e,GC}
Max peak-to-peak amplitude, after removal of frequencies < 4 Hz [μ V, $\mu \pm \sigma$]	53.87 \pm 15.99	45.24 \pm 14.17	54.38 \pm 16.85	44.16 \pm 14.89	54.89 \pm 16.34	45.43 \pm 14.18	62.40 \pm 16.64	47.51 \pm 13.13	46.20 \pm 16.62	40.15 \pm 13.92	54.78 \pm 16.24	45.29 \pm 14.41	<0.001 ^{a,b,c,d,f,GC}
Percent-to-peak amplitude [$\mu \pm \sigma$]	0.49 \pm 0.23	0.47 \pm 0.24	0.46 \pm 0.23	0.46 \pm 0.23	0.46 \pm 0.24	0.46 \pm 0.23	0.46 \pm 0.22	0.46 \pm 0.21	0.45 \pm 0.25	0.45 \pm 0.25	0.47 \pm 0.23	0.46 \pm 0.23	NS
Density [per min, $\mu \pm \sigma$]	1.29 \pm 2.44	1.87 \pm 1.56	0.91 \pm 1.36	1.60 \pm 1.27	1.16 \pm 1.95	2.01 \pm 1.82	0.30 \pm 0.51	0.64 \pm 0.84	2.91 \pm 2.52	4.21 \pm 2.14	1.15 \pm 2.06	1.86 \pm 1.57	<0.05 ^{a,b,c,GC}

Wilcoxon rank sum tests were used to test for significance between patients with Parkinson's disease (PD) and control subjects (C). ^asignificant for expert 1, ^bsignificant for expert 2, ^csignificant for expert 3, ^dsignificant for expert 4, ^esignificant for expert 5, ^fsignificant for group consensus.

patients and 0 controls had no SS. Only 3 patients compared to 10 controls had more than 20 SS in the group consensus.

As a supplementary check, the significance tests were performed on SS identifications from each of the five experts individually. The maximum peak-to-peak amplitude was, for all five experts, both before and after removal of frequencies below 4 Hz, significantly different in patients with PD compared to controls. The duration and oscillation frequency were also significantly different between the two groups for 4/5 of the experts, and density significantly different between the two groups for 3/5 of the experts. The mean and standard deviations of the SS morphology measures and the results from the significance tests are summarized in Table 2.

Figure 2 illustrates the relation between the SS measures and disease duration for the patients, and Figure 3 illustrates the relation between the SS measures and Addenbrooke's Cognitive Examination (ACE) score for the patients. Note that the x-axes are not continuous, but denote disease duration in years (Figure 2) and ACE score (Figure 3) for 15/15 and 13/15 of the patients, respectively. The three subjects with highest SS density are all females, and the one with the highest SS density is a patient with PD later diagnosed with MSA-P (indicated as PD+MSA in the figures). She is illustrated with black in Figures 1, 2, 3. No clear visual tendency between SS characteristics and disease duration or ACE score was seen for any of the measures. Supplementary Figure 1 illustrates the relation between SS measures and Hoehn and Yahr (H and Y) stage and Supplementary Figure 2 illustrates the relation between SS measures and the Unified Parkinson's Disease Rating Scale (UPDRS) Part III. No clear visual trends were seen.

Considering that the outlier PD patient with a very high spindle density (highest of all subjects in the study) later developed MSA, we reanalyzed the SS included in the group consensus when results from this outlier patient were left out, and found the same measures to be as significant different between the groups. Specifically, patients now have an even bigger decrease in SS density ($-61.29\%/ -1.14$ SS/min), a longer SS duration ($+11.69\%/ +0.09$ s), a slower frequency ($-4.14\%/ -0.53$ Hz) and a higher max peak-to-peak amplitude, both before ($+16.93\%/ 8.16$ μ V) and after ($+17.95\%/ 8.13$ μ V) removal of low frequencies when compared to controls. The results for this analysis are summarized in Table 3.

Figure 4 shows scatterplots for the individual SS, where the maximum peak-to-peak amplitude (before removal of low frequencies) defines the y-axis and the oscillation frequency and duration defines the x-axis, respectfully. Linear trend lines are added on top of the scatterplots in order to see differences between groups. We found a trend of a positive correlation between the duration and maximum peak-to-peak amplitude. Interestingly, SS from patients showed this tendency to a lesser degree (slope of $+11.74$ μ V/s) compared to SS from controls (slope of $+18.09$ μ V/s). Also, we found a negative correlation of oscillation frequency and maximum peak-to-peak amplitude, and found this tendency to be less apparent for SS from patients (slope of -1.02 μ V/Hz) compared to SS from controls (slope of -4.10 μ V/Hz).

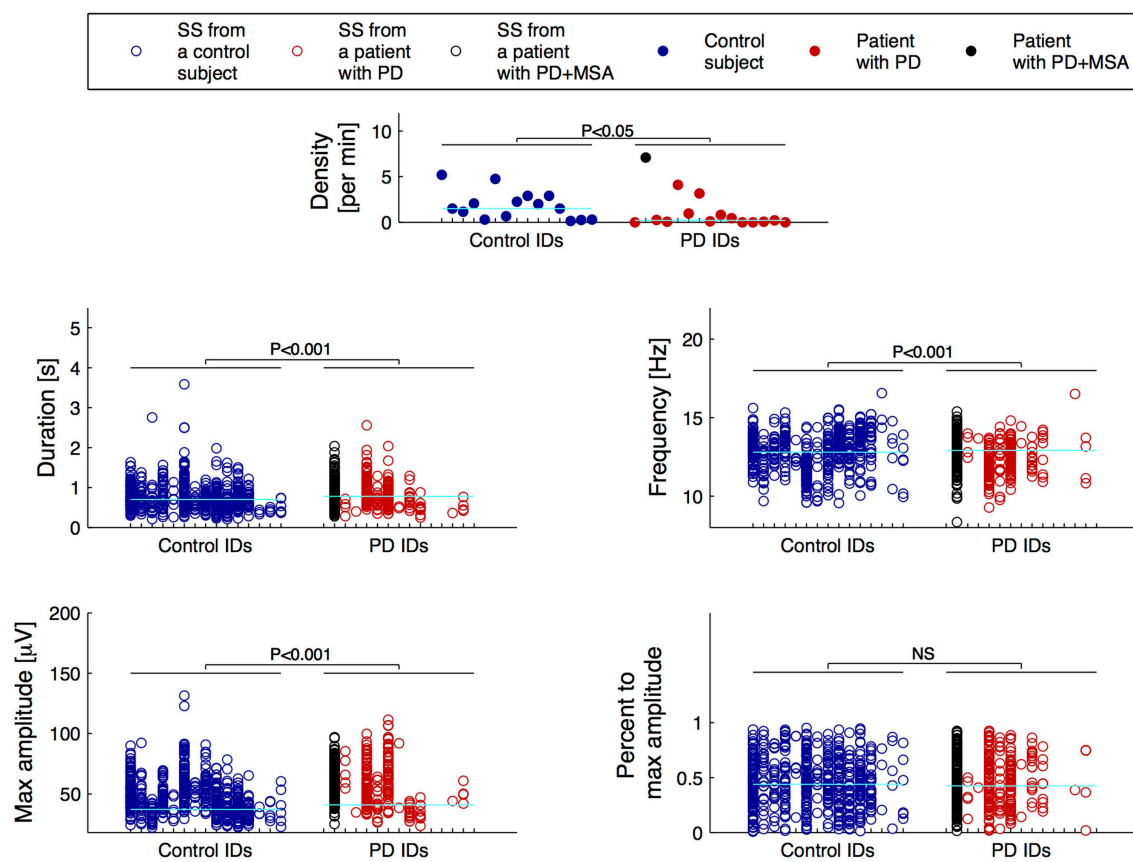


FIGURE 1 | Distributions of the morphology measures for the spindles included in the group consensus. From left to right, the first eight IDs in both groups are females ranging from the youngest to the oldest, and the following seven IDs are males also ranging

from the youngest to the oldest. One patient with Parkinson's disease (PD) later developed Multiple System Atrophy (MSA) and is indicated with black. The cyan horizontal lines indicate the group median for each of the measures.

Table 4 summarizes the fraction of SS included in the group consensus that do not strictly pass AASM criteria for a spindle (11–16 Hz, 0.5–3.0 s). Overall, 25.3% of the SS identified by experts and included in the group consensus did not meet AASM criteria. Most of these “abnormal” SS would have been excluded because their duration is too short (16.9%) or have an oscillation frequency that is too slow (9.7%).

In order to determine if there was a difference between PD and controls in the frequency of “abnormal” spindles not meeting AASM criteria, we compared the groups. All 15/15 control subjects had SS, whereas only 11/15 patients with PD had some SS. It was found that control subjects show significantly more “abnormal” spindles not meeting AASM criteria, i.e., more spindles with a too short duration compared to patients with PD (**Table 4**). No significant difference was however found between groups when the outlier patient with PD + MSA was left out of the analysis.

When computing the SS characteristic based on AASM criteria, the same SS characteristics were found to be significantly different between PD patients and controls (**Table 5**). Analysis of these SS showed that patients with PD have a decreased density ($-32.84\%/-0.44$ SS/min), and their SS are longer

($+9.41\%/+0.08$ s), have a lower frequency ($-2.69\%/-0.35$ Hz) and higher max peak-to-peak amplitude before removal of low frequencies ($+21.34\%/+10.37$ μ V) and after ($+22.51\%/+10.30$) compared to controls. These differences are similar to those found based on all SS in the group consensus.

Table 6 summarizes inter-expert reliabilities of SS scoring, where the SS are grouped according to their confidence score. The mean inter-expert reliability of scoring “definite SS” computed by κ was found to be significant lower for patients compared to controls. Although not significant, a trend for a lower κ was found for “probable/definite SS” in patients compared to controls ($P = 0.054$). In all cases, the inter-expert reliability is lower for scoring SS in patients compared to controls.

Discussion

Based on a group consensus of manually scored SS from five independent sleep experts, this study investigates morphological changes of SS in a central EEG lead of patients with PD compared to age- and sex-matched control subjects. The main findings of this study are that patients with PD have a decreased SS density, and that their SS have a longer duration, a slower oscillation

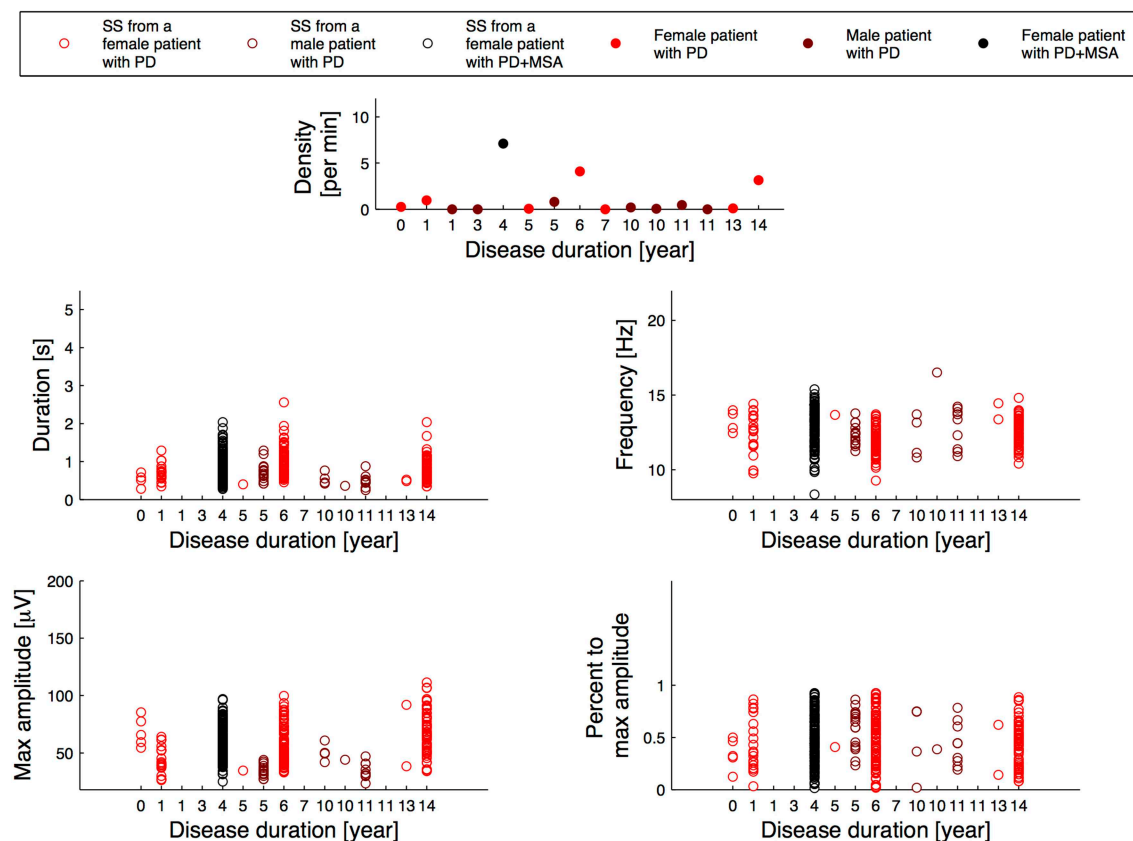


FIGURE 2 | Distribution of the morphology measures for the spindles from patients with Parkinson's disease (PD), where the patients are sorted according to their disease duration. PD+MSA indicates a patient with PD, that later developed Multiple System Atrophy (MSA).

frequency and higher maximum peak-to-peak amplitude. These results suggest that not only SS density but also specific morphological changes in SS have potential clinical utility when diagnosing PD. Further, the data suggests that the disease process affect directly or indirectly the brain regions responsible for the generation of SS. Future studies including more subtypes of PD and NDDs in general are however needed to investigate whether the specific morphological changes in SS can be used to differentiate different PD subtypes as well as different NDDs.

The results illustrate the fact that there are fewer SS in patients with PD, and that the few that are remaining are more pronounced when compared to those seen in controls. There could be several explanations for this. First, patients with PD have a more “blurred” EEG in general with either a lack of or an abnormal mixture of micro- and macro-sleep structures (Petit et al., 2004; Christensen et al., 2014b). This pattern may make it more difficult to identify distinct SS, as they would be buried within other undefined EEG microstructural changes. In this case, only the obvious SS would rise over background and be marked. Second, it could be that the neurodegenerative process has affected the thalamic neurons responsible for generating and controlling SS in such a way, that SS are only generated when very strong signals from pre-thalamic fibers reaches the thalamus resulting in more pronounced SS. Third, we cannot

rule out that these SS changes could be the result of treatment with dopaminergic agents affecting the morphology of SS, although a previous report suggests that these drugs should increase spindle density (Puca et al., 1973), which is not what we observed.

It was found that patients with PD have a lower SS density compared to age and sex-matched controls. This finding is consistent with our and other groups' prior findings (Emser et al., 1988; Christensen et al., 2014a; Latreille et al., 2015), but contradicts those of other studies (Happe et al., 2004). According to Braak et al. (2003), the neurodegenerative progress in PD shows a progressive ascending course starting from the brain stem and spreading to additional brain structures. At some point, the neurodegeneration may affect or destroy the SS generator of the thalamus, resulting in fewer or no spindles. Interestingly, (Roth et al., 2000) found that medial thalamotomy abolishes spindle activity in N2 sleep systematically, but that pallido-thalamic tractotomy attenuate spindle activity only to a varying degree, with spindles reemerging after 3 months. It is therefore likely that neurodegenerative involvement of prethalamic fibers from the brain stem may affect spindle activity to a certain degree. In **Figure 1**, it is apparent that for four of the patients, no SS are included in the group consensus, and that for six other patients, less than 10 spindles were identified.

Surprisingly, a PD patient showing an abnormally high SS density was later diagnosed with MSA-P. Although only a single case, it is an interesting finding which support the hypothesis that spindles can be used as a marker of diagnostic subgroups of PD. Latreille et al. (2015) reported a decline in SS activity paralleling cognitive decline in patients with PD, suggesting that SS activity could be used as an early marker of Dementia. The number of patients included in present study is, however, too small to perform further subgroup analysis. Additionally, in both groups, younger subjects and females trend in showing slightly higher spindle densities when compared to older and male subjects. The three oldest male control subjects have negligible SS densities. These observations suggest that reduced SS density is not specific for PD, in agreement with the fact that many conditions such as cognitive function, memory consolidation, pharmacological interventions and pre-PSG conditions have been reported to influence SS density (De Gennaro and Ferrara, 2003; Caporro et al., 2012). Further analysis including more PD and iRBD patients, together with a more in-depth investigation of cognitive decline and disease severity would be needed to evaluate the relation of abnormalities in SS development in the disease process, and the use of SS as a prognostic marker. Additionally, SS density has also been reported decreased for other conditions such as Dementia, Alzheimer's disease (AD) and mild cognitive impairment (Rauchs et al., 2008; Westerberg et al., 2012; Latreille

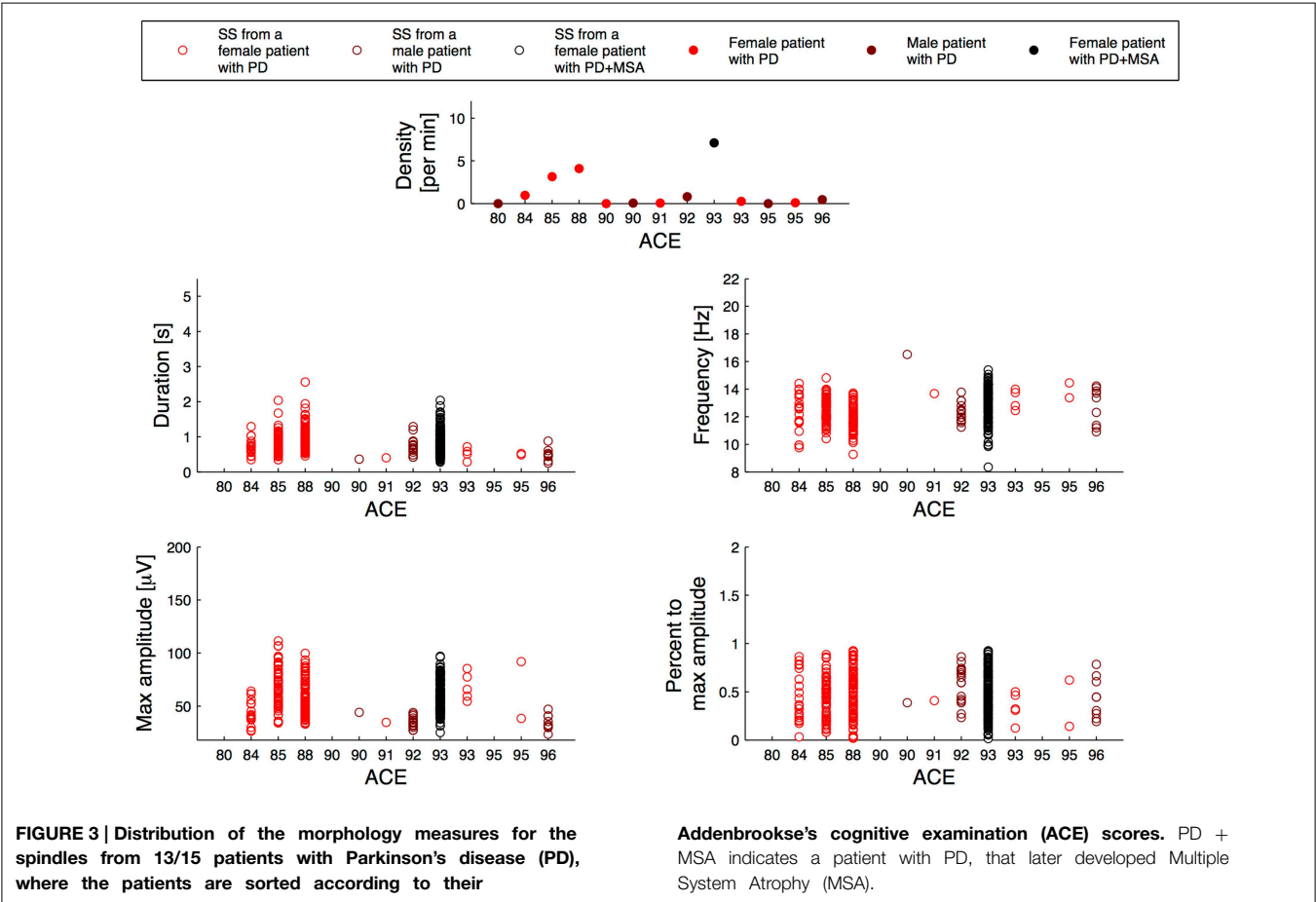
et al., 2015), and is also a sign of normal aging (Wauquier, 1993; De Gennaro and Ferrara, 2003; Ktonas et al., 2009).

To our knowledge, no studies have investigated the impact of L-DOPA on SS morphology. Previous studies have reported that

TABLE 3 | Mean (μ) and standard deviation (σ) for characteristics of spindles in patients with Parkinson's disease (PD) compared to controls (C).

Spindle characteristic	Group consensus (759 SS)		P
	PD(-MSA)	C	
Duration [sec, $\mu \pm \sigma$]	0.86 \pm 0.35	0.77 \pm 0.36	<0.001
Frequency [Hz, $\mu \pm \sigma$]	12.27 \pm 1.07	12.80 \pm 1.23	<0.001
Max peak-to-peak amplitude [μ V, $\mu \pm \sigma$]	56.35 \pm 18.97	48.19 \pm 15.55	<0.001
Max peak-to-peak amplitude After removal of frequencies < 4 Hz [μ V, $\mu \pm \sigma$]	53.42 \pm 17.84	45.29 \pm 14.41	<0.001
Percent-to-peak amplitude [$\mu \pm \sigma$]	0.47 \pm 0.23	0.46 \pm 0.23	NS
Density [per min, $\mu \pm \sigma$]	0.72 \pm 1.28	1.86 \pm 1.57	<0.007

In this case, the patient that later was diagnosed with Multiple System Atrophy (MSA) was excluded from the PD group [PD (-MSA)]. P-values for the Wilcoxon rank sum tests between the two groups are shown. Only spindles in the group consensus are included in the comparison.



SS density is increased in patients with PD taking dopaminergic treatment compared to non-treated patients, but the study lacks a comparison to controls, and evaluation of spindle morphology (Puca et al., 1973). As dopaminergic treatments were not discontinued in this study, we cannot rule out that the changes in SS morphology observed are due to the dopaminergic interactions from the treatments, although we do not believe so, as we did not see increases in SS density in these subjects. Future studies will have to investigate this further including a potential

association between amount and duration of L-DOPA and/or dopamine agonist treatment and SS morphological changes.

Surprisingly, SS in patients with PD had a longer duration and a higher maximum peak-to-peak amplitude. To our knowledge, no other studies have reported differences in SS duration in patients with PD when compared to controls. The maximum peak-to-peak amplitude significantly differ for SS identifications in the group consensus as well as for each of the individual expert's identifications. This finding was also significant after we

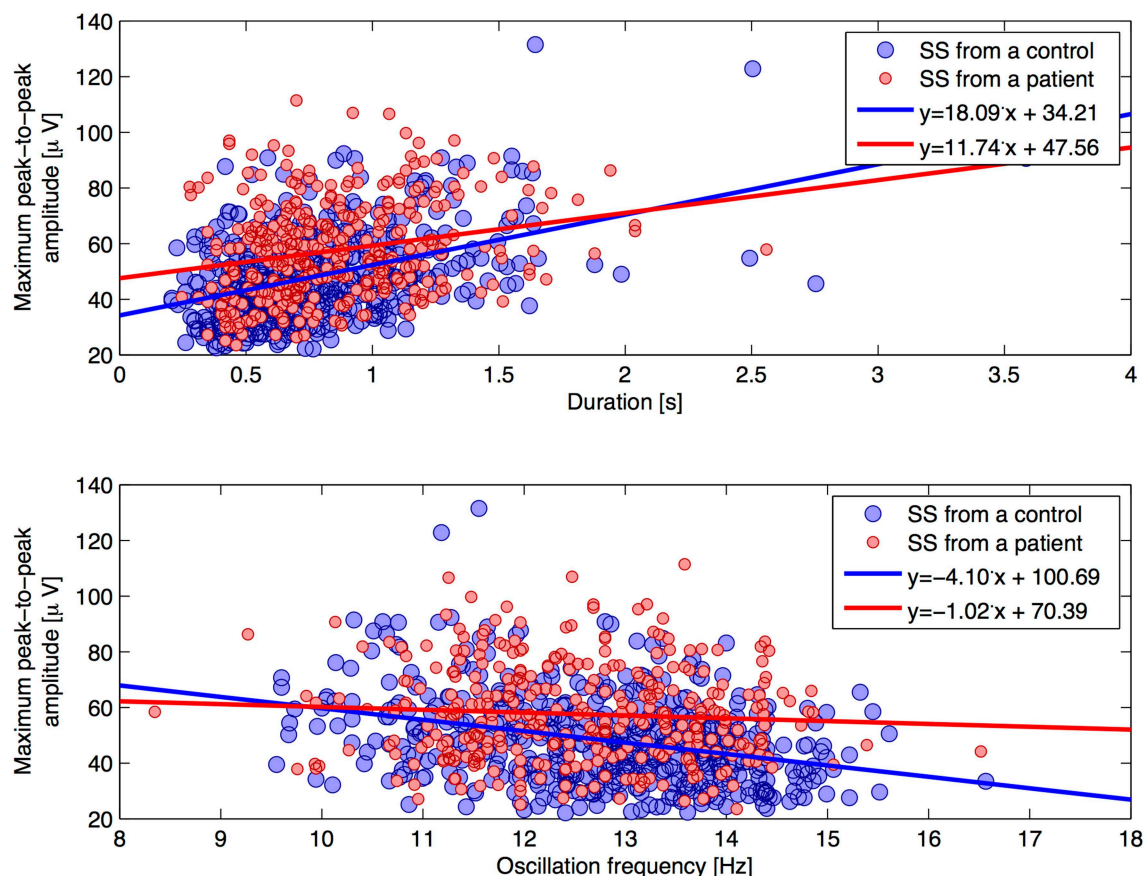


FIGURE 4 | Two scatterplots for individual SS characteristics. The plot illustrates the maximum peak-to-peak amplitude (without removal of frequencies below 4 Hz) as a function of (1) duration (top plot) and (2) oscillation frequency (lower plot), respectively. Trend lines are added for each group.

TABLE 4 | Percent of sleep spindles (SS) identified in the group consensus that do not strictly meet AASM criteria Iber et al. (2007).

AASM criteria	Total SS	PD SS	PD-MSA SS	Control SS	P-value PD vs. controls	P-value PD-MSA vs. controls
Duration too short (<0.5 s)	0.169	0.128	0.134	0.194	0.010	NS
Duration too long (>3 s)	0.001	0	0	0.002	NS	NS
Oscillation frequency too slow (<11 Hz)	0.097	0.090	0.099	0.101	NS	NS
Oscillation frequency too high (>16 Hz)	0.002	0.003	0.005	0.002	NS	NS
At least one criteria not met	0.253	0.212	0.228	0.278	0.027	NS

There were a total of 344 SS from 11 patients with Parkinson's disease (PD) and 557 SS from 15 control subjects. There were 202 SS from 10 patients when one patient with PD, who later was diagnosed with Multiple System Atrophy (MSA) [PD(-MSA)] was left out. χ^2 -tests were used to test for significance between spindles from PD patients (including and excluding the one with MSA) and control subjects.

filtered the data to eliminate the impact of low frequency, high amplitude waves. This was surprising, and contradicts the idea that polygraphic features such as SS and K-complexes are less well formed in various NDDs (Petit et al., 2004; Ktonas et al., 2009). By computing maximum peak-to-peak amplitude both without any further filtration and after elimination of low frequencies, our data show that patients with PD show SS with higher amplitudes, regardless of the EEG patterns surrounding them. Margis et al. (2015) reports increased sigma power in N2 sleep of patients with PD vs. controls. Increased sigma power is consistent with our findings of increased duration and amplitude of spindles, which would overpower the decrease in spindle density we and others have reported in PD. Interestingly, SS morphology was unchanged in schizophrenia patients compared to controls, even though they had a significant decrease in SS density (Wamsley et al., 2012).

Enhanced maximum peak-to-peak amplitude is also not consistent with the findings of Latreille et al. (2015), who reports no significant differences of SS amplitude between PD patients

and controls, and significantly reduced SS amplitude in patients with PD, who later developed Dementia when compared with controls. The SS in Latreille et al. (2015) were found automatically and mandated a duration criteria of least 0.5 s to be included. Also, the spindle detection method includes a filtration of the signal (11–15 Hz) and a threshold determined based on root-mean-square (RMS) values of the background NREM activity (Martin et al., 2013). Lastly, the SS in Latreille et al. (2015) were detected in all NREM stages, and the individual SS characteristics (amplitude and frequency) were computed as the mean of both hemispheres, as they found no significant hemispheric interaction. The definition of SS is thus not the same in the two studies, and the different results could be due to the fact that automatic detectors detect SS that humans cannot see. Another explanation could be that the detector in Latreille et al. (2015) lack to identify the smaller SS in controls, thereby enlarging the mean spindle amplitude in controls. If the threshold used is based on values across all NREM sleep stages, different amount of NREM stages between controls and patients influences the threshold, maybe resulting in harder thresholds to cross for control spindles. Lastly, taking into account the fact that PD patients show more mixed sleep patterns making sleep stages more difficult to distinguish (Danker-Hopfe et al., 2004; Jensen et al., 2010), it could also be that more N3 sleep is present in the annotated data of patients compared to controls, although we did select data from N2 sleep according to each hypnogram. Whether the contradicting findings are due to methodological reasons only, have to be investigated in future studies, e.g., by applying different automatic spindle detectors on the same dataset and on data from different derivations, and see if the morphological alterations are consistent across detectors, manually scorings and derivations.

EEG slowing has been frequently reported in PD (Petit et al., 2004; Rodrigues Brazête et al., 2013), including slowing in occipital, temporo-occipital and frontal regions (Sirakov and Mezan, 1963; Soikkeli et al., 1991; Primavera and Novello, 1992). It is therefore not surprising that we found slower SS oscillation frequencies in PD patients. Whether or not this is specific for PD or generalizable to other NDDs will need further investigations.

TABLE 5 | Mean (μ) and standard deviation (σ) for the spindle characteristics found for the spindles in the group consensus meeting the AASM criteria.

Spindle characteristic	Group consensus (673 SS)		<i>P</i>
	PD	C	
Duration [sec, $\mu \pm \sigma$]	0.93 \pm 0.33	0.85 \pm 0.31	1.95 $\cdot 10^{-4}$
Frequency [Hz, $\mu \pm \sigma$]	12.65 \pm 1.01	13.00 \pm 0.96	9.04 $\cdot 10^{-6}$
Max peak-to-peak amplitude [μ V, $\mu \pm \sigma$]	58.97 \pm 16.64	48.60 \pm 14.92	3.90 $\cdot 10^{-16}$
Max peak-to-peak amplitude After removal of frequencies < 4 Hz [μ V, $\mu \pm \sigma$]	56.06 \pm 15.75	45.76 \pm 13.89	5.27 $\cdot 10^{-18}$
Percent-to-peak amplitude [$\mu \pm \sigma$]	0.47 \pm 0.23	0.45 \pm 0.23	NS
Density [per min, $\mu \pm \sigma$]	0.90 \pm 1.71	1.34 \pm 1.25	4.50 $\cdot 10^{-2}$

Wilcoxon rank sum tests were used to test for significance between patients with PD and control subjects (C).

TABLE 6 | Mean (μ) and standard deviation (σ) for the inter-expert reliability measure F_1 -scores and Cohen's Kappa (κ) for scoring sleep spindles (SS).

SS group definition	F_1 -score		κ		<i>P</i>
	PD	C	PD	C	
Low confidence "maybe"	0.12 \pm 0.11	0.17 \pm 0.12	0.14 \pm 0.11 "slight"	0.16 \pm 0.12 "slight"	NS
Medium confidence "probably"	0.13 \pm 0.10	0.19 \pm 0.11	0.15 \pm 0.10 "slight"	0.18 \pm 0.11 "slight"	NS
High confidence "definitely"	0.24 \pm 0.13	0.32 \pm 0.13	0.21 \pm 0.13 "fair"	0.32 \pm 0.13 "fair"	4.76 $\cdot 10^{-2\kappa}$
Medium or high confidence "probably/definitely"	0.34 \pm 0.15	0.39 \pm 0.17	0.28 \pm 0.15 "fair"	0.39 \pm 0.17 "fair"	NS
All SS	0.41 \pm 0.16	0.45 \pm 0.15	0.32 \pm 0.17 "fair"	0.43 \pm 0.15 "moderate"	NS

The mean and standard deviations are taken across the ten expert-pairs available. Wilcoxon rank sum tests were used to test for significantly lower inter-expert reliability for scoring SS in patients with Parkinson's disease (PD) compared to control subjects (C). κ indicates significance for κ and *F* indicates significance for F_1 -score.

In AD, Rauchs et al. (2008) found no change in spindle density but found that fast spindles (defined as having frequencies of 13–15 Hz) were significantly reduced when compared to age-matched controls. Consistently, Westerberg et al. (2012) found that patients with amnesic mild cognitive impairment had fewer N2 spindles compared to age-matched controls, and that the reduction was seen in fast spindles (13–15 Hz) and not in slow spindles (11–13 Hz). Latreille et al. (2015) found significant lower SS frequency in patients with PD who later developed Dementia compared to controls, but not in Dementia-free patients with PD compared to controls. This last study might however suffer from a selection bias as they automatically defined SS within a certain frequency range, as stated by the AASM. Nonetheless, as in this study, we found that PD patients had a slower SS frequency, both when looking at SS included in the group consensus, but also when looking at SS strictly meeting AASM criteria.

Figures 2, 3 and Supplementary Figures 1, 2 report on SS measures for the PD group consensus, but with subjects sorted according to their disease duration (**Figure 2**), their ACE score (**Figure 3**), their H and Y stage (**Supplementary Figure 1**) and UPDRS part III score (**Supplementary Figure 2**). Although no clear tendency was seen for any of the SS measures for disease duration, ACE score, H and Y stage or UPDRS part III score, longitudinal studies are likely needed to determine whether SS morphology measures can provide prognostic value. Indeed, the patients included here may have had a PD diagnosis for various amounts of time, and inter-subject variation of disease progression and severity makes such a relationship very complicated to analyze. ACE is a brief assessment of cognitive functions and is in this study used as a screening tool to determine Dementia, which none of the patients had at inclusion. A more in-depth examination of cognitive functions as well as a follow-up study of the patients is needed to determine the subject-specific progression and severity rate. These rates can be compared to the SS morphology measures to investigate the prognostic value.

A biomarker does not have to be specific to a disease to have clinical utility, and combining the different SS measures may reveal that different diseases show different trends or different combinations of changes in SS morphology measures. If a trend is found, it is important to also look at SS that might fall out of the stated AASM criteria, as not doing that may misrepresent the data. **Table 4** shows that a rather high proportion of SS in both groups do not meet AASM criteria. Additionally, when looking at inter-expert reliability, it was found that experts are less likely to agree on definite SS in patients when compared to controls. Considering that automatic SS detectors are likely to be used in patients with NDDs, it is highly encouraged to build detectors capable of detecting atypical SS as well. Such atypical SS could be spindles with abnormal duration or frequency or spindles surrounded by EEG that is not typically seen in N2 sleep. Because of this, detectors should not be constrained or designed to perform well only in the context of a single expert or for normal EEG. Ideally, automatic detectors should give a confidence score for each detected SS and group subtypes of SS using specific parameters describing their morphology. Specifically, description of “probable SS” in different patient

groups may give a better idea of the specific morphological changes that can be observed for each disease. Also, such studies should investigate how disease duration and/or severity impact morphology. Such in-depth studies would be beneficial to better understand the pathological differences between the NDDs and also see if any of the morphology measures hold potential for separating diseases or subtypes of them.

In conclusion, we investigated SS in an objective way and found that the oscillation frequency and duration of SS manually scored in clinical settings are not necessarily bound to the limits given by AASM. The shorter or slower SS must have had an ability to stand out from the background EEG, and we believe that these per-definition-not-SS should be included in studies analyzing SS morphology changes, particularly when searching for disease biomarkers.

Based on a group consensus of five individual experts' identification of SS in N2 sleep, we compared 15 patients with PD with 15 age-matched control subjects and found that patients show a lower SS density and that their SS have a longer duration, a higher maximum peak-to-peak amplitude and a slower oscillation frequency. All the included patients were taking dopaminergic treatment, and we can therefore not rule out that the significant differences found could be due to treatment effects. We conclude that SS are significantly altered in patients with PD, but that due to high inter-subject variability in disease progression and severity, future longitudinal studies are needed to investigate the clinical utility of the SS morphology changes as well as their value as prognostic biomarkers.

Financial Support

The PhD project is supported by grants from H. Lundbeck A/S, the Lundbeck Foundation, the Technical University of Denmark and the Center for Healthy Aging, University of Copenhagen.

Acknowledgments

The authors would like to thank the five experts for their time and effort in annotating and giving confidence scores of the sleep spindles analyzed in this study. The PhD project is supported by grants from H. Lundbeck A/S, the Lundbeck Foundation, the Technical University of Denmark, Center for Healthy Aging, University of Copenhagen and Stanford Center for Sleep Sciences and Medicine.

Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fnhum.2015.00233/abstract>

Supplementary Figure 1 | Distribution of the morphology measures for the spindles from 11/15 patients with Parkinson's disease (PD), where the patients are sorted according to their Hoehn and Yahr (H and Y) stage.

Supplementary Figure 2 | Distribution of the morphology measures for the spindles from 11/15 patients with Parkinson's disease (PD), where the patients are sorted according to their Unified Parkinson's Disease Rating Scale (UPDRS) part III score.

References

- American Academy of Sleep Medicine. (2014). *International Classification of Sleep Disorders - Third Edition (ICSD-3)*, 3rd Edn. Westchester, IL: American Academy of Sleep Medicine.
- Braak, H., Tredici, K., Del Rüb, U., Vos, R. A., de Steur, E. N. H. J., and Braak, E. (2003). Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol. Aging* 24, 197–211. doi: 10.1016/S0197-4580(02)00065-9
- Caporro, M., Haneef, Z., Yeh, H. J., Lenartowicz, A., Buttinelli, C., Parvizi, J., et al. (2012). Functional MRI of sleep spindles and K-complexes. *Clin. Neurophysiol.* 123, 303–309. doi: 10.1016/j.clinph.2011.06.018
- Chaudhuri, K. R., Healy, D. G., and Schapira, A. H. (2006). Non-motor symptoms of Parkinson's disease: diagnosis and management. *Lancet Neurol.* 5, 235–245. doi: 10.1016/S1474-4422(06)70373-8
- Chaudhuri, K. R., Odin, P., Antonini, A., and Martinez-Martin, P. (2011). Parkinson's disease: the non-motor issues. *Parkinsonism Relat. Disord.* 17, 717–723. doi: 10.1016/j.parkrel.2011.02.018
- Christensen, J. A. E., Frandsen, R., Kempfner, J., Arvastson, L., Christensen, S. R., Jennum, P., et al. (2012). Separation of Parkinson's patients in early and mature stages from control subjects using one EOG channel. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2012, 2941–2944. doi: 10.1109/EMBC.2012.6346580
- Christensen, J. A. E., Kempfner, J., Zoetmulder, M., Leonthin, H. L., Arvastson, L., Christensen, S. R., et al. (2014a). Decreased sleep spindle density in patients with idiopathic REM sleep behavior disorder and patients with Parkinson's disease. *Clin. Neurophysiol.* 125, 512–519. doi: 10.1016/j.clinph.2013.08.013
- Christensen, J. A. E., Koch, H., Frandsen, R., Kempfner, J., Arvastson, L., Christensen, S. R., et al. (2013). Classification of iRBD and Parkinson's patients based on eye movements during sleep. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2013, 441–444. doi: 10.1109/EMBC.2013.6609531
- Christensen, J. A. E., Zoetmulder, M., Koch, H., Frandsen, R., Arvastson, L., Christensen, S. R., et al. (2014b). Data-driven modeling of sleep EEG and EOG reveals characteristics indicative of pre-Parkinson's and Parkinson's disease. *J. Neurosci. Methods* 235, 262–276. doi: 10.1016/j.jneumeth.2014.07.014
- Comella, C. L., Tanner, C. M., and Ristanovic, R. K. (1993). Polysomnographic sleep measures in Parkinson's disease patients with treatment-induced hallucinations. *Ann. Neurol.* 34, 710–714. doi: 10.1002/ana.410340514
- Danker-Hopfe, H., Kunz, D., Gruber, G., Klösch, G., Lorenzo, J. L., Himanen, S. L., et al. (2004). Interrater reliability between scorers from eight European sleep laboratories in subjects with different sleep disorders. *J. Sleep Res.* 13, 63–69. doi: 10.1046/j.1365-2869.2003.00375.x
- De Gennaro, L., and Ferrara, M. (2003). Sleep spindles: an overview. *Sleep Med. Rev.* 7, 423–440. doi: 10.1053/smr.2002.0252
- Emser, W., Brenner, M., Stober, T., and Schimrigk, K. (1988). Changes in nocturnal sleep in Huntington's and Parkinson's disease. *J. Neurol.* 235, 177–179. doi: 10.1007/BF00314313
- Fogel, S., Martin, N., Lafortune, 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
- Fogel, S. M., and Smith, C. T. (2011). The function of the sleep spindle: a physiological index of intelligence and a mechanism for sleep-dependent memory consolidation. *Neurosci. Biobehav. Rev.* 35, 1154–1165. doi: 10.1016/j.neubiorev.2010.12.003
- Galvin, J. E., Lee, M.-Y., and Trojanowski, J. Q. (2001). Synucleinopathies: clinical and pathological implications. *Arch. Neurol.* 58, 186–190. doi: 10.1001/archneur.58.2.186
- Garcia-Ruiz, P. J., Chaudhuri, K. R., and Martinez-Martin, P. (2014). Non-motor symptoms of Parkinson's disease A review...from the past. *J. Neurol. Sci.* 338, 30–33. doi: 10.1016/j.jns.2014.01.002
- Happe, S., Anderer, P., Pirker, W., Klösch, G., Gruber, G., Saletu, B., et al. (2004). Sleep microstructure and neurodegeneration as measured by [123I]beta-CIT SPECT in treated patients with Parkinson's disease. *J. Neurol.* 251, 1465–1471. doi: 10.1007/s00415-004-0564-3
- Iber, C., Ancoli-Israel, S., Chesson, A. L., and Quan, S. F. (2007). *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology, and Technical Specification*. Westchester, IL: American Academy of Sleep Medicine.
- Iranzo, A. (2011). Sleep-wake changes in the premotor stage of Parkinson disease. *J. Neurol. Sci.* 310, 283–285. doi: 10.1016/j.jns.2011.07.049
- Jensen, P. S., Sorensen, H. B. D., Leonthin, H. L., and Jennum, P. (2010). Automatic sleep scoring in normals and in individuals with neurodegenerative disorders according to new international sleep scoring criteria. *J. Clin. Neurophysiol.* 27, 296–302. doi: 10.1097/WNP.0b013e3181eaa4b
- Kempfner, J., Jennum, P., Nikolic, M., Christensen, J. A. E., and Sorensen, H. B. D. (2013). Sleep phenomena as an early biomarker for Parkinsonism. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2013, 5773–5776. doi: 10.1109/EMBC.2013.6610863
- Koch, H., Christensen, J. A. E., Frandsen, R., Arvastson, L., Christensen, S. R., Jennum, P. J., et al. (2014). Automatic sleep classification using a data-driven topic model reveals latent sleep states. *J. Neurosci. Methods* 235, 130–137. doi: 10.1016/j.jneumeth.2014.07.002
- Ktonas, P. Y., Golemati, S., Xanthopoulos, P., Sakkalis, V., Ortigueira, M. D., Tsekou, H., et al. (2009). Time-frequency analysis methods to quantify the time-varying microstructure of sleep EEG spindles: possibility for dementia biomarkers? *J. Neurosci. Methods* 185, 133–142. doi: 10.1016/j.jneumeth.2009.09.001
- Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174. doi: 10.2307/2529310
- Latreille, V., Carrier, J., Lafortune, M., Postuma, R. B., Bertrand, J.-A., Panisset, M., et al. (2015). Sleep spindles in Parkinson's disease may predict the development of dementia. *Neurobiol. Aging* 36, 1083–1090. doi: 10.1016/j.neurobiolaging.2014.09.009
- Latreille, V., Carrier, J., Montplaisir, J., Lafortune, M., and Gagnon, J.-F. (2011). Non-rapid eye movement sleep characteristics in idiopathic REM sleep behavior disorder. *J. Neurol. Sci.* 310, 159–162. doi: 10.1016/j.jns.2011.06.022
- Leiser, S. C., Dunlop, J., Bowlby, M. R., and Devilbiss, D. M. (2011). Aligning strategies for using EEG as a surrogate biomarker: a review of preclinical and clinical research. *Biochem. Pharmacol.* 81, 1408–1421. doi: 10.1016/j.bcp.2010.10.002
- Margis, R., Schönwald, S. V., Carvalho, D. Z., Gerhardt, G. J. L., and Rieder, C. R. M. (2015). NREM sleep alpha and sigma activity in Parkinson's disease: evidence for conflicting electrophysiological activity? *Clin. Neurophysiol.* 126, 951–958. doi: 10.1016/j.clinph.2014.07.034
- Martin, N., Lafortune, 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
- Micallef, J., Rey, M., Eusebio, A., Audebert, C., Rouby, F., Jouve, E., et al. (2009). Antiparkinsonian drug-induced sleepiness: a double-blind placebo-controlled study of L-dopa, bromocriptine and pramipexole in healthy subjects. *Br. J. Clin. Pharmacol.* 67, 333–340. doi: 10.1111/j.1365-2125.2008.03310.x
- Micanovic, C., and Pal, S. (2014). The diagnostic utility of EEG in early-onset dementia: a systematic review of the literature with narrative analysis. *J. Neural Transm.* 121, 59–69. doi: 10.1007/s00702-013-1070-5
- Myslobodsky, M., Mintz, M., Ben-Mayor, V., and Radwan, H. (1982). Unilateral dopamine deficit and lateral EEG asymmetry: sleep abnormalities in hemi-Parkinson's patients. *Electroencephalogr. Clin. Neurophysiol.* 54, 227–231. doi: 10.1016/0013-4694(82)90164-X
- Nguyen, L., Bradshaw, J. L., Stout, J. C., Croft, R. J., and Georgiou-Karistianis, N. (2010). Electrophysiological measures as potential biomarkers in Huntington's disease: review and future directions. *Brain Res. Rev.* 64, 177–194. doi: 10.1016/j.brainresrev.2010.03.004
- Nicolas, A., Petit, D., Rompré, S., and Montplaisir, J. (2001). Sleep spindle characteristics in healthy subjects of different age groups. *Clin. Neurophysiol.* 112, 521–527. doi: 10.1016/S1388-2457(00)00556-3
- Petit, D., Gagnon, J.-F., Fantini, M. L., Ferini-Strambi, L., and Montplaisir, J. (2004). Sleep and quantitative EEG in neurodegenerative disorders. *J. Psychosom. Res.* 56, 487–496. doi: 10.1016/j.jpsychores.2004.02.001
- Postuma, R. B., Gagnon, J. F., Rompré, S., and Montplaisir, J. Y. (2010). Severity of REM atonia loss in idiopathic REM sleep behavior disorder predicts Parkinson disease. *Neurology* 74, 239–244. doi: 10.1212/WNL.0b013e3181ca0166
- Primavera, A., and Novello, P. (1992). Quantitative electroencephalography in Parkinson's disease, dementia, depression and normal aging. *Neuropsychobiology* 25, 102–105. doi: 10.1159/000118817
- Puca, F., Bricolo, A., and Turella, G. (1973). Effect of L-DOPA or amantadine therapy on sleep spindles in parkinsonism. *Electroencephalogr. Clin. Neurophysiol.* 35, 327–330. doi: 10.1016/0013-4694(73)90245-9

- Rauchs, G., Schabus, M., Parapatics, S., Bertran, F., Clochon, P., Hot, P., et al. (2008). Is there a link between sleep changes and memory in Alzheimer's disease? *Neuroreport* 19, 1159–1162. doi: 10.1097/WNR.0b013e32830867c4
- Rodrigues Brazête, J., Montplaisir, J., Petit, D., Postuma, R. B., Bertrand, J.-A., Génier Marchand, D., et al. (2013). Electroencephalogram slowing in rapid eye movement sleep behavior disorder is associated with mild cognitive impairment. *Sleep Med.* 14, 1059–1063. doi: 10.1016/j.sleep.2013.06.013
- Roth, C., Jeanmonod, D., Magnin, M., Morel, A., and Achermann, P. (2000). Effects of medial thalamotomy and pallido-thalamic tractotomy on sleep and waking EEG in pain and parkinsonian patients. *Clin. Neurophysiol.* 111, 1266–1275. doi: 10.1016/S1388-2457(00)00295-9
- Salawu, F., Danburam, A., and Olokoba, A. (2010). Non-motor symptoms of Parkinson's disease: diagnosis and management. *Niger. J. Med.* 19, 126–131. doi: 10.4314/njm.v19i2.56496
- Schabus, M., Hödlmoser, K., Gruber, G., Sauter, C., Anderer, P., Klösch, G., et al. (2006). Sleep spindle-related activity in the human EEG and its relation to general cognitive and learning abilities. *Eur. J. Neurosci.* 23, 1738–1746. doi: 10.1111/j.1460-9568.2006.04694.x
- Schenck, C. H., Boeve, B. F., and Mahowald, M. W. (2013a). Delayed emergence of a parkinsonian disorder or dementia in 81% of older men initially diagnosed with idiopathic rapid eye movement sleep behavior disorder: a 16-year update on a previously reported series. *Sleep Med.* 14, 744–748. doi: 10.1016/j.sleep.2012.10.009
- Schenck, C. H., Bundlie, S. R., and Mahowald, M. W. (1996). Delayed emergence of a parkinsonian disorder in 38 % of 29 older men initially diagnosed with idiopathic rapid eye movement sleep behavior disorder. *Neurology* 46, 388–393. doi: 10.1212/WNL.46.2.388
- Schenck, C. H., Montplaisir, J. Y., Frauscher, B., Hogl, B., Gagnon, J.-F., Postuma, R., et al. (2013b). Rapid eye movement sleep behavior disorder: devising controlled active treatment studies for symptomatic and neuroprotective therapy—a consensus statement from the International Rapid Eye Movement Sleep Behavior Disorder Study Group. *Sleep Med.* 14, 795–806. doi: 10.1016/j.sleep.2013.02.016
- Sirakov, A., and Mezan, I. (1963). EEG findings in parkinsonism. *Electroencephalogr. Clin. Neurophysiol.* 15, 321–322. doi: 10.1016/0013-4694(63)90101-9
- Soikkeli, R., Partanen, J., Soininen, H., Pääkkönen, A., and Riekkinen, P. (1991). Slowing of EEG in Parkinson's disease. *Electroencephalogr. Clin. Neurophysiol.* 79, 159–165. doi: 10.1016/0013-4694(91)90134-P
- Sorensen, G. L., Kempfner, J., Zoetmulder, M., Sorensen, H. B. D., and Jennum, P. (2012). Attenuated heart rate response in REM sleep behavior disorder and Parkinson's disease. *Mov. Disord.* 27, 888–894. doi: 10.1002/mds.25012
- Sorensen, G. L., Mehlsen, J., and Jennum, P. (2013). Reduced sympathetic activity in idiopathic rapid-eye-movement sleep behavior disorder and Parkinson's disease. *Auton. Neurosci.* 179, 138–141. doi: 10.1016/j.autneu.2013.08.067
- Steriade, M., McCormick, D., and Sejnowski, T. (1993). Thalamocortical oscillations in the sleeping and aroused brain. *Science* 262, 679–685. doi: 10.1126/science.8235588
- Steriade, M., and Timofeev, I. (2003). Neuronal plasticity in thalamocortical networks during sleep and waking oscillations. *Neuron* 37, 563–576. doi: 10.1016/S0896-6273(03)00065-5
- Stevens, S., and Comella, C. L. (2013). "Rapid eye movement sleep behavior disorder," in *Parkinson's Disease and Nonmotor Dysfunction*, 2nd Edn., eds R. F. Pfeiffer and Bodis-Wollner (New York, NY: Humana Press Inc.), 257–265.
- Wamsley, E. J., Tucker, M. A., Shinn, A. K., Ono, K. E., McKinley, S. K., Ely, A. V., et al. (2012). Reduced sleep spindles and spindle coherence in schizophrenia: mechanisms of impaired memory consolidation? *Biol. Psychiatry* 71, 154–161. doi: 10.1016/j.biopsych.2011.08.008
- Warby, S. C., Wendt, S. L., Welinder, P., Munk, E. G. S., Carrillo, O., Sorensen, H. B. D., 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
- Wauquier, A. (1993). Aging and changes in phasic events during sleep. *Physiol. Behav.* 54, 803–806. doi: 10.1016/0031-9384(93)90095-W
- Wendt, S. L., Welinder, P., Sorensen, H. B. D., Jennum, P. J., Peppard, P. E., 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].
- Wermuth, L., Lassen, C. F., Himmelslev, L., Olsen, J., and Ritz, B. (2012). Validation of hospital register-based diagnosis of Parkinson's disease. *Dan. Med. J.* 59:A4391.
- Westerberg, C. E., Mander, B. A., Florczak, S. M., Weintraub, S., Mesulam, M.-M., Zee, P. C., et al. (2012). Concurrent impairments in sleep and memory in amnesic mild cognitive impairment. *J. Int. Neuropsychol. Soc.* 18, 490–500. doi: 10.1017/S135561771200001X

Conflict of Interest Statement: The Reviewer Veronique Latreille declares that, despite being affiliated to the same institution as the author Simon C. Warby, the review process was handled objectively and no conflict of interest exists. 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.

Copyright © 2015 Christensen, Nikolic, Warby, Koch, Zoetmulder, Frandsen, Moghadam, Sorensen, Mignot and Jennum. 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.



Sleep spindles predict stress-related increases in sleep disturbances

Thien Thanh Dang-Vu^{1,2,3,4,5,6*}, Ali Salimi^{1,2,3,4}, Soufiane Boucetta^{1,2,3,4}, Kerstin Wenzel⁶, Jordan O'Byrne^{1,2,3,4}, Marie Brandewinder⁷, Christian Berthomier⁷ and Jean-Philippe Gouin^{3,5,6}

¹ Department of Exercise Science, Concordia University, Montréal, QC, Canada

² Center for Studies in Behavioral Neurobiology, Concordia University, Montréal, QC, Canada

³ PERFORM Center, Concordia University, Montréal, QC, Canada

⁴ Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal, Montréal, QC, Canada

⁵ Center for Clinical Research in Health, Concordia University, Montréal, QC, Canada

⁶ Department of Psychology, Concordia University, Montréal, QC, Canada

⁷ Physip SA, Paris, France

Edited by:

Christian O'Reilly, McGill University, Canada

Reviewed by:

Gary N. Garcia-Molina, Philips

Research North America, USA

Julio Fernandez-Mendoza, Penn State

Milton S. Hershey Medical Center,

USA

*Correspondence:

Thien Thanh Dang-Vu, Department of Exercise Science, Center for Studies in Behavioral Neurobiology,

PERFORM Center, Concordia

University, 7141 Sherbrooke Street

West, SP 165-27, Montreal, QC H4B

1R6, Canada

e-mail: tt.dangvu@concordia.ca

Background and Aim: Predisposing factors place certain individuals at higher risk for insomnia, especially in the presence of precipitating conditions such as stressful life events. Sleep spindles have been shown to play an important role in the preservation of sleep continuity. Lower spindle density might thus constitute an objective predisposing factor for sleep reactivity to stress. The aim of this study was therefore to evaluate the relationship between baseline sleep spindle density and the prospective change in insomnia symptoms in response to a standardized academic stressor.

Methods: Twelve healthy students had a polysomnography recording during a period of lower stress at the beginning of the academic semester, along with an assessment of insomnia complaints using the insomnia severity index (ISI). They completed a second ISI assessment at the end of the semester, a period coinciding with the week prior to final examinations and thus higher stress. Spindle density, amplitude, duration, and frequency, as well as sigma power were computed from C4–O2 electroencephalography derivation during stages N2–N3 of non-rapid-eye-movement (NREM) sleep, across the whole night and for each NREM sleep period. To test for the relationship between spindle density and changes in insomnia symptoms in response to academic stress, spindle measurements at baseline were correlated with changes in ISI across the academic semester.

Results: Spindle density (as well as spindle amplitude and sigma power), particularly during the first NREM sleep period, negatively correlated with changes in ISI ($p < 0.05$).

Conclusion: Lower spindle activity, especially at the beginning of the night, prospectively predicted larger increases in insomnia symptoms in response to stress. This result indicates that individual differences in sleep spindle activity contribute to the differential vulnerability to sleep disturbances in the face of precipitating factors.

Keywords: spindles, sleep, insomnia, stress, EEG

INTRODUCTION

The natural history of insomnia is hypothesized to involve three categories of factors: predisposing factors placing certain individuals at higher risk for insomnia complaints, precipitating factors triggering the onset of insomnia, and perpetuating factors maintaining the insomnia over time (Spielman, 1986). The characterization of predisposing and precipitating factors is of prime importance not only to understand the pathophysiology of insomnia but also to implement optimal preventive sleep interventions. In terms of predisposing factors, longitudinal studies have shown that the rate of new onset insomnia is higher among individuals with depressive or anxiety symptoms, a family history of insomnia, high arousability predisposition, poor general health condition, and pain syndrome (LeBlanc

et al., 2009; Harvey et al., 2014). On the other hand, precipitating factors have been shown to vary with age; in particular, for younger individuals (<30 years old), stress-related factors at work/school constitute the most frequent precipitating events triggering insomnia onset (Bastien et al., 2004). Given one's profile of predisposing factors, individuals are not equally vulnerable to the development of sleep disturbances in the face of common precipitating factors such as stressful events (Drake et al., 2004).

Beyond medical and psychological history, there has been no investigation of the inter-individual variations in sleep architecture – and sleep oscillations – as predisposing factors for the insomnia symptoms. Among the different components of sleep architecture, sleep spindles have been the subjects of intense

research over the past decade (De Gennaro and Ferrara, 2003; Fogel and Smith, 2011). Spindles are defined as waxing-and-waning electroencephalography (EEG) waves oscillating at a frequency of 11–16 Hz and predominant over central EEG derivations; spindles characterize stage N2 of non-rapid-eye-movement (NREM) sleep but can also be found during stage N3 (De Gennaro et al., 2000; Iber et al., 2007). Animal and human studies converge to demonstrate that sleep spindles are generated through the interplay between specific populations of thalamic (particularly thalamic reticular) and cortical neurons (Steriade and McCarley, 2005; Schabus et al., 2007). While the density of sleep spindles varies considerably between individuals, it has been shown that spindle density remains remarkably stable within a same individual across different nights, thus constituting an individual trait (Gailard and Blois, 1981; De Gennaro et al., 2005). Spindles have been shown correlated with measures of intellectual ability as well as with the overnight retention of various types of memory traces, suggesting an important role for spindles in brain plasticity and sleep-related memory consolidation (Gais et al., 2002; Schabus et al., 2004, 2006; Morin et al., 2008; Fogel and Smith, 2011). EEG and functional neuroimaging studies have also demonstrated that the cortical transmission of external – particularly acoustic – stimulation during sleep is drastically diminished during sleep spindles (Cote et al., 2000; Dang-Vu et al., 2011). These findings indicate that spindles isolate the cortex from the environment during sleep, contributing to the preservation of sleep stability. It was then inferred that spindle density, as a trait, might constitute a biomarker of sleep stability in the face of noise. This was confirmed by a study showing that individuals with lower spindle density were more vulnerable to sleep disruption from sounds presented throughout the night (Dang-Vu et al., 2010). Because spindles constitute an index of sleep stability, individuals with reduced spindle density might be more vulnerable to develop insomnia complaints, particularly when confronted to triggering factors such as stress. To the best of our knowledge, no study has investigated the role of sleep spindle as predisposing factor to insomnia onset.

The aim of this study was to prospectively assess whether spindle density would predict the worsening of sleep disturbances in response to a standardized stressor. We chose to follow a population of undergraduate university students during a period of increasing academic stress. In this context, assessing students at the beginning of the semester, corresponding to a lower stress period, and reevaluating them during a follow-up in the week preceding the final examinations, a period of higher stress, provides a unique opportunity to examine individual differences in the evolution of insomnia symptoms in response to a standardized stressor. The validity of this model is supported by data showing an increase of sleep disturbances in response to increased academic stress (Jernelov et al., 2009; Lund et al., 2010). Here, we hypothesized that a lower spindle density at baseline during the low stress period would prospectively predict larger increases in sleep complaints from the low to the high stress periods, therefore defining a neurophysiological vulnerability factor predisposing to the increase of insomnia symptoms.

MATERIALS AND METHODS

PARTICIPANTS

This study is part of a larger project investigating the psychophysiological predictors of stress-induced sleep disturbances. Participants were young healthy students enrolled in full-time undergraduate programs in Psychology or Exercise Science at Concordia University. They were recruited at the beginning of the winter semester through local advertisements posted on the campus. Potential participants were screened using a semi-structured interview for the absence of exclusion criteria, i.e., current insomnia syndrome (APA, 2013), acute or chronic medical condition including psychiatric and sleep disorders, current use of prescribed medication (other than oral contraceptives), current use of over-the-counter sleep medication, cigarette smoking, age >30 years old, working on night shifts. Those deemed eligible then had a baseline assessment during the lower stress period, i.e., within the first 4 weeks of the 15-week academic winter semester of 2014 (January). This baseline evaluation included a self-reported assessment of sleep disturbances using the insomnia severity index (ISI), which is a 7-item questionnaire assessing the nature, severity, and impact of insomnia symptoms over the past month (Bastien et al., 2001), and the Pittsburgh sleep quality index (PSQI), which is a 19-item questionnaire assessing subjective sleep quality in the past month (Buysse et al., 1989). The baseline evaluation also involved an overnight in-lab sleep recording with polysomnography (PSG), in order to confirm the absence of sleep disorders (e.g., sleep apnea) as well as to detect sleep spindles and quantify their parameters (e.g., density). Besides sleep, a self-reported evaluation of psychological distress was obtained using the depression anxiety stress scales (DASS), which includes depression (DASS-D), anxiety (DASS-A), and stress (DASS-S) subscales (Lovibond and Lovibond, 1995). Furthermore, participants also completed at baseline the Ford insomnia response to stress test (FIRST), a questionnaire developed to assess trait sleep reactivity to stress (Drake et al., 2004). All eligible participants then had a second ISI, PSQI, and DASS assessment during the higher stress period, i.e., in the week prior to the final examination period. Participants signed an informed consent form before entering the study, which was approved by Concordia University Human Research Ethics Committee.

PSG RECORDING AND SLEEP SPINDLE ANALYSIS

Overnight PSG recordings were conducted at the PERFORM Center Sleep Laboratory, using 34-channel systems (Embla Titanium, Natus Medical, San Carlos, CA, USA) with EEG referenced to linked-mastoids (bandpass filter 0.3–100 Hz, sampling rate 256 Hz), electrooculography (EOG), electromyography (EMG; submental), nasal–oral thermocouple airflow, and transcutaneous finger pulse oximeter. Participants arrived to the sleep laboratory at least 1 h before their habitual bedtime, in order to allow sufficient time for the PSG setup. They were asked to refrain from alcohol and caffeine consumption and avoid strenuous physical exercise for at least 8 h prior to the PSG recording. Participants went to bed at their habitual bedtime (or at midnight at latest) and slept until they spontaneously woke up the next morning (or at 9 a.m. at latest). PSG was recorded and sleep stages were scored according

to standard criteria (Iber et al., 2007). Sleep apnea syndrome was defined by an apnea–hypopnea index $>5/h$ (exclusion criterion). Sleep spindles were automatically detected during stages N2 and N3 of NREM sleep on EEG C4–O2 derivation. This derivation was chosen given the well-described central predominance of sleep spindles (De Gennaro et al., 2000). The spindle detection method (Aseega software, Physip, Paris, France) used data-driven criteria in order to cope with both inter-subject and inter-recording condition variabilities (Berthomier et al., 2007). It was based on an iterative approach. The first iteration aimed at determining recording-specific thresholds, based on EEG power ratios in delta, alpha, and sigma bands. The second iteration provided precise temporal localization of the events. The final iteration enabled the validation of detected events based on frequency and duration criteria (>0.5 s). Iteration 1 and 3 dealt with raw EEG data, while iteration 2 was applied on the EEG filtered in the spindle (sigma) frequency range using frequency bands adapted to each individual based on his/her global spectral profile (median values for low and high bands were 11.9 and 15.9 Hz, respectively). The density of spindles was computed as the average number of detected spindles per 30 s EEG epoch for each subject. In addition to spindle density, other spindle parameters were also computed for each participant in order to comprehensively evaluate spindle activity: average maximum spindle amplitude (in microvolts), average spindle duration (in seconds), and average spindle frequency (in hertz). After cleaning of the main EEG artifacts, the EEG power in the adapted sigma frequency range (in squared microvolts, per 30 s epoch) was also computed using Hanning window. Spindle parameters and sigma power were first calculated considering the N2–N3 NREM sleep of the entire night. In addition, in order to take into account the variation of spindle activity across successive NREM sleep periods (De Gennaro et al., 2000), spindle parameters, and sigma power were also calculated for each NREM sleep period. These periods were defined according to standard criteria: a NREM sleep period was defined by a period of at least 15 min of NREM sleep followed by at least 5 min of REM sleep, and the first four NREM sleep periods were considered given that there are usually four NREM sleep periods during overnight sleep in adults (Feinberg and Floyd, 1979). For completeness, spindle parameters and sigma power were also calculated during stage N2 only, for the entire night as well as for each NREM sleep period, and the data from these additional analyses are presented in the supplementary material given that they showed results quite similar to those from N2–N3 combined.

STATISTICAL ANALYSIS

Changes in self-reported sleep quality and stress over time were first evaluated using dependent *t*-tests. The evolution of spindle parameters across NREM sleep periods was then examined using one-way repeated measures analysis of variance (ANOVA) with Bonferroni *post hoc* tests. In order to test our main hypothesis, correlations between spindle density (during N2–N3 for the whole night and for each NREM sleep period) and changes in self-reported sleep quality (score during high stress period minus score during low stress period) were calculated using Pearson product-moment correlation. Secondary analyses likewise tested the Pearson correlations between other spindle parameters

(spindle amplitude, duration, frequency) or sigma power and sleep quality changes. FIRST scores were also correlated with sleep quality changes to evaluate the predictive potential of this self-reported measure of sleep reactivity to stress in the context of academic stress. All analyses were considered significant at a *p* value <0.05 , and were conducted with SPSS Statistics 22.0 (IBM, New York, NY, USA).

RESULTS

Out of 22 potential participants, 12 were confirmed eligible and presented at least 4 NREM sleep periods during their PSG recording. The majority of them were female students (10/12). Seven of the participants were psychology majors; the remainders were exercise science majors. Characteristics of the participants, including age, PSG parameters, PSQI, and ISI values are presented in **Table 1**. There was a significant increase in total ISI score from the low to the high stress periods ($t = 2.23$, $p = 0.047$), as illustrated in **Figure 1**, but no significant change in any of the seven individual items of the ISI. There was no significant change in PSQI ($t = 0.44$, $p = 0.67$). The total DASS score significantly increased ($t = 2.9$;

Table 1 | Characteristics of participants (*N* = 12, 10 females).

Parameters	M (SD)	Range
Age	21.08 (2.43)	17–25
PSG total sleep time (min)	417.64 (62.81)	326–527.5
PSG sleep efficiency (%)	85.86 (8.81)	66.3–96.8
PSG stage N1 (% of TST)	13.01 (8.15)	3.8–30.9
PSG stage N2 (% of TST)	51.58 (6.30)	42.5–61.4
PSG stage N3 (% of TST)	18.36 (5.10)	11.8–27.5
PSG stage REM (% of TST)	17.33 (2.39)	14.1–21.3
NREM period 1 (min)	91.13 (48.10)	44–176
NREM period 2 (min)	83.83 (40.50)	30.5–175.5
NREM period 3 (min)	87.17 (37.83)	43.5–169.5
NREM period 4 (min)	61.04 (28.08)	22–123.5
Apnea-hypopnea index (nb/h)	0.31 (0.39)	0–1
Arousal index (nb/h)	8.11 (3.05)	3.3–12.5
ISI low stress	5.75 (5.82)	0–15
ISI high stress	7.75 (6.05)	0–20
PSQI low stress	5.17 (3.16)	1–10
PSQI high stress	5.42 (3.85)	0–14
DASS-D low stress	3.08 (4.56)	0–17
DASS-D high stress	4.83 (4.76)	0–17
DASS-A low stress	3 (4.24)	0–13
DASS-A high stress	4.17 (4.43)	0–15
DASS-S low stress	5.17 (4.17)	0–16
DASS-S high stress	7.92 (5.3)	0–17
DASS-total low stress	11.25 (12.39)	0–46
DASS-total high stress	16.92 (13.73)	0–49
FIRST	21.83 (7.60)	11–32

DASS, depression anxiety stress scales (D, depression subscale; A, anxiety subscale; S, stress subscale); FIRST, Ford insomnia response to stress test; ISI, insomnia severity index; NREM period 1–4, first to fourth NREM sleep period; PSG, polysomnography; PSQI, Pittsburgh sleep quality index; TST, total sleep time.

$p = 0.014$), including increases in depression [DASS-D: $t = 2.55$; $p = 0.027$] and perceived stress (DASS-S: $t = 3.48$; $p = 0.005$) subscales, but not anxiety (DASS-A: $t = 1.7$; $p = 0.12$). There was a significant change in spindle density across NREM sleep periods ($F = 4.58$, $p = 0.033$); *post hoc* tests showed that spindle density during the first NREM sleep period was significantly lower than during each of the three following NREM sleep periods ($p = 0.031$, 0.025 , and 0.026 , respectively) (Figure 2A). Spindle duration also

demonstrated a significant change across NREM sleep periods ($F = 11.62$, $p = 0.002$), with *post hoc* tests revealing a significantly lower spindle duration during NREM sleep period 1 compared to each of the next periods ($p = 0.002$) (Figure 2C). There was no significant change in spindle amplitude ($F = 0.432$, $p = 0.735$), spindle frequency ($F = 1.664$, $p = 0.243$), and sigma power ($F = 0.869$, $p = 0.492$) across NREM sleep periods (Figure 2).

Given that PSQI did not show significant increase in response to academic stress, bivariate correlations were performed between spindle parameters or sigma power and ISI change only (Table 2). When examining spindle density, there was a significant negative correlation between spindle density during the first NREM sleep period and ISI change (Figure 3A), i.e., lower spindle density at the beginning of the night was associated with higher increases in sleep complaints in response to academic stress. The correlation was not significant for spindle density during the whole night or during any other NREM sleep period. When looking at the other spindle parameters, duration, and frequency did not correlate with ISI change. Spindle amplitude, however, negatively correlated with ISI change, when considering either the first (Figure 3B) or the third NREM sleep period. Finally, there was a significant negative correlation between sigma EEG power, for the whole night as well as for each NREM sleep period, and ISI change. This correlation with sigma power was the most significant during the first NREM sleep period (Figure 3C). FIRST score at baseline was not correlated with ISI change ($r = -0.10$; $p = 0.75$).

DISCUSSION

Taken together these results indicate that spindle activity constitutes a predisposing factor for the future aggravation of insomnia complaints in response to stress. They confirm our hypothesis that lower spindle density is associated with a higher vulnerability to

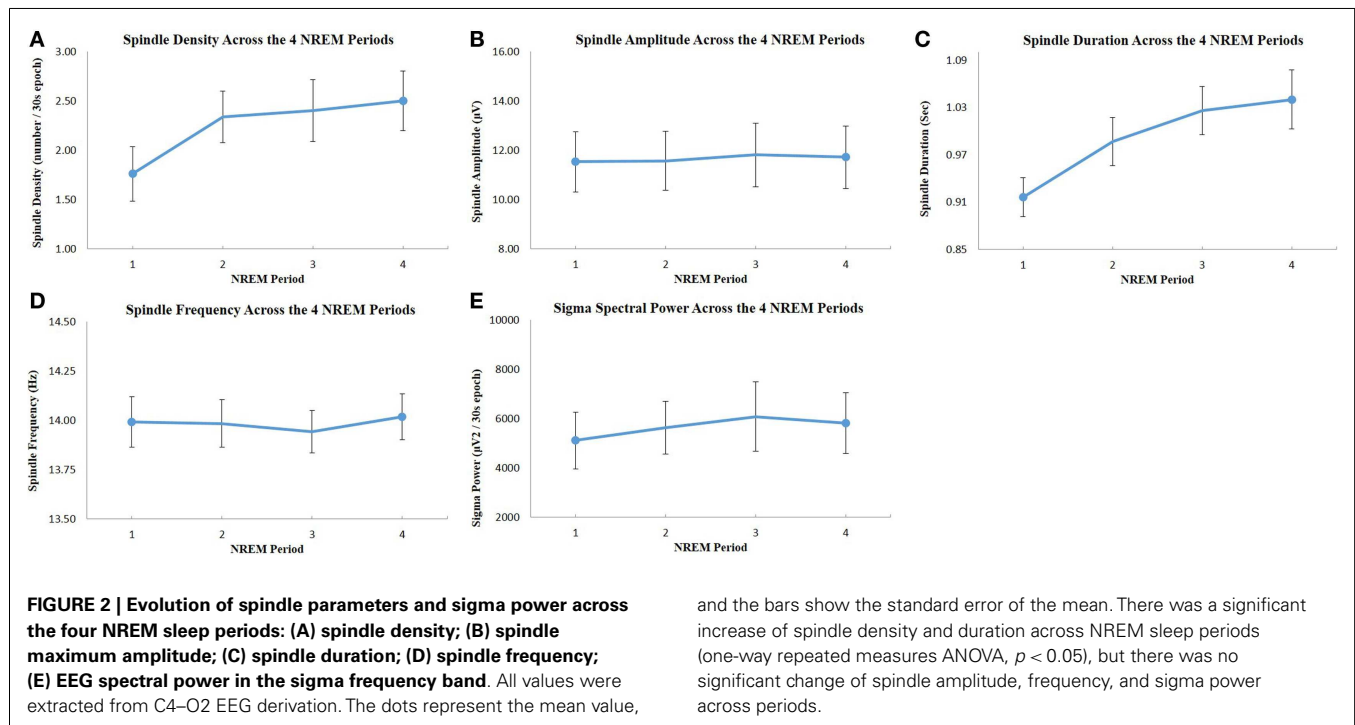
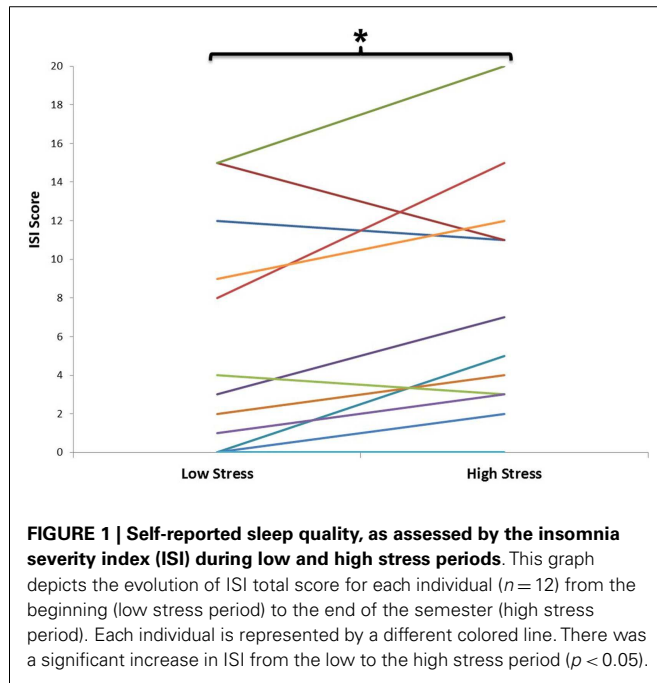


Table 2 | Correlations between baseline spindle parameters or sigma power and change in insomnia severity index from low to high stress period.

Parameters	Δ ISI	
	Pearson's <i>r</i>	<i>p</i> value
Spindle density during total NREM sleep	-0.322	0.308
Spindle density in NREM period 1	-0.578	0.049*
Spindle density in NREM period 2	-0.156	0.628
Spindle density in NREM period 3	-0.378	0.226
Spindle density in NREM period 4	-0.192	0.549
Spindle amplitude during total NREM sleep	-0.502	0.097
Spindle amplitude in NREM period 1	-0.588	0.044*
Spindle amplitude in NREM period 2	-0.486	0.109
Spindle amplitude in NREM period 3	-0.591	0.043*
Spindle amplitude in NREM period 4	-0.545	0.067
Spindle duration during total NREM sleep	0.104	0.748
Spindle duration in NREM period 1	0.010	0.975
Spindle duration in NREM period 2	0.136	0.672
Spindle duration in NREM period 3	-0.074	0.819
Spindle duration in NREM period 4	0.164	0.610
Spindle frequency during total NREM sleep	-0.167	0.603
Spindle frequency in NREM period 1	-0.257	0.420
Spindle frequency in NREM period 2	-0.091	0.779
Spindle frequency in NREM period 3	-0.182	0.572
Spindle frequency in NREM period 4	-0.210	0.512
Sigma spectral power during total NREM sleep	-0.685	0.014*
Sigma spectral power in NREM period 1	-0.761	0.004**
Sigma spectral power in NREM period 2	-0.597	0.041*
Sigma spectral power in NREM period 3	-0.710	0.010**
Sigma spectral power in NREM period 4	-0.631	0.028*

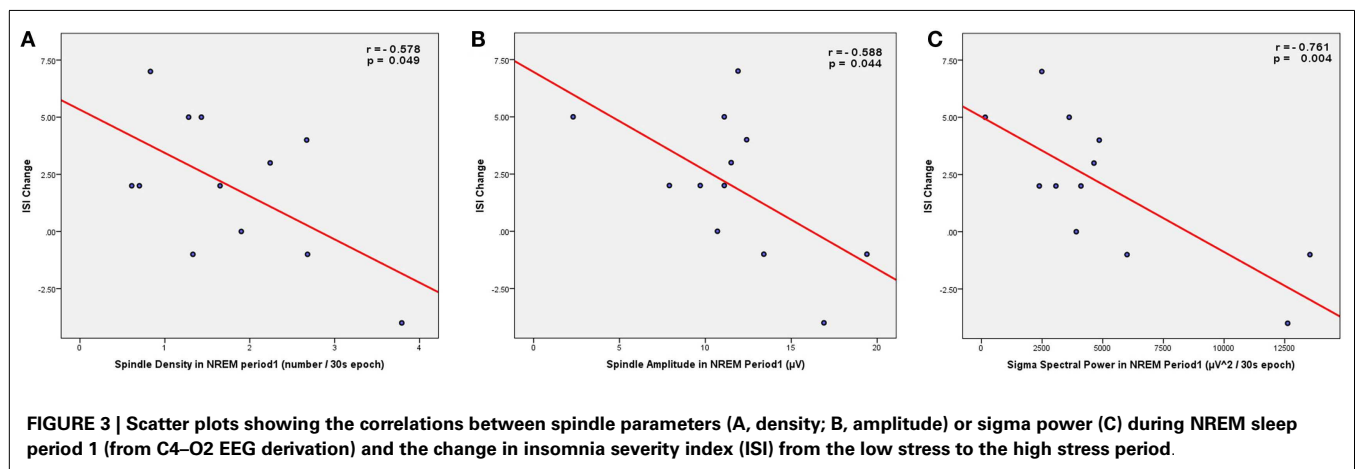
Δ ISI, insomnia severity index change from low to high stress period; NREM period 1–4, first to fourth NREM sleep period.

*Significance at $p < 0.05$.

**Significance at $p < 0.01$.

sleep disturbances triggered by stress. In this study, the severity of insomnia symptoms assessed by ISI increased from the beginning to the end of the semester – a period coinciding with intense preparations for final examinations – thus validating our chosen model of stress-induced sleep disturbances. The validity of this academic stress model is further supported by the increase of self-reported perceived stress across the academic semester, as shown by the significant DASS-S increase with time. In addition, these results extended our initial hypothesis in two directions. First, we found that this predictive relationship between spindle density and ISI was restricted to the spindles during the first NREM sleep period, i.e., at the beginning of the night. Second, besides spindle density, we also observed significant correlations between ISI change and other spindle parameters, such as spindle amplitude and EEG power in the sigma frequency range, suggesting that spindle activity in general (and not only the mere presence of spindles) prospectively affects the evolution of insomnia symptoms.

The possible mechanisms underlying the relationship between spindle and stress-triggered insomnia complaints can be discussed in the light of previous studies investigating the functional properties of human sleep spindles. On the one hand, the filtering of external information at the thalamic level during sleep spindles might provide a potential mechanism for this predictive relationship. Tones presented during most of NREM sleep were found to activate the thalamus and the primary auditory cortex in an EEG/functional magnetic resonance imaging study; however, tones presented in coincidence with sleep spindles did not consistently activate thalamocortical auditory circuits (Dang-Vu et al., 2011). This result suggests that spindles provide a gating process to preserve the sleeping brain from disruption by sounds (and presumably also by other types of environmental stimulation). Based on this finding, a further study investigated the relationship between spindle density and the probability of maintaining sleep continuity under presentation of sounds with increasing intensities (Dang-Vu et al., 2010). At any sound intensity level, individuals with higher spindle density were more likely to preserve the continuity of sleep than subjects with lower spindle density. The ability of individuals with higher spindle density to more efficiently sleep throughout noise might provide them with a better capacity to



resist sleep disturbances in response to stress. Exposure to acute stress is indeed known to enhance sensitivity to noise (Hasson et al., 2013), and is thus likely to increase vulnerability to sounds during sleep. Through the gating mechanisms associated with spindles, individuals with higher spindle density might be in a better position to counter the deleterious consequences of stress on noise sensitivity during sleep, leading to a lower propensity to insomnia complaints in response to stress. Future studies investigating sleep spindles in relation to acoustic stimulation in periods of high stress are needed to support this interpretation.

Spindles have also been shown associated with a variety of cognitive measures. Higher number of spindles and higher EEG sigma power have been shown positively correlated with better perceptual and analytical skills measured by the performance intellectual quotient (Fogel et al., 2007), as well as with higher score at the Raven progressive matrices test reflecting general cognitive abilities (Bodizs et al., 2005; Schabus et al., 2006). These findings suggest that spindles constitute a neurophysiological biomarker of intellectual capacities. Beyond these correlations with broad scores of cognitive abilities, previous studies also found that sleep spindle activity increased following procedural memory tasks such as motor sequence learning (Barakat et al., 2011) and mirror tracing task (Tamaki et al., 2008), as well as following declarative learning tasks (word pairs) (Gais et al., 2002; Schabus et al., 2004), and these increases correlated with overnight improvements of performance. Therefore, an alternative explanation for the relationship found in the present study is that individuals with higher spindle density or activity, given their higher cognitive abilities, might be more capable of efficiently learning their course materials and thus managing academic stress, which might ultimately make them less susceptible to sleep disturbances in such context. Both mechanisms – sleep-protective and cognitive – are not mutually exclusive and might act synergistically to confer individuals with higher spindle activity the ability to maintain sleep stability in the face of stress.

In this study, spindle density and other spindle parameters during stages N2–N3 were computed separately for each NREM sleep period, in addition to their quantification throughout the whole night. Indeed, spindle parameters do not remain constant throughout the night (De Gennaro et al., 2000). We observed a significant increase of spindle density and spindle duration over the course of the night (particularly between NREM sleep period 1 and each of the following NREM sleep periods). These observations are in line with previous studies showing increases of spindle density and duration across successive NREM sleep periods (De Gennaro et al., 2000; Martin et al., 2013). Interestingly, the modulation of spindle parameters throughout the night had an impact on the predictive relationships with sleep complaints, since spindle density only in the first NREM sleep period was significantly correlated with ISI change. This result suggests that sleep spindles during the early phase of the night have a predominant influence on the perception of changes in insomnia complaints. The reasons for such effects of spindles during early night remain unclear. Interestingly, a differential effect of spindle activity as a function of NREM sleep period was also observed in a study comparing spindle density between teenagers with major depressive disorder, teenagers at risk for depression and healthy controls (Lopez et al.,

2010). In contrast to our results, depressed and at-risk teenagers had lower spindle density than controls during the third and fourth NREM sleep period, but not at the beginning of the night. These various results suggest a differential clinical significance for spindle activity depending on the corresponding NREM sleep period.

Our findings also demonstrated that, besides spindle density, other parameters reflecting spindle activity (spindle amplitude, sigma power) correlated with change of sleep quality in response to stress. The correlations with spindle amplitude and EEG sigma power suggest that spindle activity (i.e., intensity) – reflecting the degree of thalamocortical synchronization – also modulates the perception of stress-induced sleep disturbances. Other studies investigating the functional properties of sleep spindles also resorted to measures of spindle intensity: EEG sigma power correlated with scores of intellectual abilities (Fogel et al., 2007), and spindle amplitude modulated the reactivation during sleep spindles of brain regions involved in the encoding of a declarative learning task (Bergmann et al., 2012). In our study, correlation between sigma power and stress-induced sleep complaints was significant for all NREM sleep periods, while it was limited to the first NREM sleep period for spindle density (as well as the third for spindle amplitude). Such discrepancies between the effects of spindle density compared to those of sigma power were also observed in previous studies (Gais et al., 2002), which further underlines that EEG spectral power in the sigma frequency range cannot be fully equated to the detection of sleep spindles as discrete events. Indeed, sigma power also captures activities that do not meet the standard criteria for full-blown sleep spindles, and thus constitutes a more sensitive (but less specific) method for spindle quantification. Taken together these effects of spindle density, amplitude, and sigma power suggest that, although the contribution of spindles during early night seems more predominant, spindle activity throughout the whole night affects the perception of stress-induced changes in sleep quality.

If lower spindle density (and activity) constitutes a predisposing factor for the surge of insomnia complaints, it could be expected that chronic insomniacs would tend to demonstrate lower spindle measures compared to good sleepers. A previous study analyzed the differences in spindle density between chronic primary insomniacs and good sleepers, by performing a visual detection of sleep spindles on C3 derivation (Bastien et al., 2009). Surprisingly, no significant difference in spindle density was found between groups. Further studies are needed to replicate this result, and to extend it to other (more sensitive) modalities of spindle detection as well as to other measures of spindle activity such as spindle amplitude and sigma power. If confirmed, the absence of change in spindle activity in chronic insomniacs as a group might reflect the large heterogeneity in the clinical presentation and sleep characteristics even within primary insomniacs. For instance, the presence of objective sleep disturbances, as defined by PSG decreases in total sleep time and sleep efficiency, is not observed consistently across chronic insomniacs (Vgontzas et al., 1994). However, the presence of objective short sleep duration defines a subgroup of insomniacs with a distinct clinical profile exemplified by a higher risk for hypertension, diabetes, cognitive impairment, and mortality (Vgontzas et al., 2013). Likewise, it is possible that there is a subgroup of insomniacs characterized by a higher vulnerability

to environmental disturbances due to a lesser amount of sleep-protective factors such as sleep spindles. In contrast, another subgroup might instead be characterized by less objective sleep disruption and a larger contribution of cognitive-emotional factors such as dysfunctional beliefs about sleep and higher levels of anxiety and worry. Considering the chronic insomnia population as a single group might dilute the alterations of sleep microarchitecture that possibly affect a subpopulation of insomniacs only. Future studies should further explore the quantification of sleep-protective mechanisms in chronic insomniacs and subgroups of insomniacs.

While the current study was primarily focused on the neurophysiological predictors of sleep disturbances through the assessment of spindle activity, it should be reminded that psychological and medical factors also play an important role in the incidence of insomnia complaints. For instance, mental health problems, maladaptive personality traits, a positive family history of insomnia, and an objectively shorter sleep duration on PSG were associated with a higher risk of evolution of poor sleep toward chronic insomnia (Fernandez-Mendoza et al., 2012). As for insomnia complaints in response to stress, specific questionnaires of vulnerability to stress-induced insomnia have been developed, such as the FIRST (Drake et al., 2004). Surprisingly, the FIRST score did not predict changes in ISI from low to high stress periods in our analysis. This might be explained by the high correlation between the FIRST score and ISI at baseline in this sample ($r_{\text{FIRST-ISI}} = 0.86$), leaving little room to predict change over time. Nevertheless, it has been previously shown that individuals with higher score on the FIRST (Drake et al., 2004) are more vulnerable to the first night effect (i.e., worse sleep quality during the first night of sleep recording in lab) and to the sleep-disrupting effects of caffeine (Drake et al., 2004, 2006), and demonstrate higher risk of developing persistent insomnia over time (Drake et al., 2014; Jarrin et al., 2014). Interestingly, the differential vulnerability to stress-induced insomnia may emerge from differences in hyperarousal predisposition, given the association between FIRST scores and indices of cognitive-emotional hyperarousal (Fernandez-Mendoza et al., 2010), which has recently been demonstrated to be (at least partially) heritable (Fernandez-Mendoza et al., 2014). Because sleep spindles modulate sleep stability in response to environmental stimulation (Dang-Vu et al., 2010, 2011), lower spindle activity – which has been found in the present study to predispose to higher increase of sleep disturbances – might be considered as a trait predisposing to a state of neurobiological hyperarousal in which individuals are more vulnerable to externally driven sleep disruption. Therefore, these various findings on the vulnerability to stress-induced insomnia can be integrated within the framework of the hyperarousal model for insomnia viewed from a psychophysiological perspective (Riemann et al., 2010).

There are several limitations to the current study. First, larger samples are needed to confirm these findings. Due to the limited number of participants, correction for multiple comparisons was not applied in the present data set, and thus our findings need replication. Second, only undergraduate university students were included in the present study due to the need of a naturally occurring stressor encompassing well-defined periods of lower and higher stress, as provided by the model of academic stress.

Future studies should extend these findings to other populations and other types of stressors, including chronic stressors that may impact the persistence of insomnia complaints over time. Third, assessment of sleep quality and insomnia complaints was evaluated through self-reported questionnaires only: ISI and PSQI. The absence of significant PSQI change across the semester in our study might indicate that the impact of academic stress on sleep predominantly affects insomnia complaints rather than general sleep quality. Furthermore, the nature of the stressor (academic stress) precluded the repetition of objective sleep measurements with PSG, given the difficulty of having participants coming at the sleep laboratory during busy periods of final examinations. The use of more practical objective measures of sleep such as actigraphy measurements might constitute an interesting complement in future studies, in order to obtain not only objective but also prolonged assessments of sleep over several days or weeks. Finally, we restricted our analyses to spindle activity over central derivations (C4), given the centroparietal predominance of sleep spindle activity (De Gennaro et al., 2000). In order to avoid additional comparisons in our limited sample, distinction between fast and slow spindles was not performed in this present analysis. Our results, however, suggest that the frequency of spindles did not affect the change in insomnia symptoms, given the absence of correlation between spindle frequency and ISI change (Table 2). Future studies on larger samples could further evaluate the role of spindle frequency on sleep quality changes by analyzing the role of slow and fast spindles separately. The study of other EEG oscillations during sleep might also be of interest given previous results indicating the contribution of brain oscillations in other frequency bands, such as alpha rhythms (McKinney et al., 2011) and slow wave activity (Dang-Vu et al., 2011; Schabus et al., 2012), to the preservation of sleep continuity in the face of external stimulation.

CONCLUSION

Our study provides the first evidence for the contribution of sleep neurophysiological activity to the prospective increase of sleep disturbances in response to a standardized stressor in a sample of young healthy volunteers. In line with previous findings indicating that sleep spindle constitutes a biomarker of sleep stability, our results suggest that spindle activity also represents a predisposing factor modulating the vulnerability to sleep disruption in conditions of stress. These results have implications for the understanding of the neural mechanisms underlying the evolution of sleep disturbances and particularly insomnia. They might also have clinical implications, by providing a biomarker for the identification of individuals at risk for future sleep disruption. Finally, our findings emphasize the potential importance of future therapeutic interventions aimed at enhancing sleep spindle activity in order to preserve sleep quality.

AUTHOR CONTRIBUTIONS

TDV and JPG designed the study. AS, SB, KW, and JOB acquired the data. TDV, AS, SB, MB, CB, and JG analyzed the data. TDV, AS, SB, and JPG interpreted the results. TDV wrote the manuscript. AS prepared the tables and figures. All the authors revised

and commented the manuscript, gave their final approval of the manuscript, and agree to be accountable for all aspects of the work.

ACKNOWLEDGMENTS

This research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The authors thank Ms. Ruby Bedi, Ms. Marilia Bedendi, Ms. Neressa Noel, and the Clinique Sommeil Santé for their contribution to the setup and scoring of sleep recordings. Dr. TDV receives research support from the Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council of Canada (NSERC), the Fonds de Recherche du Québec – Santé (FRQ-S), the Canada Foundation for Innovation (CFI), the Sleep Research Society Foundation (SRSF), the Fonds Québécois de Recherche sur le Vieillessement (RQRV), the Institut Universitaire de Gériatrie de Montréal, and Concordia University. Dr. JPG receives research support from the Canada Research Chair program, the CFI, the CIHR, and the Social Sciences and Humanities Research Council of Canada (SSHRC). Mr. JOB receives scholarship support from the CIHR.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://www.frontiersin.org/Journal/10.3389/fnhum.2015.00068/abstract>

REFERENCES

- APA. (2013). *Diagnostic and Statistical Manual of Mental Disorders*, 5th Edn. Arlington, VA: American Psychiatric Publishing.
- Barakat, M., Doyon, J., Debas, K., Vandewalle, G., Morin, A., Poirier, G., et al. (2011). Fast and slow spindle involvement in the consolidation of a new motor sequence. *Behav. Brain Res.* 217, 117–121. doi:10.1016/j.bbr.2010.10.019
- Bastien, C. H., St-Jean, G., Turcotte, I., Morin, C. M., Lavallee, M., and Carrier, J. (2009). Sleep spindles in chronic psychophysiological insomnia. *J. Psychosom. Res.* 66, 59–65. doi:10.1016/j.psychres.2008.05.013
- Bastien, C. H., Vallieres, A., and Morin, C. M. (2001). Validation of the insomnia severity index as an outcome measure for insomnia research. *Sleep Med.* 2, 297–307. doi:10.1016/S1389-9457(00)00065-4
- Bastien, C. H., Vallieres, A., and Morin, C. M. (2004). Precipitating factors of insomnia. *Behav. Sleep Med.* 2, 50–62. doi:10.1207/s15402010bsm0201_5
- Bergmann, T. O., Molle, M., Diedrichs, J., Born, J., and Siebner, H. R. (2012). Sleep spindle-related reactivation of category-specific cortical regions after learning face-scene associations. *Neuroimage* 59, 2733–2742. doi:10.1016/j.neuroimage.2011.10.036
- Berthomier, C., Drouot, X., Herman-Stoica, M., Berthomier, P., Prado, J., Bokar-Thire, D., et al. (2007). Automatic analysis of single-channel sleep EEG: validation in healthy individuals. *Sleep* 30, 1587–1595.
- Bodizs, R., Kis, T., Lazar, A. S., Havran, L., Rigo, P., Clemens, Z., et al. (2005). Prediction of general mental ability based on neural oscillation measures of sleep. *J. Sleep Res.* 14, 285–292. doi:10.1111/j.1365-2869.2005.00472.x
- Buyse, D. J., Reynolds, C. F. III, Monk, T. H., Berman, S. R., and Kupfer, D. J. (1989). The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry Res.* 28, 193–213. doi:10.1016/0165-1781(89)90047-4
- Cote, K. A., Epps, T. M., and Campbell, K. B. (2000). The role of the spindle in human information processing of high-intensity stimuli during sleep. *J. Sleep Res.* 9, 19–26. doi:10.1046/j.1365-2869.2000.00188.x
- Dang-Vu, T. T., Bonjean, M., Schabus, M., Boly, M., Darsaud, A., Desseilles, M., et al. (2011). Interplay between spontaneous and induced brain activity during human non-rapid eye movement sleep. *Proc. Natl. Acad. Sci. U.S.A.* 108, 15438–15443. doi:10.1073/pnas.1112503108
- 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/smr.2002.0252
- De Gennaro, L., Ferrara, M., and Bertini, M. (2000). Topographical distribution of spindles: variations between and within NREM sleep cycles. *Sleep Res. Online* 3, 155–160. doi:10.1046/j.1365-2869.2000.00193.x
- De Gennaro, L., Ferrara, M., Vecchio, F., Curcio, G., and Bertini, M. (2005). An electroencephalographic fingerprint of human sleep. *Neuroimage* 26, 114–122. doi:10.1016/j.neuroimage.2005.01.020
- Drake, C., Richardson, G., Roehrs, T., Scofield, H., and Roth, T. (2004). Vulnerability to stress-related sleep disturbance and hyperarousal. *Sleep* 27, 285–291.
- Drake, C. L., Jefferson, C., Roehrs, T., and Roth, T. (2006). Stress-related sleep disturbance and polysomnographic response to caffeine. *Sleep Med.* 7, 567–572. doi:10.1016/j.sleep.2006.03.019
- Drake, C. L., Pillai, V., and Roth, T. (2014). Stress and sleep reactivity: a prospective investigation of the stress-diathesis model of insomnia. *Sleep* 37, 1295–1304. doi:10.5665/sleep.3916
- Feinberg, I., and Floyd, T. C. (1979). Systematic trends across the night in human sleep cycles. *Psychophysiology* 16, 283–291. doi:10.1111/j.1469-8986.1979.tb02991.x
- Fernandez-Mendoza, J., Shaffer, M. L., Olavarrieta-Bernardino, S., Vgontzas, A. N., Calhoun, S. L., Bixler, E. O., et al. (2014). Cognitive-emotional hyperarousal in the offspring of parents vulnerable to insomnia: a nuclear family study. *J. Sleep Res.* 23, 489–498. doi:10.1111/jsr.12168
- Fernandez-Mendoza, J., Vela-Bueno, A., Vgontzas, A. N., Ramos-Platon, M. J., Olavarrieta-Bernardino, S., Bixler, E. O., et al. (2010). Cognitive-emotional hyperarousal as a premorbid characteristic of individuals vulnerable to insomnia. *Psychosom. Med.* 72, 397–403. doi:10.1097/PSY.0b013e3181d75319
- Fernandez-Mendoza, J., Vgontzas, A. N., Bixler, E. O., Singareddy, R., Shaffer, M. L., Calhoun, S. L., et al. (2012). Clinical and polysomnographic predictors of the natural history of poor sleep in the general population. *Sleep* 35, 689–697. doi:10.5665/sleep.1832
- Fogel, S. M., Nader, R., Cote, K. A., and Smith, C. T. (2007). Sleep spindles and learning potential. *Behav. Neurosci.* 121, 1–10. doi:10.1037/0735-7044.121.1.1
- Fogel, S. M., and Smith, C. T. (2011). The function of the sleep spindle: a physiological index of intelligence and a mechanism for sleep-dependent memory consolidation. *Neurosci. Biobehav. Rev.* 35, 1154–1165. doi:10.1016/j.neubiorev.2010.12.003
- Gaillard, J. M., and Blois, R. (1981). Spindle density in sleep of normal subjects. *Sleep* 4, 385–391.
- Gais, S., Molle, M., Helms, K., and Born, J. (2002). Learning-dependent increases in sleep spindle density. *J. Neurosci.* 22, 6830–6834.
- Harvey, C. J., Gehrmann, P., and Espie, C. A. (2014). Who is predisposed to insomnia: a review of familial aggregation, stress-reactivity, personality and coping style. *Sleep Med. Rev.* 18, 237–247. doi:10.1016/j.smrv.2013.11.004
- Hasson, D., Theorell, T., Bergquist, J., and Canlon, B. (2013). Acute stress induces hyperacusis in women with high levels of emotional exhaustion. *PLoS ONE* 8:e52945. doi:10.1371/journal.pone.0052945
- Iber, C., Ancoli-Israel, S., Chesson, A. L., and Quan, S. F. (2007). *The AASM Manual for the Scoring of Sleep and Associated Events*. Westchester, NY: American Academy of Sleep Medicine.
- Jarrin, D. C., Chen, I. Y., Ivers, H., and Morin, C. M. (2014). The role of vulnerability in stress-related insomnia, social support and coping styles on incidence and persistence of insomnia. *J. Sleep Res.* 23, 681–688. doi:10.1111/jsr.12172
- Jernelov, S., Hoglund, C. O., Axelsson, J., Axen, J., Gronneberg, R., Grunewald, J., et al. (2009). Effects of examination stress on psychological responses, sleep and allergic symptoms in atopic and non-atopic students. *Int. J. Behav. Med.* 16, 305–310. doi:10.1007/s12529-008-9020-6
- LeBlanc, M., Merette, C., Savard, J., Ivers, H., Baillargeon, L., and Morin, C. M. (2009). Incidence and risk factors of insomnia in a population-based sample. *Sleep* 32, 1027–1037.
- Lopez, J., Hoffmann, R., and Armitage, R. (2010). Reduced sleep spindle activity in early-onset and elevated risk for depression. *J. Am. Acad. Child Adolesc. Psychiatry* 49, 934–943. doi:10.1016/j.jaac.2010.05.014
- Lovibond, P. F., and Lovibond, S. H. (1995). The structure of negative emotional states: comparison of the depression anxiety stress scales (DASS) with the beck depression and anxiety inventories. *Behav. Res. Ther.* 33, 335–343. doi:10.1016/0005-7967(94)00075-U

- Lund, H. G., Reider, B. D., Whiting, A. B., and Prichard, J. R. (2010). Sleep patterns and predictors of disturbed sleep in a large population of college students. *J. Adolesc. Health* 46, 124–132. doi:10.1016/j.jadohealth.2009.06.016
- Martin, N., Lafortune, 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
- McKinney, S. M., Dang-Vu, T. T., Buxton, O. M., Solet, J. M., and Ellenbogen, J. M. (2011). Covert waking brain activity reveals instantaneous sleep depth. *PLoS ONE* 6:e17351. doi:10.1371/journal.pone.0017351
- Morin, A., Doyon, J., Dostie, V., Barakat, M., Hadj Tahar, A., Korman, M., et al. (2008). Motor sequence learning increases sleep spindles and fast frequencies in post-training sleep. *Sleep* 31, 1149–1156.
- Riemann, D., Spiegelhalter, K., Feige, B., Voderholzer, U., Berger, M., Perlis, M., et al. (2010). The hyperarousal model of insomnia: a review of the concept and its evidence. *Sleep Med. Rev.* 14, 19–31. doi:10.1016/j.smrv.2009.04.002
- Schabus, M., Dang-Vu, T. T., Albouy, G., Balet, 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
- Schabus, M., Dang-Vu, T. T., Heib, D. P., Boly, M., Desseilles, M., Vandewalle, G., et al. (2012). The fate of incoming stimuli during NREM sleep is determined by spindles and the phase of the slow oscillation. *Front. Neurol.* 3:40. doi:10.3389/fneur.2012.00040
- Schabus, M., Gruber, G., Parapatics, S., Sauter, C., Klosch, G., Anderer, P., et al. (2004). Sleep spindles and their significance for declarative memory consolidation. *Sleep* 27, 1479–1485.
- Schabus, M., Hodlmoser, K., Gruber, G., Sauter, C., Anderer, P., Klosch, G., et al. (2006). Sleep spindle-related activity in the human EEG and its relation to general cognitive and learning abilities. *Eur. J. Neurosci.* 23, 1738–1746. doi:10.1111/j.1460-9568.2006.04694.x
- Spielman, A. J. (1986). Assessment of insomnia. *Clin. Psychol. Rev.* 6, 11–25. doi:10.1016/0272-7358(86)90015-2
- Steriade, M., and McCarley, R. W. (2005). *Brain Control of Wakefulness and Sleep*. New York, NY: Springer.
- Tamaki, M., Matsuoka, T., Nittono, H., and Hori, T. (2008). Fast sleep spindle (13–15 Hz) activity correlates with sleep-dependent improvement in visuomotor performance. *Sleep* 31, 204–211.
- Vgontzas, A. N., Bixler, E. O., Kales, A., Manfredi, R. L., and Tyson, K. (1994). Validity and clinical utility of sleep laboratory criteria for insomnia. *Int. J. Neurosci.* 77, 11–21. doi:10.3109/00207459408986015
- Vgontzas, A. N., Fernandez-Mendoza, J., Liao, D., and Bixler, E. O. (2013). Insomnia with objective short sleep duration: the most biologically severe phenotype of the disorder. *Sleep Med. Rev.* 17, 241–254. doi:10.1016/j.smrv.2012.09.005

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: 28 November 2014; accepted: 27 January 2015; published online: 10 February 2015.

Citation: Dang-Vu TT, Salimi A, Boucetta S, Wenzel K, O'Byrne J, Brandewinder M, Berthomier C and Gouin J-P (2015) Sleep spindles predict stress-related increases in sleep disturbances. *Front. Hum. Neurosci.* 9:68. doi: 10.3389/fnhum.2015.00068 This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2015 Dang-Vu, Salimi, Boucetta, Wenzel, O'Byrne, Brandewinder, Berthomier and Gouin. 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.



Sleep spindle deficits in antipsychotic-naïve early course schizophrenia and in non-psychotic first-degree relatives

Dara S. Manoach^{1,2,3*}, Charmaine Demanuele^{1,2,3}, Erin J. Wamsley^{3,4†}, Mark Vangel^{2,3}, Debra M. Montrose⁵, Jean Miewald⁵, David Kupfer⁵, Daniel Buysse⁵, Robert Stickgold^{3,4} and Matcheri S. Keshavan^{3,4,5}

¹ Department of Psychiatry, Massachusetts General Hospital, Charlestown, MA, USA

² Athinoula A. Martinos Center for Biomedical Imaging, Charlestown, MA, USA

³ Harvard Medical School, Boston, MA, USA

⁴ Department of Psychiatry, Beth Israel Deaconess Medical Center, Boston, MA, USA

⁵ Department of Psychiatry, Western Psychiatric Institute and Clinic, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

Edited by:

Simon C. Warby, Stanford University, USA

Reviewed by:

Carlyle Smith, Trent University, Canada

Roger Godbout, Université de Montréal, Canada

*Correspondence:

Dara S. Manoach, Psychiatric Neuroimaging, Massachusetts General Hospital Charlestown Navy Yard, 149 13th Street, Room 1.111, Charlestown, MA 02129, USA
e-mail: dara@nmr.mgh.harvard.edu

† Present address:

Erin J. Wamsley, Program in Neuroscience, Department of Psychology, Furman University, Greenville, USA

Introduction: Chronic medicated patients with schizophrenia have marked reductions in sleep spindle activity and a correlated deficit in sleep-dependent memory consolidation. Using archival data, we investigated whether antipsychotic-naïve early course patients with schizophrenia and young non-psychotic first-degree relatives of patients with schizophrenia also show reduced sleep spindle activity and whether spindle activity correlates with cognitive function and symptoms.

Method: Sleep spindles during Stage 2 sleep were compared in antipsychotic-naïve adults newly diagnosed with psychosis, young non-psychotic first-degree relatives of schizophrenia patients and two samples of healthy controls matched to the patients and relatives. The relations of spindle parameters with cognitive measures and symptom ratings were examined.

Results: Early course schizophrenia patients showed significantly reduced spindle activity relative to healthy controls and to early course patients with other psychotic disorders. Relatives of schizophrenia patients also showed reduced spindle activity compared with controls. Reduced spindle activity correlated with measures of executive function in early course patients, positive symptoms in schizophrenia and IQ estimates across groups.

Conclusions: Like chronic medicated schizophrenia patients, antipsychotic-naïve early course schizophrenia patients and young non-psychotic relatives of individuals with schizophrenia have reduced sleep spindle activity. These findings indicate that the spindle deficit is not an antipsychotic side-effect or a general feature of psychosis. Instead, the spindle deficit may predate the onset of schizophrenia, persist throughout its course and be an endophenotype that contributes to cognitive dysfunction.

Keywords: sleep, sleep spindles, schizophrenia, cognition, IQ, polysomnography, endophenotype, relatives

INTRODUCTION

Sleep disturbances in schizophrenia have been described since Kraepelin (1919) and are common throughout its course (Lieberman et al., 2005), including in the prodrome (Miller et al., 2003). The presence of sleep disturbances in antipsychotic-naïve and unmedicated patients indicate that they are not merely a side-effect of medications (for meta-analysis see Chouinard et al., 2004). While often viewed as secondary to schizophrenia, as the accompanying psychological distress may itself diminish sleep quality (Benca, 1996), sleep deprivation can precipitate psychosis in vulnerable individuals (Tyler, 1955; Wright, 1993; but see, Kahn-Greene et al., 2007), and there is growing evidence that sleep disturbances can trigger or aggravate a range of psychiatric conditions (Wehr et al., 1987; Ford and Kamerow, 1989; Breslau et al., 1996; Turek, 2005; Huang et al., 2007; Germain et al., 2008; Sateia, 2009). In schizophrenia, sleep disturbances are seen in

high-risk samples (Keshavan et al., 2004; Lunsford-Avery et al., 2013), are anecdotally associated with the initial onset of psychosis and predict psychotic decompensation in remitted patients (Dencker et al., 1986; Benson, 2006). If specific sleep abnormalities that contribute to the initial onset, relapse and manifestations of schizophrenia can be identified, they may serve as targets for intervention to prevent the emergence of schizophrenia, remediate its course and ameliorate core features.

Recent studies have reported that chronic, medicated patients with schizophrenia show a deficit in sleep spindles (Ferrarelli et al., 2007, 2010; Manoach et al., 2010; Seeck-Hirschner et al., 2011; Wamsley et al., 2012), which are a defining feature of non-rapid eye movement (NREM) Stage 2 sleep that are seen on the electroencephalogram (EEG) as brief (~1 s) bursts of synchronous activity in the 12–15 Hz range. This sleep spindle deficit occurred in the context of normal sleep architecture and

Stage 2 spectral power, except in the sigma frequency band, which corresponds to the frequency range of sleep spindles. Here, we analyzed archival sleep data to determine whether young individuals at high genetic risk for schizophrenia (Keshavan et al., 2004) and antipsychotic-naïve early course patients with schizophrenia (Keshavan et al., 2011) have reduced sleep spindles and whether sleep spindle activity is related to cognitive function.

Animal studies point to sleep spindles as a key mechanism of synaptic plasticity, which may mediate memory consolidation during sleep (Rosanova and Ulrich, 2005; Werk et al., 2005). In humans, sleep spindles correlate with measures of intelligence and with sleep-dependent consolidation of both procedural and declarative memory (for review see, Fogel and Smith, 2011). In antipsychotic-naïve patients with schizophrenia, spindle activity is inversely related to reaction time on tests of attention (Forest et al., 2007). In chronic medicated patients with schizophrenia, reduced spindle activity predicts poorer recognition memory for words that were learned prior to sleep (Goder et al., 2008), impaired sleep-dependent motor procedural memory consolidation (Wamsley et al., 2012) and increased severity of positive symptoms (Ferrarelli et al., 2010; Wamsley et al., 2012). In a randomized placebo-controlled trial, chronic, medicated patients with schizophrenia were treated with eszopiclone (Lunesta®, a non-benzodiazepine hypnotic agent that acts on γ -aminobutyric acid (GABA) neurons in the thalamic reticular nucleus (TRN) where spindles are generated) showed a significant increase in spindle number, density and Stage 2 sigma power (Wamsley et al., 2013). These findings suggest that the spindle deficit in schizophrenia is a specific and treatable sleep abnormality that is related to cognitive dysfunction and symptoms.

Prior reports of decreased sleep spindles in chronic medicated patients with schizophrenia leave a number of important questions unresolved. For example, it is not known whether the spindle deficit is related to the pathophysiology of schizophrenia or to treatment with antipsychotic drugs. One study found that only antipsychotic-treated patients with schizophrenia, but not those with other psychotic disorders, showed deficient spindle activity (Ferrarelli et al., 2010) suggesting that the spindle deficit is neither an antipsychotic side-effect nor a general feature of psychosis. In contrast, several studies of unmedicated schizophrenia patients did not find evidence of a spindle deficit: Two studies reported normal spindle density during Stage 2 sleep in 11 (Poulin et al., 2003) and eight (Forest et al., 2007) antipsychotic-naïve patients; another reported normal spindle density in six unmedicated patients (Van Cauter et al., 1991); and one reported increased spindle density in five unmedicated patients (Hiatt et al., 1985). The latter two studies analyzed only selected NREM sleep segments and neither study distinguished between Stage 2 and slow wave sleep, making them difficult to compare with studies measuring spindle activity during all of Stage 2 sleep. Another unresolved question is whether spindle deficits are present in first-degree relatives. To address these questions we examined sleep spindles in young first-degree relatives of patients with schizophrenia and in antipsychotic-naïve patients recently diagnosed with psychosis. Sigma power (12–15 Hz), which is the frequency band of spindles, shows high heritability in twin studies, high inter-individual variability and within-individual

stability over time suggesting that it is a genetically-mediated trait (Ambrosius et al., 2008; De Gennaro et al., 2008). The presence of spindle deficits in first degree relatives, early course and chronic schizophrenia patients would suggest that it is an endophenotype of schizophrenia. Endophenotypes are heritable traits that indicate genetic vulnerability to illness (Gottesman and Gould, 2003). They are associated with illness but are also present in some syndromally-unaffected relatives.

We also investigated the association of sleep spindles with cognitive performance, functional assessments and symptom ratings. Since sleep spindles positively correlate with performance on a range of cognitive measures in both healthy individuals (Fogel and Smith, 2011) and patients with schizophrenia (Forest et al., 2007; Goder et al., 2008; Seeck-Hirschner et al., 2011; Wamsley et al., 2012), we expected to observe similar relations in our experimental and control samples. Based on prior findings, we also expected reduced spindle activity to correlate with positive symptoms in schizophrenia (Ferrarelli et al., 2010; Wamsley et al., 2012).

METHODS

PARTICIPANTS

Demographic and descriptive data are given in **Table 1**. For all samples, potential participants were excluded if they met DSM-IV criteria (American Psychiatric Association, 2000) for current substance abuse or dependence.

Early course participants and controls

Twenty-six inpatients and outpatients were recruited from the Western Psychiatric Institute and Clinic based on having a newly diagnosed psychotic disorder confirmed in consensus meetings led by senior clinicians (MSK, DM) using all clinical data including Structured Clinical Interviews for DSM-IV (SCID, First et al., 1997). Patients were diagnosed with schizophrenia ($n = 15$); major depression ($n = 4$); delusional disorder ($n = 2$); schizoaffective disorder ($n = 2$); bipolar disorder ($n = 2$); mood disorder, NOS ($n = 1$). Patients were characterized with the Scales for the Assessment of Positive and Negative Symptoms (SAPS and SANS, Andreasen, 1983, 1990) and the Global Assessment of Functioning Scale (GAF, American Psychiatric Association, 2000) within a week of the sleep studies. The following neuropsychological assessments were administered: Ammon's Quick Test, a pointing picture vocabulary test, to estimate verbal IQ (Otto and McMenemy, 1965); the Wisconsin Card Sort Test (WCST, Berg, 1948); Trail Making Tests Parts A and B (Reitan, 1958); the Block Design subtest of the Wechsler Adult Intelligence Scale-Revised (WAIS-R, Wechsler, 1981); the Wide Range Achievement Test-Revised, Reading portion (WRAT-R, Jastak and Wilkinson, 1984) and immediate recall of the California Verbal Learning Test (Delis et al., 1988). Supplemental Table 1 presents neuropsychological data.

The 15 early course patients diagnosed with schizophrenia were similar in age, sex, and estimated IQ to the 11 patients with other psychotic disorders, but had completed one less year of education, a difference that was statistically significant (**Table 1**). The early course groups did not differ significantly in ratings of positive or negative symptom severity or global functioning

Table 1 | Demographic characteristics and description of study samples.

	Patients <i>n</i> = 26		Controls <i>n</i> = 25	<i>p</i>	Relatives <i>n</i> = 19	Controls <i>n</i> = 12	<i>p</i>
	SZ <i>n</i> = 15	Other <i>n</i> = 11					
Age	27 ± 7	27 ± 7	27 ± 7	0.96	14 ± 4	15 ± 5	0.61
	28 ± 8	27 ± 7		0.74			
Sex (#/% M) ^a	17/65	16/64	16/64	0.92	9/47	7/58	0.82
	11/73	6/55		0.56			
Education (years)	14 ± 2	16 ± 2	16 ± 2	<0.001*	8 ± 4	9 ± 4	0.63
	13 ± 2	14 ± 2		0.04*			
Est. IQ	99 ± 13	96 ± 12	110 ± 16	0.01*			
	97 ± 10	101 ± 17		0.43			
Parental SES					29 ± 11	52 ± 10	<0.001*

Means ± SD, SZ, schizophrenia; M, male; SES, socio-economic status. *p*-values are based on *t*-tests.

*Significant at *p* ≤ 0.05.

^a*p*-values are based on chi-square tests.

(GAF) (Supplemental Table 2). While they did not differ significantly on neuropsychological measures (Supplemental Table 1), with the exception of WRAT-R reading, which is often used to estimate pre-morbid verbal IQ, schizophrenia patients generally performed at a lower level. Two patients with schizophrenia and three patients with other psychotic disorders reported current cigarette use.

Sleep and cognitive data on these patients were presented in a prior publication (Keshavan et al., 2011). The present report includes a healthy control group for comparison, considers patients with schizophrenia separately from those with other psychoses and measures sleep spindles rather than sigma frequency power.

Twenty-five healthy individuals, screened to exclude a personal history of mental illness and present substance abuse (SCID-Non-patient edition; First et al., 2002), were recruited from the local community by advertisement, word of mouth and presentations to community groups. The healthy controls were matched to the patients for age and sex but had completed significantly more years of education. They were not administered neuropsychological assessments and no information on parental socioeconomic status or cigarette use was available.

First-degree relatives and their controls

A total of 19 children (*n* = 17) and siblings (*n* = 2) of patients with SCID confirmed diagnoses of schizophrenia were recruited by first asking the patient's permission to approach their relative. Relatives were included if they never had a psychotic disorder and were not taking antipsychotic drugs. Thirteen of the high-risk sample were diagnosed with a lifetime history of other disorders: Attention Deficit Disorder (*n* = 5); major depression (*n* = 2); separation anxiety disorder (*n* = 2); oppositional defiant disorder (*n* = 2); and conduct disorder (*n* = 2). One relative was taking amphetamine and dextroamphetamine (Adderall) and another was taking sertraline at the time of the sleep study.

Twelve age, sex and education matched healthy individuals without a personal history of mental illness or any first-degree family members with an Axis I disorder (confirmed by SCID interviews), were recruited from the community (as above) as control participants (Table 1). Control participants had a significantly higher IQ estimates and parental socioeconomic status (SES, Hollingshead, 1965) than the high-risk relatives. Two of the relatives and none of the control participants reported current cigarette use.

Relatives and controls were administered a SCID, SCID-Non patient version or, for children under 15, the children's epidemiological version of the Schedule for Affective Disorders and Schizophrenia (K-SADS-E, Orvaschel and Puig-Antich, 1987). Potential participants with substance abuse within 4 weeks of the initial assessment or alcohol dependence within the previous 2 years were excluded.

Relatives and controls were characterized with the Chapman Psychosis Proneness Scales of Magical Ideation (Eckblad and Chapman, 1983), Perceptual Aberration (Chapman et al., 1978), and Social Anhedonia (Mishlove and Chapman, 1985), the GAF, and the Premorbid Adjustment Scale (PAS, Cannon-Spoor et al., 1982). Relatives had significantly higher ratings of Magical Ideation and Perceptual Aberration and significantly lower GAF and PAS ratings (Supplemental Table 2). The following neuropsychological measures were administered to relatives and controls: the Ammon's Quick Test, the WCST, and the Continuous Performance Test—Identical Pairs version (CPT-IP, Cornblatt et al., 1988) (Supplemental Table 1).

Consent

Experimental protocols were approved by the University of Pittsburgh School of Medicine Institutional Review Board. All participants provided written informed consent (or assent if under 18) following a full description of the study. The parent or guardian also provided informed consent for participants younger than 18.

PROCEDURES

Polysomnography (PSG)

Sleep studies were conducted at the Western Psychiatric Institute and Clinic sleep lab over two consecutive nights. For several days prior to the sleep study, participants were asked to refrain from napping during the day. Sleep times were based on habitual “good night” and “good morning” times, determined using a participant diary of recent sleep patterns. PSG electrodes were placed approximately 1 h before bedtime. Sleep data were acquired at 128 Hz using Grass Telefactor M15 bipolar Neurodata amplifiers and locally-developed collection software. The recording montage consisted of bilateral central (C3 and C4) electroencephalogram (EEG) leads referenced to the linked mastoids (A1+A2); right and left electrooculogram (EOG) referenced to A1+A2; and bipolar submental chin electromyogram (EMG). We analyzed data from the second night. Each 30 s epoch of PSG data was visually classified into stages (Wake, NREM 1, 2, slow wave sleep, and REM) according to standard criteria (Rechtschaffen and Kales, 1968) by a rater blind to diagnostic group. The classified sleep data were segmented into 30 s segments for subsequent data analyses.

Sleep spindle analysis

As in our prior studies, we analyzed spindles during Stage 2 sleep (Manoach et al., 2010; Wamsley et al., 2012, 2013). PSG data were preprocessed and analyzed using BrainVision Analyzer (version 2.0.2, BrainProducts, Munich Germany) and MATLAB (version R2009b, The MathWorks, Natick MA) software. Prior to analysis, data were filtered at 0.3–35 Hz and artifacts were rejected by manual inspection. Discrete sleep spindle events were automatically detected at the C4 lead, which was the only lead available for all participants, using a wavelet-based algorithm that was previously validated against both hand-counted spindles and 12–15 Hz sigma power in both healthy individuals and patients with schizophrenia (Wamsley et al., 2012) and outperformed other available automated spindle detectors by most closely approximating expert consensus spindle counts (Warby et al., 2014).

For each spindle, measures of amplitude, sigma power, duration, and peak frequency were based on analysis of 2 s EEG epochs centered on the point of spindle detection. Within the sigma range (12–15 Hz), *amplitude* was the maximal voltage following 12–15 Hz band pass filtering, *peak frequency* was defined as the spectral peak of the spindle following Fast Fourier transform (FFT) decomposition, and *sigma power* was defined as the mean FFT-derived power spectral density in the 12–15 Hz range ($\mu\text{V}^2/\text{Hz}$). To examine the time-frequency characteristics of individual spindles, wavelet analysis was conducted. A complex Morlet wavelet was applied separately to each spindle epoch. The *duration* of each spindle was calculated as the half-height width of wavelet energy within the spindle frequency range.

We chose spindle density (events/min) and individual spindle amplitude as our primary dependent variables for regressions with cognitive and symptom measures. Spindle density was chosen because it is more resistant to group differences in total sleep time (TST) than spindle number, was deficient in our prior studies of chronic medicated patients and correlated with

sleep-dependent memory consolidation (Manoach et al., 2010; Wamsley et al., 2012). Spindle amplitude was chosen because it negatively correlated with positive symptoms in our prior study (Wamsley et al., 2012) and contributes to the measurement of “integrated spindle activity,” which negatively correlated with positive symptoms in a study from another group (Ferrarelli et al., 2010).

Spectral characterization of stage 2 sleep

The power spectral density ($\mu\text{V}^2/\text{Hz}$) was calculated by FFT, using a Hanning window with 50% overlap applied to successive 3 s epochs of Stage 2 sleep. Spectral power in the slow oscillation (0.5–1 Hz), delta (1–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), sigma band (12–15 Hz), and beta (15–30 Hz) frequency bands was measured.

Spindle density and amplitude in relation to cognition, function, and symptom ratings

Cognitive, function or symptom measurements were regressed on the primary spindle parameters (density and amplitude) using robust regression models, which limit the influence of outliers on the results (Andersen, 2008), as implemented in MATLAB. Group (x_2) and its interaction with spindle parameter (x_1) were included in the model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$. In the early course sample, group refers to schizophrenia patients vs. those with other psychotic disorders (for early course controls cognitive data were not available). In the high-risk sample, group refers to relatives vs. controls. If the group factor (difference in intercepts) and the group by spindle parameter interaction (difference in slopes) are not significant, we report the relations for the pooled group data without factors for group and its interaction with the sleep parameter ($y = \beta_{0p} + \beta_{1p} x$). Otherwise we also report standard linear regression results for each group separately.

RESULTS

EARLY COURSE PARTICIPANTS (TABLE 2)

Sleep quality

Early course patients showed worse sleep quality than controls with significantly less TST, more wake time after sleep onset (WASO), and lower sleep efficiency (TST/total time in bed). Although both groups of early course patients showed disrupted sleep compared with controls, in schizophrenia patients the disruption tended to be worse as indexed by trends toward more WASO and lower sleep efficiency.

Sleep architecture

Relative to controls, early course patients showed a greater percentage of Stage 2 sleep and a reduced percentage of slow wave sleep. This was true of both schizophrenia patients and those with other psychotic disorders.

Spectral characteristics of stage 2 sleep

Relative to controls, early course patients showed reduced slow oscillation, delta, and theta power. Of these, only the reduction in theta power significantly differentiated schizophrenia patients from those with other psychotic disorders.

In the sigma frequency band, which corresponds to the frequency range of sleep spindles, schizophrenia patients showed

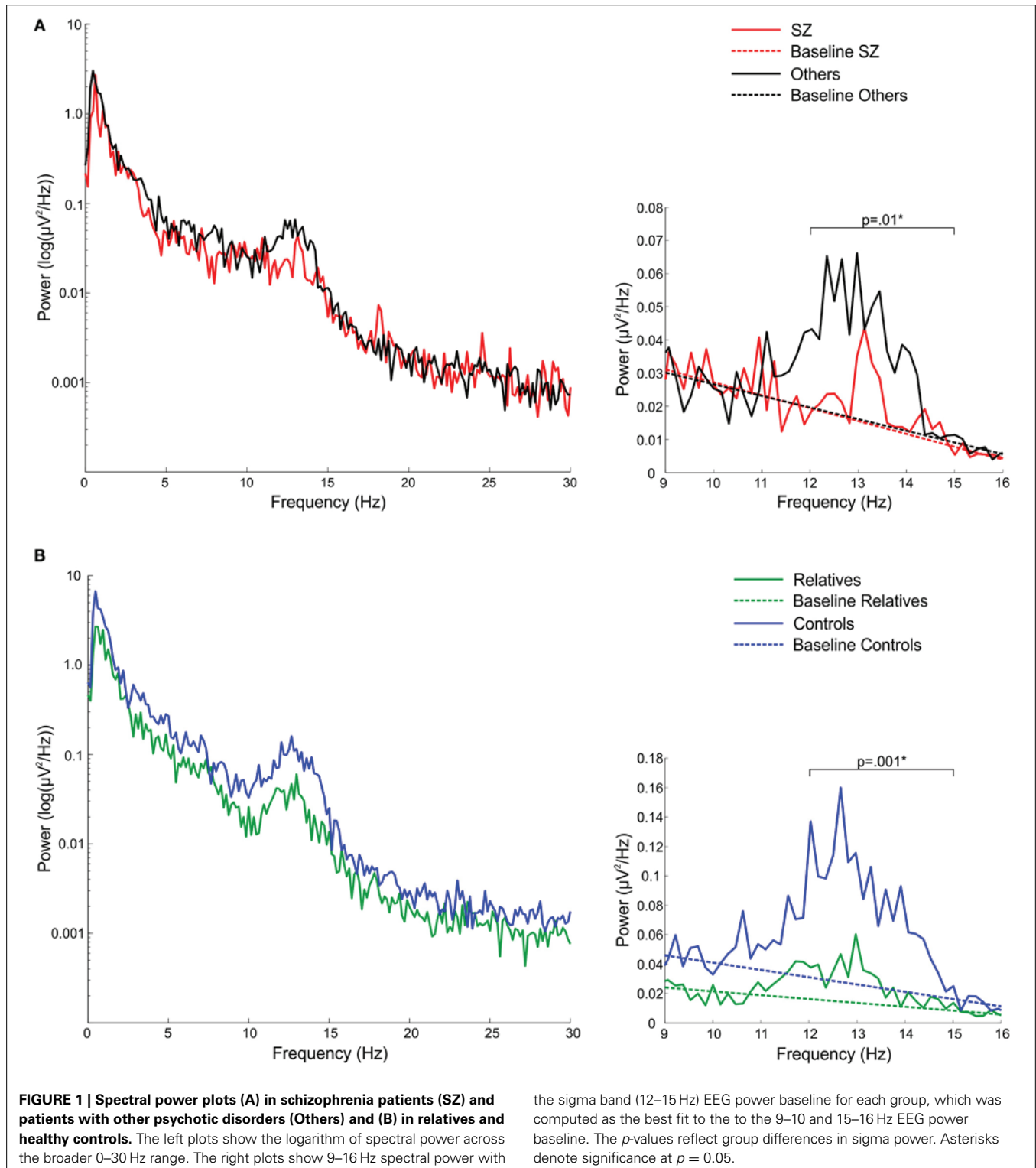
Table 2 | Sleep data for early course patients and their controls reported as means \pm SD.

	Patients <i>n</i> = 26		Controls <i>n</i> = 25	Patients vs. controls		Sz vs. controls		Sz vs. others		Others vs. controls	
	SZ <i>n</i> = 15	Other <i>n</i> = 11		<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>
SLEEP QUALITY											
TST (min)	403 ± 76 411 ± 80	392 ± 73	443 ± 30	−2.51	0.02*	−1.86	0.07	0.62	0.54	−3.08	0.004*
WASO (min)	37 ± 39 49 ± 40	21 ± 33	15 ± 16	2.61	0.01*	3.79	< 0.001*	1.93	0.07	0.69	0.49
Sleep efficiency %	85 ± 11 81 ± 12	89 ± 9	94 ± 4	−3.77	< 0.001*	−4.68	< 0.001*	−1.69	0.10	−2.23	0.03*
SLEEP ARCHITECTURE											
Stage 1 %	4 ± 2 5 ± 2	4 ± 2	4 ± 3	0.04	0.97	0.45	0.65	1.14	0.27	−0.48	0.63
Stage 2 %	62 ± 8 64 ± 7	60 ± 9	52 ± 9	4.41	< 0.001*	4.52	< 0.001*	1.19	0.25	2.56	0.02*
SWS %	7 ± 6 6 ± 5	9 ± 7	17 ± 7	−5.19	< 0.001*	−5.15	< 0.001*	−1.23	0.23	−3.03	0.005*
REM %	26 ± 5 26 ± 6	28 ± 4	27 ± 6	−0.54	0.59	−0.87	0.39	−0.88	0.39	0.14	0.89
SPECTRAL POWER DURING STAGE 2 SLEEP (μV ² /HZ)											
Slow (0.5–1 Hz)	6.2 ± 2.5 6.2 ± 2.2	6.3 ± 3.0	8.4 ± 4.6	−2.12	0.04*	−1.79	0.08	−0.20	0.84	−1.37	0.18
Delta (1–4 Hz)	1.23 ± 0.37 1.15 ± 0.33	1.35 ± 0.41	1.55 ± 0.59	−2.29	0.03*	−2.38	0.02*	−1.33	0.20	−1.03	0.31
Theta (4–8 Hz)	0.16 ± 0.06 0.14 ± 0.05	0.19 ± 0.05	0.23 ± 0.10	−2.92	0.005*	−3.01	0.005*	−2.16	0.04*	−1.26	0.22
Alpha (8–12 Hz)	0.09 ± 0.05 0.08 ± 0.06	0.10 ± 0.04	0.12 ± 0.08	−1.43	0.16	−1.54	0.13	−1.11	0.28	−0.54	0.60
Sigma (12–15 Hz)	0.16 ± 0.1 0.12 ± 0.08	0.22 ± 0.09	0.22 ± 0.12	−1.82	0.08	−2.72	0.01*	−2.68	0.01*	−0.03	0.98
Beta (15–30 Hz)	0.007 ± 0.003 0.007 ± 0.004	0.007 ± 0.002	0.007 ± 0.004	−0.42	0.67	−0.40	0.69	−0.16	0.87	−0.25	0.81
SPINDLE MEASURES											
Spindle number	346 ± 124 314 ± 133	389 ± 101	374 ± 92	−0.92	0.36	−1.68	0.10	−1.56	0.13	0.44	0.66
Spindle density/min	1.41 ± 0.47 1.22 ± 0.48	1.67 ± 0.30	1.64 ± 0.32	−2.01	0.05*	−3.32	0.002*	−2.78	0.01*	0.32	0.75
CHARACTERISTICS OF INDIVIDUAL SPINDLES											
Amplitude (μV)	16.1 ± 4.1 14.8 ± 4.4	17.7 ± 3.1	18.0 ± 4.6	−1.62	0.11	−2.17	0.04*	−1.89	0.07	−0.19	0.85
Frequency (Hz)	13.1 ± 0.9 13.0 ± 0.9	13.2 ± 0.8	13.1 ± 0.9	−0.17	0.87	−0.44	0.66	−0.62	0.54	0.26	0.80
Duration (s)	0.813 ± 0.044 0.800 ± 0.047	0.835 ± 0.031	0.826 ± 0.044	−0.87	0.39	−1.74	0.09	−2.16	0.04*	0.65	0.52
Sigma power (μV ² /Hz)	0.607 ± 0.917 0.637 ± 1.20	0.566 ± 0.315	0.519 ± 0.363	0.45	0.66	0.46	0.65	0.19	0.85	0.37	0.71

TST, total sleep time; WASO, wake after sleep onset; SWS, slow wave sleep; REM, rapid eye movement sleep. Asterisks denote significance at $p \leq 0.05$.

significantly reduced sigma power compared with both other psychotic patients and controls. Psychotic patients with other disorders did not differ from controls in sigma power. So while both patient groups showed reduced spectral power in multiple frequency bands during Stage 2 sleep, only schizophrenia patients

showed a sigma deficit. When calculated relative to the EEG power baseline for each group, computed as the best fit to the 9–10 and 15–16 Hz data, the sigma power (12–15 Hz) in schizophrenia patients was only 27% of that seen in patients with other psychoses (**Figure 1A**).



Sleep spindle parameters

Relative to controls, early course patients showed significantly reduced spindle density (Figure 2A). This reduction was entirely due to the subset of patients diagnosed with schizophrenia who had significantly lower spindle density than both controls and patients with other psychotic disorders whose spindle density was nearly identical to that of controls. Schizophrenia patients also showed reduced spindle amplitude (Figure 2B) and duration compared with controls (trend for duration) and patients with other psychotic disorders (trend for amplitude). Patients with other psychotic disorders did not differ from controls on any spindle parameter.

Spindle density and amplitude in relation to cognition and symptom ratings (Table 3, Figure 3)

In the pooled group of early course patients, lower spindle density was associated with worse cognitive performance on all cognitive measures except immediate recall of the CVLT word list. Lower spindle density significantly predicted slower completion

of Trails A and B, increased perseverative errors on the WCST, lower WRAT-R reading scores and lower estimated verbal IQ. Lower spindle density also predicted lower scaled scores on the Block Design subtest of the WAIS-R, but at a trend level. With the exception of estimated verbal IQ, these relations did not differ significantly as a function of group (schizophrenia, other psychotic disorders). Although the relation with IQ was in the same direction in both groups [significant in the non-schizophrenia psychotic patients: $t_{(9)} = 2.29, p = 0.05$; at a trend level in schizophrenia: $t_{(13)} = 1.95, p = 0.07$], the regression lines differed significantly (Table 3, Figure 4).

Spindle amplitude also correlated with cognitive performance. Like spindle density, lower spindle amplitude was associated with slower performance on Trails B, increased WCST perseverative errors and a lower score on Block Design and these relations did not differ by group. Reduced spindle amplitude also correlated with lower estimated verbal IQ and WRAT-R reading scores in the pooled data, but there was an effect of group reflecting that only the non-schizophrenia psychotic patients showed significant relations of amplitude with estimated IQ [others: $t_{(9)} = 2.76, p = 0.02$; schizophrenia: $t_{(13)} = 0.97, p = 0.35$] and WRAT-R reading [others: $t_{(9)} = 4.44, p = 0.002$; schizophrenia: $t_{(13)} = 0.61, p = 0.56$].

No significant relations between spindle density or amplitude with symptom rating scores or GAF were observed. Because we and another group previously found relations between reduced spindle amplitude (or “integrated spindle activity,” which is influenced by amplitude) and increased severity of positive symptoms in chronic, medicated schizophrenia patients (Ferrarelli et al., 2010; Wamsley et al., 2012), we examined the schizophrenia group alone and found a significant relation in the opposite direction: increased amplitude of individual spindles correlated with increased severity of positive symptoms [$t_{(13)} = 2.21, p = 0.05$] (Figure 5).

Control analyses

In addition to showing lower spindle density and amplitude than both healthy controls and psychotic patients with other diagnoses (trend for amplitude), the sleep of schizophrenia patients was also more disrupted. Sleep efficiency (a general measure of sleep quality), however, did not significantly correlate with spindle density or amplitude in schizophrenia, healthy controls or other psychotic patients, suggesting that sleep disruption is unlikely to account for the spindle deficits. Spindle density and amplitude correlated with multiple measures of cognition in the pooled group of early course patients. Sleep efficiency also correlated with several cognitive measures (Table 3), but notably not with estimated premorbid verbal IQ from the Ammons Quick Test or WRAT-R single word reading, which is also an estimate of premorbid verbal IQ.

FIRST-DEGREE RELATIVES OF SCHIZOPHRENIA PATIENTS

Sleep quality, architecture, and spectral characteristics (Table 4 presents sleep data)

Compared with controls, relatives showed significantly worse sleep quality as indicated by increased WASO and reduced sleep efficiency. Sleep architecture was also disrupted in relatives who showed a greater percentage of time in lighter sleep (Stages 1

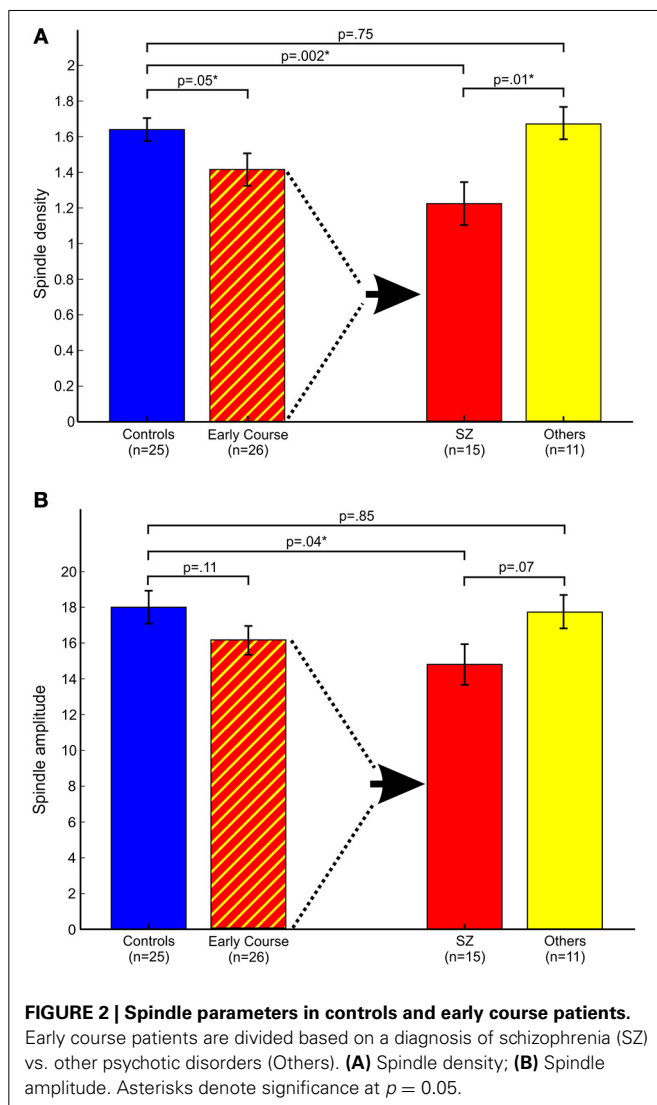


Table 3 | Regressions of cognitive and symptom measures on sleep parameters (spindle density, spindle amplitude, or sleep efficiency) in early course patients.

		Group		Sleep parameter x group		Sleep parameter (pooled data)	
		<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>
Trails A (s)	Spindle density	−0.98	0.34	0.78	0.44	−2.02	0.05*
	Spindle amplitude	−1.34	0.19	1.28	0.21	−0.80	0.43
	Sleep efficiency	−0.54	0.59	0.62	0.54	0.03	0.98
Trails B (s)	Spindle density	0.54	0.60	−0.18	0.86	−5.36	< 0.0001*
	Spindle amplitude	0.21	0.84	0.11	0.91	−5.22	< 0.0001*
	Sleep efficiency	0.45	0.66	−0.19	0.84	−2.13	0.04*
WCST perseverative errors	Spindle density	−0.76	0.46	0.71	0.49	−2.49	0.02*
	Spindle amplitude	−1.95	0.06	2.02	0.06	−2.06	0.05*
	Sleep efficiency	1.28	0.22	−1.22	0.24	−4.70	0.0001*
IQ estimate	Spindle density	2.05	0.05*	−2.14	0.04*	2.96	0.007*
	Spindle amplitude	2.34	0.03*	−2.35	0.03*	2.29	0.03*
	Sleep efficiency	1.57	0.13	−1.58	0.13	1.39	0.18
WRAT-R reading	Spindle density	1.42	0.17	−1.31	0.21	3.19	0.004*
	Spindle amplitude	2.85	0.01*	−2.77	0.01*	2.09	0.05*
	Sleep efficiency	1.54	0.14	−1.52	0.14	1.51	0.14
Block design	Spindle density	−0.12	0.90	0.13	0.90	1.90	0.07
	Spindle amplitude	1.73	0.10	−1.71	0.10	3.12	0.005*
	Sleep efficiency	0.64	0.53	−0.69	0.51	2.37	0.03*
CVLT	Spindle density	−0.06	0.95	−0.20	0.84	0.27	0.79
	Spindle amplitude	0.07	0.94	−0.24	0.81	0.85	0.40
	Sleep efficiency	−0.46	0.65	0.40	0.69	1.84	0.08
SANS	Spindle density	0.89	0.38	−0.57	0.58	0.66	0.52
	Spindle amplitude	−1.27	0.22	1.49	0.15	−0.78	0.44
SAPS	Spindle density	0.99	0.33	−0.79	0.44	0.44	0.66
	Spindle amplitude	−0.81	0.42	1.10	0.28	1.18	0.25
GAF	Spindle density	−0.68	0.50	0.38	0.71	0.25	0.80
	Spindle amplitude	0.32	0.75	−0.60	0.56	0.40	0.69

Group: schizophrenia vs. other psychotic disorders. The Group (x_2) and Sleep parameter (x_1) x Group columns are based on the following model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$. The sleep parameter column is based on the regression using pooled group data without factors for group and its interaction with the sleep parameter ($y = \beta_0 + \beta_1 x_1$). WCST, Wisconsin Card Sort Test; WRAT-R, Wide Range Achievement Test-Revised; Block Design scaled score; CVLT, California Verbal Learning Test standard score for total immediate word recall; SANS, Scale for the Assessment of Positive Symptoms global total; SAPS, Scales for the Assessment of Positive Symptoms global total; GAF, Global Assessment of Functioning.

*Significant at $p \leq 0.05$.

and 2) and trends toward lower percentages of slow wave and REM sleep. During Stage 2 sleep, relatives showed significant power reductions in all frequency bands except delta. To control for this shift in global power, we calculated sigma power relative to the EEG power baseline for each group, computed as the best fit to the 9–10 and 15–16 Hz data (Figure 1B). Sigma power (12–15 Hz) in relatives was only 25% of that seen in healthy controls.

Sleep spindle parameters

Relatives showed significantly reduced amplitude and sigma power of individual spindles, as well as a trend toward reduced

spindle density. Relatives with and without psychiatric diagnoses did not differ in spindle density [$t_{(1, 17)} = 1.16$, $p = 0.26$] or amplitude [$t_{(1, 17)} = 1.16$, $p = 0.26$].

Spindle density and amplitude in relation to cognitive function and symptom ratings

Spindle density correlated with WCST perseverative errors in controls [$t_{(10)} = 3.07$, $p = 0.01$] but not relatives [$t_{(18)} = 0.08$, $p = 0.94$] and not in the pooled data of controls and relatives (Table 5). For the combined groups, spindle amplitude significantly correlated with GAF ($t_{df} = 2.72$, $p = 0.01$), but neither group alone showed this relation and the plot suggested it was

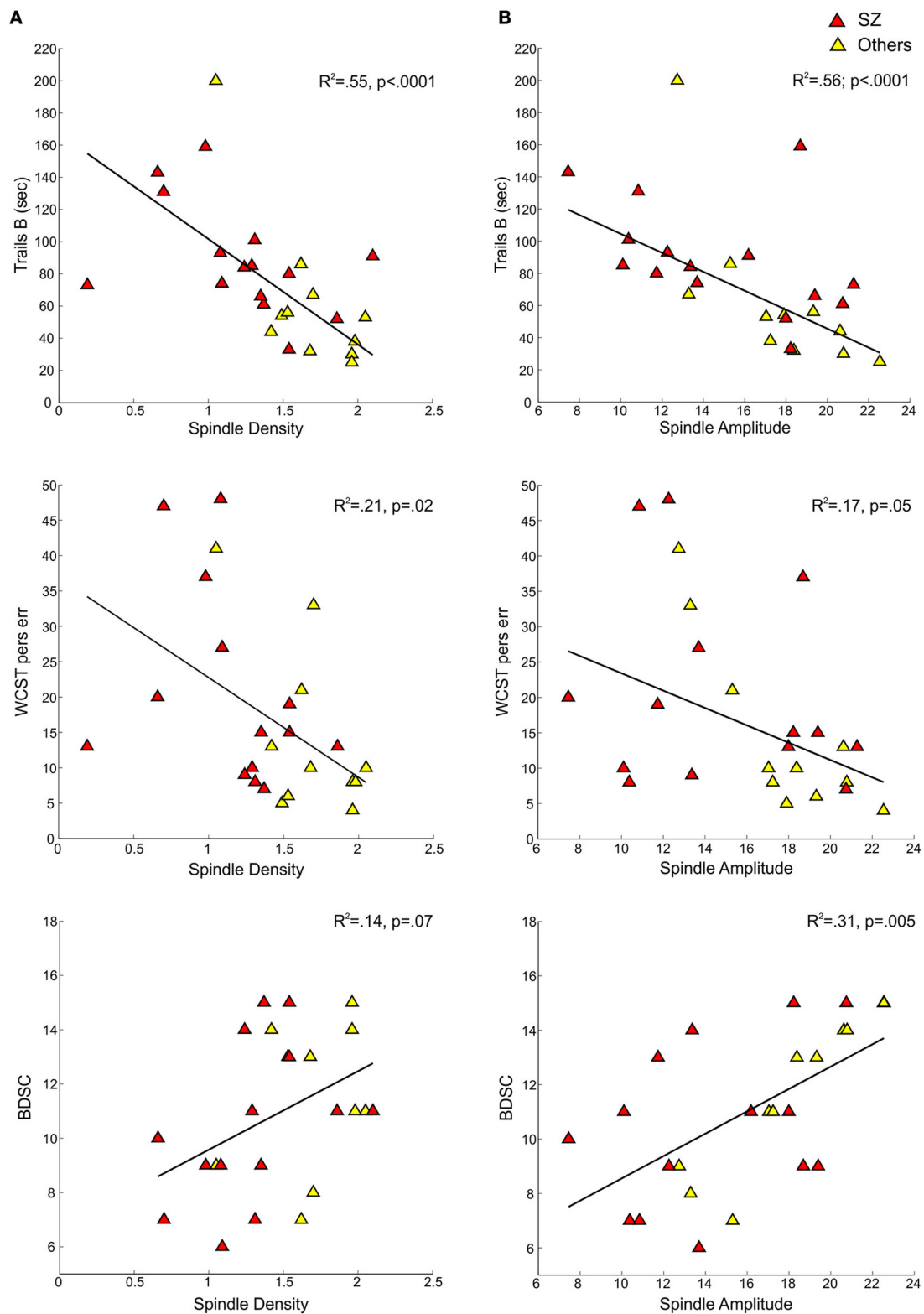


FIGURE 3 | Regressions of cognitive measures on spindle density and amplitude for early course schizophrenia patients (SZ) and those with other psychotic disorders (Others). (A) Shows cognitive measures—Trails

B, WCST perseverative errors, and Block Design scaled score—regressed on spindle density. (B) Shows the same cognitive measures regressed on spindle amplitude.

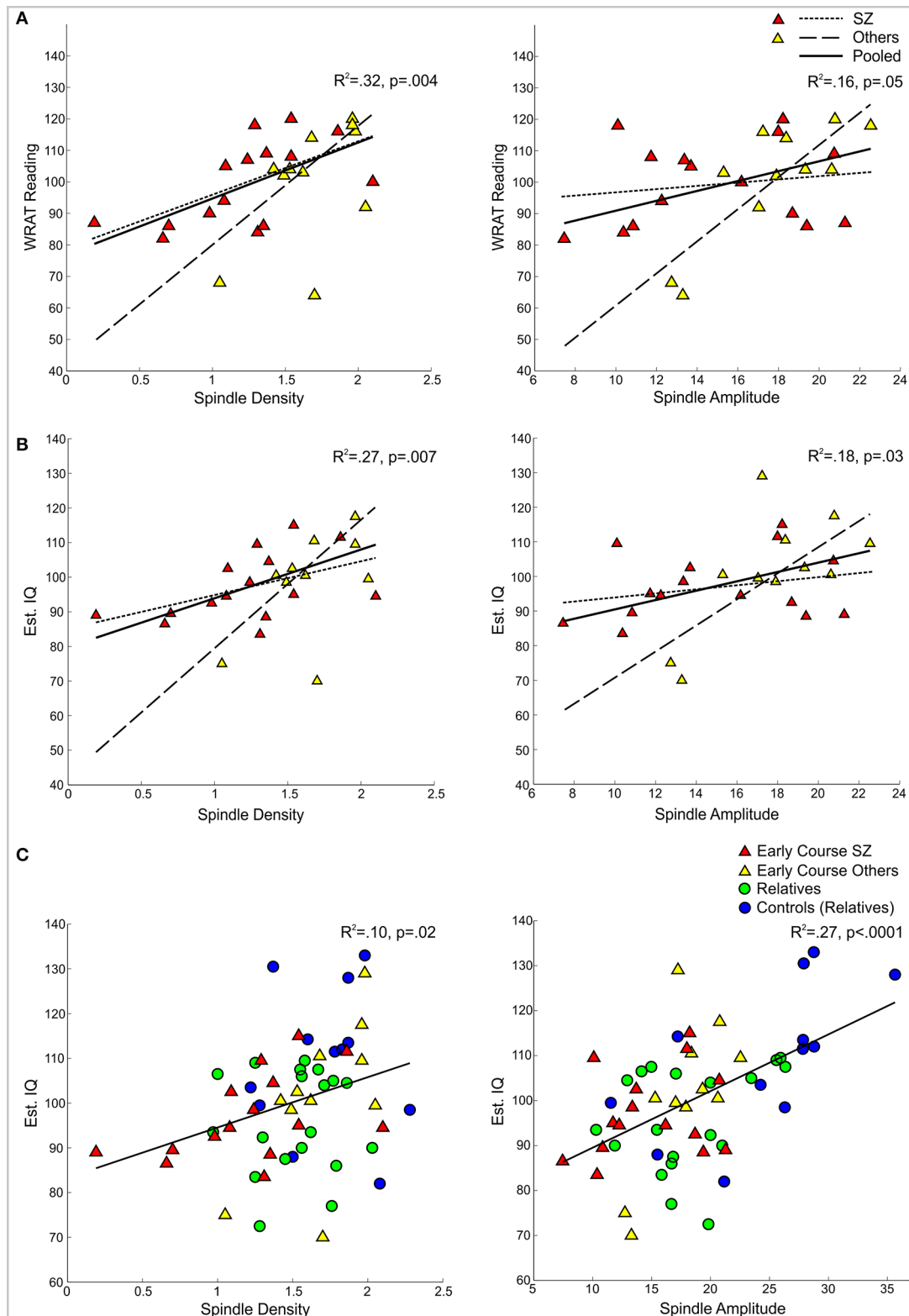
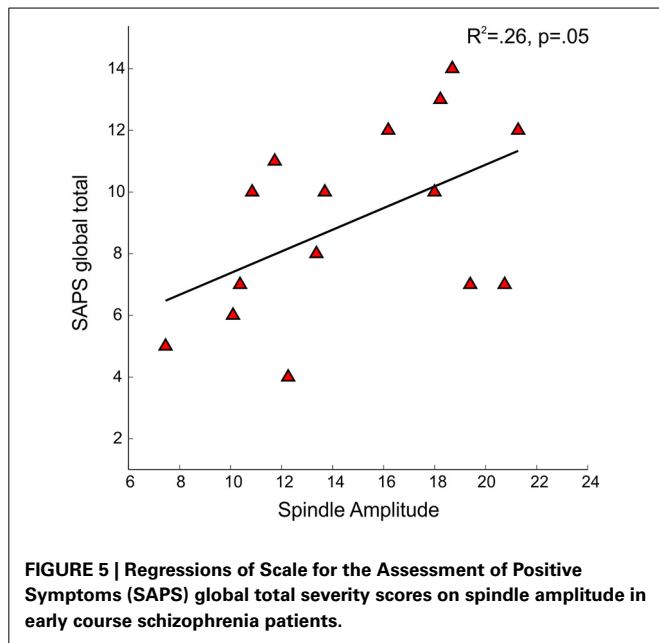


FIGURE 4 | Regressions of WRAT-R Reading standard scores and estimated verbal IQ from the Ammons Quick Test on spindle density and amplitude. (A,B) Early course patients with schizophrenia (SZ) and other psychotic disorders (Others) with regression lines for each group

and the pooled group data. **(C)** Regression of estimated verbal IQ on spindle density and amplitude for the pooled group data from early course schizophrenia, other early course psychotic patients, relatives, and relatives' controls.



due to group differences in both parameters. Spindle amplitude significantly correlated with IQ **Table 5**, (**Figure 3**) and showed a trend level relation with premorbid adjustment. These relations did not differ by group. None of the psychosis proneness ratings correlated with spindle density or amplitude in either the pooled group data or in either group alone.

Control analyses

Like the early course patients with schizophrenia, relatives showed reduced spindle density (trend) and amplitude relative to healthy controls, but their sleep quality was also more disrupted. Sleep efficiency, however, did not significantly correlate with spindle density or amplitude in relatives or their healthy controls, suggesting that the sleep disruption is unlikely to account for the spindle deficits. Nor did sleep efficiency correlate significantly with cognitive measures in relatives, controls, or the pooled group data.

DISCUSSION

The present study provides the first demonstration that both young first-degree relatives of patients with schizophrenia and antipsychotic-naïve patients early in the course of schizophrenia show reduced sleep spindle activity. In contrast, early course psychotic patients with other diagnoses showed normal spindle activity. These findings indicate that the spindle deficit, which was previously reported in chronic, medicated patients with schizophrenia (Ferrarelli et al., 2007, 2010; Manoach et al., 2010; Seeck-Hirschner et al., 2011; Wamsley et al., 2012), is not due to antipsychotic medications, is not a product of chronic illness and is not a general feature of psychosis. Moreover, consistent with growing evidence that links sleep spindles to a range of cognitive functions including intellectual ability in healthy individuals (Fogel and Smith, 2011), the present study found that sleep spindle activity correlated with multiple cognitive measures including estimates of verbal IQ in young healthy controls, early

Table 4 | Sleep data in relatives and controls.

	Relatives <i>n</i> = 19	Controls <i>n</i> = 12	<i>t</i>	<i>p</i>
SLEEP QUALITY				
TST (min)	477 ± 61	511 ± 63	−1.36	0.19
WASO (min)	14 ± 14	5 ± 3	2.29	0.03*
Sleep efficiency %	93 ± 3	96 ± 2	−2.21	0.04*
SLEEP ARCHITECTURE				
Stage 1 %	4 ± 2	2 ± 1	2.01	0.05*
Stage 2 %	52 ± 7	45 ± 11	2.02	0.05*
SWS %	21 ± 8	27 ± 10	−1.78	0.09
REM %	23 ± 4	26 ± 4	−1.70	0.10
SPECTRAL POWER DURING STAGE 2 SLEEP (μV²/HZ)				
Slow (0.5–1 Hz)	10.6 ± 3.8	15.2 ± 6.3	−2.55	0.02*
Delta (1–4 Hz)	2.39 ± 0.97	3.04 ± 1.37	−1.54	0.13
Theta (4–8 Hz)	0.33 ± 0.17	0.49 ± 0.21	−2.38	0.02*
Alpha (8–12 Hz)	0.10 ± 0.05	0.17 ± 0.06	−3.29	0.003*
Sigma (12–15 Hz)	0.19 ± 0.10	0.44 ± 0.27	−3.60	0.001*
Beta (15–30 Hz)	0.008 ± 0.004	0.012 ± 0.004	−2.34	0.03*
SPINDLE MEASURES				
Spindle number	377 ± 89	399 ± 148	−0.52	0.61
Spindle density/min	1.52 ± 0.29	1.72 ± 0.33	−1.76	0.09
CHARACTERISTICS OF INDIVIDUAL SPINDLES				
Amplitude (μV)	18.2 ± 4.7	24.4 ± 6.8	−3.01	0.005*
Frequency (Hz)	12.9 ± 0.6	13.1 ± 0.8	−0.64	0.52
Duration (s)	0.844 ± 0.057	0.857 ± 0.049	−0.61	0.54
Sigma power(μV ² /Hz)	0.555 ± 0.379	1.283 ± 0.847	−3.28	0.003*

Means ± SD; TST, total sleep time; WASO, wake after sleep onset; SWS, slow wave sleep; REM, rapid eye movement sleep.

*Significant at $p \leq 0.05$.

course psychotic patients, and young relatives of schizophrenia patients. Thus, spindle activity was related to cognitive function regardless of diagnosis. Together with prior work documenting a spindle deficit in chronic, medicated patients with schizophrenia that correlates with sleep-dependent memory consolidation (Wamsley et al., 2012), the present findings are consistent with the hypothesis that the spindle deficit is an endophenotype of schizophrenia that predates the onset of schizophrenia, is present throughout its course and affects cognitive function. Although suggestive, our findings are correlative and it is not possible to draw strong conclusions about causal relationships between spindles and cognitive function.

Recent work suggests sleep spindle activity as a potential target for the remediation of cognitive deficits in schizophrenia. Eszopiclone—a non-benzodiazepine sedative hypnotic that acts on the TRN, which generates sleep spindles (Jia et al., 2009)—significantly increased spindle activity compared with placebo in a small sample of chronic medicated schizophrenia patients (Wamsley et al., 2013). While its effect on sleep-dependent memory consolidation was not significant, only the eszopiclone group showed significant overnight improvement on the motor sequence task (Walker et al., 2002). Moreover, in the combined eszopiclone and placebo groups, spindle density predicted this overnight consolidation. These findings raise the possibility that

Table 5 | Regressions of cognitive and symptom measures on spindle parameters in relatives and their controls.

		Group		Spindle × group		Spindle	
		<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>
CPT	Density	0.68	0.50	−0.97	0.34	−1.13	0.27
Verbal	Amplitude	−0.28	0.78	−0.05	0.96	−0.20	0.84
CPT	Density	0.73	0.47	−1.02	0.32	0.50	0.62
Visual	Amplitude	−0.94	0.36	0.34	0.74	−0.13	0.90
WCST	Density	−1.78	0.09	2.01	0.06	−1.29	0.21
Pers err	Amplitude	0.61	0.55	−0.32	0.75	−1.34	0.19
IQ estimate	Density	−0.50	0.62	0.07	0.94	0.63	0.54
	Amplitude	0.44	0.66	−0.79	0.44	3.80	0.0007*
GAF	Density	−2.28	0.03*	0.58	0.57	1.26	0.22
	Amplitude	−1.63	0.12	−0.40	0.69	2.72	0.01*
PAS	Density	1.59	0.12	−1.09	0.28	−0.67	0.51
	Amplitude	0.99	0.33	−0.39	0.70	−1.80	0.08
Magical	Density	0.13	0.90	0.24	0.81	−0.23	0.82
Ideation	Amplitude	0.47	0.64	0.01	0.99	−0.78	0.44
Perceptual	Density	0.95	0.35	−0.64	0.53	−0.51	0.62
Aberration	Amplitude	1.40	0.17	−1.12	0.27	−1.69	0.10
Social	Density	−0.17	0.87	0.19	0.85	−2.01	0.07
Anhedonia	Amplitude	0.15	0.89	−0.01	0.99	−1.62	0.13

Spindle refers to spindle parameter, density, or amplitude. Continuous Performance Test (CPT)—Identical Pairs version; WCST, Wisconsin Card Sort Test perseverative errors; GAF, Global Assessment of Functioning; PAS, Premorbid Adjustment Scale; Chapman Scales of Magical Ideation, Perceptual Aberration, and Social Anhedonia. *Significant at $p \leq 0.05$.

spindle deficits can be effectively treated and that treatment may remediate cognitive deficits. This body of work, identifying abnormal sleep spindles as a potentially treatable candidate endophenotype of schizophrenia that is related to cognitive deficits, opens new avenues for research aimed at understanding, treating, and preventing schizophrenia.

The sleep spindle deficit in schizophrenia implicates dysfunction of thalamocortical circuitry. Sleep spindles are generated in the TRN (Guillery and Harting, 2003) and reduced spindle activity may reflect TRN and/or cortical dysfunction. There is evidence of TRN abnormalities in schizophrenia (Smith et al., 2001) and of reduced thalamic volume in antipsychotic-naïve first-episode schizophrenia (Gilbert et al., 2001). The TRN is comprised entirely of GABAergic neurons (Houser et al., 1980) that primarily inhibit glutamatergic thalamic neurons that project to the cortex. Cortical neurons, in turn, send glutamatergic inputs back to N-methyl-D-aspartate (NMDA) receptors on TRN neurons. Thus, spindles are mediated by a thalamocortical feedback loop that is regulated by both GABAergic and NMDA-receptor mediated glutamatergic neurotransmission (Jacobsen et al., 2001), which are implicated in current models of schizophrenia. In schizophrenia there is evidence of GABA deficits (Thompson et al., 2009) and abnormal expression of NMDA receptors and glutamate transporters in the thalamus (Ibrahim et al., 2000; Smith et al., 2001).

The correlations of spindle activity with IQ in the present samples are similar to what has been reported for healthy individuals in prior work (Fogel and Smith, 2011). Sleep spindles have been linked to a range of cognitive abilities in healthy individuals, particularly to the sleep-dependent consolidation of both procedural

(Walker et al., 2002; Fogel and Smith, 2006; Nishida and Walker, 2007; Peters et al., 2008; Rasch et al., 2008; Tamaki et al., 2008) and declarative (Clemens et al., 2005, 2006; Schabus et al., 2008) memory. Converging evidence suggests that neocortical slow oscillations temporally group thalamocortical sleep spindles with hippocampal ripples thus enabling the redistribution of recently encoded memories from temporary hippocampal to long-term neocortical storage sites (Molle and Born, 2011). The coherent expression of spindles across wide areas of cortex could support the synchronous “reactivation” of recent memory traces across cortical regions (Buzsaki, 1998; O’Neill et al., 2010). In addition to reduced spindle activity, we previously found less coherent spindle activity across the cortex in chronic medicated schizophrenia (Wamsley et al., 2012). This may reflect dysfunction in thalamocortical circuits that could interfere with sleep-dependent memory processing preventing the simultaneous reactivation of memory components stored across visual, spatial, emotional, and goal-representation networks, resulting in the fragmentation of memories and cognition.

Consistent with this, in addition to its relations with estimates of premorbid verbal IQ (the Ammons Quick Test and WRAT-R Reading), sleep spindles correlated with multiple measures of cognitive performance. Sleep efficiency, a general measure of sleep quality, also correlated with cognitive measures in early course patients, but not in relatives, and it was not significantly correlated with IQ estimates or with spindle density or amplitude. This may reflect that while generalized sleep disruption affects the performance of many effortful and attentionally-demanding tasks (Van Dongen et al., 2003), the performance of tasks that primarily tap crystallized knowledge specifically relates to spindles.

Unlike chronic, medicated patients with schizophrenia in whom the sleep spindle reduction was found to be specific [i.e., with the exception of increased sleep onset latency in two studies (Ferrarelli et al., 2007, 2010)], it occurred in the context of normal sleep quality, architecture, and other spectral characteristics of sleep (Manoach et al., 2010; Wamsley et al., 2012)] in both the early course schizophrenia patients and the relatives of schizophrenia patients, sleep was more generally disrupted. Early course patients with other psychotic disorders also showed disrupted sleep relative to controls, but the schizophrenia patients showed greater disruption as indicated by trends toward poorer sleep quality and significantly lower theta power during Stage 2 sleep. But the most compelling difference between schizophrenia patients and those with other psychotic disorders was the significantly reduced spindle activity including spindle density, sigma power and individual spindle duration and amplitude (trend). While schizophrenia patients significantly differed from healthy controls on multiple measures of spindle activity, those with other psychoses did not differ on any. In addition, as sleep efficiency was not significantly correlated with spindle density or amplitude in any group, a general sleep disruption is unlikely to fully account for the spindle deficits observed in schizophrenia patients or in young non-psychotic first-degree relatives. These findings suggest that sleep is disrupted in early course psychotic patients, but only those with schizophrenia show a spindle deficit. Not only was spindle density reduced, but schizophrenia patients also showed abnormal morphology of individual spindles (reduced amplitude and a trend to shorter duration) consistent with some (Ferrarelli et al., 2007, 2010) but not all (Wamsley et al., 2012) studies of chronic medicated patients.

A surprising observation was that positive symptoms were *positively* correlated with spindle amplitude in early course antipsychotic-naïve schizophrenia patients. This contrasts with the negative correlations previously observed in chronic medicated schizophrenia patients (Ferrarelli et al., 2010; Wamsley et al., 2012). This may reflect that the pathophysiological underpinnings of positive symptoms differ in these two populations. In chronic medicated patients, residual positive symptoms have not fully responded to standard dopamine blocking medications and may therefore arise from non-dopaminergic mechanisms such as GABA or NMDA hypofunction (Demjaha et al., 2014), which may also contribute to spindle deficits. In contrast, positive symptoms in early untreated schizophrenia typically respond well to antipsychotics and may reflect dopamine hyperactivity (Keshavan, 1999). These correlations suggest that, in addition to their putative role in cognition, sleep spindles may be related to the expression of schizophrenia symptoms, though the mechanisms of these relations are unknown. Spindle parameters did not correlate with measures of psychosis proneness in the combined group of relatives and their controls, or in the relatives alone.

There are several important limitations of the present study. First, we note that two prior studies of antipsychotic-naïve patients with schizophrenia did not show reduced spindle density during Stage 2 sleep. As in the present study, the sample sizes were relatively small $n = 11$ (Poulin et al., 2003) and $n = 8$ (Forest et al., 2007). Unlike the present study, the spindles were hand counted. This is unlikely to be the source of the discrepancy

since the wavelet-based spindle counting algorithm used for the present study was previously validated against both hand-counted spindles and 12–15 Hz sigma power in both healthy individuals and patients with schizophrenia (Wamsley et al., 2012) and outperformed other available automated spindle detectors by most closely approximating expert consensus spindle counts (Warby et al., 2014). Given this discrepancy it will be important to replicate our findings in larger samples. The small sample sizes of the present study also left us underpowered for some analyses including those involving more complex models that could adjust for the effects of sleep efficiency or IQ on group differences in spindles. As this was an archival study, we were limited to available data and lacked information such as whether the time of day of cognitive and other functional measures was standardized across participants and groups. Because we were missing cognitive and some demographic measures for early course controls we also do not know whether they were well-matched to the early course patients on important demographic features such as parental socioeconomic status. This is a potential confound since the heritability of IQ varies as a function of parental socioeconomic status (e.g., Turkheimer et al., 2003) and IQ correlates with sleep spindles (e.g., Fogel and Smith, 2011). We do know, however, that the early course schizophrenia patients did not differ from other psychotic patients in age, sex, estimated IQ, positive, or negative symptom severity, or on a global functional assessment, yet only the schizophrenia patients showed a spindle deficit.

The group of young relatives had lower parental socioeconomic status than their controls. This may reflect socioeconomic slippage of the parents as a consequence of schizophrenia. The relatives also had lower estimated IQs, worse global function and more magical ideation and perceptual aberration, which may all be reflections of genetic vulnerability to schizophrenia and/or the psychosocial effects of having a first-degree family member with schizophrenia. Given these group differences, we cannot exclude the possibility that rather than reflecting genetic vulnerability to schizophrenia, the spindle deficit in relatives reflects differences in other factors such as IQ.

The present findings raise a number of important questions. Is reduced sleep spindle activity a genetic risk factor that predicts psychosis in high-risk individuals and in the prodromal phase? And, if so, would treating the spindle deficit improve cognition and/or reduce the probability of conversion to frank psychosis? And does the sleep spindle deficit help to illuminate the pathophysiology of pre-morbid stages of schizophrenia? Our findings implicate abnormal function in thalamocortical circuitry even before the onset of illness, which is consistent with a recent report of reduced volume of the thalamus bilaterally that correlated with sleep disturbance in adolescents at ultra high risk for psychosis (Lunsford-Avery et al., 2013). In chronic patients, would treating the spindle deficit improve cognition and symptoms and thereby reduce the risk of relapse?

These questions highlight important directions for future research. Sleep studies are non-invasive and the potential to remediate abnormal sleep for the prevention and treatment of schizophrenia should be examined. The detection of reduced spindle activity as a risk marker for conversion to schizophrenia in high-risk individuals and during the prodromal period

would allow treatment of this deficit. In schizophrenia patients, treatment of the spindle deficit could potentially reduce the clinical, neurocognitive, and functional consequences of illness. In summary, we propose sleep spindles as a potential novel endophenotype and target for research and treatment development.

ACKNOWLEDGMENTS

K24 MH099421, R01 MH092638 (Dara S. Manoach); R01 MH048832, R01 MH092638; R21 MH098171 (Robert Stickgold); R01 MH45203, K02 01180 (Matcheri S. Keshavan).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fnhum.2014.00762/abstract>

REFERENCES

- Ambrosius, U., Lietzenmaier, S., Wehrle, R., Wichniak, A., Kalus, S., Winkelmann, J., et al. (2008). Heritability of sleep electroencephalogram. *Biol. Psychiatry* 64, 344–348. doi: 10.1016/j.biopsych.2008.03.002
- American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders, (4th Edn., text revision)*. Washington, DC: Author.
- Andersen, R. (2008). *Modern Methods for Robust Regression. Series on Quantitative Applications in the Social Sciences*. Thousand Oaks, CA: SAGE Publications, 7–152.
- Andreasen, N. C. (1983). *Scale for the Assessment of Negative Symptoms (SANS)*. Iowa City, IA: University of Iowa.
- Andreasen, N. C. (1990). Methods for assessing positive and negative symptoms. *Mod. Probl. Pharmacopsychiatry* 24, 73–88.
- Benca, R. M. (1996). Sleep in psychiatric disorders. *Neurol. Clin.* 14, 739–764. doi: 10.1016/S0733-8619(05)70283-8
- Benson, K. L. (2006). Sleep in schizophrenia: impairments, correlates, and treatment. *Psychiatr. Clin. North Am.* 29, 1033–1045; abstract: ix–x. doi: 10.1016/j.psc.2006.08.002
- Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *J. Gen. Psychol.* 39, 15–22. doi: 10.1080/00221309.1948.9918159
- Breslau, N., Roth, T., Rosenthal, L., and Andreski, P. (1996). Sleep disturbance and psychiatric disorders: a longitudinal epidemiological study of young adults. *Biol. Psychiatry* 39, 411–418. doi: 10.1016/0006-3223(95)00188-3
- Buzsaki, G. (1998). Memory consolidation during sleep: a neurophysiological perspective. *J. Sleep Res.* 7(Suppl. 1), 17–23. doi: 10.1046/j.1365-2869.7.s1.3.x
- Cannon-Spoor, H. E., Potkin, S. G., and Wyatt, R. J. (1982). Measurement of pre-morbid adjustment in chronic schizophrenia. *Schizophr. Bull.* 8, 470–484. doi: 10.1093/schbul/8.3.470
- Chapman, L. J., Chapman, J. P., and Raulin, M. L. (1978). Body-image aberration in Schizophrenia. *J. Abnorm. Psychol.* 87, 399–407. doi: 10.1037/0021-843X.87.4.399
- Chouinard, S., Poulin, J., Stip, E., and Godbout, R. (2004). Sleep in untreated patients with schizophrenia: a meta-analysis. *Schizophr. Bull.* 30, 957–967. doi: 10.1093/oxfordjournals.schbul.a007145
- 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
- Clemens, Z., Fabo, D., and Halasz, P. (2006). Twenty-four hours retention of visuospatial memory correlates with the number of parietal sleep spindles. *Neurosci. Lett.* 403, 52–56. doi: 10.1016/j.neulet.2006.04.035
- Cornblatt, B. A., Risch, N. J., Faris, G., Friedman, D., and Erlenmeyer-Kimling, L. (1988). The Continuous Performance Test, identical pairs version (CPT-IP): I. New findings about sustained attention in normal families. *Psychiatry Res.* 26, 223–238. doi: 10.1016/0165-1781(88)90076-5
- De Gennaro, L., Marzano, C., Fratello, F., Moroni, F., Pellicciari, M. C., Ferlazzo, F., et al. (2008). The electroencephalographic fingerprint of sleep is genetically determined: a twin study. *Ann. Neurol.* 64, 455–460. doi: 10.1002/ana.21434
- Delis, D. C., Freeland, J., Kramer, J. H., and Kaplan, E. (1988). Integrating clinical assessment with cognitive neuroscience: construct validation of the California Verbal Learning Test. *J. Consult. Clin. Psychol.* 56, 123–130. doi: 10.1037/0022-006X.56.1.123
- Demjaha, A., Egerton, A., Murray, R. M., Kapur, S., Howes, O. D., Stone, J. M., et al. (2014). Antipsychotic treatment resistance in schizophrenia associated with elevated glutamate levels but normal dopamine function. *Biol. Psychiatry* 75, e11–e13. doi: 10.1016/j.biopsych.2013.06.011
- Dencker, S. J., Malm, U., and Lepp, M. (1986). Schizophrenic relapse after drug withdrawal is predictable. *Acta Psychiatr. Scand.* 73, 181–185. doi: 10.1111/j.1600-0447.1986.tb10584.x
- Eckblad, M., and Chapman, L. J. (1983). Magical ideation as an indicator of schizotypy. *J. Consult. Clin. Psychol.* 51, 215–225. doi: 10.1037/0022-006X.51.2.215
- Ferrarelli, F., Huber, R., Peterson, M. J., Massimini, M., Murphy, M., Riedner, B. A., et al. (2007). Reduced sleep spindle activity in schizophrenia patients. *Am. J. Psychiatry* 164, 483–492. doi: 10.1176/appi.ajp.164.3.483
- Ferrarelli, F., Peterson, M. J., Sarasso, S., Riedner, B. A., Murphy, M. J., Benca, R. M., et al. (2010). Thalamic dysfunction in schizophrenia suggested by whole-night deficits in slow and fast spindles. *Am. J. Psychiatry* 167, 1339–1348. doi: 10.1176/appi.ajp.2010.09121731
- First, M. B., Spitzer, R. L., Gibbon, M., and Williams, J. B. W. (1997). *Structured Clinical Interview for DSM-IV Axis I Disorders, Research Version, Patient Edition with Psychotic Screen (SCID-I/P W/PSY SCREEN)*. New York, NY: Biometrics Research, New York State Psychiatric Institute.
- First, M. B., Spitzer, R. L., Gibbon, M., and Williams, J. B. W. (2002). *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Nonpatient Edition*. New York, NY: Biometrics Research, New York State Psychiatric Institute.
- Fogel, S. M., and Smith, C. T. (2006). Learning-dependent changes in sleep spindles and Stage 2 sleep. *J. Sleep Res.* 15, 250–255. doi: 10.1111/j.1365-2869.2006.00522.x
- Fogel, S. M., and Smith, C. T. (2011). The function of the sleep spindle: a physiological index of intelligence and a mechanism for sleep-dependent memory consolidation. *Neurosci. Biobehav. Rev.* 35, 1154–1165. doi: 10.1016/j.neubiorev.2010.12.003
- Ford, D. E., and Kamerow, D. B. (1989). Epidemiologic study of sleep disturbances and psychiatric disorders. An opportunity for prevention? *JAMA* 262, 1479–1484. doi: 10.1001/jama.1989.03430110069030
- Forest, G., Poulin, J., Daoust, A. M., Lussier, I., Stip, E., and Godbout, R. (2007). Attention and non-REM sleep in neuroleptic-naïve persons with schizophrenia and control participants. *Psychiatry Res.* 149, 33–40. doi: 10.1016/j.psychres.2005.11.005
- Germain, A., Buysse, D. J., and Nofzinger, E. (2008). Sleep-specific mechanisms underlying posttraumatic stress disorder: integrative review and neurobiological hypotheses. *Sleep Med. Rev.* 12, 185–195. doi: 10.1016/j.smrv.2007.09.003
- Gilbert, A. R., Rosenberg, D. R., Harenski, K., Spencer, S., Sweeney, J. A., and Keshavan, M. S. (2001). Thalamic volumes in patients with first-episode schizophrenia. *Am. J. Psychiatry* 158, 618–624. doi: 10.1176/appi.ajp.158.4.618
- Goder, R., Fritzer, G., Gottwald, B., Lippmann, B., Seeck-Hirschner, M., Serafin, I., et al. (2008). Effects of olanzapine on slow wave sleep, sleep spindles and sleep-related memory consolidation in schizophrenia. *Pharmacopsychiatry* 41, 92–99. doi: 10.1055/s-2007-1004592
- Gottesman, I. I., and Gould, T. D. (2003). The endophenotype concept in psychiatry: etymology and strategic intentions. *Am. J. Psychiatry* 160, 636–645. doi: 10.1176/appi.ajp.160.4.636
- Guillery, R. W., and Harting, J. K. (2003). Structure and connections of the thalamic reticular nucleus: advancing views over half a century. *J. Comp. Neurol.* 463, 360–371. doi: 10.1002/cne.10738
- Hiatt, J. F., Floyd, T. C., Katz, P. H., and Feinberg, I. (1985). Further evidence of abnormal non-rapid-eye-movement sleep in schizophrenia. *Arch. Gen. Psychiatry* 42, 797–802. doi: 10.1001/archpsyc.1985.01790310059007
- Hollingshead, A. B. (1965). *Two Factor Index of Social Position*. New Haven, CT: Yale University Press.
- Houser, C. R., Vaughn, J. E., Barber, R. P., and Roberts, E. (1980). GABA neurons are the major cell type of the nucleus reticularis thalami. *Brain Res.* 200, 341–354. doi: 10.1016/0006-8993(80)90925-7
- Huang, Y. S., Guilleminault, C., Li, H. Y., Yang, C. M., Wu, Y. Y., and Chen, N. H. (2007). Attention-deficit/hyperactivity disorder with obstructive sleep apnea:

- a treatment outcome study. *Sleep Med.* 8, 18–30. doi: 10.1016/j.sleep.2006.05.016
- Ibrahim, H. M., Hogg, A. J. Jr., Healy, D. J., Haroutunian, V., Davis, K. L., and Meador-Woodruff, J. H. (2000). Ionotropic glutamate receptor binding and subunit mRNA expression in thalamic nuclei in schizophrenia. *Am. J. Psychiatry* 157, 1811–1823. doi: 10.1176/appi.ajp.157.11.1811
- Jacobsen, R. B., Ulrich, D., and Huguenard, J. R. (2001). GABA(B) and NMDA receptors contribute to spindle-like oscillations in rat thalamus *in vitro*. *J. Neurophysiol.* 86, 1365–1375.
- Jastak, S., and Wilkinson, G. (1984). *The Wide Range Achievement Test-Revised: Administration Manual*. Wilmington, DE: Jastak Associates.
- Jia, F., Goldstein, P. A., and Harrison, N. L. (2009). The modulation of synaptic GABA(A) receptors in the thalamus by eszopiclone and zolpidem. *J. Pharmacol. Exp. Ther.* 328, 1000–1006. doi: 10.1124/jpet.108.146084
- Kahn-Greene, E. T., Killgore, D. B., Kamimori, G. H., Balkin, T. J., and Killgore, W. D. (2007). The effects of sleep deprivation on symptoms of psychopathology in healthy adults. *Sleep Med.* 8, 215–221. doi: 10.1016/j.sleep.2006.08.007
- Keshavan, M. S. (1999). Development, disease and degeneration in schizophrenia: a unitary pathophysiological model. *J. Psychiatr. Res.* 33, 513–521. doi: 10.1016/S0022-3956(99)00033-3
- Keshavan, M. S., Diwadkar, V. A., Montrose, D. M., Stanley, J. A., and Pettegrew, J. W. (2004). Premorbid characterization in schizophrenia: the Pittsburgh High Risk Study. *World Psychiatry* 3, 163–168.
- Keshavan, M. S., Montrose, D. M., Miewald, J. M., and Jindal, R. D. (2011). Sleep correlates of cognition in early course psychotic disorders. *Schizophr. Res.* 131, 231–234. doi: 10.1016/j.schres.2011.05.027
- Kraepelin, E. (1919). *Dementia Praecox and Paraphrenia*. Edinburgh: E.S. Livingston.
- Lieberman, J. A., Stroup, T. S., McEvoy, J. P., Swartz, M. S., Rosenheck, R. A., Perkins, D. O., et al. (2005). Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N. Engl. J. Med.* 353, 1209–1223. doi: 10.1056/NEJMoa051688
- Lunsford-Avery, J. R., Orr, J. M., Gupta, T., Pelletier-Baldelli, A., Dean, D. J., Smith Watts, A. K., et al. (2013). Sleep dysfunction and thalamic abnormalities in adolescents at ultra high-risk for psychosis. *Schizophr. Res.* 151, 148–153. doi: 10.1016/j.schres.2013.09.015
- Manoach, D. S., Thakkar, K. N., Stroynowski, E., Ely, A., McKinley, S. K., Wamsley, E., et al. (2010). Reduced overnight consolidation of procedural learning in chronic medicated schizophrenia is related to specific sleep stages. *J. Psychiatr. Res.* 44, 112–120. doi: 10.1016/j.jpsychires.2009.06.011
- Miller, T. J., Zipursky, R. B., Perkins, D., Addington, J., Woods, S. W., Hawkins, K. A., et al. (2003). The PRIME North America randomized double-blind clinical trial of olanzapine versus placebo in patients at risk of being prodromally symptomatic for psychosis. II. Baseline characteristics of the “prodromal” sample. *Schizophr. Res.* 61, 19–30. doi: 10.1016/S0920-9964(02)00440-1
- Mishlove, M., and Chapman, L. J. (1985). Social anhedonia in the prediction of psychosis proneness. *J. Abnorm. Psychol.* 94, 384–396. doi: 10.1037/0021-843X.94.3.384
- Molle, M., and Born, J. (2011). Slow oscillations orchestrating fast oscillations and memory consolidation. *Prog. Brain Res.* 193, 93–110. doi: 10.1016/B978-0-444-53839-0.00007-7
- Nishida, M., and Walker, M. P. (2007). Daytime naps, motor memory consolidation and regionally specific sleep spindles. *PLoS ONE* 2:e341. doi: 10.1371/journal.pone.0000341
- O'Neill, J., Pleydell-Bouverie, B., Dupret, D., and Csicsvari, J. (2010). Play it again: reactivation of waking experience and memory. *Trends Neurosci.* 33, 220–229. doi: 10.1016/j.tins.2010.01.006
- Orvaschel, H., and Puig-Antich, J. (1987). *Schedule for Affective Disorders and Schizophrenia for School-Age Children: Epidemiologic 4th version*. Ft. Lauderdale, FL: Nova University, Center for Psychological Study.
- Otto, W., and McMenemy, R. A. (1965). An appraisal of the ammons quick test in a remedial reading program. *J. Educ. Meas.* 2, 193–198. doi: 10.1111/j.1745-3984.1965.tb00415.x
- Peters, K. R., Ray, L., Smith, V., and Smith, C. (2008). Changes in the density of stage 2 sleep spindles following motor learning in young and older adults. *J. Sleep Res.* 17, 23–33. doi: 10.1111/j.1365-2869.2008.00634.x
- Poulin, J., Daoust, A. M., Forest, G., Stip, E., and Godbout, R. (2003). Sleep architecture and its clinical correlates in first episode and neuroleptic-naïve patients with schizophrenia. *Schizophr. Res.* 62, 147–153. doi: 10.1016/S0920-9964(02)00346-8
- Rasch, B., Pommer, J., Diekelmann, S., and Born, J. (2008). Pharmacological REM sleep suppression paradoxically improves rather than impairs skill memory. *Nat. Neurosci.* 12, 396–397.
- Rechtschaffen, A., and Kales, A. (1968). *A Manual Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*. Bethesda, MD: National Institute of Neurological Diseases and Blindness, Neurological Information Network.
- Reitan, R. M. (1958). Validity of the Trail Making Test as an indication of organic brain damage. *Percept. Mot. Skills* 8, 251–276. doi: 10.2466/pms.1958.8.3.271
- Rosanova, M., and Ulrich, D. (2005). Pattern-specific associative long-term potentiation induced by a sleep spindle-related spike train. *J. Neurosci.* 25, 9398–9405. doi: 10.1523/JNEUROSCI.2149-05.2005
- Sateia, M. J. (2009). Update on sleep and psychiatric disorders. *Chest* 135, 1370–1379. doi: 10.1378/chest.08-1834
- Schabus, M., Hoedlmoser, K., Pecherstorfer, T., Anderer, P., Gruber, G., Parapatics, S., et al. (2008). Interindividual sleep spindle differences and their relation to learning-related enhancements. *Brain Res.* 1191, 127–135. doi: 10.1016/j.brainres.2007.10.106
- Seeck-Hirschner, M., Baier, P. C., Sever, S., Buschbacher, A., Aldenhoff, J. B., and Goder, R. (2011). Effects of daytime naps on procedural and declarative memory in patients with schizophrenia. *J. Psychiatr. Res.* 44, 42–47. doi: 10.1016/j.jpsychires.2009.05.008
- Smith, R. E., Haroutunian, V., Davis, K. L., and Meador-Woodruff, J. H. (2001). Expression of excitatory amino acid transporter transcripts in the thalamus of subjects with schizophrenia. *Am. J. Psychiatry* 158, 1393–1399. doi: 10.1176/appi.ajp.158.9.1393
- Tamaki, M., Matsuoka, T., Nittono, H., and Hori, T. (2008). Fast sleep spindle (13–15 Hz) activity correlates with sleep-dependent improvement in visuomotor performance. *Sleep* 31, 204–211.
- Thompson, M., Weickert, C. S., Wyatt, E., and Webster, M. J. (2009). Decreased glutamic acid decarboxylase(67) mRNA expression in multiple brain areas of patients with schizophrenia and mood disorders. *J. Psychiatr. Res.* 43, 970–977. doi: 10.1016/j.jpsychires.2009.02.005
- Turek, F. W. (2005). Insomnia and depression: if it looks and walks like a duck. *Sleep* 28, 1362–1363.
- Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., and Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychol. Sci.* 14, 623–628. doi: 10.1046/j.0956-7976.2003.psci.1475.x
- Tyler, D. B. (1955). Psychological changes during experimental sleep deprivation. *Dis. Nerv. Syst.* 16, 293–299.
- Van Cauter, E., Linkowski, P., Kerkhofs, M., Hubain, P., L'Hermite-Baleriaux, M., Leclercq, R., et al. (1991). Circadian and sleep-related endocrine rhythms in schizophrenia. *Arch. Gen. Psychiatry* 48, 348–356.
- Van Dongen, H. P., Maislin, G., Mullington, J. M., and Dinges, D. F. (2003). The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep* 26, 117–126.
- Walker, M. P., Brakefield, T., Morgan, A., Hobson, J. A., and Stickgold, R. (2002). Practice with sleep makes perfect: sleep-dependent motor skill learning. *Neuron* 35, 205–211. doi: 10.1016/S0896-6273(02)00746-8
- Wamsley, E. J., Shinn, A. K., Tucker, M. A., Ono, K. E., McKinley, S. K., Ely, A. V., et al. (2013). The effects of eszopiclone on sleep spindles and memory consolidation in schizophrenia: a randomized placebo-controlled trial. *Sleep* 36, 1369–1376. doi: 10.5665/sleep.2968
- Wamsley, E., Tucker, M. A., Shinn, A. K., Ono, K. E., McKinley, S., Ely, A. V., et al. (2012). Reduced sleep spindles and spindle coherence in schizophrenia: mechanisms of impaired memory consolidation? *Biol. Psychiatry* 71, 154–161. doi: 10.1016/j.biopsych.2011.08.008
- 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
- Wechsler, D. (1981). *The Wechsler Adult Intelligence Scale-Revised*. New York, NY: Psychological Corporation.
- Wehr, T. A., Sack, D. A., and Rosenthal, N. E. (1987). Sleep reduction as a final common pathway in the genesis of mania. *Am. J. Psychiatry* 144, 201–204.

- Werk, C. M., Harbour, V. L., and Chapman, C. A. (2005). Induction of long-term potentiation leads to increased reliability of evoked neocortical spindles *in vivo*. *Neuroscience* 131, 793–800. doi: 10.1016/j.neuroscience.2004.12.020
- Wright, J. B. (1993). Mania following sleep deprivation. *Br. J. Psychiatry* 163, 679–680. doi: 10.1192/bjp.163.5.679

Conflict of Interest Statement: Dr. David Kupfer has the following disclosures: Consultant to the American Psychiatric Association (as Chair of the DSM-5 Task Force); holds joint ownership of copyright for the Pittsburgh Sleep Quality Index (PSQI); received honorarium for manuscript submission to *Medicographia* (Servier); he is a member of the Valdoxan Advisory Board of Servier International; he is a stockholder in AliphCom; and he and his spouse, Dr. Ellen Frank are stockholders in Psychiatric Assessments, Inc.

Received: 23 May 2014; accepted: 09 September 2014; published online: 07 October 2014.

Citation: Manoach DS, Demanuele C, Wamsley EJ, Vangel M, Montrose DM, Miewald J, Kupfer D, Buysse D, Stickgold R and Keshavan MS (2014) Sleep spindle deficits in antipsychotic-naïve early course schizophrenia and in non-psychotic first-degree relatives. *Front. Hum. Neurosci.* 8:762. doi: 10.3389/fnhum.2014.00762 This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2014 Manoach, Demanuele, Wamsley, Vangel, Montrose, Miewald, Kupfer, Buysse, Stickgold and Keshavan. 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.



Correlations between adolescent processing speed and specific spindle frequencies

Rebecca S. Nader^{1,2*} and Carlyle T. Smith¹

¹ Department of Psychology, Trent University, Peterborough, ON, Canada

² Department of Psychology, Queen's University, Kingston, ON, Canada

Edited by:

Tore Nielsen, Université de Montréal, Canada

Reviewed by:

Steffen Gais, University of Munich, Germany

Erin J. Wamsley, Furman University, USA

*Correspondence:

Rebecca S. Nader, Department of Psychology, Trent University, 1600 West Bank Drive, Peterborough, ON K9J 7B8, Canada
e-mail: rebeccanader@trentu.ca

Sleep spindles are waxing and waning thalamocortical oscillations with accepted frequencies of between 11 and 16 Hz and a minimum duration of 0.5 s. Our research has suggested that there is spindle activity in all of the sleep stages, and thus for the present analysis we examined the link between spindle activity (Stage 2, rapid eye movement (REM) and slow wave sleep (SWS)) and waking cognitive abilities in 32 healthy adolescents. After software was used to filter frequencies outside the desired range, slow spindles (11.00–13.50 Hz), fast spindles (13.51–16.00 Hz) and spindle-like activity (16.01–18.50 Hz) were observed in Stage 2, SWS and REM sleep. Our analysis suggests that these specific EEG frequencies were significantly related to processing speed, which is one of the subscales of the intelligence score, in adolescents. The relationship was prominent in SWS and REM sleep. Further, the spindle-like activity (16.01–18.50 Hz) that occurred during SWS was strongly related to processing speed. Results suggest that the ability of adolescents to respond to tasks in an accurate, efficient and timely manner is related to their sleep quality. These findings support earlier research reporting relationships between learning, learning potential and sleep spindle activity in adults and adolescents.

Keywords: intelligence, processing speed, stage 2, REM, SWS, spindles, sleep

Sleep spindles are an often used hallmark of Stage 2 sleep; these waxing and waning oscillations are commonly observed with frequencies of between 11 and 16 Hz and have durations of between 0.5 and 3 s (Zeitlhofer et al., 1997; DeGennaro and Ferrara, 2003; Schabus et al., 2007; Peters et al., 2008). This frequency range has been further divided by researchers into slow spindles (often between 11 and 13.5 Hz) and fast spindles (often between 13.6 and 16 Hz; Fogel and Smith, 2011). While these are commonly used ranges, there seems to be no clearly defined frequency range for each type (Fogel and Smith, 2011).

It is, however, generally accepted that sleep spindles occur primarily in Stage 2 sleep. They are considered to be markedly reduced in Stage 3 and virtually absent in Stage 4 when EEG records are visually scored (Rechtschaffen and Kales, 1968; Steriade and McCarley, 2005). Steriade and McCarley (2005) have suggested that spindle activity declines as individuals enter deep slow wave sleep (SWS) and only begins to resume when the individual is entering the lighter stages of sleep and preparing for rapid eye movement (REM). Further, the presence of more than a single spindle, without intervening REMs, during an epoch of REM sleep has been considered to be a Stage 2 arousal (Carskadon and Rechtschaffen, 2000).

Spindles have typically been counted visually as opposed to using an automated methodology (DeGennaro and Ferrara, 2003; Ray et al., 2009) and are generally only counted in Stage 2 sleep. Automated spindle counters have become more reliable over the years. One of the advantages of these systems is the ability to filter

frequencies that are not of interest, and to allow researchers to visually observe frequencies that are of interest. Ray et al. (2009) validated an automated spindle detection system that modified the settings for each individual subject. By assessing each subject's average spindle amplitude and setting the minimum amplitude for that subject at 1.96 standard deviations below that mean, the spindle counter was personalized for everyone, allowing for more accurate assessment. Ray et al. (2009) found an overall sensitivity of 98.96% and a specificity of 88.49% using this personalized method.

This automatic spindle counting technique also allows for easy detection of different spindle types (i.e., slow and fast spindles), as well as allowing for the analysis of spindles during different sleep stages. Visual counts of sleep spindles in SWS often lead to the conclusion that there are very few (if any) sleep spindles in SWS. The possibility exists that the sleep spindles are more prominent in SWS than visual inspection would suggest, due to the large amplitude slow waves that make up SWS. Further, examining REM sleep for spindle activity is normally not considered and only two papers actually report a spindle density during REM sleep (Gaillard and Blois, 1981; Zeitlhofer et al., 1997).

Spindle activity in SWS has been observed by other researchers (e.g., Gaillard and Blois, 1981; Zeitlhofer et al., 1997; Steriade and McCarley, 2005; Peter-Derex et al., 2012), but traditionally, spindle activity is assessed primarily in Stage 2 sleep. Using automatic spindle detectors, which can filter out undesired frequencies, researchers can more easily observe the spindle

activity that occurs in SWS sleep. Gaillard and Blois (1981) for example, found that sleep spindle activity showed no changes from Stage 2 to Stages 3 and 4. In contrast, Zeitlhofer et al. (1997) found that spindle activity showed a significant decrease from Stage 2 to SWS. Both groups of researchers found that there was spindle activity present in REM sleep, although in both cases it was significantly lower than in Stage 2 and SWS (Gaillard and Blois, 1981; Zeitlhofer et al., 1997).

Sleep spindles do show changes across the lifespan, but the majority of studies have focused on young adults. Nicolas et al. (2001) studied spindle characteristics in individuals from 10 years of age to 69 years of age. Nicolas et al. (2001) found that the number, density, and duration of sleep spindles (in Stage 2) declined with age from early adolescence on. The declines that they observed occurred primarily in the first four decades of life, and Nicolas et al. suggested that these changes are due to a long maturation, rather than aging *per se*. Jenni and Carskadon (2004) examined sigma activity in adolescence, and observed a decrease in the power of the sigma frequency range as adolescents mature from pre- to post-puberty. They also observed a shift in the predominant peak of sigma activity, which they suggest is due to maturation of the thalamocortical system (Jenni and Carskadon, 2004). Tarokh and Carskadon (2010) also found that the peak frequency in the sigma band increased from childhood to early adolescence and that there was a decline in the absolute EEG spectral power across both NREM and REM sleep. Tarokh and Carskadon (2010) suggest that this decline is due to the synaptic pruning which is occurring during adolescence (and beyond). The adolescent period seems to be a time when the brain is engaged in a substantial amount of maturation and the spindle activity that occurs during this period, may not be equivalent to the adult activity.

The current study was designed to investigate two aspects of spindle activity. The first goal was to observe any spindle activity occurring in Stage 2 sleep, as well as SWS and REM in order to determine whether the spindle system really is inhibited during these alternate stages or whether this activity is simply obscured by the other frequencies. The second goal was an attempt to identify whether spindle activity is a marker for cognitive ability or intelligence in children. Some research in adults has supported the idea that certain sleep characteristics such as spindle frequency activity (SFA) and even stage 2 sleep itself, are related to intelligence (e.g., Bódizs et al., 2005; Geiger et al., 2011). Other research has suggested that baseline sleep spindles are related to an individual's capacity for memory processing and perhaps an inherent learning aptitude or intelligence (e.g., Nader and Smith, 2001, 2003; Schabus et al., 2006; Fogel et al., 2007; Fogel and Smith, 2011). Indeed a number of studies examining sleep spindle development support this idea. Nicolas et al. (2001) for example, observed that spindle activity decreases from adolescence up to the late 60 s and Petit et al. (2004) report that sleep spindles decline with age in terms of their formation, their frequency and their number. It is possible that these declines are related to an age-related decline in cognitive processing. Petit et al. also report that spindle activity declines in patients with dementia, again linking spindle activity with

cognition. Another observation that led to the examination of spindles being a possible marker for intelligence/ability is that spindles display significant inter-individual variability but are very consistent within the individual (e.g., Gaillard and Blois, 1981; DeGennaro et al., 2005; Fogel and Smith, 2006).

Research conducted by Nader and Smith (2003) and Fogel et al. (2007) demonstrated that spindle activity was positively correlated with Performance IQ, but not with Verbal IQ (see also, Fogel and Smith, 2011). We predicted that our adolescents would show a similar pattern of results, with spindle activity positively correlated with the more procedural scales of the Wechsler Intelligence Scale for Children (WISC), but with no relationship between the verbal scales of the WISC and spindle activity.

METHOD

PARTICIPANTS

The participants were 32 adolescents (17 female) between the ages of 12 and 19 years ($M = 15.36$ years) recruited from the Peterborough community. Participants were all considered to be healthy and medication free, as assessed by their parents, with no indication of sleep disorders. All subjects were assessed for pubertal development in order to control for hormonal effects on spindle activity.

MEASURES

EEG recordings

In-home recordings were made using Suzanne™ (Tyco-Healthcare Group LP, Mansfield, MA, USA) portable polysomnographic systems. The sampling rate was 120 Hz and data were stored on PC flash memory cards, and then downloaded off-line onto a PC computer for further analysis. We recorded EEG, electrooculogram (EOG) (horizontal eye movements only), and EMG using silver-plated electrodes. The EEG (C3, C4, FZ, and PZ) and the EOG (right and left eyes) were monopolar recordings and referenced to contralateral electrodes at A1 and A2. The EMG channel was bipolar. For the EEG and EOG channels, the low- and high-pass software filters were set at 0.03 and 30 Hz. For the EMG channel, only frequencies above 10 Hz were recorded.

Sleep stages were generally scored according to standard criteria (Rechtschaffen and Kales, 1968). However, we sometimes deviated slightly from traditional protocol when scoring the REM sleep stage. The appearance of spindles during REM sleep in the raw EEG was rare, and they only became more visible in the filtered channel. However, according to standard criteria, the observation of a spindle would normally signal an ending to the REM period and the beginning of a period of Stage 2 with the appearance of other Stage 2 indicators. It would also be expected that there would be some increased activity in the EMG channel. If there was absolutely no change in the EMG, no other sign of a Stage 2 intrusion (such as a K-complex) and further REM bursts, the epoch was counted as REM sleep despite the appearance of a spindle. Sleep spindles were counted using the automated spindle counter PRANA® (PhiTools, Strasbourg, France). For each spindle type, an expert technologist identified and recorded the peak amplitudes of 15

spindles in each of the first and second halves of the night for Stage 2 (30 spindles in total for each spindle type). Values were then used to calculate the mean and standard deviation of peak amplitude for each subject. The minimal amplitude criterion for the automated spindle counter was determined by subtracting 1.96 SD units from each mean. This procedure was repeated for each subject. Included in the study were spindle-like waves in the 16–18.50 Hz range. We will use the term “spindle-like” rather than spindle throughout. While these waves share many characteristics of the spindle, their frequencies are in the 16.01–18.50 Hz range. This EEG activity appears to varying degrees in all individuals (Nader et al., 2012a,b,c). The same minimum spindle amplitudes were used in each of the sleep stages (Stage 2, SWS, and REM).

Intelligence was assessed using the Wechsler Intelligence Scale for Children- Fourth Edition (WISC-IV) Canadian Edition. Tests were administered individually by a registered psychometrist. Five participants were assessed by the same psychometrist using the Wechsler Adult Intelligence Scale—Third Edition (WAIS-III) as they were above the age for the WISC-IV.

A number of correlations were performed on sleep spindle activity as it relates to age and IQ in the different sleep stages. Since the measures examined were of the ratio order (spindle density, age) and interval order (IQ scores), we utilized the Pearson correlation. Despite the possible non-normality of some of the data, the non-parametric Spearman's correlation was rejected because reduction of the data to an ordinal level would result in considerable loss of information and power. The Pearson is known to be quite robust, even with non-normal distributions and with our relatively small sample size, was considered the most appropriate.

All subjects were assessed for pubertal development, using the Tanner Scale, in order to control for hormonal effects on spindle activity.

This study was approved by the Trent University Research Ethics Board.

RESULTS

Sleep spindles were detected in all sleep stages (see **Table 1** for densities), not just in Stage 2. Spindle counts varied among the different stages of sleep, with Stage 2 having the highest density and REM having the lowest density. Despite substantial variability (particularly in REM where one individual may have exhibited no spindles at a specific electrode site while another may have exhibited over 100 spindles), a significant number of our adolescents showed spindle activity during REM. In fact, the number of individuals showing more than 30 spindles during REM sleep was substantial, with eight individuals (25%) displaying more than 30 slow spindles and six individuals (19%) showing more than 30 fast spindles during REM sleep. From the visual EEG, it was clear that these young healthy subjects were not exhibiting Stage 2 intrusions into REM sleep during the night (see **Figure 1**).

Interestingly, the 16.01–18.50 Hz spindle-like activity was also observed in all subjects in every phase of the sleep night. While the number of these incidents was less than that for conventional spindle activity, the relative frequency of occurrence

in each sleep stage and density of this event showed several similarities (**Table 1**). For example, the 16.01–18.50 Hz activity showed a similar pattern to the slow spindle (11.00–13.50 Hz) activity, with the greatest number appearing in the frontal region and fewer in the parietal region. This is in contrast to the fast spindle (13.51–16.00 Hz) activity, which was most prominent in the parietal region.

We do not think these waves are artifacts for several reasons. During scoring, all epochs with movement artifacts were discarded. They do not appear in any time locked form that we can see in relation to the other spindle types. Thus we do not think they are some kind of “echo” of the other spindles. Our system is capable of separating these frequency bands such that we do not think they are scoring errors related to spillover activity from spindles in the 13.50–16.00 Hz range. These 16.01–18.50 Hz waves appear to occur on their own time, unrelated to the other two spindle types and can even be seen to occur simultaneously on occasion, suggesting that they are governed by an independent generator. They are more prevalent at the Fz derivation, although they are present at C3, C4 and Pz as well, suggesting their origin is more frontal. They also appear to be smaller in size than spindles measured between 11.00 and 16.00 Hz. There were differences in the amplitudes of the three spindle types in Stage 2. At Fz, for example, the 16.01–18.50 Hz waveforms have significantly smaller average amplitudes (34.27 ± 8.21 uV) than slow (43.95 ± 8.73 uV) or fast (44.84 ± 7.39 uV) spindles which do not differ [$F_{(2,6)} = 45.04$, $p < 0.000001$]. They also have quite different densities (spindles/minute) as can be seen from the **Table 1**. The three frequency ranges were found to have significantly different densities (spindles/minute), with slow spindles being most prevalent (7.14 ± 1.87), then fast spindles (1.45 ± 1.05) and finally the 16.01–18.50 Hz range (0.16 ± 0.18), [$F_{(2,60)} = 300.20$, $p < 0.000001$].

The three frequency ranges were also compared for mean duration in Stage 2 sleep. An ANOVA showed that there was a significant main effect of frequency range, $F_{(2,62)} = 345.415$, $p < 0.000001$. Slow spindles ($M = 1.74$ s) had significantly longer durations than fast spindles ($M = 1.381$ s) and the 16.01–18.50 Hz range ($M = 0.892$ s). Fast spindles also had a significantly longer duration than the 16.01–18.50 Hz waveform. All of these factors lead us to believe that the 16.01–18.50 Hz activity is a separate waveform worthy of further investigation.

SLEEP SPINDLE ACTIVITY AND AGE

Sleep spindle density was correlated with age in order to determine whether the appearance of spindles in the various stages varied with age. The density of slow spindles in Stage 2 sleep was negatively correlated with age in the frontal region (FZ; $r_{(30)} = -0.37$, $p < 0.05$). The density of fast spindles at C4 in Stage 2 was positively correlated with age ($r_{(30)} = 0.35$, $p < 0.05$). This suggests that there may be a tendency for the density of slow spindles in Stage 2 to decline with age and a tendency for the density of fast spindles to increase with age across adolescence. There were no other significant correlations with age, suggesting that the spindle densities in SWS and REM are not strongly related to the age range in

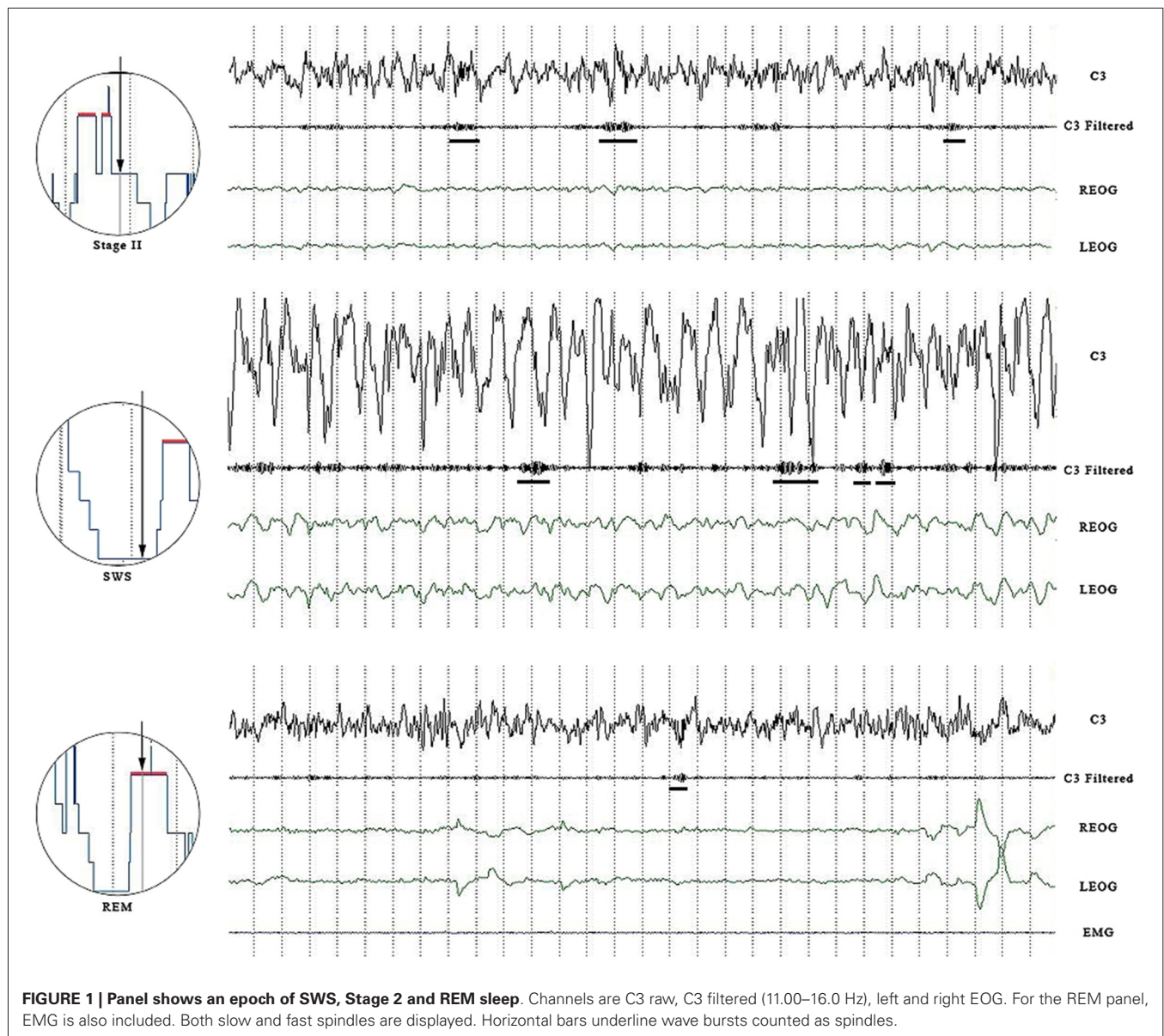


FIGURE 1 | Panel shows an epoch of SWS, Stage 2 and REM sleep. Channels are C3 raw, C3 filtered (11.00–16.0 Hz), left and right EOG. For the REM panel, EMG is also included. Both slow and fast spindles are displayed. Horizontal bars underline wave bursts counted as spindles.

this adolescent group. It also suggests that age may not be a factor in the appearance of the activity in the 16.01–18.50 Hz range.

We also examined the relationships between age and spindle duration in Stage 2 sleep. Correlations showed that the duration of the slow spindles showed a significant decline with age in three of our four derivations [C3: $r_{(30)} = -0.38$, $p < 0.05$; C4: $r_{(30)} = -0.48$, $p < 0.01$; FZ: $r_{(30)} = -0.50$, $p < 0.005$]. Spindle amplitude showed a similar pattern of results, slow spindle amplitude declined at C4 ($r_{(30)} = -0.48$, $p < 0.01$). Fast spindle amplitude declined significantly at all four of our electrode locations [C3: $r_{(30)} = -0.41$, $p < 0.05$; C4: $r_{(30)} = -0.56$, $p < 0.001$; FZ: $r_{(30)} = -0.36$, $p < 0.05$; PZ: $r_{(30)} = -0.49$, $p < 0.01$] in Stage 2 sleep. The 16.01–18.50 Hz waveform showed a similar trend toward an age related decline in Stage 2 sleep, but was only significant at C4 ($r_{(30)} = -0.39$, $p < 0.05$). Taken together,

these results suggest that there is an age related decline in spindle amplitude for all frequencies during the adolescent period.

Similar to age, there were no significant correlations between pubertal development (Tanner Stages) and spindle density (11.00–13.50 Hz and 13.51–16.00 Hz) in Stage 2, SWS and REM. However, when the activity in the 16.01–18.50 Hz range was correlated with the Tanner stages, there were some significant relationships observed. The spindle density in this frequency range was found to be significantly, positively related to pubertal development in Stage 2 in C3 ($r_{(27)} = 0.39$, $p < 0.05$), C4 ($r_{(28)} = 0.42$, $p < 0.05$) and PZ ($r_{(27)} = 0.42$, $p < 0.05$) and to show a trend toward a positive relationship in FZ ($r_{(28)} = 0.34$, $p < 0.10$).

A similar pattern was observed in SWS, where the density of activity in the 16–18.5 Hz range was significantly, positively related to Tanner Stage. This positive relationship was observed

Table 1 | Mean density (spindles/minute) and mean number (\pm SD) of sleep spindles or spindle-like activity in Stage 2, SWS and REM.

	Number								
	11.00–13.5 Hz			13.51–16 Hz			16.01–18.5 Hz		
	Stage 2	SWS	REM	Stage 2	SWS	REM	Stage 2	SWS	REM
C3	1656.36 (519.84)	793.68 (604.42)	21.69 (34.38)	348.97 (291.76)	104.34 (104.95)	12.55 (17.12)	36.16 (41.12)	17.06 (33.46)	10.74 (20.56)
C4	1557.19 (545.73)	688.28 (535.08)	19.25 (29.03)	442.41 (368.60)	139.09 (107.94)	16.25 (32.36)	40.41 (46.47)	20.59 (35.58)	12.19 (23.63)
FZ	1776.09 (448.50)	1019.88 (624.16)	19.63 (18.85)	365.66 (280.72)	126.63 (88.21)	19.06 (32.44)	42.31 (54.49)	12.28 (19.10)	18.13 (33.38)
PZ	1568.94 (468.79)	646.97 (540.44)	15.71 (23.72)	437.94 (362.27)	142.61 (155.82)	5.16 (5.96)	19.26 (24.17)	11.68 (20.13)	4.03 (6.90)

	Density								
	11–13.5 Hz			13.51–16 Hz			16.01–18.5 Hz		
	Stage 2	SWS	REM	Stage 2	SWS	REM	Stage 2	SWS	REM
C3	6.65 (2.22)	5.86 (3.77)	0.18 (0.28)	1.38 (1.09)	0.77 (0.63)	0.11 (0.14)	0.14 (0.16)	0.11 (0.18)	0.10 (0.19)
C4	6.21 (2.10)	5.05 (2.94)	0.15 (0.21)	1.74 (1.35)	1.08 (0.78)	0.13 (0.21)	0.15 (0.18)	0.14 (0.20)	0.11 (0.21)
FZ	7.14 (1.87)	7.54 (3.77)	0.17 (0.16)	1.45 (1.05)	0.97 (0.55)	0.18 (0.33)	0.16 (0.18)	0.08 (0.11)	0.18 (0.34)
PZ	6.24 (1.82)	4.74 (3.19)	0.12 (0.15)	1.73 (1.38)	1.05 (0.98)	0.05 (0.06)	0.07 (0.09)	0.08 (0.11)	0.04 (0.07)

in C4 ($r_{(28)} = 0.39$, $p < 0.05$) and PZ ($r_{(27)} = 0.40$, $p < 0.05$) and a trend toward this positive relationship was observed in C3 ($r_{(27)} = 0.33$, $p < 0.10$). The activity in REM sleep did not vary with pubertal development as measured by the Tanner Stages.

SPINDLE ACTIVITY AND IQ

Spindle activity in Stage 2, SWS and REM was correlated with the Full Scale IQ from the WISC-IV. We did not expect to see any significant correlations between spindle density and any of the Verbal subscales (e.g., verbal comprehension), although we did expect that there would be significant correlations with Picture Completion (perceptual organization) and Processing Speed (see Fogel and Smith, 2011). Consequently, we confined our correlations to Full Scale IQ and these two procedural traits. There was only one significant correlation between the EEG activity and Full Scale IQ (in SWS, density of 13.5–16.00 Hz activity was negatively related to Full Scale IQ; $r_{(32)} = -0.351$, $p < 0.05$).

However, a pattern of significant correlations emerged when the procedural IQ subscales were examined. Since Age was observed to be positively related to both Processing Speed and percentage of SWS, partial correlations, controlling for age, were conducted between these subscales and the various EEG frequencies. Processing speed appeared to be highly related to Spindle Density, particularly during REM and SWS. **Table 2** presents the pattern of significant correlations between Processing Speed and EEG activity. An estimated total of 108 Pearson correlations were run [Derivation (4), Spindle Type (3), Sleep Stage (3)]. While this could be considered to be a large number of correlations requiring some kind of correction for Type

Table 2 | Partial correlations (controlling for Age) between spindle densities (spindles/minute) and processing speed, organized by frequency and sleep stage.

	Stage 2	SWS	REM
C3 11.00–13.5 Hz	0.47*	0.09	0.42*
C3 13.5–16 Hz	−0.04	−0.01	0.43*
FZ 13.5–16 Hz	0.03	0.01	0.53*
C3 16–18.5 Hz	0.33	0.45*	0.19
C4 16–18.5 Hz	0.26	0.32*	0.17
FZ 16–18.5 Hz	0.27	0.37*	0.31
PZ 16–18.5 Hz	0.22	0.46*	0.08

* $p < 0.05$.

I Error, we did not do so for several reasons. Our sample size was quite small and thus applying such corrections as Bonferroni would have been too conservative. Further, the spindle types, EEG derivations and sleep states are undoubtedly not completely independent of each other and this reduces the need for correction. Also, the consistent patterns in the results suggest that these findings are not random and do warrant further examination. While our predictions were partially confirmed, this is an exploratory study and the data provide new research directions.

As the bulk of the significant correlations seemed to be between EEG activity in REM and SWS, with only a single significant correlation in Stage 2, a regression analysis was conducted using SWS and REM activity. A regression analysis was performed on Processing speed, with age being entered first to control for its effects ($R = 0.484$, $p < 0.01$). Proportion of SWS and proportion of REM sleep were entered into the equation next ($R = 0.654$, $p < 0.01$). Finally the variables C3

(11.00–13.50 Hz) REM, C3 (13.5–16.00 Hz) REM, FZ (13.5–16.00 Hz) REM, C3 (16–18.50 Hz) SWS, FZ (16–18.50 Hz) SWS, PZ (16–18.50 Hz) SWS were entered into the equation ($R = 0.803$, $p < 0.01$). The regression analysis suggests that the measures of age and sleep EEG account for 64.6% (adj. $R^2 = 0.486$) of the variance in processing speed. While it is not surprising that age contributes a large proportion of the variance, the results underline the importance of the activity in REM and SWS rather than Stage 2.

DISCUSSION

Spindles are not limited to Stage 2 sleep and appear in all the sleep stages of healthy adolescents. It is possible that the appearance of spindles in SWS and REM could be due to a developmental process of adolescence, but correlations of spindle density with age and puberty showed few consistent significant relationships, with the exception of a positive relationship between density of the faster wavelengths in SWS and Tanner stage. This suggests that the appearance of spindles in REM and SWS may not be a consequence of development, and instead may be a consistent robust phenomenon. It is possible that the faster wavelengths increase in density with pubertal development. As the brain undergoes its substantial maturation during adolescence, it may become more physically able to produce these faster wavelengths during SWS.

Our results do suggest that, despite the lack of consistent changes in density, other measures of spindle activity may be changing across the adolescent age range. Slow spindle duration declines over adolescence in Stage 2, while there is no change in the duration of the fast and 16.01–18.5 Hz waveforms. Amplitude declined in all three frequency bands across adolescence in Stage 2. These results seem to be in agreement with those of Nicolas et al. (2001), Jenni and Carskadon (2004) and Tarokh and Carskadon (2010).

The density of spindles during SWS is significantly less than the spindle density observed during Stage 2, but it is still quite substantial (see **Figure 1**). Our results of a decreased density in SWS is consistent with the findings of both Zeitlhofer et al. (1997) and Andrillon et al. (2011) who observed a significant decrease in spindle density from Stage 2 to Stage 3 to Stage 4. Despite a difference in the electrodes used to measure EEG (Andrillon et al. used depth electrodes), the results from the present study are in agreement with the findings of Andrillon et al. (2011) who observed more slow spindles in the frontal region than in the parietal region and more fast spindles in the parietal region than in the frontal region.

The density of sleep spindles in REM is low, but certainly not completely absent. Our data corresponds very closely to data reported by Gaillard and Blois (1981) who examined spindle activity in adults. Using a filter system, which isolated frequencies between 11.6 and 17.2 Hz, Gaillard and Blois found spindle densities in REM sleep that were very similar to the results presented here. These researchers found a spindle density of 0.87 spindles per minute (± 1.74), supporting the idea that while there is great variability in the number of spindles that occur during REM, they are certainly not absent during this stage of sleep. Our results are also similar to those found by

Zeitlhofer et al. (1997), although our spindle densities in REM ($M = 0.18$ at C3) were lower than their findings ($M = 1.3$ spindles/minute).

The appearance of spindles in REM suggests that the mechanism that produces sleep spindles is not completely inhibited or absent during REM. It is possible that the separate phasic systems that produce eye movements and spindle activity cannot occur simultaneously, but apparently they can occur in close succession. While it has previously been accepted that the appearance of sleep spindles in REM is actually a Stage 2 intrusion (Carskadon and Rechtschaffen, 2000), it seems unlikely that one-quarter of our healthy, young subjects would have more than 30 Stage 2 intrusions during the REM period. In fact, our visual scoring procedure revealed no spindles and certainly no Stage 2 intrusions (Carskadon and Rechtschaffen, 2000). It was only when we filtered out the other frequencies that we were able to count the spindles occurring in REM sleep. The spindles in REM sleep met the same criteria for amplitude and duration, as did the spindles in Stage 2 in order to be counted.

The data suggest that, with the advent of more sophisticated measuring techniques, spindles that occur during SWS and REM are phenomena that have been mostly overlooked, because they were not easily observable. In future, it would be valuable to include all of the sleep stages as well as to examine possible sleep spindle activity in frequency ranges from 11–19 Hz.

While spindle density, in any frequency range, was not very strongly related to age (at least within our adolescent age group), it does appear that pubertal development plays a role in the appearance of some of these spindles. An increase in density of the activity in the 16.01–18.50 Hz range was observed in conjunction with an increase in pubertal development in both Stage 2 and SWS. It is possible that this activity is related to maturity and brain development and may be involved in the establishment of higher order cognitive abilities. A tentative hypothesis is that the link between IQ measures and sleep states develops over the adolescent period, as the brain matures to its adult state.

The sleep measures were correlated with Full Scale IQ, and its subscales to try and assess whether there are any biological markers for intelligence in adolescents. Research in adults has suggested that sleep spindle activity may be linked with an aptitude for learning (Nader and Smith, 2001; Schabus et al., 2006; Fogel et al., 2007; Fogel and Smith, 2011). Given the substantial development in the brain that occurs over the adolescent period, we were interested in whether there was any support for this relationship in adolescents. Dang-Vu et al. (2010) observed that the faster spindle activity was associated with more extensive cortical activation; the results from our adolescents suggest that it is the faster (or higher frequency) brain activity that is associated with some forms of intelligence.

Using the WISC-IV (or WAIS-III), Full scale IQ and the procedural subscales were correlated with brain wave activity in three frequency ranges. The first two ranges, 11.00–13.50 Hz, and 13.5–16.00 Hz, are traditional spindle frequencies; the third frequency range (16–18.50 Hz) is above the normal spindle range, but we observed consistent activity in these frequencies

in the current sample and in an earlier study (Nader and Smith, 2001) and felt it was important to include this activity in our examination. While Full Scale IQ was not related to any of the measured brain activity, some of the subscales were. Processing Speed in particular, seemed to be strongly related to the brain wave activity during sleep. Processing speed is a skill that is linked to executive functioning (Jacobson et al., 2011) and requires individuals to be able to complete a task accurately and as quickly as possible. Processing Speed was observed to be positively associated with age in this adolescent group; this positive association may be due to the development of executive functioning that occurs in adolescence. Executive functioning involves the ability to plan, coordinate, and execute behavior (Blakemore and Choudhury, 2006), processing speed requires the individual to not only perform at a rapid pace, but to be able to respond both efficiently and accurately (Jacobson et al., 2011). This ability requires the individual to plan and prepare for stimulus orientation and appropriate responses (Jacobson et al., 2011).

Performance on the Processing Speed task was observed to be positively related to age, the proportion of both SWS and REM sleep, and the EEG activity that occurs during these stages. Due to these observed relationships, a regression analysis was run to determine how much variance in Processing Speed scores could be accounted for by these sleep variables. Using this exploratory analysis, we were able to account for 64.5% of the variance in Processing Speed by knowing age, proportion of REM and SWS, the density of activity in the 16.00–18.50 Hz range during SWS and the density of activity in the 11.00–13.50 Hz, and 13.51–16.00 Hz ranges during REM. This suggests that the ability of adolescents to respond in an efficient and accurate manner to the task at hand is related to sleep quality. In fact, it may be that the spindle activity during REM and SWS is indicative of their Processing Speed abilities. Since we were not able to predict Full Scale IQ, these results suggest that only specific components of intelligence are related to sleep state activity. This supports research conducted with adolescents and adults, which has suggested that some measures of IQ are related to sleep and also that sleep spindles are related to how well we learn (e.g., Nader and Smith, 2003; Schabus et al., 2006; Fogel et al., 2007).

There are some things that should be considered in future studies. As this was an exploratory study with a limited number of participants, further research needs to be done to confirm the findings presented here. The small number of participants did limit power, and we did not apply any correction procedures for Type I error. However, we did limit the number of correlations performed and only examined the relationships between spindle activity and the Performance IQ/Procedural tasks, along with Full Scale IQ. Further research would be able to use a larger sample and correct for Type I error. Also, the scoring system (Ray et al., 2009) was developed using the EEG from young adults and validating it for younger participants would be valuable. It is possible that the number of false positives might have been different in our younger participants. We can only say that there was no consistent increase in the spindle count as we looked at younger subjects. Depending on spindle type, some were positively correlated with age while

some were negatively correlated with age or not correlated at all. This suggests that there was no general increase in false positives.

REFERENCES

- Andrillon, T., Nir, Y., Staba, R. J., Ferrarelli, F., Cirelli, C., and Tonini, G. (2011). Sleep spindles in humans: insights from intracranial EEG and unit recordings. *J. Neurosci.* 31, 17821–17834. doi: 10.1523/jneurosci.2604-11.2011
- Blakemore, S.-J., and Choudhury, S. (2006). Development of the adolescent brain: implications for executive function and social cognition. *J. Child Psychol. Psychiatry* 47, 296–312. doi: 10.1111/j.1469-7610.2006.01611.x
- Bódizs, R., Kis, T., Lazar, A. S., Havran, L., Rigo, P., Clemens, Z., et al. (2005). Prediction of general mental ability based on neural oscillation measures of sleep. *J. Sleep Res.* 14, 285–292. doi: 10.1111/j.1365-2869.2005.00472.x
- Carskadon, M. A., and Rechtschaffen, A. (2000). “Monitoring and staging of human sleep,” in *Principles and Practice of Sleep Medicine*, 3rd Edn. eds M. H. Kryger, T. Roth and W. C. Dement (Philadelphia: W.B. Saunders Company), 1197–1215.
- Dang-Vu, T. T., Schabus, M., Desseilles, M., Sterpenich, V., Bonjean, M., and Maquet, P. (2010). Functional neuroimaging insights into the physiology of human sleep. *Sleep* 33, 1589–1603.
- DeGennaro, L., and Ferrara, M. (2003). Sleep spindles: an overview. *Sleep Med. Rev.* 7, 423–440. doi: 10.1053/smr.2002.0252
- DeGennaro, L., Ferrara, M., Vecchio, F., Curcio, G., and Bertini, M. (2005). An electroencephalographic fingerprint of human sleep. *Neuroimage* 26, 114–122. doi: 10.1016/j.neuroimage.2005.01.020
- Fogel, S. M., Nader, R., Cote, K. A., and Smith, C. T. (2007). Sleep spindles and learning potential. *Behav. Neurosci.* 121, 1–10. doi: 10.1037/0735-7044.121.1.1
- Fogel, S. M., and Smith, C. T. (2006). Learning-dependent changes in sleep spindles and Stage 2 sleep. *J. Sleep Res.* 15, 250–255. doi: 10.1111/j.1365-2869.2006.00522.x
- Fogel, S. M., and Smith, C. T. (2011). The function of the sleep spindle: a physiological index of intelligence and a mechanism for sleep-dependent memory consolidation. *Neurosci. Biobehav. Rev.* 35, 1154–1165. doi: 10.1016/j.neubiorev.2010.12.003
- Gaillard, J.-M., and Blois, R. (1981). Spindle density in sleep of normal subjects. *Sleep* 4, 385–391.
- Geiger, A., Huber, R., Kurth, S., Ringli, M., Jenni, O. G., and Achermann, P. (2011). The sleep EEG as a marker of intellectual ability in school age children. *Sleep* 34, 181–189.
- Jacobson, L. A., Ryan, M., Martin, R. B., Ewen, J., Mostofsky, S. H., Denckla, M. B., et al. (2011). Working memory influences processing speed and reading fluency in ADHD. *Child Neuropsychol.* 17, 209–224. doi: 10.1080/09297049.2010.532204
- Jenni, O. G., and Carskadon, M. A. (2004). Spectral analysis of the sleep encephalogram during adolescence. *Sleep* 27, 774–783.
- Nader, R., and Smith, C. (2001). The relationship between stage 2 sleep spindles and intelligence. *Sleep* 24, A160.
- Nader, R., and Smith, C. (2003). “A role for stage 2 sleep in memory processing,” in *Sleep and Synaptic Plasticity*, eds P. Maquet, C. Smith and R. Stickgold (Oxford: Oxford University Press), 87–98.
- Nader, R. S., Smith, C. T., Murkar, A., and Sabbagh, M. A. (2012a). Spindles through the stages. *J. Sleep Res.* 21(Suppl. 1), 132–133. doi: 10.1111/j.1365-2869.2012.01044.x
- Nader, R., Smith, C. T., and Sabbagh, M. (2012b). Evidence of a superfast spindle in the 16–19 Hz range. *Sleep* 35(Abstract supplement), A21–A22.
- Nader, R., Smith, C. T., and Sabbagh, M. (2012c). The intelligence quotient and spindle activity. *J. Sleep Res.* 21(Suppl. 1), 71. doi: 10.1111/j.1365-2869.2012.01044.x
- Nicolas, A., Petit, D., Rompre, S., and Montplaisir, J. (2001). Sleep spindle characteristics in healthy subjects of different age groups. *Clin. Neurophysiol.* 112, 521–527. doi: 10.1016/s1388-2457(00)00556-3
- Peter-Derex, L., Comte, J.-C., Mauguier, F., and Salin, P. A. (2012). Density and frequency caudo-rostral gradients of sleep spindles recorded in the human cortex. *Sleep* 35, 69–79. doi: 10.5665/sleep.1588
- Peters, K. R., Ray, L., Smith, V., and Smith, C. (2008). Changes in the density of stage 2 sleep spindles following motor learning in young and older adults. *J. Sleep Res.* 17, 23–33. doi: 10.1111/j.1365-2869.2008.00634.x

- Petit, D., Gagnon, J. F., Fantini, M. L., Ferini-Strambi, L., and Montplaisir, J. (2004). Sleep and quantitative EEG in neurodegenerative disorders. *J. Psychosom. Res.* 56, 487–496. doi: 10.1016/j.jpsychores.2004.02.001
- Ray, L. B., Fogel, S. M., Smith, C. T., and Peters, K. R. (2009). Validating an automated sleep spindle detection algorithm using an individualized approach. *J. Sleep Res.* 19, 374–378. doi: 10.1111/j.1365-2869.2009.00802.x
- Rechtschaffen, A., and Kales, A. (1968). *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*. Bethesda, MD: US Department of Health, Education and Welfare Public Health Service-National Institute of Health.
- Schabus, M., Dang-Vu, T. T., Albouy, G., Albouy, G., Balteau, E., Boly, M., 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
- Schabus, M., Hödlmoser, K., Gruber, G., Sauter, C., Anderer, P., Klosch, G., et al. (2006). Sleep spindle-related activity in the human EEG and its relation to general cognitive and learning abilities. *Eur. J. Neurosci.* 23, 1738–1746. doi: 10.1111/j.1460-9568.2006.04694.x
- Steriade, M., and McCarley, R. (eds) (2005). “Synchronized brain oscillations leading to neuronal plasticity during waking and sleep states,” in *Brain Control of Wakefulness and Sleep* (New York: Kluwer Academic/Plenum Publishers), 255–344.
- Tarokh, L., and Carskadon, M. A. (2010). Developmental changes in the human sleep EEG during early adolescence. *Sleep* 33, 801–809.
- 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: 11 September 2014; accepted: 12 January 2015; published online: 09 February 2015.

Citation: Nader RS and Smith CT (2015) Correlations between adolescent processing speed and specific spindle frequencies. *Front. Hum. Neurosci.* 9:30. doi: 10.3389/fnhum.2015.00030

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

Copyright © 2015 Nader and Smith. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and 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.

Age-related changes in sleep spindles characteristics during daytime recovery following a 25-hour sleep deprivation

T. Rosinvil^{1,2,3}, M. Lafortune^{1,2,3}, Z. Sekerovic^{1,2,3}, M. Bouchard^{1,2,3}, J. Dubé^{1,2,3}, A. Latulipe-Loiselle², N. Martin^{1,2,3}, J. M. Lina^{1,4} and J. Carrier^{1,2,3*}

¹ Center for Advanced Research in Sleep Medicine, Hôpital du Sacré-Coeur de Montréal, Montréal, QC, Canada,

² Department of Psychology, Université de Montréal, Montréal, QC, Canada, ³ Research Center, Institut Universitaire Gériatrique de Montréal, Montréal, QC, Canada, ⁴ Department of Electrical Engineering, École de Technologie Supérieure, Montréal, QC, Canada

OPEN ACCESS

Edited by:

Srikantan S. Nagarajan,
University of California,
San Francisco, USA

Reviewed by:

Juliana Yordanova,
Institute of Neurobiology, Bulgarian
Academy of Sciences,
Bulgaria

Heidi E. Kirsch,
University of California,
San Francisco, USA

*Correspondence:

J. Carrier,
Center for Advanced Research in
Sleep Medicine, Hôpital du
Sacré-Coeur de Montréal, 5400
Gouin Blvd West Montreal, Montréal,
QC H4J 1C5, Canada
julie.carrier.1@umontreal.ca

Received: 06 January 2015

Accepted: 19 May 2015

Published: 03 June 2015

Citation:

Rosinvil T, Lafortune M, Sekerovic Z, Bouchard M, Dubé J, Latulipe-Loiselle A, Martin N, Lina JM and Carrier J (2015) Age-related changes in sleep spindles characteristics during daytime recovery following a 25-hour sleep deprivation. *Front. Hum. Neurosci.* 9:323. doi: 10.3389/fnhum.2015.00323

Objectives: The mechanisms underlying sleep spindles (~11–15 Hz; >0.5 s) help to protect sleep. With age, it becomes increasingly difficult to maintain sleep at a challenging time (e.g., daytime), even after sleep loss. This study compared spindle characteristics during daytime recovery and nocturnal sleep in young and middle-aged adults. In addition, we explored whether spindles characteristics in baseline nocturnal sleep were associated with the ability to maintain sleep during daytime recovery periods in both age groups.

Methods: Twenty-nine young (15 women and 14 men; 27.3 y ± 5.0) and 31 middle-aged (19 women and 13 men; 51.6 y ± 5.1) healthy subjects participated in a baseline nocturnal sleep and a daytime recovery sleep after 25 hours of sleep deprivation. Spindles were detected on artifact-free Non-rapid eye movement (NREM) sleep epochs. Spindle density (nb/min), amplitude (μV), frequency (Hz), and duration (s) were analyzed on parasagittal (linked-ears) derivations.

Results: In young subjects, spindle frequency increased during daytime recovery sleep as compared to baseline nocturnal sleep in all derivations, whereas middle-aged subjects showed spindle frequency enhancement only in the prefrontal derivation. No other significant interaction between age group and sleep condition was observed. Spindle density for all derivations and centro-occipital spindle amplitude decreased whereas prefrontal spindle amplitude increased from baseline to daytime recovery sleep in both age groups. Finally, no significant correlation was found between spindle characteristics during baseline nocturnal sleep and the marked reduction in sleep efficiency during daytime recovery sleep in both young and middle-aged subjects.

Conclusion: These results suggest that the interaction between homeostatic and circadian pressure modulates spindle frequency differently in aging. Spindle characteristics do not seem to be linked with the ability to maintain daytime recovery sleep.

Keywords: aging, sleep spindles, circadian process, sleep loss, homeostatic sleep pressure

Non-rapid-eye movement (NREM) sleep is a global brain process commonly defined by an absence of interaction with the environment, altered awareness, reduced external information processing and enhanced cortical synchronization. High levels of cortical synchronization during slow-wave sleep (SWS or N3 NREM sleep) is characterized by high-amplitude (>75 mV) electroencephalographic (EEG) slow waves (<4 Hz; SW). SW have two phases at the cellular level: a hyperpolarization phase (surface EEG SW negative phase), during which cortical neurons are mostly silent, and a depolarization phase (surface EEG SW positive phase), during which most cortical neurons fire intensively (Steriade, 2006). Sleep spindles (waxing and waning EEG waves of 12–15 Hz; >0.5 s) occur mostly during N2 NREM sleep but still persist in N3 NREM sleep to be eventually replaced by SWs. Hence, several observations support a reciprocal relationship between sleep spindles and SW in NREM sleep (Dijk et al., 1993; Steriade et al., 1993; for a review: De Gennaro and Ferrara, 2003).

Aging is associated with less time asleep, more frequent awakenings of longer duration and shallower sleep (Buysse et al., 1992; Hoch et al., 1994; Landolt et al., 1996; Carrier et al., 1997, 2001; Landolt and Borbély, 2001). These changes are part of the normal aging process and occur gradually during the middle years of life (Carrier et al., 2001). Moreover, NREM sleep changes drastically in the middle years of life through a substantial reduction in SWS and an increase in lighter NREM sleep stages (Hoch et al., 1994; Landolt et al., 1996; Carrier et al., 1997). Studies have shown considerable changes in NREM sleep from age 20 to 60 years, including significant decreases in slow-wave activity (SWA; i.e., spectral power between 0.5–4.5 Hz) and low sigma activity (spectral power between 13–14 Hz), during NREM (Carrier et al., 2001; Landolt and Borbély, 2001). Our group has also shown that middle-aged subjects exhibit lower density and amplitude of SW and spindles when compared to younger participants, especially in prefrontal/frontal brain areas (Carrier et al., 2011; Lafortune et al., 2012; Martin et al., 2013).

The sleep-wake cycle is regulated by the interaction between the homeostatic and the circadian processes (Dijk and Czeisler, 1994). The homeostatic process represents the sleep pressure accumulated by the time spent awake and dissipated during a sleep episode (Achermann et al., 1993). In humans, the intensity and dynamics of slow wave activity (SWA; spectral power between 0.5–4 Hz in NREM) model the time course of the homeostatic process (i.e., more time awake produces more SWA, whereas more time asleep is associated with less SWA; Achermann et al., 1993). A few studies showed lower rebound of SWA as well as SW density and amplitude after sleep deprivation in middle-aged and older subjects when compared to younger participants, particularly in anterior brain areas (Gaudreau et al., 2001a; Münch et al., 2004; Carrier et al., 2009; Lafortune et al., 2012). The latter results suggest that there is a reduction in homeostatic sleep pressure as age increases starting in the middle years of life. On the other hand, a biological “clock” located in the suprachiasmatic nucleus controls the circadian process of sleep regulation. Circadian wake propensity increases during the day and maximizes in

the evening (Czeisler et al., 1980; Zulley et al., 1981; Lavie, 1985). Studies have shown that sleep in middle-aged and older subjects is particularly vulnerable to circadian phases of high wake propensity, which means that it is more difficult with aging to maintain sleep at the “wrong” circadian phase (e.g., in the daytime), even after sleep deprivation (Cajochen et al., 1999; Gaudreau et al., 2001b). The mechanisms underlying this stronger enhancement of wakefulness during daytime recovery sleep in middle-aged and older participants compared to younger subjects remain unknown. Recently, we tested whether age-related modifications in SW could be linked to enhanced wakefulness during daytime recovery sleep, but none of the SW characteristics at baseline were associated with daytime recovery sleep efficiency in young and middle-aged subjects (Lafortune et al., 2012).

One of the functional roles attributed to sleep spindles is to prevent afferent signals from being transmitted to the cortex, thus allowing cortical unresponsiveness to stimulation during sleep (Steriade et al., 1993; Steriade, 1994, 2006; Bazhenov et al., 1999; Born et al., 2002; Czisch et al., 2002; Dang-Vu et al., 2011). Hence, age-related changes in spindles may be linked to the ability to maintain sleep at an abnormal circadian phase. Interestingly, spindle characteristics are also regulated by the interaction between the homeostatic and the circadian processes. Compared to conditions of lower homeostatic sleep pressure prior to nocturnal sleep, studies have shown a reduction in spindle density and in spindle mean frequency under higher sleep homeostatic pressure in young subjects (Curcio et al., 2003; Knoblauch et al., 2003a). However, to our knowledge, no study has evaluated age-related effects of sleep deprivation on spindles. Studies have also reported lower spindle density and higher spindle mean frequency when sleep occurred at a circadian time corresponding to daytime in comparison to night-time (Wei et al., 1999; Knoblauch et al., 2003b, 2005). Importantly, this circadian modulation of spindles is reduced in older subjects when compared to younger subjects (Wei et al., 1999; Knoblauch et al., 2005).

The main aim of this study is to compare sleep spindles characteristics between baseline nocturnal sleep and daytime recovery sleep after 25 h of total sleep deprivation in both young and middle-aged subjects. We also evaluated whether sleep spindles are associated with the ability to maintain sleep during daytime recovery sleep. We predict that middle-aged subjects will have a lower reduction of spindles during daytime recovery sleep compared to younger subjects and that higher spindle density during baseline sleep will be associated with a smaller decrease in sleep efficiency during daytime recovery sleep in young and older subjects.

Methods

Subjects and Procedure

Twenty-nine young (15 women and 14 men; 20–38 years old, mean = 27.3 years, SD = 5.0) and 31 middle-aged (19 women and 13 men, 40–60 years old, mean = 51.6 years, SD = 5.1) healthy subjects were recruited for this study. Data from participants

were drawn from two studies conducted between 1999 and 2006 in our laboratory, all following similar recording procedures and free from active pharmacological manipulation (Gaudreau et al., 2001b; Carrier et al., 2009). All subjects signed an informed consent form and received monetary compensation for their participation. All research studies were approved by the ethical committee of the Hôpital du Sacré-Coeur de Montréal.

A semi-structured interview using a homemade questionnaire was performed to exclude potential subjects who smoked, used sleep-affecting medication and reported sleep complaints or unusual sleep duration (i.e., <7 h and >9 h). Participants who engaged in night work or transmeridian travel 3 months prior to the study were also excluded. No subjects reported neurological or psychiatric illness history using our homemade questionnaire, nor showed indication of depression (Beck Depression Inventory, short version >3 or long version >9; Beck and Steer, 1987). Moreover, to rule out any significant medical condition, certified physicians evaluated blood sample analysis (complete blood count, serum chemistry, including hepatic and renal functions; prolactin level; testosterone level in men; and estrogen, follicle stimulating hormone (FSH) and luteinizing hormone levels in women) and urinalysis results. Perimenopausal women and women using hormonal contraception or receiving hormonal replacement therapy were excluded. Premenopausal women reported regular menstrual cycles (25–32 days) in the year preceding the experiment, had no vasomotor complaints (i.e., night sweats, hot flashes) and showed low FSH levels (<20 iU/L). All postmenopausal women reported an absence of menses in the past year and showed high FSH levels (>20 iU/L).

Prior to data acquisition, all subjects underwent a polysomnographic (PSG) adaptation and screening night; including nasal/oral thermistor and an electromyogram (EMG) leg electrode recordings to screen for sleep disturbances. The presence of sleep disorders such as sleep apneas, hypopneas and periodic leg movements (index per hour >10) resulted in the participant's exclusion.

All subjects came to the laboratory for a baseline nocturnal sleep episode (BSL). The following night, subjects were sleep deprived. A morning recovery sleep episode (REC) was initiated one hour after their habitual wake time (after 25 h of wakefulness). During the night of sleep deprivation, all subjects remained awake in a semi-recumbent position in dim light (<15 lux) until the next morning. Bedtime and wake time in the laboratory were determined using averaged regular schedules obtained from sleep diary entries (recorded 7 days prior to BSL).

Polysomnographic Recordings

PSG recordings included EEG electrodes (10–20 system, referential montage with linked ears), chin EMG and left and right electrooculography (EOG). PSG was recorded using a Grass Model 15 amplifier system (gain 10,000; bandpass 0.3–100 Hz). Signals were digitalized at a sampling rate of 256 Hz using commercial software (Harmonie, Stellate System). Sleep stages were visually scored on C3 in 20-s epochs on a computer screen according to standard criteria (Rechtschaffen and Kales, 1968).

EEG artifacts were detected automatically (Brunner et al., 1996) and then inspected visually to ensure appropriate rejection from analysis.

Automatic Algorithm Detection of Sleep Spindles

Sleep spindles were detected automatically on artifact-free NREM epochs for left and right parasagittal scalp derivations (i.e., Fp1, F3, C3, P3, O1 and Fp2, F4, C4, P4, O2). EEG data were first bandpass filtered from 11 to 15 Hz with a linear phase Finite Impulse Response filter (−3 dB at 11.1 and 14.9 Hz). Forward and reverse filtering was performed to obtain zero-phase distortion and to double the filter order. The root mean square (RMS) of the filtered signal was then calculated with a 0.25 s time window and thresholded at its 95th percentile (Schabus et al., 2007). A spindle was identified when at least two consecutive RMS time-points exceeded the threshold, reaching duration criterion (0.5 s; no superior limit but 98% of spindles were ≤1 s). Four spindle characteristics were derived: density (number of spindles/minutes of NREM sleep, expressed in nb/min), amplitude (peak-to-peak difference in voltage, expressed in μ V), frequency (number of cycles/second, expressed in Hz), and duration (expressed in seconds). Spindle characteristics were assessed over the entire night. Spindle characteristics from left and right electrodes were averaged together (prefrontal: FP1–FP2, Frontal: F3–F4, Central: C3–C4, Parietal: P3–P4, Occipital: O1–O2).

Statistical Analyses

Preliminary Analyses

To evaluate possible interaction between sex, age and sleep conditions, 3-way mixed design analysis of variance (ANOVA) with two independent factors (age groups: young and middle-aged; sex groups: men and women) and one repeated measure (2 sleep conditions: BSL, REC) were performed on PSG variables and spindle characteristics for each topographical site (prefrontal, frontal, central, parietal, occipital). No significant interactions between age group, sex and sleep condition were found for PSG characteristics and all spindle characteristics, except for sleep spindle density in the central area ($F_{(1,57)} = 5.64$, $p = 0.02$). *Post hoc* analysis revealed an age group by sleep condition interaction for women ($F_{(1,32)} = 5.14$, $p = 0.03$) and not for men ($F_{(1,25)} = 1.39$, $p = 0.25$). Middle-aged women showed a stronger decrease in spindle density in the central area compared to young women. In men, spindle density was lower from BSL to REC in both age group resulting in only a main effect of sleep condition ($F_{(1,25)} = 72.58$, $p < 0.0001$). Consequently, data from men and women were pooled together, except for spindle density, which was analyzed separately in men and women.

Analyses

Two-way ANOVAs with one independent factor (2 age groups) and one repeated measure (2 sleep conditions: BSL, REC) were performed on PSG sleep variables. Mixed ANOVAs with one independent factor (2 age groups) and two repeated measures (2 sleep conditions: BSL, REC; 5 derivations: Prefrontal, Frontal, Central, Parietal and Occipital) were performed for each spindle characteristic.

P values for repeated measures with more than two levels were adjusted for sphericity with Huynh-Feldt corrections, but original degrees of freedom were reported. Differences in main effects and in interactions were assessed with *post hoc* multiple mean comparisons, and effect size (ES) were measured using the partial ETA square and Wilk's Lambda partial ETA square when applicable. Results were considered significant when $p \leq 0.05$.

Pearson correlations were performed between all-night spindle characteristics during baseline sleep and the change in sleep efficiency between BSL and REC sleep (absolute and % in change) in the young and the middle-aged groups separately and in the two groups pooled together with age as a control variable. In these analyses, we applied a more severe level of significance (i.e., $p \leq 0.01$) to correct for multiple comparisons.

Results

Sleep Architecture

Sleep efficiency and duration was lower during REC sleep as compared to BSL sleep in both age groups. However, this reduction of sleep efficiency and duration was more prominent in the middle-aged than in the young subjects. SWS was higher during REC sleep compared to BSL sleep, but this effect was less prominent in middle-aged compared to young subjects. As for sleep latency, % of stages 2 and REM sleep, they were all lower in REC sleep when compared to BSL sleep. Finally, middle-aged subjects showed a higher percentage of stage 2 sleep in comparison to younger participants (see **Table 1** for all effects).

All-Night Spindle Characteristics

Significant interactions between sleep conditions and derivations were found for spindle density in men ($F_{(4,128)} = 6.35$, $p < 0.0001$) and in women ($F_{(4,128)} = 18.32$, $p < 0.0001$; see **Figure 1**). For both men and women, spindle density was lower in REC sleep compared to BSL sleep in all derivations. The effect was stronger in the central area and weaker in the prefrontal region. No significant effect of age or interaction between age groups and sleep conditions was found for spindle density in men (interaction: $F_{(1,25)} = 0.24$, $p = 0.63$; age only: $F_{(1,25)} = 0.79$, $p = 0.38$) or in women

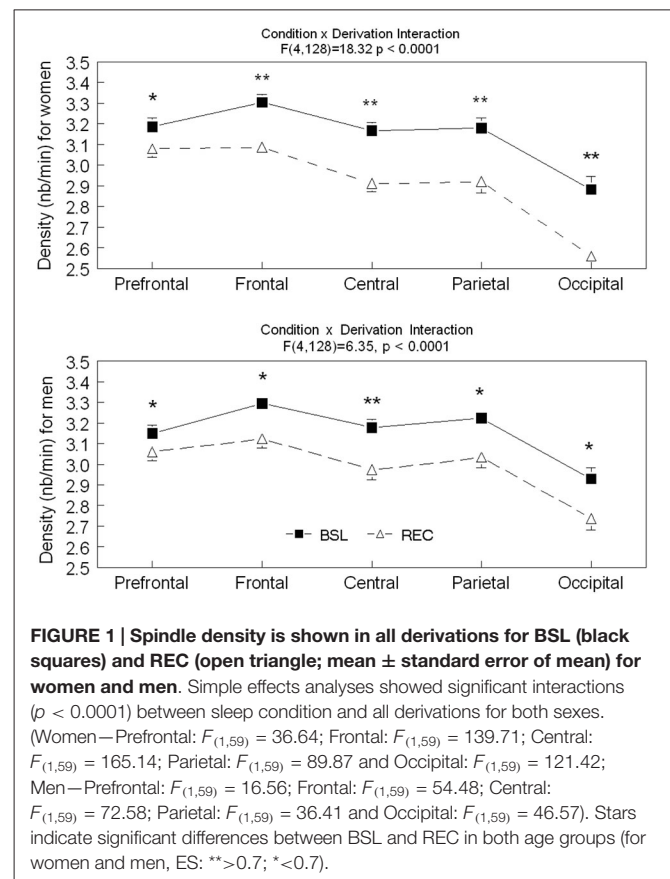


FIGURE 1 | Spindle density is shown in all derivations for BSL (black squares) and REC (open triangle; mean \pm standard error of mean) for women and men. Simple effects analyses showed significant interactions ($p < 0.0001$) between sleep condition and all derivations for both sexes. (Women—Prefrontal: $F_{(1,59)} = 36.64$; Frontal: $F_{(1,59)} = 139.71$; Central: $F_{(1,59)} = 165.14$; Parietal: $F_{(1,59)} = 89.87$ and Occipital: $F_{(1,59)} = 121.42$; Men—Prefrontal: $F_{(1,59)} = 16.56$; Frontal: $F_{(1,59)} = 54.48$; Central: $F_{(1,59)} = 72.58$; Parietal: $F_{(1,59)} = 36.41$ and Occipital: $F_{(1,59)} = 46.57$). Stars indicate significant differences between BSL and REC in both age groups (for women and men, ES: ** >0.7 ; * <0.7).

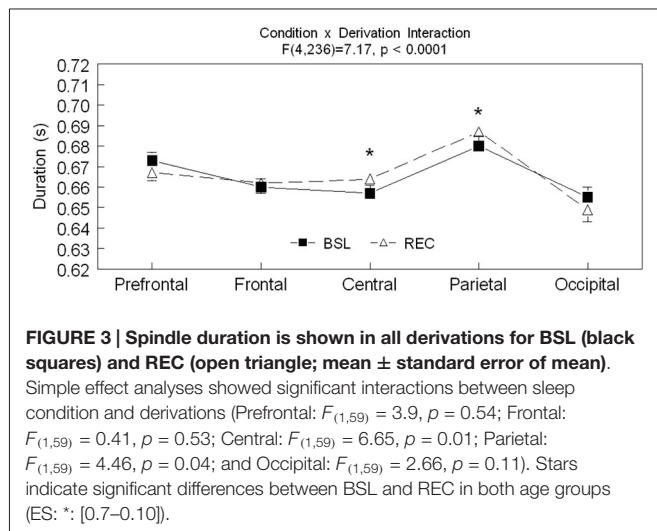
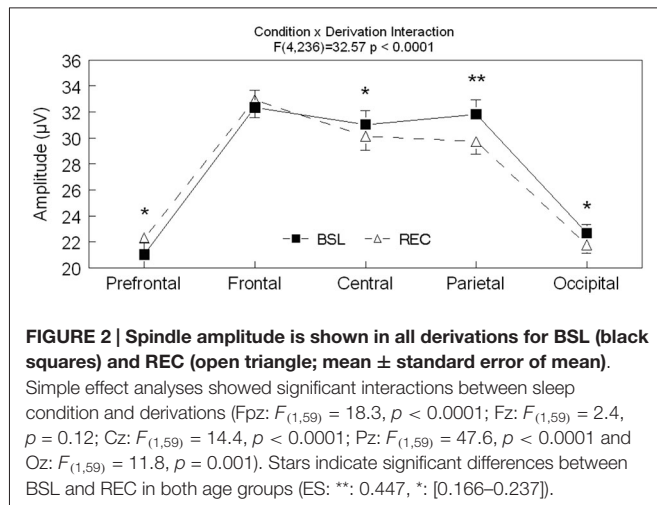
(interaction: $F_{(1,32)} = 1.21$, $p = 0.28$; age only: $F_{(1,32)} = 1.59$, $p = 0.22$).

Significant interactions between sleep conditions and derivations were also found for spindle amplitude ($F_{(4,236)} = 32.57$, $p < 0.0001$) and spindle duration ($F_{(4,236)} = 7.17$, $p < 0.0001$; see **Figures 2–4** for *post hoc* analyses). Compared to BSL sleep, spindle amplitude was higher during REC sleep for the prefrontal area but lower for central, parietal, and occipital areas. Finally, compared to REC sleep, spindles lasted longer only in the central and parietal areas in BSL sleep. No significant effect of age groups ($F_{(1,59)} = 2.0$, $p = 0.16$) or interaction

TABLE 1 | Polysomnographic variables for young and middle-aged subjects in both sleep conditions.

PSG variables	Young		Middle-Aged		Age effect	Condition effect	Age \times Condition interaction
	BSL	REC	BSL	REC	$F(p)$	$F(p)$	$F(p)$
Sleep latency (min)	9.2 (5.8)	3.5 (3.6)	10 (5.9)	5.2 (4.8)	n.s.	$F = 46.5$ ($p < 0.0001$)	n.s.
Sleep duration (min)	437.3 (38.5)	377.9 (65.1)	428.8 (46.2)	335.1 (59.7)	$F = 5.5$ ($p = 0.02$)	$F = 83.5$ ($p < 0.0001$)	$F = 4.3$ ($p = 0.04$)
Efficiency (%)	91.1 (5.7)	79.8 (14.4)	87.7 (6.1)	69.2 (13)	$F = 10.2$ ($p = 0.002$)	$F = 89.2$ ($p < 0.0001$)	$F = 5.1$ ($p = 0.02$)
Stage 1 (%)	7.7 (3.7)	7.9 (5.4)	7.9 (3.4)	8.6 (3.6)	n.s.	n.s.	n.s.
Stage 2 (%)	59.8 (5.6)	56.5 (9.5)	66.4 (5.6)	65.8 (6.7)	$F = 25.1$ ($p < 0.0001$)	$F = 5$ ($p = 0.03$)	n.s.
SWS (%)	9.1 (6.4)	16.8 (9.8)	3.9 (4.1)	8.7 (7.3)	$F = 15.2$ ($p < 0.0001$)	$F = 94$ ($p < 0.0001$)	$F = 5$ ($p = 0.03$)
REM (%)	23.4 (4.9)	18.7 (6.7)	21.8 (4.4)	17 (6.4)	n.s.	$F = 45.6$ ($p < 0.0001$)	n.s.
NREM (min)	301.9 (35.6)	276.5 (51.5)	297.2 (39.9)	252.0 (51.9)	n.s.	$F = 28.3$ ($p < 0.0001$)	n.s.

Note. Untransformed mean (standard deviation).

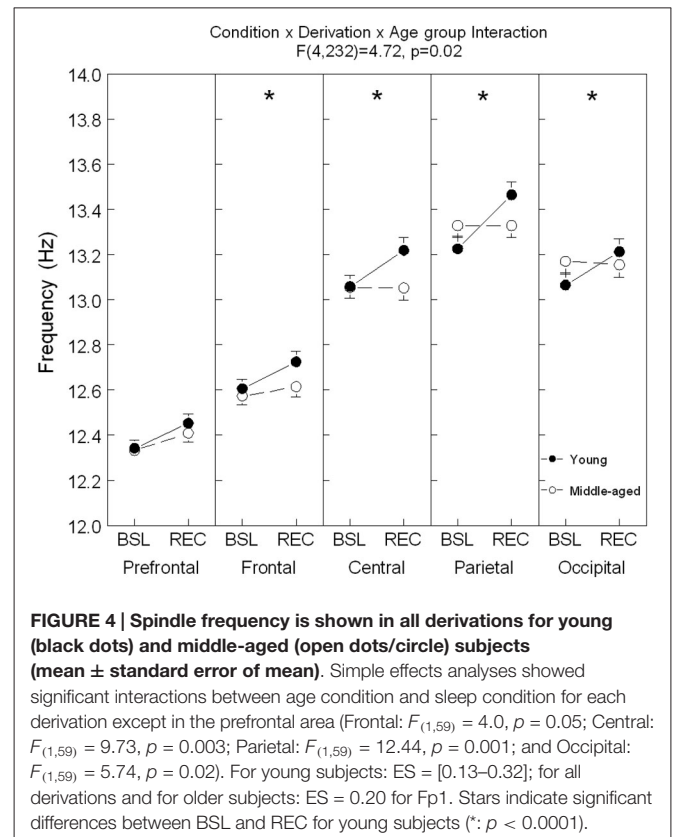


between age groups and sleep conditions ($F_{(1,59)} = 0.06, p = 0.81$) were found for spindle amplitude. For spindle duration, a main effect of age groups was found ($F_{(1,59)} = 11.5, p = 0.001$) with no significant interaction between age groups and sleep conditions ($F_{(1,59)} = 0.49, p = 0.49$). Hence, middle-aged subjects showed shorter spindle duration compared to young subjects.

A significant interaction between age groups, sleep conditions and derivations was found for spindle frequency ($F_{(4,236)} = 4.72, p = 0.02$; see **Figure 4** for contrast analyses). In comparison to BSL sleep, an increase of spindle frequency was observed during REC sleep for young subjects in all derivations, whereas in the middle-aged subjects, spindle frequency was higher only in the prefrontal area.

Spindles Characteristics and Sleep Efficiency

No significant correlations were found between spindle density, frequency and amplitude at BSL and change in sleep efficiency from BSL to REC sleep (absolute change and percent of change) for young subjects. Only a few moderate counterintuitive negative correlations were found between spindle density in



the prefrontal area and the decrease of sleep efficiency in the middle-aged subjects (absolute change and % of change: $r = -0.46, p < 0.01$) and in both age groups combined (absolute change: $r = -0.37, p < 0.01$; % of change: $r = -0.36, p < 0.01$).

Discussion

Young and middle-aged adults showed comparable differences in spindle density, spindle amplitude and spindle duration during REC sleep compared to BSL sleep. Only spindle frequency showed a differential effect of age between BSL and REC sleep. Although our results illustrated a marked reduction of sleep efficiency during the day associated with aging, spindle characteristics were not linked with the ability to maintain REC sleep.

In our study, during REC sleep compared to BSL sleep, homeostatic sleep propensity was higher (due to sleep loss) and circadian wake propensity increased (due to daytime sleep). Studies evaluating the circadian modulation of sleep spindles have reported higher spindle frequency during daytime sleep as compared to nighttime (Wei et al., 1999; Knoblauch et al., 2003b, 2005). On the other hand, studies showed a reduction in spindle frequency under higher compared to lower sleep homeostatic pressure in young participants (Knoblauch et al., 2003a). During REC sleep, young subjects showed faster spindle frequency compared to BSL sleep over all derivations. This result suggests that in young subjects,

the enhancement of spindles frequency by the circadian modulation during daytime overrides the homeostatic pressure for a reduction in spindle frequency induced by the 25-h sleep deprivation. In the middle-aged participants, faster spindle frequency during REC sleep was observed only in the prefrontal area. This observation supports a previous study that showed an age-related reduction in time-of-day modulation of spindle frequency using a 40-h multiple-nap paradigm under constant-routine conditions (Knoblauch et al., 2005).

In the present study, spindle density was lower in REC sleep compared to BSL in all derivations, but this decrease was more prominent in central and frontal areas. These results confirm a previous study, which showed lower spindle density, especially in the frontal derivation, after a 40-h sleep deprivation in young subjects (Knoblauch et al., 2003a). Our results are also congruent with studies showing that spindle incidence and density are lower during daytime compared to night-time sleep (Wei et al., 1999; Knoblauch et al., 2005). Compared to BSL sleep, women showed a stronger decrease in spindle density in the central derivation than men. Higher sigma power and spindle density in women compared to men has been reported in previous studies (Gaillard and Blois, 1981; Carrier et al., 2001; Huupponen et al., 2002; Lafortune et al., 2014). However, no studies have yet evaluated whether homeostatic and circadian modulations of sleep spindles differ between men and women.

Compared to BSL sleep, spindle amplitude was higher during REC sleep for the prefrontal area but lower for central, parietal and occipital areas. These results do not support one previous study, which reported higher spindle amplitude in central, parietal and occipital areas during nocturnal sleep after a 40-h sleep deprivation (Knoblauch et al., 2003a). However, circadian studies showed lower spindle amplitude when sleep is initiated at a circadian time corresponding to daytime (Wei et al., 1999; Knoblauch et al., 2005). Hence, circadian modulation of spindle

amplitude probably explains the decrease in spindle amplitude in central, parietal and occipital derivations during daytime REC sleep in our study.

Finally, compared to REC sleep, spindles lasted longer only in the central and parietal areas in BSL sleep. No change in spindle duration was previously reported in nocturnal recovery sleep after a 40-h sleep deprivation (Knoblauch et al., 2003a). Studies evaluating the circadian modulation of spindle duration found conflicting results. One forced desynchrony study reported shorter spindle duration in the central derivation when sleep was initiated at a circadian time corresponding to daytime compared to night-time (Wei et al., 1999), whereas a 40-h nap study showed longer spindle duration in the frontal derivation but shorter duration in the parietal derivation when naps occurred during daytime compared to night-time (Knoblauch et al., 2005).

Middle-aged subjects showed lower sleep efficiency when compared to younger subjects. No significant positive relationship was found between sleep spindles characteristics during the BSL night and change in sleep efficiency between BSL and REC. Our study suggests that individual spindle characteristics do not predict the ability to override the circadian waking signal after sleep loss. Similarly, Knoblauch et al. (2005) did not observe any relationship between the day-night difference in spindle frequency and the day-night difference in wake time. Our results are also in line with our previous results showing no relationship between SW and change in sleep efficiency between BSL and REC sleep (Lafortune et al., 2012). Taken together, these results indicate that individual characteristics in NREM sleep oscillations do not predict the increased wakefulness during daytime recovery sleep. Further studies should aim at understanding the mechanisms that explain the greater sensitivity in older individuals to circadian challenges.

References

- Achermann, P., Dijk, D. J., Brunner, D. P., and Borbély, A. A. (1993). A model of human sleep homeostasis based on EEG slow-wave activity: quantitative comparison of data and simulations. *Brain Res. Bull.* 31, 97–113. doi: 10.1016/0361-9230(93)90016-5
- Bazhenov, M., Timofeev, I., Steriade, M., and Sejnowski, T. J. (1999). Self-sustained rhythmic activity in the thalamic reticular nucleus mediated by depolarizing GABA_A receptor potentials. *Nat. Neurosci.* 2, 168–174. doi: 10.1038/5729
- Beck, A. T., and Steer, R. A. (1987). *The Beck Depression Inventory*. London: Psychological Corporation.
- Born, A. P., Law, I., Lund, T. E., Rostrup, E., Hanson, L. G., Wildschjodt, G., et al. (2002). Cortical deactivation induced by visual stimulation in human slow-wave sleep. *Neuroimage* 17, 1325–1335. doi: 10.1006/nimg.2002.1249
- Brunner, D. P., Vasko, R. C., Detka, C. S., Monahan, J. P., Reynolds, C. F. 3rd, 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
- Buyse, D. J., Browman, K. E., Monk, T. H., Reynolds, C. F. 3rd, Fascizka, A. L., and Kupfer, D. J. (1992). Napping and 24-hour sleep/wake patterns in healthy elderly and young adults. *J. Am. Geriatr. Soc.* 40, 779–786. doi: 10.1111/j.1532-5415.1992.tb01849.x
- Cajochen, C., Foy, R., and Dijk, D. J. (1999). Frontal predominance of a relative increase in sleep delta and theta EEG activity after sleep loss in humans. *Sleep Res. Online* 2, 65–69.
- Carrier, J., Land, S., Buysse, D. J., Kupfer, D. J., and Monk, T. H. (2001). The effects of age and gender on sleep EEG power spectral density in the middle years of life (ages 20–60 years old). *Psychophysiology* 38, 232–242. doi: 10.1111/1469-8986.3820232
- Carrier, J., Monk, T. H., Buysse, D. J., and Kupfer, D. J. (1997). Sleep and morningness-eveningness in the 'middle' years of life (20–59 y). *J. Sleep Res.* 6, 230–237. doi: 10.1111/j.1365-2869.1997.00230.x
- Carrier, J., Paquet, J., Fernandez-Bolanos, M., Girouard, L., Roy, J., Selmaoui, B., et al. (2009). Effects of caffeine on daytime recovery sleep: a double challenge to the sleep-wake cycle in aging. *Sleep Med.* 10, 1016–1024. doi: 10.1016/j.sleep.2009.01.001
- Carrier, J., Viens, I., Poirier, G., Robillard, R., Lafortune, M., Vandewalle, G., et al. (2011). Sleep slow wave changes during the middle years of life. *Eur. J. Neurosci.* 33, 758–766. doi: 10.1111/j.1460-9568.2010.07543.x
- Curcio, G., Ferrara, M., Pellicciari, M. C., Cristiani, R., and De Gennaro, L. (2003). Effect of total sleep deprivation on the landmarks of stage 2 sleep. *Clin. Neurophysiol.* 114, 2279–2285. doi: 10.1016/s1388-2457(03)00276-1
- Czeisler, C. A., Weitzman, E. D., Moore-Ede, M. C., Zimmerman, J. C., and Knauer, R. S. (1980). Human sleep: its duration and organization depend on its circadian phase. *Science* 210, 1264–1267. doi: 10.1126/science.7434029

- Czisch, M., Wetter, T. C., Kaufmann, C., Pollmächer, T., Holsboer, F., and Auer, D. P. (2002). Altered processing of acoustic stimuli during sleep: reduced auditory activation and visual deactivation detected by a combined fMRI/EEG study. *Neuroimage* 16, 251–258. doi: 10.1006/nimg.2002.1071
- Dang-Vu, T. T., Bonjean, M., Schabus, M., Boly, M., Darsaud, A., Desseilles, M., et al. (2011). Interplay between spontaneous and induced brain activity during human non-rapid eye movement sleep. *Proc. Natl. Acad. Sci. U S A* 108, 15438–15443. doi: 10.1073/pnas.1112503108
- De Gennaro, L., and Ferrara, M. (2003). Sleep spindles: an overview. *Sleep Med. Rev.* 7, 423–440. doi: 10.1053/smr.2002.0252
- Dijk, D. J., and Czeisler, C. A. (1994). Paradoxical timing of the circadian rhythm of sleep propensity serves to consolidate sleep and wakefulness in humans. *Neurosci. Lett.* 166, 63–68. doi: 10.1016/0304-3940(94)90841-9
- Dijk, D. J., Hayes, B., and Czeisler, C. A. (1993). Dynamics of electroencephalographic sleep spindles and slow wave activity in men: effect of sleep deprivation. *Brain Res.* 626, 190–199. doi: 10.1016/0006-8993(93)90579-c
- Gaillard, J. M., and Blois, R. (1981). Spindle density in sleep of normal subjects. *Sleep* 4, 385–391.
- Gaudreau, H., Carrier, J., and Montplaisir, J. (2001a). Age-related modifications of NREM sleep EEG: from childhood to middle age. *J. Sleep Res.* 10, 165–172. doi: 10.1046/j.1365-2869.2001.00252.x
- Gaudreau, H., Morettini, J., Lavoie, H. B., and Carrier, J. (2001b). Effects of a 25-h sleep deprivation on daytime sleep in the middle-aged. *Neurobiol. Aging* 22, 461–468. doi: 10.1016/s0197-4580(00)00251-7
- Hoch, C. C., Dew, M. A., Reynolds, C. F. 3rd, Monk, T. H., Buysse, D. J., Houck, P. R., et al. (1994). A longitudinal study of laboratory- and diary-based sleep measures in healthy “old old” and “young old” volunteers. *Sleep* 17, 489–496.
- Huupponen, E., Himanen, S. L., Värri, A., Hasan, J., Lehtokangas, M., and Saarinen, J. (2002). A study on gender and age differences in sleep spindles. *Neuropsychobiology* 45, 99–105. doi: 10.1159/000048684
- Knoblauch, V., Martens, W. L., Wirz-Justice, A., and Cajochen, C. (2003a). Human sleep spindle characteristics after sleep deprivation. *Clin. Neurophysiol.* 114, 2258–2267. doi: 10.1016/s1388-2457(03)00238-4
- Knoblauch, V., Martens, W., Wirz-Justice, A., Kräuchi, K., and Cajochen, C. (2003b). Regional differences in the circadian modulation of human sleep spindle characteristics. *Eur. J. Neurosci.* 18, 155–163. doi: 10.1046/j.1460-9568.2003.02729.x
- Knoblauch, V., Münch, M., Blatter, K., Martens, W. L., Schröder, C., Schnitzler, C., et al. (2005). Age-related changes in the circadian modulation of sleep-spindle frequency during nap sleep. *Sleep* 28, 1093–1101.
- Lafortune, M., Gagnon, J. F., Latreille, V., Vandewalle, G., Martin, N., Filipini, D., et al. (2012). Reduced slow-wave rebound during daytime recovery sleep in middle-aged subjects. *PLoS One* 7:e43224. doi: 10.1371/journal.pone.0043224
- Lafortune, M., Gagnon, J. F., Martin, N., Latreille, V., Dubé, J., Bouchard, M., et al. (2014). Sleep spindles and rapid eye movement sleep as predictors of next morning cognitive performance in healthy middle-aged and older participants. *J. Sleep Res.* 23, 159–167. doi: 10.1111/jsr.12108
- Landolt, H. P., and Borbély, A. A. (2001). Age-dependent changes in the sleep EEG topography. *Clin. Neurophysiol.* 112, 369–377. doi: 10.1016/s1388-2457(00)00542-3
- Landolt, H. P., Dijk, D. J., Achermann, P., and Borbély, A. A. (1996). Effect of age on the sleep EEG: slow-wave activity and spindle frequency activity in young and middle-aged men. *Brain Res.* 738, 205–212. doi: 10.1016/s0006-8993(96)00770-6
- Lavie, P. (1985). “Ultradian rhythms: gates of sleep and wakefulness,” in *Ultradian Rhythms in Physiology and Behavior*, eds H. Schulz and P. Lavie (Berlin: Springer Verlag), 148–164.
- Martin, N., Lafortune, 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
- Münch, M., Knoblauch, V., Blatter, K., Schröder, C., Schnitzler, C., Kräuchi, K., et al. (2004). The frontal predominance in human EEG delta activity after sleep loss decreases with age. *Eur. J. Neurosci.* 20, 1402–1410. doi: 10.1111/j.1460-9568.2004.03580.x
- Rechtschaffen, A., and Kales, A. (1968). *A Manual of Standardized Terminology, Techniques and Scoring System of Sleep Stages in Human Subjects*. Los Angeles: Brain Information Service/Brain Research Institute, University of California.
- Schabus, M., Dang-Vu, T. T., Albouy, G., Baletau, 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
- Steriade, M. (1994). Sleep oscillations and their blockage by activating systems. *J. Psychiatry Neurosci.* 19, 354–358.
- Steriade, M. (2006). Grouping of brain rhythms in corticothalamic systems. *Neuroscience* 137, 1087–1106. doi: 10.1016/j.neuroscience.2005.10.029
- Steriade, M., McCormick, D. A., and Sejnowski, T. J. (1993). Thalamocortical oscillations in the sleeping and aroused brain. *Science* 262, 679–685. doi: 10.1126/science.8235588
- Wei, H. G., Riel, E., Czeisler, C. A., and Dijk, D. J. (1999). Attenuated amplitude of circadian and sleep-dependent modulation of electroencephalographic sleep spindle characteristics in elderly human subjects. *Neurosci. Lett.* 260, 29–32. doi: 10.1016/s0304-3940(98)00851-9
- Zulley, J., Wever, R., and Aschoff, J. (1981). The dependence of onset and duration of sleep on the circadian rhythm of rectal temperature. *Pflugers Arch.* 391, 314–318.

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.

Copyright © 2015 Rosinvil, Lafortune, Sekerovic, Bouchard, Dubé, Latulipe-Loiselle, Martin, Lina and Carrier. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and 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.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read,
for greatest visibility



COLLABORATIVE PEER-REVIEW

Designed to be rigorous
– yet also collaborative,
fair and constructive



FAST PUBLICATION

Average 85 days from
submission to publication
(across all journals)



COPYRIGHT TO AUTHORS

No limit to article
distribution and re-use



TRANSPARENT

Editors and reviewers
acknowledged by name
on published articles



SUPPORT

By our Swiss-based
editorial team



IMPACT METRICS

Advanced metrics
track your article's impact



GLOBAL SPREAD

5'100'000+ monthly
article views
and downloads



LOOP RESEARCH NETWORK

Our network
increases readership
for your article

Frontiers

EPFL Innovation Park, Building I • 1015 Lausanne • Switzerland
Tel +41 21 510 17 00 • Fax +41 21 510 17 01 • info@frontiersin.org
www.frontiersin.org

Find us on

