HHS Public Access

To Score or Not to Score? A look at the distinguishing power of micro EEG analysis on an annotated sample of PSG studies conducted in an HIV cohort

Yu Min Kang,
Department of Biomedical Engineering, Johns Hopkins University, 3400 Charles St., Baltimore, MD 21218 USA

Kristin M. Gunnarsdottir,
Department of Biomedical Engineering, Johns Hopkins University, 3400 Charles St., Baltimore, MD 21218 USA

Matthew S. D. Kerr [Member, IEEE],
Department of Biomedical Engineering, Johns Hopkins University, 3400 Charles St., Baltimore, MD 21218 USA

Rachel M. E. Salas,
Department of Neurology, Johns Hopkins Medicine, 1800 Orleans St. Baltimore, MD 21287 USA

Joshua Ewen,
Department of Neurology, Johns Hopkins Medicine, 1800 Orleans St. Baltimore, MD 21287 USA

Richard Allen,
Department of Neurology, Johns Hopkins Medicine, 1800 Orleans St. Baltimore, MD 21287 USA

Charlene Gamaldo, and
Department of Neurology, Johns Hopkins Medicine, 1800 Orleans St. Baltimore, MD 21287 USA

Sridevi V. Sarma [Member, IEEE]
Department of Biomedical Engineering, Johns Hopkins University, 3400 Charles St., Baltimore, MD 21218 USA

Abstract

In this study, we used the Pittsburgh Sleep Quality Index to divide the subjects into two groups, good sleepers and bad sleepers. We computed sleep *behavioral* (macro-sleep architectural) features and sleep *spectral* (micro-sleep architectural) features in order to observe if the annotated EEG data can be used to distinguish between good and bad sleepers in a more quantitative manner. Specifically, the macro-sleep features were defined by sleep stages and included sleep transitions, percentage of time spent in each sleep stage, and duration of time spent in each sleep stage. The micro-sleep features were obtained from the power spectrum of the EEG signals by computing the total power across all channels and all frequencies, as well as the average power in each sleep stage and across different frequency bands. We found that while the scoring-independent micro

corresponding author: kgunnar1@jhu.edu.

features are significantly different between the two groups, the macro features are not able to significantly distinguish the two groups. The fact that the macro features computed from the scoring files cannot pick up the expected difference in the EEG signals raises the question as to whether human scoring of EEG signals is practical in assessing sleep quality.

I. Introduction

In the modern world, sleep deprivation is a common occurrence in the general population. In the United States alone, about 40 million Americans suffer from a chronic sleep disorder [1]. Chronic sleep restriction has been shown to be associated with a variety of physiological consequences including increased heart rate and blood pressure [2], increased inflammation as measured by C-reactive protein [3], impaired glucose tolerance [4], and increased hunger/ appetite [5]. Insomnia is considered the most common and disabling sleep disorder [6] and is a condition associated with chronic sleep disruption and associated daytime function symptoms such as fatigue, difficulty concentrating, and mood disturbance [7].

Currently, the gold standard for discriminating between a patient suffering from a sleep disorder and a healthy sleeper is through the Pittsburgh Sleep Quality Index (PSQI). The PSQI is computed from a questionnaire that assesses sleep quality and disturbances, with a high PSQI indicating poor sleep quality [8]. As such, the PSQI is subjected to survey bias, and hence not entirely objective. Subjective assessments may lead to inaccurate diagnostics, which calls for a great necessity to implement a more objective, quantitative means for diagnosing sleep disturbances. One important diagnostic tool used to characterize sleep and diagnose sleep disorders is overnight polysomnography (PSG). At present, the traditional PSG measures are limited to diagnosing sleep apnea and involve costly human scoring with poor inter-scorer reliability. A polysomnogram collects, amongst other things, EEG data in a non-invasive way. The EEG data are subjected to a laborious and subjective "scoring" process by sleep specialists who assign a sleep stage to every 30-second interval of the EEG data. Clinicians then make a diagnosis based on the annotated data. Consequently, the standard procedure is currently heavily dependent upon human factors. Furthermore, the lack of utility of traditional PSG approaches for non-apnea sleep disorders has led to a reliance on, and general clinical acceptance of, purely subjective diagnostic criteria such as questionnaires and clinician interviews.

Spectral analysis is not a standard of care method for diagnosing or managing any sleep disorder. It remains a technique exclusively explored for its utility in further characterizing the underlying sleep neurophysiology in the research arena only. While some quantitative EEG methods have been utilized for some time to better understand basis of poor subjective sleep complaints notoriously reported in patients with fibromyalgia, chronic fatigue syndrome and insomnia, very few studies have evaluated sleep EEG in individuals demonstrating poor sleep quality. We are not aware of any studies that have attempted to operationalize spectral features as a diagnostic tool for sleep disorders such as insomnia [9].

Analysis of EEG signals present a possible alternative method to a more objective way of viewing PSG data. In this study, we used patients with extreme values of PSQI (very low vs. very high) to divide the subjects into two groups, good sleepers and bad sleepers. We looked

at sleep *behavioral* (macro-sleep architectural) features and sleep *spectral* (micro-sleep architectural) features from the annotated EEG data with the goal of utilizing current measures to create a more rigorous and quantitative means for the diagnostics of sleep disorders. We found that many of the scoring-independent spectral features are significantly different between the two groups, whereas none of the behavioral features or the scoring-dependent spectral features is able to significantly distinguish good sleepers from bad sleepers. The fact that the behavioral features computed from the scoring files cannot identify the expected difference in the EEG signals calls into question whether human scoring of EEG signals is practical in assessing sleep quality.

II. Methods

A. Experimental Setup

1) Study Population—Our study population consisted of 15 patients without any sleep disorder and 21 patients suffering from one or more sleep disorders, many of which had insomnia. All participants were African American males older than 35 years old. The groups were age-matched and heterogeneous in that each group contained both seronegative controls and seropositive HIV participants (Table 1).

All of the seropositive HIV participants were recruited at Johns Hopkins Medical Institutions (JHMI) from an established HIV-research cohort at JHU [the Northeastern AIDS Dementia (NEAD)], Central Nervous System HIV Antiretroviral Therapy Effects Research (CHARTER) and other available seropositive HIV patient research cohorts. Control participants were recruited from other JHMI research cohorts, advertisements, and from personal referral of established participants. This study was approved by the JHMI IRB and all participants provided informed consent prior to enrollment. A full medical evaluation was conducted to ensure that each participant was medically, cognitively, and psychologically stable to participate. Seropositive HIV participants were required to have a relatively low HIV viral load (3000copies/ml), and those whose cART regimen included efavirenz were excluded from the study due to its potential sleep-altering effects [10]. Participants were also dropped from the study if they screened positive for recreational drug use during the 2-week protocol.

2) EEG acquisition—The study data was collected between August 2008 and April 2011 for the HIV group and between May 2010 and April 2011 for the control group. The raw data was collected at a sampling rate of 500 Hz, which we downsampled to 80 Hz in our analysis. EEG signals were collected in a contralateral ear reference montage using 6 scalp electrodes and 2 ear electrodes (F3A2, F4A1, C3A2, C4A1, O1A2 and O2A1) (Fig. 1). A PSG was conducted in the Johns Hopkins Clinical Research Unit followed by 2-week-in-home functional assessments with questionnaires and actigraphy monitoring of their sleep and wake activity. The PSG device model used was the same across all participants.

3) Sleep Stage Scoring—The EEG data was visually scored according to the 2007 American Academy of Sleep Medicine (AASM) Manual for Scoring Sleep [11] by assigning a sleep stage to every 30 second epoch of the EEG data. The sleep cycle consists of five sleep stages, three non-REM stages; stage N1, stage N2, and stage N3, the REM stage, and

wake. A certified sleep specialist reviewed and finalized all of the studies, which were conducted and scored by a registered technician.

B. Data Analysis

1) Sleep Behavioral features—Three types of sleep behavioral features were computed from the scoring files: (i) sleep transitions, (ii) the percentage of time spent in each sleep stage, and (iii) the duration of time spent in each sleep stage. For each participant, a 5-by-5 sleep transitions matrix was created. Each cell in the matrix had a nonnegative integer value representing the count of a specific transition. The direction of a transition flowed from row to column (i.e. cell (1,2) stands for the number of transitions from stage N1 to stage N2. The diagonal of the matrix represents being in a certain sleep stage and staying in that stage for the subsequent epoch). Once a transitions matrix was created, it was normalized by the sum of the matrix to eliminate bias due to total sleep time. From these normalized transitions matrices, the percentage of time spent in each sleep stage was calculated for each patient by summing across the columns. The output was a vector of length 5 with the values representing the percentage of time spent in each of the five sleep stages. Lastly, assuming a patient goes into a sleep stage, a distribution of the duration of time spent in that sleep stage can be extracted. The values of the duration distributions were all non-zero positive multiples of 30 due to the nature of the scoring process.

2) Sleep Spectral features—The sleep spectral features computed were the total power over the entire night, as well as the average power in each of the five sleep stages and across five equally sized frequency bands of 5 Hz between 0–25 Hz. The features were obtained by inspecting the power spectrum of the EEG signals over the entire night (~8.5 hours). The spectrogram of the EEG signal for each subject was computed using the mtspecgramc command from the Chronux toolbox in MATLAB (R2014b) [12]. This uses a multi-taper estimate scheme based on Slepian functions for calculating the power spectrum of the signal. A sliding window of 3 seconds was used, incrementing by 1 second per step. The time-bandwidth product was 3, with 5 tapers used for estimation. The spectrogram was computed for the frequency of the signal ranging from 0–25 Hz.

From the log power spectrum for each patient, we computed the total average power over the entire night by averaging across all six channels and all frequencies of the signal. Although we observed nonstationaities in the data, we wanted to first test whether a very coarse metric would significantly differ between the two groups. Furthermore, we divided the log power spectrum into 5 bands of 5 Hz width (0–5 Hz, 5–10 Hz, …, 20–25 Hz) and calculated the average power in each frequency band. Finally we used the staging files to find the average power in each of the five sleep stages by averaging across all annotated 30 second windows for each sleep stage. All the measures were averaged across all channels since the channels all gave very similar results when observed separately.

3) Feature Visualization—We performed principal component analysis on the features for visualization purposes. Feature vectors were created by concatenating all cells of the transitions matrix. Other visualization methods include boxplots and histograms (Fig. 2).

4) Likelihood Ratio Test—In order to test the accuracy of each feature to discriminate between good and bad sleepers, we derived a distribution for each subject group and then employed a likelihood ratio classifier. Distributions were created with a bin size of 20. The likelihood ratio test combines the priors of each group with the probability of being in that group derived from a distribution obtained by leaving one patient out at a time. The patient left out is the patient tested. It then assigns the patient to the group with a higher likelihood [13].

5) Testing for Significance—P-values were calculated for sleep behavioral features by bootstrapping with no replacements. We ran a total of 10,000 iterations. For the spectral features the p-values were calculated using a two-sample t-test assuming unequal variances.

III. Results

The difference between the two groups was statistically insignificant ($P > 0.05$) for all the behavioral features, with the exception of four sleep transitions, stage N3 to stage N1, stage N3 to stage N2, stage N2 to stage N3, and wake to wake, without correction. However, with a Bonferroni correction, all features became insignificant (Table 2). For the spectral features, the total average power over the entire night was significantly different ($P = 0.0289$) between the two groups. Furthermore, the total average powers in the frequency bands 15–20 Hz and 20–25 Hz showed even more significant difference between the two groups ($P=0.0103$ and $P = 0.0159$, respectively). The average powers in the other frequency bands along with all scoring-dependent spectral features did not show a significant difference between the two groups ($P > 0.05$).

A. Accuracy of Likelihood Ratio Test

1) Sleep Behavioral features—Accuracy values for all sleep behavior features were too low to be considered a good classifier. Accuracy values ranged from 38.89% to 69.44% (Fig. 3). Typically, a feature with accuracy above 70% is considered a good classifier.

2) Sleep Spectral features—The accuracy of the likelihood ratio test using the scoring independent spectral features ranged from 40.54% – 72.97% whereas the accuracy for the scoring dependent features ranged from 45.95% – 56.76% (Fig. 4). A feature with accuracy above 70% is considered a good classifier.

IV. Conclusion

Sleep disturbances are known to be highly correlated to serious health concerns. In this study, we objectively looked at human-scored components of PSG as well as those that are human-independent. We demonstrated that while scoring-independent spectral features of EEG data can be used to significantly discriminate good sleepers from bad sleepers, scoring-dependent spectral features and sleep stage defined behavioral features are not able to identify the expected difference between the two groups. This brings up the issue of whether the costly and time-consuming scoring process is the best approach to analyzing data obtained through PSG. One possible feature we propose that yields a better separation between good sleepers and bad sleepers is average total power. Further research in this area

may lead to a less expensive, quicker, and objective measure that can assist physicians in making a diagnosis in the realm of sleep disorders.

Acknowledgments

References

1. [Retrieved March 29, 2015] Brain Basics: Understanding Sleep. 2014 Jul 25. from http://www.ninds.nih.gov/disorders/brain_basics/understanding_sleep.htm

2. Tochikubo O, Ikeda A, Miyajima E, Ishii M. Effects of Insufficient Sleep on Blood Pressure Monitored by a New Multibiomedical Recorder. Hypertension. 1996:1318–1324. [PubMed: 8641742]

3. Meier-Ewert HK, Ridker PM, Rifai N, Regan MM, Price NJ, Dinges DF, Mullington JM. Effect of sleep loss on C-Reactive protein, an inflammatory marker of cardiovascular risk. Journal of the American College of Cardiology. 2004; 43(4):678–683. http://doi.org/10.1016/j.jacc.2003.07.050. [PubMed: 14975482]

4. Spiegel K, Leproult R, Van Cauter E. Impact of sleep debt on metabolic and endocrine function. The Lancet. 1999; 354(9188):1435–1439. http://doi.org/10.1016/S0140-6736(99)01376-8.

5. Spiegel K, Tasali E, Penev P, Cauter EV. Brief Communication: Sleep Curtailment in Healthy Young Men Is Associated with Decreased Leptin Levels, Elevated Ghrelin Levels, and Increased Hunger and Appetite. Annals of Internal Medicine. 2004; 141(11):846–850. http://doi.org/10.7326/0003-4819-141-11-200412070-00008. [PubMed: 15583226]

6. Gamaldo CE, Spira AP, Hock RS, Salas RE, Mcarthur JC, David PM, Smith MT. Sleep, Function and HIV: A Multi-Method Assessment. AIDS and Behavior. 2013; 17(8):2808–2815. http://doi.org/10.1007/s10461-012-0401-0. [PubMed: 23299876]

7. Symptoms. (n.d.). [Retrieved March 29, 2015] from http://sleepfoundation.org/insomnia/content/symptoms.

8. Buysse DJ, Reynolds CF III, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index (PSQI): A new instrument for psychiatric practice and research. Psychiatry Research. 1989; 28(2):193–213. http://doi.org/10.1016/0165-1781(89)90047-4. [PubMed: 2748771]

9. Reid S, Dwyer J. Insomnia in HIV infection: A systematic review of prevalence, correlates, and management. Psychosomatic Medicine. 2005; 67(2):260–269. http://doi.org/10.1097/01.psy.0000151771.46127.df. [PubMed: 15784792]

10. Moyle G, Fletcher C, Brown H, Mandalia S, Gazzard B. Changes in sleep quality and brain wave patterns following initiation of an efavirenz-containing triple antiretroviral regimen. HIV Medicine. 2006; 7(4):243–247. http://doi.org/10.1111/j.1468-1293.2006.00363.x. [PubMed: 16630036]

11. Iber, C.; Ancoli-Israel, S.; Chesson, A.; Quan, SF. for the American Academy of Sleep Medicine. The AASM manual for the scoring of sleep and associated events: Rules, terminology and technical specifications. Westchester, IL: American Academy of Sleep Medicine; 2007.

12. Bokil H, Andrews P, Kulkarni JE, Mehta S, Mitra P. Chronux: A Platform for Analyzing Neural Signals. Journal of Neuroscience Methods. 2010; 192(1):146–151. [PubMed: 20637804]

13. [Retrieved March 28, 2015] 8.2.3.3. Likelihood ratio tests. (n.d.). from http://www.itl.nist.gov/div898/handbook/apr/section2/apr233.htm
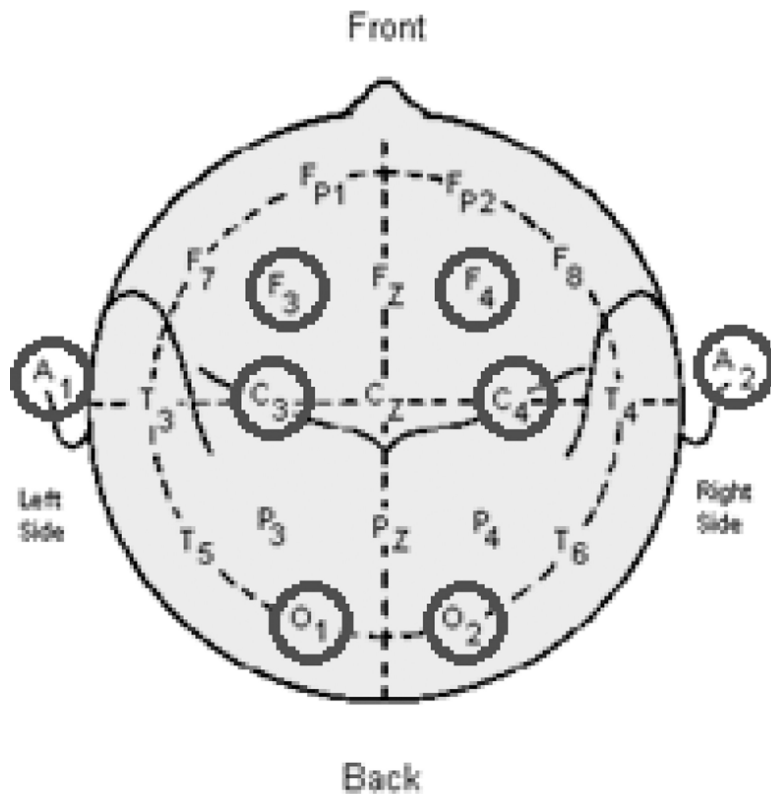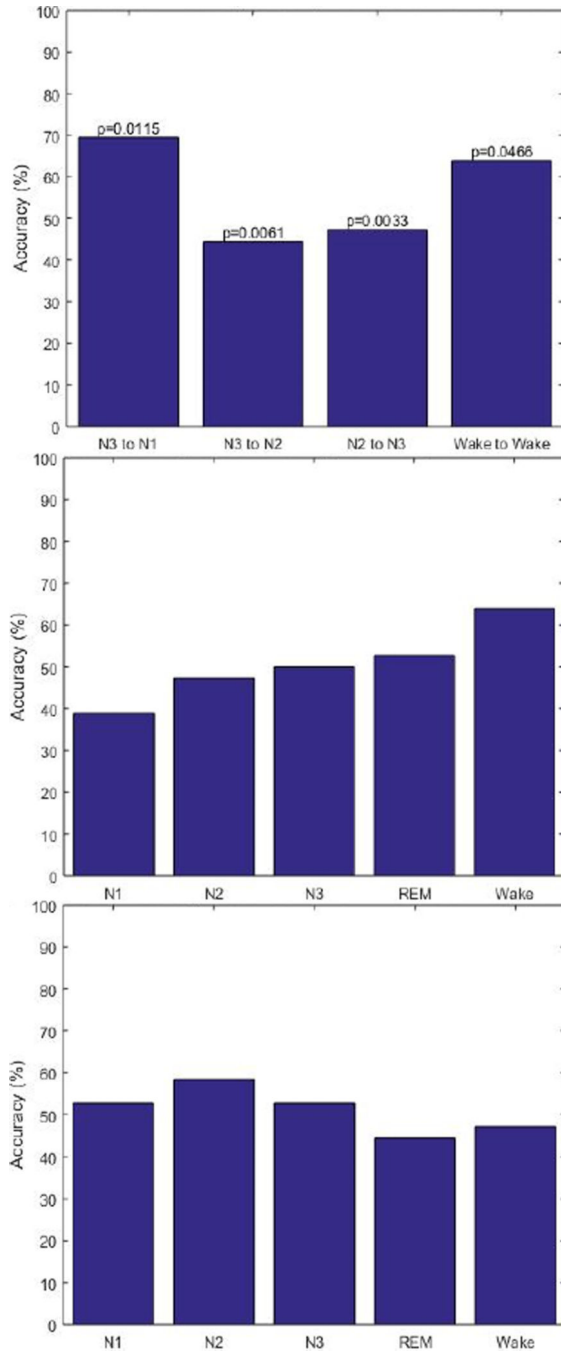
Figure 1.
EEG electrode placements in PSG
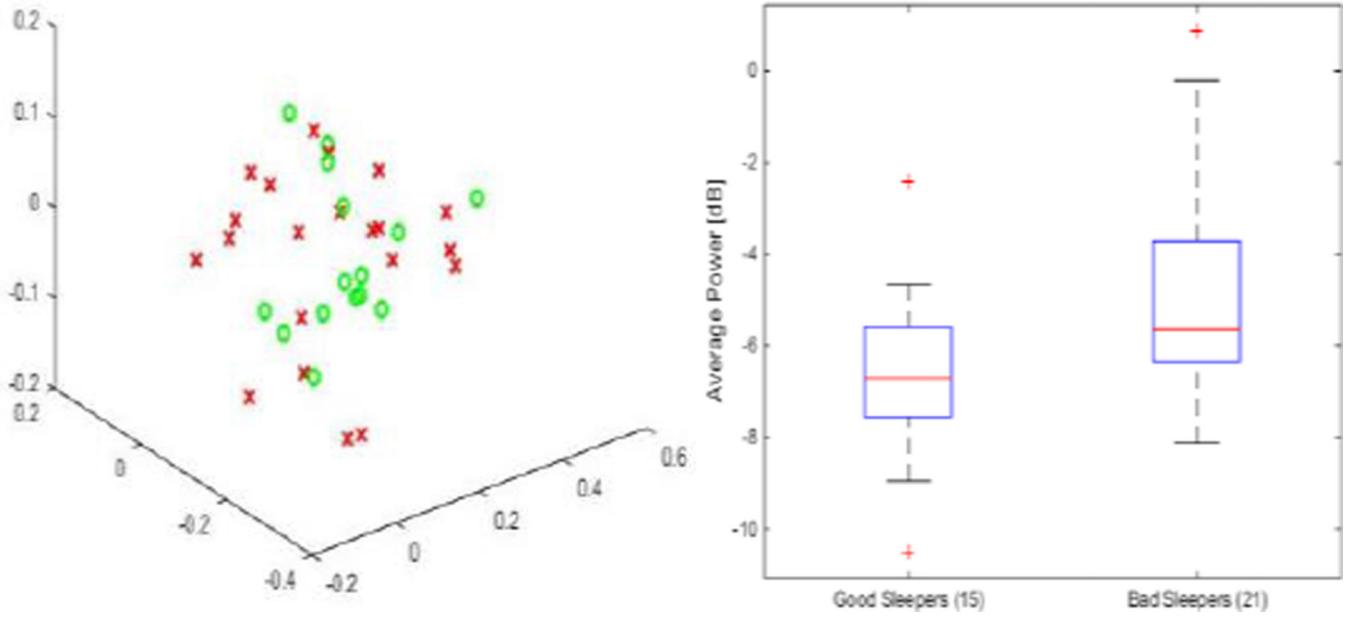
Figure 2.
Normalized counts of all sleep transitions as a feature vector in PC space. There is little to no separation between good sleepers and bad sleepers (left). The average power across the frequency band 20–25 Hz is one of the features that is significantly different between the two groups and gives the highest accuracy of classifying subjects into groups of good sleepers and bad sleepers (right)

Figure 3.
The accuracy of the likelihood ratio classification for sleep transitions with P < 0.05 without correction (top), the percentage of time spent in each sleep stage (middle) and the duration of time spent in each sleep stage (bottom).
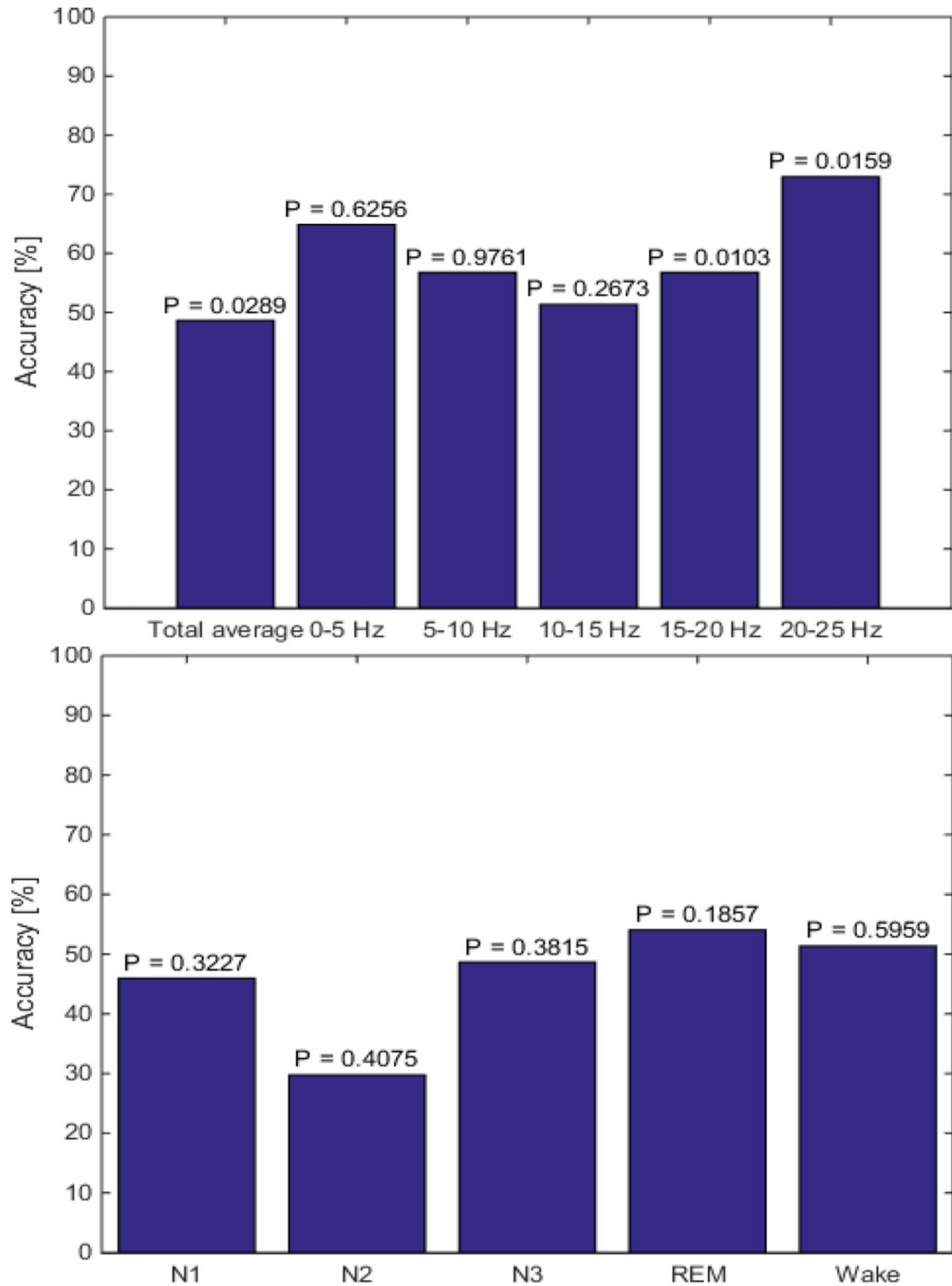
Figure 4.
The accuracy of the likelihood ratio classification for the scoring-independent spectral features (top). The highest accuracy was obtained using the average power across 20–25 Hz. The accuracy of the likelihood ratio classification for the scoring-dependent spectral features (bottom). None of these features were able to significantly discriminate good sleepers from bad sleepers.

TABLE I

Patient population statistics

Subject statistics	Good sleepers (n = 15)	Bad Sleepers (n = 21)
Number of HIV individuals	5	16
Number of controls	10	6
Age (P = 0.1249)	48.67 ± 6.35	52 ± 6.08
PSQI	3.33 ± 1.18	9.24 ± 2.91

TABLE II

P-values of sleep behavior features without correction. Underlined cells have p-values < 0.05

	Transitions Matrix					% Time	Duration
	N1	N2	N3	REM	wake		
N1	0.3265	0.8735	0.1595	0.6262	0.9695	0.4384	0.5691
N2	0.8194	0.3954	0.0033	0.2370	0.7202	0.2870	0.8970
N3	0.0115	0.0061	0.1902	0.4160	0.5497	0.1032	0.5024
REM	0.4710	0.0740	1.0000	0.9181	0.8876	0.8867	0.6480
wake	0.9810	0.1939	0.4200	0.8043	0.0466	0.0613	0.6708