Sleep Stage Classification with Cross Frequency Coupling*

Teresa H. Sanders[1], Mark McCurry[2], and Mark A. Clements[3]

Abstract— **Sleep is a key requirement for an individual's health, though currently the options to study sleep rely largely on manual visual classification methods. In this paper we propose a new scheme for automated offline classification based upon cross-frequency-coupling (CFC) and compare it to the traditional band power estimation and the more recent preferential frequency band information estimation. All three approaches allowed sleep stage classification and provided whole-night visualization of sleep stages. Surprisingly, the simple average power in band classification achieved better overall performance than either the preferential frequency band information estimation or the CFC approach. However, combined classification with both average power and CFC features showed improved classification over either approach used singly.**

I. INTRODUCTION

During a single night, a healthy individual transitions through multiple stages of sleep that can be classified as either rapid eye movement (REM) or non-rapid eye movement (NREM) sleep. REM sleep makes up approximately 20-25% of nightly sleep in adults, occurs in cycles of generally increasing length, and is characterized by rapid and random eye movement. NREM sleep can be divided into 3 or 4 stages, labeled as Stage 1 through Stage 3 or 4, with increasing numbers indicating deeper levels of sleep. It is well known that deep sleep (stages 3 or 4) and rapid eye movement (REM) sleep are important for brain health. Recently, new studies have underscored the importance of adequate amounts of certain sleep stages by identifying specific processes occurring during sleep that contribute to learning and memory [1], accelerate clearance of toxins [2], and promote healthy brain structure and function.

Electrophysiologically instrumented studies are commonly used to measure the stages and quality of sleep for diagnosis and treatment of sleep-related disorders. Typically the data from these studies are evaluated visually; however, some approaches for automatic sleep stage classification have been developed. Wavelet methods have been used in some studies to characterize sleep stages [3], [4]. In this study, we tested two standard methods along with a new method for classifying sleep stages that extends continuous wavelet transform (CWT) methods by evaluating CWT cross-frequency-coupling (CFC). Testing was performed using the "ST" (Sleep Telemetry) files from a 1994 study of temazepam effects on sleep available in the Sleep-EDF Database on the Physionet.org website [5].

*This work was sponsored by the Texas Instruments Leadership University.

[1]T.H. Sanders graduated in May 2014 with a Bioengineering PhD from the Georgia Institute of Technology. tsanders7 at gatech.edu

[2]M. McCurry is a Graduate Student in the School of Electrical and Computer Engineering, Georgia Institute of Technology, USA.

[3]M.A. Clements is a Professor in the School of Electrical and Computer Engineering, Georgia Institute of Technology, USA.

II. METHODS

The Physionet Sleep-EDF Database "ST" files contain recordings from 22 individuals (males and females) who had some difficulty falling asleep but were otherwise healthy. Data was collected during an overnight hospital visit using a miniature telemetry system described in [6]. The files analyzed in this paper were from the placebo nights when no temazepam was administered. The recordings analyzed were taken from the Fpz-Cz EEG electrodes. Sleep stages 1 through 4, REM sleep, and wakefulness were scored according to the hypnogram method described in [7]. For the analysis in this paper, sleep stages 3 and 4 were combined into one category (slow wave sleep (SWS) / deep sleep) to reflect the newer guidelines suggested by the American Academy of Sleep Medicine (AASM) [8]. In order to allow preliminary comparisons of multiple sleep stage classification approaches at different time scales, data from ten sleep study patients were selected at random from the database. A description of the data parameters for each subject is shown in Table I.

TABLE I
SLEEP SUBJECT DATA DESCRIPTIONS

Patient	Length	epochs	Awake	Stage 1	2	3/4	REM
ST7061	8.28 hr	994	36	89	589	10	270
ST7082	7.70 hr	924	148	170	450	45	111
ST7101	8.69 hr	1043	126	72	303	179	231
ST7112	8.18 hr	981	45	36	388	64	293
ST7121	8.58 hr	1030	65	34	452	120	267
ST7162	8.18 hr	982	153	102	429	128	80
ST7192	7.71 hr	925	16	21	545	59	231
ST7201	7.63 hr	916	14	65	609	66	134
ST7221	8.53 hr	1023	141	211	438	2	231
ST7241	8.43 hr	1012	20	38	670	38	236

The classification performance of three sets of features extracted from the data for all 10 patients were compared in order to determine the whether the CFC features offered any advantages over current methods.

A. Average spectral power method

The first set of features were based on the common power spectrum bands. The data was segmented into 1 second frames without overlap over which the power within each of the Delta (3-4 Hz), Theta (4-8 Hz), Alpha (8-13 Hz),

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

Low Beta (13-16 Hz), High Beta (16-30 Hz), and Gamma (30-58 Hz) frequency bands were recorded. The filters were implemented using short-time Fourier transforms (STFTs) followed by Hamming window smoothing. For each of these bands the features were normalized to unit variance and zero mean. For visualization, these features were averaged over many frames to generate a coarse resolution version. The graphs shown in Fig. 1 use a 150 second window. For classification, a 30 second epoch window was used.

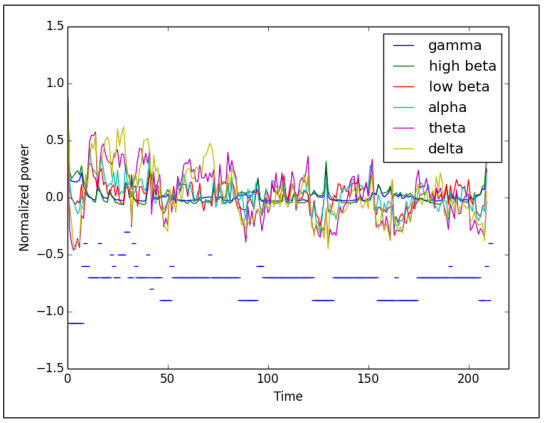

Fig. 1. Two minute average spectral power features for ST7241. The dots at bottom indicate the expert identified sleep state: waking (lowest y-axis value), and sleep stages (REM and stages 1-4) respectively shown with increasing y-axis value.

B. Preferential frequency band method

The second set of features were created as an adaptation of [9] using maximal frequency indicies within STFT time frames (Fig. 2). This proved insufficient for classification on the given data, so the procedure was modified as follows.

Let X be the STFT of the signal with no overlap and five second frames. Next, normalize the z-score of every frequency bin across the patient's entire recording. Let $Y_{i,j} < -1$ if $argmax_j X_{j,i} \forall i$ and 0 otherwise. Lastly, blur Y columnwise s.t. the energy placed in one frequency bin is spread to adjacent ones.

To obtain better frequency estimates, additional blurring can be done through time along each row as seen below [9]. This however was not done on the features used for classification to avoid any over optimistic results through nearest neighbor cases.

C. Cross-frequency-coupling method

For this method, the EEG data was z-scored for each epoch. Next, the continuous wavelet transform with a complex Morlet mother wavelet was used to extract phase and amplitude features corresponding to 6 frequency sub-bands from each 30 second epoch. The frequency bands used were the same as used in the average spectral power method. For each epoch, Phase-amplitude cross frequency coupling (CFC) features were calculated as described in [10] for each possible phase, amplitude sub-band combination (36 total). The length-36 column feature vectors for the epochs were then concatenated in an array and the rows of the

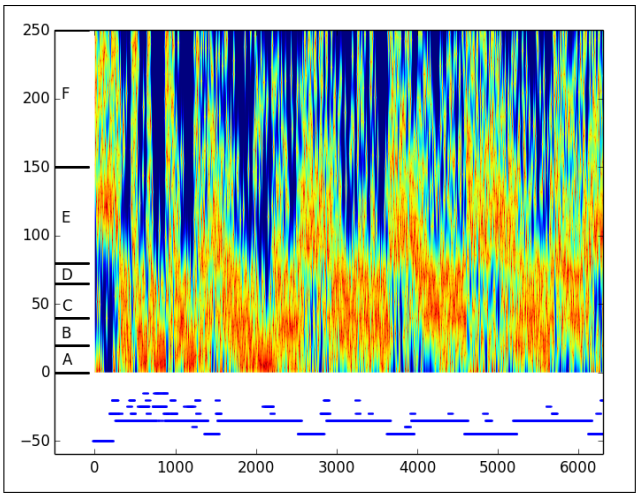

Fig. 2. Preferential frequency band features versus epoch for ST7241. On the y-axis, the letters refer to the approximate region of the Delta (A), Theta (B), Alpha (C), Low beta (D), High beta (E), and Gamma (F) frequency bands. The black dots at bottom indicate the expert identified sleep state: waking (lowest y-axis value), REM sleep, sleep stages 1-4 (shown with respectively increasing y-axis value).

array were z-scored in order to avoid variations in feature strength between rows due to differences in the sub-band powers. Variations between different rows occur due to power differences in the frequency bands.

Fig. 3 columns correspond to the CFC features for each recording epoch in chronological order. Colors represent the z-scored degree of phase-amplitude CFC in each of the amplitude sub-bands shown on the y-axis. The 6 rows in each amplitude sub-band group indicate the degree of CFC between the amplitude in the frequency band shown on the y-axis and the phase for each of the 6 sub-bands (listed in the same order as the amplitude sub-bands).

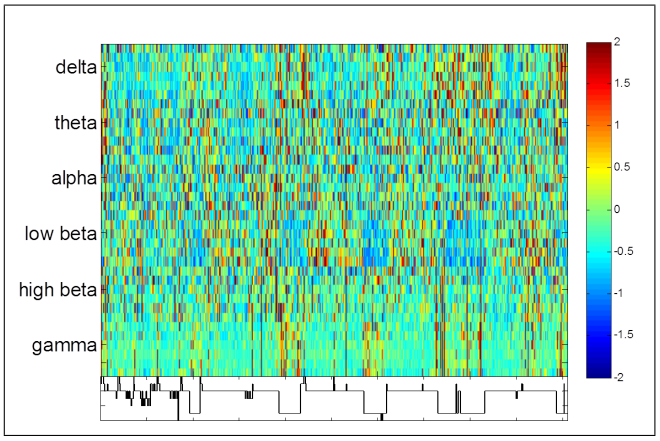

Fig. 3. Phase-Amplitude Coupling features versus 30 second epoch for ST7241. The black stair-step line at bottom indicates the expert identified stages: movement, REM, stages 4,3,2,1, and waking (from bottom to top stair-steps, respectively).

Each column in the image represents a 30 second time interval (epoch). Several regions of reduced (blue) or enhanced (red) modulation activity can be seen. In particular, blue patches in the low beta band are visible during REM sleep (broad low steps on black "stair-step" line).

D. Classification

Each of the 10 files analyzed in this study contained approximately 8 hours of sleep. Sleep scoring annotations (hypnograms) were available for each file. Annotations indicated one of the following for each 30 second epoch in the file: Movement, Waking, Rapid Eye Movement (REM) sleep, and non-REM sleep stages 1-4. The single stream of EEG data from the Fpz-Cz electrodes were used to perform the classification analysis.

The three types of features described previously in the methods were extracted from the EEG data. The corresponding labels from the annotation files were then used to enable Linear Discriminant Analysis (LDA) classification with 5-fold cross-validation. To test the performance of the three feature sets, each individual dataset was used as a test set based on training data from the other 9 individual datasets. These performance scores are shown as icons in Fig. 4. In addition to the simple LDA approach, a bootstrap aggregating (bagging) algorithm was used to improve the classification accuracy, and to test whether combining two of the feature sets might improve overall classification performance (Fig. 5). The bagging classifier used ensembles of decision trees with up to 200 trees.

E. ANOVA multiple comparison analysis

Multiple comparison analysis using pairwise ANOVA with Bonferroni correction was used to evaluate whether any individual features showed statistically significant differences in the sleep stage data populations. The resulting plots in Fig. 6, 7, and 8, show the mean values of the most statistically significant features for the four sleep stages considered (small filled circles). Lines indicate statistical significance boundaries. None of the lines in the three figures overlap, indicating these features have statistically significant differences for each sleep stage.

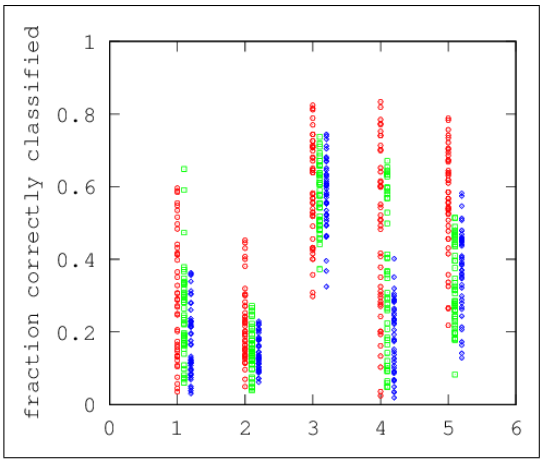

Fig. 4. Classification of 4 sleep stages (waking, Stage 1, Stage 2, Stage 3/4, and REM) using 3 different feature selection methods: average spectral power (red), peak powers (green), and cross-frequency-coupling (blue). Each dot represents the classification rate for one individual for the class indicated on the x-axis. For this figure, each individual was tested based on classifiers trained on the other 9 individuals.

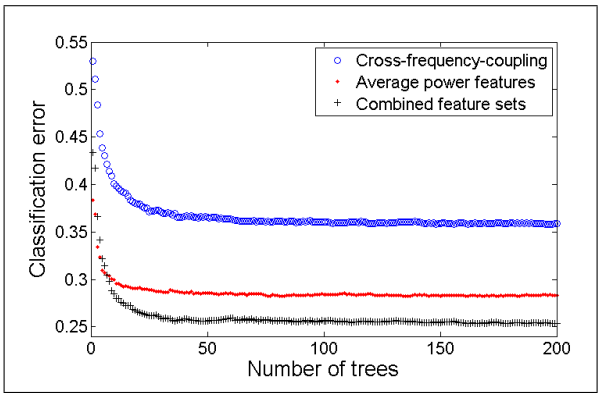

Fig. 5. Overall classification performance for 4 sleep stages using bagged classification with average spectral power (blue), cross-frequency-coupling (red), and combined features (green).

III. Results

Surprisingly, the simple average band power features (shown in red in Fig. 4) allowed better mean overall performance than did the preferential frequency region features (green) and the CFC features (blue). However, the performance rate varied more for different individuals with these features than with the other two feature types.

All methods performed better on average for stage 2 sleep, likely because of the large number of datapoints for stage 2 as seen in Table I. Stage 1 sleep was the least well classified. This may have been due to the fact that Stage 1 sleep is more transient in nature as can be see by the fluctuations near the beginning of the sleep label indicators underneath the example feature images in Fig. 2 and 3.

In general, no individual features allowed enough separation of the data for classification. However, for each of the three feature sets, multiple comparison analysis with Bonferroni correction was used to determine if any features showed statistically significant difference in the sleep stage data populations. The feature or group of features that showed the most statistically significant difference in the data populations for each stage occurred in different frequency regions as shown in Fig. 6, 7, and 8 for the average band power, preferential frequency band, and CFC feature selection methods. For the average band power, the most statistically significant feature was the alpha band power. For the preferential frequency band, the feature grouping with the most statistical significance was the high beta band, while for the CFC approach, the most statistical significance was found in the low beta amplitude features. Note that the proximity of the lines denoting statistical significance indicate that for the three features, the CFC feature differences are more statistically significant, while the average power method feature differences are the least statistically significant.

Since the most statistically significant features differed for the three methods, the performance of the combined average power and CFC MI features was compared to the overall performance for each feature set used separately. The bagging classification results, calculated with 5-fold cross-validation, show that the combined average power and CFC MI features performed better than either feature set on its own (Fig. 5). The preferential frequency band feature set

was not included in the combined testing due to it's large size (250 features).

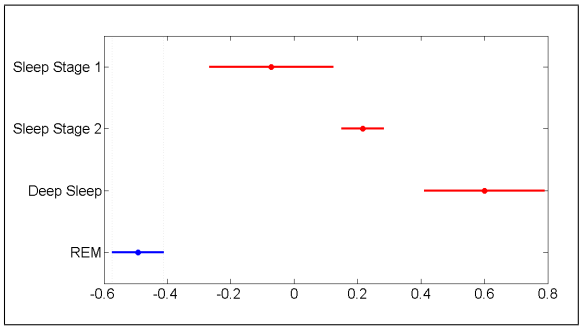

Fig. 6. Multiple ANOVA comparisons showing statistically significant differences between sleep stages in alpha band for average power features.

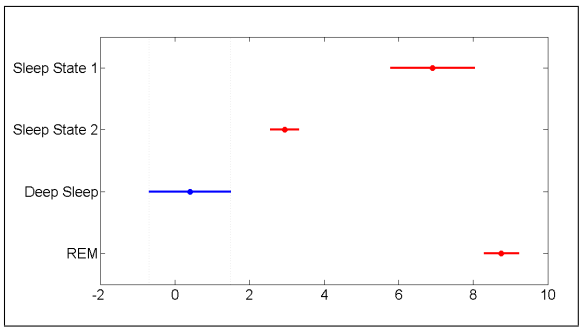

Fig. 7. Multiple ANOVA comparisons showing statistically significant differences between sleep stages in high beta band for preferential frequency band features.

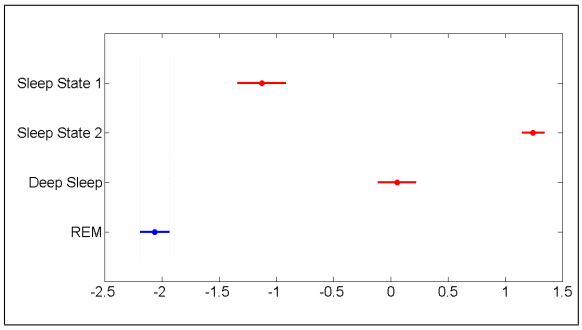

Fig. 8. Multiple ANOVA comparisons showing statistically significant differences between sleep stages in low beta band for CFC features.

IV. DISCUSSION

Although the best sleep classification performance shown in the results reached only 75% accuracy, it is important to note that the ground truth hypnogram sleep scores generated by expert sleep scorers can vary a great deal. For example, in one recent study [11], the interrater agreement for the R&K standard sleep-scored data was reported to be 80.6%.

Interestingly, the frequency bands yielding the most statistically significant differences between features for the 4 sleep states were unique for each of the three feature sets, suggesting that the information contained in each feature set may not be simply related to underlying average power changes, but may be conveying different information about the EEG signal for the sleep stages. The improved performance of the bagged classifier test with the combined average power and CFC MI features supports this hypothesis.

Since the CFC methods described in this study were previously shown to allow assessment of parkinsonism severity in epidural EEGs in non-human primates [10], it may be possible to apply the methods as a tool for gaining understanding about neural changes in individuals who suffer from a combination of sleep disorders and parkinsonism.

V. CONCLUSIONS

Sleep stages were classified at greater than chance levels using each of the three feature selection methods tested in this study. Combining average power and cross-frequency-coupling features in a bagging classifier with ensembles of decision trees yielded 75% accurate classification. However, based on a qualitative review of the sleep study hypnogram annotations compared to our automated scoring, manual visual sleep staging appeared to provide better annotations. Thus, the best use of the feature sets analyzed here may be to assist the sleep scoring expert, by allowing quantitative frequency band analysis, and visualization of the entire night's sleep in a single image.

REFERENCES

[1] R. Huber, M. F. Ghilardi, M. Massimini, and G. Tononi, "Local sleep and learning," *Nature*, vol. 430, no. 6995, pp. 78–81, 2004.

[2] L. Xie, H. Kang, Q. Xu, M. J. Chen, Y. Liao, M. Thiyagarajan, J. O'Donnell, D. J. Christensen, C. Nicholson, and J. J. Iliff, "Sleep drives metabolite clearance from the adult brain," *science*, vol. 342, no. 6156, pp. 373–377, 2013.

[3] C. R. Pinnegar, H. Khosravani, and P. Federico, "Time-frequency phase analysis of ictal EEG recordings with the S-transform," *Biomedical Engineering, IEEE Transactions on*, vol. 56, no. 11, pp. 2583–2593, 2009.

[4] M. R. Azim, M. S. Amin, S. Haque, M. Ambia, and M. A. Shoeb, "Feature extraction of human sleep EEG signals using wavelet transform and Fourier transform," in *Signal Processing Systems (ICSPS), 2010 2nd International Conference on*, vol. 3. IEEE, pp. 701–705.

[5] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[6] B. Kemp, A. Janssen, and M. Roessen, "A digital telemetry system for ambulatory sleep recording," *Sleep-Wake Research in The Netherlands 4, eds ML Coenen and J Arends, Dutch Society for Sleep Wake Research*, pp. 129–132, 1993.

[7] A. Rechtschaffen and A. Kales, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects," 1968.

[8] "The American Academy of Sleep Medicine manual for the scoring of sleep and associated events: Rules," *Terminology and Technical Specifications. AASM*, 2007.

[9] P. S. Low, *A New Way to Look at Sleep*. United States Library of Congress, 2007.

[10] T. H. Sanders, A. Devergnas, T. Wichmann, and M. A. Clements, "Canonical correlation to estimate the degree of parkinsonism from local field potential and electroencephalographic signals," in *Neural Engineering (NER), 2013 6th International IEEE/EMBS Conference on*. IEEE, pp. 158–161.

[11] H. Danker-hopfe, P. Anderer, J. Zeitlhofer, M. Boeck, H. Dorn, G. Gruber, E. Heller, E. Loretz, D. Moser, and S. Parapatics, "Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard," *Journal of sleep research*, vol. 18, no. 1, pp. 74–84, 2009.