Unsupervised Sleep and Wake State Identification in Long-Term Electrocorticography Recordings

Samantha Sun[1,6], Linxing Preston Jiang[2,6], Steven M. Peterson[3,6], Jeffrey Herron[4,6], Kurt Weaver[5,6], Andrew Ko[4,6], Jeffrey Ojemann[4,6], Rajesh P. N. Rao[2,6]

***Abstract*— Studying the neural correlates of sleep can lead to revelations in our understanding of sleep and its interplay with different neurological disorders. Sleep research relies on manual annotation of sleep stages based on rules developed for healthy adults. Automating sleep stage annotation can expedite sleep research and enable us to better understand atypical sleep patterns. Our goal was to create a fully unsupervised approach to label sleep and wake states in human electrocorticography (ECoG) data from epilepsy patients. Here, we demonstrate that with continuous data from a single ECoG electrode, hidden semi-Markov models (HSMM) perform best in classifying sleep/wake states without excessive transitions, with a mean accuracy (n=4) of 85.2% compared to using K-means clustering (72.2%) and hidden Markov models (81.5%). Our results confirm that HSMMs produce meaningful labels for ECoG data and establish the groundwork to apply this model to cluster sleep stages and potentially other behavioral states.**

I. INTRODUCTION

Sleep remains an elusive topic in neuroscience, having intricate relationships with learning and memory consolidation[1]. Disturbances of normal sleep patterns frequently emerge and interact with other medical conditions, such as epilepsy[2], [3] and many mental disorders[4], and the reasons and effects are poorly understood. Studying the dynamic interactions between sleep and neurological impairments can aid towards building a comprehensive understanding of the disorders and improve patient outcomes.

To study sleep, researchers commonly perform overnight electroencephalography (EEG), and trained technicians label the different sleep stages: N1, N2, N3, and REM[5]. Manual sleep staging involves visually examining 30-second segments of neural data and labeling each segment based on established sleep scoring rules. These rules include decision-based logic and identifying specific neural features. A key challenge for sleep staging is that some rules are subject to interpretation, leading to labeling differences between technicians[6]. These rules were also developed based on healthy adult subjects, allowing little extrapolation to other age groups or to those with atypical sleep patterns.

*This work was supported by the National Science Foundation (1630178 and EEC-1028725) and the University of Washington Big Data for Genomics and Neuroscience Training Grant (1T32CA206089-01A1)

[1]Department of Bioengineering
[2]Paul G. Allen School of Computer Science & Engineering
[3]Department of Biology
[4]Department of Neurological Surgery
[5]Department of Radiology
[6]Center for Neurotechnology
All authors within the University of Washington, Seattle, WA 98195 USA

With rapid developments in machine learning, automated sleep staging can resolve these issues by mitigating human bias and providing flexibility to variable sleep patterns. Current efforts in automation primarily use EEG in combination with supervised or unsupervised machine learning models[7], [8], [9]. However, little work has been done with electrocorticography (ECoG), with increased spatial resolution and signal-to-noise ratio to capture and reveal subtle brain dynamics related to sleep. Kremen et al. [10] demonstrated using a single intracranial electrode to classify Wake, N2, and N3 stages with 94% accuracy. Their unsupervised model used decision trees with numerical thresholds that required tuning from pre-existing sleep stage-labeled data, which may not always be available. Questions remain regarding whether a threshold-based approach is generalizable to variable sleep patterns. We believe that leveraging sequence-based methods can overcome these limitations and contribute to automation tools for sleep staging.

One nuance is that our ECoG recordings are of patients with intractable epilepsy, a neurological disease known to have complex reciprocal interactions with sleep and often presents with disrupted sleep patterns [2]. Successful development of a sleep stage classifier in these patients must be robust towards variable sleep dynamics between subjects. An unsupervised approach is ideal due to self-learning of features with high distinguishing properties, providing flexibility to differences in sleep patterns or electrode placement. Unsupervised methods also alleviate the impracticality of labeling sleep stages in every patient for model training.

In this paper, we built an unsupervised hidden semi-Markov model (HSMM) to label sleep and wake states in epilepsy patients, using spectral power features from a single ECoG electrode. We show that HSMMs produce higher labeling accuracy and fewer extraneous transitions than k-means clustering and hidden Markov models. Our results provide initial confirmation of HSMM as an effective tool for automated sleep staging, and motivate future applications of HSMM to neural behavioral state classification.

II. METHODS

A. Data Collection

Our data consist of continuous neural recordings from patients at Harborview Medical Center (Seattle, Washington) undergoing clinical monitoring for intractable epilepsy. As part of their monitoring procedure, an 8x8 electrocorticography (ECoG) grid (2.3 mm exposed diameter, Ad-tech Medical, Racine, WI, USA) was implanted on the patient's

978-1-7281-1990-8/20/$31.00 ©2020 IEEE

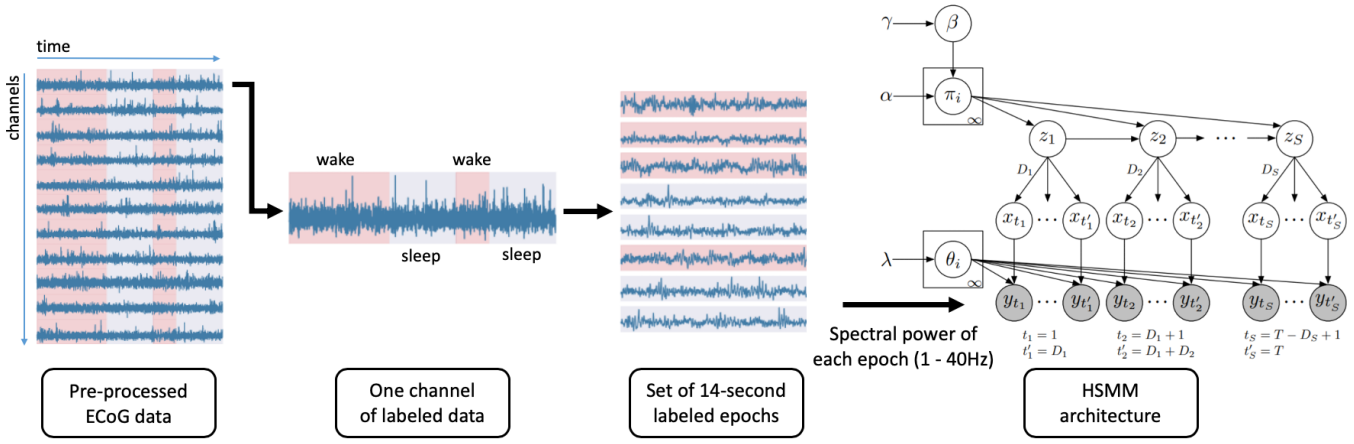

Fig. 1. **Data processing pipeline:** ECoG data were pre-processed (bandpassed 1-200Hz, notch filtered at noise harmonics, downsampled to 500 Hz), and a single channel was used for sleep/wake clustering. Data from this channel were epoched to 14-second segments with corresponding sleep/wake labels. Spectral power of each epoch (1-40Hz) was used as input to the HSMM. **HSMM Architecture** (adopted from Johnson et al.)[11]: β is drawn from $GEM(\gamma)$, α and β define the Dirichlet process (DP) prior, π represents the transition probability drawn from $DP(\alpha, \beta)$, λ represents a uniform Gaussian prior for observations, and D represents a Poisson prior for duration states. x and y represent the states and observations.

cortical surface, providing 24-hour continuous ECoG data. Data collection was approved by the Institutional Review Board of the University of Washington, and all subjects gave written, informed consent. We sampled approximately 4-14 hour continuous segments of ECoG data from two patients over three days and one day from two additional patients, ensuring that there was no seizure activity during the recording. Sleep and wake labels were coarsely annotated based on patient room video monitoring and synchronized to neural data using timestamp information.

Our signal processing pipeline is summarized in Figure 1. We minimally processed the ECoG signal to remove noise and baseline drift, including band-pass filtering 1-200Hz, notch filtering at noise harmonics (60Hz, 120Hz, 180Hz), and down-sampling to 500Hz. Electrodes with excessive noise or signal artifacts were removed from analysis. We selected the first grid channel, typically near the superior frontal gyrus, across patients for clustering to minimize computational cost. We split the continuous data into 14-second epochs with their corresponding sleep/wake labels. Total sleep and wake times are summarized in Table I.

TABLE I

SUBJECT SLEEP/WAKE TOTAL TIMES. TOTAL NUMBER OF EPOCHS (T = 14S) ALSO REPORTED.

	Subj 1	Subj 2	Subj 3	Subj 4
total wake time (hr)	25.2	9.5	4.6	5.5
total sleep time (hr)	17.7	11.1	3.5	2.7
total epochs (n)	11053	5295	2074	2117

B. Feature Selection

Neural analysis commonly involves decomposing the time-series signal into its frequency components due to the association between specific frequency bands and neural activity[12]. Sleep and its progression through different sleep stages are often characterized by shifting towards more low frequency rhythms. For the purpose of sleep/wake clustering, we used frequencies 1–40 Hz to capture a broadband range of frequencies, which we believe is sufficient for distinguishing between wake and sleep. We calculated the spectral power from each integer frequency within that range using Welch's method[13] across epochs. These vectors of length 40 were the inputs to our unsupervised models.

C. Model Architecture

Sleep has sequential staging with distinct spectral properties, so computational models that leverage sequential properties may better model sleep patterns. We propose an unsupervised implementation of hidden semi-Markov models (HSMMs) as seen in Figure 1[14]. Compared to hidden Markov models (HMMs), HSMMs allow hidden states to have various lengths instead of single-unit transitions in HMMs. Given that individual sleep stages can range from minutes to almost an hour, and sleep and wake periods commonly last for extended periods of time, we believe HSMMs are most suitable for our sleep/wake classification task. We additionally implemented K-means clustering and HMMs to compare across other unsupervised methods.

To determine the prediction label of each cluster, we followed the purity measure[15] of clustering performance – each cluster was assigned the label having the most counts within that cluster. We evaluated and reported accuracy across our unsupervised methods and across subjects.

1) K-means Clustering: We first ignored the sequential nature of the data and used the performance of k-means clustering as a baseline. k-means finds the best k points that minimize the overall Euclidean distance between our data and the corresponding centroids, and in our case, we set there to be two centroids to represent the clustering of sleep and wake epochs.

2) Hidden Markov Model: Our HMM was initialized with two states (sleep, wake) with transition probabilities of 0.9 (self) and 0.1 (change). Emission probabilities were set as a multivariate Gaussian, and transition and emission parameters were updated using expectation-maximization (EM). We

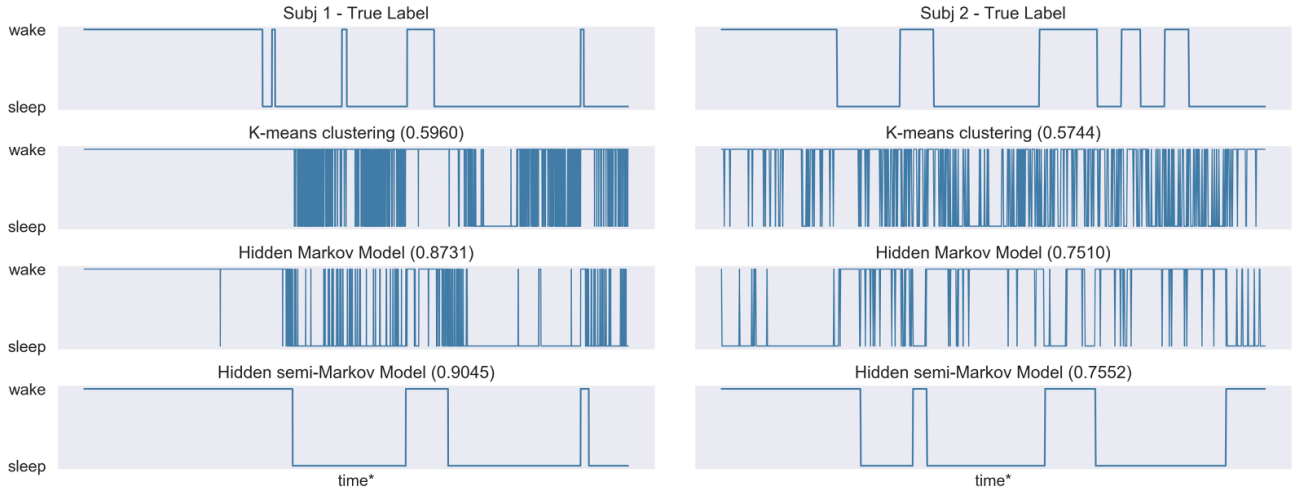

Fig. 2. Subject 1, 2 sleep/wake true labels and clustering comparison across three methods over one day. Model accuracy for Subject 1 and 2 are shown in parentheses next to model used. HSMM produced the highest accuracy across models and minimized unnecessary transitions.

used Viterbi inference[16] to produce the best state sequence for the expectation step, and maximum likelihood estimation (MLE) to update parameter values for the maximization step, iterating until convergence. We ran Viterbi inference one final time to produce the predicted sleep/wake sequence.

3) Hidden Semi-Markov Model: We followed a Bayesian non-parametric approach[11] and used a Dirichlet process (DP) prior for sampling the transition probability between states and a Poisson prior for the state duration length. Figure 1 shows a graphical representation of the model. To keep the computational time tractable, we limited the maximum number of states the model can use to four while allowing the model to select the ideal number of states. We set the conjugate prior to $Gamma(1000,5)$ for inferring the Poisson state duration. We used $\gamma = 0.6, \alpha = 0.6$ for Dirichlet process priors. The parameters were estimated through posterior samples from Gibbs sampling[17] for 150 iterations. While we did not define a convergence criterion, the model converged after an average of 110 re-samplings.

III. RESULTS

Figure 2 summarizes sleep/wake clustering performance across our three unsupervised models on one day of data from Subjects 1 and 2. As expected, k-means clustering resulted in the lowest classification accuracy of 59.6% (Subject 1) and 57.4% (Subject 2). HMM performed better with an accuracy of 87.3% (Subject 1) and 75.1% (Subject 2). The HSMM performed best out of the three, with an accuracy of 90.4% (Subject 1) and 75.5% (Subject 2). While the HSMM accuracy was only slightly higher than the HMM accuracy for Subject 2, the HSMM visually provided more accurate transitions between sleep and wake states. We attributed these infrequent transitions to the additional duration component in the HSMM, which seemed to create more strict requirements in order for transitions to occur. However, these stricter guidelines may explain why the HSMM did not capture some short-duration wake states.

When comparing model performance across all days for all subjects (Table II), HSMM produced the best accuracy except in Subject 4. All models tended to perform worse for Subjects 3 and 4. Visualizing the HSMM performance across subjects (Figure 3), we noticed that while the HSMM captured the general pattern of Subjects 1 and 2, albeit missing some short duration transitions, it poorly matched the the pattern for Subjects 3 and 4. One explanation for variable performance across subjects is that we froze model parameters, and these parameters may perform better in subjects with longer state durations than in others. These results suggest that generating subject-specific parameters is necessary to better handle variable sleep patterns.

TABLE II

ACCURACY OF UNSUPERVISED MODELS ACROSS INDIVIDUAL SUBJECTS AND OVERALL (WEIGHTED ON SUBJECT DATA LENGTH).

	Subj 1	Subj 2	Subj 3	Subj 4	Overall
k-means	0.7331	0.7014	0.6451	0.6476	0.7217
HMM	0.8523	0.8042	0.7623	0.6840	0.8146
HSMM	0.8939	0.8410	0.7956	0.6736	0.8517

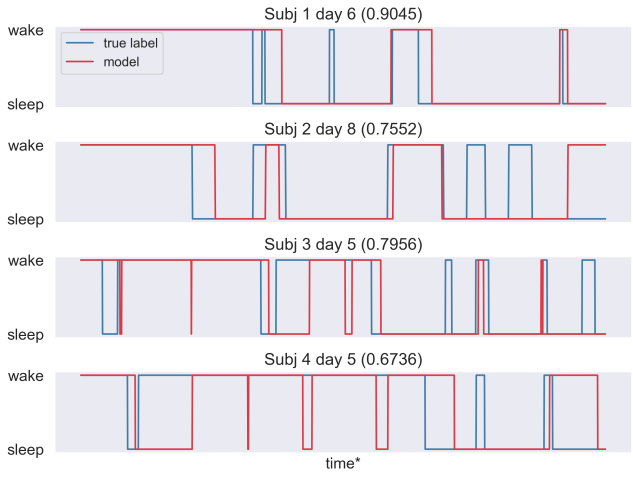

Fig. 3. Inter-subject comparison of HSMM performance in classifying sleep/wake states. Performance varies across subjects, capturing major sleep/wake transitions but failing to capture some short-duration states.

To determine if electrode selection was critical in HSMM performance, we ran the model for one day of Subject 1 using each electrode individually and reported model accuracy (Figure 4). Accuracy ranged from 60.2% to 97.5%, with more than half the electrodes having greater than 90% accuracy. The lesser performing electrodes were located along or just superior to the Sylvian fissure, which suggests a regional dependence of whether the electrode is a good candidate for model input.

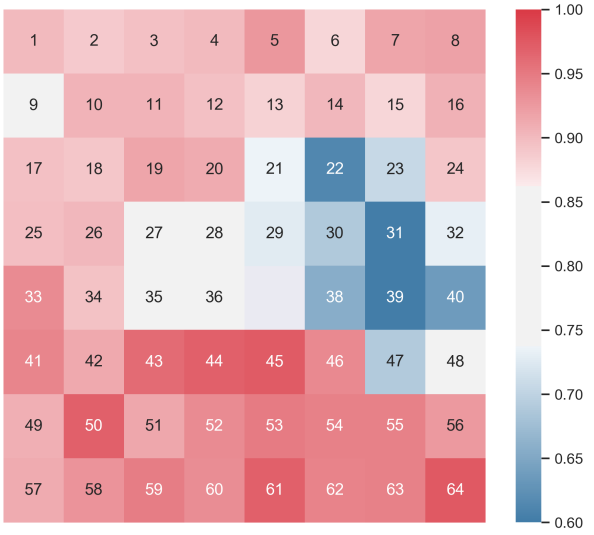

Fig. 4. Comparison of HSMM accuracy when using different ECoG electrodes (1-64) in Subject 1. Most electrodes produced an accuracy greater than 90%, while a smaller set in the center-right produced lower accuracies. Electrode 37 was omitted due to excessive signal artifacts.

IV. DISCUSSION

We demonstrated a Bayesian non-parametric HSMM approach for unsupervised labeling of sleep and wake states while incorporating smooth state transitions. While HSMM performance varied among subjects, HSMM consistently performed best across the unsupervised methods. We also showed that in one subject, HSMM accuracy remained consistent when using different electrodes, noting that electrode location had considerable importance to model performance and should be further explored. These results serve as a benchmark of using a Bayesian non-parametric approach for unsupervised classification of behavioral states in ECoG.

One limitation of using advanced machine learning models is that the abstraction tends to shift away from biological interpretability. The benefit of HSMM is that it explicitly models a duration distribution, which enables us to provide prior knowledge for the duration of sleep and wake segments. Such an approach is highly interpretable in terms of biological significance but may require modification to account for variability in sleep patterns for successful state classification across a variety of people.

With the goal of automating sleep staging in ECoG, future work will employ this HSMM on data with sleep stage labels. The next step would be to stratify the sleep states for further clustering into sleep stages. Since this dataset is uniquely from epilepsy patients, additional seizure state identification is critical to ensure seizure events are classified separately. Due to the high comorbidity between epilepsy and impaired sleep, labeling both sleep stages and seizure events could provide new insights into the interplay between sleep and epilepsy and translate clinically as a tool for seizure labeling, prediction, and localization. Improving tools in automated behavioral state classification will allow broader application to other neurological states and impairments, and will open possibilities in characterizing a range of behavioral states and their interactions.

V. ACKNOWLEDGMENT

We thank the patients for their participation in this research and Dr. Su-in Lee for providing initial project feedback.

REFERENCES

[1] B. Rasch and J. Born, "About sleeps role in memory," *Physiological Reviews*, vol. 93, pp. 681–766, Apr. 2013.
[2] M. Méndez and R. A. Radtke, "Interactions between sleep and epilepsy," *Journal of Clinical Neurophysiology*, vol. 18, pp. 106–127, Mar. 2001.
[3] L. Dorris, N. Scott, S. Zuberi, N. Gibson, and C. Espie, "Sleep problems in children with neurological disorders," *Developmental Neurorehabilitation*, vol. 11, pp. 95–114, Jan. 2008.
[4] T. Roth, S. Jaeger, R. Jin, A. Kalsekar, P. E. Stang, and R. C. Kessler, "Sleep problems, comorbid mental disorders, and role functioning in the national comorbidity survey replication," *Biological Psychiatry*, vol. 60, pp. 1364–1371, Dec. 2006.
[5] R. B. Berry, R. Budhiraja, D. J. Gottlieb, D. Gozal, C. Iber, V. K. Kapur, C. L. Marcus, R. Mehra, S. Parthasarathy, S. F. Quan, S. Redline, K. P. Strohl, S. L. D. Ward, and M. M. Tangredi, "Rules for scoring respiratory events in sleep: Update of the 2007 AASM manual for the scoring of sleep and associated events," *Journal of Clinical Sleep Medicine*, vol. 08, pp. 597–619, Oct. 2012.
[6] H. Danker-hopfe, P. Anderer, J. Zeitlhofer, M. Boeck, H. Dorn, G. Gruber, E. Heller, E. Loretz, D. Moser, S. Parapatics, B. Saletu, A. Schmidt, and G. Dorffner, "Interrater reliability for sleep scoring according to the rechtschaffen & kales and the new AASM standard," *Journal of Sleep Research*, vol. 18, pp. 74–84, Mar. 2009.
[7] H. Sun, J. Jia, B. Goparaju, G.-B. Huang, O. Sourina, M. T. Bianchi, and M. B. Westover, "Large-scale automated sleep staging," *Sleep*, vol. 40, Sept. 2017.
[8] A. Malafeev, D. Laptev, S. Bauer, X. Omlin, A. Wierzbicka, A. Wichniak, W. Jernajczyk, R. Riener, J. Buhmann, and P. Achermann, "Automatic human sleep stage scoring using deep neural networks," *Frontiers in Neuroscience*, vol. 12, Nov. 2018.
[9] M. Längkvist, L. Karlsson, and A. Loutfi, "Sleep stage classification using unsupervised feature learning," *Advances in Artificial Neural Systems*, vol. 2012, pp. 1–9, 2012.
[10] V. Kremen, B. H. Brinkmann, J. J. V. Gompel, M. Stead, E. K. S. Louis, and G. A. Worrell, "Automated unsupervised behavioral state classification using intracranial electrophysiology," *Journal of Neural Engineering*, vol. 16, p. 026004, Jan. 2019.
[11] M. J. Johnson and A. S. Willsky, "Bayesian nonparametric hidden semi-markov models," 2012.
[12] G. Buzsaki, *Rhythms of the Brain*. Oxford University Press, 2006.
[13] P. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, pp. 70–73, June 1967.
[14] S.-Z. Yu, "Hidden semi-markov models," *Artificial Intelligence*, vol. 174, pp. 215–243, Feb. 2010.
[15] C. D. Manning, *Introduction to Information Retrieval*. Cambridge University Press, Jul 2008.
[16] G. Forney, "The viterbi algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, 1973.
[17] B. Walsh, "Markov chain monte carlo and gibbs sampling," 2004.