EEG AND ECG CHARACTERISTICS OF HUMAN SLEEP COMPOSITION TYPES

Amro Khasawneh[1], Sergio A. Alvarez[2], Carolina Ruiz[1], Shivin Misra[1] and Majaz Moonis[3]
[1]*Dept. Computer Science, Worcester Polytech. Inst., Worcester, MA 01609, U.S.A.*
[2]*Dept. Computer Science, Boston College, Chestnut Hill, MA 02467, U.S.A.*
[3] *Dept. Neurology, U. Massachusetts Medical School, Worcester, MA 01655, U.S.A.*

Keywords: Clustering, EEG, ECG, HRV, Sleep.

Abstract: Unsupervised clustering of staged human polysomnographic recordings reveals a hierarchy of sleep composition types described primarily by sleep efficiency and slow wave sleep content. Associations are found between these sleep clusters and health-related variables including BMI, smoking habits, and heart disease, showing that sleep types correspond to objective and medically relevant groupings. The present work describes the sleep type hierarchy, and studies the EEG and ECG correlates of sleep composition type. It is found that measures of EEG variation such as δ, θ, and α spectral content, as well as average heart rate, and measures of heart rate variability, including the standard deviation of the sequence of RR intervals, and Hjörth activity and mobility of the ECG signal, differ significantly among sleep composition type clusters. EEG analysis is shown to allow approximate reconstruction of sleep type without the need for ECG data, while ECG alone is found to be insufficient for accurate sleep type classification.

1 INTRODUCTION

The idea that human sleep may be segmented into identifiable stages based on electrical potentials measured on the surface of the scalp dates back at least as far as the work of Loomis et al in the 1930's (Loomis et al., 1937). Contemporary all-night human sleep studies employ not only electroencephalography (EEG), which records electrical brain potentials, but also electrocardiography (ECG), which records heart potentials, electrooculography (EOG), which records eye movements, and electromyography (EMG), which records chin muscle movements (Kryger et al., 2005). Following standard rules of sleep scoring, a sleep technician scores these polysomnographic recordings into sleep stages one epoch (typically 30 sec) at a time, resulting in a sequence of sleep stage labels known as a hypnogram. Fig. 1 shows a hypnogram, from one of the polysomnographic recordings used for the present paper, that is labeled according to the classical Rechstchaffen and Kales (R&K) staging standard (Rechtschaffen and Kales, 1968). As illustrated in Fig. 1, sleep typically follows an overall temporal pattern of sleep stages, with several cycles involving alternation between REM (Rapid Eye Movement) and NREM (non-REM) sleep during the night, with a greater fraction of slow-wave sleep (stages NREM 3, 4) during the first half of the night, and a greater fraction of stage NREM 2 sleep during the second half of the night. Occasional stage fragmentations, including periods of wakefulness after the beginning of sleep, are also observed.

1.1 Scope of the Paper

The present paper describes work by the authors that shows that staged human sleep studies may be grouped naturally into a small number of distinct sleep types, each of which is characterized by a different sleep stage composition, with sleep efficiency and time in slow wave sleep differentiating among sleep types. Associations between sleep type and health-related factors such as age, Body-Mass Index (BMI), and smoking are also found, which supports the objectivity and medical relevance of sleep types. The previous work (Khasawneh et al., 2010) only considers all-night summaries of staged sleep studies, while full polysomnographic recordings comprise more detailed electroencephalogram (EEG) and electrocardiogram (ECG) time series data. The present paper also addresses the underlying EEG and ECG data, fo-

Khasawneh A., A. Alvarez S., Ruiz C., Misra S. and Moonis M..
EEG AND ECG CHARACTERISTICS OF HUMAN SLEEP COMPOSITION TYPES.
DOI: 10.5220/0003173900970106
In *Proceedings of the International Conference on Health Informatics* (HEALTHINF-2011), pages 97-106
ISBN: 978-989-8425-34-8

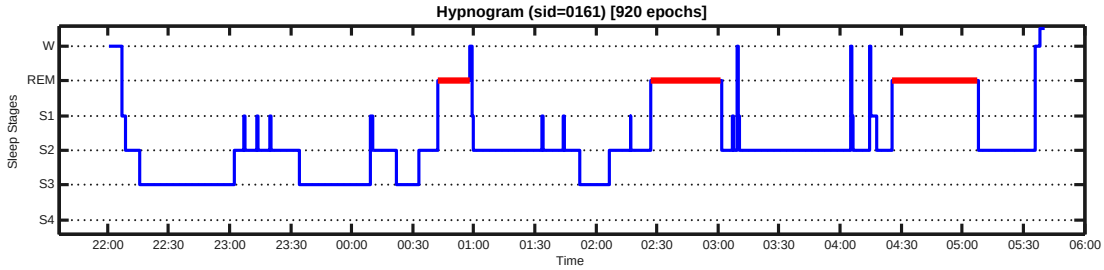

Figure 1: Hypnogram from the present study, staged by Rechstchaffen and Kales standard.

cusing on variations in summary statistics of the EEG and ECG signals among sleep composition types. It is found that heart rate, Hjörth activity and mobility, and normalized low frequency content of ECG all differ significantly among sleep types. These results provide further evidence that sleep types constitute objective and medically relevant concepts. Furthermore, it is found that analysis of EEG signals alone, in the absence of ECG data, is sufficient for approximate reconstruction of sleep type based on unstaged sleep data. On the other hand, ECG data alone does not provide accurate sleep type classification.

1.2 Related Work

Previous works consider variations in sleep composition associated with factors such as medication (Smith et al., 2006), smoking (Zhang et al., 2006), the practice of yoga (Sulekha et al., 2006), body composition (Rao et al., 2009), handedness (Propper et al., 2007), and autism spectrum conditions (Limoges et al., 2005). Work that relates sleep stage structure to subjective assessments of sleep quality include (Bonnet and Johnson, 1978) and (Keklund and Akerstedt, 1997). Such works use groupings in sleep structure that are explicitly guided by existing sleep measures, such as the Pittsburgh sleep quality index (Buysse et al., 1989), or measures extracted from the Karolinska Sleep Diary (Keklund and Akerstedt, 1997). The work described in the present paper is the only one of which we are aware that addresses a general description of intrinsic sleep types based on sleep architecture itself. Specifically, the present work uncovers a natural hierarchy of sleep types based on differences in measures of overall sleep stage composition such as sleep efficiency and time in slow wave sleep. Associations between sleep type and health-related indicators, including age and BMI, support the medical objectivity of these sleep types.

Plan of the Paper

Study data and methodology are described in section 2. Section 3 summarizes the results described in the prior work (Khasawneh et al., 2010), followed by the new results of the present paper involving EEG and ECG. Section 4 concludes with a summary of findings and ideas for future work.

2 METHODOLOGY

2.1 Data

We conduct our study on a dataset extracted from 244 recorded polysomnographic overnight sleep studies performed in the Sleep Disorder Center at Day Kimball Hospital in Putnam, CT, corresponding to 122 male and 122 female subjects. Summary statistics for patients associated with the dataset appear in Table 1.

2.1.1 Staging

Most of the sleep studies used for the present study were conducted prior to 2007, and hence the available staging information follows the Rechstchaffen and Kales (R&K) recommendations (Rechtschaffen and Kales, 1968), which were the standard at that time. In 2007, the American Academy of Sleep Medicine (AASM) revised their staging recommendations (Iber et al., 2007). The rationale for the new staging recommendations is discussed in (Silber et al., 2007). We briefly discuss the relationship between the two standards below.

The most immediately noticeable change in the AASM system as compared with the R&K system is the use of only three non–REM stages N1, N2, N3 instead of four, with the new N3 stage essentially replacing R&K NREM stages 3 and 4 corresponding to slow wave sleep (SWS). The AASM system also contains a revision to the rule for scoring stage N2, namely the elimination of the “3 minute rule” that allows continuation of NREM stage 2 labeling for up

Table 1: Summary statistics of the sleep dataset used in this paper.

	Age (years)	BMI ($kg\,m^{-2}$)	ESS (score)	BDI (score)	Mean SaO_2 (%)	Heart rate (bpm)
Male (n=122) $\mu\pm\sigma$	47.4±15.1	33.7±8.1	7.6±5.4	11.5±8.8	93.5±2.9	68.5±11.3
Female (n=122) $\mu\pm\sigma$	48.4±14.5	33.7±8.3	7.1±4.8	13.0±7.8	94.6±1.9	70.8±9.6
Overall (n=244) $\mu\pm\sigma$	47.9±14.8	33.7±8.2	7.4±5.1	12.2±8.3	94.1±2.5	69.7±10.5
min±max	20±85	19.2±64.6	0±23	0±48	70.2±97.9	46±99

to three minutes in the absence of the K–complex and sleep spindle EEG features that characterize stage 2. The AASM standard also includes clarified specifications for electrode placement. The variation in stage content between the AASM and R&K systems is studied in (Moser et al., 2009). Because of the new guidelines for scoring stage 2, AASM N2 content in a given sleep recording, as a fraction of total sleep time, is found to be 4.9% less than R&K NREM2 content, with the increase balanced by greater AASM N1 and N3 content: N1 content is 2.8% greater than NREM 1, and N3 content is 2.4% greater than NREM 3, 4 combined (Moser et al., 2009). No statistically significant differences are found in (Moser et al., 2009) between R&K and AASM stagings in total sleep time, sleep efficiency, or REM duration.

Given the relatively small differences between the AASM and R&K systems in terms of relative time spent in each stage (with combined time in R&K stages NREM 3, 4 corresponding to time in AASM stage N3), the results of this work should be similarly relevant to sleep studies staged by the AASM system.

2.1.2 Descriptive Variables

We summarize sleep composition in terms of seven summary measurements, as listed in Table 2. Total sleep time in minutes, fraction of time in bed spent sleeping, fraction of time in bed awake, and fraction of sleep period time in each of the sleep stages NREM $1,2$, in SWS (NREM$3+4$), and REM, are used. See Fig. 2 for an illustration. Only the variables described in Table 2 are used for clustering. A dataset is constructed in which each polysomnographic sleep study is summarized as a feature vector consisting of the values of the variables in Table 2 for that study. Clustering is performed over this dataset as described in section 2.2. Additional descriptive variables are used to study associations of other factors with sleep stage composition. These variables describe health history information such as age, Body Mass Index (BMI), and habitual smoking.

EEG-related variables are obtained by applying a short-time Fourier transform to extract spectral content from the C3-A2 EEG time signals on a 30-second epoch by epoch basis, at a sampling rate of 200 Hz.

Table 2: Summary descriptors of sleep composition.

Variable name	range of values
Total sleep time (TST)	$0-500$ min
Sleep efficiency (TST / Time-in-bed)	$0-1$
Percentage of sleep period time (SPT) in each of the stages NREM1, NREM2, SWS (NREM$3+4$), REM, wake	$0-100\,\%$

EEG spectra are then binned into δ $(0.5-4$ Hz), θ $(4-7$ Hz), α $(8-12$ Hz) and β $(12-30$ Hz) ranges. Summary descriptors of the EEG spectral variables are generated by taking the all-night mean and standard deviation of the collection of epoch-specific spectra for each spectral band, as well as by computing other measures such as spectral entropy and Hjörth activity, mobility, and complexity.

For the ECG-related variables, RR intervals, the time durations between consecutive R peaks in the QRS complexes of the ECG time signal, are extracted and likewise described in terms of all-night summary variables including the mean RRm, standard deviation SDRR, entropy RRentr, autocorrelation RRxcorr, and mean absolute linear predictability error LPCerror of the sequence of RR interval durations, Hjörth activity, mobility, and complexity measures, the standard deviation SDSD of the sequence of differences between successive RR intervals, the root mean squared difference RMSSD of successive RR intervals, the fraction pNNx of consecutive RR intervals that differ by more than x milliseconds (Mietus et al., 2002) for x = $10,20,30,40,50$, and the lengths SD1, SD2 of the principal axes of the Poincaré plot, which provides graphical information on parasympathetic nervous system activity and sympathovagal balance (Kamen et al., 1996). The Lomb-Scargle method (Moody, 1993) is used to compute spectral content of the RR interval sequence, and normalized RR spectral components within low frequency $(0.04-0.15$ Hz) and high frequency $(0.15-0.4$ Hz) bands are extracted. See (M Malik et al, 1996) for further information on the above and other commonly used measures of Heart Rate Variability (HRV).

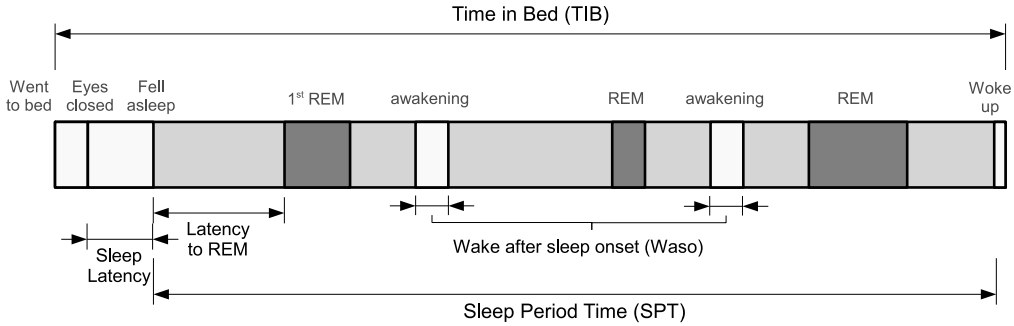

Figure 2: Summary-level descriptors of sleep composition. NREM sleep lightly shaded.

2.2 Clustering

Expectation-maximization (EM) (Dempster et al., 1977; Neal and Hinton, 1998) is a powerful iterative search technique for finding members of parameterized families of probabilistic models that locally maximize the likelihood of a given set of data. We use the EM clustering implementation in the Weka machine learning toolkit (Hall et al., 2009), version 3.7.1, which uses Gaussian mixture components as a basis for the models, and initializes the mixture parameters by k-means clustering. Unsupervised EM clustering is applied to the set of sleep composition feature vectors consisting of the values of the summary variables of Table 2 for the available staged PSG studies, yielding k clusters. The values $k = 3,4,5$ are considered in the present paper. Consistency of clustering results is evaluated by using two metrics, the cluster purity and the normalized mutual information (NMI) (Manning et al., 2008; Strehl, 2002). Each of these measures the agreement between two clusterings on a scale of 0 to 1, with higher values indicating better agreement.

2.3 Statistical Significance

Statistical significance is assessed by using a χ^2 test for independence in the case of nominal variables; for continuous variables, a t-test or Wilcoxon signed rank test are used for pairwise comparisons, and ANOVA or a Kruskal-Wallis test for multifactor comparisons. Control of the increased risk of false positives that is associated with simultaneous multiple tests of significance is accomplished by using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995). Given n individual findings with corresponding p-values $p_1 < p_2 < \cdots < p_n$, and given a desired overall level of significance (p-value) p, the Benjamini-Hochberg procedure declares as significant the first k findings, where k is the largest index i, $1 \leq i \leq n$, for which $p_i i/n < p$. The Benjamini-Hochberg method provides a rigorous bound on the false discovery rate (FDR), the fraction of predicted positives that are actually negatives. Indeed, the procedure as described here guarantees an overall FDR below the desired level p (Benjamini and Hochberg, 1995). The FDR is distinct from the traditional type I error rate, or familywise error rate (FWR), which is the probability of one or more false positives, regardless of the number of positive predictions. The FDR is generally considered to be a better choice of significance criterion than the FWR for exploratory data analysis tasks in which there are a large number of findings to evaluate.

3 RESULTS

We discuss the results of EM clustering of the sleep composition instance data as described in section 2, beginning with a summary of the prior work (Khasawneh et al., 2010), and show that the clusters found are stable, that they can be described by sleep efficiency and fraction of sleep time in slow wave sleep, and that the family of clusterings for $k = 3,4,5$ has a hierarchical structure. Health-related association results suggest that the clusters represent medically meaningful groups of distinct sleep behaviors. We then proceed to describe new results involving differences in EEG and ECG variables among the sleep type clusters. We find significant differences among clusters in EEG spectral band content, mean heart rate, Hjörth activity and mobility of both the EEG and ECG signals, and ECG normalized low frequency content, among others.

3.1 Salient Properties of Sleep Type Clusters

Kruskal-Wallis multiway comparison analysis of the clusters found by the EM algorithm for $k = 3,4,5$

clusters indicate significant differences among clusters in the means of all sleep composition variables used for clustering (Benjamini-Hochberg FDR $p < 0.01$, cf. section 2.3). Because of these differences in sleep composition, the clusters are referred to as *sleep types*. Wilcoxon pairwise test results confirm that the mean values of total sleep time, sleep efficiency, and fraction of sleep period time in NREM stages 2 and SWS, and in stages REM and wake, differ significantly (FDR $p < 0.01$) between all pairs of the three clusters in the case $k = 3$. However, the only variables found to be significantly different among all pairs of clusters for $k = 4,5$ are sleep efficiency and fraction of time in SWS.

3.1.1 Visualization of Sleep Types in Sleep Composition Space

Fig. 3 displays sleep types for three prespecified numbers of EM clusters: $k = 3$ (left), $k = 4$ (center), and $k = 5$ (right), in terms of sleep efficiency and fraction of sleep period time in SWS. Classification models were constructed to predict the cluster label based on sleep efficiency and SWS. Pruned decision trees achieved a classification accuracy of at least 0.91 for $k = 3,4,5$, confirming that these variables provide good separation among clusters in all cases. Cluster decision boundaries for $k = 3$ are vertical lines of constant sleep efficiency, while the cases $k = 4,5$ require the use of time in SWS also.

3.1.2 Hierarchical Structure of Sleep Types

Fig. 3 suggests that the family of clusterings for $k = 3,4,5$ has an approximately hierarchical structure: clusters 1 and 2 are relatively stable across the family, while the cluster labeled 3 in the leftmost image ($k = 3$) splits into the two clusters labeled 3 and 4 in the middle image ($k = 4$); in turn, the cluster labeled 4 in the middle image ($k = 4$) splits into the two clusters labeled 4 and 5 in the rightmost image ($k = 5$). Further evidence of the existence of this hierarchical structure is provided by a visualization of the decision boundaries between cluster regions that are found using Linear Discriminant Analysis (LDA). See Fig. 4. Classification accuracies are 90%, 86%, and 86% for $k = 3,4,5$, respectively. Despite some variation in the sizes and boundaries of the clusters for $k = 3,4,5$, support for the hierarchical structure of sleep types, including the stability of clusters 1 and 2 as conceptual entities across different values of k, derives from the summary statistics of the clusters (Khasawneh et al., 2010). Sleep efficiency and stage composition of clusters 1 and 2 are seen to remain nearly constant across the values $k = 3,4,5$. Cluster 2 has the lowest sleep efficiency across values of k, followed by cluster 1. The remaining clusters, which we loosely associate with subgroups of cluster 3 for $k = 3$, consistently have higher sleep efficiency than clusters 2 and 1. The transition from $k = 3$ to $k = 4$ produces an approximate subdivision of cluster 3 into a new cluster 3 that is SWS–heavy, and a cluster 4 that is stage NREM2–heavy. Stage composition of these two clusters is similar in other regards. Likewise, the transition from $k = 4$ to $k = 5$ generates a new cluster, 5, characterized by higher sleep efficiency and SWS content than cluster 4, but similar NREM2 content.

3.1.3 Relationships between Sleep Type and Health History Factors

For $k = 3$, Kruskal-Wallis multiway analysis reveals significant differences among clusters in mean patient age, collar size, BMI, smoking habit, and heart disease (Benjamini-Hochberg FDR $p < 0.05$). These results show that sleep types are meaningfully connected with overall health.

Pairwise differences in numerical variables are assessed using a Wilcoxon test. The cluster with highest sleep efficiency (cluster 3 on the left in Fig. 3) has significantly lower mean patient age (42.2 ± 12.7 vs 52.0 ± 14.3 and 57.0 ± 15.2), collar size (15.5 ± 2.1 vs 16.3 ± 2.1 and 16.5 ± 1.6), and habitual smoking than clusters 1 and 2, respectively. Body Mass Index (BMI) differs significantly between the two clusters at opposite ends of the sleep efficiency scale (2 and 3). Heart disease is significantly more frequent in cluster 2 than in the two clusters with higher sleep efficiency.

Pairwise comparison tests for $k = 4$ refine these results. The mean ages in cluster 3 (47.0 ± 14.7) and 4 (42.0 ± 12.0), the clusters into which cluster 3 for $k = 3$ approximately splits, are significantly lower than in cluster 2 (57.5 ± 15.2), the cluster with lowest sleep efficiency. Age is also significantly lower in cluster 4 than in the cluster with second lowest sleep efficiency, cluster 1 (age 52.1 ± 14.7). Collar size is significantly lower in cluster 4 (15.6 ± 2.1) than in cluster 2 (16.5 ± 1.6). Smoking is significantly less frequent in cluster 4 than in cluster 2. Heart disease is significantly more frequent in cluster 2 than in any other cluster.

3.2 EEG Characteristics of Sleep Type Clusters

We consider the variation of EEG signal summary descriptors among sleep type clusters.

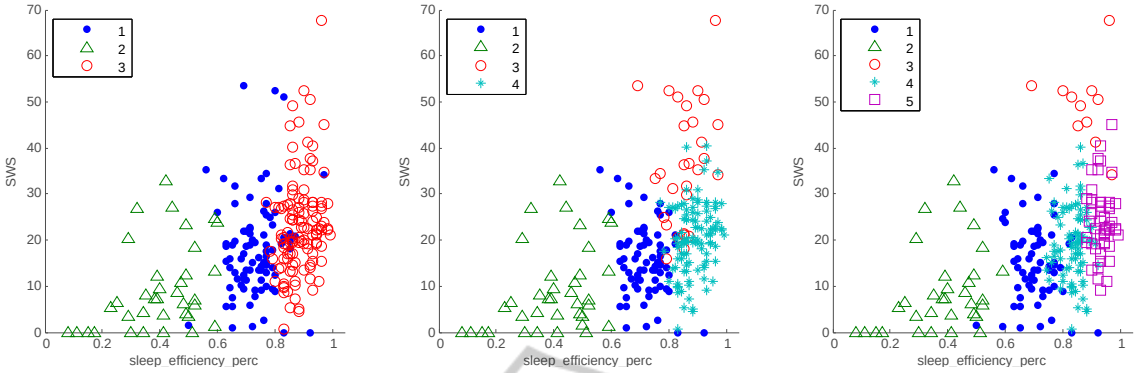

Figure 3: Sleep types in terms of sleep efficiency and fraction of SWS, $k = 3, 4, 5$.

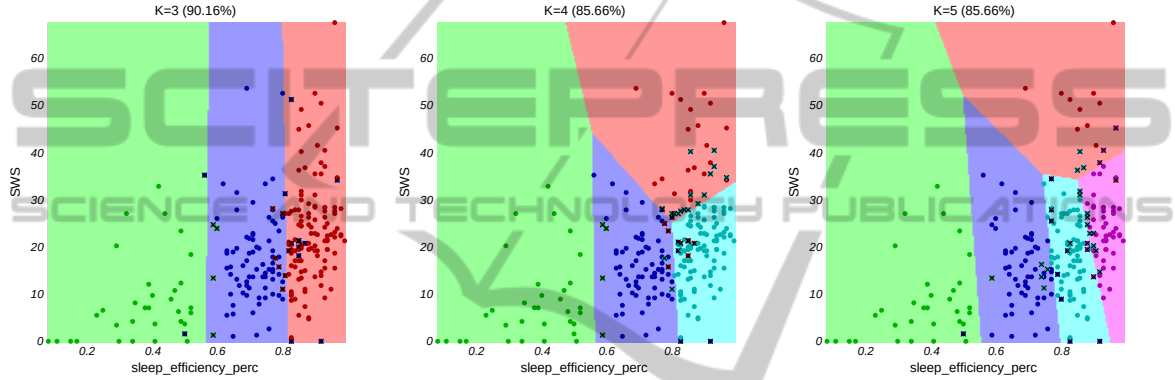

Figure 4: LDA decision boundaries in stage composition space. Classification accuracies along top.

3.2.1 EEG Summary Statistics

We begin with the case $k = 3$. ANOVA / Kruskal-Wallis multiway comparison finds significant differences among clusters in the variables listed in Table 3. Given the differences in stage composition among clusters as described in section 3.1, the corresponding differences in spectral content seen in table 3 are not entirely unexpected. For example, SWS content increases from cluster 2 to cluster 3 to cluster 1, which is the same cluster ordering obtained according to δ band (slow wave) spectral content. Likewise, α band spectral content is highest in cluster 2, which corresponds to the fact that cluster 2 has the lowest sleep efficiency, and hence the highest occurrence of wakefulness after sleep onset.

3.2.2 Sleep Type Classification based on EEG Alone

The MultiDimensional Scaling (MDS) visualization in Fig. 5 shows considerable separation among clusters in the space described by the EEG variables. This suggests that EEG may provide enough information to characterize the sleep type clusters. However, separation is not as marked as for the stage composition attributes used by EM to produce the clusters originally (see section 3.1).

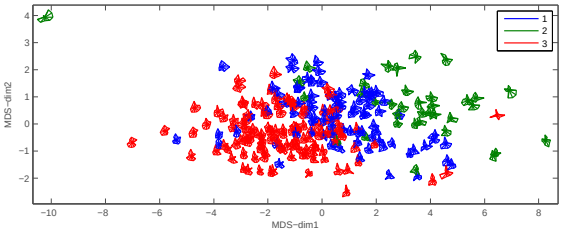

Figure 5: MDS visualization of sleep type clusters in EEG space, $k = 3$.

The result of Linear Decision Analysis (LDA) applied to the δ band power and spectral entropy attributes for $k = 3, 4, 5$ clusters exhibits the characteristic hierarchical structure discussed in section 3.1, as shown in Fig. 6. Classification accuracies for $k = 3, 4, 5$ are 62%, 50%, and 42%, respectively. Slightly higher LDA classification accuracies (75%, 59%, and 54%, respectively) are obtained for the at-

Table 3: EEG variables that differ significantly among sleep types, $k = 3$.

	Cluster 1	Cluster 2	Cluster 3	Kruskal-Wallis p-value
No. instances	87	36	121	to 3 digits:
mu_deltaRelPower	0.550±0.079	0.438±0.112	0.634±0.069	0.000*
mu_thetaRelPower	0.100±0.022	0.088±0.026	0.106±0.028	0.000*
mu_alphaRelPower	0.110±0.045	0.141±0.073	0.084±0.026	0.000*
mu_betaRelPower	0.115±0.041	0.135±0.042	0.092±0.036	0.000*
mu_spectralEntropy	0.568±0.049	0.611±0.081	0.516±0.048	0.000*
sd_deltaRelPower	0.211±0.032	0.226±0.037	0.184±0.027	0.000*
sd_alphaRelPower	0.081±0.037	0.091±0.042	0.064±0.025	0.000*
sd_totalPower	2.076E5±12.530E5	2.489E5±9.980E5	0.603E5±2.429E5	0.000*
sd_medianFrequency	6.777±0.422	7.215±0.414	6.286±0.445	0.000*
sd_spectralEntropy	0.136±0.025	0.146±0.026	0.125±0.021	0.000*
md_deltaRelPower	0.575±0.094	0.414±0.141	0.654±0.082	0.000*
md_alphaRelPower	0.089±0.037	0.125±0.077	0.071±0.023	0.000*
md_betaRelPower	0.105±0.044	0.134±0.045	0.080±0.040	0.000*
md_spectralEntropy	0.567±0.053	0.629±0.091	0.520±0.056	0.000*
mu_hjorthMobility	0.230±0.035	0.265±0.049	0.197±0.032	0.000*
mu_hjorthComplexity	2.524±0.384	2.407±0.762	2.714±0.353	0.000*
sd_hjorthActivity	5.315E5±34.771E5	5.816E5±25.234E5	1.434E5±6.339E5	0.000*
sd_hjorthMobility	0.084±0.019	0.090±0.023	0.073±0.014	0.000*
md_hjorthMobility	0.218±0.039	0.269±0.055	0.189±0.036	0.000*
md_hjorthComplexity	2.360±0.390	2.134±0.776	2.544±0.352	0.000*
md_thetaRelPower	0.097±0.023	0.084±0.028	0.102±0.030	0.004*
md_hjorthActivity	0.197E5±0.173E5	0.219E5±0.346E5	0.258E5±0.429E5	0.005*
sd_betaRelPower	0.071±0.021	0.073±0.021	0.064±0.017	0.010*
mu_totalPower	0.388E5±2.102E5	0.573E5±2.333E5	0.150E5±0.267E5	0.042*

tribute pair consisting of mean δ relative power and median Hjörth complexity (not shown). These results confirm that the EEG attributes allow relatively accurate sleep type labeling if full polysomnography is not available.

3.3 ECG and Heart Rate Variability (HRV)

We proceed to discuss the variation of ECG variables among sleep type clusters.

3.3.1 ECG Summary Statistics

In the case $k = 3$, ANOVA / Kruskal-Wallis multiway comparison finds significant differences among clusters in the variables listed in Table 4. The mean duration of RR intervals, that is, the time lapse between successive R peaks in the ECG signal (corresponding to overall heart rate), occupies the top position on this list. However, many of the remaining variables with the greatest significance, that is, with lowest p-values in Table 4, represent measures of the difference between consecutive RR intervals. Examples include SD2, SDRR, and S. In contrast, the values of variables such as SD1 that measure longer term variation in RR interval duration, do not differ significantly among sleep types. Consistent with this, the variables pNNx that measure the fraction of consecutive RR intervals that differ by more than x milliseconds are more significant the smaller the value of x, and the variable pNN50 with the largest value of x does not even meet the significance threshold $p < 0.05$.

The mean duration of RR intervals is found by a t test to be significantly higher in cluster 3 than in both cluster 1 and cluster 2. Recall that the latter two clusters are the ones with lower sleep efficiency. Note also that mean RR interval duration increases with sleep efficiency across the three clusters. Hjörth activity likewise is significantly higher in cluster 3 than in both cluster 1 and cluster 2. Overall heart rate variability as measured by the standard deviation SDRR of the sequence of RR intervals, is lowest in cluster 2, intermediate in cluster 1, and highest in cluster 3, that is, it increases with sleep efficiency (although the difference in SDRR between clusters 1 and 2 is not significant at the level $p < 0.05$). Also, RRxcorr, the autocorrelation of the RR interval sequence, is significantly lower in cluster 3 than in cluster 2, and linear prediction error LPCerror is correspondingly higher. Significance statements are at an overall Benjamini-Hochberg FDR level $p < 0.05$. Decreased heart rate variability is known to be associated with an increased risk of coronary heart disease and mortality from multiple causes (Dekker et al., 2000). Thus, our HRV findings are consistent with the distribution of heart disease among the clusters described in section 3.1.3. These results support the view that, for $k = 3$, the sleep type clustering is linearly ordered by overall health and sleep quality.

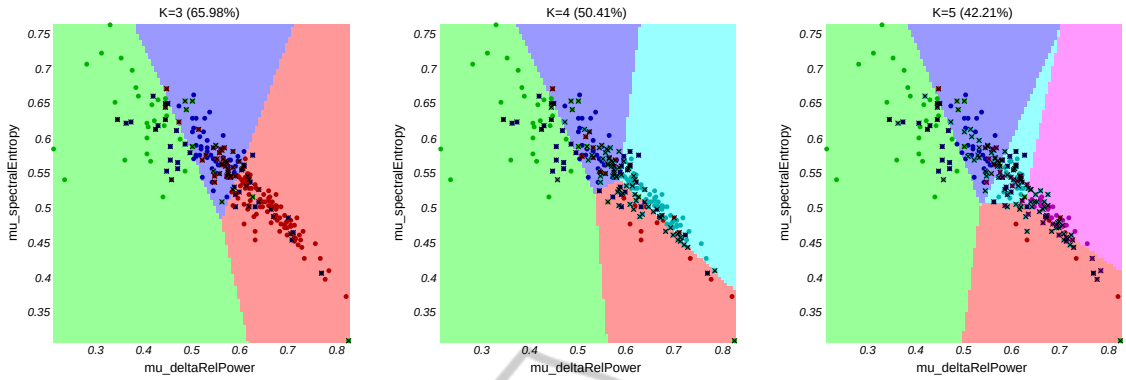

Figure 6: LDA decision boundaries in EEG space. Classification accuracies along top of figure.

Table 4: ECG variables that differ significantly among sleep types, $k = 3$.

	Cluster 1	Cluster 2	Cluster 3	Kruskal-Wallis p-value
No. instances	87	36	121	to 3 digits:
heartrate_mean	72.046 ± 9.749	74.722 ± 14.561	66.438 ± 8.507	0.000*
RRm	0.852 ± 0.123	0.840 ± 0.179	0.924 ± 0.125	0.000*
pNN10	70.565 ± 18.404	68.674 ± 17.588	77.529 ± 12.783	0.003*
SD2	0.101 ± 0.041	0.087 ± 0.039	0.111 ± 0.040	0.004*
SDRR	0.075 ± 0.031	0.065 ± 0.028	0.083 ± 0.030	0.004*
hjorthActivity	0.007 ± 0.006	0.005 ± 0.004	0.008 ± 0.006	0.004*
SDNNindex	0.053 ± 0.027	0.046 ± 0.022	0.058 ± 0.025	0.008*
S	0.009 ± 0.008	0.008 ± 0.007	0.012 ± 0.010	0.008*
SDANNindex	0.048 ± 0.022	0.043 ± 0.023	0.054 ± 0.021	0.009*
pNN20	47.104 ± 22.316	45.909 ± 22.383	55.495 ± 19.315	0.010*
nLF	33.031 ± 7.146	30.938 ± 7.788	34.781 ± 7.402	0.017*
pNN30	32.744 ± 21.494	32.761 ± 20.596	40.317 ± 20.122	0.018*
hjorthMobility	0.482 ± 0.190	0.590 ± 0.241	0.508 ± 0.161	0.022*
RRentr	5.080 ± 0.590	4.828 ± 0.688	5.169 ± 0.379	0.030*
LPCerror	0.025 ± 0.017	0.024 ± 0.014	0.028 ± 0.016	0.032*
FreqRatio	1.944 ± 1.308	1.514 ± 1.167	1.920 ± 1.141	0.034*
pNN40	23.671 ± 19.357	24.008 ± 17.906	29.595 ± 18.917	0.034*
RRxcorr	0.996 ± 0.005	0.997 ± 0.003	0.997 ± 0.003	0.046*

It is interesting that FreqRatio, the ratio between low and high frequency power, is significantly lower in cluster 2, which has the lowest sleep efficiency among the three clusters, than in the other two clusters. This is surprising at first sight, since cluster 2 has the greatest proportion of wakefulness after sleep onset, and hence one would expect increased sympathetic nervous system activity in this cluster than in the others, and hence a higher ratio of low to high frequency power. However, as found in (Vanoli et al., 1995), this expected behavior is reversed in patients that have suffered a myocardial infarction. Since cluster 2 contains the greatest incidence of heart disease among the clusters as discussed in section 3.1.3, this reversal is likely behind the lowered FreqRatio in cluster 2 here.

For $k = 4$, mean RR interval duration, standard deviation of the RR interval sequence, ratio of low to high frequency power, and ECG Hjörth activity are again significantly higher in the cluster with the highest sleep efficiency, cluster 4, than in the two clusters with the lowest sleep efficiency, 1 and 2.

The findings for $k = 3,4$ persist for $k = 5$, as mean RR interval duration, standard deviation of the sequence of RR intervals, ratio of low to high frequency power, and ECG Hjörth activity are significantly higher in cluster 5, which has the highest sleep efficiency, than in the three clusters with the lowest sleep efficiency, clusters $2,1,3$. Autocorrelation of the RR sequence is significantly lower in cluster 5 than in clusters $2,1,3$.

3.3.2 Sleep Type Classification based on ECG

An attempt to discriminate among the sleep type clusters in terms of ECG-related variables yields mixed results, with considerable overlaps between clusters (visualization not shown due to lack of space). For example, if cluster labels are assigned by LDA classification, using mean RR interval duration and Hjörth

mobility as predictive variables, the resulting classification accuracy is only 47% for $k = 3$ clusters, and decreases to 32% for $k = 5$. Equally significantly, while the hierarchical structure of the clustering family is reflected in the relationship between the cases $k = 3, 4$, the structure breaks down for $k = 5$. Indeed, as expected based on the discussion in section 3.1.2 and section 3.2.2, the LDA-predicted cluster that is most closely associated with intermediate sleep efficiency in the case $k = 3$ persists relatively unchanged for $k = 4$, while the cluster with the highest sleep efficiency splits into two clusters from $k = 3$ to $k = 4$. However, in the transition from $k = 4$ to $k = 5$, the LDA-predicted cluster with intermediate sleep efficiency changes substantially. Comparable results are obtained using other pairs of ECG variables for prediction. In view of these results, it is apparent that the existing ECG data alone is insufficient to fully characterize the family of sleep composition types.

Since sleep type classification relies on staged sleep recordings, a natural route toward ECG-based sleep type classification is to first construct a sleep stager that uses ECG signals alone, without the need for EEG or EMG. Classification of ECG time signals into the two classes sleep and wake has been addressed without additional data (Lewicke et al., 2005) and with the aid of respiratory signals (Karlen et al., 2009). However, we are not aware of techniques that successfully perform full sleep staging based only on ECG. The classification results based on ECG variables described in the preceding paragraph are therefore consistent with the current state of the art. It is an open problem whether alternative descriptions of ECG time series will allow accurate sleep type predictions from ECG data alone.

4 CONCLUSIONS AND FUTURE WORK

Unsupervised clustering of all-night human polysomnographic studies reveals a hierarchy of sleep composition types determined primarily by sleep efficiency, and subordinately by time in SWS. The present paper has described these sleep type clusters and their associations with health-related variables, and has compared the behavior of EEG and ECG signals among clusters. Statistically significant differences have been found among clusters in BMI, age, and heart disease incidence, supporting the medical objectivity of sleep types. Significant differences have also been found in δ, θ, α, and β EEG spectral band content, and it has been shown that relatively accurate prediction of sleep type is possible based on EEG alone. Analysis of the ECG signals has revealed significant differences among sleep types in RR interval duration, Hjörth activity and mobility, and in overall heart rate variability as measured for example by the standard deviation of the sequence of RR intervals. These findings are consistent with the health-related associations described above, the incidence of heart disease in particular. Despite the ECG findings, we have found that only limited information about sleep type may be extracted from ECG recordings alone based on the variables considered in the present paper. Future work should further explore the possibility of determining sleep type based on alternative descriptions of the ECG signals. Work in progress by the authors of the present paper involves modeling the dynamics of sleep stage transitions during sleep, and in particular studying the differences in dynamics among sleep type clusters.

REFERENCES

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57(1):289–300.

Bonnet, M. and Johnson, L. (1978). Relationship of arousal threshold to sleep stage distribution and subjective estimates of depth and quality of sleep. *Sleep*, 1(2):161–168.

Buysse, D., Reynoldsiii, C., Monk, T., Berman, S., and Kupfer, D. (1989). The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Research*, 28(2):193–213.

Dekker, J. M., Crow, R. S., Folsom, A. R., Hannan, P. J., Liao, D., Swenne, C. A., and Schouten, E. G. (2000). Low heart rate variability in a 2-minute rhythm strip predicts risk of coronary heart disease and mortality from several causes: The aric study. *Circulation*, 102(11):1239–1244.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The WEKA data mining software: an update. *SIGKDD Explor. Newsl.*, 11(1):10–18.

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*. American Academy of Sleep Medicine, Westchester, Illinois, USA.

Kamen, P., Krum, H., and Tonkin, A. (1996). Poincaré plot of heart rate variability allows quantitative display of

parasympathetic nervous activity in humans. *Clin Sci (Lond)*, 91(2):201–8.

Karlen, W., Mattiussi, C., and Floreano, D. (2009). Sleep and wake classification with ecg and respiratory effort signals. *IEEE Transactions on Biomedical Circuits and Systems*, 3(2):71–78.

Keklund, G. and Akerstedt, T. (1997). Objective components of individual differences in subjective sleep quality. *J Sleep Res.*, 6(4):217–220.

Khasawneh, A., Alvarez, S. A., Ruiz, C., Misra, S., and Moonis, M. (2010). Discovery of sleep composition types using expectation-maximization. In *Proc. 23rd IEEE International Symposium on Computer-Based Medical Systems (CBMS 2010)*, Perth, Australia.

Kryger, M., Roth, T., and Dement, W. (2005). *Principles and Practice of Sleep Medicine*. Elsevier Saunders, Philadelphia, PA, USA, 4th edition.

Lewicke, A., Sazonov, E., Corwin, M., and Schuckers, S. (2005). Reliable determination of sleep versus wake from heart rate variability using neural networks. In *Neural Networks, 2005. IJCNN '05. Proceedings. 2005 IEEE International Joint Conference on*, volume 4.

Limoges, E., Mottron, L., Bolduc, C., Berthiaume, C., and Godbout, R. (2005). Atypical sleep architecture and the autism phenotype. *Brain*, 128(5):1049–1061.

Loomis, A., Harvey, E., and Hobart, G. (1937). Cerebral states during sleep, as studied by human brain potentials. *Journal of Experimental Psychology*, 21(2):127–144.

M Malik et al (1996). Heart rate variability: standards of measurement, physiological interpretation and clinical use. task force of the european society of cardiology and the north american society of pacing and electrophysiology. *Circulation*, 93(5):1043–1065.

Manning, C. D., Raghavan, P., and Schutze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press. Web publication at informationretrieval.org.

Mietus, J., Peng, C.-K., I, I. H., Goldsmith, R., and Goldberger, A. (2002). The pNNx files: re-examining a widely used heart rate variability measure. *Heart*, 88(4):378–380.

Moody, G. (1993). Spectral analysis of heart rate without resampling. *Computers in Cardiology*, 20:715–718.

Moser, D., Anderer, P., Gruber, G., Parapatics, S., Loretz, E., Boeck, M., Kloesch, G., Heller, E., Schmidt, A., Danker-Hopfe, H., Saletu, B., Zeitlhofer, J., and Dorffner, G. (2009). Sleep classification according to AASM and Rechtschaffen & Kales: Effects on sleep scoring parameters. *Sleep*, 32(2):139–149.

Neal, R. and Hinton, G. E. (1998). A view of the EM algorithm that justifies incremental, sparse, and other variants. In *Learning in Graphical Models*, pages 355–368. Kluwer Academic Publishers.

Propper, R., Christman, S., and Olejarz, S. (2007). Home-recorded sleep architecture as a function of handedness II: Consistent right- versus consistent left-handers. *J Nerv Ment Dis.*, 195(8):689–692.

Rao, M., Blackwell, T., Redline, S., Stefanick, M., Ancoli-Israel, S., and Stone, K. (2009). Association between sleep architecture and measures of body composition. *Sleep*, 32(4):483–90.

Rechtschaffen, A. and Kales, A., editors (1968). *A Manual of Standardized Terminology, Techniques, and Scoring System for Sleep Stages of Human Subjects*. US Department of Health, Education, and Welfare Public Health Service – NIH/NIND.

Silber, M., Anconi-Israel, S., Bonnet, M., Chokroverty, S., Grigg-Damberger, M., Hirshkowitz, M., Kapen, S., Keenan, S., Kryger, M., Penzel, T., Pressman, M., and Iber, C. (2007). The visual scoring of sleep in adults. *Journal of Clinical Sleep Medicine*, 3(2):121–131.

Smith, S., Dingwall, K., Jorgensen, G., and Douglas, J. (2006). Associations between the use of common medications and sleep architecture in patients with untreated obstructive sleep apnea. *Journal of Clinical Sleep Medicine*, 2(2):156–162.

Strehl, A. (2002). *Relationship-based Clustering and Cluster Ensembles for High-dimensional Data Mining*. PhD thesis, The University of Texas at Austin.

Sulekha, S., Thennarasu, K., Vedamurthachar, A., Raju, T., and Kutty, B. (2006). Evaluation of sleep architecture in practitioners of Sudarshan Kriya yoga and Vipassana meditation. *Sleep and Biological Rhythms*, 4(3):207–214.

Vanoli, E., Adamson, P. B., Ba-Lin, Pinna, G. D., Lazzara, R., and Orr, W. C. (1995). Heart rate variability during specific sleep stages: A comparison of healthy subjects with patients after myocardial infarction. *Circulation*, 91(7):1918–1922.

Zhang, L., Samet, J., Caffo, B., and Punjabi, N. (2006). Cigarette smoking and nocturnal sleep architecture. *Am J Epidemiol.*, 164(6):529–537.