Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Sleep State Scoring in Infants from Respiratory and Activity Measurements

Nadezhda A. Sazonova,, *Student member, IEEE,* Edward E. Sazonov, *Member, IEEE,*
Bozhao Tan, Stephanie A. C. Schuckers, *Member, IEEE,* and CHIME study group

***Abstract*—Sleep state scoring usually relies on polysomnographic measurements, which include Electroencephalogram (EEG), Electromyogram (EMG), Electrooculogram (EOG), two or three lead chest Electrocardiogram (ECG), and may include other measurements. Overall, polysomnography is an intrusive procedure not well tolerated by infants and elderly. The goal of this research is to study possibility of automatic sleep state scoring from less intrusive measurements such as activity measurements and respiratory measurements by inductive plethysmography. The study is based on the Collaborative Home Infant Monitoring Evaluation (CHIME) dataset. Results demonstrate that the suggested approach is capable of scoring sleep states (awake, rapid eye movement and quiet sleep) with good accuracy.**

I. Introduction

Sleep state scoring is widely used for diagnostic of sleep disorders and in physiological studies since many physiological variables have properties dependent on sleep state. For example, sleep state was an important predictor in a study of apparent life-threatening events in infants [1]. A typical sleep scoring procedure is based on polysomnographic (PSG) measurements [2] that include Electroencephalogram (EEG), Electromyogram (EMG), Electrooculogram (EOG), two or three lead chest Electrocardiogram (ECG), and may include other measurements. Overall, the high intrusiveness of the polysomnographic measurements creates problems with applications in sensitive groups such as infants and nursing home patients. Sleep/wake state scoring via activity measurements (actigraphy) is procedure with low intrusiveness that has been used in a variety of studies, including studies on infants [3, 4], adults [5], demented nursing home patients [6]. Sleep/wake identifications made in adults by using actigraphy have shown 85-95% agreement rates between actigraphy and polysomnography [7]. In infants, agreement rates varied from 54% to 87% at different ages [4].

The authors have investigated actigraphic sleep/wake detection in infants from Collaborative Home Infant Monitoring Evaluation (CHIME) database and received results comparable to traditional polysomnographic scoring [8]. One of the major differences from traditional actigraphy was in placement of the accelerometer on infant's diaper (The sensor was originally designed to detect infant's position rather than measure activity). The major drawback of the sleep/wake classification in is its inability to differentiate between quiet (NREM) and Rapid Eye Movement (REM) sleep states.

CHIME dataset also includes respiratory recordings performed by inductive plethysmography on RespiTrace PT home monitor [9]. Previous studies have shown that respiratory variability is a good predictor of sleep state (REM/NREM) if the sleep/wake state is known [10]. Therefore, it seems natural to attempt sleep scoring on a combined data set including activity measurements as a primary predictor for sleep/wake classification and respiratory measurements for REM/NREM classification. This paper presents some preliminary results of this study.

II. Methodology

A. Data

Data used in this paper represent a part of the CHIME dataset. Each infant in the CHIME study also had a standard in-hospital 8-hour PSG with EEG, EMG and EEOG performed on Healthdyne ALICE3 system versions 1.17 and 1.19 [9]. Sleep state scoring identified four primary sleep states: awake (AWK), REM sleep, quiet sleep (NREM) and indeterminate sleep (IND). Sleep state scoring on 30-second intervals (epochs) was conducted by trained technicians and used in this study as a "gold standard" in training and evaluating performance of the automatic scoring algorithms. Additionally, a home monitor (NIMS RespiTrace PT) recorded electrocardiogram, respiratory inductance plethysmography (RIP), pulse oximetry, and an accelerometer to detect motion and infant position.

Data from the home monitor (accelerometer signal, tidal volume and breath marker) were extracted, preprocessed and used as predictors for sleep state scoring. Accelerometer signal (Acc) was sampled at 50Hz as an 8-bit value; tidal

Manuscript received April 24, 2006.

N.A. Sazonova is with the Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV, 26506 USA (phone: 315-265-4532; e-mail: nsazonov@stat.wvu.edu).

E.S. Sazonov is with the Department of Electrical and Computer Engineering, Clarkson University, (email: esazonov@clarkson.edu).

S.A.C. Schuckers is with the Department of Electrical and Computer Engineering, Clarkson University (email: sschucke@clarkson.edu).

B. Tan is with the Department of Electrical and Computer Engineering, Clarkson University (email: tanb@clarkson.edu).

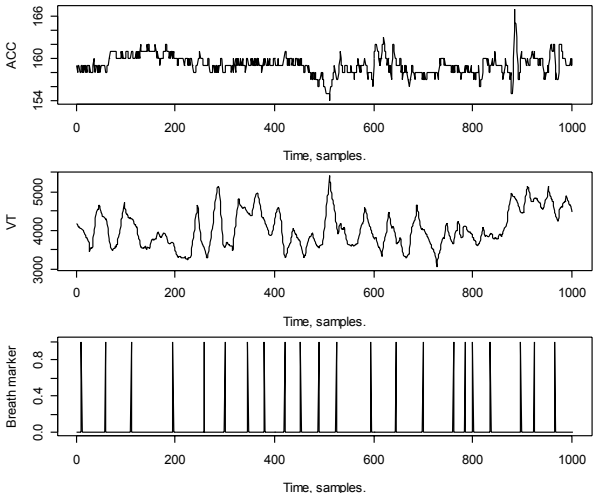

Fig. 1. Raw recordings of acceleration, tidal volume and breath marker signals.

volume (Vt) was sampled at 50 Hz and 8 bits; breath marker (Bm) was computed from the respiratory signal to indicate breath duration. Fig. 1 shows unprocessed Acc, Vt, and Bm signals extracted from the monitor recordings. Preprocessing of the signals was performed in several steps:

Accelerometer signal:

1. The signal was aligned with the PSG sleep score on the 30-second epoch boundary.
2. An artifact (a momentary transition of the sensor signal to zero every 5021 samples) was removed from the signal.
3. DC levels of the characterizing infant's position in the crib were removed by demeaning of every epoch.
4. Maximum absolute value and standard deviation were computed for every epoch.

Tidal volume:

1. The signal was aligned with the PSG sleep score on the 30-second epoch boundary.
2. Maximum absolute value after demeaning, mean, standard deviation and coefficient of variation were computed for every epoch.

Breath marker:

1. The signal was aligned with the PSG sleep score on the 30-second epoch boundary.
2. Mean and standard deviation of respiratory cycle time Ttot, and mean respiratory frequency were computed from the signal.

Fig. 2 shows some of the extracted features vs. sleep state.

B. Classification with Support Vector Machines

The predictor metrics (mean, standard deviation, etc.) extracted from the Acc, Vt and Bm signals were used to produce feature sets for classification.

The value of a given metric, taken for the last n consecutive periods produced a set of features F_{-n} , … , F_{-2} , F_{-1} , F_0 associated with a given response variable. Lagged data are commonly utilized in sleep research due to the fact that sleep state transitions usually are not sudden. For this

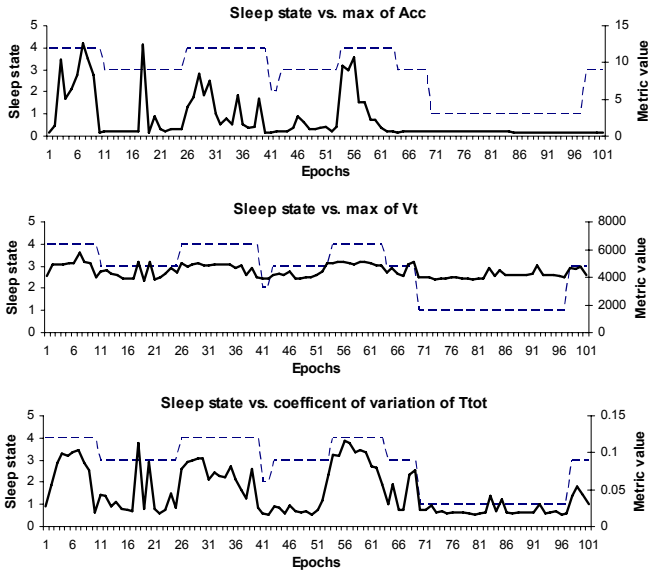

Fig. 2. Three different predictor metrics plotted vs. sleep state. Solid line shows the values of respective metric. Dashed line represents sleep state (4-awake, 3-REM, 2-Indeterminate, 1-NREM).

study, the value of n was chosen to be 4.

The response variable was represented by SLEEP/WAKE categories for sleep/wake detection (SLEEP represents a union of NREM and REM categories) or NREM/REM categories for sleep state identification. Indeterminate sleep state was not included into analysis. Sleep labels in the PSG score for each infant (A,Q,I,W) that were converted to numerical labels.

Classification was performed in R statistical package with Support Vector Machines (package e1071). The model was built as a C-classifier using radial basis kernels $K(u,v) = e^{-\gamma|u-v|^2}$.

Two types of classification models were built on the same set of feature variables: a model for prediction of the Sleep and Wake states (SW), and a model for prediction of Active and Quiet states within the Sleep state (AQ).

A classification model was built for every feature extracted from the raw signals: 4 models for Acc signal (SW and AQ models based on epoch's maximum and on epoch's standard deviation); 6 models for Vt signal (SW and AQ models each based on one of the three extracted features); and 6 models for Bm.

The study involved 26 subjects. Initially all 16 possible classification models were built for each of the 26 subjects comprising the *individual models*. The training set for each individual model was taken as randomly selected 50% of observations with a requirement that the two classes (Sleep and Wake; Active and Quiet) were equally represented. The rest 50% of observations were left as a validation set.

The individual models allowed estimation of the quality of different metrics as predictors. The practical significance, though, lies in building the universal model that does not

TABLE I
Classification performance of the Sleep/Wake models

Predictor	Combined models		Individual models	
	Agreement rate (training sample)	Validation rate (validation sample)	Average agreement rate	Average validation rate
Acc:				
max *	**0.7947**	**0.7405**	**0.7178**	**0.8150**
std. deviation.	0.7814	0.7271	0.6879	0.8023
Vt:				
max	0.7779	0.7214	0.7511	0.7511
mean	0.7072	0.6551	0.5919	0.5391
coeff. var. *	**0.7994**	**0.7477**	**0.7589**	**0.7789**
Bm:				
mean Ttot	0.7031	0.6795	0.7140	0.6989
resp. frequency *	0.7031	0.6795	**0.7183**	**0.7234**
coeff. var. Ttot	0.7031	0.6795	0.6914	0.6569

Asterisk * indicates the best predictor in the group

TABLE II
Classification performance of the Active/Quiet models

Predictor	Combined models		Individual models	
	Agreement rate (training sample)	Validation rate (validation sample)	Average agreement rate	Average validation rate
Acc				
max *	**0.6224**	**0.5264**	**0.6109**	**0.5987**
std. deviation	0.5980	0.5008	0.5963	0.5829
Vt				
max *	**0.6906**	**0.6503**	**0.6912**	0.6372
mean	0.5460	0.5168	0.5196	0.5196
coeff. var.	0.5460	0.6471	0.6783	**0.6680**
Bm:				
mean Ttot	0.7115	0.6140	0.6176	0.6176
resp. frequency	0.7038	0.6181	0.6613	0.6462
coeff. variation Ttot *	**0.7369**	**0.7008**	**0.7328**	**0.7191**

Asterisk * indicates the best predictor in the group

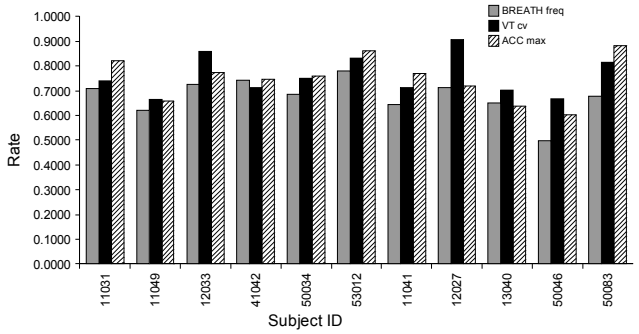

Fig. 3. Validation rate for the best predictors from the combined SW models.

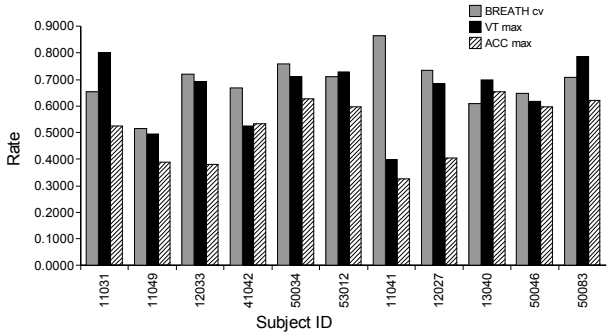

Fig. 4. Validation rate of the best predictors from the combined AQ models.

involve individual effects. This indicated the need for a model based on the pooled data from several subjects, a *combined model*.

The combined models were produced by methods identical to those used for individual models, but the training data were pooled from 13 of the subjects. Remaining 13 subjects comprised the validation set.

The choice of SVM parameters (γ and misclassification cost C) was performed on a single subject by a grid search on validation set. Since for every set of predictors the best parameters essentially coincided, these values (e^{-3} for γ and e^{-1} for cost) were then uniformly applied to all the models for all subjects.

III. Results

A. SLEEP/WAKE classification models

Individual and combined SW models were built for each of the lagged predictor metrics (n=4). Results from applying the SVM classification are given in Table I. The table shows the best predictor metric within each group of individual as well as combined models. Classification rates for individual models in Table I are averaged across the subject group.

The coefficient of variation for Vt and maximum of the Acc have shown the best and approximately equal performance both individual and combined SW models. Both of these predictors outperformed metrics derived from Bm (Fig. 3). Average rate from the combined model for the validation group of subjects based on VT's coefficient of variation was 74.8% and from the model based on ACC max was 74% (compared with respiratory frequency producing average accuracy of 68%).

B. ACTIVE/QUIET sleep classification models

Results of the active/quiet sleep states identification are represented in Table II. Agreement and validation rates are very close between individual and combined models, indicating good representation of the data by classifiers. According to the values of the agreement and validation rates the best predictor for the AQ models was unambiguously coefficient of variation for Ttot (Fig. 4).

IV. Discussion

Results of this study produce a few important observations.

First, the group of infants in this study was essentially the same as the group of infants in [8]. This gives an opportunity to compare results of sleep/wake classification on Acc signal

by support vector machines vs. previously used Learning Vector Quantization (LVQ) networks and logistic regression. Average validation rate for individual SW models reported in [8] for LVQ networks was 75.3% while the same rate for logistic regression was 75.7%. The validation rate achieved by SVM on maximum of Acc predictor in this study was 81%, which shows some improvement over earlier results and potential advantage of using SVM as a classification model.

Second, results of SW classification show significant differences in performance depending on the used predictor variable. Maximum of Acc signal and coefficient of variation for Vt signal produced comparable results in sleep/wake classification. At the same time, none of the predictor metrics derived from Bm signal performed well. It should be noted that all three of the combined models built for Bm metrics have unexpectedly produced identical results, which upon further review was caused by classification of the whole data set as SLEEP. These results show that predictors related to respiratory frequency do not perform well in SW classification, confirming some of the finding reported in [10].

Third, results of AQ classification indicate a completely opposite picture. The coefficient of variation for Ttot shows the best results, while predictors derived from Acc and Vt signals lag in performance. Validation rates for models built with Acc derived metrics were in region of 50%, indicating that Acc is not a meaningful predictor for AQ classification.

Fourth, automatic scoring using these methods which are based on limited physiologic signals (activity, respiration) achieve comparable results to traditional sleep scoring using a full PSG [9].

Last, the AQ model was trained and tested using only the active/quiet data. In a two stage classification, that is, first stage (sleep/wake) and the second stage (active/quiet), error will compound from the first to the second stage, i.e. mistakes where sleep is called awake would be automatic errors for AQ. Ultimately, a model which combines the best features for SW/AQ needs to be developed.

V. Conclusion

Results of this study show that prediction metrics derived from respiratory signals and activity measurements show different trends in prediction accuracy of SLEEP/WAKE and QUIET/ACTIVE sleep states. Predictors derived from the accelerometer signal show the best performance in SW models but are insignificant in AQ models. Predictors derived from the tidal volume measurement have performance comparable to Acc in SW models and perform better in AQ models. Predictors derived from respiratory cycle time (respiratory variability) show the best performance in AQ models but perform inadequately in SW models.

Overall, these results suggest that an automatic sleep scoring methodology can be developed by combining the predictors derived from activity and respiratory measurements. Future work will consider the best-performing features in development of a combined model of sleep stage identification (AWK,NREM,REM,IND). Such a methodology can potentially reduce invasiveness of the sleep scoring equipment.

References

[1] X. Xueyan, "*Prediction of life-threatening events in infants using heart rate variability measurements,*" Ph.D. dissertation, West Virginia University, 2002.

[2] T. Anders, R. Emde, A.H. Parmelee, "*A Manual of Standardized Terminology, Techniques, and Criteria for Scoring of States of Sleep and Wakefulness in Newborn Infants*". Los Angeles:UCLA Brain Information Services, NINDS Neurological Information Network, 1971.

[3] A. Sadeh, "*Evaluating Night Wakings in Sleep-Disturbed Infants: A Methodological Study of Parental Reports and Actigraphy*". Sleep, v.19, p.757-762, 1996.

[4] A. Sadeh, C. Acebo, R. Seifer, S. Aytur, and M.A. Carskadon, "*Activity-Based Assessment of Sleep-Wake Patterns During the 1st Year of Life*". Infant Behavior and Development, v.18, p.329-337., 1995.

[5] A. Sadeh, K. Sharkey, and M. Carskadon, "*Activity-Based Sleep-Wake Identification: An Empirical Test of Methodological Issues*". Sleep, v.17, p.201-207., 1994.

[6] I.S. Amcoli, P. Clopton, M.R.. Klauber, R. Fell, and W. Mason, "*Use of Wrist Activity for Monitoring Sleep/Wake in Demented Nursing-Home Patients.*" Sleep, v.20, p.24-27, 1997.

[7] R. J. Cole, D.F. Kripke, W. Gruen, D.J. Mullaney, and J.C. Gillin, "*Automatic Sleep/Wake Identification From Wrist Actigraphy.*" Sleep, v.15, p.461-469, 1992.

[8] E.S.Sazonov, N.A. Sazonova, S. Schuckers, M. Neuman and CHIME Study Group "*Activity-based sleep–wake identification in infants*" 2004 Physiol. Meas. 25 1291-1304.

[9] D.H. Crowell, L.J. Brooks, T.Colton, M.J. Corwin, T.T. Hoppenbrouwers, C.E. Hunt, L.E. Kapuniai, G. Lister, M.R. Neuman, M. Peucker, S.L. Ward, D.E. Weese-Mayer and M. Willenger "*Infant polysomnography: reliability.*" Collaborative Home Infant Monitoring Evaluation (CHIME) Steering Committee. Sleep, v.20, p.553-560. 1997.

[10] G.G. Haddad GG, H.J. Jeng, T.L. Lai, R.B. Mellins "*Determination of sleep state in infants using respiratory variability."* Pediatr Res. 1987 Jun;21(6):556-62.