Activity-based sleep–wake identification in infants

2004 Physiol. Meas. 25 1291

(http://iopscience.iop.org/0967-3334/25/5/018)

INSTITUTE OF PHYSICS PUBLISHING PHYSIOLOGICAL MEASUREMENT

Physiol. Meas. **25** (2004) 1291–1304 PII: S0967-3334(04)78551-2

Activity-based sleep–wake identification in infants

Edward Sazonov[1], Nadezhda Sazonova[2], Stephanie Schuckers[1], Michael Neuman[3] and CHIME Study Group

[1] Department of Electrical and Computer Engineering, Clarkson University, PO Box 5720, Potsdam, NY 13699, USA
[2] Department of Statistics, West Virginia University, PO Box 6330, Morgantown, WV 26506, USA
[3] Department of Biomedical Engineering, Michigan Technological University, 310 Minerals and Materials Building, 1400 Townsend Drive, Houghton, MI 49931, USA

E-mail: esazonov@ieee.org, nsazonov@stat.wvu.edu and mneuman@mtu.edu

Received 29 March 2004, accepted for publication 22 June 2004
Published 11 August 2004
Online at stacks.iop.org/PM/25/1291
doi:10.1088/0967-3334/25/5/018

Abstract

Actigraphy offers one of the best-known alternatives to polysomnography for sleep–wake identification. The advantages of actigraphy include high accuracy, simplicity of use and low intrusiveness. These features allow the use of actigraphy for determining sleep–wake states in such highly sensitive groups as infants. This study utilizes a motion sensor (accelerometer) for a dual purpose: to determine an infant's position in the crib and to identify sleep–wake states. The accelerometer was positioned over the sacral region on the infant's diaper, unlike commonly used attachment to an ankle. Opposed to broadly used discriminant analysis, this study utilized logistic regression and neural networks as predictors. The accuracy of predicted sleep–wake states was established in comparison to the sleep–wake states recorded by technicians in a polysomnograph study. Both statistical and neural predictors of this study provide an accuracy of approximately 77–92% which is comparable to similar studies achieving prediction rates of 85–95%, thus validating the suggested methodology. The results support the use of body motion as a simple and reliable method for determining sleep–wake states in infants. Nonlinear mapping capabilities of the neural network benefit the accuracy of sleep–wake state identification. Utilization of the accelerometer for the dual purpose allows us to minimize intrusiveness of home infant monitors.

Keywords: sleep–wake identification, actigraphy, polysomnography, logistic regression, neural networks

1. Introduction

Identification of sleep/wake states is used in several areas of medical science. For infants, sleep state is an important variable in analyzing heart rate variability and other measures in the study of life-threatening events such as bradycardia and apnea (Xueyan 2002). Polysomnography (PSG), which includes an electroencephalogram (EEG), electrooculogram (EOG) and electromyogram (EMG), is considered to be the most accurate procedure in determining sleep states. The largest shortcoming of PSG is that it is rather expensive and too complex to be used by an untrained person. Relatively high intrusiveness of the PSG method is also the cause of its low tolerance by nursing-home patients and infants. An appealing alternative is presented by the actigraphic methods. Actigraphy does not require complex equipment that has to be serviced by a trained technician and is perfectly suited to be used in home conditions. Another advantage of actigraphy is its low intrusiveness on the patient. An actigraph is a wireless portable device usually worn on a wrist or an ankle. It includes a motion sensor (an accelerometer), a microprocessor with analog/digital circuitry and a memory chip. The motion patterns are recorded throughout the day and analyzed for the information of interest. Usually, actigraphy does not aim at identifying sleep states, such as active or quiet sleep, but attempts to determine sleep–wake patterns. It provides sleep detection results comparable to those of polysomnography and behavioral response monitoring (Blood *et al* 1997, Sadeh and Acebo 2002) when applied to different population groups like adults (Sadeh *et al* 1994), demented nursing-home patients (Ancoli *et al* 1997), young children and infants (Sadeh *et al* 1991, 1995, Acebo *et al* 1999), etc. Actigraphy has been successfully used in studies of insomnia (Wicklow and Espie 2000, Shinkoda *et al* 1998, Sadeh *et al* 2000), attention deficit disorder in children (Gruber *et al* 2000), sleep disturbances in children (Sadeh 1996, Tikotzky and Sadeh 2001, Sadeh *et al* 2002) and mental development of preterm infants (Gertner *et al* 2002) and retarded children (Leitner *et al* 2002). Such a wide spectrum of subjects can be covered due to actigraphy's non-invasiveness. Sleep/wake identifications made in adults by using actigraphy have shown 85–95% agreement rates between actigraphy and polysomnography (Cole *et al* 1992). In infants, agreement rates varied from 54% to 87% at different ages (Sadeh *et al* 1995). A known limitation of actigraphy is due to the fact that an actigraphic recording provides less information on the underlying brain activity than a PSG, and is based on a single source (accelerometer) contrary to three distinct types of data (EEG, EOG, EMG) of a PSG (Ancoli *et al* 2003).

This study's aim to validate the use of actigraphy when an accelerometer normally used to determine infant's position is also used to provide motion data for actigraphic analysis. The described method also features a different position of the accelerometer on the subjects (diaper instead of an ankle). The accuracy of the sleep/wake state prediction by a statistical method (logistic regression) is compared to that of a computational intelligence method (supervised neural network). The accelerometer was used as a part of a major Collaborative Home Monitoring Evaluation (CHIME) study which studied home infant monitors for apnea and bradycardia for over 1000 infants. Sleep/wake state information would be helpful in analyzing approximately 100 000 min of cardiorespiratory data recorded in home conditions with no available PSG scoring.

2. Methods

Data used in this study were collected on a group of infants as a part of the CHIME study. Each infant had a standard, monitored 8 h PSG performed with EEG, EOG and EMG performed on a Healthdyne Alice PSG monitor followed by calculation of sleep states (Crowell *et al* 1997).

Additionally, electrocardiogram, respiratory volume measured by respiratory inductance plethysmography, pulse oximetry and accelerometer measurements were recorded on a CHIME monitor as used in the home study (Neuman *et al* 2001). Sleep state was classified for every 30 s into four sleep–wake states: quiet sleep (QS), active sleep (AS), awake (AWK) and indeterminate states. This sleep state classification was performed by technicians in the CHIME study group from the EEG, EOG, respiratory regularity and body movements of the infants based on an algorithm by Anders *et al* (1971) and modified by the CHIME group (Crowell *et al* 1997). Intra- and inter-rater reliability was assessed using k statistics which assessed agreement between observers, corrected by chance, with relative strength of agreement considered moderate for $k = 0.41$–0.60, substantial for $k = 0.61$–0.80 and almost perfect for $k = 0.81$–1.00. A k-value of 0.58 was achieved for uniformly trained technicians and with some refinements was increased to 0.68 on a second dataset (Crowell *et al* 1997). For the CHIME data, raw sleep parameters, technician-determined sleep score and smoothed sleep scores using a 30 s epoch length (Kulp *et al* 2000) are available.

Although the CHIME dataset did not include specialized actigraphic measurements, an accelerometer (ACC) attached to the diaper was utilized to determine the infant's position in the crib. The utilized accelerometer was a single axis silicon accelerometer chip molded in a $2.5 \times 2.1 \times 1.0$ cm. PVC package with the sensitive axis normal to the 2.5×2.1 cm surface. A blue dot on the package indicated the surface in the direction of the positive acceleration. The accelerometer was taped to the upper lip of the subject's diaper with the blue dot facing toward the infant's skin. The placement of the accelerometer on the diaper allowed recording of the infant's position (such as prone, lateral or supine) by registering DC levels resulting from the interaction of the sensor with the gravitational field. Such a placement of the sensor is significantly different from traditionally used placement on an extremity and may cause lesser sensitivity to motion (motion of the body is less pronounced than motion of the limbs) and lower noise tolerance (the body has motion artifacts originating from breathing, heart beat and gastro-intestinal processes). At the same time, the accelerometer may pick up enough motion to identify sleep/wake states and eliminate the need for another sensor on a limb. Utilization of an accelerometer for the dual purpose of position and activity monitoring would reduce the intrusiveness of the home monitoring device.

The infant population in this study was represented as a random sample of about 100 subjects provided by the CHIME group at that time; out of them 26 subjects had satisfactory (non-flat) accelerometer and available PSG records. The postconceptional age for the latter group varied within 34 to 46 weeks. According to the maternal and infant characteristics (Ramanathan *et al* 2001), 5 infants belonged to the group of healthy full-term infants (mean gestational age 39.5 week, mean birth weight 3311 g), 13 to the preterm asymptomatic group (29.7 week, 1252 g), 2 to the group of full-term SIDS siblings (39.7 week, 3520 g), 1 to the group of preterm SIDS siblings (35 week, 2532 g) and 5 to the group of full-term symptomatic group (39.6 week, 3367 g). The random sample containing 26 subjects was taken regardless of the distribution over these groups as group information was not available at that time. Each infant had a monitored PSG for a time duration ranging from 6 h 57 min to 8 h 19 min, with the average length of a recording 7 h 54 min.

The raw signal from the accelerometer sampled at 50 Hz (figure 1) was the primary source of the motion data. An 8-bit analog-to-digital converter was used in the CHIME monitor, discretizing the ACC signal into 256 levels, referred to as computer units. The CHIME monitor was calibrated to recognize certain ranges of the ACC signal as the indicator of the infant's position. Although the raw ACC signal was recorded by the monitor, it was stored on one of the hidden memory channels and clinicians in the CHIME study had access only to the decoded position information. This study focused on the part of the

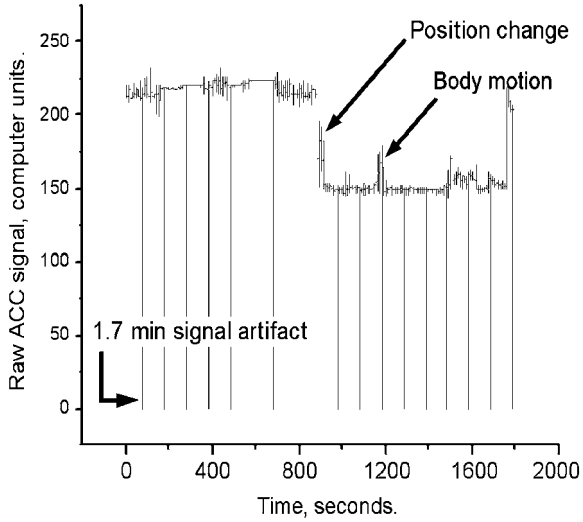

Figure 1. Raw accelerometer signal as recorded by a CHIME monitor. The artifact in the signal appears on the graph as vertical lines starting at 0.

accelerometer signal generated by body activity and normally disregarded the positional components.

Simultaneously with the acquisition of the physiological indicators, independent PSG-based sleep state identification was performed by trained technicians on the infants with 30 s intervals (epochs) (Crowell *et al* 1997). The PSG-identified sleep states (Awake, Active sleep, Quiet sleep and Indeterminate) were used as a baseline for building the regression model and training of the neural network.

In order to accommodate further analysis, the raw ACC signal for every infant was preprocessed in the following manner:

1. Both the beginning and the end of the accelerometer recording were synchronized in time with the PSG data on an epoch (30 s) boundary. Normally, the start and end times of the PSG and ACC recordings did not match. Furthermore, the ACC recording had a different sampling interval than the PSG recording. To ensure perfect alignment of the two signals, the PSG and ACC recordings were matched starting from the known timestamps saved by the CHIME and PSG monitors with subsequent trimming of unmatched tails from both recordings. The validity of the timestamps was ensured by synchronizing internal clocks of the CHIME and PSG monitors before a recording session.
2. An artifact of unknown origin (figure 1) was removed from the signal. The artifact appeared as a momentary, single-sample transition of the ACC signal to zero, occurring with the precise period of 1.7 min. This periodic and abrupt nature of the observed phenomenon qualify it as a data acquisition artifact and justify its removal.
3. Position-related DC levels in the ACC signal were removed by applying discrete Fourier transform (DFT) to an epoch of the signal (50 Hz $\times$ 30 s $=$ 1500 data points), zeroing the DC harmonic and applying the inverse DFT to the data. While a DFT conversion was used to minimize programming time due to readily available source code, other, less computationally demanding techniques such a 'floating mean' could be utilized to the same effect.

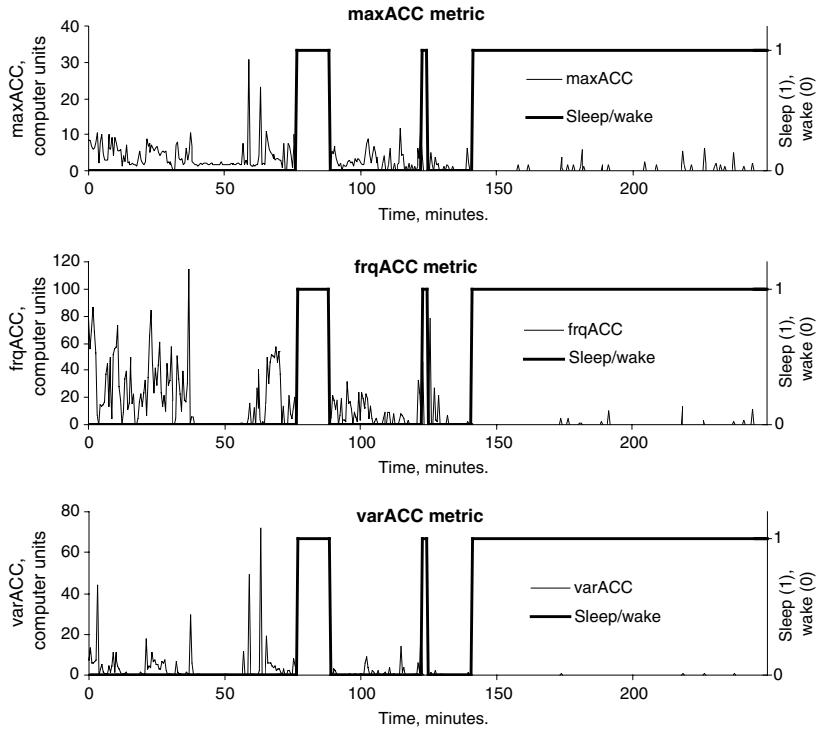

Figure 2. Plots of *maxACC*, *frqACC* and *varACC* versus sleep/wake state.

4. The noise of measurement equipment was minimized by applying a small threshold (± 1 computer units) to the signal obtained by the inverse DFT. The best performing value of the threshold was established by comparative statistical analysis of the accelerometer signal under different threshold values from the set of $\{0, \pm 1, \pm 2\}$. Details of the threshold selection are given in section 3.
5. The following metrics were extracted from the signal for each epoch: maximum and variance of accelerometer readings (*maxACC* and *varACC*), and frequency of changes in the ACC readings (*frqACC*). The *frqACC* metric was computed by counting number of zero crossings by the ACC signal during an epoch. The metrics are illustrated in figure 2.

The methods chosen to address the problem were logistic regression and neural networks as opposed to rather broadly used discriminant analysis (Blood *et al* 1997, Sadeh *et al* 1994, 1995). This choice was motivated by the advantage of these methods over discriminant analysis in the assumptions to the model. Logistic regression does not require normally distributed data while neural networks are capable of nonlinear mapping of the data. The accelerometer data clearly showed high degree of positive skewness and, therefore, the use of discriminant analysis was not justified.

Predictive power of logistic regression and neural networks was compared, first, by producing individual predictive models for each of the subjects in the dataset and, second, generalized models for the subject population.

The following variables were used in the analysis.

- The response variable is represented by the PSG records of Wake, Active or Quiet sleep for each 30 s epoch. Indeterminate states from the original PSG record and corresponding lagged predictors were removed from the dataset.

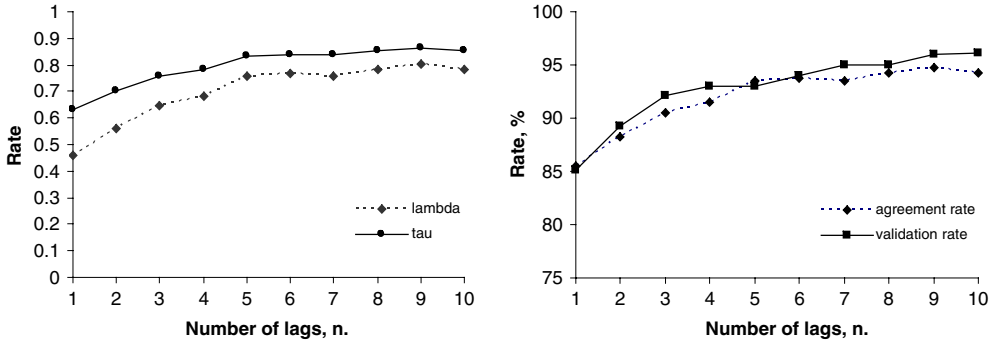

Figure 3. Comparative analysis of regression models with different number of lagged variables.

- The predictors are *maxACC*, *varACC* and *frqACC* metrics, taken for the last n consecutive periods, forming $(n + 1)$ separate predictors for each of the three metrics, so that there was a total of $3(n + 1)$ predictors included in the original model.

Utilization of lagged metrics as predictors was based on a simple reasoning that there could be no sudden change in the wake/sleep state, i.e. such changes may be expected to follow some preliminary changes in infant's conditions. Similar lagged models were used in related studies on adults (Cole *et al* 1992). The Active and Quiet sleep states were combined, so that the response vector contained only 'Wake' and 'Sleep' states.

The value of the threshold used in data preprocessing was established assuming the following form of the regression model:

$$\begin{aligned} Logit\ p = \eta = {} & \alpha + \beta_1\, maxACC_0 + \beta_2\, maxACC_1 + \cdots + \beta_{n+1}\, maxACC_n + \gamma_1\, varACC_0 \\ & + \gamma_2\, varACC_1 + \cdots + \gamma_{n+1}\, varACC_n + \theta_1\, frqACC_0 \\ & + \theta_2\, frqACC_1 + \cdots + \theta_{n+1}\, frqACC_0 + \varepsilon, \end{aligned} \tag{1}$$

$$Logit\, p = \log(p/(1-p)), \qquad p = (1 + e^{-\eta})^{-1}, \tag{2}$$

where n represents the number of previous 30 s epochs for which the accelerometer metrics are included in the predictor set and p is the probability of sleep. This model predicts a state to be 'Sleep' if $p > 0.5$ and 'Wake' otherwise, which corresponds to η being either positive or negative, respectively. The value of n was taken to be 8, which corresponds to 4.5 min of previous activity recording before a prediction is to be made. Approximately the same time interval of prior recording was used in previous studies on infants (Blood *et al* 1997, Sadeh *et al* 1995) and adults (Cole *et al* 1992). Increasing the value of n beyond 8 adversely affects the goodness of fit and the quality of prediction. A study done on a particular subject (figure 3) shows that by the time the number of lagged variables reaches 8 both agreement and validation rate flatten out and stop improving. In addition, measures representing a proportional reduction in error λ_p and τ_p (Venables 2002) show their best (highest) values for $n = 8$. Inclusion of more than eight lagged variables would result in the overfitting the data and decrease the quality of model's prediction.

The threshold was selected by analysis of the results obtained from regression models utilizing the *maxACC*, *varACC* and *frqACC* metrics which were preprocessed with different thresholds. The set of best predictors was chosen from the available metrics in order to minimize the number of predictor variables. The selection was performed starting with the original model (1) applied to each individual subject and by subsequent elimination of the

least significant or minimally contributing variables. The results showed low significance of the *varACC* and *frqACC* metrics, leading to the updated form of the regression model:

$$Logit\ p = \eta = \alpha + \beta_1\, maxACC_0 + \beta_2\, maxACC_1 + \cdots + \beta_{n+1}\, maxACC_n + \varepsilon, \tag{3}$$

$$Logit\ p = \log(p/(1-p)), \qquad p = (1 + e^{-\eta})^{-1}. \tag{4}$$

All further regression models were built using (3) and (4).

Individual regression models were developed separately for each subject by utilizing formulas (3) and (4), and taking approximately half of the observations as a training (calibration) set and another half as a validation set.

Because building an individual model requires 4–8 h of preliminary measuring with polysomnographic equipment and subsequent grading of the PSG data, practical application of individual models is limited. This fact motivated the search for a uniform model that can be applied to any subject with available activity measurements. The combined or general model was created on the data pooled from a subset of 13 (out of 26) infants: 13 infants created the training sample and the remaining 13 infants were used as a validation sample.

Learning vector quantization (LVQ) neural network (Kohonen *et al* 1995), a subclass of the so-called Kohonen networks, was used to build the neural predictor. Classification by LVQ networks is based on a codebook of vectors m_i, where each codebook vector belongs to a certain class. The input vector X is compared to each code book vector m_i and assigned to the same class to which the closest codebook vector belongs. The distance d between X and m_i is calculated according to the following formula: $d = \min_i \| X - m_i \|$.

Two variations of the supervised training procedure, known as LVQ-1 and LVQ-3, differ in the way the codebook vectors are obtained. LVQ-1 utilizes the simplest adaptation procedure and features fast convergence, while LVQ-3 represents a more robust approach that employs differential adaptation of the codebook vectors combined with stabilization factors that ensure correct approximation of the class distributions. The LVQ networks are primarily used as nonlinear classifiers, which makes this method a perfect candidate for the task at hand. More information on the LVQ neural networks can be found in the text on self-organizing maps (Kohonen 2001).

The best set of parameters for the neural network (number of codevectors, training method, etc) was established by building individual neural predictors for all the infants in the dataset. Several individual predictors, varying in the number of codevectors and the training method (LVQ-1 or LVQ-3), were trained for each subject. The training procedure consisted of 100 000 initial training iterations by LVQ-1 and subsequent 100 000 training iterations by either LVQ-1 or LVQ-3. The number of codevectors in a neural predictor varied from 2 to 256. The best set of parameters was identified by finding the maximum (a function of the neural network parameters) of the median of the training and validation rates for each infant in this group. The same training and validation sets were utilized for the neural predictors as for the logistic regression, therefore both agreement and validation rates of the neural predictors can be directly compared to the results provided by individual regression models. The same is true for the combined neural predictor.

3. Results

The value of the threshold used for preprocessing (± 1 computer units) was selected from the set of $\{0, \pm 1, \pm 2\}$, where threshold with value 0 corresponded to no threshold, i.e. to the original data. This was intended to establish acceptable threshold value as well as justify the applicability of the thresholding for improving the quality of data by reducing the noise of

Table 1. Summary statistics for a regression model with different thresholds.

	Agreement rates				
Model	Wake	Sleep	Average	AIC	R^2_{LA}
Model with threshold 0	91.7	94.7	93.2	249.6	0.6426
Model with threshold 1	92.1	95.1	93.6	230.1	0.6707
Model with threshold 2	85.8	92.6	89.2	339.4	0.5130

Table 2. Performance of models with different set of predictors.

Variables include	Agreement rate	Validation rate	R^2_{LA}
maxACC, *varACC*, *frqACC*	93.6	97.4	0.7
maxACC, *varACC*	94.2	97.4	0.7
MaxACC	93.0	97.9	0.7

the data acquisition equipment in CHIME monitors. Logistic regression with three 8-period lagged variables (1) was applied separately to three datasets, obtained on the same subject, but each preprocessed with one of the three different thresholds. The results for a particular subject are presented in table 1. All other subjects exhibited similar outcomes. As can be seen from table 1, the highest accuracy (agreement rate) and the lowest value of Akaike information criterion (AIC) (Venables 2002) were achieved by the model with data subjected to the threshold ± 1, indicating the best goodness of fit. In addition, the adjusted coefficient of determination for the logistic regression (R^2_{LA} metric) (Menard 2001) has the highest value for the model with the threshold ± 1, indicating the best general fit of the data to the regression model. Thus, the best fit was provided by the data processed with the threshold ± 1, and the rest of the analysis was conducted using this threshold.

The best set of predictors for an individual model was selected from the available pool of *maxACC*, *varACC* and *frqACC* metrics (figure 2). Table 2 demonstrates agreement and validation rates for the three models with different activity metrics acting as predictors. The models were constructed for a particular subject (the rest of the subjects from the study group demonstrated similar behavior). As table 2 shows, the predictive characteristics are slightly better for the model containing only the *maxACC* metric. The *maxACC* metric exhibited much greater significance than any of the other two variables, while the *varACC* was highly correlated to *maxACC* and did not provide much improvement. The rest of the analysis was carried out assuming only the *maxACC* metric as the predictor.

The individual regression models for 26 subjects were constructed using the *maxACC* metric as the predictor and (3) as the form of the logistic regression (table 3). On average, the accuracy of sleep state prediction was 90.6% and the accuracy of wake prediction was 54.82%. The average rate of prediction for both training and validation sets yielded an accuracy of 79.41%. This is comparable to the results of a similar study by Cole *et al* (1992).

The combined logistic regression model was constructed from the data pooled from 13 subjects constituting the training set. Splitting into the validation and training sets for the general model was performed by systematic sampling procedure, i.e. every second subject in the arbitrarily organized list was taken into the validation group, the rest subjects had formed training group. This method is known to be a very good proxy for formal randomization (Lohr 1998).

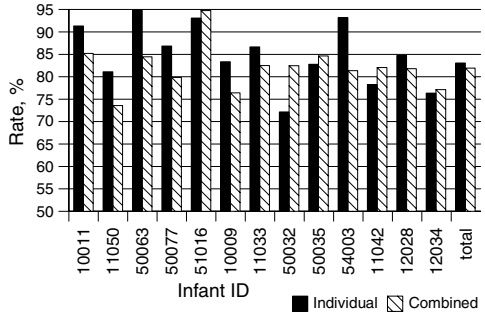

Figure 4. Comparison of agreement rates for individual and combined regression models.

Table 3. Average agreement and validation rates for the individual neural predictors versus individual regression models.

	Accuracy of sleep prediction (%)		Accuracy of awake prediction (%)		Weighted average (%)	
	Regression	Neural	Regression	Neural	Regression	Neural
Agreement rate	93.5	95.2	63.6	65.2	83.0	90.0
Validation rate	87.7	84.4	46.0	50.8	75.7	75.3
Average	90.6	89.8	54.8	58.0	79.4	82.6

Table 4. Agreement and validation rates for the combined neural predictor versus combined regression model.

	Accuracy of sleep prediction (%)		Accuracy of awake prediction (%)		Weighted average (%)	
	Regression	Neural	Regression	Neural	Regression	Neural
Agreement rate	94.7	95.1	46.5	55.4	80.5	83.4
Validation rate	93.2	91.2	41.0	48.3	75.4	76.6
Average	93.9	93.6	43.8	51.8	77.9	80.0

Running logistic regression on the model (3) resulted in the following expression for the probability of sleep prediction:

$$\begin{aligned} Logit\ p = {} & 1.996\,04 - 0.194\,50\,maxACC_0 - 0.097\,46\,maxACC_1 - 0.099\,75\,maxACC_2 \\ & - 0.101\,94\,maxACC_3 - 0.089\,17\,maxACC_4 - 0.081\,08\,maxACC_5 \\ & - 0.074\,94\,maxACC_6 - 0.073\,00\,maxACC_7 - 0.102\,07\,maxACC_8, \end{aligned} \tag{5}$$

where all coefficients were significant at 0.001 level.

The performance of the combined regression model was comparable to the performance of the individual regression models. Tables 3 and 4 show that the combined model reaches almost the same degree of accuracy as the individual models: average accuracy of prediction for the individual models is 79.41% whereas the same statistic for the combined model is 77.93%. In addition, figures 4 and 5 give a graphical representation of the individual agreement and validation rates in comparison to the agreement and validation rates for the training set of the combined model.

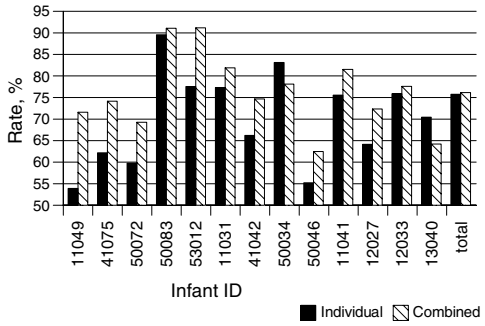

Figure 5. Comparison of validation rates for individual and combined regression models.

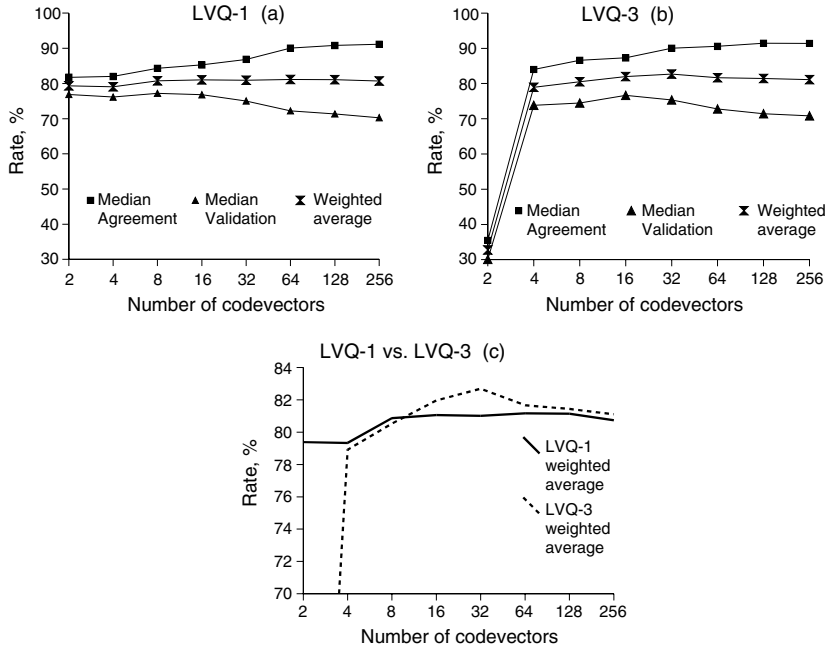

Figure 6. (a) Dependency of median agreement and validation rates on the number of codevectors for LVQ-1 training procedure; (b) same for LVQ-3 training procedure; (c) comparison of average prediction rates for LVQ-1 and LVQ-3.

Next, similar analysis was done using neural networks where the selection of the best set of parameters for the neural network was performed by training a number of neural predictors for each subject in the dataset. Figure 6 illustrates dependency of the median (across the population) agreement and validation rates on the number of codevectors in a neural network and on the utilized training procedure (LVQ-1 or LVQ-3). The graphs clearly indicate that increasing the number of codevectors over 32 overfits the neural predictors, causing higher agreement rates but lower validation rates. This tendency is more pronounced for the LVQ-1 training procedure. Based on the performed evaluation, the parameters of the neural predictor were set to be 32 codevectors, LVQ-3 training procedure.

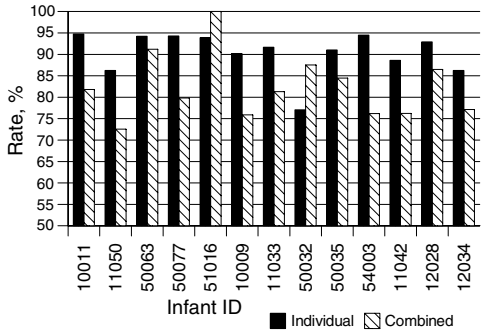

Figure 7. Comparison of agreement rates for individual and combined neural predictors.

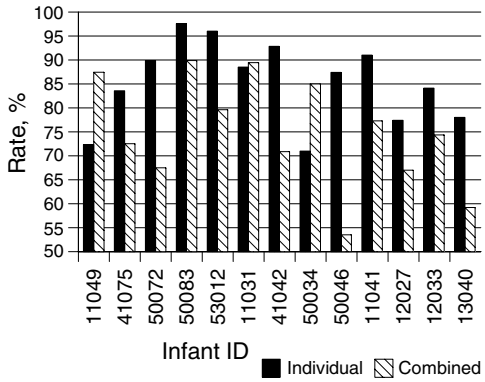

Figure 8. Comparison of validation rates for individual and combined neural predictors.

The individual neural predictors for 26 subjects were constructed using the same data as for the regression models. On average, the validation rate for the sleep state prediction was 84.46% (versus 87.7% by the individual regression models, table 3) and the accuracy of wake prediction was 50.87% (versus 46.01%). The average rate of prediction for both training and validation sets yielded accuracy of 82.69% (versus 79.41%).

The combined neural predictor was trained on the same data as the combined regression model (13 subjects constituting the training set and other 13 subjects constituting the validation set). The performance of the neural predictor was close to the performance of the combined regression model. Table 4 shows the agreement and validation rates for the combined neural predictor compared to the combined regression model. Average accuracy of prediction for the combined neural predictor is 79.97% whereas the same statistic for the combined regression model is 77.93%. However, the neural predictor has significantly higher accuracy of wake state prediction: 51.83% versus 43.79% for the combined regression model. Figure 7 presents the comparison of the agreement rates for the individual neural predictors compared to the agreement rates for the training set of combined neural predictors. Figure 8 illustrates validation rates for the individual and combined neural predictors (validation set).

4. Discussion

The results of this study have several important implications. First, body motion in infants can be used for determination of sleep/wake states and has predictive power similar to that

of the limb motion. Indeed, previous studies reported 85–95% prediction rates of sleep/wake identification achieved by actigraphy (Sadeh *et al* 1995, Cole *et al* 1992). For example, an earlier infant study (Sadeh *et al* 1995) reported the combined accuracy of sleep/wake detection for newborns to be 88.9%, with an accuracy of wake state identification of 82.8%. Both statistical and neural predictors of this study provide a comparable accuracy of approximately 77–92%. The accuracy of both sleep and wake state prediction for the best specimen was over 90%.

The results show that combined predictor models achieved both agreement and validation rates close to those from the individual models. Utilization of individual models requires preliminary 4 to 8 h screening of a subject with expensive and not always available polysomnograph equipment, followed up by building a regression or a neural model individually for each subject. In contrast, application of the combined models does not call for preliminary screening and does not involve any additional work on model construction. Therefore, the combined models can be easily applied in home conditions.

It should also be noted that both regression models and neural predictors had a tendency to overpredict the sleep state and underpredict the wake state. An explanation for this effect may lay in the behavioral aspect of sleep/wake identification using actigraphy and positioning of the motion sensor on the infant's diaper. As noted in the previous studies (Blood *et al* 1997), prolonged periods of subject's inactivity result in overprediction of the sleep state. An infant in quiet alert state, resting without motion would be classified as being asleep by the actigraphic methods and awake by PSG. The same tendency was observed in Sadeh's study (Sadeh *et al* 1995), where the accuracy of awake identification improved with infant's age and at the age of 6 month became comparable to that of an adult. Position of the motion sensor is critical for the accurate classification of the sleep/wake states (Sadeh *et al* 1994). Positioning the motion sensor on the diaper allows detection of a significantly reduced range of motion (compared to a limb). Subtle motion of the limbs and body may not be detectable on the diaper, reducing the accuracy of awake state identification.

The error in sleep state identification may be attributed to the forced motion of an infant. Unlike older children and adults, whose motion is largely independent, motion of infants may be forced without changes in the actual sleep/wake state. Such a forced motion would create a discrepancy between the PSG recording and motion recording, detecting an awake state when the infant was actually asleep, lowering the accuracy of sleep prediction.

The variability in the agreement and validation rates may also be attributed to a particular birth group (healthy full-term, preterm asymptomatic, etc) and future analysis may include stratified analysis for each such group retaining all the same variables that are represented in the current study. In addition, more explanatory variables such as birth weight, gender and postconceptional age can be included into analysis.

Second, the results of this study show that logistic regression can be successfully applied to the problem of sleep/wake detection. The results show comparable levels of agreement and validation rates as in similar rates for studies on adults where discriminant analysis was used. The combined regression model showed quality of prediction comparable to that of the individual regression models and proved itself to be a good alternative to individual models. The combined regression model considerably improves practical applicability of logistic regression for determination of sleep/wake states and allows easy use in home conditions.

Third, results provided by the neural predictors support the results of statistical analysis and display similar accuracy and trends. A promising direction of further research would be utilization of a self-organizing neural network that could bring the prediction performance of the combined neural predictor close to the performance of an individual predictor by adapting to the individual trends displayed by each subject.

Finally, the study validates the usage of an accelerometer for determining both the infant's sleep/wake state and position. Using one sensor to perform both position detection and acquisition of the motion data drives down the cost of the home monitoring system, minimizes intrusiveness on the subject and simplifies collection of data. All these features of the proposed method will facilitate development of the equipment for home infant monitoring and SIDS research. Furthermore, ability to determine sleep/wake states without traditional PSG measurements would allow for analysis of over 7000 003 min epochs of the home data (not PSG-scored) collected in the CHIME study.

5. Conclusion

This study presented a method for scoring sleep/wake states of infants based on activity measurements. The method utilized an accelerometer attached to the infant's diaper for a dual purpose: to determine an infant's position in the crib and to identify sleep/wake states. This sensor position is significantly different from commonly used attachment to an ankle. The results of the study show that such sensor attachment to the diaper is capable of generating motion activity related to the sleep states and is capable of producing accuracy of sleep/wake identification (77–92%) which is comparable to other studies. At the same time, reduced motion of the diaper leads to underprediction of the wake state.

Logistic regression and neural networks showed comparable performance in classification of the sleep states and agreement in results with other actigraphy studies, thus validating the suggested methodology. Nonlinear capabilities of the LVQ neural networks allowed for better accuracy in scoring of the wake state. The results of this support the use of body motion as a simple and reliable method for determining wake–sleep states in infants.

The results of this study will be utilized for scoring of the home data collected in the CHIME study, where no PSG data is available and actigraphy presents the only viable alternative for determining sleep/wake states from a signal not directly related to the recorded cardiorespiratory activity. Further attempts will be made to improve on the accuracy of the actigraphic sleep scoring and to reduce the effect of wake state underestimation.

References

Acebo C, Sadeh A, Seifer R, Tzischinsky O, Wolfson A R, Hafer A and Carskadon M A 1999 Estimating sleep patterns with activity monitoring in children and adolescence: how many nights are necessary for reliable measures? *Sleep* **22** 95–103

Ancoli I S, Clopton P, Klauber M R, Fell R and Mason W 1997 Use of wrist activity for monitoring sleep/wake in demented nursing-home patients *Sleep* **20** 24–7

Ancoli I S, Cole R, Alessi C, Chambers M, Moorcroftm W and Pollak C P 2003 The role of actigraphy in the study of sleep and circadian rhythms *Sleep* **26** 342–92

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

Blood M L, Sacks R L, Percy D C and Pen J C 1997 A comparison of sleep detection by wrist actigraphy, behavioral response, and polysomnography *Sleep* **20** 388–95

Cole R J, Kripke D F, Gruen W, Mullaney D J and Gillin J C 1992 Automatic sleep/wake identification from wrist actigraphy *Sleep* **15** 461–9

Crowell D H *et al* 1997 Infant polysomnography: reliability. Collaborative Home Infant Monitoring Evaluation (CHIME) Steering Committee *Sleep* **20** 553–60

Gertner S, Greenbaum C W, Sadeh A, Dolfin Z and Sirota L 2002 Sleep-wake patterns in preterm infants and six month's home environment: implication for early mental development *Early Hum. Dev.* **68** 93–102

Gruber R, Sadeh A and Raviv A 2000 Instability of sleep patterns in children with attention deficit/hyperactivity disorder *J. Am. Acad. Child Adolesc. Psychiatry* **39** 495–501

Kohonen T 1995, 1997, 2001 *Self-Organizing Maps* 3rd extended edn (Berlin: Springer) 501 pp

Kohonen T, Hynninen J, Kangas J, Laaksonen J and Torkkola K 1995 *LVQ PAK: The Learning Vector Quantization Program Package* Laboratory of Computer and Information Science, Helsinki University of Technology, Finland

Kulp T D, Corwin M J, Brooks L J, Peucker M, Fabrikant G, Crowell D H and Hoppenbrouwers T 2000 The effect of epoch length and smoothing on infant sleep and waking state architecture for term infants at 42 to 46 weeks postconceptional age. *Sleep* **23** 893–9

Leitner Y, Mimouni-Bloch A, Sadeh A, Tikotzky L, Neuderfer O and Harel S 2002 Sleep-wake patterns in children with intrauterine growth retardation (IUGR) *J. Child Neurol.* **17** 872–6

Lohr S 1998 *Sampling: Design and Analysis* 1st edn (Pacific Grove, CA: Brooks Cole) 450 pp

Menard S 2001 *Applied Logistic Regression Analysis (Sage Publications. Series: 106; Quantitative Applications in the Social Sciences)* 2nd edn 112 pp

Neuman M R *et al* (the CHIME Study Group) 2001 Cardiopulmonary monitoring at home: the CHIME monitor *Physiol. Meas.* **22** 267–86

Ramanathan R *et al* (the Collaborative Home Infant Monitoring Evaluation (CHIME) Study Group) 2001 Cardiorespiratory events recorded on home monitors: Comparison of healthy infants with those at increased risk for SIDS *JAMA* **285** 2199–207

Sadeh A 1996 Evaluating night wakings in sleep-disturbed infants: a methodological study of parental reports and actigraphy *Sleep* **19** 757–62

Sadeh A and Acebo C 2002 The role of actigraphy in sleep medicine *Sleep Med. Rev.* **6** 113–24

Sadeh A, Acebo C, Seifer R, Aytur S and Carskadon M A 1995 Activity-based assessment of sleep-wake patterns during the 1st year of life *Infant Beh. Dev.* **18** 329–37

Sadeh A, Gruber R and Raviv A 2002 Sleep, neurobehavioral functioning and behavior problems in school-age children *Child Dev.* **73** 405–17

Sadeh A, Levie P, Scher A, Tirosh E and Epstein R 1991 Actigraphic home-monitoring sleep-disturbed and control infants and young children: a new method for pediatric assessment of sleep-wake patterns *Pediatrics* **87** 494–9

Sadeh A, Raviv A and Gruber R 2000 Sleep patterns and sleep disruptions in school-age children *Dev. Psychol.* **36** 291–301

Sadeh A, Sharkey K and Carskadon M 1994 Activity-based sleep-wake identification: an empirical test of methodological issues *Sleep* **17** 201–7

Shinkoda H, Matsumoto K, Hamasaki J, Seo Y J, Park Y M and Park K P 1998 Evaluation of human activities and sleep-wake identification using wrist actigraphy *Psychiatry Clin. Neurosci* **52** 157–9

Tikotzky L and Sadeh A 2001 Sleep patterns and sleep disruptions in kindergarten children *J. Clin. Child Psychol.* **30** 579–89

Venables W N 2002 *Modern Applied Statistics with S-PLUS* 4th edn (Berlin: Springer) pp 495

Wicklow A and Espie C A 2000 Intrusive thoughts and their relationship to actigraphic measurement of sleep: towards a cognitive model of insomnia *Behav. Res. Ther.* **38** 679–93

Xueyan X 2002 Prediction of life-threatening events in infants using heart rate variability measurements *PhD dissertation* West Virginia University.