Polysomnographic pattern recognition for automated classification of sleep–waking states in infants

P. A. Estévez[1] C. M. Held[1] C. A. Holzmann[1] C. A. Perez[1]
J. P. Pérez[1] J. Heiss[1] M. Garrido[2] P. Peirano[2]

[1]Department of Electrical Engineering, Universidad de Chile, Santiago, Chile
[2]INTA, Universidad de Chile, Santiago, Chile

Abstract—*A robust, automated pattern recognition system for polysomnography data targeted to the sleep–waking state and stage identification is presented. Five patterns were searched for: slow-delta and theta wave predominance in the background electro-encephalogram (EEG) activity; presence of sleep spindles in the EEG; presence of rapid eye movements in an electro-oculogram; and presence of muscle tone in an electromyogram. The performance of the automated system was measured indirectly by evaluating sleep staging, based on the experts' accepted methodology, to relate the detected patterns in infants over four months of post-term age. The set of sleep–waking classes included wakefulness, REM sleep and non-REM sleep stages I, II, and III–IV. Several noise and artifact rejection methods were implemented, including filters, fuzzy quality indices, windows of variable sizes and detectors of limb movements and wakefulness. Eleven polysomnographic recordings of healthy infants were studied. The ages of the subjects ranged from 6 to 13 months old. Six recordings counting 2665 epochs were included in the training set. Results on a test set (2369 epochs from five recordings) show an overall agreement of 87.7% (kappa 0.840) between the automated system and the human expert. These results show significant improvements compared with previous work.*

Keywords—*Sleep analysis, Pattern recognition, Sleep–waking states, Sleep–waking classification, Fuzzy sets, Polysomnography*

Med. Biol. Eng. Comput., 2002, **40**, 105–113

1 Introduction

The determination of whether an infant is awake or asleep and, if asleep, in which state or stage of it, serves several purposes in both clinical and physiological studies. Applications include the neuro-functional assessment of the central nervous system, nutritional deficits or those associated to evaluate the risk of sudden infant death syndrome (PEIRANO *et al.*, 1986; 1989; PEIRANO and CURZI-DASCALOVA, 1995; FERBER and KRYGER, 1995).

To investigate properly the characteristics of sleep, human subjects are made to sleep connected to a polysomnograph. This non-invasive test simultaneously collects data relating to several variables, such as the bio-electric activity of the brain (electro-encephalogram (EEG)), eye movements (electro-oculogram (EOG)) and muscle tone (electromyogram (EMG)). According to the accepted standard procedure for sleep scoring, these signals are analysed to extract patterns that allow the identification of the subjects' sleep–waking states. Other signals, such as body movements, provide context information that is useful for discarding artifacts (RECHTSCHAFFEN and KALES, 1968; GUILLEMINAULT and SOUQUET, 1979).

The EEG is analysed to establish the predominant background activity and the presence or absence of sleep spindles. The EEG background activity is defined by the predominant frequency range: alpha (8–12 Hz), theta (3–7 Hz), and slow (0.5–2 Hz) or fast (2–3 Hz) delta. The frequency of sleep spindles is in the sigma (10–16 Hz) range. The EOG is analysed to determine the presence or absence of episodes of rapid eye movements, and the EMG is analysed to determine the presence or absence of muscle tone.

Two distinct sleep states are defined as: rapid-eye movement (REM) or active sleep, and non-REM (NREM) or quiet sleep. NREM is usually subdivided into four stages, denoted by I–IV. However, as the main difference between stages III and IV is the required percentage (density) of the slow, large-amplitude delta waves, in this work they are considered together. Thus, the set of sleep–waking classes is: wakefulness (WA), NREM-I, NREM-II, NREM-III and IV and REM.

Traditionally, polysomnographic recordings are visually scored by well-trained personnel. This evaluation is extremely time-consuming and prone to inaccuracies, mainly owing to observer fatigue or distraction and inconsistencies between different observers.

The convenience of developing a computerised system for automated analysis and classification of sleep states has been recognised by different authors (BESSET, 1998; COLLURA *et al.*,

Correspondence should be addressed to Dr P. A. Estévez; email: pestevez@cec.uchile.cl

Paper received 20 March 2001 and in final form 13 November 2001

MBEC online number: 20023646

1993; GUILLEMINAULT, 1998; HARPER *et al.*, 1987; JANSEN and DAWANT, 1989; PARK *et al.*, 1990; 2000; PRINCIPE *et al.*, 1989; SMITH, 1986; TAFTI, 1998). A few commercial systems are also available (KUBICKI *et al.*, 1989; WHITE and GIBB, 1998). WHITE and GIBB (1998) evaluated the computerised system ALICE 3 using 50 subjects. The results of automated computer scoring showed substantial differences from the visually scored paper polysomnography in the distribution of the sleep stages. The manually edited computer sleep staging had 75.7% agreement with the paper polysomnography scoring.

The objective of the present work was to develop and improve pattern detection schemes for sleep staging. The research focused on the automated identification of the relevant patterns to perform the sleep staging. These patterns are: predominance of slow, large-amplitude delta (SD) waves and theta (TH) waves in the background activity of the EEG, presence of sleep spindles (SSs) in the EEG, presence of rapid eye movements (REM_{ov}) in the EOG and presence of muscle tone (MT) in the EMG. Our approach was to emulate the experts' procedure to identify these patterns, because it allows simpler explanations and justifications of the system results. Other approaches, for instance, use neural networks to detect characteristic waves in sleep EEG (BANKMAN *et al.*, 1992; ROBERTS and TARASSENKO, 1992; SHIMADA *et al.*, 2000).

The literature and experts generally use the term REM both for a distinct pattern in the EOG, where it is detected that the patient's eyes are moving rapidly, and for a sleep state, in which the pattern appears. In this paper, the pattern of movement is called REM_{ov}, and the sleep state is called REM.

The real data acquisition operation introduces substantial noise and artifacts into the recordings. Human experts look for context information that alerts them about the plausible noise content of a recording at a particular point in time. Similarly, the developed system looks for relevant noise factors, e.g. it senses limb movement (LM) and has a separate algorithm to detect WA.

The present work is the continuation of that presented in HOLZMANN *et al.* (1999). In our previous work, a significant number of recorded data obtained from infants had to be discarded (14.6%) because of artifacts, missing scales to apply to the data or incomplete classification marks. In other studies, noise was also manually eliminated on-line (SCHER, 1997). In the present study, no expert pre-processing of the recordings is performed, i.e. raw data are used, the recordings are analysed chronologically, and no portion of the data is eliminated.

The EEG patterns of newborns and infants differ from those of adults (ANDERS *et al.*, 1971; CURZI-DASCALOVA and MIRMIRAN, 1996). However, after 4 months of post-term age, the EEG patterns that allow the differentiation of NREM stages (I–IV), particularly SS and SD, are already identifiable. Based on this evidence, GUILLEMINAULT and SOUQUET (1979) proposed a classification method for infants beyond 3 months of post-term age that uses an epoch-by-epoch approach, based on the RECHTSCHAFFEN and KALES (1968) visual scoring system. Table 1 summarises these criteria, which require simultaneous concordance of the above mentioned activities and episodes to identify the sleep–waking states and stages in infants. Notice that, when the criteria established in Table 1 are not met, epochs are classified as indeterminate sleep (IS). Although IS classification is relevant in the sleep analysis of infants (CURZI-DASCALOVA and MIRMIRAN, 1996; PEIRANO *et al.*, 1989), this pseudo-state occurred very rarely in the available data. Thus IS classification was not targeted in the present work.

2 Methods

2.1 *Subjects and recordings*

Eleven polysomnographies of healthy infants, recorded at the Sleep Laboratory of the Instituto de Nutrición y Tecnología de Alimentos, INTA, Universidad de Chile, were studied. The post-term ages of the subjects ranged from 6 to 13 months old. The procedures were standardised to limit the potential influences of environment, circadian rhythms and/or food intake on sleep–waking patterns and related physiological activities.

Studies took place in a special, quiet and comfortable room, during the infant's spontaneous afternoon nap. Infants and their mothers were transported the short distance from their homes to the laboratory so that they arrived 1 h before the habitual noon meal. During this 1 h period, electrodes were attached according to routine procedures. Mothers then fed their infants and put them down for a nap following their usual routines. The mothers stayed in the room throughout the data acquisition period. Infants were normally dressed and lying in the supine position. Ambient temperature was maintained constant at 22–24°C.

Recordings were made continuously on a TECA 1A97 18-channel polygraph as follows: EEG channels with electrode placement adapted for infants from the international 10-20 system (FP1-C3, C3-O1, FP2-C4, C4-O2 and C3-C4); EOG for REM_{ov} detection; tonic chin and diaphragmatic EMGs, and ECG (electrocardiogram) using surface electrodes; detection of motor activity of both upper and lower limbs recorded independently by piezo-electric crystal transducers; abdominal ventilatory movements, recorded using a mercury strain gauge; and nostrils airflow, recorded by means of a thermistor. All data were simultaneously recorded on paper and digitally, the signals being converted on-line from analogue to digital at a 250 Hz sampling rate. The digital data were collected on hard disk and then stored in magnetic or laser media for off-line analyses. Infant behaviour was also observed directly and annotated on the polygraph paper.

An experienced clinical neurophysiologist, who was a sleep expert, scored the paper records in a blinded fashion using standard criteria (GUILLEMINAULT and SOUQUET, 1979). The analysis was made on a 'one-page' frame, i.e. the expert looked at one page at a time for an initial assessment of the sleep–waking state.

Although page-by-page scoring is performed, state stability throughout successive pages is required before a recording period can be assigned to a given state (or an NREM stage). The minimum duration is 1 min (ANDERS *et al.*, 1971;

Table 1 Pattern concordance for sleep–waking state classification in infants, according to expert criteria

Patterns \ State/stage	WA	NREM-I	NREM-II	NREM-III and IV	REM sleep
EEG: slow delta	no*	no	no	yes	no
EEG: theta	X	yes	X	X	yes
EEG: sleep spindles	no	no	yes	X	no
EOG: REM_{ov} signal	yes	no	no	no	yes
EMG: muscle tone	yes	X	X	X	no

Yes = present; no = absent; X = do not care
* Although physiological slow delta waves are absent during WA, artifacts resembling slow delta waves are not uncommon. Other disturbances caused by signals from others sources, movements etc. are particularly pervasive in WA

RECHTSCHAFFEN and KALES, 1968; CURZI-DASCALOVA and MIRMIRAN, 1996). As the speed used to record on paper is typically 20 or 30 s per page (15 mm s^{-1} or 10 mm s^{-1}, respectively), experts consider three or two consistent consecutive pages, respectively, to determine a change in state or stage.

Moreover, the 'one-page analysis' was not isolated but was put in context, as some phasic patterns, such as sleep spindles, are sporadic in nature and thus do not need to be present on each page to assert their presence, as will be explained later.

The proposed automated system followed the expert procedure for analysing pages. To standardise the algorithms, 20 s 'epochs' were always considered, neglecting the actual paper record that had 20 or 30 s per page.

Out of the 11 recordings available, six of them were used for system training, and five served as an independent test set. Table 2 shows some characteristics of the recordings and their corresponding subjects.

2.2 *Digital filter pre-processing*

The EEG and EOG channels were pre-processed using a 10th-order digital FIR filter that acted as a comb filter to eliminate the power-line frequency of 50 Hz and its harmonics (PROAKIS and MANOLAKIS 1996). The digital filter applied to the EEG and OG channels also had a smoothing effect, as the main lobe of the filter has a -3 dB gain at about 16 Hz. The EMG channel was filtered using a high-pass 10th-order digital FIR filter, with a cutoff frequency of 16 Hz.

2.3 *EEG signal processing*

The EEG in humans aims to record the electrical activity inside the skull, particularly of the brain, in a non-invasive way. The electrodes are placed on the head skin. Transmission of the electrical waves of interest between inside the skull and the electrodes is hindered and contaminated with noise and artifacts originating from deficient electrical contact, physical distance and barriers (skin, tissue, bone), and other electromagnetic sources. Hence, the EEG signals are considerably weak and particularly noisy.

Out of the five EEG signals available in the polysomnography, four were considered in this study. Two of these were 'posterior derivations' (C-O: channels 2 and 4), and two were 'anterior derivations' (FP-C: channels 1 and 3). As explained below, the EEG recordings were analysed to find a predominant background activity and to detect SS. Usually, posterior derivations (C-O) are the primary reference for background activity, and the anterior derivations (FP-C) are preferred for SS detection, although the others can also be considered.

2.3.1 *Slow delta wave background activity (SD):* SD waves are defined by frequencies between 0.5 and 2.0 Hz and peak-to-peak amplitudes of at least 120 µV. Our first approach to detect SD waves involved mimicking the expert procedure, e.g. a zero-crossing method (HOLZMANN *et al.*, 1999).

In the current version of the algorithm, a 'predominant presence' of SD is defined based on the power spectrum contained in EEG channel 4 (C3-O1). This is obtained by applying the fast Fourier transform (FFT) to the EEG signal, using a moving window of 1024 samples with a step size of 128 samples. The spectra of all windows within an epoch are averaged, yielding one spectrum per epoch. A Hamming filter with a moving window of 7 epochs is applied to each discrete FFT frequency component in the delta wave range (0.49 Hz – 3.17 Hz). Then, a heuristic power index is calculated as the difference between the mean power contained in the SD range (0.49 Hz –1.95 Hz) and the mean power in the fast delta wave range (2.20 Hz –3.17 Hz). During a long SD activity period, if the SD power decays, experts still consider it as present. Therefore a filter that enhances the heuristic SD power index between high values within 1 min is applied. Then, the heuristic power index is normalised to the [0, 1] interval.

Limb movements and other artifacts can cause false positive (FP) detection of SD in the EEG channels, particularly during WA. To avoid these FPs, a WA detection algorithm was used, which is explained in Section 2.7.

To determine the predominant presence of SD during an epoch, as required to define NREM III and IV, a threshold was applied over the normalised heuristic power index. The threshold was determined for each recording as follows. The average heuristic power index for the whole recording was calculated, excluding the previously detected WA epochs. If the average heuristic power index for the recording was greater than 0.2, then the threshold value was set at 0.23; otherwise, a threshold of 0.16 was used.

2.3.2 *Theta wave background activity (TH):* Experts rely on visual judgment to assess TH predominance; there is no measurement procedure ready to automate. A 'predominant presence' index for TH is defined based on relative energy. This relative power index shows the amount of power within the band of 3–7 Hz with respect to the total power calculated from the FFT of the EEG signal, considering data taken from a moving window of 512 samples with a step size of 64 samples. The size of the window, about 2 s, was chosen to take into account frequencies as low as 0.5 Hz. Data were acquired from channel 4, as for SD detection.

A two-step criterion to assess TH predominance was established, where thresholds were set based on training data. TH

Table 2 Characteristics of 11 polysomnographic recordings used in study and corresponding subjects. Number of epochs indicated in different sleep states or stages correspond to expert visual scoring. Very small fraction of recordings were classified as indeterminate sleep (IS), when criteria of Table 1 were not met in periods lasting at least 1 min

	Recording identification	Sex	Age, months	Number of epochs	WA epochs	NREM-I epochs	NREM-II epochs	NREM-III and IV epochs	REM epochs	IS epochs
Training recordings	AC100394	M	7	558	73	151	77	196	61	0
	CR082995	F	12	459	215	28	65	135	16	0
	CV061493	M	12	668	122	80	229	131	101	5
	FH120594	M	13	396	83	30	80	179	24	0
	PG031693	F	12	390	58	89	188	0	55	0
	PG091592	F	6	194	48	42	80	0	21	3
Test recordings	AM102793	F	6	464	110	54	50	178	72	0
	BH072093	M	12	356	122	35	58	120	19	2
	KD032493	F	6	501	107	43	290	4	57	0
	SS051893	F	7	510	239	17	148	76	25	5
	SS101893	F	12	538	54	55	99	280	49	1

predominant presence in each moving window was determined if at least 32% of the total energy lay in the TH range. TH epoch predominance was established at a threshold of 0.1 of the moving windows of the epoch, using the [0, 1] range. In other words, TH predominance was established if at least 10% of the moving windows of the epoch had at least 32% of the energy concentrated in the TH range.

2.3.3 *Sleep spindles (SSs):* Experts describe a sleep spindle as a sequence (train) of fast (sigma) EEG waves, lasting more than 0.5 s and with a magnitude above 10 μV. Spindle activity has been reported to reside between 10 and 16 Hz (Jankel and Niedermeyer, 1985). Usually, spindles are distinguished from background sigma activity using an amplitude threshold. The SSs are also usually mounted on other slower waves with larger amplitudes in the EEG; this renders inadequate the signal analysis methods applied in SD and TH.

In our method, the search began with identification of three consecutive peaks determined by three consecutive sign changes in the instantaneous slope of the signal. The slope was determined by adjusting a straight line by the least square error method, using five consecutive points. The peaks were labelled as left (L), centre (C) and right (R), and identified by (amplitude, time) pairs: (A_L, t_L), (A_C, t_C), (A_R, t_R), as shown in Fig. 1*a*.

The signal period was given by the time elapsed between two consecutive peaks of the same sign, i.e. $(t_R - t_L)$. As in the other cases, noise can form patterns that will fall in the sigma frequency range. One way to diminish this interference was to require a certain degree of symmetry between the two half-waves, which is expressed as

$$(t_C - t_L) \geq K(t_R - t_L) \quad \text{and} \quad (t_R - t_C) \geq K(t_R - t_L)$$

where K is a symmetry factor. The value for K was fixed, using training data, at $K = 0.43$.

Neighbouring signals complying with the previous criteria were candidates for forming sigma trains. It was checked that they had peak-to-peak amplitudes of at least 10 μV and were standing less than 0.45 s apart. Trains lasting less than 0.56 s were discarded, and trains closer than 0.5 s were merged. These parameters were found empirically using the training data.

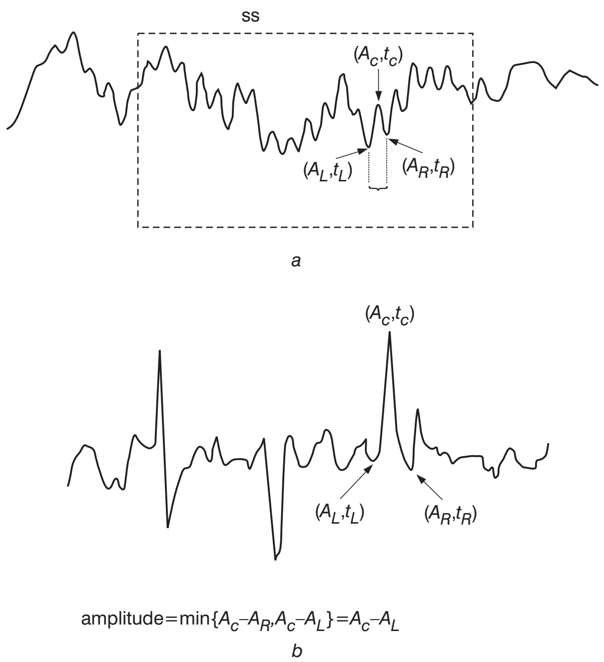

Fig. 1 *Time-domain analysis for pattern detection: (a) SS search in EEG signal; (b) REM_{ov} search in EOG. Three consecutive sign changes in slope identified three consecutive peaks, labelled as left (L), centre (C) and right (R), and denoted by (amplitude, time) pairs: (A_L, t_L), (A_C, t_C), (A_R, t_R). Period = $(t_R - t_L)$*

Although SSs are normally detected in the anterior EEG channels, tests using training data showed better results when all four EEG channels were searched. The presence of an SS in one channel was enough to assert its presence.

Following expert reasoning, a normalised SS quality index considering the train amplitude and duration was developed. The amplitude of a signal train is defined as the average of the tallest 30% peak-to-peak values of the signals in the train.

Amplitude and train duration were evaluated against recording-specific parameters, normalised by the 70% percentile of train amplitude and 70% percentile of train duration, respectively. These values were saturated, i.e. amplitudes within the tallest 30% were assigned a normalised amplitude of 1.0. To avoid noise contamination in determining the peak values, strict thresholds for duration and amplitude were considered. The quality index I_{SS} for each SS is defined by

$$I_{ss} = \left\{ \begin{array}{lll} A_{ss}{}^2 T_{ss} & \text{if} & A_{ss} > T_{ss} \\ A_{ss} T_{ss}{}^2 & \text{if} & T_{ss} > A_{ss} \end{array} \right\}$$

where A_{ss} and T_{ss} correspond to normalised amplitude and duration, respectively. The definition of I_{ss} takes into account that the quality index of short-time, high-amplitude SSs should be as good as that of long-time, low-amplitude SSs. If there was more than one sigma train in the same epoch, the I_{ss} value for the epoch was given by the largest quality index.

Analysis of the scored training recordings showed that sigma trains towards the beginning and the end of NREM-II were declared SS by the expert, in spite of not complying with the minimum standard parameters. They were described as 'weak' SS. Therefore the SS detection parameters were made more permissive in limited time frames before and after a series of 'solid' SS were detected, to detect the 'weak' SSs but minimise false SSs caused by noise.

Limb movement is one source of false positive SSs. Another is fast alpha activity during WA periods (Huuponen *et al.*, 2000). Information from the limb movement channels (see Section 2.6) and the WA detection algorithm (see Section 2.7) were used to avoid FP SS detection.

SSs are phasic events that are sporadic by nature. This has to be taken into account when classifying sleep–waking states. It is quite normal that, when SSs occur, more than one episode appears in each epoch. SS presence can be much sparser though; even occurrences of single SS episodes up to 5 min apart are enough to assert their presence (Guilleminault and Souquet, 1979). Accordingly, a 5 min SS memory was applied. However, as even sparse FP SS detection would considerably degrade classification performance, additional SS validation criteria were developed based on training data and expert advice

(i) If a widely isolated SS occurred, i.e. there was no other SS during 5 min before or after it, the widely isolated SS was discarded as noise.
(ii) If a single SS was detected in a group of consecutive epochs with repeated and consistent REM_{ov}, then the SS memory was not applied, which meant that, unless there was SS presence in successive epochs, SS detection was discarded.
(iii) Similarly, if a single SS was detected in a group of consecutive epochs with confirmed SD predominance (consistent with NREM-III and IV), then the SS memory was not applied.

2.4 *Rapid eye movements (REM$_{ov}$)*

The EOG shows the time-sequence of potential differences between the front and back of the ocular globe. Eye movements correspondingly show a change in potential. REM$_{ov}$ appear as isolated or in-burst pulses in the EOG.

In the proposed method, the REM$_{ov}$ search began by identifying three consecutive peaks determined by three consecutive sign changes in the instantaneous slope of the signal, as shown in Fig. 1*b*. The peaks were labelled as left (L), centre (C) and right (R), and identified by (amplitude, time) pairs: (A_L, t_L), (A_C, t_C), (A_R, t_R). Once these points had been established, three values were determined in this REM$_{ov}$ candidate: amplitude A_{REM}, duration t_{REM} and the RMS factor RMS_{REM}. The first two values are defined as

$$A_{REM} = \min\{|A_L - A_C|, |A_C - A_R|\}$$

$$t_{REM} = t_R - t_L \text{ [s]}$$

where || stands for absolute value.

RMS_{REM} is a value to discriminate distinctive peak pulses, as opposed to pulse trains (as in SS). The RMS power contained in the REM$_{ov}$ pulse candidate C was compared with the RMS power contained in the neighbouring signal, considering 1.5 s wide windows immediately before (L) and after (R) the candidate

$$RMS_{REM} = \frac{2^* RMS_c}{1 + RMS_L + RMS_R}$$

where a constant 1 was added in the denominator to avoid near-zero divisions.

Fuzzy variables for REM$_{ov}$ evaluation, to describe relative amplitude, power content and duration, were defined as FA, FRMS and Ft, respectively, as shown in Fig. 2. REM$_{ov}$ characterise the REM sleep state, although REM$_{ov}$ also appear during WA. Analysis of training data showed distinctive differences in REM$_{ov}$ patterns between REM and WA. During REM sleep, REM$_{ov}$ patterns tended to have medium A_{REM} and high RMS_{REM} values, whereas during WA, A_{REM} was high and RMS_{REM} was relatively low.

This fuzzy differentiation allows better noise rejection, e.g. a high A_{REM} during a state other than WA can be discarded as noise. The duration criterion is the same in REM and WA. The meanings for 'high' (h), 'medium' (m) and 'low' (l) for each of the three values were defined as fuzzy numbers, as shown in Fig. 2. The parameters of the fuzzy numbers were fixed based on the training data set.

Applying the fuzzy concepts on the values for amplitude, duration and RMS factor, each REM$_{ov}$ pulse candidate obtained two quality indices Q, related to the expected values in REM sleep and WA, respectively,

$$REM_{ov}Q_{REM} = FA_m(A_{REM})*FRMS_h(RMS_{REM})* Ft_m(t_{REM})$$

$$REM_{ov}Q_{WA} = FA_h(A_{REM})*FRMS_l(RMS_{REM})* Ft_m(t_{REM})$$

As a single occurrence of an REM$_{ov}$ episode is enough to define its presence, the highest $REM_{ov}\,Q_{REM}$ and $REM_{ov}\,Q_{WA}$ indices in each epoch were recorded. In the normalised scale used, zero means absent and 1 means 100% (total assurance) presence.

For the purpose of classifying sleep–waking states, it must be taken into account that REM$_{ov}$ are sporadic by nature. Usually, more than a single episode occurs in a time window when they appear. However, occurrence of a single REM$_{ov}$ episode in 1 min asserts its presence for REM state classification purposes (Curzi-Dascalova and Mirmiran, 1996; Crowell *et al.*, 1997). REM$_{ov}$ can be even sparser, especially at the onset or at the end of an REM state. A 1 min memory was applied accordingly.

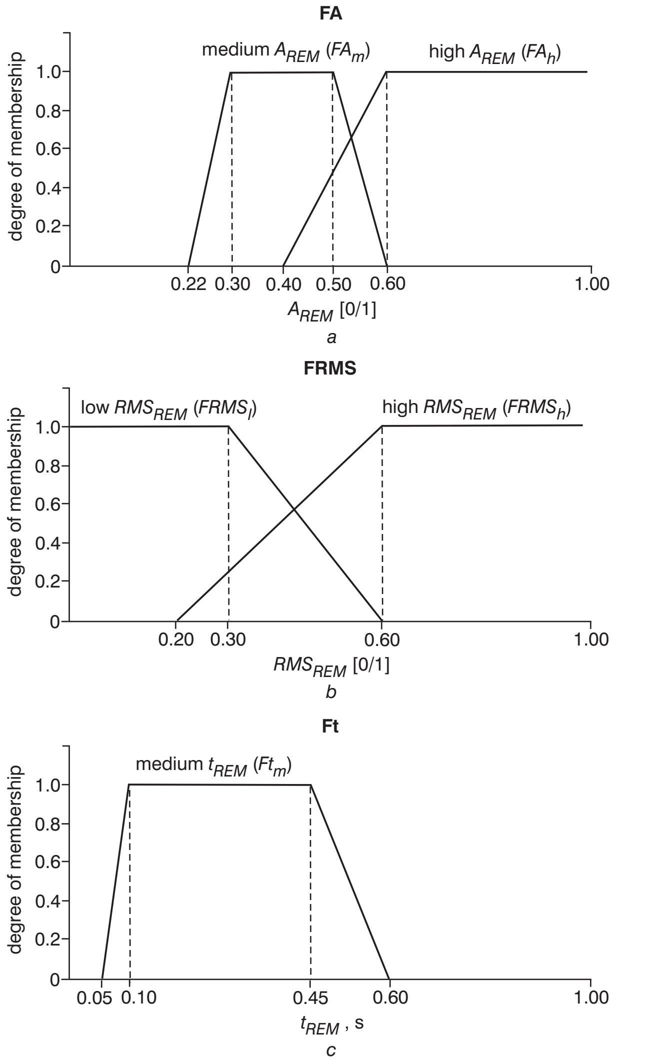

Fig. 2 *Fuzzy quality variables defined for REM$_{ov}$ pattern detection: (a) amplitude; (b) relative power; (c) event duration*

2.5 *Muscle tone (MT)*

The EMG recorded the time-sequence of potential differences of electrodes on the chin, reflecting the chin muscle tone. MT detection was performed on a moving window applied to the EMG channel. Once the moving window was stationed on a piece of signal trace, the average value $\bar{x}$ of the EMG signal in the window was calculated. Then, the RMS power was obtained as

$$RMS_{MT} = \sqrt{\frac{\sum_{x_i \subset W} (x_i - \overline{x})^2}{card(W)}}$$

where W is the moving window, $card(W)$ is the number of samples in the window, and x_i are the EMG signal samples. The initial window size was set to 4 s (1000 samples), and the moving step was set to 2 s. Tests on training data showed that the best classification results were obtained using a wider window and step during high MT periods, and a narrower window and step during low MT periods. Therefore the instant window size W_t and moving step S_t at time t were determined dynamically, using a simple linear proportionality rule on the variation of the calculated RMS_{MT}

$$W_{t+1} = \frac{(RMS_{MT})_t}{(RMS_{MT})_{t-1}} W_t \quad 20 < W_t < 2500$$

(i.e. not more than a half-epoch),

$$S_{t+1} = \frac{(RMS_{MT})_t}{(RMS_{MT})_{t-1}} S_t \quad 10 < S_t < 1250$$

(i.e. not more than a quarter of an epoch).

An empirically determined, recording-specific algorithm set the threshold on RMS_{MT}, indicating instant presence of MT. The algorithm reviewed the complete EMG recording to be evaluated and established the dynamical limits of RMS_{MT} variation, finding a maximum $MAX(RMS_{MT})$ and a minimum $MIN(RMS_{MT})$ value. The threshold was set at $THRES(RMS_{MT}) = MIN(RMS_{MT}) + 0.2\,(MAX(RMS_{MT}) - MIN(RMS_{MT}))$. The output of the MT algorithm was the relative MT presence in the epoch, expressed as a value between 0 and 1.

2.6 *Limb movement (LM)*

The RMS power of the recorded signals of movements at the upper and lower limbs (obtained from the two respective recording channels) were obtained using a fixed moving window of 1000 samples (4 s)

$$RMS_{LM} = \sqrt{\frac{\sum_{x_i \subset W} (x_i - \bar{x})^2}{card(W)}}$$

where x_i are the LM signal samples, and $card(W) = 1000$. LM needed to appear in just one channel to assert its presence. Also, LM signals are usually weak, and therefore the measured RMS value was boosted using an exponential function. The algorithm output was defined as a real function on RMS_{LM}, as

$$FRMS_{LM} = \exp(max(RMS_{ULM}, RMS_{LLM}))$$

where ULM stands for upper-limb movement, and LLM stands for lower-limb movement.

The threshold to determine the presence of LM was calculated for each recording. First, a minimum non-zero $MIN(FRMS_{LM})$ value was established. Then, the threshold was set at $20 * MIN(FRMS_{LM})$, a value that was empirically determined using training data.

2.7 *Wakefulness detection (WAD) algorithm*

WA is characterised by MT, REM_{ov} and LM. Each of these signals, and particularly LM, can frequently add noise and artifacts to the polysomnography. To reduce this negative effect on pattern detection, a special algorithm to detect WA was developed.

If the WAD algorithm detects WA, it triggers other algorithms that revise the detection of specific patterns and confirm or discard these as artifacts. If no WA is detected, no revision is performed. Therefore the WAD algorithm was not aimed at optimising WA detection, which was left for a later stage in the process, but at avoiding FPs of SD and SS during WA, and to diminish WA misclassification if an electrode became loose.

The WAD index calculation considered LM, REM_{ov}, MT and TH, as the inclusion of TH was shown to be useful. Context information was considered in two aspects: LM detection on neighbouring epochs was used to assert WA, and isolated WA or non-WA detection was discarded as ‘noise’ by the WAD algorithm, i.e. a single epoch was not enough to determine WA or non-WA.

2.8 *System evaluation*

The algorithms described in Sections 2.3–2.7 determined a figure of merit in the [0, 1] interval for each of five patterns: SD, TH, SS, REM_{ov} and MT. Once the pattern detection stage was finished, each epoch was independently analysed. Thresholds to define which signals corresponded to the sought activity patterns were adjusted by tuning the detection algorithms with training data.

The performance of the automated system was measured in an indirect way by evaluation of sleep staging, using the expert criteria shown in Table 1. It is an indirect evaluation, because it does not evaluate event detection against the expert, but the effect of those detections on sleep staging. This follows the final objective of this research work, which did not focus on the detection of every single episode in the EEG, EOG and EMG, which is very difficult because of noise and other types of signal corruption, but aimed to classify the sleep–waking states and stages at the level of a human expert.

As each state should last at least 1 min, in cases where there were less than three consecutive epochs with the same state, the system included rules to determine the prevailing state, following expert criteria. Hence, the final output showed only states or stages lasting at least three epochs, except at transitions. When the system detected a transition between two states/stages, it considered correct either of these two (within a window of seven epochs), as alternating epochs are normal occurrences at transitions.

When no definitive benchmark is available, such as in sleep medicine, the validation of automated systems is made against human experts. Some degree of error is inevitable when an observer’s judgment is used to determine whether the trait is present or not. The use of the kappa statistic (Cook, 1998; Fleiss, 1981) to assess observer variability is typical of studies in which there is no ‘gold standard’ (Goldman, 1992). The kappa statistic is frequently used to measure inter-rater agreement between human experts and also between automated medical systems and human experts (Martin-Baranera *et al.*, 2000; Ohayon *et al.*, 1999). The kappa index can be interpreted as a chance-corrected agreement. In this work, validation results are presented using inter-rater agreement and the kappa statistic.

3 Algorithm tuning and results on training data

Six polysomnographic recordings involving five different patients were scored by the expert, who identified the characteristic patterns and associated each page with a sleep state or stage accordingly. The procedure was for the expert to review each paper recording and put pencil markings on each page. These markings defined the current sleep–waking state or stage for each page, but, in several instances, also defined the predominant EEG background activity and several specific episodes, such as REM_{ov} and SS, LM, etc. The recordings were double-checked by the expert to avoid errors during the manual scoring of each epoch.

The training data were the digital recordings, associated with the available paper records, and the expert markings of the corresponding sleep–waking states, which were fed into the database. The sleep classification by page, carried out by the expert, was used as target.

Our approach was to mimic the expert’s procedure for sleep scoring. This led to an iterative process of algorithm tuning with training data, adding new criteria to the algorithms when the automated system showed recurring discrepancies. Several iterations trying different methods and parameters were performed on the training data, which led to the algorithms described in Section 2. The goal was to detect the different patterns, while rejecting noise-induced false negative (FN) and FP detections.

As an example, early results on recording FH120594 showed several consecutive FP outputs for the SD detection algorithm. Expert analysis of these pages determined that it was EEG signal corruption originating in the ventilation signal that made it resemble SD. Improvements, such as smoothing the output of the SD detection algorithm and enhancing threshold values, eliminated most of these errors.

To avoid overfitting to a single class or recording over the others, the algorithm parameters were adjusted to reach at least an 80% agreement for every sleep category on each recording in the training set.

Table 3 shows a contingency table obtained with the training data, once all the algorithms had been tuned. The rows of the Table show the classification of the automated system, and the columns show the expert classification. These classifications consider a minimum state duration of 1 min. Each unit represents an epoch that was classified in one of the sleep–waking categories. A total of 2665 epochs were included in the training set. Figures in the diagonal of Table 3 represent the agreement cases. Summing over these agreements and dividing by the total number of epochs gives the overall agreement, also called observed agreement. The overall agreement between the automated system and the human expert was 0.915 or 91.5%. Less than 0.4% of the total number of epochs were classified as IS by the expert; this was insufficient to attempt classification of this pseudo-state, but they were included in the results for completeness.

4 Results on test recordings

Once the system had been tuned with the training data, an independent test was performed using five polysomnographic recordings involving four different subjects. Table 4 shows the contingency table for the classifications carried out by the automated system and the human expert, on test data including a total of 2369 epochs.

Table 5 Observed agreement and kappa index obtained for each test recording

Recording identification	Observed agreement	Chance agreement	Kappa index	Standard error
AM102793	0.907	0.268	0.873	0.026
BH072093	0.902	0.263	0.867	0.030
KD032493	0.828	0.350	0.736	0.027
SS051893	0.888	0.317	0.836	0.027
SS101893	0.870	0.319	0.809	0.024
Overall	0.877	0.231	0.840	0.011

Table 5 shows the observed agreement, chance agreement, kappa index κ and standard error of kappa, obtained for each test recording. Chance agreement measures the degree of agreement to be expected by chance alone. If the observed agreement is greater than the chance agreement, then $\kappa > 0$; if the observed agreement is explained by pure chance, then $\kappa = 0$. According to FLEISS (1981), $\kappa > 0.75$ represents excellent agreement beyond chance. In Table 5, all recordings show $\kappa > 0.75$, except recording KD032493, which has $\kappa = 0.736$, and therefore show a good level of agreement.

The overall observed agreement between the automated system and the human expert, on the test data, was 0.877 ($\kappa = 0.840$). The ratio between κ and its standard error was used to test the hypothesis that the underlying value of kappa was equal to zero (FLEISS, 1981). The hypothesis was rejected at the level of significance 0.01 for each kappa index in Table 5, indicating a good degree of agreement beyond chance.

To measure the degree of agreement on each sleep state or stage, the original contingency Table 4 was collapsed into 2×2 tables in which all categories other than the one of current interest were combined into a single 'all others' category (FLEISS, 1981). Table 6 presents the results obtained for each individual sleep state or stage. All categories show $\kappa > 0.75$, except NREM-I and IS. The kappa index for NREM-I (0.690)

Table 3 Contingency table for classification of sleep–waking states and stages by automated system and human expert, obtained with training data

System \ Expert	WA	NREM-I	NREM-II	NREM-III and IV	REM	IS	Total
WA	560	2	0	6	6	0	574
NREM-I	19	393	41	13	12	3	481
NREM-II	6	10	665	49	0	0	730
NREM-III and IV	0	7	6	567	2	0	582
REM	6	6	7	0	254	5	278
IS	8	2	0	6	4	0	20
Total	599	420	719	641	278	8	2665

Table 4 Contingency table for classification of sleep–waking states and stages by automated system and human expert, obtained with test data

System \ Expert	WA	NREM-I	NREM-II	NREM-III and IV	REM	IS	Total
WA	614	4	0	0	1	0	619
NREM-I	0	174	51	26	27	1	279
NREM-II	0	12	486	15	0	0	513
NREM-III and IV	0	14	85	607	2	2	710
REM	18	0	10	0	192	0	220
IS	0	0	13	10	0	5	28
Total	632	204	645	658	222	8	2369

Table 6 *Observed agreement and kappa index obtained for each sleep-waking state or stage with test data*

State/stage	Observed agreement	Chance agreement	Kappa index	Standard error
WA	0.990	0.611	0.975	0.021
NREM-I	0.943	0.816	0.690	0.020
NREM-II	0.921	0.629	0.788	0.020
NREM-III and IV	0.935	0.589	0.842	0.021
REM	0.976	0.831	0.855	0.021
IS	0.989	0.985	0.274	0.017

shows a good agreement. Although IS classification was not targeted in this work, it has been included for completeness. As the IS category is relatively rare, the observed agreement on its absence is large. However, κ, which corrects chance, shows a poor agreement for IS, as expected.

5 Discussion and conclusions

The overall agreement between the automated system and the human expert using the test data was 87.7% ($\kappa = 0.840$). According to the results shown in Table 6, the agreement on the WA state was the highest ($\kappa = 0.975$), followed by REM sleep ($\kappa = 0.855$) and NREM III and IV ($\kappa = 0.842$). The automated system showed a tendency to overscore NREM-I ($\kappa = 0.690$) and to underscore NREM II ($\kappa = 0.788$), as can be seen in Table 4. These disagreements can be traced to FN events of REM_{ov}, FN events of SS and FP events of SD. Although the SD detection algorithm showed a tendency to FP detection with test data, this was not the case with training data, where, as a whole, more FN than FP detections occurred.

A special WA detection algorithm was developed (WAD, Section 2.7) to discard FPs of SD and SS during WA. Other artifact rejection strategies were also developed, such as the LM detection algorithm (Section 2.6). To account for inter-patient variations, several algorithm thresholds were automatically adjusted to each particular recording data set, instead of fixed values being used. The system made a first reading of the whole recording to establish threshold values and applied the algorithms on a second pass.

To compare the automated system with visual paper scoring, White and Gibb (1998) suggested using the expert intra-rater agreement as a standard. They reported an 88.0% agreement in paper polysomnography score-rescore results (i.e. agreement rate of expert scoring of the same recording at two different instances).

In a previous work (Holzmann *et al.*, 1999), we independently reported an 88.2% expert classification agreement in score-rescore, which is consistent with White and Gibb (1998). These intra-rater performances are quite similar to the current system performance, which reaches an overall agreement of 87.7% with test data. According to it, disagreements between automated system and paper visual scoring could be related to score-rescore differences rather than errors of the system.

This work aimed at developing a more robust pattern detection system for sleep staging, compared with the one available in Holzmann *et al.* (1999). No data were eliminated in the present work, making the system closer to a truly automatic sleep–waking state classifier. The number of data employed to train and test the system was increased by a factor of five to a total of 5034 epochs.

In Holzmann *et al.* (1999), the results showed significant improvements in the classification agreement between the system and the expert owing to the application of an expert system. When criteria similar to Table 1 were simply applied, only a 50.3% agreement was obtained in the test set, whereas that figure rose to 84.9% with the expert system. Later on, using modular neural networks on the same test data, the results were further enhanced to 94.3% (Estévez *et al.*, 2000). Both results suggest that the criteria presented in Table 1 are too simplistic to explain all the details of expert criteria, and therefore the current system performance could be enhanced using neural networks or fuzzy expert systems.

A related issue is that the current classification standard was developed for manual scoring of paper recordings more than 30 years ago (Rechtschaffen and Kales, 1968). Nowadays, no recognised standard exists for computerised polysomnography, which usually follows a different procedure from that of the expert, trying to achieve similar results. Recent efforts such as the SIESTA project (Klösch *et al.*, 2001) aimed to define a new standard, including the creation of a reference database for adults. Many different algorithms exist for pattern detection and artifact processing, based on modern and conventional techniques. Our system could serve as a reference for performance comparison of automated sleep staging in infants.

Acknowledgments—This work has been partially supported by the Department of Electrical Engineering and INTA, Universidad de Chile, and by grants from CONICYT-Chile (FONDECYT 1980909 and 1000657) and NIH-USA (HD 33487).

References

Anders, T., Emde, R., and Parmelee, A. (Eds) (1971): 'A manual of standardized terminology, techniques and criteria for scoring of states of sleep and wakefulness in newborn infants' (UCLA Brain Information Service, NINDS, Neurological Information Network, Los Angeles, USA, 1971)

Bankman, I. N., Sigillito, V. G., Wise, R. A., and Smith, P. L. (1992): 'Feature-based detection of the K-complex wave in the human electroencephalogram using neural networks', *IEEE Trans. Biomed. Eng.*, **39**, pp. 1305–1310

Besset, A. (1998): 'L'analyse automatique du sommeil', in Billiard, M. (Ed.): 'Le sommeil normal et pathologique, 2nd edn' (Masson, Paris, France, 1998), pp. 126–133

Collura, T. F., Jacobs, E. C., Braun, D. S., and Burgess, R. C. (1993): 'EVIEW—A workstation-based viewer for intensive clinical electroencephalography', *IEEE Trans. Biomed. Eng.*, **40**, pp. 736–744

Cook, R. J. (1998): 'Kappa', in Armitage, P., and Colton, T. (Eds): 'The encyclopedia of biostatistics' (Wiley, New York, USA, 1998) pp. 2160–2166

Crowell, D., Brooks, L. J., Colton, T., Corwin, M. J., Hoppenbrouwers, T., Hunt, C. E., Kapuniai, L. E., Lister, G., Neuman, M. R., Peucker, M., Davidson Ward, S. L., Weese-Mayer, D. E., Willinger, M., and Collaborative Home Infant Monitoring Evaluation (CHIME) Steering Committee (1997): 'Infant polysomnography: reliability', *Sleep*, **20**, pp. 553–560

Curzi-Dascalova, L., and Mirmiran, M. (1996): 'Manual of methods of recording and analyzing sleep-wakefulness states in premature and full-term infants' (INSERM Publ., Paris, France, 1996)

Estévez, P. A., Fernández, M. E., Held, C. M., Holzmann, C. A., Perez, C. A., and Pérez, J. P. (2000): 'Classification of sleep-waking states using modular neural networks'. IEEE International Conference on Systems, Man and Cybernetics, Nashville, Tennessee, USA, 2580–2585

Ferber, R., and Kryger, M. (1995): 'Principles and practice of sleep medicine in the child' (Saunders, Philadelphia, USA, 1995)

Fleiss, J. (1981): 'Statistical methods for rates and proportions, 2nd edn' (John Wiley, New York, USA, 1981)

Goldman, L. R. (1992): 'The kappa statistic – in reply', *JAMA*, **268**, pp. 2513–2514

Guilleminault, C., and Souquet, M. (1979): 'Sleep states and related pathology' in Korobkin, R., and Guilleminault, C. (Eds): 'Advances in perinatal neurology, Vol. 1', (Spectrum, New York, USA, 1979)

GUILLEMINAULT, C. (1998): 'Le sommeil normal de l'homme', in BILLIARD, M. (Ed).: 'Le sommeil normal et pathologique, 2nd edn' (Masson, Paris, France, 1998) pp. 3–11

HARPER, R. M., SCHECHTMAN, V. L., and KLUGE, K. A. (1987): 'Machine classification of infant sleep state using cardio-respiratory measures', *Electroenceph. Clin. Neurophysiol.*, **67**, pp. 379–387

HOLZMANN, C. A., PEREZ, C. A., HELD, C. M., SAN MARTÍN, M., PIZARRO, F., PÉREZ, J. P., GARRIDO, M., and PEIRANO, P. (1999): 'Expert-system classification of sleep-waking states in infants', *Med. Biol. Eng. Comput.*, **37**, pp. 466–476

HUUPPONEN, E., VÄRRI, A., HIMANEN, S. L., HASAN, J., LEHTOKANGAS, M., and SAARINEN, J. (2000): 'Optimization of sigma amplitude threshold in sleep spindle detection', *J. Sleep Res.*, **9**, pp. 327–334

JANKEL, W. R., and NIEDERMEYER, E. (1985): 'Sleep spindles', *J. Clin. Neurophysiol.*, **2**, pp. 1–35

JANSEN, B. H., and DAWANT, B. M. (1989): 'Knowledge-based approach to sleep EEG analysis – A feasibility study', *IEEE Trans. Biomed. Eng.*, **36**, pp. 510–518

KLÖSCH, G., KEMP, B., PENZEL, T., SCHLÖGL, A., RAPPELSBERGER, P., TRENKER, E., GRUBER, G., ZEITLHOFER, J., SALETU, B., HERRMANN, W. M., HIMANEN, S. L., KUNZ, D., BARBANOJ, M. J., RÖSCHKE, J., VÄRRI, A., and DORFFNER, G. (2001): 'The SIESTA project polygraphic and clinical database' *IEEE Eng. Med. Biol. Mag.*, **20**, pp. 51–57

KUBICKI, ST., HÖLLER, L., BERG, I., PASTELAK-PRICE, C., and DOROW, R. (1989): 'Sleep EEG evaluation: A comparison of results by visual scoring and automatic analysis with the Oxford sleep stager', *Sleep*, **12**, pp. 140–149

MARTIN-BARANERA, M., SANCHO, J. J., and SANZ, F. (2000): 'Controlling for chance agreement in the validation of medical expert systems with no gold standard: PNEUMON-IA and RENOIR revisited', *Comput. Biomed. Res.*, **33**, pp. 380–397

OHAYON, M. M., GUILLEMINAULT, C., ZULLEY, J., PALOMBINI, L., and RAAB, H. (1999): 'Validation of the sleep-EVAL system against clinical assessments of sleep disorders and polysomnographic data', *Sleep*, **22**, pp. 925–930

PARK, S., PRINCIPE, J. C., SMITH, J. R., and REID, S. A. (1990): 'TDAT - Time domain analysis tool for EEG analysis', *IEEE Trans. Biomed. Eng.*, **37**, pp. 803–811

PARK, H. J., OH, J. S., JEONG, D. U., and PARK, K. S. (2000): 'Automated sleep stage scoring using hybrid rule- and case-based reasoning', *Comput. Biomed. Res.*, **33**, pp. 330–349

PEIRANO, P., CURZI-DASCALOVA, L., and KORN, G. (1986): 'Influence of sleep state and age on body motility in normal premature and full-term neonates', *Neuropediatrics*, **17**, pp. 186–190

PEIRANO, P., FAGIOLI, I., SINGH, B. B., and SALZARULO, P. (1989): 'Effect of early human malnutrition on waking and sleep organization', *Early Hum. Dev.*, **20**, pp. 67–76

PEIRANO, P., and CURZI-DASCALOVA, L. (1995): 'Motor activity patterns modulation and sleep states in low-risk prematurely born infants reaching normal term: A comparison with full-term newborns', *Neuropediatrics*, **26**, pp. 8–13

PRINCIPE, J. C., GALA, S. K., and CHANG, T. G. (1989): 'Sleep staging automaton based on the theory of evidence', *IEEE Trans. Biomed. Eng.*, **36**, pp. 503–509

PROAKIS, J. G., and MANOLAKIS, D. G. (1996): 'Digital signal processing: principles, algorithms and applications, 3rd edn' (Prentice Hall, New Jersey, USA, 1996)

RECHTSCHAFFEN, A., and KALES, A. (1968): 'A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects' (UCLA, Brain Research Institute/ Brain Information Service, Los Angeles, USA, 1968)

ROBERTS, S., and TARASSENKO, L. (1992): 'Analysis of the sleep EEG using a multilayer network with spatial organization', *IEE Proc. F*, **139**, pp. 420–425

SCHER, M. S. (1997): 'Computer classification of state in healthy preterm neonates', *Sleep*, **20**, pp. 132–141

SHIMADA, T., SHIINA, T., and SAITO, Y. (2000): 'Detection of characteristic waves of sleep EEG by neural network analysis', *IEEE Trans. Biomed. Eng.*, **47**, pp. 369–378

SMITH, J. R. (1986): 'Automated analysis of sleep EEG data', in LOPES DE SILVA, F. H., STORM VAN LEEUWEN, W., and RÉMOND, A. (Eds): 'Handbook of electroencephalography and clinical neurophysiology, Vol. 2' (Elsevier Publ., 1986) pp. 131–147

TAFTI, M. (1998): 'Analyse numèrique du sommeil', in BILLIARD, M. (Ed.): 'Le sommeil normal et pathologique, 2nd edn' (Masson, Paris, France, 1998)

WHITE, D. P., and GIBB, T. J. (1998): 'Evaluation of a computarized polysomnographic system', *Sleep*, **21**, pp. 188–196

Authors' biographies

PABLO A. ESTÉVEZ received his degree in electrical civil engineering from Universidad de Chile and PhD from the University of Tokyo, Japan, in 1995. He is an Assistant Professor in the Electrical Engineering Department, Universidad de Chile. His research interests include the application of neural networks and genetic algorithms to pattern recognition and biomedical issues.

CLAUDIO M. HELD received his degree in electrical civil engineering from Universidad de Chile and PhD from the Rensselaer Polytechnic Institute, USA, in 1995. He is an Assistant Professor in the Electrical Engineering Department, Universidad de Chile. His research interests include the applications of fuzzy logic and the development of expert systems applied to medical as well as other issues.

CARLOS A. HOLZMANN, B.S.E.E. (U.C. Valparaíso, Chile), PhD (Notre Dame, Indiana, USA, 1966) is a Professor at the Electrical Engineering Department, Universidad de Chile. He is the founder of the Biomedical Engineering Research Group and the Graduate Program in Biomedical Engineering.

CLAUDIO A. PÉREZ received his degree in electrical civil engineering from Universidad de Chile and PhD from Ohio State University, USA, in 1991. He is an Associate Professor in the Electrical Engineering Department, Universidad de Chile. His research interests include the development of new technology for the sensory impaired and computational vision models.

JUAN PABLO PÉREZ, is pursuing an electrical civil engineering degree at Universidad de Chile. He is a Research Assistant in the Biomedical Engineering Laboratory.

JAIME HEISS is pursuing his MSc in Biomedical Engineering at Universidad de Chile. He is a Research Assistant in the Biomedical Engineering Laboratory.

MARCELO GARRIDO received his MSc from the Universidad de Chile and is an Associate Researcher at the Sleep Laboratory, INTA. His current research includes the study of motor activity and heart rate patterns during the sleep-wake cycle in early human development.

PATRICIO PEIRANO received his MD from Universidad de Chile and PhD from the University of Paris in 1989. He returned to Chile after ten years of research activity in France. He is currently Associate Professor and Head of the Sleep Laboratory at INTA, Universidad de Chile.