A Method for the Automatic Detection of Arousals During Sleep

*Fabrizio De Carli, †Lino Nobili, †Paola Gelcich, †Franco Ferrillo

**Center for Cerebral Neurophysiology, National Research Council, Genoa, Italy. †Sleep Disorder Center, DISM, University of Genoa, Italy.*

Summary: A method for the automatic detection of arousals in digital polysomnographic recordings is described. The computer program analyzed two EEG and one EMG derivations marking variable length segments as arousals. The processing of EEG data started from the wavelet transform, which characterizes the signal in the time-frequency domain, and resulted in a set of indices used to discriminate possible arousal segments. Transient increases in muscle activity were also identified, while a multichannel and context sensitive analysis allowed arousal detection. Out of 11 overnight recordings, 3 were used as the training set and 8 as the program testing set. In the first stage of the study two EEG experts inspected the tracings independently to score arousals. They then reviewed all recordings and jointly examined each event for validation, both those scored by themselves and those scored by the computer. A reference set of definite arousals (1125 in the testing set) and a number of uncertain events (266) were thus obtained. The sensitivity of the automatic system (88.1 %) was higher than that of the human experts (72.4 and 78.4 %) while the selectivity was lower (74.5 % for the automatic system, 83.0 and 82.0 % for the experts). This suggested that automatic detection, followed by an expert's validation, may render the analysis of arousals more widely feasible as well as support the study of arousal features.

Key words: Arousal; wavelet analysis; automatic sleep analysis; sleep microstructure; phasic events

INTRODUCTION

TRANSIENT AROUSALS OFTEN LAST ONLY A FEW SECONDS and do not generally cause awakenings or sleep stage shifts; rather, they belong to a group of microstructural patterns characterizing non-stationary EEG segments that, according to conventional scoring, are commonly found either in stage 2 or in REM sleep.[1,2] A certain number of spontaneous arousals seem to be an intrinsic component of physiological sleep.[3,4] Arousals may be induced either by exogenous or by endogenous stimuli and can be associated with unstable sleep conditions. In respiratory sleep disorders and nocturnal myoclonus the arousing stimulus may be identified in snoring, apnea or leg movement and results in sleep fragmentation. In severe clinical conditions arousals become increasingly frequent and are associated with daytime sleepiness.[5,6] A number of experimental studies, in which arousals were purposely induced by the presentation of tones, confirmed that sleep fragmentation raises daytime sleepiness;[7,8,9] others reported that sleep fragmentation may also influence the impairment of cognitive functions.[10]

Standard staging rules[11] do not highlight sleep fragmentation caused by such transient changes in arousal level; therefore, an alternative approach is required when taking such microstructural features into account. A set of guidelines for arousal scoring has been proposed by the American Academy of Sleep Medicine.[12] However, the visual inspection of the microstructure of sleep is cumbersome and time consuming; in this respect the support of automatic or semi-automatic pattern recognition procedures could foster the extensive application of arousal analysis.

A computerized method for the detection of episodes of wakefulness during sleep based on the alpha slow-wave index has been proposed;[13] however, the work in question still considered 30-second epochs with accordance to conventional sleep staging. An adaptive segmentation approach in which the signal is "continuously" analyzed and the boundaries between segments are set when significant changes are detected,[14,15] appears to be more suitable for the analysis of transient patterns, such as arousals.

A preliminary study was performed[16] to select a set of indices marking EEG changes accompanying arousal. This study suggested that parameters related to the EEG frequency were the most suitable for arousal detection. Another approach was proposed[17] in which the relative power of EEG was computed for five bands and acted as the input within a multilayer neural network.

The main aim of this study was the development of an automatic procedure for arousal detection according to

Accepted for publication February 1999
Corresponding author: Dr. Fabrizio De Carli, Centro di Studio per la Neurofisiologia Cerebrale, C.N.R., Via de Toni, 5, 16132 - Genova, Italy, Telephone: +39.10.3537466, Fax: +39.10.354180, E-mail: fabrizio@dism.unige.it

AASM criteria and its testing on a group of subjects in different pathological conditions. The algorithm was tuned by means of a training set of polygraphic recordings on which two experts had previously marked the start and duration of each arousal, based on visual inspection. Another set of previously classified recordings was used to test the method and simultaneously verify the interscorer agreement.

MATERIALS AND METHODS

Subjects

In order to explore a wide range of arousal patterns, a set of recordings belonging to 11 subjects affected by different pathological conditions, was extracted from the database of the Center for Sleep Disorders (University of Genoa). Five subjects suffered from breathing disorders, one from nocturnal myoclonus, two were epileptic patients, two were affected by psychophysiologic insomnia and one was a narcoleptic patient. The age range was 19 - 67 years (mean: 45). More details about the subjects are reported in Table 1.

TABLE 1. Patient information for each recording in the training set (1-3) and in the testing set (4-11).

Recording	Sex	Age	Diagnosis
1	F	67	OSA (AI: 84)
2	M	63	Nocturnal myoclonus
3	F	19	Partial epilepsy
4	M	49	OSA (AI:55)
5	M	60	OSA (AI:54)
6	M	31	Partial epilepsy
7	M	39	Psychophysiological insomnia
8	M	63	Simple snoring
9	M	49	OSA (AI:14)
10	M	24	Narcolepsy
11	M	31	Psychophysiological insomnia

OSA: obstructive sleep apnea. AI: apnea index.

Data Recording

Each subject underwent an overnight recording in our sleep laboratory, where the data were collected by a digital polygraph (Galileo, ESAOTE). The EEG was acquired from at least four electrodes in common reference (F4, C4, T4, O4, according to the 10-20 international electrode placement system), the time constant was set at 0.1, the low-pass filter at 70 Hz and the notch filter was switched on. Two EOG channels (time constant: 0.3 sec, filter 30 Hz) and one submental EMG (0.01 sec, 70 Hz) were recorded in each case. Additional data were collected, whenever requested, to evaluate breathing disorders and leg movements, including oronasal flow, thoracic respiration and tibialis EMG. All signals were sampled with 512-Hz frequency and 12-bit resolution and stored with 128-Hz frequency and 8-bit size, after the application of an antialiasing digital filter and a procedure for the dynamic adjustment of a scale factor for the signal amplitude.

Each record was scored by an expert in accordance to conventional rules for sleep staging[11] and the hypnogram was filed, with a link to the recording.

Computer data analysis

The software program analyzed the data from two bipolar EEG channels (F4-C4 and C4-O2) and one EMG.

EEG analysis. The processing of the EEG data proceeded from the wavelet transform - a technique that enables the representation of a signal in the time frequency domain.[18,19] The algorithm for arousal detection performed the wavelet transform of thirty-two-second overlapping epochs, analyzing the time course of the signal in different frequency bands, with variable frequency and time resolution, as described in the Appendix (a). The program then processed the transformed data to evaluate the signal power, with a time resolution of 0.125 sec, for six frequency bands: 0.0-0.5 (slow delta), 0.5-4 (delta), 4-8 (theta), 8-12 (alpha), 12-16 (sigma), and 16-64 (beta).

The result of the time-frequency analysis was then used to evaluate a set of indices describing the EEG changes accompanying arousals, mainly consisting in an abrupt shift in EEG frequency and an increase in theta, alpha and/or beta activity. In order to detect these variations, a long-term weighed moving average of power was computed for each band to estimate a background reference value, while a short-term moving average stood for the actual trend. The first six indices were the ratios between short-term and long-term average, indicating the actual variation for each band; the other indices involved average power in different bands and evaluated particular features that may be important in arousal detection (see section b of the Appendix for details).

The selection of these indices was performed by using heuristic criteria in the early stage of the program development. After that, by using a training set of sleep recordings previously scored for arousals, a stepwise discriminant analysis was performed to verify usefulness of such indices in arousal detection. For this purpose the Stepdisc procedure of the SAS package was used.[20]

A linear discriminant function was then estimated and inserted in the algorithm. In analyzing an EEG channel for arousal detection, the program evaluated the discriminant function for each 0.125-sec epoch and a possible arousal was marked when it remained positive for more than three and up to thirty seconds (after which an awakening was detected). A score was assigned to the arousal resulting from the mean value of the discriminant function.

EMG analysis. The EMG signal was processed in parallel and filtered digitally to eliminate low frequency components. The time course of its power was estimated by

using the same weighed moving average operator as per the EEG case; the ratio between short-term and long-term average was utilized to detect transient increases in muscle activity, which were marked when lasting more than one and less than 30 seconds. A score was also evaluated as a function of the mean value and duration of the increase.

Multi-channel data integration. In the next stage the patterns marked in the three channels were compared: overlapping events were linked; a weighed sum of the scoring was performed and compared with a threshold to select probable arousals. These were then filed with their start, duration and score. The weights and threshold were tuned during the training phase, with the constraint that no arousals could be detected without EEG changes as the increase in EMG alone was not enough. Tuning was performed by running the program through a number of times, during which the threshold was arbitrarily fixed at one hundred; four values were tested for the weight of the EMG, whereas the weight of the EEG was slowly changed within a wide range. Sensitivity and selectivity were computed at each step, as described in the Appendix (c). As expected, the number of detected arousals increased together with the weight of the EEG or the EMG, thus generating both correct and incorrect detections. Sensitivity consequently increased while selectivity decreased. The analysis of the sensitivity-versus-selectivity pattern enabled the selection of suitable weights.

Arousal sequence revision. In the last step the program reviewed the sequence of arousals to delete some possibly incorrect detections:

- the events occurring during REM sleep without a transient increase in muscle activity: the data from visual sleep staging were used for this purpose;
- each event less than ten seconds distant from a previous one.

Visual inspection and validation

Two researchers with experience in the clinical analysis of polygraphic sleep tracings separately inspected every recording included in this study to score arousals. Each event was filed on the computer memory and linked to the tracing with its start and duration. According to AASM criteria, arousal scoring was independent of Rechtschaffen and Kales epoch scoring; consequently an arousal could be marked within an epoch classified as wake. For that reason the arousal index of each tracing was computed as events per hour of recording.

Three recordings involving different pathological conditions and arousal patterns were selected. These were then revised by the two scorers jointly to solve disagreements and set up the training set for the program, consenting to the tuning of the algorithm, the estimation of the coefficients for the linear discriminant function and the choice of weights and threshold for the arousal detection. The remaining eight recordings were then processed by the program, arousals were detected by the tuned algorithm and filed with a start, a duration, and a score associated to each.

At this stage, the agreement between the two experts was evaluated — for the testing set — as the rate between the number of arousals detected by both raters and the total number of arousals marked by either one. The agreement between each expert and the computer was estimated in the same way.

A further step was then performed with the aim of checking the disagreements and setting up a "gold standard" reference set. A computer program collected all the marked arousals from each recording into a single set that concealed the source of the detection — either computer or human — and merged overlapping events, (i.e., marked by different scorers), into one. When revising this unified event set the human experts, by mutual consent, marked each arousal as definite, possible (uncertain) or wrong. They also reviewed each recording entirely for arousals that had been overlooked by all during the first pass and marked them as definite or possible. This validated arousal set provided the reference evaluation to estimate the sensitivity and selectivity of human and computer scorings.

Global sensitivity and selectivity, evaluated for the program and the two experts, was compared by means of the chi-square test: the distribution between true positive and false negative, arranged into a 3 x 2 table, was analyzed to evaluate differences in sensitivity. The distribution between true and false positive was instead used to evaluate differences in selectivity.

The procedure adopted for the validation of arousals did not take the exact start and duration of each event into consideration; consequently, an exhaustive analysis of time relationship between overlapping events was not performed. Nevertheless, a preliminary analysis was carried out solely on events in the training set, as each arousal in this set had a definite start and duration. The difference between the computer and the experts' evaluation of the start and duration of each overlapping event was computed and the distributions of such differences are reported.

RESULTS

The training set included different arousal patterns, which are briefly illustrated with example results of the wavelet analysis.

One recording concerned a patient with severe breathing disorders (AI=84), whose very frequent obstructive apneas or hypopneas were mostly overcome by arousals. The arousals were characterized by the disappearance of delta waves, the increase in beta activities and, often, by a transient increase in muscle activity (see Fig. 1a,). These changes in the frequency composition of the EEG signal are reported in Fig. 1b. The second recording involved a

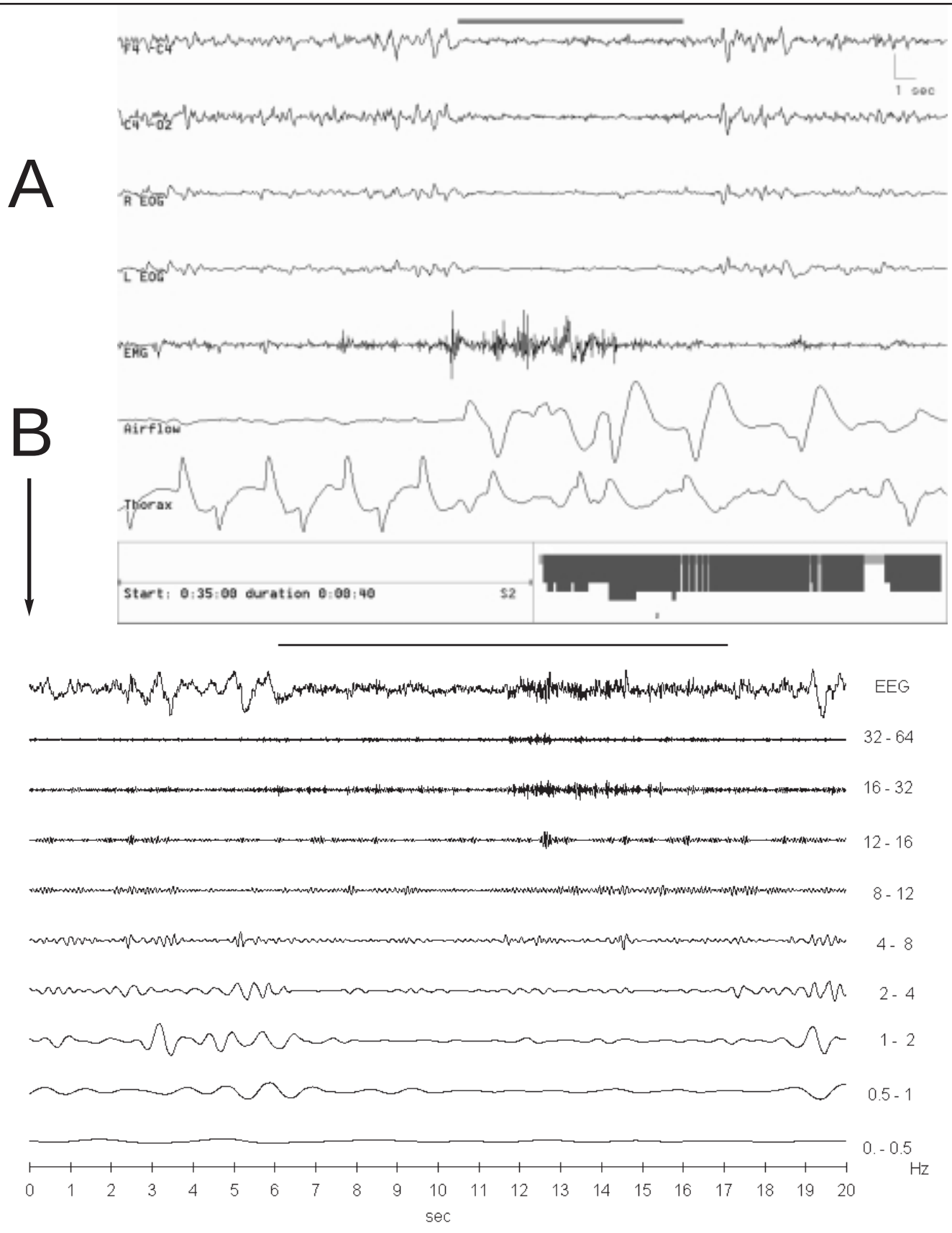

Figure 1 — Arousal in a patient with severe breathing disorders. a) 40-sec epoch of the polygraphic recording, including two EEG derivations, two EOG, one EMG, the airflow and thorax movement. The arousal, marked by the upper bar, ends an obstructive apnea and includes an increase in muscle activity. b) Wavelet decomposition of the first EEG derivation (F4-C4) presented in the form of the output of a filter bank. The first trace is the original signal: 20 seconds surrounding the arousal; the following lines present the signal band-pass filtered for each frequency band. The arousal, marked by the upper bar, is characterized by a frequency shift: slow waves disappear while fast activities increase, especially in 16-32-Hz band.

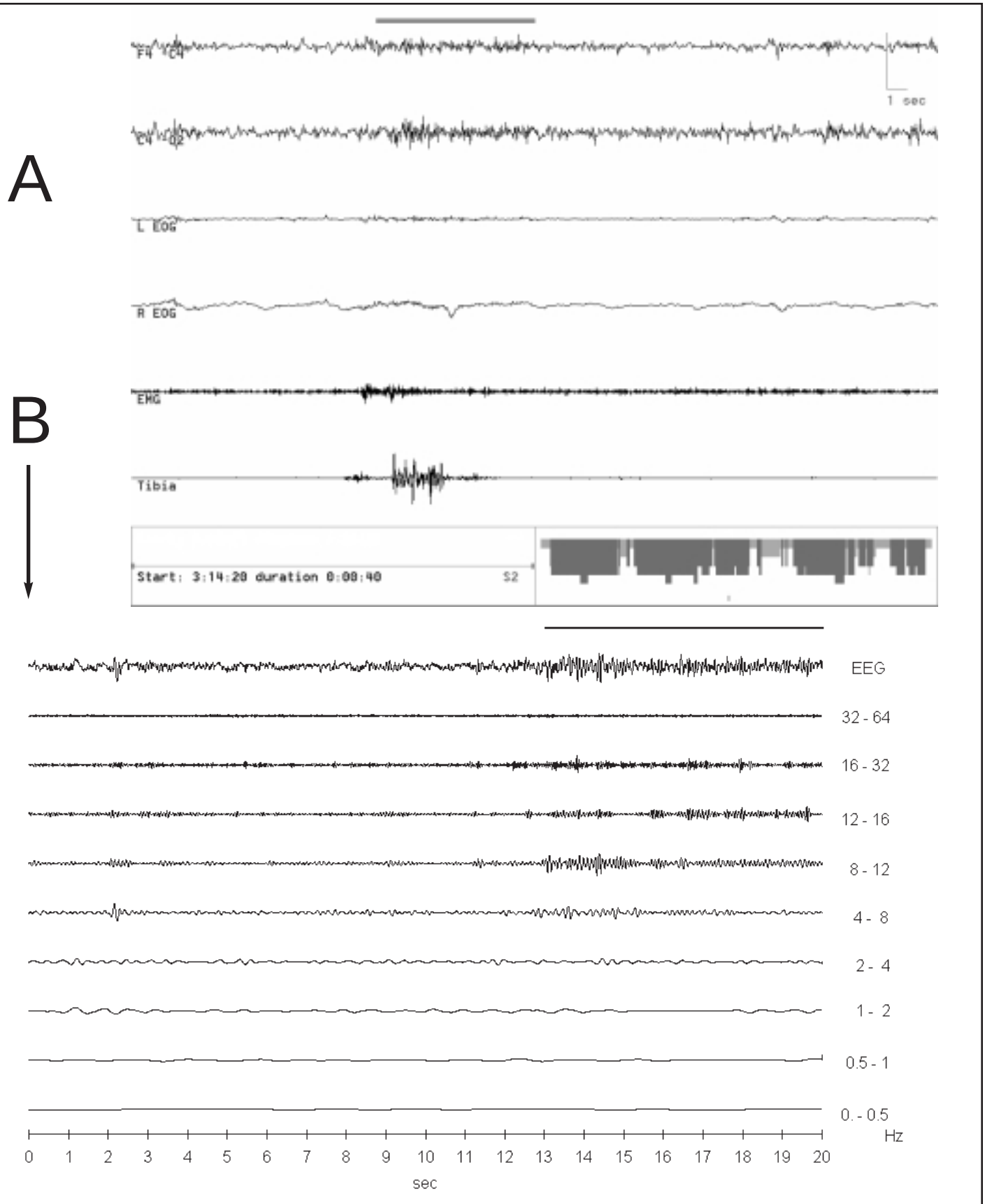

Figure 2 — Arousal in a patient with periodic leg movements a) 40-sec epoch of the polygraphic recording, including two EEG derivations, two EOG, one chin and one tibialis EMG. The arousal, marked by the upper bar, is associated to a leg movement. b) Wavelet decomposition of the first EEG derivation (F4-C4) presented in the form of the output of a filter bank. The first trace is the original signal: 20 seconds surrounding the arousal; the following lines present the signal band-pass filtered for each frequency band. The arousal, marked by the upper bar, is characterized by an increase in the theta, alpha and beta bands.

patient with rare hypopneas and a nocturnal myoclonus: the hypnogram showed poor slow-wave sleep and arousals were characterized by an increase of the alpha, theta and beta power, often accompanied by a transient increase in submental and/or tibialis muscle activity (see Fig. 2a and 2b). The third recording, from an epileptic patient, showed a regular sleep macrostructure. Arousals were characterized by a significant increase in alpha activity, occasionally accompanied by an increase in the EMG. The training set included two hours from the first recording, with 158 arousals (79 arousals per hour), six hours and forty minutes from the second recording, with 96 arousals (14.4 per hour) and eight hours from the third recording, with 89 arousals (11.1 per hour); the mean arousal length being 10.9 sec.

Table 2 — Distribution of the arousals in the training set and performance of the tuned computer program.

		Reference arousals		Computer detected			
Recording	Minutes	#	Index	#	Index	Sensitivity	Selectivity
1	120	158	79.00	160	80.00	97.47	96.25
2	400	96	14.40	100	15.00	94.79	91.00
3	480	89	11.13	97	12.13	93.26	85.57

The stepwise discriminant analysis indicated that all indices contributed to the detection of arousals: the estimated discriminant function, when applied to the training set of fixed length basic epochs, resulted in 90.2% sensitivity, 39.5% selectivity and 90.6% specificity.

Table 3 — Distribution of the arousals in the testing set obtained from the validation process in which the events indicated by experts and/or the computer program have been marked as definite, possible (uncertain) or wrong.

		Definite arousals		Possible arousals		Definite + possible	
Recording	Minutes	#	Index	#	%	#	Index
4	345	243	42.26	60	19.80	303	52.70
5	300	241	48.20	35	12.68	276	55.20
6	471	91	11.59	13	12.50	104	13.25
7	336	66	11.79	27	29.03	93	16.61
8	430	157	21.91	35	18.23	192	26.79
9	418	65	9.33	20	23.53	85	12.20
10	454	165	21.81	27	14.06	192	25.37
11	480	78	9.75	38	32.76	116	14.50

The resulting set of coefficients for the discriminant function was fed to the program which repeatedly processed the three records of the training set by using different values for the weights assigned to the EEG and the EMG. The resulting patterns of sensitivity versus selectivity are shown in Fig. 3a: selectivity slowly decreased while sensitivity increased to 90-95%, then selectivity fell under 80% and quickly dropped. The evaluation of the trend depicted in Fig. 3a does not account for the final stage of the program, in which a number of events were cancelled because occurring either in REM stage (and without muscle activity) or too close to each other. The effect of this correction is depicted in Fig. 3b for the selected weight of the EMG (0.75). On this basis a value was chosen for the EEG weight producing 95.6% sensitivity and 91.9% selectivity in the training set. The number of arousals detected in the recordings of the training set is reported in Table 2 with corresponding indices.

Table 4 — Sensitivity and selectivity resulting for each observer and each recording in the testing set, with reference to the arousal set resulting from the validation process: a) considering only definite arousals as true arousals; b) considering both definite and possible arousals as true arousals.

a) definite arousals

	First expert		Second expert		Computer	
Recording	Sensitivity	Selectivity	Sensitivity	Selectivity	Sensitivity	Selectivity
4	81.07	79.12	90.95	82.46	92.59	76.79
5	83.40	86.64	90.04	82.82	89.63	81.82
6	71.43	86.67	64.84	98.33	89.01	72.32
7	40.91	72.97	46.97	83.78	93.94	58.49
8	76.43	83.92	84.71	79.17	88.54	72.40
9	80.00	83.87	69.23	77.59	81.54	80.30
10	66.67	93.22	70.30	80.56	88.48	76.44
11	55.13	65.15	76.92	75.95	88.46	64.49
Total	73.69	82.99	79.75	81.97	89.60	74.46

b) definite + possible arousals

	First expert		Second expert		Computer	
Recording	Sensitivity	Selectivity	Sensitivity	Selectivity	Sensitivity	Selectivity
4	78.88	95.98	86.14	97.39	89.44	92.49
5	80.80	96.12	89.13	93.89	86.96	90.91
6	68.27	94.67	57.69	100.00	90.38	83.93
7	39.78	100.00	37.63	94.59	92.47	81.13
8	71.35	95.80	80.73	92.26	85.94	85.94
9	70.59	96.77	63.53	93.10	74.12	95.45
10	59.38	96.61	67.71	90.28	85.94	86.39
11	54.31	95.45	66.38	97.47	79.31	85.98
Total	69.36	96.13	74.80	94.61	86.41	88.35

The tuned program was applied to the eight polygraphic recordings of the testing set, for which the overall agreement between the program and the experts was 56.2% and 58.3% respectively, while the agreement between the two experts reached 68.8%. The whole set of arousals detected by the program and by the experts then underwent the validation process, which resulted in 1125 definite arousals and 266 possible ones (mean arousal index: definite: 20.9 per hour, definite+possible: 25.8 per hour). The number of arousals detected for each recording and corresponding indices are reported in Table 3. Sensitivity and selectivity were computed for each observer with reference to the validated arousal set and results are reported in Table 4a and 4b for each recording. Considering only definite arousals the overall sensitivity resulted in 72.4% and 78.4% for the

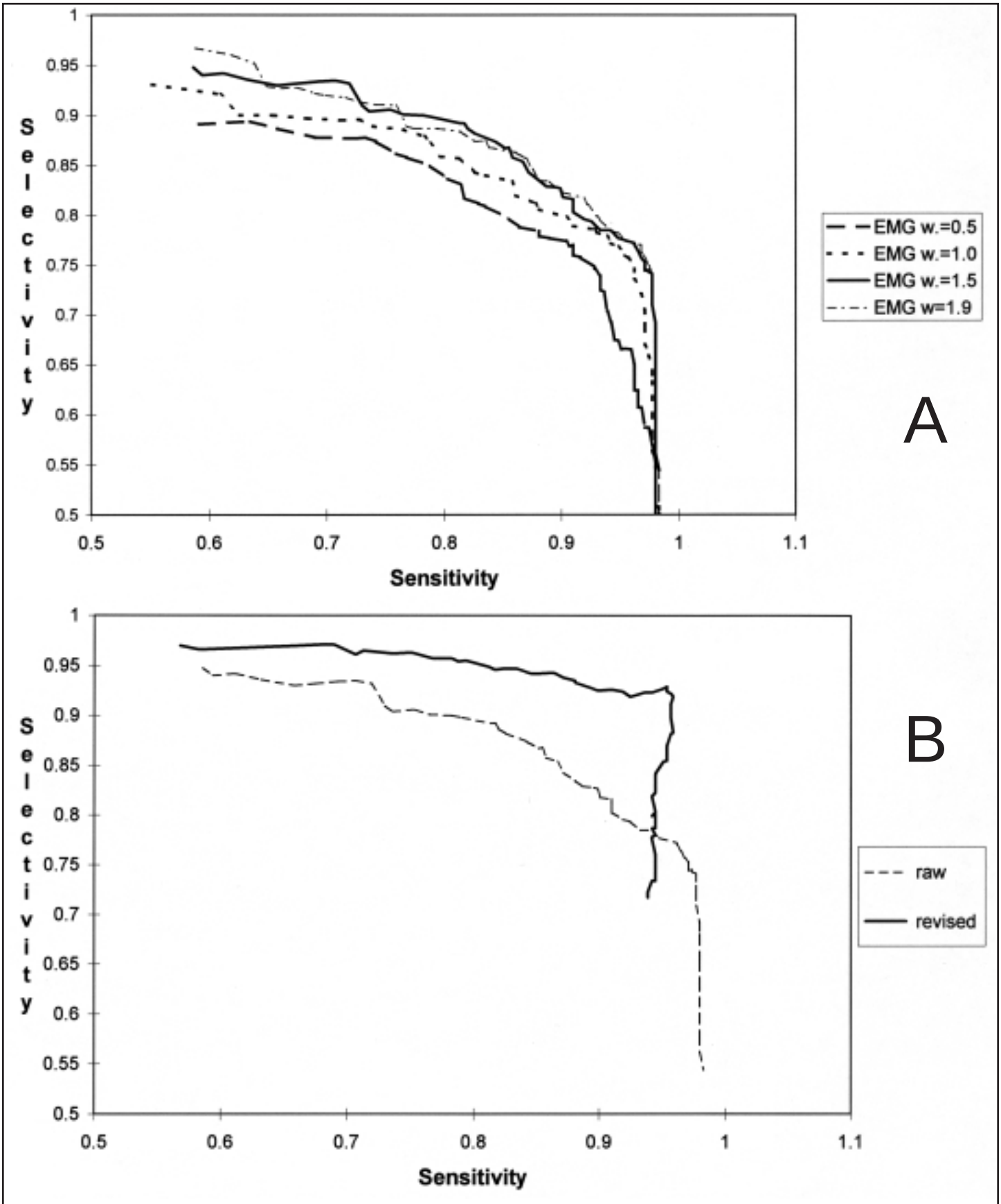

Figure 3 — Variation of selectivity versus sensitivity of the automatic system following variations in system parameters. a) The weights of EEG and EMG vary according to a fixed threshold for arousal detection: each curve refers to a particular value for EMG weight, while EEG weight slowly changes in a wide range. The number of arousals detected increases with the weight of EEG, generating both true and false positives, so selectivity decreases while sensitivity increases. When sensitivity approaches one, the selectivity quickly drops. b) A revision procedure enables the elimination of many false positives, resulting in higher selectivity values for a wide range of sensitivity: a set of weights is thus chosen leading to 95% of sensitivity and 91% of selectivity in the training set.

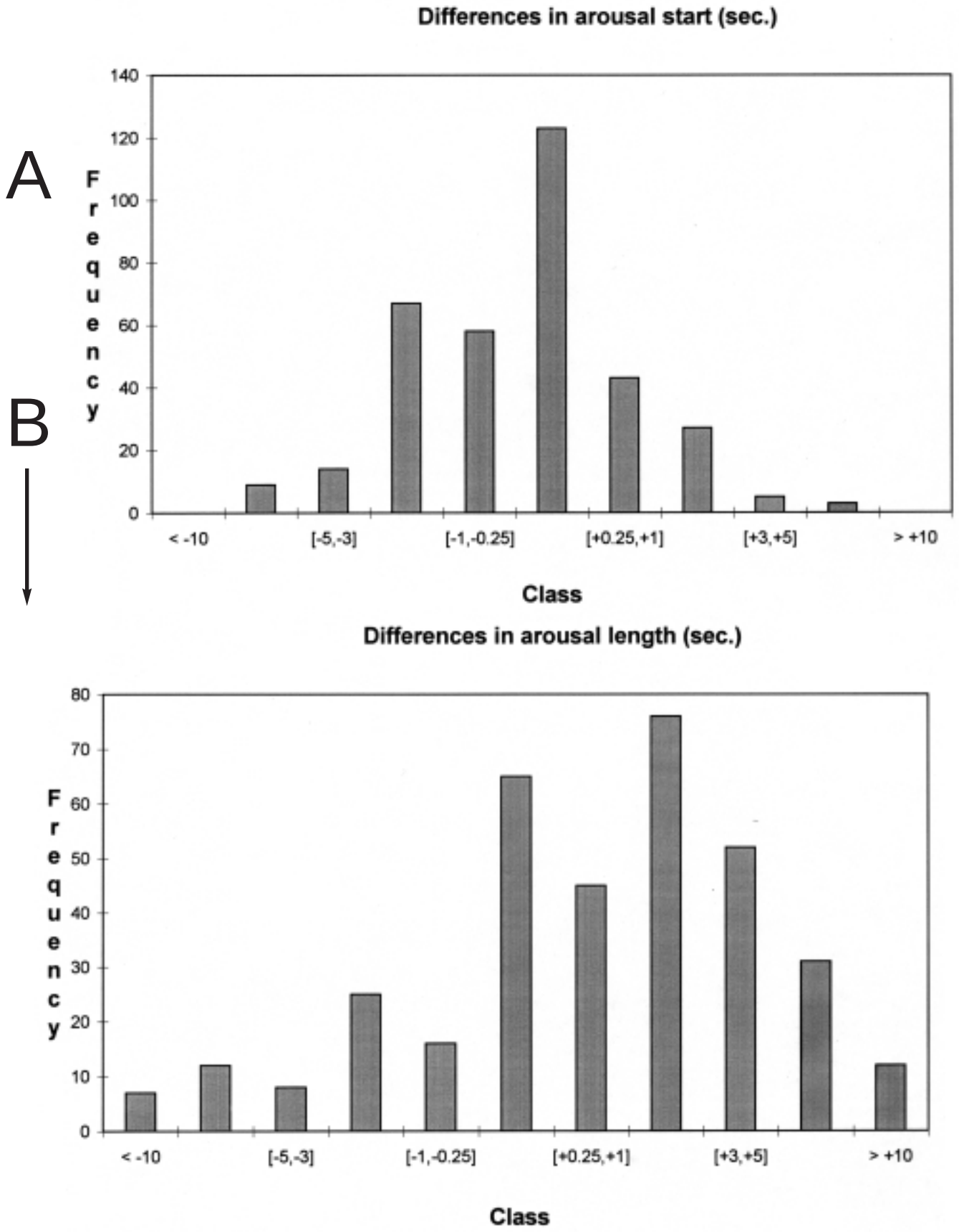

Figure 4 — Time relationship between overlapping events as resulting from the comparison of reference and system detected events in the training set. a) Histogram of the differences in arousal start. b) Histogram of the differences in arousal length.

two experts and 88.1% for the computer program while selectivity was 83.0% and 82.0% for the experts and 74.4% for the program. By including possible arousals in the reference sensitivity decreased to 67.9%, 73.2% and 84.5% while selectivity increased to 96.1%, 94.6% and 88.4%. Among events added during the validation phase, but missed by all during the first pass, 19 were classified as definite (1.69%), and 11 as possible (4.14%).

When considering only definite arousals, the null hypothesis of no difference in sensitivity was rejected at 0.001 level, and the further analysis of the distribution between true positive and false negative indicated a significant difference between the two experts and between each expert and the program. Differences in selectivity were also significant as a whole (at 0.001 level), but not between the human experts: the selectivity of the program was significantly lower than that of each human expert.

The mean arousal length, limited to the training set, resulted in 10.9+/-6.4 sec. As concerns the training set, the distribution of the differences between the computer and the experts in the evaluation of the arousal start, is shown in Fig 4a: in 35% of events the difference was lower than 0.25 seconds while in 64% of events it was lower than 1 second; the computer program tended to advance the arousal start slightly. The distribution of the differences in the evaluation of the arousal duration is shown in Fig. 4b; a greater divergence appears in this case: 19% of events coincided almost exactly (less than 0.25 seconds difference) and 36% of events were within 1 second difference, while in most cases the computer tended to assign a longer duration.

DISCUSSION

The analysis of the testing data set indicates that the sensitivity of the automatic system is higher than that of human experts while its selectivity is lower.

Unlike conventional analysis of sleep macrostructure, in which the stages are assigned to fixed length epochs, the main object of sleep microstructure analysis is the detection of transient events, among which arousals, and the analysis of their distribution in time, duration and other significant features. The system therefore used context and multichannel information to detect arousals as individual phasic events. The evaluation of the agreement between an observer and the reference set of arousals was therefore based on the identification of individual events, admitting for partial overlapping. The time relationship between overlapping events was considered of minor importance at this stage. However, the preliminary analysis performed on the training set indicates a fair agreement between computer and experts on arousal starting time but a greater divergence regarding arousal duration. This problem requires appropriate analysis that may lead to further development of the method.

The first comparison between the arousals marked by different observer in the testing set showed little agreement between the experts and an even lower agreement between each expert and the computer program. The inter-expert agreement is considerably lower than that obtained in a recent study concerning arousals in normal sleep,[4] although it is closer to the level of agreement reported by another study on arousals in respiratory sleep disorders and involving a number of sleep laboratories.[21]

In order to better the evaluation of the disagreements and the estimate of sensitivity and selectivity of each observer, a final stage of validation of the detected arousals was performed by the two human experts. The numerous events detected by the automatic system, but not by the experts, were thus analyzed enabling the distinction between true arousals, which escaped the human reader, and false positive ones. The experts also revised each tracing completely to look for events missed by all during the previous steps, though only few unmarked events were detected. The validation was performed three months after the previous scoring, by mutual consent of the two experts, in compliance to AASM criteria. No information was directly available concerning who had detected each arousal, individual patterns could hardly be remembered and only a vague perception of personal style could influence the reviewers, mainly typical errors made by the automatic system. In a way the validation process forced the readers to a closer examination of the detection criteria. The low inter-scorer agreement resulting from the first scoring and the persistence of a portion of uncertain events indicated certain objective difficulties in arousal characterization. Some of these difficulties concerned events that scarcely met AASM criteria but could definitely be included in the activation phase of what has been described as a cyclic alternating pattern - CAP [2]. Another source of uncertainty was the presence of effects of drug use by some patients originating bursts of beta waves. The prevalence of fast activities, especially in tracing number 7, induced the experts to neglect several events in their first analysis since they considered them as not having the same origin as arousals. However, in the validation phase a stricter application of the rules for arousal detection brought to the inclusion of all those events that met the conventional criteria marking borderline cases as possible.

When exploring the differences between the arousals detected by the automatic system and the reference set we came across several cases in which the arousal was marked during the wake state, owing to the misinterpretation of artifacts or changes in the vigilance level. The addition of a sleep onset detector could possibly lead to a better performance in such situations.

However, our results indicate the usefulness of wavelet analysis in the detection of EEG changes accompanying arousals, and, probably, in the characterization of other

microstructure EEG patterns. Several indices are required to take the different arousal patterns into account: some of these patterns are characterized by an increase in the alpha band and could be detected by using a single index such as the alpha slow-wave index (ASI), previously used for the detection of episodes of wakefulness during sleep.[13] Other patterns are dominated by EEG desynchronization, with low or absent alpha: a measure of mean frequency seems more appropriate in this case, as already suggested by Drinnan,[16] though overlapping slow waves may be included in arousal patterns with considerable effects on slowing the mean frequency down. The evaluation of various indices, supported by the discriminant analysis, enables the detection of patterns in which different features are present and differently combined. A further step can be made in the analysis of microstructural events by associating the wavelet analysis with other tools, such as artificial neural networks, for the automatic extraction of the main features of the patterns of interest in the training set.[1,22]

The performance of the automatic system is, however, interesting for its applications, as the arousals can be detected with high sensitivity and then validated by a human expert for the rejection of false positives. This could substantially decrease the time taken by the expert to score the records for arousals.

Although this study was only designed for methodological purposes, some impressions concerning the distribution of arousals in different pathological conditions may be drawn from the results. Patients with breathing disorders had a higher arousal index than normative physiological values for their age-span.[4] The arousal index increased with the apnea index, reaching values just over one per minute - commonly reported as extreme arousal rate.[12] In these cases conventional sleep staging was hardly applicable, as sleep onset and arousal alternated continuously. As for the remaining pathologies, we found the arousal index value to be higher than normal in the narcoleptic patient while almost physiological values were detected in the other recordings. This seemed to reflect the regular sleep structure found in the two patients with partial epilepsy, but was surprising in patients affected by sleep disturbances (psychophysiological insomnia and nocturnal myoclonus). Of course, the comparison with normative data requires caution, considering the low inter-scorer agreement reported to date.[21] Moreover, further development of automatic systems may help in providing a homogeneus basis for the confrontation of data. However, our data may suggest that sleep fragmentation and instability are more complex than what is reflected by the arousal index alone and that other microstructural features, such as the cyclic alternating pattern,[23] could support clinical purposes.

APPENDIX

Details concerning the detection algorithm, preceded by a brief introduction to the wavelet transform (i.e., the method we used to evaluate the time frequency structure of EEG) are supplied in this appendix. A wider introduction of wavelet technique can be found elsewhere, as in[24,25,26] for a general presentation and in[18,19] for its applications to EEG analysis.

a. The wavelet transform

The analysis of the frequency composition of a signal is mostly performed by the Fourier transform, which enables the evaluation of statistical features of a stationary signal. The Fourier transform was adapted to the analysis of non-stationary processes by introducing the window Fourier transform, in which the signal is seen through a sliding window of limited extent, considering it as approximately stationary for such a short period of time. The size of the window determines the analysis time and frequency resolution.

In the wavelet transform a family of functions is used to decompose the signal: each function is the translation and dilatation of a unique prototype $\psi(t)$, with few oscillations concentrated in time. The wavelet transform is then defined by:

$$Wf(s,\tau) = \int f(t) * \sqrt{s} * \Psi(s*(t-\tau))dt$$

where s is the scale factor, which, in this formulation, replaces frequency — the higher the value of s, the lower the frequency explored. As opposed to the windowed Fourier components, in which a fixed time window is used, the scale factor expands or contracts the whole wavelet, so that high frequency components are more concentrated in time than low frequency components. Consequently, the wavelet transform enables a multiresolution analysis in which the precision of time and frequency localization is automatically adapted to the frequency content of the explored patterns.

The discrete wavelet transform, in the form we apply to digital EEG, can be described as the recursive application of a pair of half-band mirror filters, which decomposes the signal into two series at each step: the low-pass filter originates the coarse approximation (slow component), while the high-pass filter originates the series of additional details (fast component). Both series are then sub-sampled by a factor two, thus halving the time resolution. In the multiresolution approach the pair of filters is applied first to the raw signal and then repeatedly to the approximation series only; the result may be viewed as the output of a filter bank with a constant relative bandwidth. The decomposition procedure may be adjusted, following the "wavelet packet" scheme, in order to enhance the discrimination of particular patterns.

In our application to the EEG the discrete wavelet trans-

form was implemented by the Daubechies wavelet filter with 12 coefficients, analyzing 32-sec overlapping epochs with 128-Hz sampling rate: this originated the set of bands: 32-64, 16-32, 8-16, 4-8 ... 0.-0.0325Hz with a bandwidth ranging between 32 to 0.0325Hz. In order to obtain a better detection of the relevant EEG patterns, we stopped the decomposition with the 0.0-0.5 band and conversely further halved the 8-16-Hz band to differentiate the alpha activity from the sleep spindles. The time resolution resulted in 0.125 sec for the 4-8, 8-12 and 12-16-Hz bands; but grew for the beta band and decreased for the delta band, which turned out to be suitably decomposed into four subbands.

b. The EEG indices for arousal detection

The results of wavelet transform were used to evaluate a set of indices describing the EEG changes accompanying arousals, mainly consisting in an abrupt shift in EEG frequency and an increase in theta, alpha and/or beta activity. In order to detect these variations, the power was computed for the following six frequency bands: 0.0-0.5 (slow delta), 0.5-4 (delta), 4-8 (theta), 8-12 (alpha), 12-16 (sigma), and 16-64 (beta). The time resolution was 0.125 sec, proper of the theta, alpha and sigma bands. A long-term weighed, moving, average of power was then computed for each band to estimate a background reference value, while a short-term moving average stood for the actual trend; in both cases the average value was estimated as:

$$y_i = \frac{N * y_{i-1} + x_i}{N+1}$$

where $\{x_i\}$ represents the input series of band powers, $\{y_i\}$ the output series of moving averages and N is a coefficient qualifying the time constant of the exponentially decaying window, thus differentiating short-term from long-term averages.

The first six indices were the ratios between short-term and long-term average, indicating the actual variation for each band; the other indices involved average power in different bands and evaluated particular features that may be important in arousal detection. They were defined as follows:

- the ratio between short-term and long-term mean frequency, where the mean frequency was computed as:

$$\bar{f} = \frac{\sum_i p_i * f_i}{\sum_i p_i}$$

where f_i is the central frequency of the band and p_i is its power; the index was sensitive to the frequency shift;

- the ratio between delta and alpha plus beta power, computed for both short-term and long-term average; this could indicate the presence of slow wave sleep;
- the ratio between short-term and long-term alpha relative power, which highlighted variations in alpha activity;
- the ratio between long-term alpha plus slow-delta and theta plus delta power; it could indicate that the subject was already awake;
- the ratio between sigma and alpha plus beta power, which could suggest the presence of sleep spindles;
- the ratio between beta and delta variations (each expressed as the ratio between short-term and long-term average power), which could suggest a desynchronization of the tracing.

The linear discriminant function was estimated by maximizing the sensitivity and the specificity when applied to the training set of fixed length basic epochs. It has the form:

$$F(X) = \sum_i w_i * x_i - \theta$$

where X is the index vector $\{x_i\}$, $\{w_i\}$ is the weight vector and θ is the threshold.

The algorithm for arousal detection evaluated the set of indices for every 0.125-sec basic epoch. When the discriminant function became positive an arousal detection mode was activated in which long-term averages were not updated. Possible arousals were marked when the discriminant function remained positive for more than three and up to thirty seconds (after which an awakening was detected); a score was then assigned to each arousal resulting from the mean value of the discriminant function.

The program proceeded by analyzing EMG data and by combining the results from the different channels as described in the Method section.

c. The measure of sensitivity, selectivity and specificity.

The performance of a particular run of the program or a particular reading by an expert were evaluated in terms of sensitivity, selectivity and specificity, where:

$$sensitivity = \frac{TP}{TP+FN}\ ;\ selectivity = \frac{TP}{TP+FP}\ ;\ specificity = \frac{TN}{TN+FP}$$

The specificity was computed only in the context of the discriminant analysis, in which each fixed length basic epoch was classified as true positive (*TP*), false positive (*FP*), true negative (*TN*) or false negative (*FN*). In subsequent analyses, variable length arousals, marked by one observer, were compared to the reference set and the individual events were considered as *TP* (if an overlapping occurred), *FP* or *FN*. We believed that in this case *TN* counting, and consequently specificity evaluation, was nonsensical.

REFERENCES

1. Kubicki S, Herrmann WM. The future of computer assisted investigation of the polysomnogram: sleep microstructure. JCfin-Neurophysiol 1996;13:285-94.
2. Terzano MG, Halasz P, Declerck AC, Eds. Phasic events and dynamic organization of sleep. New York: Raven Press, 1991.
3. Mathur R, Douglas NJ. Frequency of EEG arousals from nocturnal sleep in normal subjects. Sleep, 1995;18:330-3.
4. Boselli M, Parrino L, Smerieri A, Terzano MG. Effect of age on EEG arousalsin normal sleep. Sleep, 1998;21:351-7.
5. Guilleminault C, Stoohs R, Clerk A, Cetel M, Maistros P. A cause of excessive sleepiness: the upper airway resistance syndrome. Chest 1993;104:781-7.
6. Rosental L, Roehrs T, Sicklesteel Y, Zorick F, Wittig R, Roth T. Periodic legmovements during sleep. Sleep 1984;7:326-30.
7. Bonnet MH. Effect of sleep disruption on sleep performance and mood. Sleep1985;8:11-9.
8. Philip P, Stoohs 1~, Guilleniinault C. Sleep fragmentation in normals: a model for sleepiness associated with upper airway resistance syndrome. Sleep 1994;17:242-7.
9. Roehrs T, Merlotti L, Petrucelli N, Stepanski E, Roth T. Experimental sleep fragmentation. Sleep 1994;17:438-43.
10. Cheshire K, Engleman H, Deary 1, Ahapiro C, Douglas NJ. Factors impairing daytime performance in patients with sleep apnea/hypopnea syndrome. Chest 1992;152:538-41.
11. Rechtschaffen A, Kales A, eds. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. Los Angeles: Brain Information Service/Brain Research Institute, University of California Los Angeles, 1968.
12. Atlas Task Force of American Sleep Disorders Association. EEG arousals: scoring rules and examples. Sleep 1992;15:174-84.
13. Jobert M, Schulz H, Jahnig P, Tismer C, Bes F, Escola H. A computerizedmethod for detecting episodes of wakefulness during sleep based on the alpha slow wave index (ASI). Sleep 1994;17:37-46.
14. Creutzfeldt OD, Bodenstein G, Barlow JS. Computerized EEG pattern classification by adaptive segmentation and probability density function classification. Clinical evaluation. Electroencephalog Clin Neurophysiol 1985;60:373-93.
15. Hasan J, Hirvonen K, Varri A, Hakkinen V, Loula P. Validation of computer analysed polygraphic patterns during drowsiness and sleep onset. Electroencephalogr Clin Neurophysiol 1993;87:117-27.
16. Drinnan MJ, Murray A, White JES, Smithson AJ, Griffiths CJ, Gibson GJ. Automated recognition of EEG changes accompanying arousal in respiratory sleep disorders. Sleep 1996;19:296-303.
17. Huupponen E, Hasan J, Varri A, Saarinen J, Kaski K. Arousal detection with a neural network. J Sleep Res 1996;5(l):97.
18. Jobert M, Tismer C, Poiseau E, Schulz H. Wavelets — a new tool in sleep biosignal analysis. JSleep Res 1994;3:223-32.
19. Shiff SJ, Aldroubi A, Unser M, Sato S. Fast wavelet transformation of EEG. Electroencephalogr Clin Neurophysiol 1994;91:442-55.
20. SAS Institute Inc. SASISTAT User's Guide, Release 6.03 Edition. Cary, NC: SAS Institute Inc, 1988.
21 Drinnan MJ, Murray A, Griffiths CJ, Gibson J. Interobserver variability in recognizing arousal in respiratory sleep disorders. Am J Respir Crit Care Med 1998;158:358-62.
22. Pardey J, Roberts S, Tarassenko L, Stradling J. A new approach to the analysis of the human sleep/wakefulness continuum. J Sleep Res 1996;5:20 1 -10.
23. Terzano MG, Parrino L. Clinical applications of cyclic alternating pattern. Physiol Behav 1993;54:807-13.
24 . Mallat SG. Multifrequency channel decompositions of images and wavelet models. IEEE Trans Acoust Speech Signal Proc 1989;37:209 1 -110.
25 Rioul 0, Vetterli M. Wavelets and signal processing. IEEE SP Magazine 1991; October:14-38.
26 Daubechies 1. The wavelet transform. Time-frequency localization and signal analysis. IEEE Trans Info Yheory 1990;36:961-1005.