

VISUAL AND AUTOMATIC CYCLIC ALTERNATING PATTERN (CAP) SCORING

Inter-rater reliability study

Agostinho Rosa¹, Gabriela Rodrigues Alves², Magneide Brito²,
Maria Cecília Lopes^{1,2}, Sérgio Tufik²

ABSTRACT - The classification of short duration events in the EEG during sleep, as the A stage of the cyclic alternating pattern (CAP) is a tedious and error prone task. The number of events under normal conditions is large (several hundreds), and it is necessary to mark the limits of the events with precision, otherwise the time sensitive classification of the CAP phases (A and B) and specially the scoring of different types of A phases will be compromised. The objective of this study is to verify the feasibility of visual CAP scoring with only one channel of EEG, the evaluation of the inter-scoring agreement in a variety of recordings, and the comparison of the visual scorings with a known automatic scoring system. Sixteen hours of one channel (C4-A1 or C3-A2) of NREM sleep were extracted from eight whole night recordings in European Data Format and presented to the different scorers. The average inter-scoring agreement for all scorers is above 70%, the pair wise inter-scoring agreement found was between 69% up to 77.5%. These values are similar to what has been reported in different type studies. The automatic scoring system has similar performance of the visual scorings. The study also has shown that it is possible to classify the CAP using only one channel of EEG.

KEY WORDS: cyclic alternating pattern, CAP, sleep, visual scoring, automatic scoring.

Estagiamento visual e automático do padrão alternante cíclico (CAP)

RESUMO - A classificação dos eventos de curta duração no EEG durante o sono, como a fase A do padrão alternante cíclico (CAP) é tarefa tediosa e propensa ao erro. O número de eventos é geralmente grande (centenas ao milhar), e é necessário marcar os limites dos eventos com precisão, caso contrário tanto a classificação das duas fases do CAP como a distinção dos diferentes tipos de fases A podem ser comprometidas. O objetivo deste estudo é a avaliação da concordância inter-estagiadores no CAP tanto visual quanto automático usando um único canal de EEG, a avaliação do acordo inter-estagiadores em uma variedade das gravações, e a comparação das classificações visuais com um sistema classificador automático. Dezesesseis horas de um canal (C4-A1 ou C3-A2) do sono de NREM foram extraídas de oito gravações inteiras da noite no Formato de dados Europeus e apresentadas a diferentes classificadores, o acordo médio inter-estagiadores para todas as classificações está acima de 70%, o acordo inter-estagiadores do par encontrado entre 69% até 77,5%. Estes valores são similares ao que foi relatado nos diferentes tipos de estudos. O sistema marcador automático tem o desempenho similar aos dos classificadores visuais. O estudo também mostrou que é possível classificar o CAP usando somente um canal de EEG.

PALAVRAS-CHAVE: padrão alternante cíclico, CAP, sono, estagiamento visual, estagiamento automático.

The classification of short duration events in sleep EEG like the A phase of the cyclic alternating pattern¹ (CAP) is a tedious and error prone task. The number of events is usually large (several hundreds up to thousands), it is necessary to mark the events boundary with high accuracy, otherwise the classifi-

cation of the events into different A subtypes and the accuracy of the phase durations, cyclic and the alternating conditions may be compromised.

There is no published data about CAP sequences inter-scoring and *intra-laboratory* scoring agreement, or comparison with published automatic scoring sys-

¹Laboratório de Sistemas Evolutivos e de Engenharia Biomédica, ISR, IST, UTL - Lisboa, Portugal; ²Departamento de Psicobiologia - Instituto do Sono, EPM - UNIFESP, São Paulo, Brasil.

Received 30 November 2005, received in final form 6 March 2006. Accepted 22 April 2006.

Dr. Agostinho Rosa - Laseeb-ISR-IST - Av. Rovisco Pais 1 - 1049-001 Lisboa, Portugal. E-mail: acrosa@isr.ist.utl.pt

Table 1. Truth table for events scoring.

	Phase A (1=Ref)	Phase B (1=Ref)	
Phase A (2)	T	Fp	SR2 =T/(T+Fp)
Phase B (2)	Fn	D=(Fp+Fn)/2	
	S2R=T/(T+Fn)		C=T/(T+D)

Ref, reference; 1, scorer 1; 2, scorer 2; T, true positive; Fp, false positive; Fn, false negative; S21, scorer to reference sensitivity; S12, reference to scorer sensitivity; C, mutual agreement. The value C (in percentage) is always between SR2 and S2R. It is calculated as sensitivity between 2 classifiers without taking any of them as reference.

tems²⁻⁹. The effects of age, patient group, recording characteristics on the scoring agreement is also not completely known¹⁰⁻¹². There was also concern about the feasibility, accuracy and reliability of sleep stage independent CAP classification using single EEG lead (C4-A1 or C3-A2).

The main objective of this work is the assessment of viability and performance of visual and automatic scoring on single EEG channel, as suggested by the published CAP scoring rules¹, using recordings from different laboratories, age, and gender and sampling frequency of normal and diseased subjects.

METHOD

Sixteen hours with one EEG channel, either C4-A1 or C3-A2 channel selected from eight full night polysomnographic recordings from different sleep laboratories (sampling frequencies of 100, 128 and 256Hz) in EDF format¹³ where used in this study. Three visual and one automatic scoring system² participated in this process.

All the phase A scorings where converted into ISO time format (and decimated to the same sampling frequency if appropriate) for agreement analysis. Two type of analysis can be performed; one is the time based agreement, where the scorings are compared for every sampling point and the second is event based, where an agreement is scored whenever 2 events have an intersection longer than 0.5 seconds. In this work only the event based analysis will be presented. There are different possibilities for event based analysis. When 2 events intersects more than a specific period of time it is considered a hit (T), else events are marked either a miss (Fn) or a false (Fp). In this situation there is no much sense scoring the true negatives, since it is always in the neighbor of the total number of events of the score r taken as reference.

Under these circumstances a specific agreement analysis was proposed in Table 1, where we can only calculate de sensitivity of scorer to reference (S2R) or vice-versa (SR2) when we invert the role. The "mutual agreement" (C) is the measure of the agreement between any two scorers without taking any of them as reference.

Table 2. All files average mutual agreement. The diagonal values represent the inter-scorers average C.

	A	1	2	3
A	72.09			
1	68.95	71.85		
2	70.51	75.17	74.39	
3	76.79	71.43	77.49	75.24

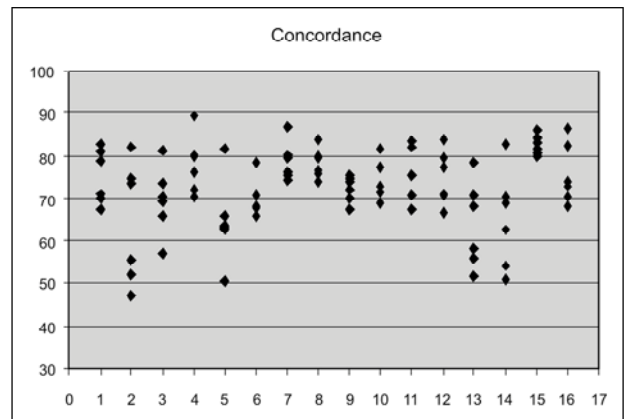


Fig 1. The scatter plot of all C values in different files.

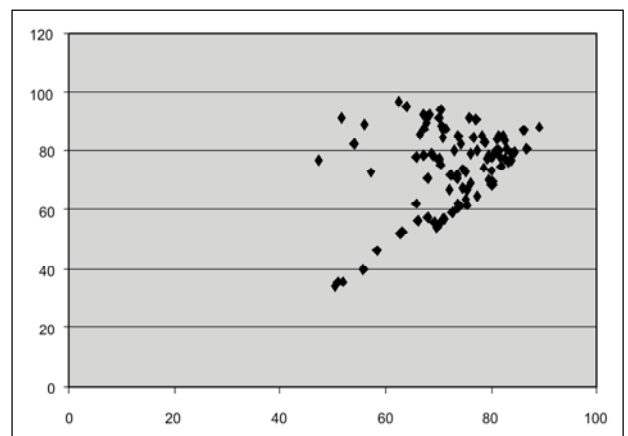


Fig 2. X axis sensitivity, Y axis mutual agreement scatter plot of all pair wise file by file comparisons. This graph shows high density of sensitivity or mutual agreement between 80 up to 100% for individual files.

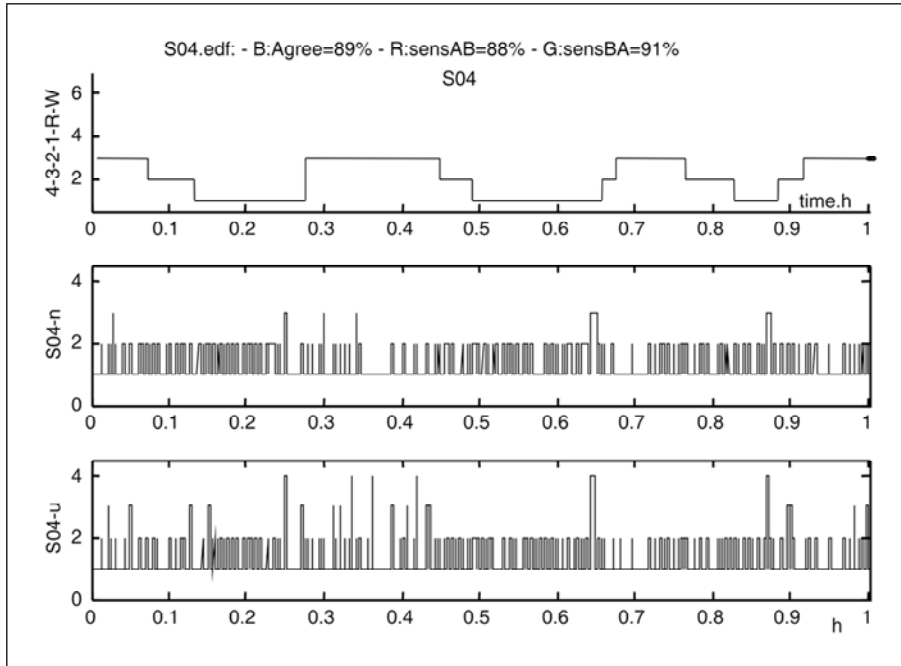


Fig 3. Example of scoring comparison of an 1 hour file. Top trace is the sleep stage, middle trace is scorer 2 and bottom is scorer 3. The mutual agreement is 89.4%.

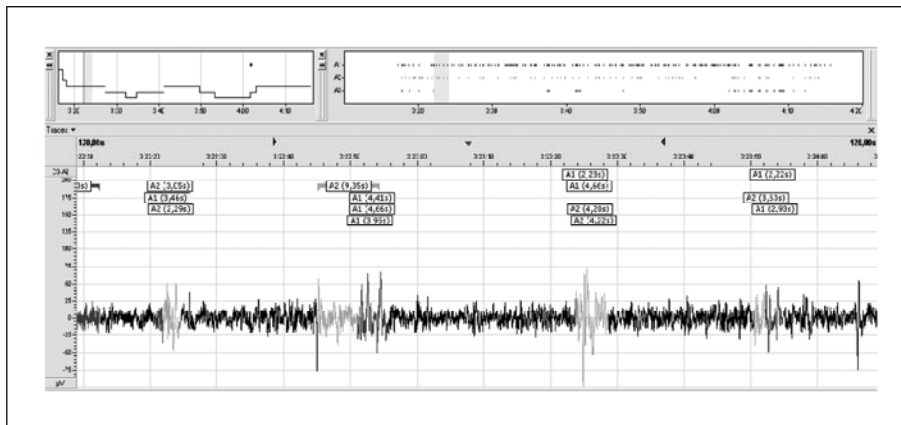


Fig 4. Stage 2, 2 minutes epoch scorings.

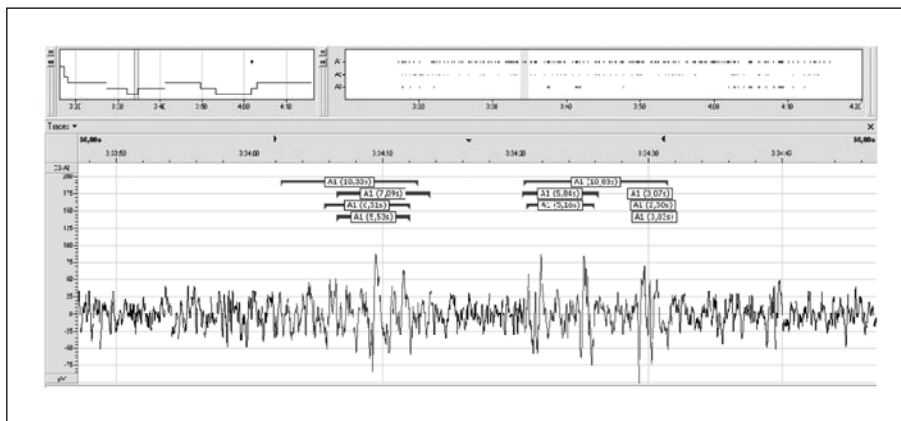


Fig 5. Stage 3-4, 1 minute epoch.

RESULTS

The total number of events scored by the automatic scoring system and the 3 visual classifiers are 1358, 1292, 1341 and 1450 respectively. All pair wise scorers (automatic or visual) were compared file by file and the average results of mutual agreement are presented in Table 2. The table is symmetrical therefore only the upper triangular matrix is presented. Highlighted values are highest inter-scorer C of 77.49% for scorers 2-3 (red color), highest overall inter-scorers C is 75.2% for scorer 3 (blue).

Figure 1 shows the scatter plot of all C values in different files. Roughly half of the files have a small spread (10-20%) and the other half about the double (30-40%). The four lowest spread (files 7, 8 and 15 are from lab D and (files 1 and 9 from lab B) are from the same labs. So it seems that there are specific factors that may influence the mutual agreement C.

Figure 2 shows on the X axis the sensitivity and on Y axis the mutual agreement scatter plot of all pair wise file by file comparisons. This graph shows high density of sensitivity or mutual agreement above 70% for individual files. The limit curve of this graph suggest a quadratic curve for the putative "pareto front" for the scoring process, indicating also the best averaged mutual agreement and sensitivity are both around 90%.

Figure 3 shows an example of scoring comparison between scorer's 2 and 3 of an one hour file. The mutual agreement is 89% and sensitivity of 88 and 91%.

Figures 4 and 5 show several examples of scorings in stages 2 and SWS, indicating high clustering of agreements, but also pointing boundary decision problems. This boundary (start and stop) problem could be the main cause of phase A subtypes classification mismatch.

DISCUSSION

From this study some conclusions could be drawn. The first one is the possibility of classifying CAP using only a single EEG channel. Adding more channels with associated and clearly defined criteria may im-

prove further the agreement rates. Although the experience of different scorers varies a lot (from newcomers to experienced ones) the results are in the range of what is seen for inter-scorer agreement in similar type of studies.

Another important result is that the automatic scoring performance is similar to visual scorings.

Acknowledgments – Rogerio Largo of Laseeb-ISR who provided the scoring comparison tools. CAP Validation Group for the organization model of this work. Medcare for providing the manual scoring program. Dra. Márcia Pradella-Hallinan for encouraging the research work.

REFERENCES

1. Terzano M, Parrino L, Smerieri A, et al. Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep. *Sleep Medicine* 2002;3:187-199.
2. Rosa AC, Allen Lima J. Fuzzy classification of microstructural dynamics of human sleep. *Proceedings of the 1996 IEEE International Conference on Systems, Man and Cybernetics - Information, Intelligence and Systems. SMC'96* 1996;2:1108-1113.
3. Allen Lima J, Rosa AC. Maximum likelihood based classification for the microstructure of human sleep. *ACM SigBio Newsletter* 1997;17 Issue 3.
4. Ferrillo F, Gabarra M, Nobili L, et al. Comparison between visual scoring of cyclic alternating pattern (CAP) and computerized assessment of slow EEG oscillations in the transition from light to deep non-REM sleep. *J Clin Neurophysiol* 1997;14:210-216.
5. Rosa AC, Parrino L, Terzano M. Automatic detection of cyclic alternating pattern (CAP) sequences in sleep: preliminary results. *Electroenceph Clin Neurophysiol* 1999;110:585-592.
6. Navona C, Barcaro U, Bonanni E, Di Martino F, Maestri M, Murri L. An automatic method for the recognition and classification of the A-phases of the cyclic alternating pattern. *Clin Neurophysiol* 2002;113:1826-1831.
7. Largo R, Rosa A. Wavelets based detection of a phases in sleep EEG. *Proc. II Int Conf Computacional Bioengineering* 2005;2:1105-1115.
8. Ferri R, Bruni O, Miano S, Smerieri A, Spruyt K, Terzano MG. Inter-rater reliability of sleep cyclic alternating pattern (CAP) scoring and validation of a new computer-assisted CAP scoring method. *Clin Neurophysiol* 2005;116:696-707.
9. Largo R, Munteanu C, Rosa A. CAP event detection by wavelet and GA tuning. *Proc of IEEE - WISP Evolutionary Computation* 2005. CDROM.
10. Parrino L, Boselli M, Spaggiari MC, Smerieri A, Terzano MG. Cyclic alternating pattern (CAP) in normal sleep: polysomnographic parameters in different age groups. *Electroenceph Clin Neurophysiol* 1998;107:439-450.
11. Lopes MC, Rosa A, Poyares D, Roizenblatt S, Passarelli C, Tufik S. Cyclic alternating pattern in peripubertal children. *J Sleep* 2005;28:215-219.
12. Bruni O, Ferri R, Miano S, et al. Sleep cyclic alternating pattern in normal school-age children. *Clin Neurophysiol* 2002;113:1806-1814.
13. Kemp B, Varri A, Rosa A, Nielsen K, Gade J. A simple format for exchange of digitized polygraphic recordings. *Electroenceph Clin Neurophysiol* 1992;82:391-393.