Automatic artifact detection for whole-night polysomnographic sleep recordings

P541

D. Coppieters 't Wallant[1], S.L. Chellappa[1], G. Gaggioni[1], M. Jaspar[1], C. Meyer[1], V. Muto[1], G. Vandewalle[1], P. Maquet[1,2,3], C. Phillips[1,4]

[1]Cyclotron Research Centre, University of Liège, Belgium; [2]Dep. of neurology, CHU, University of Liège, Belgium; [3]Walloon Excellence in Lifesciences and Biotechnology (WELBIO); [4]Dep. of Electrical Engineering and Computer Science, University of Liège, Belgium.

INTRODUCTION

Analyses of sleep electro-encephalographic data (EEG) have first to detect artifacts in order to reject the corresponding time points before further examination.
Manual artifact detection has two main shortcomings. It is
- A very time consuming and tedious task
- A subjective procedure leading to disagreements between experts

Ideally artefact detection should be *automatic*, *fast*, *reproducible* and *accurate*.

There are currently no such method. Proposed approaches face different issues:
- Lack of specificity (detection of some but not all artifacts)
- Methodological weaknesses (e.g., need of a training set, arbitrarily fixed thresholds)
- Computational burden (huge computing time for a whole night).

See review for artifacts processing in sleep EEG in [1].

The aim of this project was to develop an automatic artifact detection method for whole-night polysomnographic sleep recordings.

METHODS

Principle

Artifacts are marked either:
- per short (1 second) epoch, over all channels
- per channel, over a "scoring window" (20s [2] or 30s [3])

Processing pipeline

The raw data are processed through different modules (Figure 1) that are applied successively and have a specific task: ``pre-processing", ``bad channel detection", and ``artifact detection" (Figure 2).

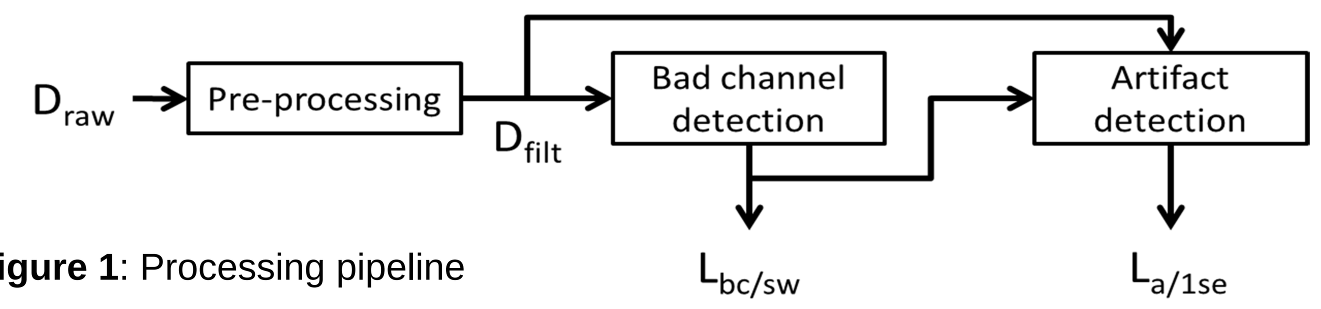

Figure 1: Processing pipeline

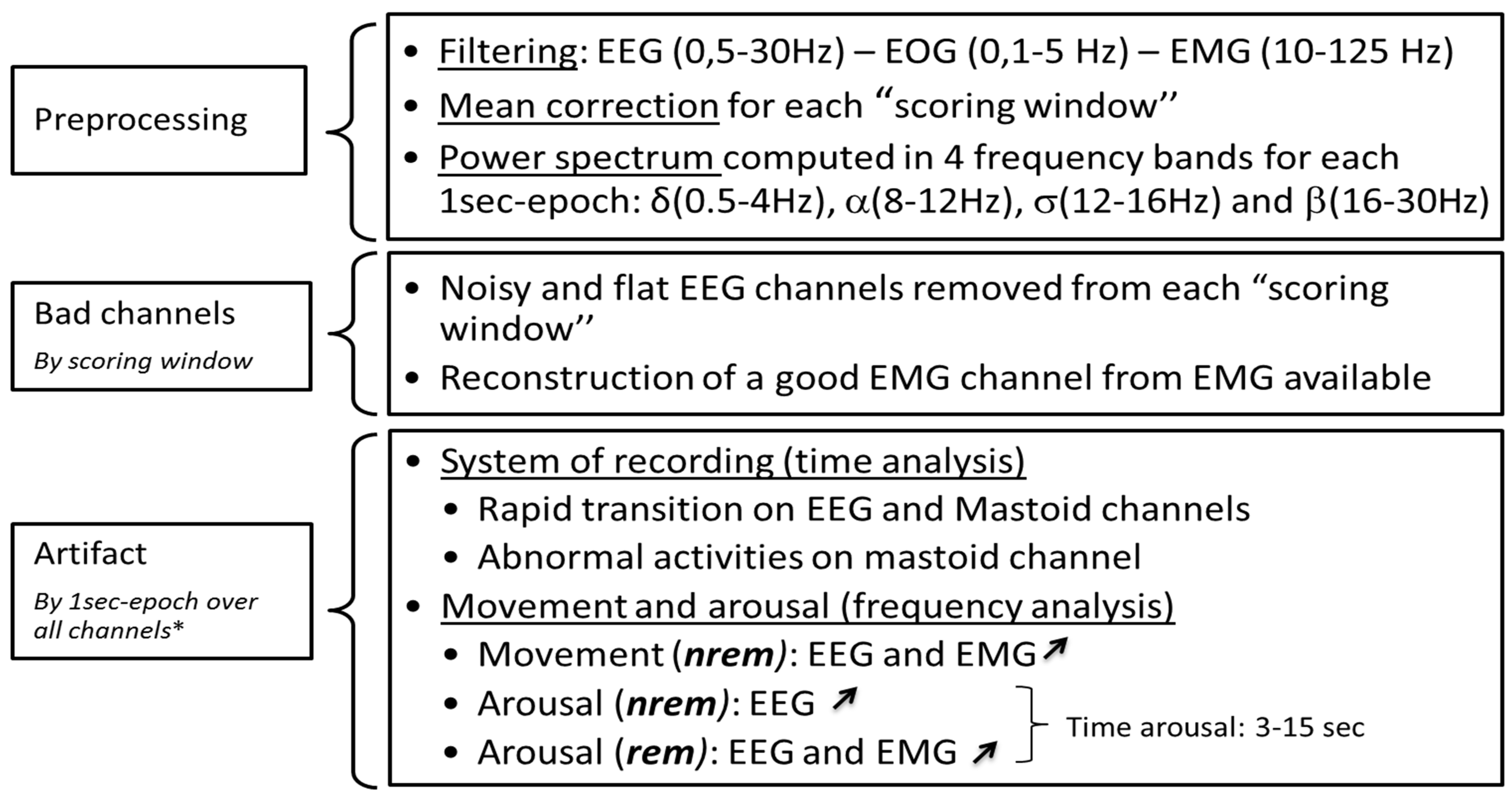

Figure 2: Modules details

Output
- list $L_{bc/sw}$ of 'bad channels' per scoring window and
- list $L_{a/1se}$ of artifacted episodes defined by 1 second epochs (Figure 3).

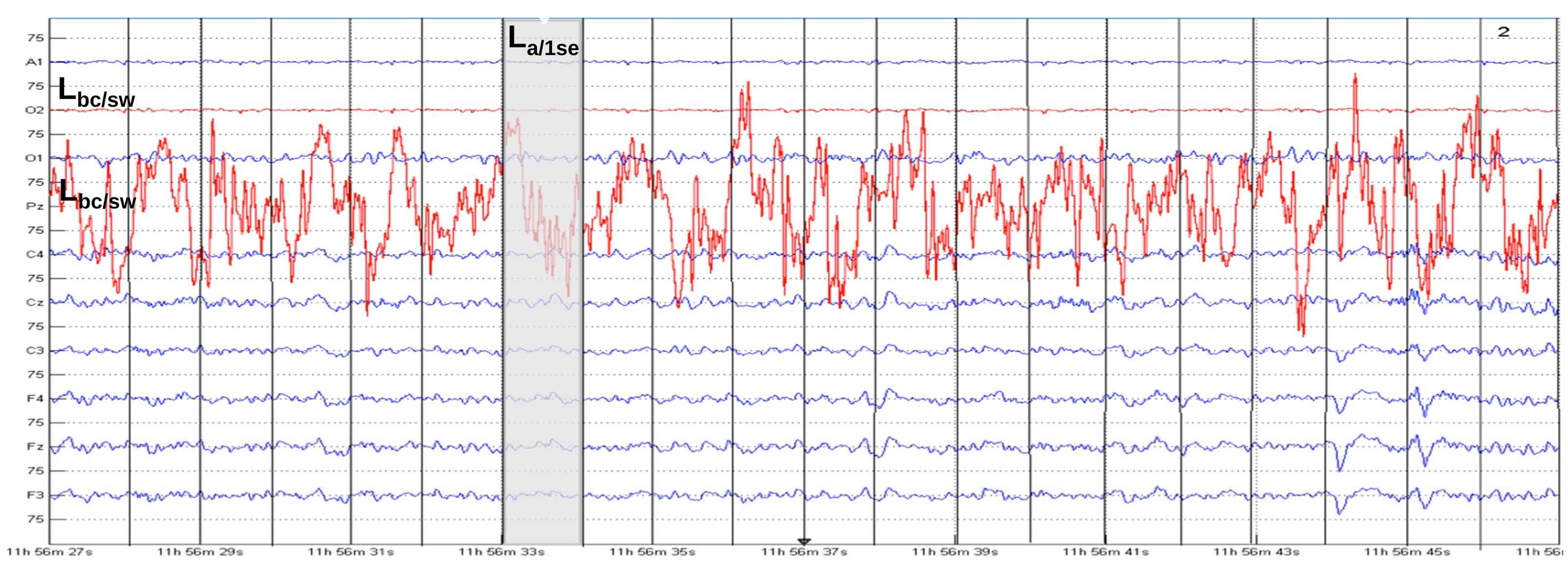

Figure 3: Bad channels (**$L_{bc/sw}$**) and bad epochs (**$L_{a/1se}$**). In this case, **$L_{a/1se}$** is empty, no short artifacts whereas **$L_{bc/sw}$** is composed of two bad channels: 'O_2' and 'P_z'.

Thresholds definition

The thresholds used in the two detection modules are directly derived from data, making the automatic method
- robust in front of inter- and intra- subject variability
- expert independent and reproducible

CONCLUSION

Tests performed over two datasets show that our automatic method is *robust* and *reproducible* (phase I), as well as *more reliable* than different experts between them (phase II). Moreover it works (on a standard PC) *much faster than manual detection* and can used with FASST, an open source software available at http://www.montefiore.ulg.ac.be/~phillips/FASST.html

RESULTS

Data

Data consist in whole night sleep multichannel EEG recording (10-20 system)

Table 1: Dataset Characteristics

	#SR	#SE	Age	Gender	Scoring	
Dataset 0	6	1	21,3±1,7	3m 1g	R&K	Parameters fixation
Dataset I	35	2	21±1,3	13m 12f	R&K	Evaluation: Phase I
Dataset II	4	6	26	1m	AASM	Evaluation: Phase II

= number, SR = total Sleep Recordings, SE = Sleep Experts, Scoring = rules used to score the SR, m = male and f : female.

Statistic parameters

- Agreement measurement **(Pabak)** [4]
- Sensitivity **(S):** # artifacted events detected / total # events detected by the gold standard
- Episode Overlap **(EO):** Overlap (L_{AD}/L_{EX}) averaged for each artifacted episode detected
- False Discovery Ratio **(FDR)**

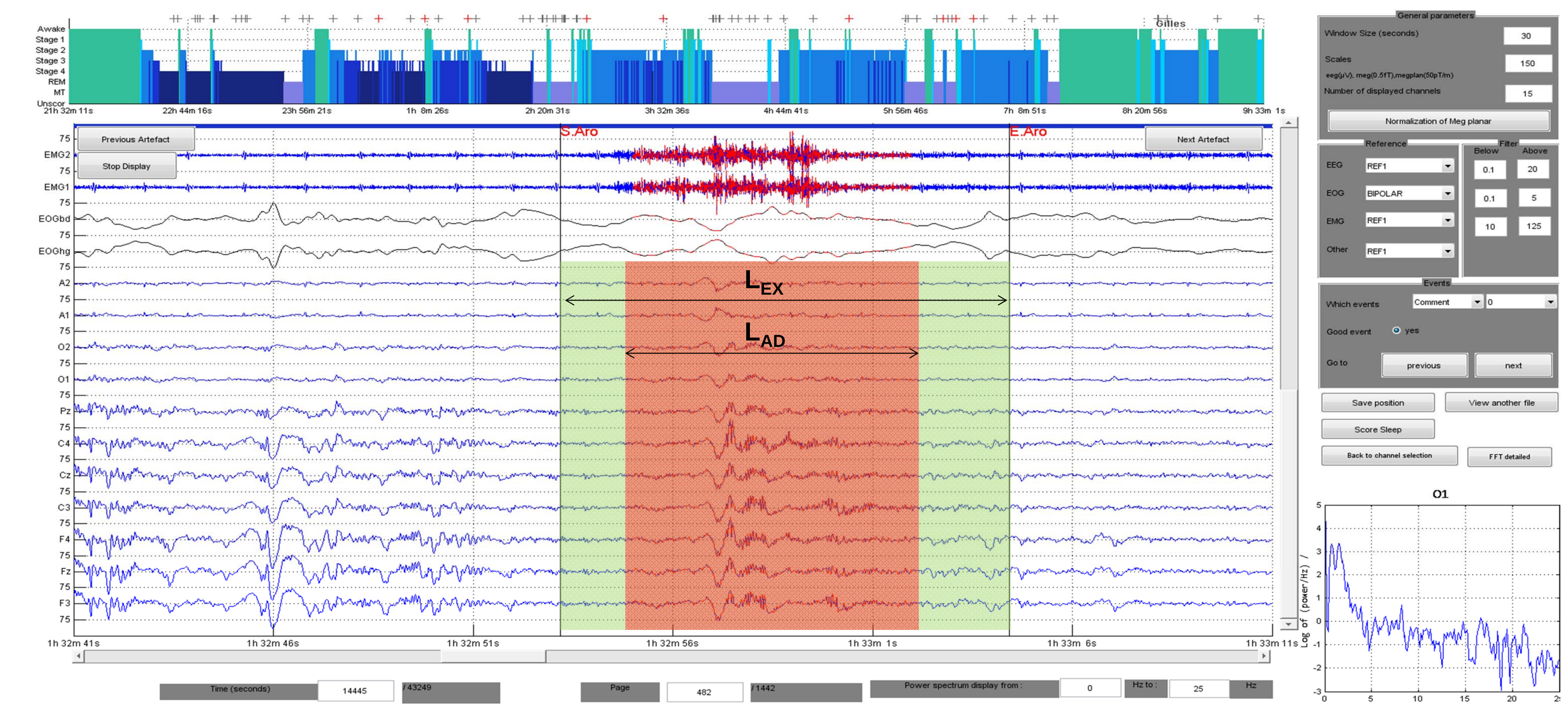

Figure 3: Artifacted episode detected by the expert (L_{EX}, green) and by the automatic detection (L_{AD}, red).

Phase I: Robustness through assessment with 35 sleep recordings (Dataset I)

S1 Artefacts scored independently by an expert (VM) or the "automatic detection" (AD):
- AD is compared to VM (gold standard), Figure 5a (S1).
- AD detects 81,7% of the artifacted episodes considered by VM

S2 All the artifacts detected by AD but not VM have been reviewed and reassessed (false detection or oversight) by another expert (GG), Figure 5a (S2).
- AD reached finally 91,8% sensitivity
- FDR decreased from 37,6% to 22,2%.

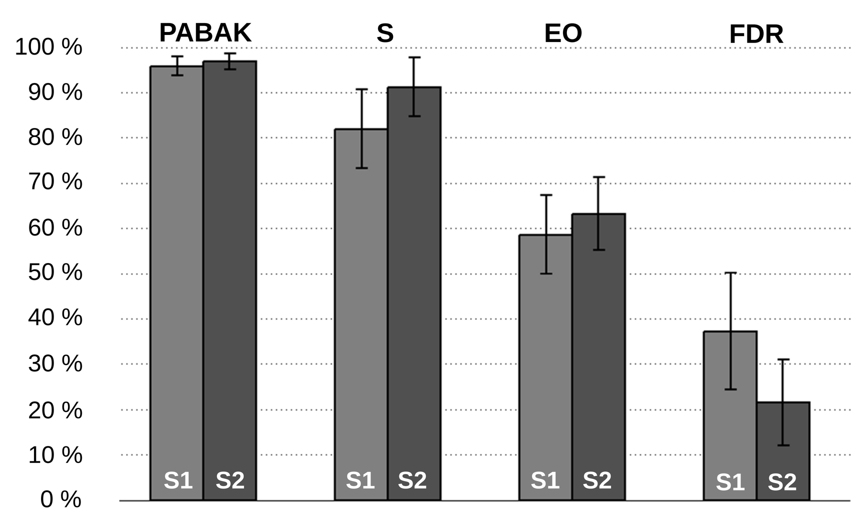

Figure 5a : Phase I (dataset I). The automatic detection (AD) VS the gold standard VM (S1) and AD VS the gold standard after the review by the second expert GG (S2)

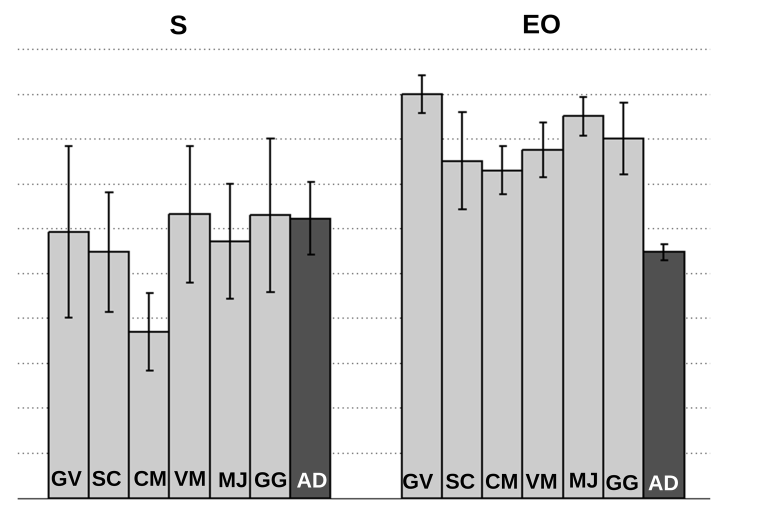

Figure 5b: Phase II (dataset II). 6 experts (GV,SC,CM,VM,MJ,GG) compared to the gold standard created from the union of them and the AD VS these same gold standard

Phase II: Robustness in front of six different sleep experts (Dataset II)

- Gold standard created by the union of 6 experts scoring.
- Each expert scored and the AD compared to the gold standard, Figure 5b.
- AD's artefact detection is similar to that of the best two experts with smaller standard deviation over the 4 recordings (S_{AD} = 62.21% ± 8.1%, S_{VM} = 63.20% ±15.25% and S_{GG}=62.95% ±17.12%)
- Episode overlap is smaller as AD is more conservative (over time) than the experts

REFERENCES

[1] P. Anderer et al., Neuropsychobiology, 40(3):150-157, 1999; [2] E.A. Wolpert, Archives of General Psychiatry, 20(2): 246-247, 1969; [3] C. Iber et al., American Academy of Sleep Medecine, 2007; [4] T. Byrt et al., Journal of clinical epidemiology, 46(5):423-429,1993.

ACKNOWLEDGEMENTS & SPONSORS

Cyclotron Research Centre (CRC); Belgian National Funds of Scientific Research (FNRS); Actions de Recherches Concertées (ARC, ULg) – Fondation Médicale Reine Elisabeth (FMRE); Walloon Excellence in Lifesciences and Biotechnology (WELBIO)