

EOG-based auto-staging: less is more

Christian Berthomier · Marie Brandewinder

Received: 3 November 2014 / Revised: 21 January 2015 / Accepted: 22 January 2015 / Published online: 6 February 2015
© Springer-Verlag Berlin Heidelberg 2015

This issue of *Sleep and Breathing* presents the validation results of a new automated wake/sleep staging method based on EOG activity, developed by Jussi Virkkala from the Finnish Institute of Occupational Health. Classically, the automated method is compared to visual analysis, on an epoch by epoch basis. It reaches a level of global concordance of 88 % with a Kappa of 0.57. In other words, on the 248,696 epochs of the validation dataset, 212,138 were scored correctly in wake/sleep, that is as the human expert did it, and on 36,558 epochs, the two scorings differ.

This level is considered good in literature focusing on evaluating automated methods. It shows as follows:

1. Automated analysis methods are continuously developing [1].
2. Performances increase.

Performance wise, two trends in literature coexist: one aiming at evaluating inter-expert agreement (the percentage of epochs of a recording or a set of recordings for which two human scorers give exactly the same score), when not intra-expert agreement (the percentage of epochs of a recording or a set of recordings for which a human scorer give the same score, when scoring data twice within a given period of time) [2–9]. The other one aiming at evaluating performances of automated methods [10–12], compared to visual analysis. A recent publication demonstrated that on a dataset of 70 recordings, an automated method did not differ more than visual

analysis from a reference scoring [13]. In other words, automated analysis can reach accuracy comparable to visual analysis. These levels of performances are new. Let us remember what automated analysis looked like only a few years ago. There was some vicious circle: automated analysis was disregarded, thus attracted only little attention and effort, and was therefore doomed to be unsatisfactory as it obviously takes talent and time to learn a machine to mimic the extremely complex operations that an experienced scorer does when scoring sleep. The vicious circle seems to turn virtuous as automated analysis becomes a topic of interest where high-profile research teams get involved.

Now that the accuracy of this method is established, let us consider how it works. Indeed, when visual analysis is standardized, automated methods are very diverse: many alternative approaches to PSG are explored.

As stated in the AASM manual, conventional PSG, which is necessary even for the not so simple discrimination between wake and sleep, requires a minimum of seven channels. Here, the proposed montage is respiratory polygraphy+3 sensors (2 EOG+ref). The EOG-based method validated in this paper belongs to a set of methods which all tend to reduce the number of sensors on the patient: actimetry [14], peripheral arterial tone and pulse transit time [15], motion analysis [16], EOG [17], and EEG only [18–24]. One question immediately appears: could the discrepancies observed between visual staging and the validated methods be explained by this reduction in the number of signals? Probably not, as in this study, the automated-visual agreement is nearing the upper-bound of visual-visual (inter-scorer) agreement reached in the above cited literature. Discrepancies with visual analysis, when they reach this level, can be considered as an effect of the irreducible imprecision of scoring sleep—due to the content itself, difficult patterns, transitional epochs—as well as to the

C. Berthomier (✉) · M. Brandewinder
Physip, 6, rue Gobert, 75011 Paris, France
e-mail: C.Berthomier@physip.fr

M. Brandewinder
e-mail: M.Brandewinder@physip.fr

intellectual process—interpretation of rules, limits of human perception, and fatigue.

Good validation results of new scoring methods are not an endpoint, but an invitation to imagine new applications. And in that perspective, as far as sensors are concerned, less is more.

Indeed, these alternative methods reducing the number of sensors pave the way for new diagnostic approaches particularly relevant and interesting for the diagnosis of OSA. OSA is largely and increasingly prevalent, has consequences, and can be treated with real results, even if difficulties should not be underestimated. But when it comes to diagnosis, clinicians face a frustrating alternative between home sleep testing with a relatively simple, cheap and comfortable respiratory polygraphy, which also comes with questionable reliability, and a very reliable but expensive, complex, and considered invasive by patients full PSG. This alternative gives way to pitfalls and suboptimal diagnosis schemes, when a non-conclusive HST ends up either with an additional PSG—at high cost—or an untreated patient, when PSG cannot be performed, for technical, operational, or financial reasons.

At risk of displaying the obvious, why does PV lack reliability? Because it misses crucial information when it comes to diagnosing a sleep-disordered breathing—is the patient awake or asleep? This allows false negative results, when long periods void of respiratory events can indicate an absence of respiratory events—hence a sound sleep—as well as long WASO periods indicating disordered sleep. With methods able to identify wake/sleep periods, new enhanced PV protocols become possible, as a third way between HST and PSG. Identifying automatically wake periods could ease the tedious and imprecise operation of excluding wake portions of the traces based on body position, light, and sound.

One caveat though. As tempting as it may seem, this perspective could come true only if these methods

1. Demonstrate their ability to cope with real-life data, not only data provided by studies where conditions are carefully controlled: patients, variable montage, movement artifacts, and environment artifacts as they can occur when PV is performed in ambulatory mode
2. Keep their rejection level low enough to remain a sensible option in everyday practice

If they fulfill these conditions, innovative methods allowing higher reliability with reduced number of sensors could help in addressing the challenges of contemporary health systems: provide care at a larger scale in order to address increasing prevalence of OSA and its consequences while keeping resources at a reasonable and sustainable level.

Seen as a succession of performance records, the topic of validating automated diagnosis methods is repetitive, arid, and technical. But it is not only technical. Technique is a bad

master but a good servant. It opens perspectives. Behind a good agreement level reached with less sensors, it is possible to see more patients provided with better care, more performing and efficient sleep labs, and more satisfaction for health professionals thanks to higher quality tools.

Conflicts of interest Both authors have ownership and directorship in Physip Company.

References

1. Roebuck A, Monasterio V, Geder E, Osipov M, Behar J, Malhotra A, Penzel T, Clifford GD (2014) A review of signals used in sleep analysis. *Physiol Meas* 35(1):R1–R57
2. Norman RG, Pal I, Stewart C, Walsleben JA, Rapoport DM (2000) Interobserver agreement among sleep scorers from different centers in a large dataset. *Sleep* 23(7):901–908
3. Danker-Hopfe H, Kunz D, Gruber G, Klösch G, Lorenzo JL, Himanen SL, Kemp B, Penzel T, Rösche J, Dorn H, Schlögl A, Trenker E, Dorffner G (2004) Interrater reliability between scorers from eight European sleep laboratories in subjects with different sleep disorders. *J Sleep Res* 13:63–69
4. Moser D, Anderer P, Gruber G, Parapatics S, Loretz E, Boeck M, Kloesch G, Heller E, Schmidt A, Danker-Hopfe H, Saletu B, Zeitlhofer J, Dorffner G (2009) Sleep classification according to AASM and Rechtschaffen & Kales: effects on sleep scoring parameters. *Sleep* 32(2):139–149
5. Rosenberg RS, Van Hout S (2013) The American Academy of Sleep Medicine inter-scorer reliability program: sleep stage scoring. *J Clin Sleep Med* 9(1):81–87
6. Penzel T, Zhang X, Fietze I (2013) Inter-scorer reliability between sleep centers can teach us what to improve in the scoring rules. *J Clin Sleep Med: JCSM: Off Publ Am Acad Sleep Med* 9(1):89–91
7. Magalang UJ, Chen NH, Cistulli PA, Fedson AC, Gislason T, Hillman D, Penzel T, Tamisier R, Tufik S, Phillips G, Pack AI, Investigators SAGIC (2013) Agreement in the scoring of respiratory events and sleep among international sleep centers. *Sleep* 36(4):591–596
8. Redline S, Dean D 3rd, Sanders MH (2013) Entering the era of “big data”: getting our metrics right. *Sleep* 36(4):465–469
9. Zhang X, Dong X, Kantelhardt JW, Li J, Zhao L, Garcia C, Glos M, Penzel T, Han F (2014) Process and outcome for international reliability in sleep scoring. *Sleep Breathing*. doi:10.1007/s11325-014-0990-0
10. Pittman SD, MacDonald MM, Fogel RB, Malhotra A, Todros K, Levy B, Geva AB, White DP (2004) Assessment of automated scoring of polysomnographic recordings in a population with suspected sleep-disordered breathing. *Sleep* 27(7):1394–1403
11. Anderer P, Gruber G, Parapatics S, Woertz M, Miazhyńska T, Klosch G, Saletu B, Zeitlhofer J, Barbanj MJ, Danker-Hopfe H, Himanen SL, Kemp B, Penzel T, Grozinger M, Kunz D, Rappelsberger P, Schlögl A, Dorffner G (2005) An E-health solution for automatic sleep classification according to Rechtschaffen and Kales: validation study of the Somnolyzer 24x 7 utilizing the Siesta database. *Neuropsychobiology* 51(3):115–133
12. Svetnik V, Ma J, Soper KA, Doran S, Renger JJ, Deacon S, Koblan KS (2007) Evaluation of automated and semi-automated scoring of polysomnographic recordings from a clinical trial using zolpidem in the treatment of insomnia. *Sleep* 30(11):1562–1574
13. Kuna ST, Benca R, Kushida CA, Walsh J, Younes M, Staley B, Hanlon A, Pack AI, Pien GW, Malhotra A (2013) Agreement in

- computer-assisted manual scoring of polysomnograms across sleep centers. *Sleep* 36(4):583–589
14. Malhotra A, Younes M, Kuna ST, Benca R, Kushida CA, Walsh J, Hanlon A, Staley B, Pack AI, Pien GW (2013) Performance of an automated polysomnography scoring system versus computer-assisted manual scoring. *Sleep* 36(4):573–582
 15. O'Reilly C, Gosselin N, Carrier J, Nielsen T (2014) Montreal Archive of Sleep Studies: an open-access resource for instrument benchmarking and exploratory research. *J Sleep Res*. doi:10.1111/jsr.12169
 16. García-Díaz E, Quintana-Gallego E, Ruiz A, Carmona-Bernal C, Sánchez-Armengol A, Botebol-Benhamou G, Capote F (2007) Respiratory polygraphy with actigraphy in the diagnosis of sleep apnea-hypopnea syndrome. *Chest* 131:725–732
 17. Pépin JL, Tamisier R, Borel JC, Baguet JP, Lévy P (2009) A critical review of peripheral arterial tone and pulse transit time as indirect diagnostic methods for detecting sleep disordered breathing and characterizing sleep structure. *Curr Opin Pulm Med* 15(6):550–558
 18. De Chazal P, Fox N, O'Hare E, Heneghan C, Zaffaroni A, Boyle P, Smith S, O'Connell C, McNicholas WT (2011) Sleep/wake measurement using a non-contact biomotion sensor. *J Sleep Res* 20:356–366
 19. Levendowski DJ, Popovic D, Berka C, Westbrook PR (2012) Retrospective cross-validation of automated sleep staging using electroocular recording in patients with and without sleep disordered breathing. *Int Arch Med* 5:1–9
 20. Berthomier C, Drouot X, Herman-Stoïca M, Berthomier P, Prado J, Bokar-Thire D, Benoit O, Mattout J, d'Ortho MP (2007) Automatic analysis of single-channel sleep EEG: validation in healthy individuals. *Sleep* 30:1587–1595
 21. Shambroom JR, Fábregas SE, Johnstone J (2012) Validation of an automated wireless system to monitor sleep in healthy adults. *J Sleep Res* 21:221–230
 22. Koley B, Dey (2012) An ensemble system for automatic sleep stage classification using single channel EEG signal. *Comput Biol Med* 42(12):1186–1195
 23. Popovic D, Khoo M, Westbrook P (2014) Automatic scoring of sleep stages and cortical arousals using two electrodes on the forehead: validation in healthy adults. *J Sleep Res* 23(2):211–221
 24. Kaplan RF, Wang Y, Loparo KA, Kelly MR, Bootzin RR (2014) Performance evaluation of an automated single-channel sleep-wake detection algorithm. *Nat Sci Sleep*. doi:10.2147/NSS.S71159