Real-Time Detection of Sleep Arousals with a Head-Mounted Accelerometer

Tugce Canbaz Gumussu *Student Member, IEEE*, Giulia Da Poian,
Silvano Cortesi *Student Member, IEEE*, Walter Karlen *Senior Member, IEEE*

***Abstract*— Wearable electroencephalography (EEG) enables real-time interactions with the sleeping brain in real-life settings. An important parameter to monitor during these interactions are sleep arousals, i.e. temporary increases in EEG frequency, that compose sleep dynamics, but are challenging to detect without delay. We describe the development of an EEG- and accelerometer(ACC)-based sensing approach to detect arousals in real-time. We investigated the ability of these sensing modalities to timely and accurately detect arousals. When evaluated on 6 nights of mobile recordings, ACC had a median real-time delay of 2 s and was therefore better suited for an early detection of arousals than EEG (4.7 s). The detection performance was independent of sleep stages, but worked better on longer arousals. Our results demonstrate that a head-mounted ACC might be a cost-effective and easy-to-integrate solution for arousal detection where short delays are important or EEG signals are not available.**

I. INTRODUCTION

Sleep arousals are transient intrusions of wakefulness into sleep that do not result in behavioral awakening [1]. These sleep microstructures ensure the reversibility of sleep, distinguishing it from coma [2]. Yet, higher rates of arousals can interrupt the continuity of sleep and lead to sleep fragmentation. This fragmentation impacts sleep quality and is associated with an increased risk for health issues in older adults, such as cognitive decline and neurodegenerative disease [3].

In patients with mild cognitive impairment, acoustic stimulation during sleep can induce deeper sleep [4]. Furthermore, our research group demonstrated auditory stimulation in older adults over several weeks in a fully remote setting [5], using an autonomous mobile electroencephalogram (EEG) monitoring and stimulation system [6]. The real-time detection of sleep arousals was essential during the study, as an auditory stimulation on top of an arousal might cause awakening and alter sleep dynamics. Therefore, there is a need for validated arousal detection methods that are accurate, real-time, and also applicable to wearable technology.

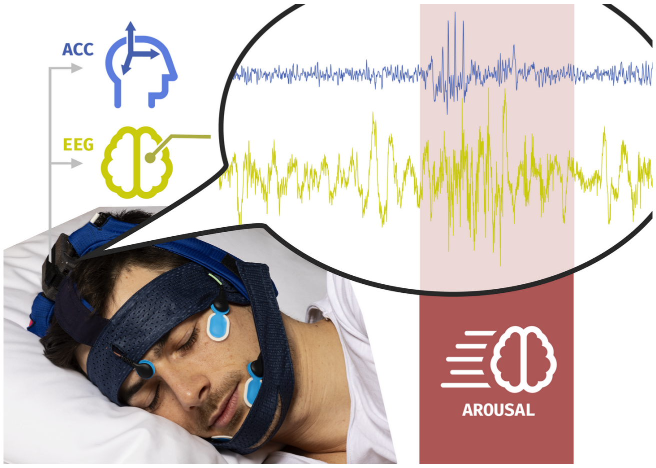

Fig. 1. The objective of this work was to use signals from a head-mounted accelerometer (ACC, blue) to detect sleep arousals and benchmark the real-time performance against the detection from electroencephalography (EEG, yellow) and expert reference labels (red shaded box).

In sleep medicine, arousals are labeled manually during sleep scoring according to the standards of American Academy of Sleep Medicine (AASM) [7]. Arousals are identified based on an abrupt shift of frequency in the EEG, which lasts at least 3 s and has at least 10 s of stable sleep preceding the shift. During rapid eye movement (REM) sleep, at least 1 s increase of electromyography (EMG) activity is additionally required to score arousals. However, this process of manual inspection is cumbersome and time consuming. Therefore, several automated arousal detection algorithms have been designed to reduce the burden of sleep experts. Many available algorithms work with pre-recorded data and replicate the definitions of AASM, notably by identifying the shifts in certain EEG frequency bands [8], [9]. There are also alternative approaches that do not use the confirmation of EMG during REM sleep [10], [11]. However, all these algorithms are based on frequency analysis and therefore, introduce delays that make them less suitable for real-time wearable systems.

Wearable systems offer a large range of alternative sensors to detect sleep arousals. A promising approach is the use of accelerometers (ACCs), as they are low-cost, low-power, versatile, and ubiquitous (Figure 1). For example, Lamprecht et al. investigated the relationship between limb movements and arousals in children [12]. Also, ACCs are useful to predict arousals in mice [13]. Although ACCs have been widely used in sleep monitoring, to our knowledge, they have not been reported for real-time arousal detection in humans.

In this work, we propose a simple algorithm to detect arousals from head movements in real-time and explore the relationship between ACC- and EEG-based arousal de-

This work was conducted as part of the SleepLoop Flagship of Hochschulmedizin Zürich supported by the Schweizerische Hirn Stiftung and the ETH Foundation. This work was funded by the SleepLoop coronavirus bridging grant.

T. Canbaz Gumussu and W. Karlen are with the Institute of Biomedical Engineering, University of Ulm, 89081 Ulm, Germany `tugce.canbaz@uni-ulm.de`

S. Cortesi and G. Da Poian are with ETH Zurich, 8092 Zurich, Switzerland

tection. By comparing an ACC- and an EEG-based algorithm against labels from a human expert, we investigate the temporal relationship between both sensing modalities (Figure 1). This work contributes to the design methodology of novel head-mounted wearables which can be used to interact with the sleeping brain in real-time and study sleep micro-structures.

II. METHODS

A. Algorithm Design

Two real-time algorithms were designed to detect arousal events during REM and non-REM sleep with a head wearable. Each algorithm was based on a different sensing modality: 1) A 3-axis ACC to detect movements related to an arousal, and 2) the frontal channel (Fpz) of a wearable EEG.

1) ACC-Based Arousal Detection ($\mathrm{AD_{ACC}}$*):* Each signal of the three ACC axes was high-pass filtered with an exponentially weighted moving average to remove the effect of the gravitational force. This was followed by a low-pass moving average filter to eliminate high frequency noise. After filtering, the algorithm applied a threshold of 0.025 g to the absolute value of the signal and classified all values above this threshold as head movements. The algorithm then reduced the dimension by applying the disjunction on all axes. After the first passing of the threshold, the cumulative duration of head movement was calculated. If the duration exceeded 0.2 s, the event was classified as an arousal event. The event ended when no head movement was detected within 5 s. When the minimal duration was not attained, the event was ignored. This approach guaranteed a dynamic response time to an arousal, based on the head movement intensity. Thresholds were selected empirically based on 8 nights that were drawn from a different source than the validation data.

2) EEG-Based Arousal Detection ($\mathrm{AD_{EEG}}$*):* To benchmark the performance of the $\mathrm{AD_{ACC}}$ method, we implemented an EEG-based algorithm as EEG frequency shifts are the main source for arousal detection according to the AASM rules [7]. We adapted the algorithm published by [14] to improve real-time performance. First, the EEG signal was pre-processed with a 6th order high-pass Butterworth filter with a cutoff frequency of 0.5 Hz to eliminate the high-power low-frequency components in the signal. A Short-Time Fourier Transform was applied with a sliding window (Hamming, 3 s, step 0.2 s). Power in alpha (8 - 12 Hz) and beta (16 - 40 Hz) bands were calculated for each window and compared to an adaptive threshold. This threshold was a multiple of the mean power over all windows in the past 10 s. Any interval with alpha or beta power higher than the respective threshold was classified as a high-power interval. After a high-power interval was detected, the algorithm calculated the cumulative duration and an arousal event was classified if the event lasted at least 1.6 s within 3 s. The event ended if no high-power interval was detected within 10 s. The main differences compared to the algorithm of [14] were the absence of electrocardiography and EMG signals, less delay-inducing filtering and post-processing, and the required adjustment of thresholds.

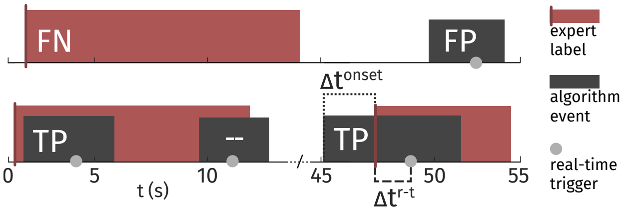

Fig. 2. Illustrative examples of how algorithm-based arousal event detection (grey box) was evaluated. If no event was detected within an expert labeled event (red box), a FN was registered (top). An event that did not overlap with an arousal label was classified as false positive (FP). Only the first event within an arousal label is registered as true positive (TP, bottom). The time difference between label onset and event onset indicated the timing agreement $\Delta \mathrm{t}^{\mathrm{onset}}$ (dotted line). The time difference between onset of label and real-time trigger (circle) indicated real-time delay $\Delta \mathrm{t}^{\mathrm{r-t}}$ (dashed line).

B. Validation

We collected synchronized ACC and EEG data from healthy subjects to validate the algorithms. Data were then analyzed to characterize algorithm properties that are important for real-time detection of arousals.

1) Data Collection: ACC and EEG data were acquired with a wearable biosignal recorder (*MHSL-Sleepband* v3) [15]. It featured a 3-axis ACC (LIS3DH, STMicroelectronics, Geneva, Switzerland) which was configured with a sampling rate of 50 Hz, a 12-bit resolution, and a measurement range of ±4 g. The biopotentials were measured with a 24-bit analog-to-digital converter (ADS1299, Texas Instruments Inc, USA) with a sampling rate of 250 Hz. Both signals were recorded to an embedded SD card together with a common timestamp for accurate signal alignment and then uploaded to a secure data storage using a health middleware (DeviceHub, Leitwert AG, Zurich, Switzerland).

During a study conducted for system configuration and evaluation, volunteers did wear the *MHSL-Sleepband* overnight in a home setting where they self-applied the wearable. Sleep stages and arousals were scored by a sleep expert according the definition of AASM [7], but with 20 s epochs. Data collection and analysis was performed according to the Declaration of Helsinki and the study was approved by the institutional ethics committee of ETH Zurich (ETH-EK2017-N-67).

2) Performance Metrics: We compared the agreement between algorithm event detections and expert arousal labels. Therefore, a true positive (TP) was defined as the first algorithm event detected within the labeled arousal (Figure 2). If an arousal did not overlap with any detected algorithm event, a false negative (FN) was registered. An event detection without the presence of an expert label was a false positive (FP). Sensitivity (SE) and precision (PR) were calculated such as

$$SE = \frac{TP}{TP + FN} \tag{1}$$

$$PR = \frac{TP}{TP + FP}. \tag{2}$$

TABLE I
Arousal Distribution across Sleep Stages

	N1	N2	N3	REM	Total
Labeled arousals	14	119	10	93	236
Mean arousals per night	2.3	19.8	1.7	15.5	39.3

TABLE II
Classification Performance of Algorithm Events

	TP	FN	FP	SE	PR	FNR	FDR
ACC	192	44	114	81.4%	62.8%	18.6%	37.3%
EEG	201	35	176	85.2%	53.3%	14.8%	46.7%

Thus, SE was used as an indicator for the correct identification of arousals, while PR indicated how conservative the algorithm was. In addition, to analyze the impact of arousal duration on detection performance, we compared the duration distributions of TP and FN. Arousals might have different phenotypes across sleep stages that might impact detection. Therefore, we compared the false negative rate (FNR) and the false discovery rate (FDR) across different sleep stages. For this analysis, we assumed stable sleep preceding the arousal. Therefore, to determine the sleep stage of the arousal, we selected the sleep stage of the epoch immediately before the epoch containing the arousal. This prevented the mislabeling with a post-arousal sleep stage and corresponded to a real-time system where only previous epoch information would be available. To evaluate the timing agreement of an arousal onset, we calculated the onset difference between the labeled arousal and the event ($\Delta t^{\mathrm{onset}}$) for all TP (Figure 2). Finally, to evaluate the real-time performance of $\mathrm{AD}_{\mathrm{ACC}}$ and $\mathrm{AD}_{\mathrm{EEG}}$, we measured the real time delay $\Delta t^{\mathrm{r-t}}$, which was the time difference between the labeled arousal onset and the timestamp where the algorithms triggered an event detection.

We performed the algorithm development and analysis with MATLAB R2021b (The MathWorks, Massachusetts, USA).

III. Results

We obtained 237 labeled reference arousals (Table I) in 409 hours of sleep, which originated from 6 recordings from 3 subjects (mean age: 29.6 ± 2 y, all female). One arousal of duration 50 s was considered an outlier and excluded. Sleep stage N2 contained the highest number of arousals (119). Labeled arousals had a mean duration of 10.72 s (SD 5.78 s). Performance metrics had similar range for both algorithms.

SEs were 81.4% and 85.2% whereas PRs were 62.8% and 53.3% for $\mathrm{AD}_{\mathrm{ACC}}$ and $\mathrm{AD}_{\mathrm{EEG}}$, respectively (Table II). The FNR was above 13% for both modalities in all sleep stages except in N3, where all arousals were detected by $\mathrm{AD}_{\mathrm{EEG}}$ (Figure 3). The FDR was lower for $\mathrm{AD}_{\mathrm{ACC}}$ which was due to better rates in N2 and REM sleep (Figure 4). For both $\mathrm{AD}_{\mathrm{ACC}}$ and $\mathrm{AD}_{\mathrm{EEG}}$, the median arousal durations of FN (5.4 and 5.7 s) were lower than the ones of TP (10.8 and 10.5 s) (Figure 5). The median timing agreements between event and arousal onsets were $\widetilde{\Delta t}^{\mathrm{onset}}_{\mathrm{ACC}} = 1.26\,\mathrm{s}$ and $\widetilde{\Delta t}^{\mathrm{onset}}_{\mathrm{EEG}} = -0.02\,\mathrm{s}$ (Figure 6).

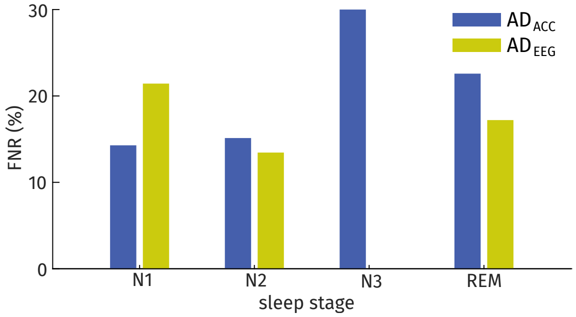

Fig. 3. False negative rate (FNR) across sleep stages. FNR is calculated as FNs over the total number of labeled arousals (FN+TP) in a sleep stage.

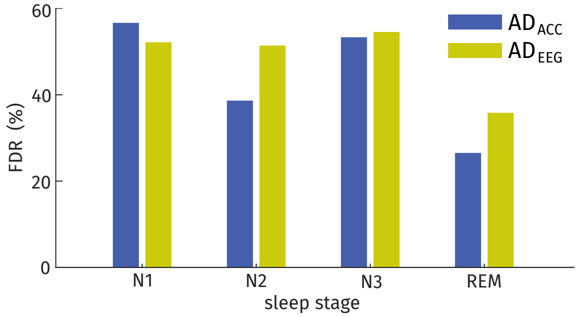

Fig. 4. False discovery rate (FDR) across sleep stages. FDR is calculated as FPs over the total number of algorithm events (FP+TP) in a sleep stage.

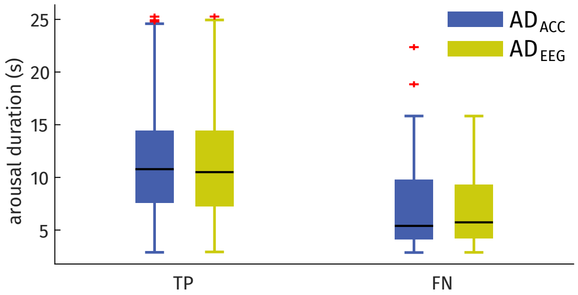

Fig. 5. Distributions of arousal durations derived from expert labels for correct (TP) and missed (FN) detections of $\mathrm{AD}_{\mathrm{ACC}}$ and $\mathrm{AD}_{\mathrm{EEG}}$. FN detections tended to occur in shorter arousals than the TP detections.

The median real-time delays were $\widetilde{\Delta t}^{\mathrm{r-t}}_{\mathrm{ACC}} = 1.99\,\mathrm{s}$ and $\widetilde{\Delta t}^{\mathrm{r-t}}_{\mathrm{EEG}} = 4.66\,\mathrm{s}$ (Figure 7).

IV. Discussion

We proposed a real-time ACC-based algorithm to detect arousals with head movements during sleep and compared the performance with an EEG-based method. Both methods achieved similar performance. Although the event onset times of $\mathrm{AD}_{\mathrm{EEG}}$ had better agreement with the labeled arousals than $\mathrm{AD}_{\mathrm{ACC}}$, the mean real-time delay was higher, which highlighted the early detection potential of $\mathrm{AD}_{\mathrm{ACC}}$.

The head-mounted ACC showed the capability for arousal detection and SE and PR was not inferior than $\mathrm{AD}_{\mathrm{EEG}}$. However, the discrepancy in identifying the arousal onsets

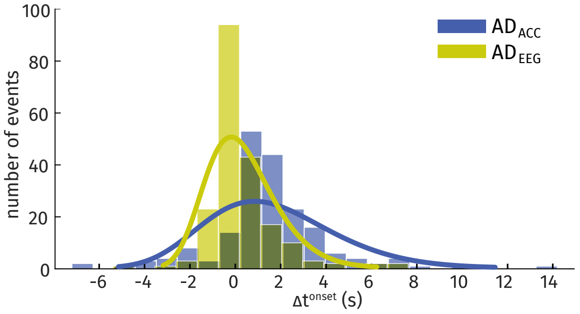

Fig. 6. Distribution of timing agreement between the onset of expert labeled arousal and algorithm TP event Δt^{onset}. The histogram is overlayed with a generalized extreme value approximation.

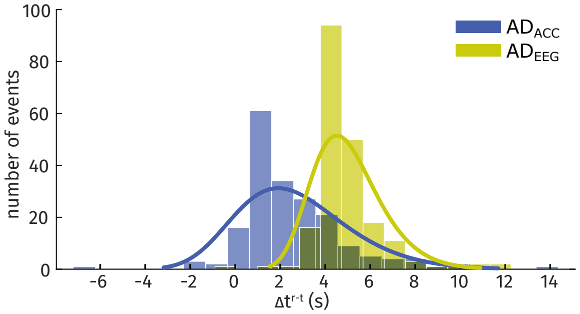

Fig. 7. Distribution of real-time detection delays Δt^{r-t} of AD_{ACC} and AD_{EEG}. Δt^{r-t} is the time difference between the onset of expert labeled arousal and the timestamp at which an algorithm triggered a TP event.The histogram is overlayed with a generalized extreme value approximation.

suggests that arousal information is present in the EEG signal earlier than in the ACC signal. This could be an indication that head movements are induced by arousals and hence start after the arousal onset. However, EEG is also the main source for human scoring and therefore, the reference labels are not independent from AD_{EEG}. Nevertheless, this could lead to sensor fusion algorithms that feature both, short detection delays and exact onset timing. Additional post-processing could indeed lead to better PR [16]. Whether this fusion could also increase the overall performance and improve SE and PR further, remains to be demonstrated.

The real-time performance of the AD_{ACC} algorithm was superior to AD_{EEG}, with AD_{ACC} detecting arousals in average $2.7\,s$ earlier than AD_{EEG}. Still, AD_{ACC} is not entirely delay-free, which needs to be taken into consideration when designing real-time critical systems. Lima et al. observed that head-mounted ACC can predict arousals ahead of EEG in mice [13], but we could not replicate this observation in human subjects. We must speculate that this is due to different sleep physiology.

For this study we purposely designed an algorithm of low complexity to gain insights into the real-time potential of each sensing modality, as each processing step introduces delays or masks challenges. As already introduced above, there are numerous opportunities to improve performance with additional processing steps or by optimizing configuration parameters. Furthermore, our evaluation was based on only 3 subjects, with limited age range and single sex. To generalize our findings, additional validation on larger and more diverse data sets is required.

This work presents a sleep arousal detection algorithm using head movements. The proposed ACC-based method showed a similar performance compared to the reference EEG-based approach and most importantly, it was earlier in detecting an arousal in real-time. Therefore, we believe ACC is a promising sensor modality to detect arousals in real-time or in applications where EEG signals are not available.

V. ACKNOWLEDGEMENTS

We thank all the participants for contributing to this research and all the SleepLoop consortium members for valuable discussions, feedback, and sharing data. We thank Ian Clark for scoring sleep.

REFERENCES

[1] M. H. Bonnet et al., “A Preliminary Report from the Sleep Disorders Atlas Task Force of the American Sleep Disorders Association,” Sleep, vol. 15, no. 2, pp. 174–184, Mar. 1992, doi: 10.1093/sleep/15.2.174.
[2] P. Halász, M. Terzano, L. Parrino, and R. Bódizs, “The nature of arousal in sleep,” Journal of Sleep Research, vol. 13, no. 1, pp. 1–23, 2004, doi: 10.1111/j.1365-2869.2004.00388.x.
[3] A. S. P. Lim et al., “Sleep Fragmentation and the Risk of Incident Alzheimer’s Disease and Cognitive Decline in Older Persons,” Sleep, vol. 36, no. 7, pp. 1027–1032, Jul. 2013, doi: 10.5665/sleep.2802.
[4] N. A. Papalambros et al., “Acoustic enhancement of sleep slow oscillations in mild cognitive impairment,” Annals of Clinical and Translational Neurology, vol. 6, no. 7, pp. 1191–1201, 2019.
[5] C. Lustenberger et al., “Auditory deep sleep stimulation in older adults at home: a randomized crossover trial,” Communications Medicine, vol. 2, no. 1, p. 30, Dec. 2022, doi: 10.1038/s43856-022-00096-6.
[6] M. L. Ferster, C. Lustenberger, and W. Karlen, “Configurable Mobile System for Autonomous High-Quality Sleep Monitoring and Closed-Loop Acoustic Stimulation,” IEEE Sensors Letters, vol. 3, no. 5, pp. 1–4, May 2019, doi: 10.1109/LSENS.2019.2914425.
[7] R.B. Berry et al.; for the American Academy of Sleep Medicine, ”The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications Version 2.4,” Darien, IL: American Academy of Sleep Medicine; 2017.
[8] F. de Carli, L. Nobili, P. Gelcich, and F. Ferrillo, “A Method for the Automatic Detection of Arousals During Sleep,” Sleep, vol. 22, no. 5, pp. 561–572, Aug. 1999, doi: 10.1093/sleep/22.5.561.
[9] T. Sugi, F. Kawana, and M. Nakamura, “Automatic EEG arousal detection for sleep apnea syndrome,” Biomedical Signal Processing and Control, vol. 4, no. 4, pp. 329–337, Oct. 2009.
[10] R. Agarwal, “Automatic Detection of Micro-Arousals,” in 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, Jan. 2005, pp. 1158–1161. doi: 10.1109/IEMBS.2005.1616628.
[11] P. Gouveia, R. Oliveira, and A. Rosa, “Sleep apnea related micro-arousal detection with EEG analysis,” in Proceedings of the BioEng 2003 7th Portuguese Conference on Biomedical Engineering, Lisbon, Portugal., 2003, vol. 15.
[12] M. L. Lamprecht, A. P. Bradley, G. Williams, and P. I. Terrill, “Temporal associations between arousal and body/limb movement in children with suspected obstructed sleep apnoea,” Physiol. Meas., vol. 37, no. 1, pp. 115–127, Jan. 2016, doi: 10.1088/0967-3334/37/1/115.
[13] G. Z. dos S. Lima et al., “Predictability of arousal in mouse slow wave sleep by accelerometer data,” PLOS ONE, vol. 12, no. 5, p. e0176761, May 2017, doi: 10.1371/journal.pone.0176761.
[14] I. Fernández-Varela, D. Alvarez-Estevez, E. Hernández-Pereira, and V. Moret-Bonillo, “A simple and robust method for the automatic scoring of EEG arousals in polysomnographic recordings,” Computers in Biology and Medicine, vol. 87, pp. 77–86, Aug. 2017.
[15] M. L. Ferster et al., “Benchmarking real-time algorithms for in-phase auditory stimulation of low amplitude slow waves with wearable EEG devices during sleep,” IEEE Transactions on Biomedical Engineering, vol. 69, no. 9, pp. 2916–2925, 2022.
[16] D. Alvarez-Estevez and I. Fernández-Varela, “Large-scale validation of an automatic EEG arousal detection algorithm using different heterogeneous databases,” Sleep Medicine, vol. 57, pp. 6–14, May 2019, doi: 10.1016/j.sleep.2019.01.025.