Machine Learning detection of arousals from at-home sleep EEG device demonstrates high reliability compared to Experts

Franz Fürbass, Jack Sampiere, Jay Pathmanathan, Antoine Guillot, Alexander M. Chan
Beacon Biosignals, Boston, MA, USA

BEACONBIOSIGNALS

P-01 Board 18, #356

An at-home sleep EEG device with an associated machine learning algorithm is able to detect arousals with a high epoch-level agreement and provides an reliable estimate of arousal index for people with disturbed sleep

Introduction

Cortical arousals provide insights into sleep pathologies and sleep quality through relationships to daytime impairment, respiratory event related arousals (RERAs), and hyperarousal patterns in patients with various sleep and psychiatric conditions. Scoring arousals from in-lab polysomnography (PSG) by expert readers is the gold standard, but is burdensome, labor-intensive, and subject to significant inter/intra-rater variability. Here, we compare the detection of arousals from PSG to Waveband, an FDA-cleared, at-home, dry-electrode EEG device that captures cortical arousals and other sleep metrics.

Methods

Population: 88 participants in total, 50 participants enrolled in a sleep clinic having confirmed sleep disorders, 38 participants from a community insomnia cohort with self-reported insomnia symptoms, ages 23-66, 73% female

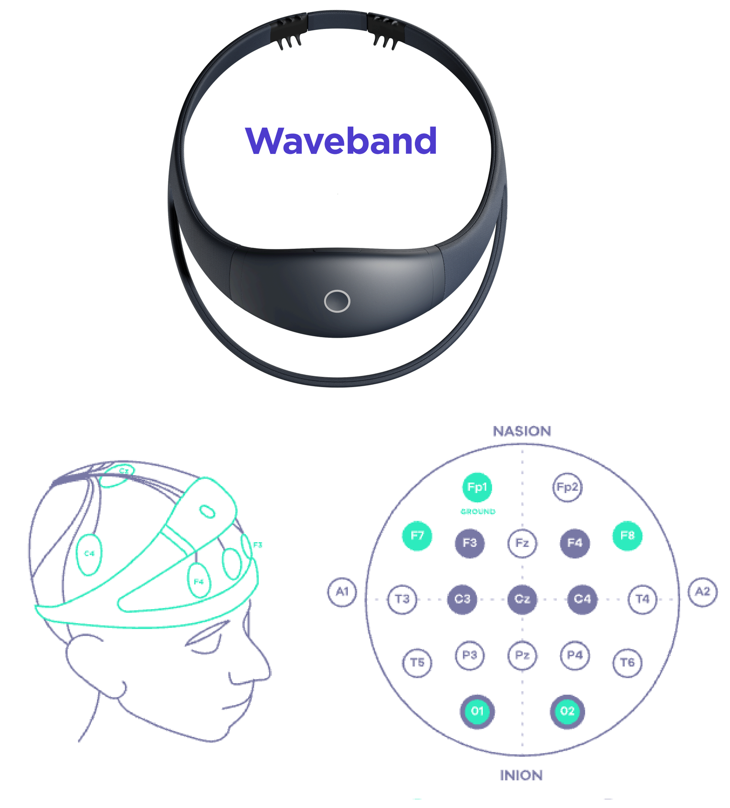
Study Designs:

- One overnight in-lab polysomnography (PSG) while also wearing the Waveband
- Five Registered PSG Technologists (RPSGT) scored arousals in PSG only
- Arousals on Waveband were detected by ArousalML, a machine-learning algorithm trained on over 15k nights

Evaluations:

- Intraclass correlation coefficient (ICC) of the model's Arousal index (ArI) to the scorers' average ArI
- 30-second epochs with and without arousals to the scorers' majority vote (at least 3 out of 5 scorers voting for an arousal in the same epoch) to compute Cohen's Kappa

Purpose: To evaluate performance of ArousalML on subjects with disturbed sleep


Waveband
NASION
INION

Nights were collected in-lab with polysomnography (PSG) and the at-home, dry-electrode EEG device (Waveband). Experts scored arousals on PSG. ArousalML detected arousals on Waveband. The figure shows an arousal in the signals of both devices.

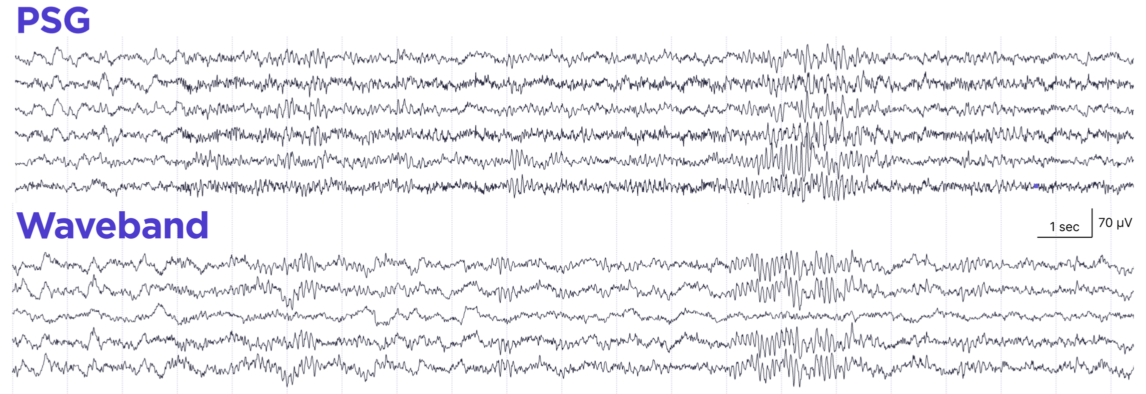
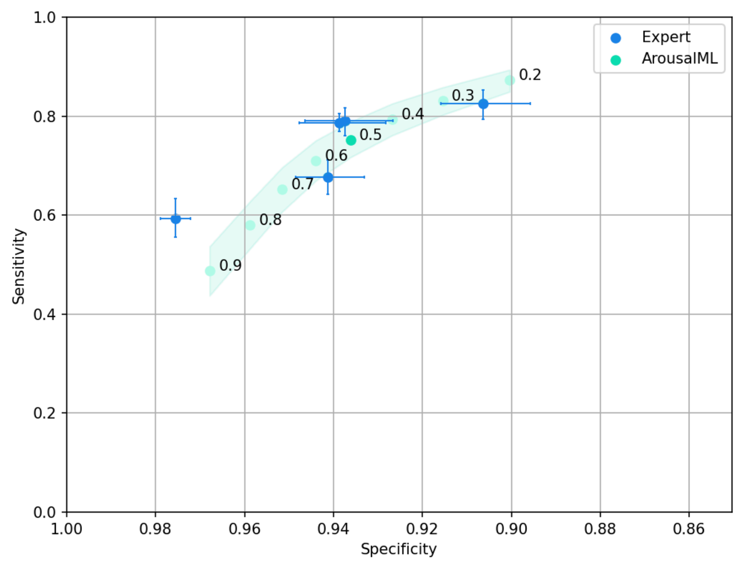

PSG
Waveband
1 sec
70 µV

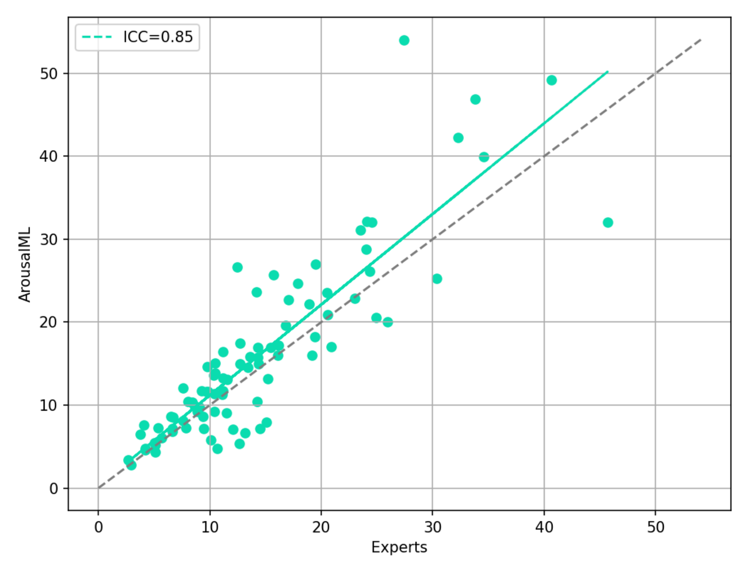
Figure 1: Arousal on simultaneous Waveband and PSG recording

Results

[A] Epoch-level Agreement


Expert
ArousalML
Sensitivity
Specificity

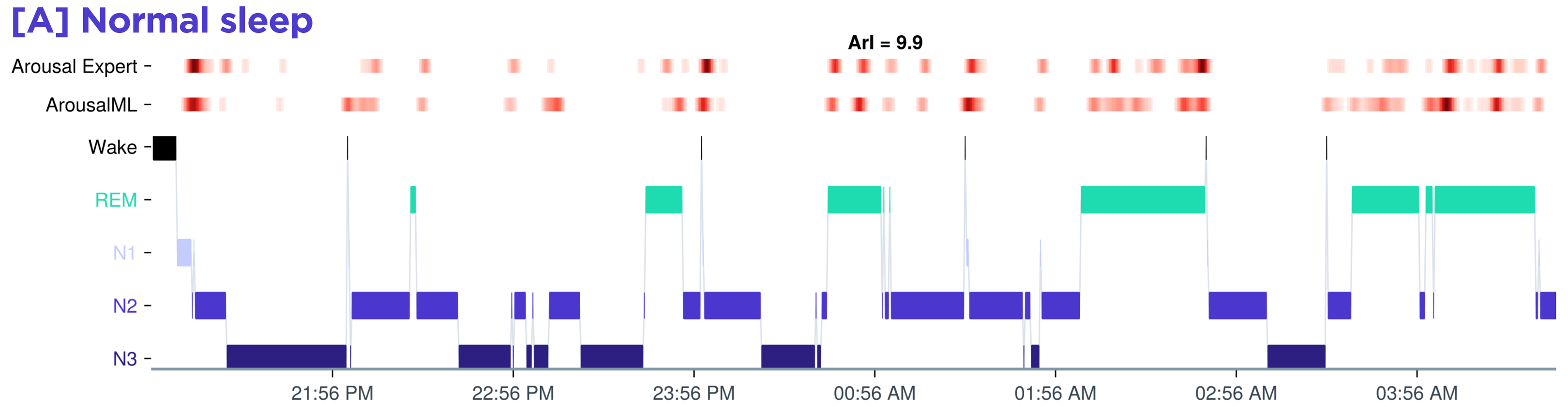
[B] Arousal Index Agreement


ICC=0.85
ArousalML
Experts

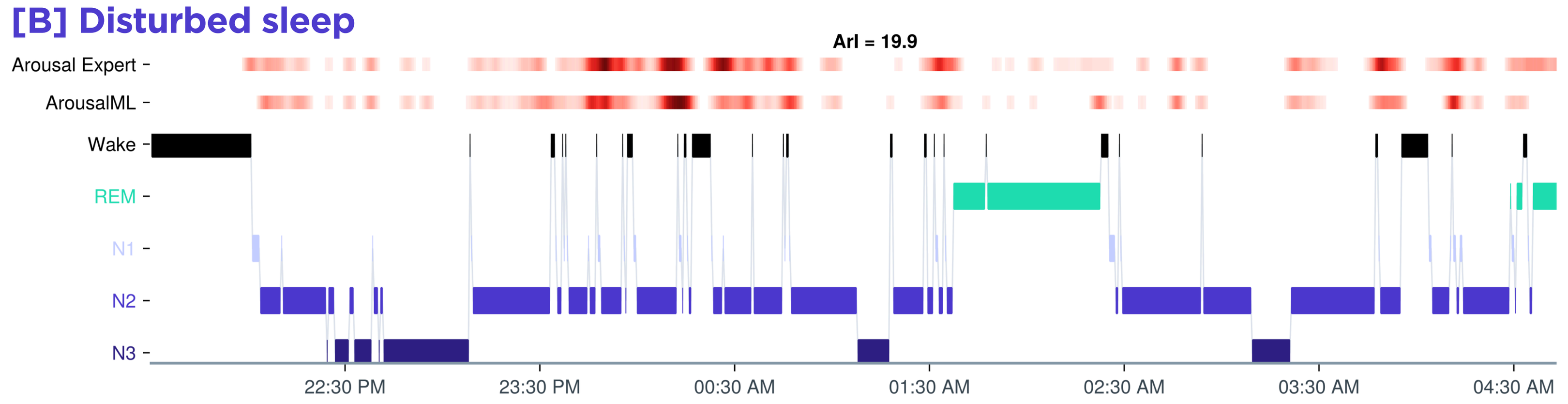
[A] Waveband's epoch-level kappa of 0.58 (95% CI:0.55-0.60) exceeded most experts' agreement (0.47 to 0.59). Different thresholds for the machine learning model are indicated by light green dots. Expert to other expert's agreement shown in blue. [B] Comparing the Waveband ArI to the average ArI of all experts resulted in an ICC of 0.85 (95% CI: 0.77-0.91). A tendency towards lower expert-to-expert and model-to-expert agreement was found for nights with ArI larger than 25.

Figure 2: Comparing performance of ArousalML on Waveband to Experts on PSG

[A] Normal sleep


ArI = 9.9
Arousal Expert
ArousalML
Wake
REM
N1
N2
N3
21:56 PM
22:56 PM
23:56 PM
00:56 AM
01:56 AM
02:56 AM
03:56 AM

[B] Disturbed sleep


ArI = 19.9
Arousal Expert
ArousalML
Wake
REM
N1
N2
N3
22:30 PM
23:30 PM
00:30 AM
01:30 AM
02:30 AM
03:30 AM
04:30 AM

Example hypnograms with arousals densities of the best performing expert (kappa=0.59 compared to consensus of other experts) and ArousalML shown in red. [A] Hypnogram of a night with normal sleep architecture and an arousal index of 9.9. [B] Hypnogram of a night with sleep disturbances, showing disrupted sleep in the first half of the night and an increased arousal index of 19.9 events per hour.

Figure 3: Waveband derived hypnograms and arousal interaction

Summary

- Mean ICC of Arousal Index (ArI) between pairs of human scorers was 0.77 (SD=0.104, range: 0.61-0.92).
- Mean ICC between Waveband's ArI and individual scorers was 0.78 (SD=0.067, range: 0.70-0.86).
- Comparing the Waveband ArI to the average ArI of all scorers resulted in an ICC of 0.85 (95% CI: 0.77-0.91).
- Waveband's epoch-level kappa of 0.58 (95% CI:0.55-0.60) exceeded most experts' agreement (0.47 to 0.59).
- Our results indicate an overall high reliability of automated derived arousal index endpoints compared to the average human expert.
- Nights with an overall high ArI (>25) show a tendency towards lower expert-to-expert and model-to-expert agreement (Figure 2B).

Conclusions

→ An at-home sleep EEG device with an associated machine learning algorithm can detect arousals with a high epoch-level agreement with expert scorers

→ A reliable estimate of arousal index in individuals with disrupted sleep is provided

→ Comprehensive multi-centre dataset including 88 participants and gold standard annotations from 5 experts

→ These findings demonstrate the ability to objectively quantify arousal index across disease conditions

→ Waveband potentially enables more efficient evaluation of therapies and future diagnostic applications in the at-home setting

Contact
Franz Fürbass, PhD
Research Scientist, Beacon Biosignals