Pupil size reveals arousal level dynamics in human sleep

Manuel Carro-Domínguez[1], Stephanie Huwiler[1], Stella Oberlin[1], Timona Leandra Oesch[1], Gabriela Badii[2], Anita Lüthi[3], Nicole Wenderoth[1,4,5*], Sarah Nadine Meissner[1*], Caroline Lustenberger[1,5,6*]

[1] Neural Control of Movement Lab, Institute of Human Movement Sciences and Sport, Department of Health Sciences and Technology, ETH Zurich, 8092 Zurich, Switzerland

[2] FMH Ophthalmology, Zurich, Switzerland

[3] Department of Fundamental Neurosciences, University of Lausanne, Lausanne, Switzerland

[4] Future Health Technologies, Singapore-ETH Center, Campus for Research Excellence and Technological Enterprise (CREATE), Singapore

[5] Neuroscience Center Zurich (ZNZ), University of Zurich, ETH Zurich, Zurich, Switzerland

[6] Center of Competence Sleep & Health Zurich, University of Zurich, Zurich, Switzerland

*These authors have contributed equally to this paper and are shared last authors, correspondence can be addressed to: Caroline Lustenberger, Neural Control of Movement Lab, Department of Health Sciences and Technology, ETH Zurich, Zurich, 8092, Switzerland.

Email: caroline.lustenberger@hest.ethz.ch

Keywords: pupil size, sleep, arousal, infraslow fluctuations, auditory stimulation, sleep spindle, K-complex, heart rate, spectral slope

Competing interests

The authors have declared no competing interests.

Abstract

Changes in arousal level are a determining characteristic of healthy and pathological sleep in humans, but how they are conventionally measured lack sensitivity and temporal precision. Recent animal research has revealed the intricate dynamics of arousal levels that are potentially crucial for maintaining proper sleep resilience and memory consolidation. To investigate whether similar dynamics exist in humans, we measured pupil size, an established indicator of arousal levels, for the first time during overnight sleep and tested the cortical response to auditory stimulation at different levels of arousal. Here we show that pupil size dynamics correspond to sleep macrostructure and microstructural events, while exhibiting an inverse relationship with fluctuations of sleep spindles, a marker of sleep resilience and memory consolidation. Additionally, pupil size prior to auditory stimulation influences its effectiveness to enhance delta power, a marker of several restorative and regenerative functions of sleep. These pupil size dynamics provide novel insights into the interplay between arousal levels and sleep, opening avenues for future research and clinical applications in diagnosing and treating pathological sleep associated with abnormal arousal levels.

Introduction

Sleep is typically described as a state of low arousal level and disconnection from the environment, characterized by reduced responsiveness to external stimuli. Previous work in rodents has shown that arousal levels are high during alert wakefulness and are strongly suppressed during non-rapid eye movement (NREM) and rapid eye movement (REM) sleep[1–4]. However, little is known about whether similar arousal level dynamics exist also during human sleep and how they relate to sleep macro- and microarchitecture. To answer this question and test several predictions from the rodent literature in human sleep, we used pupillometry (Figure 1), a well-accepted method for estimating arousal levels during wake, together with polysomnography (PSG) and electrocardiography (ECG). Recently it has been discovered in rodents that arousal levels exhibit infraslow fluctuations (ISFs) during NREM sleep, representing vigilant and consolidated periods of sleep. The Locus Coeruleus (LC), a small brainstem nucleus and the main source of noradrenergic transmission (NA), seems to be causally mediating these infraslow arousal level fluctuations[1,3].

In mice, ISFs in arousal level occur with a periodicity of approximately 30 seconds to 1 minute[3,5,6]. They show an anti-correlation with spindle activity in the cortex (sigma oscillations in the 12-16 Hz range) and a positive correlation with heart rate (HR). Specifically, consolidated periods of low arousal levels associated with low LC-NA activity coincide with an increased likelihood of sleep spindle clusters, which are believed to protect sleep from external disturbances, and a decrease in HR[3,5–7]. Vigilant periods of high arousal levels associated with high LC-NA activity, by contrast, coincide with low sleep spindle probability and elevated HR. Such periods of elevated arousal levels during NREM sleep might be essential for survival in mice due to the environmental threats present at all times. Furthermore, recent studies in rodent models indicate that the role of arousal level dynamics during sleep goes beyond changes in vigilance and contributes fundamentally to the macro- and micro-architecture of sleep, coordination of the brain and autonomic system, and memory consolidation potentially playing a fundamental role in restorative processes of sleep (for review, see [3]).

Yet, it is unclear whether ISFs of arousal levels are preserved in human sleep. Initial evidence indicates that sleep spindles in NREM sleep exhibit an ISF in humans similar to the rodent model[8,9]. However, establishing a systematic link between these spindle fluctuations and changes in arousal level most likely driven by the LC has been challenging due to the lack of appropriate methods for estimating LC activity during human sleep. Here, we propose the use of pupillometry as an alternative approach to overcome this barrier. It is well-established that the size of the pupil's eye can serve as an indirect marker of arousal levels when light conditions are kept constant[10]. Even though several neuromodulatory systems have been associated with these non-luminance-related changes in pupil size, the LC-NA system has so far received the strongest evidence[11–17]. Furthermore, pupil size and cortical activity have been measured simultaneously during sleep in mice, revealing ISFs of pupil size during NREM sleep that are anti-correlated with electrocortical activity in the spindle range [7]. This finding aligns with previous studies showing LC-NA activity fluctuating with spindle ISFs and indirectly suggests that pupillometry is a valuable tool for tracking arousal levels during sleep and providing a proxy of LC-NA activity during NREM sleep phases[3,5,6,13,18].

Even though pupillometry is a standard research tool in human participants during wakefulness, there has virtually been no data published reporting human pupil dynamics during sleep. We have therefore developed a safe method for measuring changes in pupil size during sleep using a conventional, low-dose infrared system. Here we synchronized our approach with PSG and ECG. We were able to demonstrate that arousal levels fluctuate in human sleep as revealed by pupil size dynamics across and within sleep stages and established their presence surrounding markers of cortical arousal, sleep electroencephalography (EEG) micro-architectures such as spindles and K-complexes, and HR. Using auditory stimulation during NREM sleep, we further demonstrate that the brain response to administered acoustical stimuli changes depending on arousal levels indicated by differences in pupil size.

Results

We developed a new approach for measuring pupil size together with PSG and ECG in healthy participants (n=17) during sleep (Figure 1A-B, Supplementary Video). Portable pupillometry via a low-dose infrared system was applied to the right eye, which was kept open by applying three strips of tape on the upper eyelid and one, wider, strip of tape on the lower eyelid. To reduce discomfort, the eye was taped open without exceeding the subject's palpebral fissure during wake. To prevent eye dryness, vitamin A-containing eye ointment was applied to the eye immediately before the eye was covered by a transparent eye bandage. Eye bandages are routinely used in nocturnal lagophthalmos patients who cannot close their eyelids during sleep to protect the eye from dust and mechanical damage[19]. All participants could tolerate having their eyes taped open and were able to fall asleep when the room was completely dark except for the infrared light source emitted by the pupillometer. In the debriefing, 2 out of 17 participants reported waking up due to the unfamiliarity of sleeping with one eye open. Participants slept in the lab for 7.5 hours with sleep onset having been adjusted to each individual's typical bedtime (around 10-11pm). Auditory stimulation was applied during NREM sleep throughout the night. PSG and ECG were acquired during the whole sleep session but pupillometry was limited to 4 hours, after which the experimenter gently woke the participant up, and removed the eye-tracking equipment. This was a precaution to avoid unwanted side effects such as drying of the right eye. Note that none of our participants reported negative effects, suggesting that longer measurement durations would be achievable. The videos generated by the pupillometer were processed with DeepLabCut[20,21] to detect pupil size, gaze, and some anatomical features of the eye. Since the eye was taped open without exceeding the subject's palpebral fissure during wake, there were some invalid video frames when the pupil was occluded by the eyelids. For 15 participants, pupil size could be calculated for at least 44.61% of the video recording (on average, mean±sd, 75.69±21.90% valid samples, n=15). Two participant's pupils could not be accurately tracked for a large portion of the recording due to head movements resulting in shifting the pupillometer (69.82% invalid samples, n=1) and tape dislodgement resulting in an insufficiently large palpebral fissure (70.37% invalid samples, n=1).

Pupil size and heart rate reflect dynamic changes in arousal levels at different moments of sleep

Arousal levels can be used to define sleep depth and have been shown to depend on the sleep macro-architecture. To test this basic principle, we determined whether (i) pupil size (Figure 1C), an established indicator of central arousal levels during wake (ii) EEG spectral slope (Figure 1D), a neural marker of cortical arousal[22], and (iii) HR (Figure 1D), a marker for cardiovascular activation, differed across sleep stages. Pupil size, measured as the ratio between the pupil radius and the iris radius , was largest when being awake and significantly decreased for NREM stage 1 (N1), NREM stage 2 (N2), NREM stage 3 (N3), and REM sleep (Figure 1F, $F(4, 57.43) = 168.90$, $p < .001$). The strongest changes were observed from wake to N1 and from N2 to N3, while pupil size was consistently small for N3 and REM sleep (Supplementary Table 1). These results show for the first time pupil changes across sleep stages in humans and are consistent with previous pupillometry findings in sleeping rodents[7,23].

Spectral slope was calculated in 30-s segments in the 30-45Hz range of the Fpz electrode[22]. Consistent with previous studies[22,24,25], we observed a significant decrease from wake to REM sleep in a nearly linear fashion (Figure 1G, $F(4, 68) = 41.37$, $p < .001$, Supplementary Table 2). Repeated measures correlation analyses revealed a significant positive correlation between pupil size and spectral slope across ($R = 0.34$, 95%CI = [0.313, 0.366], N=17) and within sleep stages (Figure 1I, $R >= 0.18$, $p < .001$), except for REM sleep ($R=0.05$, $p=.291$).

HR showed a U-shaped pattern from wake, through NREM sleep stages, to REM sleep ($F(4, 55.52) = 7.623$, $p < 0.001$), whereby only HR in NREM sleep stages was significantly lower compared to wake (Figure 1E, Supplementary Table 3). Repeated measures correlation analyses revealed a weak yet significant correlation between HR and spectral slope across sleep stages ($R = 0.09$, 95%CI = [0.056, 0.116], $p<.001$, $N=17$). Moreover, HR was significantly but weakly related to spectral slope for wake ($R=0.12$, $p=.001$) and REM ($R=0.14$, $p=.005$), but not for NREM sleep stages (Figure 1J, $R\leq0.07$, $p\geq0.143$).

Consistent with previous results, we showed that pupil size and spectral slope gradually decreased from wake to N1, N2, and, to a lesser degree also to N3. The small to medium relationships between pupil size and spectral slope variations suggest pupil size to be an indirect marker of arousal levels during wake and NREM sleep stages. A relationship between HR and spectral slope variations was only present during wake and REM. These findings suggest that pupil size appears to be more closely linked to cortical arousal than HR, whereas HR predominantly reflects overall cardiovascular activation state between sleep stages.

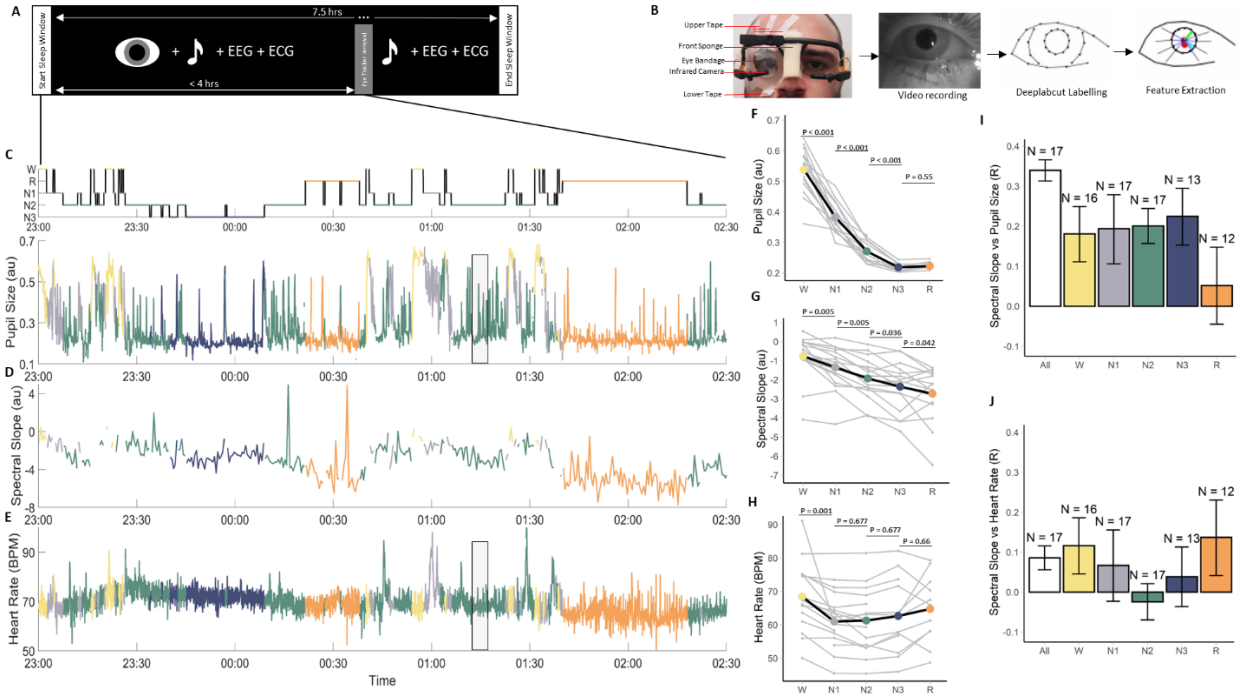


Figure 1. Experimental procedures (**A**) Overall sleep protocol. Auditory stimulation was applied the whole night during non-rapid eye movement (NREM) sleep whereas pupil size was measured only for up to 4 hours. (**B**) Sleep pupillometry protocol: The eye was taped open and protected with a transparent eye bandage while tracked with an infrared camera. Eye features were extracted from the videos: pupil and eyelid centroids (red markers), fitted ellipse around pupil circumference (black ellipse) major and minor axis of pupil ellipse (cyan and green line), iris points connected to pupil centroid (centre black lines). (**C-J**) W: wakefulness (yellow); N1: NREM stage 1 (grey); N2: NREM stage 2 (green); N3: NREM stage 3 (blue); R: rapid eye movement (REM). (**C-E**) Hypnogram while pupil size was recorded and accompanying time series from a representative participant. **(C)** Pupil size fluctuations during sleep. Values are the ratio of the average pupil radius to the average iris radius. The displayed time period refers to the first 3.5 h of a recording as shown in (A). (**D**) 30-s segments of spectral slope without artifacts. (**E**) Heart rate (HR) in beats per minute (BPM). Grey box is the time period plotted in the bottom plot of Figure 2A. **(F)** Average pupil size, (**G**) spectral slope and (**H**) Mean HR for wake, N1, N2, N3 and REM sleep. Data of each participant is depicted in grey. Statistical differences between sleep stages are depicted *p<.05; **p<.01; ***p<.001 (see supplementary tables 1-3 for more detailed information) (**I**) Repeated measures correlation within sleep stages where pupil data was aggregated in 30-s segments to match the segment lengths of the spectral slope data. "All" refers to all sleep stages together. N is the number of participants with both valid pupil size and spectral slope for each respective sleep state. (**J**) Repeated measures correlation within sleep stages where HR data was aggregated in 30s segments to match the segment lengths of the spectral slope data. N is the number of participants with both valid pupil size and HR for each respective sleep stage. 30-s segments where auditory stimulation had been applied were excluded.

Distinct cortical and cardiovascular states are present across the arousal level spectrum of N2 sleep

Spindle trains have been shown to occur according to ISFs[8,9,26]. This phenomenon is interesting because recent work in rodents suggests that LC-mediated arousal levels (e.g. noradrenergic levels in the thalamus) and HR exhibit similar ISFs which are phase-locked to the occurrence of spindle clusters.

More specifically, there is correlational and causal evidence that LC activity needs to be low for spindle trains to occur[5,6]. Here we tested whether a similar relationship between spindle ISFs, pupil size, and HR exists in humans. Spindle ISFs were detected by lowpass filtering Cz sigma power (spindle frequency range 12-16Hz) during N2 and applying a cycle-by-cycle analysis approach.

Figure 2A shows sigma power fluctuations (representing spindle ISF) together with pupil size (upper panel) and HR dynamics (lower panel) for a representative recording during N2 sleep. All three signals exhibit ISFs, and pupil size is typically low when sigma power increases. We first verified that our spindle ISF detection algorithm accurately reflected underlying spindle activity (see Supplementary Figure 1A, Methods). Then, the time points of sigma power, pupil size, and HR of each spindle ISF were renormalized to the corresponding percentage of the total spindle ISF cycle length. Pupil size and HR were normalized to baseline, the average magnitude during each ISF (Methods). Pupil size was smallest (-6.25% relative to baseline) during the increasing phase of sigma power, a more continuous sleep state, while it was largest (+7.61% relative to baseline) during the declining phase of sigma power, reportedly a more fragile sleep state[27]. Cross-correlating pupil size and sigma power revealed a negative correlation (R=-0.26, Supplementary Figure 1C), for which the trough in pupil size preceded the peak of sigma power activity by approximately 30% of the ISF length (corresponding to a 108-degree phase difference).

Contrary to pupil diameter, HR fluctuations seemed to follow sigma power fluctuations (Figure 2B Bottom panel, Supplementary Figure 1D), and a cross-correlation analysis yielded a positive correlation (R=0.26) and delay of the HR fluctuation of approximately 5% of the cycle compared to sigma power (corresponding to an 18-degree phase difference). Aside sigma fluctuations, we also examined temporal relations of our parameters with other key features of N2 sleep, notably the K-complexes. These are prominent EEG hallmarks that have long been associated to be a response to arousal level fluctuations and cause synchronized changes in excitability. Therefore, we calculated the likelihood of K-complex occurrence during spindle ISFs (for review, see[28], Supplementary Figure 1B). K-complexes were more likely to occur during the decreasing phase of sigma power (high pupil size) than during the increasing phase of sigma power (low pupil size), suggesting a preferred occurrence of K-complexes during high levels of arousal.

In summary, we show for the first time in humans that fluctuations in pupil size, HR, and K-complex likelihood are linked to spindle ISFs. Spindle ISFs are an EEG marker that has been previously established to follow LC-mediated arousal levels, being low when LC-NA levels are high and vice versa during NREM sleep in rodents[5,6]. The pupil size in our data follows a similar pattern as LC-NA levels in regard to spindle ISF and could therefore be a potential indirect readout paralleling LC-mediated arousal level.

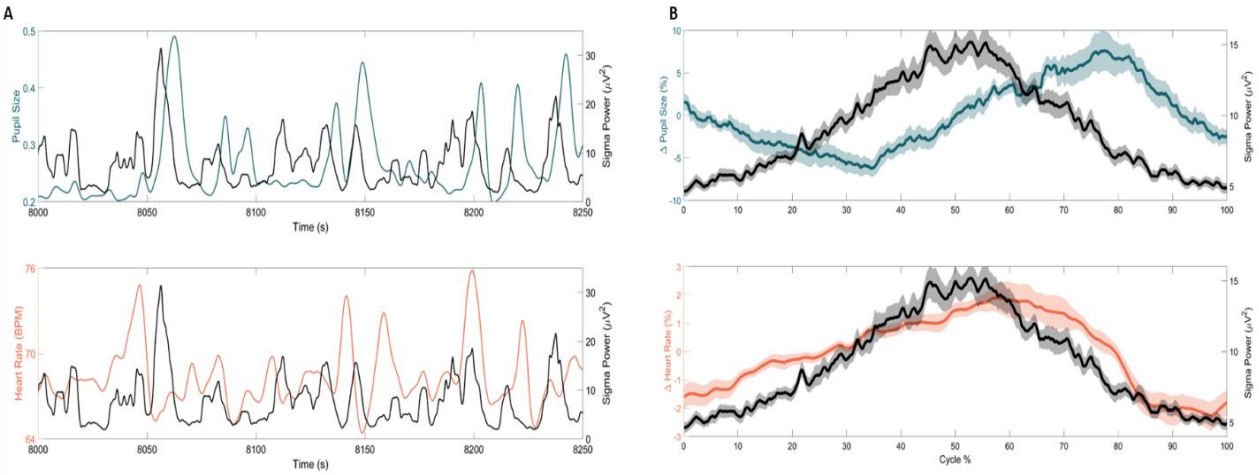


Figure 2. **Relationship between sigma power, pupil size and heart rate** (**A**) Top: Individual example of pupil size and sigma power fluctuations corresponding to the grey box in the sleep recording from Figure 1C. Bottom: Individual example of HR and sigma power fluctuations corresponding to the grey box in the sleep recording from Figure 1E. (**B**) Top: Pupil size (blue & left axis) and sigma power (black & right axis) normalized to the timing of the troughs of detected spindle infraslow fluctuations (ISF). Bottom: Heart rate (red & left axis) and sigma power (black & right axis) normalized to the troughs of detected spindle ISF. The time between troughs was renormalized in terms of cycle percentage. Centre lines plots represent the mean across participants and shadings around centre lines represent the standard error of the mean (SEM).

Pupil size and heart rate fluctuations reveal distinctly different dynamics across N2 events

Next, we investigated how pupil size relates to three specific events in N2 sleep micro-architecture that hallmark changes in arousal levels, i.e. K-complexes, spindle clusters, and sleep arousals as defined in the American Academy of Sleep Medicine (AASM) criteria[8,27,29–31]. For this analysis, we excluded epochs where auditory stimuli were presented and normalized pupil size as well as HR to the 5s before each event (Figure 3). It is worth noting that not all participants exhibited each event type together with valid pupil size or ECG measurements. To evaluate if pupil size and HR are different in relation to these sleep-microarchitecture events and therefore distinct from the overall N2 state, we investigated pupil and HR dynamics at each N2 event type and used statistical parametric mapping for one-dimensional data (SPM1D). For K-complexes, we observed a clear pupil dilation with an average increase of 13.22% from baseline which preceded the onset of the K-complex by approximately 2s ($p<.027$) and lasted up to approximately 4s ($p<.001$) after the onset (Figure 3A). Figure 3A also shows that average pupil size gradually decreased to a minimal level before the dilation started. Furthermore, HR revealed a biphasic response to K-complexes reflected in a 5.32% increase in HR from baseline ($p<.001$) followed by a 2.94% decrease below baseline in HR ($p<.001$, Figure 3D). This biphasic HR response to K-complexes aligns with previous findings[30].

Arousals from sleep as defined by AASM guidelines are brief perturbations to sleep which are characterized by large and sudden changes in the theta, alpha, or beta frequencies in the EEG power spectra, which can be accompanied by EMG activation. We found sleep arousals to be accompanied by a pupil dilation that typically started slightly before the determined AASM arousal onset and reached a significant 36% increase compared to baseline, around 5s after AASM arousal onset ($p=.038$). This pupil dilation was descriptively larger than the dilations observed during K-complexes. Similarly, HR increased by 11.1% in parallel with arousals ($p=.003$, Figure 3E) which is larger in magnitude than HR variations observed during K-complexes (Figures 2D-E). Unlike HR, pupil size did not return to baseline values within the 20s after the arousal onset suggesting a change in arousal state compared to the period prior to the sleep perturbation.

Spindles and their clustering into “trains” occur during low LC-NA activity (low arousal level) and play a sleep protecting role[27,32]. Pupil size exhibited a U-shaped pattern during the 40-s period surrounding spindle trains. Prior to the onset of the spindle train, there was a significant 10% decrease in pupil size compared to baseline ($p<.001$). Pupil size remained relatively small throughout the duration of the spindle train and only returned back to values observed 20s prior to the onset of the spindle train ($p<.001$). We found that HR started to slowly increase at the onset of spindle trains (measured at Cz) and remained at least 1% higher than baseline beyond the end of most spindle trains ($p<=.007$, Figure 2E).

In summary, pupil size fluctuations both preceded and accompany sleep spindle trains, K-complexes, and sleep arousals whereas HR show distinct reactions to these events. Therefore, pupil size provides previously undescribed quantitative relationships with the magnitude and the temporal dynamics of the changes in arousal levels surrounding key markers of sleep function and providing complementary information to HR response dynamics.

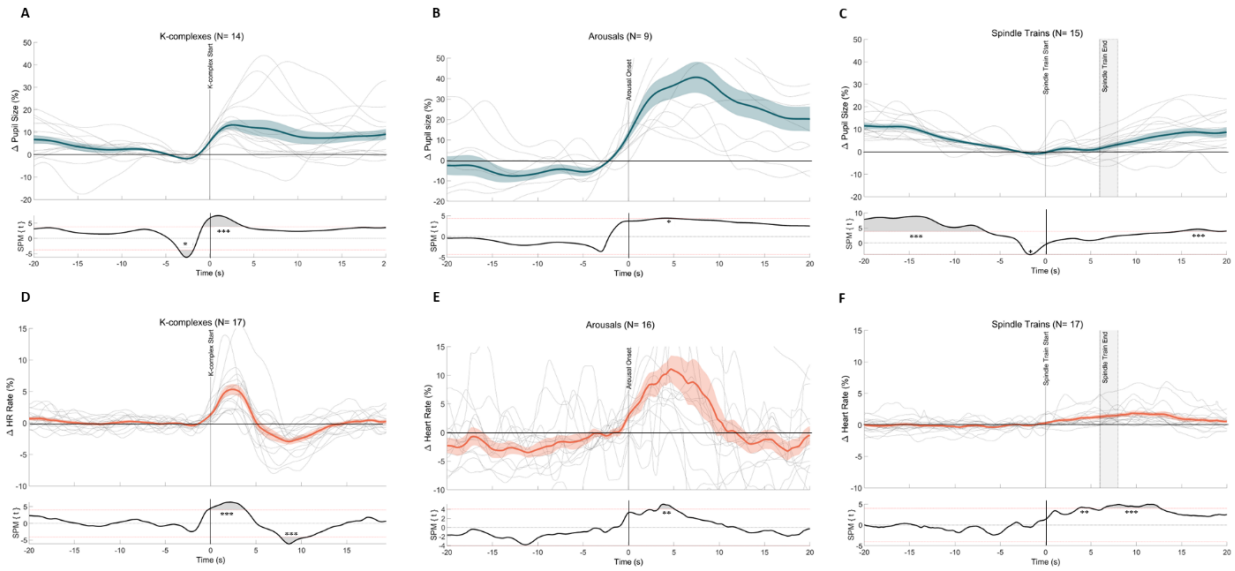


Figure 3. Pupil size and instantaneous heart rate dynamics to N2 events. (**A-C**) Top panels display pupil size colour-coded in blue and normalized to the 5s prior the detected event: K-complexes, arousals, and spindle trains (respectively). (**D-F**) Top panels display heart rate colour-coded in orange and normalized to the 5s prior to the detected event: K-Complexes, spindle trains, and arousals (respectively). (**C,F**) Grey vertical shading in top panels reflects the mean ± 1 SEM of the median spindle train across participants. In all top panels, centre lines represent the group mean and shading around the centre lines the SEM. N indicates the number of participants included in each analysis. Grey lines are the mean response of each participant. (**A-F**) Bottom panels display results of the conducted statistical parametric mapping analyses (SPM1D) with the resulting t values as the y-axis. Grey shaded areas mark significant differences from the 5s baseline prior to each event. *: p<.05, **:p<.01, ***:p<.001.

Cortical response to sensory stimulation varies with arousal levels

Next we investigated whether arousal levels as estimated from pupil size affect cortical response to external stimuli. We applied a 10-s stimulation condition of 1 Hz rhythmic auditory stimulation at 45dB (verum) and a control condition where no stimuli were applied (sham) at varying arousal levels. We found that pupil size showed a sharp pupil dilation during verum stimulation when compared to sham. This dilation reached its peak shortly after the end of the 10s stimulation window and was then followed by a gradual reduction in pupil size (Figure 4A) which, however, remained at elevated levels for at least another 10s. Sigma power, being associated with an increased likelihood of sleep spindle occurrence, increased during the first 5s of stimulation when compared to sham (Figure 4B), which is typically observed when slow wave amplitude is increased by tones[33,34]. However, after the initial peak, sigma power decreased gradually and seemed to be suppressed relative to sham during the period between 10 to 20s post-stimulus. In parallel, delta power representing slow wave activity increased during the first 5s of stimulation when compared to sham (Figure 4C) and returned then back to baseline.

Next, we investigated whether pupil size, as a potential indirect proxy of central arousal levels, prior to playing a tone modulates the response of sigma and delta power to auditory stimulation. Therefore, we separated verum trials where pupil size at baseline was large (highest quartile, representing “high” arousal level) or small (lowest quartile, representing “low” arousal level) and compared the sigma and delta power 5s prior to stimulation onset (pre) with the first 5s of stimulation (post) (Figure 4D-E; Methods). There was a percentage increase in sigma power from pre to post in the small pupil baseline (73.55±66.85%) condition but not in the large pupil baseline condition (1.03±52.16%, $F_{sigmaXpupilrange}(1, 21)=5.66$, p=.027, Figure 4D). Furthermore, there was a percentage increase in delta power from pre to post in the small pupil baseline conditions (555.55±657.48%) but not in the large pupil baseline conditions (193.76±616.25%, $F_{deltaXpupilrange}(1, 28)=6.77$, p=.015, Figure 4E).

In summary, our findings show for the first time that auditory stimulation during sleep induces a pupil response upon stimulation. Furthermore, we demonstrate distinct cortical responses to auditory stimulation at high and low arousal baseline levels as indicated by pupil size prior to stimulation onset.

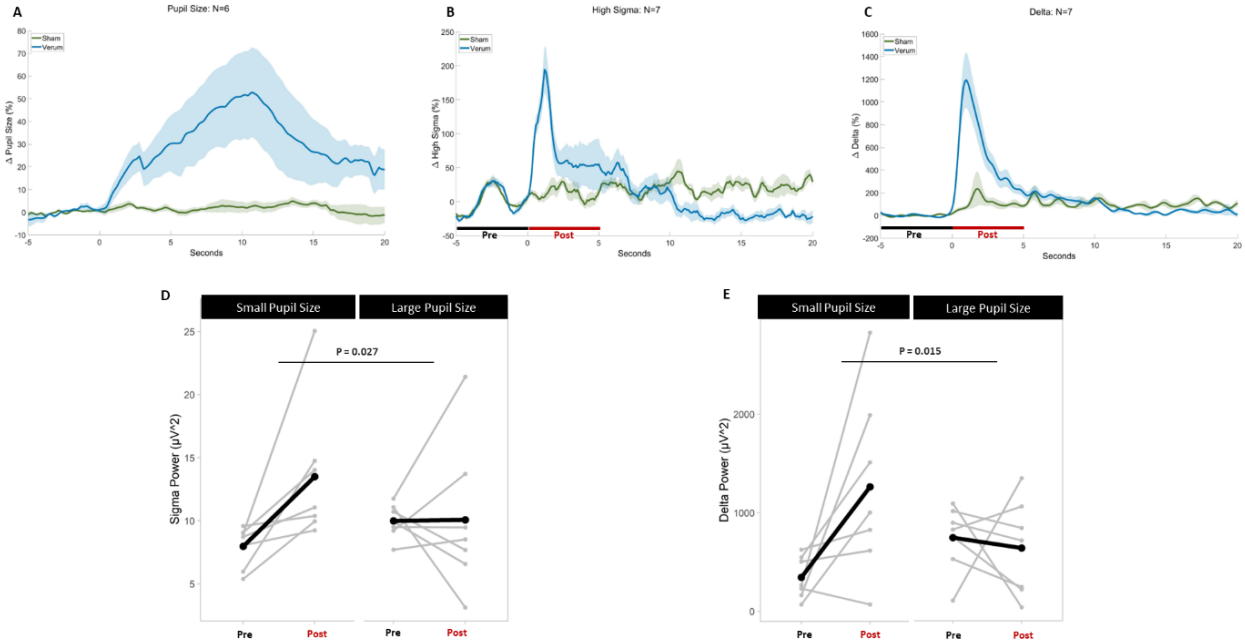


Figure 4. Pupil and EEG responses to auditory stimulation. (**A**-**C**) Pupil size, sigma power, and delta power relative to the 5s prior to the detection of verum (blue) and sham (green) stimulation windows. Centre lines represent the mean and shading around centre lines represent SEM. Time zero is the detection of the stimulation window. (**D**) Sigma power 5s before (pre) and during the first 5s (post) of verum stimulation of trials with large (highest quartile) and small pupil baselines (lowest quartile) to group into high and low arousal level trials. We found an interaction effect indicating increased sigma power from pre to post in the small pupil baseline condition but not in the large pupil baseline condition $F_{sigmaXpupilrange}(1, 21)=5.66$, p=.027, post-hoc comparison pre to post with small pupil baseline: F(1, 7)=8.60, p=.022, post-hoc comparison pre to post with large pupil baseline: F(1, 7)=0.002, p=.97. (**E**) Delta power 5s before (pre) and during the first 5s (post) of verum stimulation of trials with large (highest quartile) and small pupil baselines (lowest quartile) to group into high and low arousal level trials. We found an interaction effect indicating increased delta power from pre to post in the small pupil baseline condition but not in the large pupil baseline condition $F_{deltaXpupilrange}(1, 28)=6.77$, p=.015, post-hoc comparison pre to post with small pupil baseline: F(1, 14)=7.58, p=.032, post-hoc comparison pre to post with large pupil baseline: F(1, 14)=0.25, p=.623. The horizontal black lines connecting pupil size ranges are the interaction effects between pupil size range and pre-post power values.

Discussion

In the present study, we monitored pupil size for the first time during human sleep using a low-dose portable infrared system and compared it to cortical and cardiovascular events measured via polysomnography and ECG. We demonstrate that pupil size dynamics relate to macro- and microarchitecture of sleep and predict the brain's response to auditory stimuli. Therefore, we propose that pupil dynamics might represent a non-invasive way to track arousal levels, most likely reflecting activity of the LC-NA system, in humans at different timescales.

We showed for the first time in humans that pupil size decreases from wakefulness to deep NREM sleep, with REM sleep showing non-significant differences to the N3 deep sleep stage. Similar pupil size changes across vigilance states have previously been reported in rodent models[7]. Here, we reproduced the pronounced decrease from wakefulness to NREM sleep (N1 and N2) and then to N3 and REM sleep. Note that there is no distinction of NREM sleep stages in rodents so that the small pupil size which characterizes not only REM sleep but also NREM N3 sleep has remained undetected until now. In the present study, observed changes in pupil size were paralleled by a decrease in spectral slope, an electrophysiological marker of cortical arousal level[22], as NREM sleep deepened. Spectral slope levels decreased nearly linearly across sleep stages and, unlike pupil size, there was a significant reduction from N3 to REM sleep. Pupil size showed a correlation with spectral slope not only across all sleep stages but also within wake and NREM sleep. These findings suggest that pupil dynamics capture pronounced changes in vigilance state but also that spectral slope and pupil diameter might reflect slightly different physiological mechanisms during REM sleep. However, it cannot be ruled out that

further pupil constriction from N3 to REM sleep was prevented due to a floor effect. HR, on the other hand, showed little to no association with the spectral slope, specifically across sleep stages and within NREM and wake, indicating that this measure seems to be less closely linked to general arousal level dynamics per se but may rather be used to classify the autonomic response to arousal events[29].

Recent studies in rodents demonstrated that LC-mediated arousal levels exhibit pronounced ISF within NREM sleep which lead to sleep spindle ISFs. This has been causally linked to LC-NA activity as periods of low LC-NA activity coincided with an increased likelihood of sleep spindle clusters which protect NREM sleep from perturbations and are referred to as continuity periods. Periods of high LC activity, by contrast, were indexed by low sleep spindle probability and coincided with an elevated sensory arousability in thalamocortical and autonomic circuits and are referred to as fragility periods[5,8,13,27]. Interestingly, pupil size measured in rodents was found to show an inverse relationship to spindle ISFs and therefore parallels the trajectory of these LC-mediated fragility and continuity periods[7]. Here, we found a similar inverse relationship (a delayed anticorrelation) between pupil size and the sleep spindle ISFs during N2, providing first evidence for a potential link between sleep spindle ISFs and arousal level in human sleep. We further found that the onset of spindle trains was preceded by a significant constriction in pupil size (Figure 3C)[35]. Thus, our results are well in line with previous findings in rodents and it is tempting to speculate that pupil size reflects LC-mediated arousal levels during human NREM sleep. Although studies have found sigma (12-16 Hz) power, which reflects sleep spindle activity, to correlate positively with HR (with a time lag) in humans and negatively in rodents, there are no studies to our knowledge that have investigated HR fluctuations during individual spindle trains[7,8]. In line with initial evidence in humans[8], we found a positive correlation between HR and ISFs of sleep spindles in N2. The effect size of this correlation (R=0.26) was similar in magnitude to the anticorrelation between pupil size and sleep spindle ISFs (R=-0.26). Despite that HR fluctuations surrounding spindle trains were not as phase locked across participants as the pupil size fluctuations were, the strength of this effect may be partially explained by HR responses associated with quicker N2 events occurring during the infraslow fluctuations. For example, the preferred occurrence of spontaneous K-complexes during periods of falling sigma power where HR showed strong and robust fluctuations during spontaneous K-complexes[30].

LC activity is not just fluctuating in an infraslow manner during N2, but also shows dynamics in shorter time frames that are linked to sleep microstructural events such as K-complexes and sleep spindles[18,36]. We therefore established how pupil dynamics evolve around such microevents and observed changes in pupil size prior to, during, and after the start of K-complexes and spindle clusters. More specifically, we found that pupil size dilated before and during K-complexes and constricted before spindle clusters. Therefore, different arousal levels seem to be tightly linked to the likelihood for K-complexes and sleep spindles to occur, a relationship that can be captured by measuring pupil size. Conversely, HR remained unaffected prior to the occurrence of sleep spindles and K-complexes, but showed an increase following the onset of these events. Furthermore, the distinctive and time-locked pupil response following the onset of K-complexes suggests that LC activation may be involved in their generation in humans. In rodents, Eschenko et al. reported LC firing to be time-locked to specific phases of slow waves[37]. Here, we specifically focused on K-complexes that represent highly synchronized bottom-up slow waves that potentially involve LC activiation and the promotion of highly synchronized cortical activity[38,39]. Along these lines, Osorio-Forero et al. (2023) reported that transient LC activity increased delta power, possibly reflecting K-complexes[36]. Our results indicate that K-complexes and sleep arousals both involve a significant pupil dilation. However, sleep disruption (i.e., sleep arousal) as indicated by EEG desynchronization had a more pronounced pupil response than a K-complex. This could indicate that the more LC activity, reflected in a greater pupil dilation, is needed to promote an EEG desynchronisation than a K-complex. Furthermore, the fine-scale temporal interaction between LC and cortical events might be particularly interesting in the context of potential LC-NA contributions to memory consolidation during NREM sleep[6,18,37,40,41]. However, future studies that perturb or promote LC activity in humans are needed to understand whether LC activity is involved in the occurrence of K-complexes and sleep spindles and how this is implied in memory consolidation and pupil size fluctuations.

Assuming that different pupil levels track different arousal levels, we would expect that the cortical response to auditory stimulation differs depending on when these stimuli are played as a function of pupil size. According to previous research, we found auditory stimulation to increase delta and sigma power dynamics[33,42–48]. However, this increase in delta power was indeed only present when pupil size

was low (lowest quartile) prior to stimulation onset. Although we did not quantify K-complex occurrence during auditory stimulation as they were often not isolated, the observed time course of delta dynamics resembled the steep increase of slow wave amplitude and slope of a K-complex[49,50]. Therefore, low pupil size might be a prerequisite for the induction of sensory-evoked K-complexes and consequently, delta enhancement. It has been previously proposed that K-complexes can only be evoked during times of lower blood pressure[51] and lower HR[30], indicating decreased cardiovascular activation, or likely a state of decreased arousal. In rodents, Hayat et al. (2020) showed that tonic LC activity levels, were higher before tones that led to sensory evoked NREM sleep perturbations compared to tones resulting in maintenance of NREM sleep[13]. Importantly, in both cases the tones induced a phasic firing of LC time-locked to the tone onset, but only high prior arousal levels resulted in sleep arousals. In this regard, our results might point to the idea that during low arousal levels, as indicated by low pupil size here, slow oscillations might be induced as a consequence of the short phasic firing of LC which may act sleep protective. However, our stimulation protocol did only involve low volume sounds that are generally not inducing sleep perturbations. Overall, our results indicate that the arousal state prior to stimulation onset significantly influences the effectiveness of auditory stimulation to enhance K-complexes or slow waves, as well as spindle activity. Consequently, pupil size as a proxy for arousal level dynamics may essentially improve current auditory stimulation protocols. Tracking the likelihood of sensory evoked awakenings upon different sound volumes across pupil level fluctuations and how this relates to the occurrence of slow oscillations in addition to investigating the optimal timing of tone presentation to maximize effectiveness in enhancing delta waves and spindles may help improve auditory stimulation protocols used in sleep.

Our approach holds the potential to unlock a new research branch within the field of sleep, as it may enable us to track (LC-mediated) arousal levels during human sleep. This breakthrough allows for fresh fundamental discoveries regarding the interaction between arousal levels and sleep function, potentially offering indirect insight into the LC-NA system during human sleep. Moreover, our findings have potential clinical implications in the diagnosis of sleep disturbances. Abnormal arousal levels and LC dysfunction have been closely associated with insomnia, stress-related disorders, and neurodegenerative diseases like Alzheimer's and Parkinson's, which often manifest with disrupted sleep patterns[40,52–54]. By employing our approach, clinicians could directly diagnose whether abnormally elevated arousal levels underlie perturbed sleep in these patients, thus motivating novel treatment approaches.

In conclusion, this study highlights the potential of pupil size as a non-invasive marker of (LC-mediated) arousal levels during human sleep. Pupil size dynamics correspond to sleep macrostructure and microstructural events, while exhibiting an inverse relationship with spindle ISF. Additionally, the arousal state as indicated by pupil size prior to auditory stimulation influences its effectiveness to enhance delta power. These findings provide novel insights into the interplay between arousal levels and sleep, opening avenues for future research and clinical applications in diagnosing and treating disorders associated with abnormal arousal levels and LC dysfunction. Pupil size monitoring therefore holds promise as a new tool in both human sleep research and clinical practice.

Methods

Participants

In total 18 participants were recruited for this study. One participant was excluded because of predefined exclusion criteria, as they reported an elevated level of eye dryness (score greater than 12 in the Ocular Surface Disease Index questionnaire[55], 0 to 12 representing normal levels of eye dryness). The 17 included participants (10 female, mean±sd age; 29.83±8.69 years) were non-smokers, reported a regular sleep-wake rhythm, and had a body-mass index between 17-30. Participants reported no presence of psychiatric/neurological diseases, sleep disorders, or clinically significant concomitant diseases. The study was approved by the cantonal Ethics Committee Zurich (reference number: KEKZH, BASEC2022-00340) and conducted in accordance with the declaration of Helsinki. All participants provided written informed consent prior to study participation and received monetary compensation.

Experiment Procedure

A graphical overview of the experimental procedure is illustrated in Figure 1A. Prior to enrolment, a telephone screening was conducted to explain the procedure and address questions. On the day of testing, participants answered questionnaires about demographics, health status, handedness, sensitivity to noise, eye health, sleep habits, sleep quality, and daytime sleepiness (not reported here). Thereafter, gold electrodes (Genuine Grass electrodes, Natus Medical Inc., Pleasanton, US) were placed to record EEG (Fpz, Cz, O1, O2, M1 (Reference), according to 10-20 system), chin EMG, EOG, and ECG. All channel impedances were kept below 20kΩ. Etymotic insert earphones (Etymotic Research Inc., ER 3C) were then placed in the ear canals and taped to prevent dislodgment during the night. Subsequently, participants laid in a supine position on the bed where the sides of the pillow were wrapped with towels to restrict the head from side movements. The right eye was kept open by applying three strips of tape on the upper eyelid and one, wider, strip of tape on the lower eyelid (Figure 1B). To prevent eye dryness, vitamin A eye ointment was applied on the eye immediately before the eye was covered by a transparent eye bandage (PRO Optha S, Lohmann & Rauscher International GmbH & Co. KG, Rengsdorf, Germany). Prior to the application, baby soap was used to clean the plastic dome inside the eye bandage, to minimize the accumulation of water condensation. Finally, the eye tracker goggles (Pupil Core, Pupil Labs GmbH, Berlin, Germany) were placed as shown in Figure 1B. Lights out (approximately 10-11pm) occurred as close as possible to the time participants reported going to bed.

With the exception of the small infrared light source on the goggles, all light sources from the apparatus were covered using black tape or blackout curtains. Once the experimenter turned off the lights and left the sleep lab, participants made 3 up-and-down and 3 left-and-right eye movements that indicated the start of the subject's sleep window onset of 7.5h. During the complete sleep window, we recorded polysomnography and ECG using a BrainAmp ExG amplifier (BrainProducts GmbH, Gilching, Germany) through OpenVIBE[56] at a sampling rate of 500 Hz. Additionally, the eye tracker videostream was recorded using the manufacturer's software, Pupil Capture. The experimenter on call entered the sleep lab up to four hours from taping the eye open to remove the goggles, eye bandage, and all tape stripes. Thereafter, the experimenter left the sleep lab again, and the participant was allowed to sleep for the rest of the 7.5h sleep window. In the morning, the experimenter quietly entered the sleep lab, stopped all data acquisition devices, and woke the participant up. The session ended with a second round of sleep quality and mood questionnaires (not reported here) and reimbursement.

EEG Analysis and Sleep Scoring

All data prior to the moment of the horizontal and vertical eye movements (visually detected in the EOG electrodes) marking the beginning of the sleep window were excluded from further analysis. Then, EEG data was down-sampled to 200Hz using the EEGLAB[57] toolbox function *pop_resample* in MATLAB (R2019a, MathWorks Inc., Natick, MA). Thereafter, the automatic sleep staging tool YASA[58] was used to predict sleep stages in 30s windows. Sleep stage predictions with less than 0.6 confidence were visually inspected using the sleep stage visualiser toolbox Visbrain[59] and manually corrected by an expert scorer following the standard AASM scoring guidelines[31]. Artifacts were rejected based on a semiautomatic artifact removal procedure that was described previously[60]. To determine instantaneous delta and sigma power respectively, Fpz data was bandpass filtered in the 0.5 to 2Hz (delta) frequency range, and Cz data in the 12 to 16Hz (sigma) frequency range using MATLAB 2nd order Butterworth filters . Then, the Hilbert transform of each filtered signal was used to obtain the Hilbert amplitude. Delta and sigma power was thereafter calculated as the square of the absolute Hilbert amplitude.

To estimate the spectral slope of the signal, the Fpz signal was first 0.3Hz high-pass filtered, then 60Hz low-pass filtered, and finally 50Hz notch filtered using a Hamming windowed sinc FIR filter from the EEGLAB toolxbox function *pop_eegfiltnew*. The spectral power from 0.5Hz to 45Hz was estimated using a multitaper approach[61] in of the 30s scored epochs which were subsequently used to extract the slope of a fitted model using the FOOOF[62] algorithm in the 30 to 45Hz range, according to Lendner and colleagues[22]. FOOOF settings were kept default with the exception of minimum peak width limits set to 1 and maximum number of detectable peaks set to 4. Notably, we identified artifacts as 30s windows with power values exceeding 3dB within the frequency range of 0.5 to 45Hz, resulting in the exclusion of 5.45% of the windows.

To get metrics for sleep micro-architecture dynamics (K-complexes, sleep spindles, and arousals), K-complexes present in the Fpz electrode were visually inspected and manually labeled using Visbrain[59] according to the AASM definition of a K-complex[31]. Due to the inherent difficulty of differentiating K-complexes from increasing background slow wave activity within N3, only K-complexes occurring during N2 and without neighboring slow waves were considered valid. The negative peak of each K-complex was located using MATLAB's *findpeaks* function. As the short initial positive peak seen in K-complexes is not always present, the onset of the K-complex was defined as 550ms prior to the negative peak as this is typically the time taken for the sharp negative delineation to occur[30,50]. Fast spindles (12 to 16Hz) during N2 were automatically detected using the A7 detector from Lacourse et al.[63,64]. The spindle detection performance was visually inspected for each participant and the sigma correlation parameter for A7 detector was adjusted accordingly. Spindles lasting less than 300ms or longer than 2s were excluded. Spindle clustering in 'trains' was subsequently classified following the Boutin et Doyon[32] criteria: at least 2 consecutive spindles interspaced by at most 6s. Arousals were automatically detected using an arousal detector algorithm from Fernández-Varela and Alvarez-Estevez[65,66]. The optional "hypnogram" variable was used in the detector algorithm to ensure all detected arousals followed the AASM guidelines[31].

A cycle-by-cycle approach was applied for the detection of individual sigma power ISFs. First, the Cz sigma power was low-pass filtered at 0.04Hz using a 5th order Butterworth filter. This cut-off frequency was selected as ISFs have been reported to occur below 0.04Hz in humans[8,9]. Using the MATLAB *findpeaks* function, peaks and troughs were detected. The following criteria had to be met to consider a trough-peak-trough event an ISF: the middle peak and the two troughs occurred during uninterrupted N2; peaks and troughs were at least 15s apart; troughs were within 100s from each other; at least 5% of the fluctuation length contained spindles; no auditory stimulation was presented during the cycle; and no artifacts were present. The ISF analysis depicted in Figure 2 consisted of renormalising the time scale of sigma power, pupil size, and HR from the time duration of the ISF to the ISF cycle length in terms of percentage cycle.

Offline pupil analysis

To calculate pupil size, the eye tracker video was first resampled to 50Hz, and processed using Python 3.8 and DeepLabCut 2.2[20,21]. 36 markers were labelled in each frame as shown in Figure 1B: 12 marking the palpebral fissure circumference, 12 marking the iris circumference, and 12 marking the pupil circumference. Processing of the markers was conducted in MATLAB. First, an ellipse was fitted to the pupil markers. Then, a line was fitted from each iris marker (I) to the centroid of the ellipse (*C*). The intersection point of the fitted line with the ellipse (*P*) was then extracted. Pupil size was calculated as the ratio of the pupil radius to the iris radius using Equation 1, where N is the number of valid iris markers, DIC_i is the distance from the ith iris marker to the centroid of the ellipse, and DPC_i is the distance from the ith intersection point to the centroid of the ellipse (Figure 1B).

$$\text{Equation 1: } Pupil\ Size = \frac{\sum_i^N \frac{DPC_i}{DIC_i}}{N}$$

Pupil gaze was calculated by re-referencing the centroid of the pupil markers to the centroid of a polygon generated from the palpebral fissure markers. To account for palpebral fissure twitches (e.g. blink attempts during wake), the palpebral fissure markers were smoothed using a trailing average 40s moving window.

Right eye pupil size was systematically preprocessed using the guidelines and standardized open-source pipeline published by Kret and Sjak-Shie[67]. Invalid pupil diameter samples representing dilation speed outliers and large deviation from trend line pupil size (repeated four times in a multipass approach) were removed using a median absolute deviation (MAD; multiplier in preprocessing set to 12[67]) – an outlier resilient data dispersion metric. Further, temporally isolated samples with a maximum width of 50ms that border a gap larger than 40ms were removed. Data was then linearly interpolated using the MATLAB function *interp1* to match the 200Hz sampling rate of polysomnography data and to fill gaps with up to one second of missing data. Relative pupil size was calculated as percentage change to the mean pupil size of the last 5s before the events of interest (i.e., spindle trains, K-complexes, and arousals).

ECG analysis

ECG R peaks were automatically detected and then visually inspected using the MATLAB-based toolbox PhysioZoo[68]. Data segments consisting of non-detectable peaks or poor quality were excluded from further analyses. HR dynamicswas calculated by converting each interval between two R peaks as an inverse of its duration. HR was then linearly interpolated using the MATLAB function *interp1* to match the 200Hz sampling rate of the polysomnography data. For the K-complex, spindle train, and arousal analysis, relative HR was calculated as percentage change to the mean HR of the last 5s before the event of interest (i.e., spindle trains, K-complexes, and arousals).

Auditory stimulation protocol

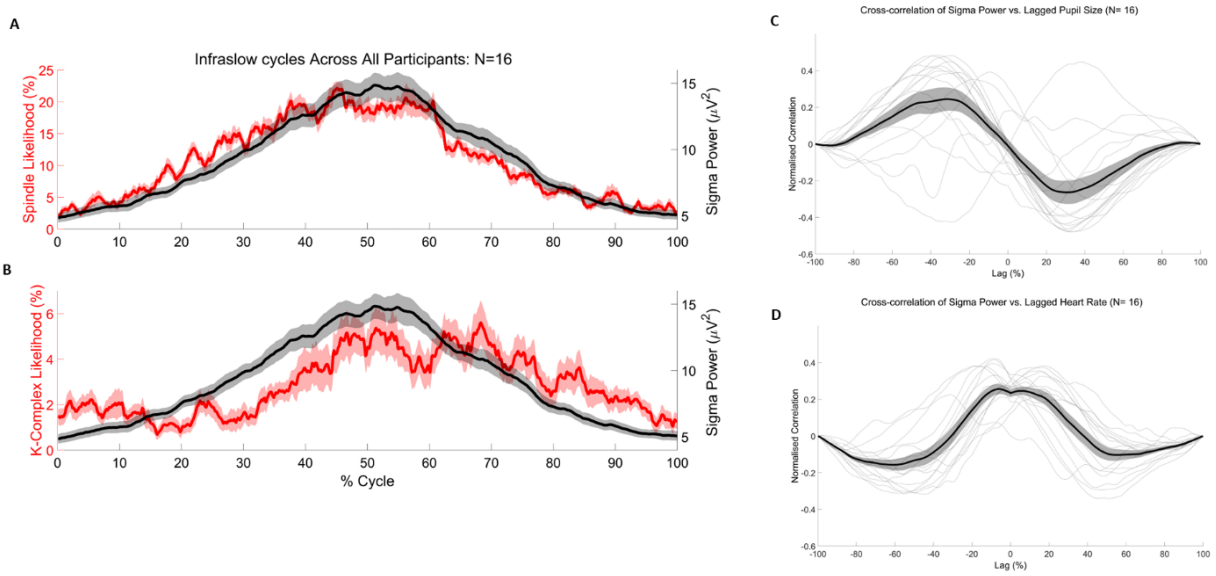
During the sleep period, we investigated brain responses to administered auditory stimuli depending on arousal levels. For that, we used our custom-developed EEG-feedback controlled stimulation protocol in OpenViBE[56] which was an extension of a previously reported protocol by our group[42,46]. As our stimulation condition (verum), we used the 45dB stimulation protocol ISI1 from Huwiler et al.[42]. This consisted of a 1Hz rhythmic EEG feedback-controlled auditory stimulation of a 50ms burst of pink noise with a sound level of 45dB in a windowed 10s ON (auditory stimulation presented) followed by a 20s OFF (no auditory stimulation presented) design. We used different approaches to apply tones for each participant. These included the verum condition (V0, n=4), and a version with the addition of an ISF detection protocol inspired by Cardis et al.[27] which included the verum condition and a SHAM condition where no auditory stimulation was presented (V1, N=10). The remaining participants were used for the parameter optimization of V1 (V0.5, N=3). For V0 stimulation prerequisites were adapted from our previous work[42] and for V1 the additional prerequisite of varying sigma power thresholds were implemented to stimulate at different levels of spindle ISF. The analysis in Figure 4A-C included the participants from the V1 protocol. Out of the 10 participants in V1, 3 were excluded from the auditory stimulation analysis reported in Figure 4A-C due to having fewer than 20 stimulation windows per condition, leading to a final n = 7.

To compare the differential effects of high and low arousal levels on responses to auditory stimuli (Figure 4D-E), we calculated the average pupil size 5s prior to each stimulation window and used it as a baseline (see offline pupil size analysis). Participants with at least 6 stimulation windows with valid pupil data during the 5s baseline were pooled together from V0 (N=2) and V1 (N=5) protocols. Then, we separated each participant into the largest and smallest 25% baseline pupil size ranges for each stimulation condition separately.

Statistics

Statistical analyses were conducted in R (v 3.6.3; R Core Team, Vienna, Austria). Using the R packages lme4[69] and lmertest[70], we computed linear mixed-effects models with several outcome variables (average pupil size, average HR, spectral slope) and fixed factors with subject as random factor. The fixed factors were sleep stage (Figure 1F-H), sleep micro-structure event (K-complexes, sleep spindles, Figure 3), stimulation condition (Figure 4A-C), or pupil size range and timing of stimulation (Figure 4D-E). If the linear mixed-effects model was significant for the fixed effect, we conducted post-hoc comparisons using the R package emmeans[71] with the Hochberg method to correct for multiple comparisons. If the linear mixed-effect model was significant for the interaction effect, we derived post-hoc p-values for each analysed contrast and corrected for multiple comparison using Bonferroni correction. Visual inspection of the residual plots of the linear models did not reveal any obvious deviations from normality or homoscedasticity. To calculate inter-subject correlations of pupil size and spectral slope in 30s epochs across sleep stages, repeated measures correlation analyses were conducted using the R package rmcorr[72]. For the cross correlation between pupil, HR, and sigma power during ISFs, each signal was zscored within their ISF and then cross-correlated using the MATLAB function *xcorr*. P-values <0.05 were considered significant. Plots were generated using the R package ggplot2[73] and MATLAB. Where applicable, percentage values reported in the main body of text(?) are in the format mean±standard deviation.

Supplementary Material



Supplementary Figure 1. Likelihoods of spindles and K-Complexes during infraslow cycles. (**A**) Spindle likelihood (red & left axis) and sigma power (black & right axis) normalized to the timing of the troughs of detected spindle infraslow fluctuations (ISF). (**B**) K-Complex likelihood (red & left axis) and sigma power (black & right axis) normalized to the timing of the troughs of detected spindle infraslow fluctuations (ISF). The time between troughs was renormalized in terms of cycle percentage. (**C**) Cross-correlation of sigma power (source) and pupil size with the lag being with respect to the cycle percentage of the ISF. (**D**) Cross-correlation of sigma power (source) and heart rate with the lag being with respect to the cycle percentage of the ISF. Centre lines plots represent the mean across participants and shadings around centre lines represent the standard error of the mean (SEM).

Supplementary Table 1. Pupil size contrasts across sleep stages. P-values are those shown in Figure 1F.

Effect	EMMeans	df	t.ratio	95% CI (Lower)	95% CI (Upper)	p-value	Effect size (Cohen's D)
AWAKE - NREM1	0.156	60.387	10.682	0.114	0.199	<0.001	3.94
AWAKE - NREM2	0.267	60.75	18.256	0.225	0.31	<0.001	6.73
AWAKE - NREM3	0.326	62.898	20.921	0.281	0.372	<0.001	8.23
AWAKE - REM	0.316	64.131	19.189	0.268	0.364	<0.001	7.97
NREM1 - NREM2	0.111	60.75	7.58	0.068	0.154	<0.001	2.8
NREM1 - NREM3	0.17	62.898	10.91	0.125	0.216	<0.001	4.29
NREM1 - REM	0.16	64.131	9.707	0.112	0.208	<0.001	4.03
NREM2 - NREM3	0.059	61.443	3.824	0.014	0.104	0.001	1.49
NREM2 - REM	0.049	62.641	2.994	0.001	0.097	0.008	1.24
NREM3 - REM	-0.01	60.699	-0.607	-0.06	0.039	0.546	-0.26

Supplementary Table 2. Spectral slope contrasts across sleep stages. P-values are those shown in Figure 1G.

Effect	EMMeans	df	t.ratio	95% CI (Lower)	95% CI (Upper)	p-value	Effect size (Cohen's D)
Wake - N1	0.585	72.25	3.292	0.07	1.099	0.005	1.16
Wake - N2	1.163	72.25	6.548	0.649	1.678	<0.001	2.32
Wake - N3	1.594	72.25	8.974	1.08	2.109	<0.001	3.17
Wake - REM	1.961	72.25	11.04	1.447	2.476	<0.001	3.9
N1 - N2	0.578	72.25	3.256	0.064	1.093	0.005	1.15
N1 - N3	1.009	72.25	5.682	0.495	1.524	<0.001	2.01
N1 - REM	1.377	72.25	7.748	0.862	1.891	<0.001	2.74
N2 - N3	0.431	72.25	2.426	-0.083	0.946	0.036	0.86
N2 - REM	0.798	72.25	4.492	0.284	1.313	<0.001	1.59
N3 - REM	0.367	72.25	2.066	-0.147	0.882	0.042	0.73

Supplementary Table 3. Heart rate contrasts across sleep stages. P-values are those shown in Figure 1H.

Effect	EMMeans	df	t.ratio	95% CI (Lower)	95% CI (Upper)	p-value	Effect size (Cohen's D)
AWAKE - NREM1	5.983	59.388	4.093	1.721	10.244	0.001	1.51
AWAKE - NREM2	6.914	59.534	4.712	2.636	11.191	<0.001	1.75
AWAKE - NREM3	5.323	59.837	3.376	0.728	9.919	0.01	1.35
AWAKE - REM	2.936	59.956	1.758	-1.931	7.803	0.419	0.74
NREM1 - NREM2	0.931	59.534	0.634	-3.347	5.208	0.677	0.24
NREM1 - NREM3	-0.66	59.837	-0.419	-5.255	3.936	0.677	-0.17
NREM1 - REM	-3.047	59.956	-1.824	-7.914	1.821	0.419	-0.77
NREM2 - NREM3	-1.591	59.573	-1.022	-6.13	2.949	0.677	-0.4
NREM2 - REM	-3.977	59.695	-2.412	-8.784	0.829	0.133	-1.01
NREM3 - REM	-2.387	59.421	-1.406	-7.337	2.563	0.66	-0.6

References

1. Poe GR, Foote S, Eschenko O, et al. Locus coeruleus: a new look at the blue spot. *Nat Rev Neurosci*. 2020;21(11):644-659. doi:10.1038/s41583-020-0360-9

2. Andrillon T, Kouider S. The vigilant sleeper: neural mechanisms of sensory (de)coupling during sleep. *Curr Opin Physiol*. 2020;15:47-59. doi:10.1016/J.COPHYS.2019.12.002

3. Osorio-Forero A, Cherrad N, Banterle L, Fernandez LMJ, Lüthi A. When the Locus Coeruleus Speaks Up in Sleep: Recent Insights, Emerging Perspectives. *Int J Mol Sci*. 2022;23(9). doi:10.3390/ijms23095028

4. Bonnet MH, Arand DL. Caffeine Use as a Model of Acute and Chronic Insomnia. *Sleep*. 1992;15(6):526-536. doi:10.1093/SLEEP/15.6.526

5. Osorio-Forero A, Cardis R, Vantomme G, et al. Noradrenergic circuit control of non-REM sleep substates. *Curr Biol*. 2021;31(22):5009-5023.e7. doi:10.1016/j.cub.2021.09.041

6. Kjaerby C, Andersen M, Hauglund N, et al. Memory-enhancing properties of sleep depend on the oscillatory amplitude of norepinephrine. *Nat Neurosci*. 2022;25(August). doi:10.1038/s41593-022-01102-9

7. Yüzgeç Ö, Prsa M, Zimmermann R, Huber D. Pupil Size Coupling to Cortical States Protects the Stability of Deep Sleep via Parasympathetic Modulation. *Curr Biol*. 2018;28(3):392-400.e3. doi:10.1016/j.cub.2017.12.049

8. Lecci S, Fernandez LMJ, Weber FD, et al. Coordinated infraslow neural and cardiac oscillations mark fragility and offline periods in mammalian sleep. *Sci Adv*. 2017;3(2):e1602026. doi:10.1126/sciadv.1602026

9. Lázár ZI, Dijk D, Lázár AS. Infraslow oscillations in human sleep spindle activity. *J Neurosci Methods*. 2019;316(October 2018):22-34. doi:10.1016/j.jneumeth.2018.12.002

10. Joshi S, Gold JI. Pupil Size as a Window on Neural Substrates of Cognition. *Trends Cogn Sci*. 2020;24(6):466-480. doi:10.1016/j.tics.2020.03.005

11. Zerbi V, Floriou-Servou A, Markicevic M, et al. Rapid Reconfiguration of the Functional Connectome after Chemogenetic Locus Coeruleus Activation Article Rapid Reconfiguration of the Functional Connectome after Chemogenetic Locus Coeruleus Activation. *Neuron*. 2019;103:702-718. doi:10.1016/j.neuron.2019.05.034

12. Megemont M, McBurney-Lin J, Yang H. Pupil diameter is not an accurate real-time readout of locus coeruleus activity. *Elife*. 2022;11:e70510.

13. Hayat H, Regev N, Matosevich N, et al. Locus coeruleus norepinephrine activity mediates sensory-evoked awakenings from sleep. *Sci Adv*. 2020;6(15). doi:10.1126/sciadv.aaz4232

14. Murphy PR, O'Connell RG, O'Sullivan M, Robertson IH, Balsters JH. Pupil diameter covaries with BOLD activity in human locus coeruleus. *Hum Brain Mapp*. 2014;35(8):4140-4154. doi:10.1002/hbm.22466

15. Alnæs D, Sneve MH, Espeseth T, Endestad T, van de Pavert SHP, Laeng B. Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. *J Vis*. 2014;14(4):1-1. doi:10.1167/14.4.1

16. de Gee JW, Colizoli O, Kloosterman NA, Knapen T, Nieuwenhuis S, Donner TH. Dynamic modulation of decision biases by brainstem arousal systems. *Elife*. 2017;6. doi:10.7554/ELIFE.23232

17. Reimer J, McGinley MJ, Liu Y, et al. Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nat Commun*. 2016;7(May):1-7. doi:10.1038/ncomms13289

18. Swift KM, Gross BA, Frazer MA, et al. Abnormal Locus Coeruleus Sleep Activity Alters Sleep Signatures of Memory Consolidation and Impairs Place Cell Stability and Spatial Memory. *Curr Biol*. 2018;28(22):3599-3609.e4. doi:10.1016/j.cub.2018.09.054

19. McNab AA. The eye and sleep. *Clin Exp Ophthalmol*. 2005;33(2):117-125. doi:10.1111/j.1442-9071.2005.00969.x

20. Weygandt Mathis M, Bethge M. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat Neurosci*. doi:10.1038/s41593-018-0209-y

21. Nath T, Mathis A, Chen AC, Patel A, Bethge M, Mathis MW. Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nat Protoc*. 2019;14(7):2152-2176. doi:10.1038/s41596-019-0176-0

22. Lendner JD, Helfrich RF, Mander BA, et al. An electrophysiological marker of arousal level in humans. *Elife*. 2020;9:1-29. doi:10.7554/eLife.55092

23. Kobayashi G, Tanaka KF, Takata N. Pupil Dynamics-derived Sleep Stage Classification of a Head-fixed Mouse Using a Recurrent Neural Network. *Keio J Med*. 2023;72(2):44-59. doi:10.2302/kjm.2022-0020-OA

24. Bódizs R, Szalárdy O, Horváth C, et al. A set of composite, non-redundant EEG measures of NREM sleep based on the power law scaling of the Fourier spectrum. *Sci Rep*. 2021;11(1):1-18. doi:10.1038/s41598-021-81230-7

25. Miskovic V, Macdonald KJ, Rhodes | L Jack, Cote KA. Changes in EEG multiscale entropy and power-law frequency scaling during the human sleep cycle. Published online 2018. doi:10.1002/hbm.24393

26. Antony JW, Piloto L, Wang M, Pacheco P, Norman KA, Paller KA. Sleep Spindle Refractoriness Segregates Periods of Memory Reactivation. *Curr Biol*. 2018;28(11):1736-1743.e4. doi:10.1016/j.cub.2018.04.020

27. Cardis R, Lecci S, Fernandez LM, et al. Cortico-autonomic local arousals and heightened somatosensory arousability during NREMS of mice in neuropathic pain. *Elife*. 2021;10:1-27. doi:10.7554/elife.65835

28. Halász P, Terzano M, Parrino L, Bódizs R. The nature of arousal in sleep. *J Sleep Res*. 2004;13(1):1-23. doi:10.1111/J.1365-2869.2004.00388.X

29. Sforza E, Jouny C, Ibanez V. Cardiac activation during arousal in humans: further evidence for

hierarchy in the arousal response. doi:10.1016/s1388-2457(00)00363-1

30. de Zambotti M, Willoughby AR, Franzen PL, Clark DB, Baker FC, Colrain IM. K-Complexes: Interaction between the Central and Autonomic Nervous Systems during Sleep. *Sleep*. 2016;39(5):1129-1137. doi:10.5665/sleep.5770

31. Iber C, Ancoli-Israel S, Chesson AL, Quan SF. The AASM manual for the scoring of sleep and associated events American Academy of Sleep Medicine. *Westchester Am Acad Sleep Med*. 2007;(October 2012):1-17.

32. Boutin A, Doyon J. A sleep spindle framework for motor memory consolidation. doi:10.1098/rstb.2019.0232

33. Ngo H V, Martinetz T, Born J, Molle M. Auditory Closed-Loop Stimulation of the Sleep Slow Oscillation Enhances Memory. *Neuron*. Published online 2013. doi:S0896-6273(13)00230-4 [pii]10.1016/j.neuron.2013.03.006

34. Ngo HV V., Seibold M, Boche DC, Mölle M, Born J. Insights on auditory closed-loop stimulation targeting sleep spindles in slow oscillation up-states. *J Neurosci Methods*. 2019;316:117-124. doi:10.1016/J.JNEUMETH.2018.09.006

35. Boutin A, Gabitov E, Pinsard B, Boré A, Carrier J, Doyon J. Clustering and temporal organization of sleep spindles underlie motor memory consolidation. doi:10.1101/2022.11.29.518200

36. Osorio-Forero A, Foustoukos G, Cardis R, et al. Locus coeruleus activity fluctuations set a non - reducible timeframe for mammalian NREM - REM sleep cycles. 2023;(184759):1-40.

37. Eschenko O, Magri C, Panzeri S, Sara SJ. Noradrenergic neurons of the locus coeruleus are phase locked to cortical up-down states during sleep. *Cereb Cortex*. 2012;22(2):426-435. doi:10.1093/cercor/bhr121

38. Siclari F, Bernardi G, Riedner BA, LaRocque JJ, Benca RM, Tononi G. Two Distinct Synchronization Processes in the Transition to Sleep: A High-Density Electroencephalographic Study. *Sleep*. 2014;37(10):1621-1637. doi:10.5665/SLEEP.4070

39. Bernardi G, Siclari F, Handjaras G, et al. Local and Widespread Slow Waves in Stable NREM Sleep: Evidence for Distinct Regulation Mechanisms. Published online 2018. doi:10.3389/fnhum.2018.00248

40. Van Egroo M, Koshmanova E, Vandewalle G, Jacobs HIL. Importance of the locus coeruleus-norepinephrine system in sleep-wake regulation: Implications for aging and Alzheimer's disease. *Sleep Med Rev*. 2022;62:101592. doi:10.1016/j.smrv.2022.101592

41. Sara SJ. Locus Coeruleus in time with the making of memories This review comes from a themed issue on Circuit plasticity and memory. *Curr Opin Neurobiol*. 2015;35:87-94. doi:10.1016/j.conb.2015.07.004

42. Huwiler S, Carro Dominguez M, Huwyler S, et al. Effects of auditory sleep modulation approaches on brain oscillatory and cardiovascular dynamics. *Sleep*. Published online July 6, 2022:zsac155. doi:10.1093/sleep/zsac155

43. Besedovsky L, Ngo H-V V, Dimitrov S, Gassenmaier C, Lehmann R, Born J. Auditory closed-loop stimulation of EEG slow oscillations strengthens sleep and signs of its immune-supportive function. *Nat Commun*. 2017;8(1):1984.

44. Krugliakova E, Skorucak J, Sousouri G, et al. Boosting recovery during sleep by means of auditory stimulation. *Front Neurosci*. 2022;16:755958.

45. Leminen MM, Virkkala J, Saure E, et al. Enhanced memory consolidation via automatic sound stimulation during non-REM sleep. *Sleep*. 2017;40(3):zsx003.

46. Lustenberger C, Ferster ML, Huwiler S, et al. Auditory deep sleep stimulation in older adults at home: a randomized crossover trial Plain language summary. doi:10.1038/s43856-022-00096-6

47. Weigenand A, Mölle M, Werner F, Martinetz T, Marshall L. Timing matters: open-loop stimulation does not improve overnight consolidation of word pairs in humans. *Eur J Neurosci*. 2016;44(6):2357-2368.

48. Grimaldi D, Papalambros NA, Reid KJ, et al. Strengthening sleep-autonomic interaction via acoustic enhancement of slow oscillations. *Sleep*. 2019;(February):1-11. doi:10.1093/sleep/zsz036

49. Bellesi M, Riedner BA, Garcia-Molina GN, Cirelli C, Tononi G. Enhancement of sleep slow waves: Underlying mechanisms and practical consequences. *Front Syst Neurosci*. 2014;8(October):1-17. doi:10.3389/fnsys.2014.00208

50. Colrain IM. The K-Complex: A 7-Decade History REVIEW. *Hist Rev SLEEP*. 2005;28(2):255. Accessed June 25, 2023. https://academic.oup.com/sleep/article/28/2/255/2696914

51. Monstad P, Guilleminault C. Cardiovascular changes associated with spontaneous and evoked K-complexes. *Neurosci Lett*. 1999;263(2-3):211-213.

52. Van Someren EJW. Brain Mechanisms of Insomnia: New Perspectives on Causes and Consequences. *Physiol Rev*. 2021;101(3):995-1046. doi:10.1152/physrev.00046.2019

53. McCall JG, Al-Hasani R, Siuda ER, et al. CRH Engagement of the Locus Coeruleus Noradrenergic System Mediates Stress-Induced Anxiety. *Neuron*. 2015;87(3):605-620. doi:10.1016/j.neuron.2015.07.002

54. Wood CS, Valentino RJ, Wood SK. Individual differences in the locus coeruleus-norepinephrine system: Relevance to stress-induced cardiovascular vulnerability. *Physiol \& Behav*. 2017;172:40-48.

55. Schiffman RM, Christianson MD, Jacobsen G, Hirsch JD, Reis BL. Reliability and validity of the Ocular Surface Disease Index. *Arch Ophthalmol (Chicago, Ill 1960)*. 2000;118(5):615-621. doi:10.1001/archopht.118.5.615

56. Renard Y, Lotte F, Gibert G, et al. OpenViBE: An open-source software platform to design, test, and use brain-computer interfaces in real and virtual environments. *Presence Teleoperators Virtual Environ*. 2010;19(1):35-53. doi:10.1162/pres.19.1.35

57. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods*. 2004;134(1):9-21.

58. Vallat R, P. Walker M. An open-source, high-performance tool for automated sleep staging. doi:10.7554/elife.70092

59. Hill SL, Chintaluri C, Bakker R, et al. Visbrain: A Multi-Purpose GPU-Accelerated Open-Source Suite for Multimodal Brain Data Visualization. *Front Neuroinformatics | www.frontiersin.org*. 2019;1:14. doi:10.3389/fninf.2019.00014

60. Lustenberger C, Wehrle F, Tüshaus L, Achermann P, Huber R. The Multidimensional Aspects of Sleep Spindles and Their Relationship to Word-Pair Memory Consolidation. *Sleep*. 2015;38(7):1093-1103. doi:10.5665/sleep.4820

61. Prerau MJ, Brown RE, Bianchi MT, Ellenbogen JM, Purdon PL. Sleep neurophysiological

dynamics through the lens of multitaper spectral analysis. *Physiology*. 2017;32(1):60-92. doi:10.1152/physiol.00062.2015

62. Donoghue T, Haller M, Peterson EJ, et al. Parameterizing neural power spectra into periodic and aperiodic components. *Nat Neurosci*. 2020;23(12):1655-1665. doi:10.1038/s41593-020-00744-x

63. Lacourse K, Yetton B, Mednick S, C. Warby S. Massive online data annotation, crowdsourcing to generate high quality sleep spindle annotations from EEG data. doi:10.1038/s41597-020-0533-4

64. Lacourse K, Delfrate J, Beaudry J, Peppard P, Warby SC. A sleep spindle detection algorithm that emulates human expert spindle scoring. Published online 2018. doi:10.1016/j.jneumeth.2018.08.014

65. Fernández-Varela I, Alvarez-Estevez D, Hernández-Pereira E, Moret-Bonillo V. A simple and robust method for the automatic scoring of EEG arousals in polysomnographic recordings. doi:10.1016/j.compbiomed.2017.05.011

66. Alvarez-Estevez D, Fern Andez-Varela I. Large-scale validation of an automatic EEG arousal detection algorithm using different heterogeneous databases. Published online 2019. doi:10.1016/j.sleep.2019.01.025

67. Kret ME, Sjak-Shie EE. Preprocessing pupil size data: Guidelines and code. *Behav Res Methods*. 2019;51(3):1336-1342. doi:10.3758/s13428-018-1075-y

68. Behar JA, Rosenberg # AA, Weiser-Bitoun I, et al. PhysioZoo: a novel open access platform for heart rate variability analysis of mammalian electrocardiographic data Running title: PhysioZoo for mammalian HRV analysis.

69. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using {lme4}. *J Stat Softw*. 2015;67(1):1-48. doi:10.18637/jss.v067.i01

70. Kuznetsova A, Brockhoff PB, Christensen RHB. {lmerTest} Package: Tests in Linear Mixed Effects Models. *J Stat Softw*. 2017;82(13):1-26. doi:10.18637/jss.v082.i13

71. Lenth R V. emmeans: Estimated Marginal Means, aka Least-Squares Means. Published online 2022. https://cran.r-project.org/package=emmeans

72. Bakdash JZ, Marusich LR. rmcorr: Repeated Measures Correlation. Published online 2021. https://cran.r-project.org/package=rmcorr

73. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York; 2016. https://ggplot2.tidyverse.org