Slow wave stimulation using a smartwatch improves sleep quality

Nathan W Whitmore, Samantha W.T. Chan*, Abigail Dulski*, Anita Podrug, Nelson Hidalgo, Nnamdi Obi, Varun K. Viswanath, Michael S. Freedman, Viswam Nathan, Pattie Maes

*Equal contributions

Abstract

Poor sleep affects millions, increasing long-term health risks. Slow wave entrainment (SWE) is a promising method for improving sleep. While typically lab-based, we developed a smartwatch app to perform SWE and tested it with 93 participants at home. We found our app significantly increased slow wave and delta power amplitude, consistent with lab-based SWE protocols. This stimulation also reduced sleep fragmentation. Furthermore, we found that the increased slow wave amplitude was linked to improvements in mood and cognition measures, indicating the induced slow waves were functional. SWE produced the largest benefits in mood for participants who reported higher levels of pre-existing sleep problems. Our results demonstrate the potential for individuals to improve their sleep and cognition using a health smartwatch that performs SWE, offering an accessible tool for millions to enhance their sleep quality and health.

Introduction

The Centers for Disease Control and Prevention (CDC) reports that poor sleep affects 70% of adults in America, often due to factors out of their control, and sleep-related problems affect 50-70 million people[1]. Poor sleep quality can affect people's mood, attention, and memory and, over long periods of time, increases the risk of adverse outcomes related to cardiovascular, cardiometabolic, and mental health[1–3]. Poor sleep quality can be secondary to other conditions, such as sleep apnea[4], driven by lifestyle and sleeping environment conditions[5,6], or idiopathic. Poor sleep quality also has a bidirectional relationship with mental health[7], aging, and neurodegenerative disease[8], where people with mental health conditions and older adults experience sleep problems, and sleep problems also worsen quality of life and contribute to disease progression.

Typical approaches to improve poor sleep include treating contributing conditions as well as behavioral/lifestyle interventions to improve sleep habits and pharmacologic interventions including sedative/hypnotic drugs when insomnia symptoms are present. While effective, lifestyle-based interventions are difficult for patients to implement, especially when patients are constrained by factors like parenting and work demands or do not have access to good sleeping conditions[9]. Current guidelines do not recommend pharmacotherapy as a first line treatment for insomnia because the adverse effects outweigh the benefits to sleep quality[10]. Therefore, new approaches to improve sleep quality are urgently needed.

One promising approach for improving sleep quality is increasing the duration and/or intensity of slow wave sleep. Slow wave sleep (otherwise known as N3 sleep) is a deep phase of non-REM sleep during which the brain generates large, rhythmic oscillations in the range of 0.5-1.5 Hz

(slow waves). Intensity of slow wave sleep reflects homeostatic sleep need[11] and slow wave sleep appears to be important for many of the restorative functions of sleep in memory, cognition, mental health, and cognitive function[3,12–15] .

Important evidence for the role of slow wave sleep has emerged through slow wave entrainment (SWE) studies, where external stimulation is used to increase the amplitude of slow waves in the brain. Slow wave entrainment has been shown to increase the restorative benefits of sleep in multiple domains, including cognitive function, immune function, and cardiovascular function[16]. These results suggest that slow wave amplitude is an important aspect of sleep quality which can be improved through noninvasive stimulation.

The most frequently performed form of SWE is closed-loop auditory stimulation (CLAS), which uses pulses of pink noise phase-locked to naturally occurring slow waves. CLAS is believed to increase the amplitude of slow waves through a resonance effect, where sensory evoked potentials and/or K complexes occurring at the peak of a slow wave increase the amplitude of the oscillation[17,18]. Because the stimulation is accomplished using ordinary sound, CLAS has an excellent safety profile and can be performed using relatively simple hardware.

A limitation in the current use of CLAS however is that it requires EEG monitoring during sleep, in order to both determine when the user has entered slow wave sleep and to present sounds locked to slow wave peaks. This represents a significant limitation to the adoption of CLAS; while EEG devices for home have been developed, they face significant challenges related to comfort[19], signal quality[20], and cost, and are not widely adopted[21].

Wrist-worn smartwatches which can sense sleep dynamics and deliver stimulation at the appropriate time present a potential more accessible and comfortable approach for slow wave stimulation. Smartwatches like the Samsung Galaxy Watch, Fitbit, and Apple Watch have been adopted by nearly 44% of Americans, and algorithms using smartwatch motion and heart rate sensor data can perform sleep staging, albeit with less accuracy than laboratory EEG systems[22]. While these systems cannot phase lock stimuli to the phase of ongoing slow waves, previous research suggests that phase locking may not be necessary so long as the system can control timing and intensity to avoid disrupting sleep[23].

In this experiment, we tested whether slow wave entrainment using a Samsung Galaxy Watch 6 could increase slow wave and delta power in the brain while participants slept at home, and also whether stimulation produced benefits for sleep quality, mood, or cognition. We tested slow wave stimulation with both conventional pink noise stimuli used in previous studies, and also with gentle vibration from the watch.

Participants slept for two nights in their own homes wearing the watch as well as a portable EEG headband to record brain activity. Participants were randomized to receive stimulation on one of the two nights and completed cognitive, mood, and sleep assessments after each night. During deep sleep in the stimulation night, participants received alternating blocks with stimulation and without stimulation. We hypothesized that EEG delta power would increase

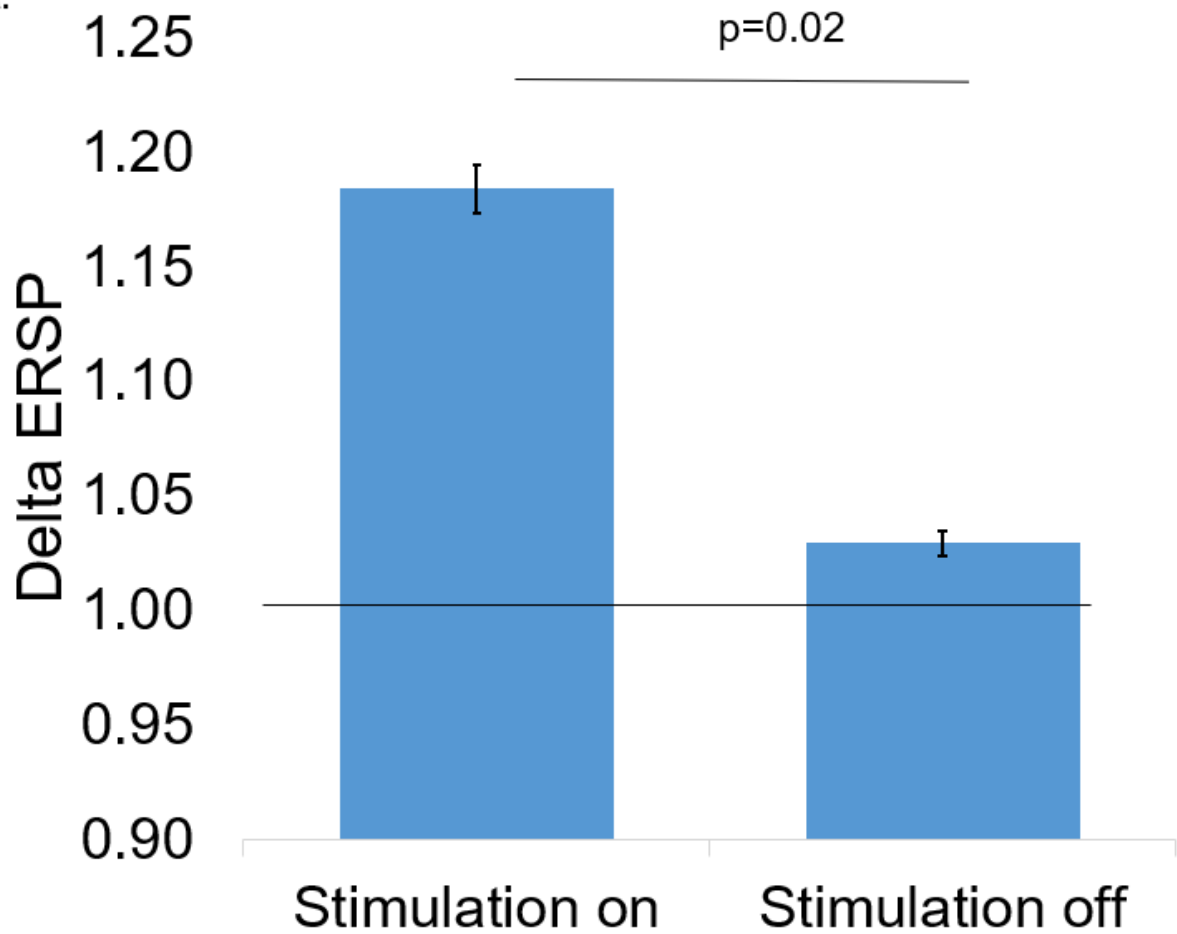
following the start of slow wave stimulation, and also that participants would have improved ratings of sleep quality, mood, and cognition following their stimulation night compared to the non-stimulation night.

Results

Slow wave entrainment increased EEG delta power

We found that frontal delta power (1-4 Hz in electrodes AF7 and AF8) increased when slow wave stimulation was initiated compared to a 30-second baseline before each stimulation period [one sample t test; $t(34)=3.04$; $p=0.005$]. Delta power also increased during stimulation blocks (Stimulation on) compared to the non-stimulation (Stimulation off) blocks [paired t test, $t(32)=2.55$, $p = 0.02$] (Figure 1A).

A.



B.

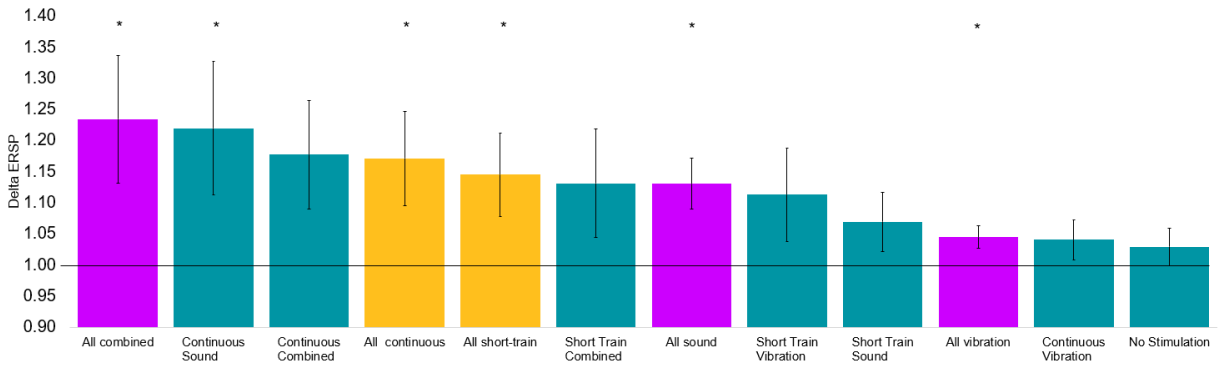
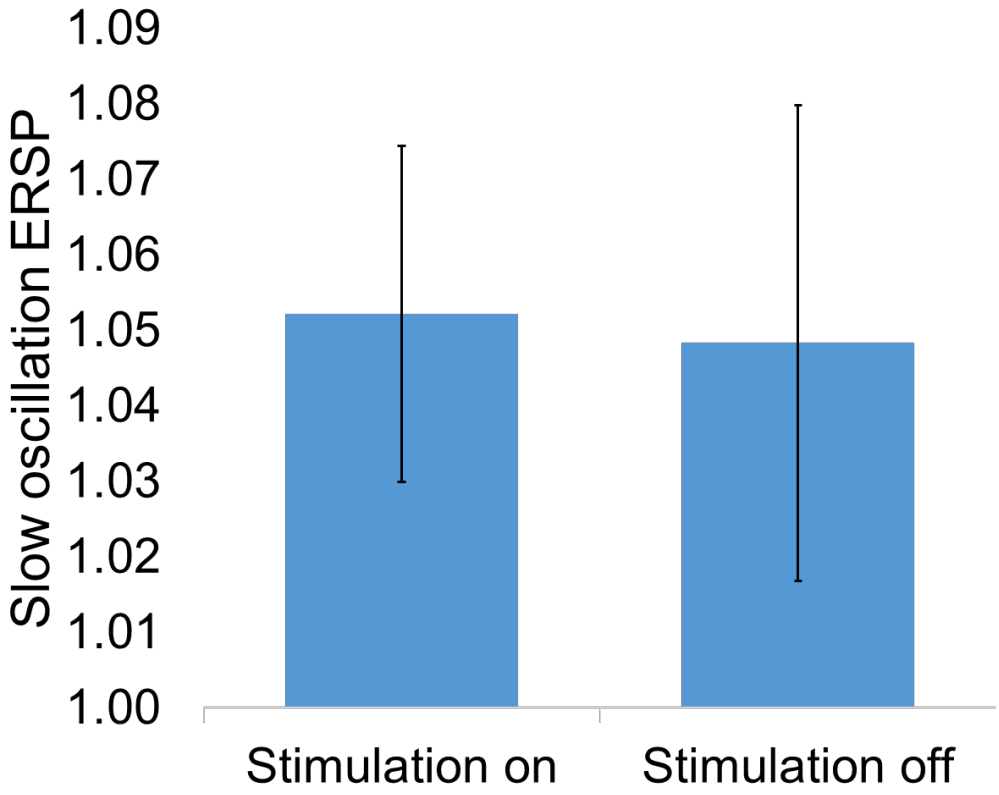


Figure 1: (A) Effects of slow wave stimulation on frontal (AF7/AF8) delta power quantified using event related spectral perturbation (ERSP) , quantified as delta power during block/delta power over previous 30 seconds. Delta power evoked by the stimulation was significantly greater than delta power evoked by the non-stimulation block. Both the stimulation and non-stimulation blocks had a significant increase in delta power over baseline (indicated by the horizontal line at Delta ERSP = 1.00); in the non-stimulation block this likely resulted from the natural increase in delta power as participants progressed into deeper N3 sleep, as delta amplitude slowly increases during transitions into N3[24]. Data are for 35 participants with delta band data available during stimulation. **(B)** Effects of specific stimulation types on frontal delta power. Blue bars indicate individual block types, whereas purple bars are aggregates of all trials using a specific modality and yellow bars are aggregates of all trials using a specific duty cycle. Asterisks indicate a significant power increase ($p < 0.05$) over baseline.

We also tested the effects of different stimulation modalities (sound, vibration and combined sound/vibration) and duty cycles (continuous and short train) in inducing delta waves. When pooling all of the trials in a particular condition (e.g., the “all sound” pool contained trials with short-train and continuous sound). Combined, vibration, and sound stimulation all elicited significant increases in delta power as did both the short-train and continuous stimulation (Figure 1B). While most individual block types did not show a significant increase due to a limited number of trials, continuous sound also significantly increased delta power.

29 participants had data with sufficient sampling rate and signal quality to perform offline measurement of the slow wave band (0.5-1.5 Hz) specifically. We measured slow wave power in the 20 seconds after the onset of each condition relative to the 20 seconds preceding it and computed the fold change in power. Power in the slow wave band increased significantly over baseline during stimulation, [$t(28)=2.35$, $p=0.03$], although this effect was only significant when aggregating over all stimulus types and no individual stimulus or pool significantly increased power (Figure 2).

A.



B.

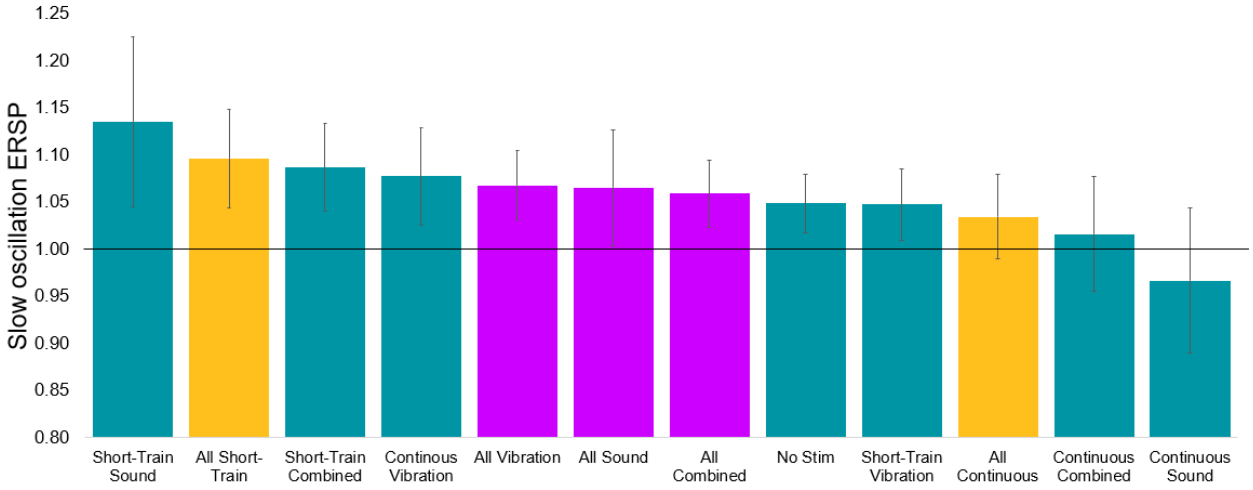


Figure 2: (A) Effects of slow wave stimulation on frontal (AF7/AF8) slow wave power. **(B)** effects of continuous and short-train stimuli on frontal slow wave power. Blue bars denote individual stimulus types while yellow bars denote multiple stimulus types grouped by duty cycle and purple bars denote stimulus types grouped by modality. Although there was an overall effect of stimulation on slow wave power, we did not see a significant effect for any individual

stimulus type. Data are for 29 participants with slow wave band data available during stimulation.

Slow wave entrainment decreased arousals in sleep

Presenting stimuli in sleep carries a risk of increased sleep fragmentation, which can decrease the overall quality of sleep[25,26]. To determine whether stimulation increased sleep fragmentation, we measured the proportion of blocks that contained a transition out of N3 for each condition, which we called the arousal rate.

We found that arousal rate was significantly lower for stimulation blocks than for non-stimulation blocks (Figure 3), indicating that stimulation did not cause sleep fragmentation.

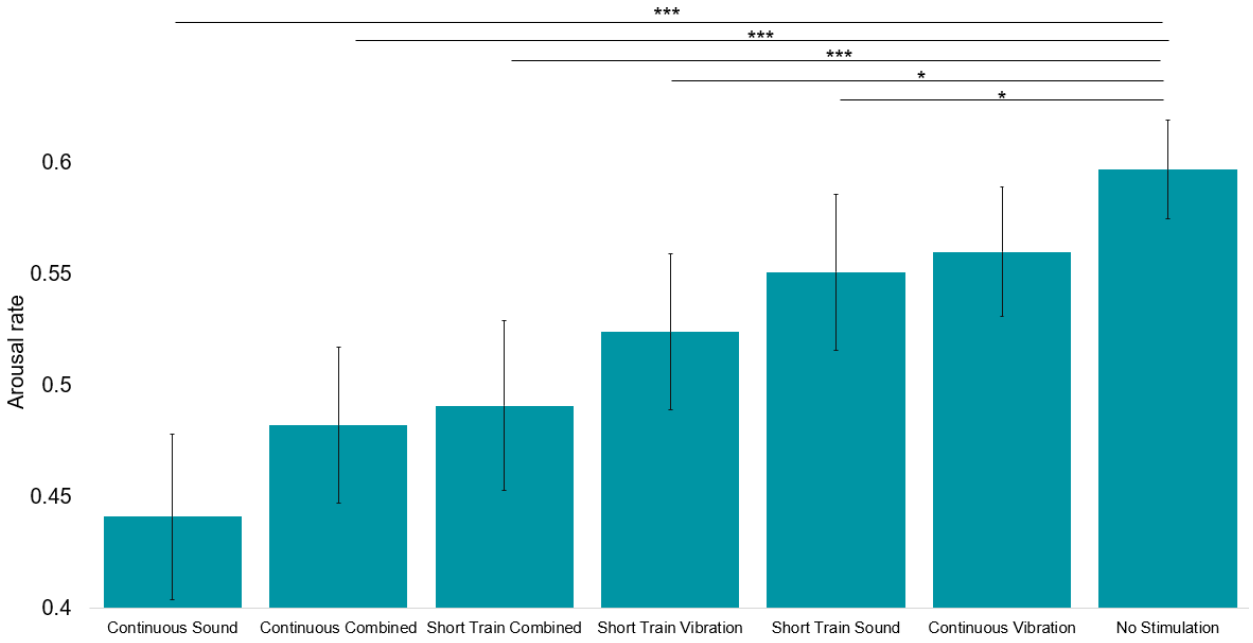


Figure 3: The effect of stimulation condition on arousal rate. Asterisks represent significant differences in arousal rate using a two-tailed t test (*: $p < 0.05$, ***: $p < 0.001$). All stimulation conditions other than the continuous vibration condition had a significantly lower arousal rate than the no-stimulation condition. Data are for 68 participants who received stimulation during the night.

Slow wave entrainment improved cognitive function and sleep in participants with large delta ERSPs

We did not find any significant main effects of stimulation on our measures of sleep quality, mood, and cognition. However further analysis revealed that the degree of delta wave

enhancement that occurred with stimulation was highly predictive of cognition and sleep improvements in the stimulation night; in particular, the size of the frontal delta ERSP to stimulation was associated with improvements in sleep quality, mood, and Trail Making Task performance (Table 1, Figure E1).

	r value x delta ERSP	P value x delta ERSP
LSEQ quality of sleep score difference	0.32	0.133
LSEQ awakening from sleep score difference	0.24	0.255
LSEQ behavior following waking score difference	0.52	0.009
Total LSEQ score difference	0.49	0.015
BRUMS anger score difference	-0.46	0.024
BRUMS confusion score difference	-0.32	0.130
BRUMS depression score difference	-0.22	0.311
BRUMS fatigue score difference	-0.40	0.052
BRUMS tension score difference	-0.37	0.074
BRUMS vigor score difference	0.18	0.394
BRUMS overall negative mood difference	-0.42	0.040
Trail making time difference	-0.23	0.277
Trail making click count difference	-0.44	0.033

Table 1: Correlation between the change in behavioral outcome measures from stimulation night to sham night and the increase in delta power evoked by stimulation on the sham night. A positive r value indicates that the stim-sham value was positively correlated with the amount of delta power induced by the stimulation. We found that increased delta power was associated with better self-reported sleep quality (LSEQ behavior following waking, total LSEQ) , reduced BRUMS anger, and reduced overall negative mood on the BRUMS). We also found increased delta power predicted a reduced number of clicks needed to complete the Trail Making Task, indicating a reduction in the number of errors on this task (since the Trail Making Task requires

a minimum of 26 clicks to complete it successfully). Data are for 24 participants who had valid EEG on both nights, stimulation on their stimulation night, and completed questionnaires after both nights.

We also tested for correlations between the number of stimuli given in the night and the effects on mood, cognition, and sleep. The total number of stimuli during the night was unrelated to most measures, but we did observe that it modulated the effect of stimulation on the awakening from sleep scale, with a higher number of stimuli associated with easier awakening (Table E1, Figure E2).

Slow wave entrainment improved mood in participants with poor baseline sleep health

We tested whether the difference between the stimulation night and non-stimulation night was correlated with participant factors including age, typical bedtime, typical sleep duration,Pittsburgh Sleep Quality Index (PSQI) PSQI and General Health Questionnaire (GHQ-12) scores. We found that participant baseline sleep health (PSQI score) modulated the effects of stimulation on confusion and tension mood scales; participants who reported more sleep problems on the PSQI at the start of the experiment showed a larger reduction of tension and confusion after their stimulation night (Table 2, Figure E3). Participant age, typical sleep duration, typical bedtime, and GHQ-12 score did not significantly modulate the effects of slow wave entrainment.

	r-Val Age	p-Val Age	r-Val Bed Time	p-Val Bed Time	r-Val Typical Sleep Duration	p-Val Typical Sleep Duration	r-Val GHQ12 Score	p-Val GHQ12 Score	r-Val PSQI Score	p-Val PSQI Score
LSEQ Quality of sleep score difference	0.013	0.938	0.060	0.709	0.061	0.707	-0.150	0.348	-0.113	0.482
LSEQ awakening from sleep score difference	0.025	0.875	-0.070	0.665	0.039	0.807	0.002	0.991	-0.060	0.711
LSEQ behavior following wakefulness score difference	0.099	0.539	0.250	0.118	0.037	0.819	0.210	0.187	-0.067	0.678
Total LSEQ score difference	0.053	0.744	0.091	0.571	0.065	0.687	-0.013	0.934	-0.114	0.476
BRUMS anger score difference	-0.226	0.155	0.049	0.761	0.213	0.182	-0.254	0.109	-0.127	0.430
BRUMS confusion score	-0.216	0.175	-0.054	0.739	0.185	0.248	-0.271	0.086	-0.317	0.044

difference										
BRUMS depression score difference	-0.215	0.176	0.111	0.489	0.244	0.124	-0.135	0.401	-0.130	0.418
BRUMS fatigue score difference	-0.106	0.511	-0.154	0.082	0.082	0.609	-0.23	0.140	0.068	0.674
BRUMS tension score difference	-0.195	0.222	0.138	0.391	0.205	0.199	-0.064	0.689	-0.402	0.009
BRUMS vigor score difference	0.036	0.823	-0.052	0.746	0.033	0.836	0.003	0.990	-0.067	0.680
BRUMS overall negative mood difference	-0.206	0.196	-0.003	0.986	0.174	0.277	-0.226	0.155	-0.124	0.440
Trail making time difference	-0.070	0.663	-0.035	0.829	0.004	0.982	0.065	0.684	0.095	0.554
Trail making click count difference	-0.017	0.918	0.028	0.915	-0.043	0.787	0.094	0.560	0.076	0.637
Stimulus-specific delta ERSP	0.096	0.600	0.130	0.478	-0.008	0.965	0.228	0.210	-0.096	0.601

Table 2: Correlations between participant demographics and the effects of slow wave stimulation. Effects of the stimulus on all measures except stimulus specific delta ERSP were measured by computing the difference (stimulation night - no stimulation night) for the metrics and are based on 41 participants who completed the initial questionnaires and both nightly questionnaires. Stimulus-specific delta ERSP reflects the difference between the delta ERSP in all stimulation conditions combined and the ERSP in the non-stimulus conditions for 32 participants who completed the initial questionnaire and also had EEG delta data available on their stimulation night for both stimulation and no-stimulation conditions.

Entrainment stimuli targeted N3 sleep

To assess which sleep stages the system targeted for stimulation, a trained sleep scorer performed manual sleep staging on the -30 to 0 second period relative to the start of stimulation blocks using the AASM criteria[27] and also staged 10 randomly selected 90 second periods of sleep in each night to create a baseline distribution of sleep stages. We considered only the period before stimulation started in order to measure which phases of sleep the system targeted for stimulation without the confound of changes in sleep physiology induced by the stimulation. We obtained baseline sleep stage distributions for 17 nights which had stageable Muse EEG and determined sleep stages of stimulation onsets in 10 nights which had both Muse EEG data and stimulations.

Stimulation was significantly more likely to occur in N3 compared to to the baseline distribution (odds ratio=1.75, $X^2(1, 229)=5.01$, $p=0.025$) and was significantly less likely to occur in wake (odds ratio=0.18, $X^2(1, 229)=10.86$, $p=0.001$) as shown in figure E4. The rate of stimulation in

N1, N2, and REM sleep did not differ significantly from the baseline occurrence of those stages.

Despite targeting N3, stimulation occurred most frequently in stage N2 (33% of stimulus onsets) due to the higher base rate of N2 compared to N3. Stimulation occurred next most frequently in N3 (29% of onsets) REM (28% of onsets), N1 (8% of onsets) and wake (3% of onsets).

Participants rarely reported perceiving entrainment stimuli

Out of 86 participants who completed the post-sleep questionnaire, 12 reported perceiving sound or vibration stimuli at some point during the night and of these 5 reported that the stimuli disturbed their sleep. The remaining participants who perceived stimuli reported that they occurred either while falling asleep or when they had awakened for some other reason.

Discussion

In this study we found that rhythmic sound and vibration delivered by a smartwatch could significantly enhance delta and slow wave power in the brain, and these increases were associated with improvements in objective and subjective sleep quality, mood, and cognitive function.

Our data show that slow wave entrainment was broadly effective in increasing delta and slow wave amplitude, but we did not observe clear differences between the different stimulus modalities. One reason for this may be a limited number of trials (mean of 3 ± 2.16 blocks per condition per night) in each modality that made it difficult to resolve small effects. In addition, all modalities and duty cycles had a mean delta ERSP above 1, suggesting that they were all at least somewhat effective at increasing delta power. For slow wave power, there was a numerical trend for short-train stimuli to be more effective than continuous stimuli.

Increasing delta power with a smartwatch also appeared to play a functional role, improving measurements of sleep quality, mood, and cognition. While we did not observe outright significant differences in questionnaire scores between the stimulation and no-stimulation nights, we did observe differences that were mediated by the amount of delta power increase induced by stimulation. This ERSP therefore reflects specifically the change in delta activity induced by stimulation, and therefore our results support the hypothesis that increasing delta power improves sleep quality and cognition.

We also found evidence for a reduced arousal rate during stimulation compared to matched no-stimulation periods. Slow wave entrainment may reduce arousability during sleep; as slow wave sleep and sensory-elicited K complexes have both been linked to deep, unarousable sleep, and K complexes are suspected to play an active role in suppressing sensory stimuli that would otherwise cause awakening[28,29].

We also found evidence that individual differences may influence how people respond to slow wave entrainment. Participants with high levels of reported sleep problems (PSQI score)

showed significantly larger improvements in tension and confusion following stimulation than participants with low PSQI scores, suggesting slow wave entrainment may be particularly beneficial for individuals with poor baseline sleep quality health. There may also be individual differences in how people respond to specific types of stimulation (such as haptic or sound) though we did not investigate these here. Future studies should explore whether the most effective stimulation methods (such as short-train and continuous sound) produce robust benefits.

A novel aspect of our system is that we present entrainment stimuli irrespective of the phase of ongoing slow wave activity in the brain. Most research on slow wave entrainment has used phase-locked stimulation[30], where stimuli are presented at the peak of naturally occurring slow waves detected using EEG; though other trials have also showed improved sleep quality with non-phase-locked stimulation[23,31] . Our data suggests that non-phase locked stimulation produced an increase in slow wave power equivalent to that seen with phase-locked stimulation, consistent with prior studies that have used non-phase-locked entrainment. While some research has suggested that phase locked entrainment is more effective [32,33], these studies did not directly compare phase locked entrainment against non-phase locked entrainment with a constant rhythm; rather the non-phase locked condition had the phase randomized on each stimulus cycle. It is well known that the brain can phase lock its own delta waves to rhythmic stimuli[34]; therefore constant rhythms may be effective because the brain adjusts the ongoing slow wave rhythms to match the stimulus.

In some cases non-phase-locked entrainment may produce larger benefits than phase-locked entrainment. Prior research has found that slow wave entrainment may be less effective for older adults than younger adults [35–37]. By contrast, our study found that age did not modulate the degree of slow wave entrainment or the effects of slow wave entrainment on mood, cognition, and sleep quality, despite including participants across ages 18-92. One possibility is that that entrainment with non-phase-locked stimuli may be more effective in older adults than phase-locked stimulation, because it does not depend on detecting slow wave peaks, which tend to be attenuated in older adults. Additionally, non-phase locked entrainment is advantageous in that it does not require EEG, allowing entrainment to be incorporated in highly usable consumer devices like smartwatches.

Comparing the stimulus onsets with manual sleep staging based on EEG revealed that the proportion of stimuli presented in N3 was greater than expected by chance (Figure E4), confirming that the system was able to detect and stimulate in N3 sleep. We also observed a substantial amount of off-target stimulation (in N2 and REM) which is likely attributable to two sources. This is likely because we used a liberal N3 detection threshold in this experiment to ensure that participants received enough stimulation to allow comparisons between different experimental conditions; increasing the threshold would increase the specificity for N3.

Despite the presence of off-target stimulation, we observed reliable increases in delta and slow wave amplitude. While slow wave amplitude is highest in N3 sleep, slow waves are also found in N2 and REM sleep[38,39]; our results suggest the possibility (which could be explored in future

studies) that slow wave entrainment in non-N3 states has similar physiological effects to stimulation in N3.

Some limitations in this study stem from the fact that participants performed the experiment in their own homes using portable EEG equipment. While this protocol allowed us to collect data from a large and diverse cohort of participants; a limitation is that some aspects of the experiment were less standardized than typical experiments conducted in a sleep laboratory. For example, a large number of participants did not obtain EEG data with sufficient quality for one or more analyses, and we were not able to standardize the times at which participants took the post-sleep surveys. Despite these limitations, we found a converging and credible pattern of results: analysis using both the delta power computed by the Muse EEG and the slow oscillation power computed manually found that stimulation increased delta/SO power, and we found that the magnitude of this increase was related to the size of the behavioral effects.

Our results also show the feasibility of using smartwatch-based sensory stimulation to study sleep. While many previous studies have used watches for passive recording of physiology during sleep[40], sensory stimulation guided by real time sleep staging can enable new types of sleep experiments to be conducted at home using watches; for example experiments using sensory stimulation to probe memory reactivation, arousability, or interventions for sleep problems. These technological advancements can in turn benefit sleep science; by expanding the pool of experiments that can be run outside a sleep lab researchers could obtain larger and more diverse samples of participants.

Overall, our results suggest that slow wave entrainment using a novel, smartwatch-based system is effective at enhancing slow wave activity during sleep, and that increasing slow wave activity can improve measures of mood, cognition, and sleep quality. Future clinical research should investigate whether optimized smartwatch entrainment protocols can produce meaningful improvements in sleep quality and health.

Methods

Participants

The study was conducted with a total of 93 participants (27 male, 66 female) ranging in age from 18-92 years old (mean=43, median=38, SEM=2). Participants were recruited from the Boston/Cambridge community and also from the Laselle Village retirement community.The study was approved by the MIT IRB and all participants gave informed consent at the beginning of the study.

Because participants recorded data in their own homes, and were not always successful at doing so, not all participants had all types of data available. However we found that the

subgroups of participants with specific data available had similar demographics to the overall sample (Table S1).

Experimental procedure

The overall experimental design is shown in Figure 4. Participants visited the laboratory on day 1 when they collected the study equipment, completed consent forms, and were oriented to the devices used. On day 1, participants also completed their initial surveys, consisting of the Pittsburgh Sleep Quality Index (PSQI)[24] and the General Health Questionnaire (GHQ-12)[,25], as well as questions about their age and gender.

Night 1 and night 2 were experimental nights; participants were randomly assigned to receive slow wave entrainment on either night 1 or night 2. On each night, participants donned the Muse EEG device and the Galaxy watch and began recording EEG data before going to sleep. Following both night 1 and night 2, participants were emailed the post-sleep questionnaire. On this questionnaire, participants rated their sleep quality the previous night (Leeds Sleep Evaluation Questionnaire (LSEQ)[43]), current mood (Brunel Mood Scale (BRUMS)[44]), and performed a modified computerized version of the Trail Making Task B[45] (Figure 4C) to measure overall cognitive function. After completing night 2, participants returned the equipment to the lab.

While participants slept, our on-watch sleep staging model performed sleep staging once per second for the duration of the whole night. Slow wave stimulation was delivered on stimulation nights when our custom on-watch sleep staging model determined the participant was likely in stage N3 sleep. The watch then began the stimulation protocol. We tested three distinct types of stimulation: stimulation with 50ms pink noise pulses[23], stimulation with 50ms vibration pulses, and stimulation with combined sound and vibration. Additionally, we tested two stimulation duty cycles: A “continuous” mode where stimuli were presented every 1.2 seconds, and a “short-train” mode where two stimuli were presented 1.2 seconds apart every 10 seconds. Therefore, there were a total of 6 stimulation conditions. Stimulation conditions were presented in blocks which lasted either 120 seconds (for the first 37 participants) or 60 seconds (for subsequent participants). After each stimulation block was an equally long no-stimulation block where no stimuli were presented. The watch continued monitoring sleep state during the stimulus and no stimulus blocks and would pause the protocol if the user left N3, resuming at the same point in the protocol once stable N3 resumed.

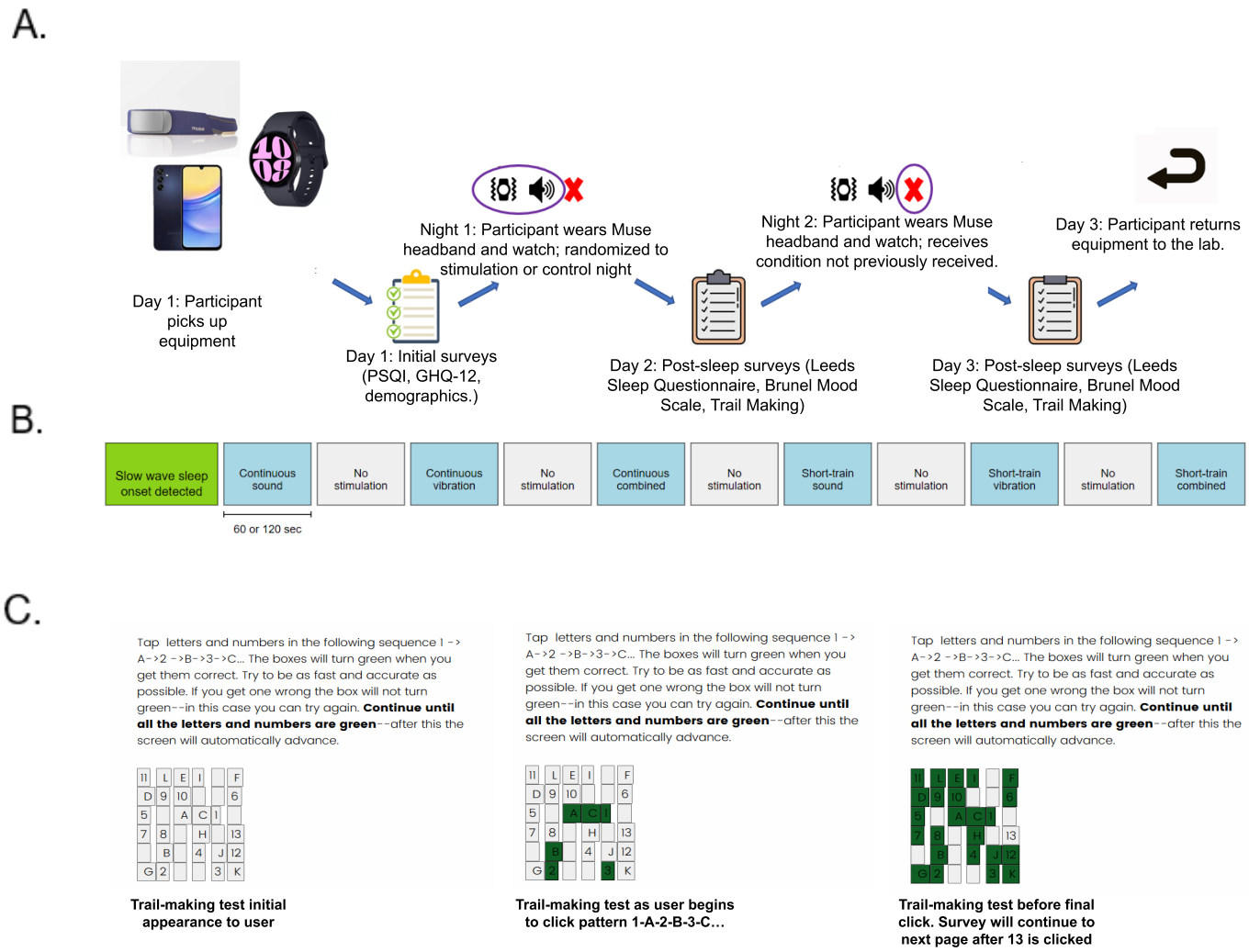


Figure 4: (A) Sequence of events in the study. **(B)** Stimulation protocol used on stimulation nights. The stimulation blocks were presented in a random order with replacement such that a participant would first receive one block of each type before repeating the blocks. The block length was 2 minutes (120 seconds) for the first group of 37 participants and shortened to 60 for

the remaining 57 participants to maximize the chances of participants receiving each block type. If an arousal occurred during stimulation, stimulation was paused and restarted at the same point in the stimulation sequence once the participant again met the threshold for stimulation. **(C)** Example of how participants complete the computerized trail making task.

When the watch detected a transition out of deep sleep (either the probability of N3 falling below 20% or a large body movement), the stimulation was immediately paused. Once the participant returned to deep sleep, stimulation was resumed in the same condition.

Smartwatch-based deep sleep detection

To deliver slow wave stimulation using the Galaxy watch 6, we developed a custom app which ran continuously on the watches. We developed a machine learning model to predict the sleep stages given by a home PSG device (Dreem 2) using the heart rate, gyroscope, and accelerometer data collected by a smartwatch. The model was trained to perform 5-stage sleep staging (Wake, N1, N2, N3 and REM) using data collected in a previous study [46]; the data consisted of approximately 12 million data points collected from participants ages 18-35. Receiver Operating Characteristic curve (ROC-AUC) for detecting N3 (versus non-N3) was 0.6 each second, indicating that the model detected N3 at above chance rates. More details about the model architecture can be found in Supplemental Note 1.

When participants slept, the app estimated sleep stages once per second based on the heart rate and motion data. Stimulation began when the following criteria were met, and was paused whenever any were not met:

1. 10-minute average probability of N3 was >=25%
2. Probability of N3 in the last second was >=25%
3. No large body movements (more than 3 degree/sec motion on gyroscope reading) were detected in the last 60 seconds
4. Participant has been asleep for at least 30 minutes
5. No more than 30 min of stimulation have been delivered this night (stimulation duration was limited as little literature exists on the effects of delivering N3 stimulation for an extended period).

Stimulus intensity control

To present stimuli at effective intensity without awakening the user, we developed an adaptive algorithm for controlling the intensity of auditory and vibration stimuli. For each stimulus type, the intensity begins at 0 at the beginning of a night of sleep and increases slightly each time a stimulus is presented. If the user transitions out of deep sleep while stimuli are being presented, the system sets a cap on the intensity of that stimulus which is 95% of the intensity where the transition out of deep sleep occurred. For the rest of the night, stimulus intensity is not permitted

to exceed the cap; the cap can also be decreased further if another transition out of deep sleep occurs during stimulation. Caps are set independently for auditory and vibration stimulation; for example if a user transitions out of deep sleep during vibration stimulation only the vibration intensity will be capped. Caps are reset at the end of each night.

EEG recording

Participants recorded their EEG while sleeping at home using an Interaxon Muse S headband linked to a smartphone provided by the lab. When participants visited the lab, we demonstrated how to use the headband and adjust it to obtain good signal quality. Participants wore the headband on both nights of the experiment. Data was recorded using the Mind Monitor app [47].

Analysis of event-related delta power

For our primary analysis, we computed event-related delta power using the absolute delta power (1-4 Hz) calculated by the Muse S on channels AF7 and AF8 once per second. We used the stimulation records recorded by the watch to insert events in the Muse time series when each block of stimulation started, allowing us to compute how delta power changed when stimulation began compared to the period before stimulation. We considered the onset of each block to be one trial, and we computed the change in delta power by dividing the delta power from 0-30 seconds after the trial start by the delta power from -30 to 0 seconds before the trial start. We computed the event-related change separately for AF7 and AF8.

We then computed an overall event related change for each trial by averaging the change in the AF7 and AF8 channels. Because channels can become disconnected from the head during sleep, we computed the overall event related spectral change by averaging AF7 and AF8 if both channels were connected; if one channel was disconnected then the overall value was the value of the remaining connected channel. If both channels were disconnected during a trial, that trial was excluded completely. We detected disconnected channels by looking for event-related spectral power (ERSP) values of 1.0, which indicated that the channel received only noise.

After obtaining a single value for each trial, we then computed the mean event related change for each participant and each stimulation condition. We then computed the mean event related change across participants and the corresponding standard error of the mean.

Analysis of event-related slow oscillation power

To measure the event-related changes in the slow wave power band specifically, we imported Muse EEG data into EEGLAB version 2023.1 [48]. We performed this analysis using a filter-Hilbert method to extract power in the 0.5-1.5 Hz range. We first epoched continuous EEG to the start of each block (time range -20 to 20 seconds relative to the block onset); we then performed amplitude-based epoch rejection where we rejected any epochs containing voltage excursions of >500 or < -500 microvolts. Epoched data were then filtered using a bandpass FIR filter (pop_eegfiltnew) with cutoff frequencies of 0.5 Hz and 1.5 Hz. After filtering, we took the

magnitude of the Hilbert transform for each channel and divided the magnitude after block onset by the magnitude before block onset to obtain the ERSP.

Analysis of sleep stages where stimulation was delivered

To determine sleep stages when stimulation was delivered, we bandpass filtered the raw Muse data (sampled at 256 Hz) between 0.1 and 40 Hz and re-referenced it to TP10 (approximating the standard montage for lab based sleep staging). If data were too noisy to allow for sleep staging, data were referenced to a different electrode (either TP9, AF7, or AF8) and/or additionally filtered with a lowpass filter at 25 Hz. Data were then epoched from -60 to +30 seconds relative to block starts. For our primary analysis, a trained sleep scorer determined the AASM sleep stage[27] in the -30-0 epoch relative to the beginning of stimulation blocks. This time period was chosen because it represents the sleep stages in which the system initiated stimulation, uninfluenced by changes in sleep induced by stimulation.

To determine if our system successfully targeted N3, we also constructed a baseline distribution for the prevalence of each sleep stage by randomly sampling 10 90-second windows from 17 night recordings with stagable Muse EEG. We used the middle 30 seconds of each window to construct a randomly baseline distribution of sleep stages for our participants. Only the middle 30 seconds were used as accurate sleep staging requires also staging 30 seconds before the epoch to be analyzed and 30 seconds after. We then used a chi squared test to test whether the percentage of stimulation in each sleep stage differed significantly from the proportion of that stage in the baseline. A significant chi-square would indicate that the system either targeted in that stage or avoided stimulating in that stage

Analysis of arousal rate

To measure whether stimulation disrupted sleep, we measured the arousal rate for each stimulation condition and the no-stimulation condition. We defined the arousal rate as the number of times the participant transitioned from deep sleep (defined by the criterion in “Smartwatch-based deep sleep detection”) to lighter sleep per block, for each block type. Transitions to light sleep can include waking, but can also include transitions to a different sleep stage such as N3 to N2 or N3 to REM. This metric therefore provides a measurement of sleep disruptions per unit of time, similar to common EEG-based metrics like the sleep fragmentation index[49].

User survey data

To examine the psychological and cognitive effects of slow wave stimulation, we collected several surveys after each night of sleep and compared results after the sham night to results after the stimulation night.

Sleep quality was obtained using the Leeds Sleep Evaluation Questionnaire (LSEQ)[43], which asks participants to rate the ease of falling asleep, quality of sleep, ease of waking up, and feeling during the day. We measured all subscales of this questionnaire except for the Getting To Sleep scale, which we considered not relevant since stimulation began after sleep onset. Participants also completed the Brunel Mood Scale (BRUMS)[44], which evaluates their current state on 7 axes (tension, depression, anger, vigor, fatigue, and confusion), plus a composite axis (overall negative mood) which reflects the sum of all other axes minus vigor.

To evaluate cognitive function, we developed an online version of the Trail Making Test part B (Trails B), a test of planning, psychomotor speed, and executive function that is sensitive to sleep deprivation [50,51]. Trails B requires participants to draw a line connecting a series of randomly spaced letters and numbers in the order 1 -> A -> 2 -> B …etc. The primary measure obtained from this task is the time taken to complete it without errors.

Trails B is typically a pen and paper task; to adapt it to an online study we created a computerized version of it (Figure 4C). In our modified trails B, 13 letters and 12 numbers are randomly distributed on a 6 x 6 grid. Participants are asked to begin with tapping the 1, and then A, continuing to 2->B until reaching L. Each letter/number turns green if it is tapped in the correct order; if a participant taps an incorrect letter/number they must find the correct one and tap it to continue. We record the total time to complete the task, and also the number of taps required to complete it.

Evaluating the sleep quality, mood, and cognitive effects of slow wave entrainment
To measure whether slow wave entrainment affected sleep quality, mood during the day, or performance on the Trails B task, we compared test performance after stimulation and no-stimulation nights using a paired t test.

We also tested whether specific stimulation parameters and individual factors predicted the effects of stimulation. For these analyses, we used linear regression to test whether each factor predicted the difference in test score following the stimulation night and test score following the sham night. Each combination of factor/test was tested individually (i.e., we did not perform multiple regression).

Works Cited

1. Nelson, K. L., Davis, J. E. & Corbett, C. F. Sleep quality: An evolutionary concept analysis. *Nurs. Forum (Auckl.)* **57**, 144–151 (2022).
2. Van Cauter, E., Spiegel, K., Tasali, E. & Leproult, R. Metabolic consequences of sleep and sleep loss. *Sleep Med.* **9**, S23–S28 (2008).

3. Chellappa, S. L. & Aeschbach, D. Sleep and anxiety: From mechanisms to interventions. *Sleep Med. Rev.* **61**, 101583 (2022).
4. Frangopoulos, F. *et al.* The Complex Interaction Between the Major Sleep Symptoms, the Severity of Obstructive Sleep Apnea, and Sleep Quality. *Front. Psychiatry* **12**, (2021).
5. Irish, L. A., Kline, C. E., Gunn, H. E., Buysse, D. J. & Hall, M. H. The role of sleep hygiene in promoting public health: A review of empirical evidence. *Sleep Med. Rev.* **22**, 23–36 (2015).
6. Xu, X., Lian, Z., Shen, J., Lan, L. & Sun, Y. Environmental factors affecting sleep quality in summer: a field study in Shanghai, China. *J. Therm. Biol.* **99**, 102977 (2021).
7. Riemann, D. *et al.* The hyperarousal model of insomnia: A review of the concept and its evidence. *Sleep Med. Rev.* **14**, 19–31 (2010).
8. Xiao, X., Rui, Y., Jin, Y. & Chen, M. Relationship of Sleep Disorder with Neurodegenerative and Psychiatric Diseases: An Updated Review. *Neurochem. Res.* **49**, 568–582 (2024).
9. Matthews, E. E., Arnedt, J. T., McCarthy, M. S., Cuddihy, L. J. & Aloia, M. S. Adherence to Cognitive Behavioral Therapy for Insomnia: A Systematic Review. *Sleep Med. Rev.* **17**, 10.1016/j.smrv.2013.01.001 (2013).
10. Matheson, E. & Hainer, B. L. Insomnia: Pharmacologic Therapy. *Am. Fam. Physician* **96**, 29–35 (2017).
11. Halász, P., Bódizs, R., Parrino, L. & Terzano, M. Two features of sleep slow waves: homeostatic and reactive aspects – from long term to instant sleep homeostasis. *Sleep Med.* **15**, 1184–1195 (2014).
12. Diekelmann, S. Sleep for cognitive enhancement. *Front. Syst. Neurosci.* **8**, (2014).
13. Zhang, Y. & Gruber, R. Can Slow-Wave Sleep Enhancement Improve Memory? A Review of Current Approaches and Cognitive Outcomes. *Yale J. Biol. Med.* **92**, 63–80 (2019).
14. Tasali, E., Leproult, R., Ehrmann, D. A. & Van Cauter, E. Slow-wave sleep and the risk of type 2 diabetes in humans. *Proc. Natl. Acad. Sci.* **105**, 1044–1049 (2008).

15. Fung, M. M. *et al.* Decreased slow wave sleep increases risk of developing hypertension in elderly men. *Hypertens. Dallas Tex 1979* **58**, 596–603 (2011).
16. Bellesi, M., Riedner, B. A., Garcia-Molina, G. N., Cirelli, C. & Tononi, G. Enhancement of sleep slow waves: underlying mechanisms and practical consequences. *Front. Syst. Neurosci.* **8**, (2014).
17. Esfahani, M. J. *et al.* Closed-loop auditory stimulation of sleep slow oscillations: Basic principles and best practices. *Neurosci. Biobehav. Rev.* **153**, 105379 (2023).
18. Ngo, H.-V. V., Martinetz, T., Born, J. & Mölle, M. Auditory closed-loop stimulation of the sleep slow oscillation enhances memory. *Neuron* **78**, 545–553 (2013).
19. Hanzal, S., Tvrda, L. & Harvey, M. An Investigation into Discomfort and Fatigue Related to the Wearing of an EEG Neurofeedback Headset. 2023.02.16.23284115 Preprint at https://doi.org/10.1101/2023.02.16.23284115 (2023).
20. Keller, L. S. *et al.* Acceptability, Feasibility, and Day-to-Day Stability of a Portable Sleep Recording Device in Adolescents versus Young Adults. *MedRxiv Prepr. Serv. Health Sci.* 2025.04.13.25325346 (2025) doi:10.1101/2025.04.13.25325346.
21. EEG Equipment Market Size, Growth, Share and Forecast 2032. *Credence Research Inc.* https://www.credenceresearch.com/report/eeg-equipment-market (2024).
22. Birrer, V., Elgendi, M., Lambercy, O. & Menon, C. Evaluating reliability in wearable devices for sleep staging. *Npj Digit. Med.* **7**, 74 (2024).
23. Ngo, H.-V. V., Claussen, J. C., Born, J. & Mölle, M. Induction of slow oscillations by rhythmic acoustic stimulation. *J. Sleep Res.* **22**, 22–31 (2013).
24. Hussain, I. *et al.* Quantitative Evaluation of EEG-Biomarkers for Prediction of Sleep Stages. *Sensors* **22**, 3079 (2022).
25. Stepanski, E. J. The Effect of Sleep Fragmentation on Daytime Function. *Sleep* **25**, 268–276 (2002).
26. Stepanski, E. J. The Effect of Sleep Fragmentation on Daytime Function. *Sleep J. Sleep*

Sleep Disord. Res. **25**, 268–276 (2002).

27. Berry, R. B. *et al.* The AASM manual for the scoring of sleep and associated events. *Rules Terminol. Tech. Specif. Darien Ill. Am. Acad. Sleep Med.* **176**, 7 (2012).
28. Forget, D., Morin, C. M. & Bastien, C. H. The Role of the Spontaneous and Evoked K-Complex in Good-Sleeper Controls and in Individuals with Insomnia. *Sleep* **34**, 1251–1260 (2011).
29. Fasiello, E. *et al.* Sleep stage transitions driven by K-complex: Electrophysiological patterns in insomnia disorder. (2024) doi:10.1111/jsr.14291.
30. Ngo, H.-V. V., Martinetz, T., Born, J. & Mölle, M. Auditory closed-loop stimulation of the sleep slow oscillation enhances memory. *Neuron* **78**, 545–553 (2013).
31. Tononi, G., Riedner, B. A., Hulse, B. K., Ferrarelli, F. & Sarasso, S. Enhancing sleep slow waves with natural stimuli. (2010).
32. 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.* **44**, 2357–2368 (2016).
33. Roüast, N. M., Kumral, D., Gais, S. & Schönauer, M. Random auditory stimulation disturbs traveling slow waves and declarative memory. 2025.02.06.636865 Preprint at https://doi.org/10.1101/2025.02.06.636865 (2025).
34. Will, U. & Berg, E. Brain wave synchronization and entrainment to periodic acoustic stimuli. *Neurosci. Lett.* **424**, 55–60 (2007).
35. Schneider, J., Lewis, P. A., Koester, D., Born, J. & Ngo, H.-V. V. Susceptibility to auditory closed-loop stimulation of sleep slow oscillations changes with age. *Sleep* **43**, zsaa111 (2020).
36. Wunderlin, M., Zeller, C. J., Wicki, K., Nissen, C. & Züst, M. A. Acoustic stimulation during slow wave sleep shows delayed effects on memory performance in older adults. *Front. Sleep* **2**, (2024).

37. Sauvé, S. A., Bolt, E. L. W., Fleming, D. & Zendel, B. R. The impact of aging on neurophysiological entrainment to a metronome. *NeuroReport* **30**, 730 (2019).
38. Langille, J. J. Human REM Sleep Delta Waves and the Blurring Distinction between NREM and REM Sleep. *J. Neurosci.* **39**, 5244–5246 (2019).
39. Guo, D. *et al.* Slow wave synchronization and sleep state transitions. *Sci. Rep.* **12**, 7467 (2022).
40. Baron, K. G. *et al.* Feeling validated yet? A scoping review of the use of consumer-targeted wearable and mobile technology to measure and improve sleep. *Sleep Med. Rev.* **40**, 151–159 (2018).
41. Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R. & Kupfer, D. J. The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Res.* **28**, 193–213 (1989).
42. Goldberg, D. P. & Blackwell, B. Psychiatric Illness in General Practice: A Detailed Study Using a New Method of Case Identification. *Br. Med. J.* **2**, 439–443 (1970).
43. Parrott, A. C. & Hindmarch, I. Factor analysis of a sleep evaluation questionnaire. *Psychol. Med.* **8**, 325–329 (1978).
44. Terry, P. C., Lane, A. M., Lane, H. J. & Keohane, L. Development and validation of a mood measure for adolescents. *J. Sports Sci.* **17**, 861–872 (1999).
45. Bucks, R. S. Trail-Making Test. in *Encyclopedia of Behavioral Medicine* 1986–1987 (Springer, New York, NY, 2013). doi:10.1007/978-1-4419-1005-9_1538.
46. Whitmore, N. W., Harris, J. C., Kovach, T. & Paller, K. A. Improving memory via automated targeted memory reactivation during sleep. *J. Sleep Res.* **31**, e13731 (2022).
47. Mind Monitor - Apps on Google Play. https://play.google.com/store/apps/details?id=com.sonicPenguins.museMonitor&hl=en_US.
48. Delorme, A. & Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **134**, 9–21

(2004).

49. Haba-Rubio, J., Ibanez, V. & Sforza, E. An alternative measure of sleep fragmentation in clinical practice: the sleep fragmentation index. *Sleep Med.* **5**, 577–581 (2004).

50. Bowie, C. R. & Harvey, P. D. Administration and interpretation of the Trail Making Test. *Nat. Protoc.* **1**, 2277–2281 (2006).

51. Wimmer, F., Hoffmann, R. F., Bonato, R. A. & Moffitt, A. R. The effects of sleep deprivation on divergent thinking and attention processes. *J. Sleep Res.* **1**, 223–230 (1992).

52. Arnal, P. J. *et al.* The Dreem Headband compared to polysomnography for electroencephalographic signal acquisition and sleep staging. *Sleep* **43**, zsaa097 (2020).

Extended data figures

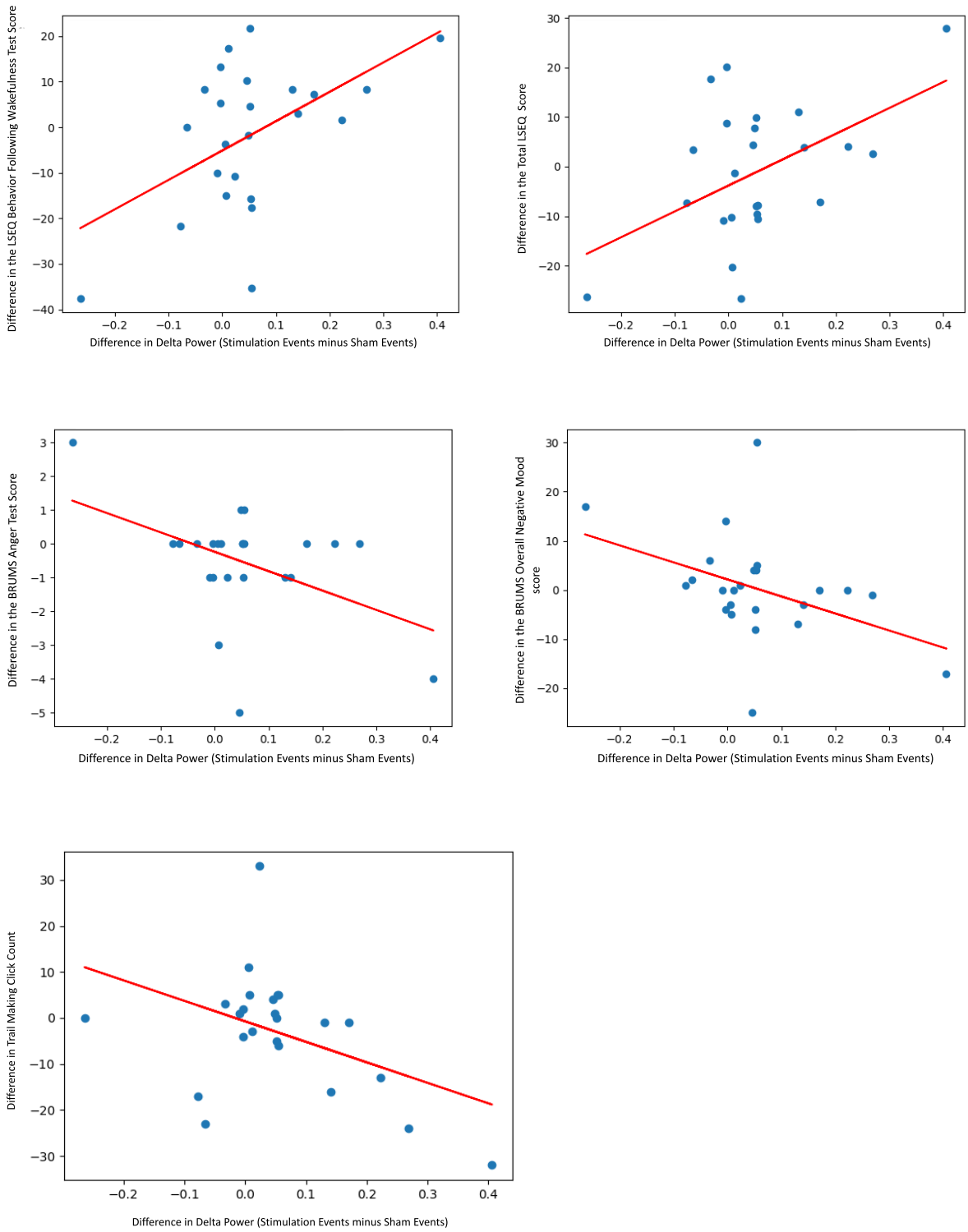


Figure E1: Scatter plots of the significant correlations in Table 1. The difference on the Y axis reflects score after the stimulation night minus score after the no-stimulation night.

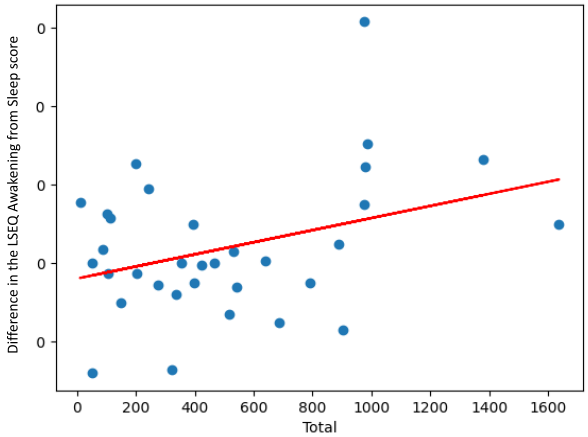


Figure E2: Scatter plot of the significant relationship between stimuli count and difference in awakening after sleep score shown in Table E1.

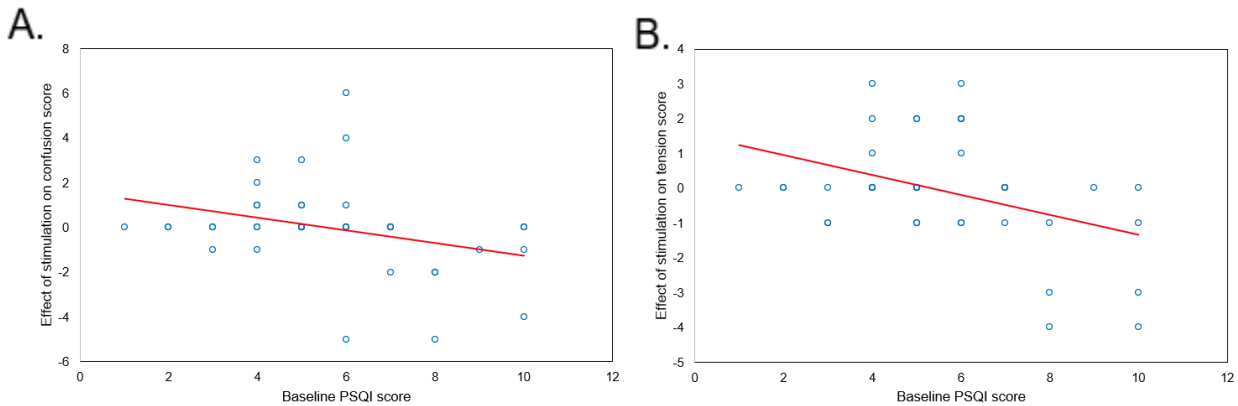


Figure E3: The significant correlations between initial PSQI score and effects of the stimulation on BRUMS confusion and tension axes. The Y axis represents the value on the stimulation night minus the value on the sham night.

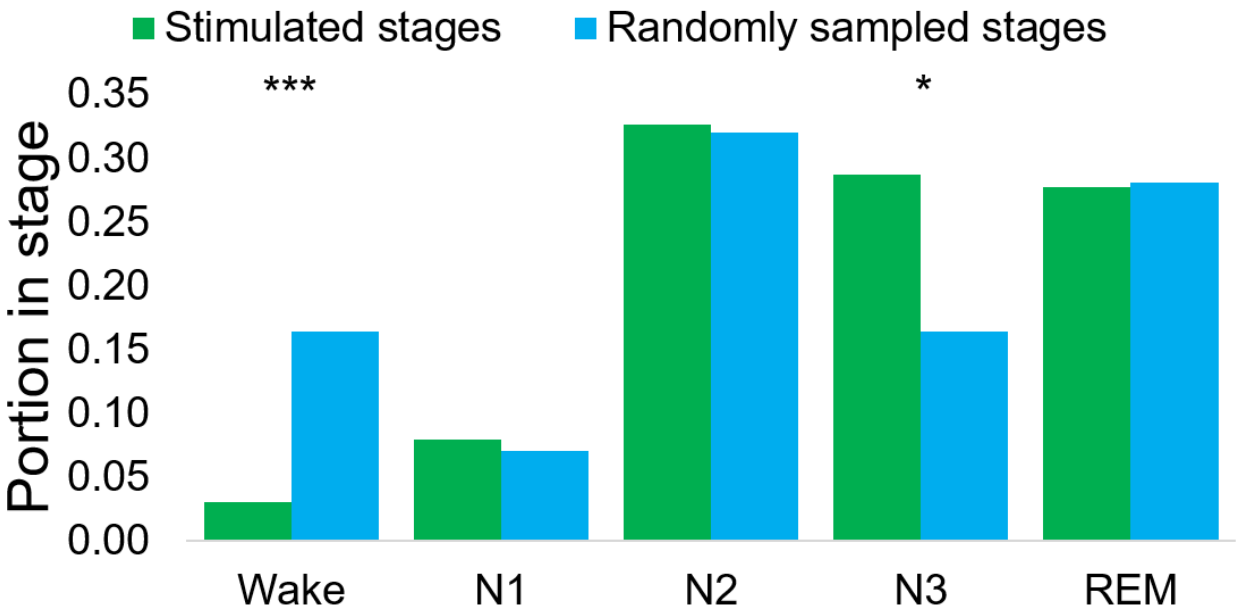


Figure E4: Portion of the stimulation delivered in each sleep stage in our manually staged subset. Asterisks indicate significant differences (chi-square $p < 0.05$) where stages were overrepresented or underrepresented in the stimulated periods compared to the overall distribution of stages. Data for the randomly sampled stages are from 17 nights that had stagable EEG data and data for the "stimulated stages" bars are from 10 nights which had both stagable EEG data and stimulations.

	r	p
LSEQ quality of sleep score difference	0.07	0.71
LSEQ awakening from sleep score difference	0.35	0.04
LSEQ behavior following waking score difference	0.02	0.92
Total LSEQ score difference	0.19	0.28
BRUMS anger score difference	-0.15	0.42
BRUMS confusion score difference	0.08	0.65
BRUMS depression score difference	-0.19	0.30
BRUMS fatigue score difference	-0.08	0.64
BRUMS tension score difference	-0.18	0.32
BRUMS vigor score difference	-0.14	0.44
BRUMS overall negative mood difference	-0.06	0.74
Trail making time difference	-0.11	0.55
Trail making click count difference	-0.06	0.74

Table E1: Correlations between total stimulation count and stimulation effect on the questionnaires (score after stimulation night - score after non-stim night). A greater number of stimuli over the course of the night was associated with a larger benefit for the LSEQ awakening from sleep scale, which reflects the ease of waking up in the morning. Data are for 33 participants who had stimulation and completed questionnaires on both nights.

Supplementary material

Supplementary Table 1: Demographics of subgroups included in analyses

Group	Percent male	Mean age	Median age	SEM age
Participants with delta power data	30.30	43.7	38	4.24
Participants with slow wave data	25.93	39.93	36	4.13
Participants with arousal rate data	33.33	44.15	38	2.97
Participants with correlations between questionnaires and demographics	26.83	38	37	2.71
Participants with correlations between questionnaires and delta power	25.00	36	32	3.99

Supplementary Note 1: Description of the real time sleep intervention system

Sleep staging model

The sleep staging algorithm (summarized in Figure S1) processes data from wearable sensors to classify each second into one of five stages: wake (S0), S1, S2, S3, and REM. It utilizes three types of sensor data: heart rate, accelerometer, and gyroscope readings (each with X, Y, and Z axes). The algorithm's first step is to transform this time-domain sensor data into the frequency domain using Fast Fourier Transforms (FFTs). An FFT is calculated for the heart rate, and also for the accelerometer signal summed across all axes and the gyroscope signal summed across

all axes. These FFTs are calculated over 64-sample windows (approximately one minute of data) at a 1 Hz frequency, resulting in 64 features each for heart rate, accelerometer (sum of X, Y, and Z axes), and gyroscope data (sum of X, Y, and Z axes). In addition, the values of each accelerometer and gyroscope axis in the previous second are also given as input, for a total of 199 features.

This frequency-domain representation serves as input to a neural network classifier. The network consists of four layers: three dense hidden layers with 128, 64, and 32 units respectively, each using ReLU activation functions, followed by a dense output layer with 5 units and a softmax activation. This structure allows the model to learn complex, non-linear relationships between the frequency-domain features and sleep stages.

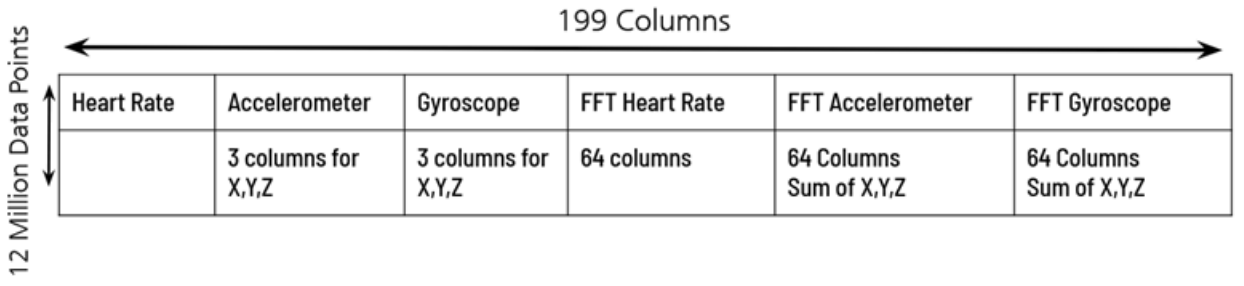


Figure S1: summary of the sleep staging model

We trained the model using a dataset obtained from a previous study[46]. The full dataset consisted of ~12 million datapoints sampled once per second, acquired using the procedures described in[46]. Briefly, gyroscope, accelerometer, and heart rate data were acquired from a Fitbit smartwatch while participants slept in their homes and received targeted memory reactivation controlled by the Fitbit. The Fitbit wearable data were co-registered with sleep stages obtained from a Dreem 2 home polysomnography device[52]. This process thus allowed us to train the model to predict the sleep stage from the motion (gyroscope+accelerometer) and heart rate data.

We evaluated the model by holding out the last 20% of the data from training and using it as a testing set. This allowed us to evaluate the model's predictive abilities on participants never seen before, as well as previously seen participants. Model performance was quantified using confusion matrices representing correct and error rates for each second of sleep, as shown in Figure S2.

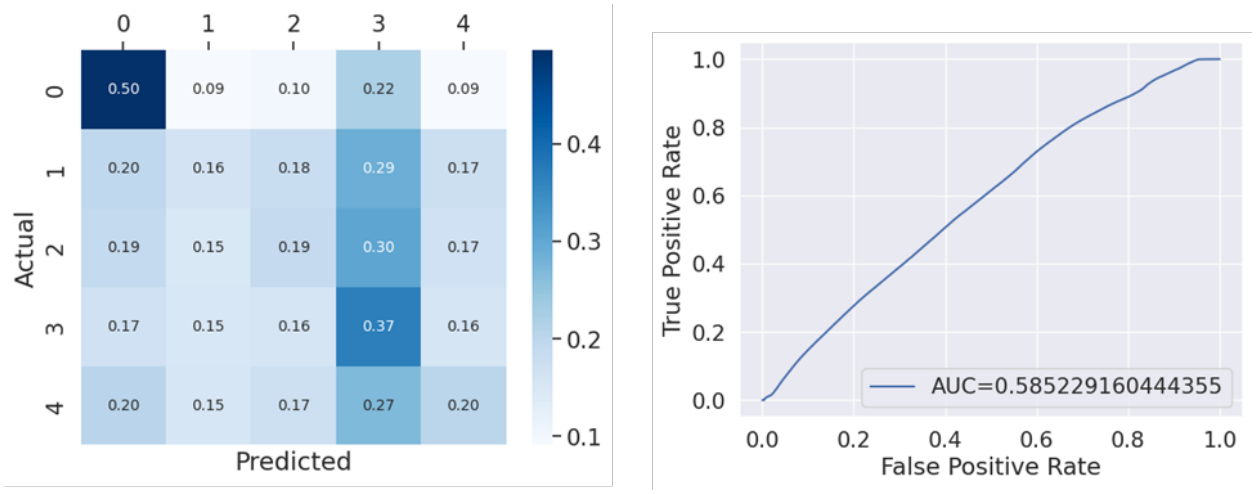


Figure S2: Performance of the sleep staging model in staging each second of sleep. (Left) Confusion matrix comparing the outputs of the model against the true sleep stages determined by the Dreem 2. Stage 0 is wake, stage 4 is REM, and stages 1-3 represent N1-N3. (Right) A ROC curve of the machine learning model in predict stage N3.

Post-processor

While the model output can be used directly (the probability of each sleep stage is saved every second and can be used to collect statistics on sleep), it can also be used to control stimulation in sleep.

With stimulation, the goal is generally to present stimuli in a particular sleep stage. The targeted sleep stage is stage N3 (otherwise known as stage 3 sleep). The post-processor thus allows stimulation when several criteria are met:

1. The probability of stage 3 in the current second is above a specified threshold

2. The average probability of stage 3 over the past X seconds is above a specified threshold (both values can be set by the user)

3. It has been at least X seconds since sleep onset (sleep onset is detected by the Wear OS Health Services platform)

4. It has been at least X seconds since the last detected arousal

5. No more than X stimuli have been presented

6. The user has been in a stable sleep state (defined using the criteria above) for at least X seconds

7. No more than X seconds have elapsed since the start of sleep

Arousals are defined when a large motion is detected on the gyroscope sensors, or the probability of stage 3 in the current seconds drops below its threshold, or the average probability of N3 drops below its threshold.

The post-processor controls the start and stop of stimulation, and also the stimulation intensity. Intensity is set using an automated calibration algorithm; it starts at 0 and increases linearly whenever stimulation is turned on. If an arousal is detected while stimulation is on, the intensity of the stimulation is capped at the value that triggered the arousal, minus a small offset. Stimulation then continues at this intensity until either the end of the sleep epoch (where the cap is reset) or until another arousal occurs (at which point the cap and intensity are decreased further).

Offline system evaluation
We performed an additional evaluation testing whether an entire system consisting of the machine learning model and the post-processor tuned to detect stage 3 sleep would accurately target stimulation to stage 3. This analysis supplements our previous analysis which evaluated the machine learning model's ability to stage each second of sleep by demonstrating in which sleep stages would stimulation occur, taking into account both the performance of the model and the performance of the sleep staging system.

We performed this evaluation on the same validation set used for the testing of the model. Because the post-processor was optimized to target N3, we tested whether the combined model/processor system was more likely to stimulate sleep staged as N3 than expected by chance. This was calculated by comparing the proportion of N3 in sleep with stimulation to the proportion of N3 in all sleep, as shown in Figure S3.

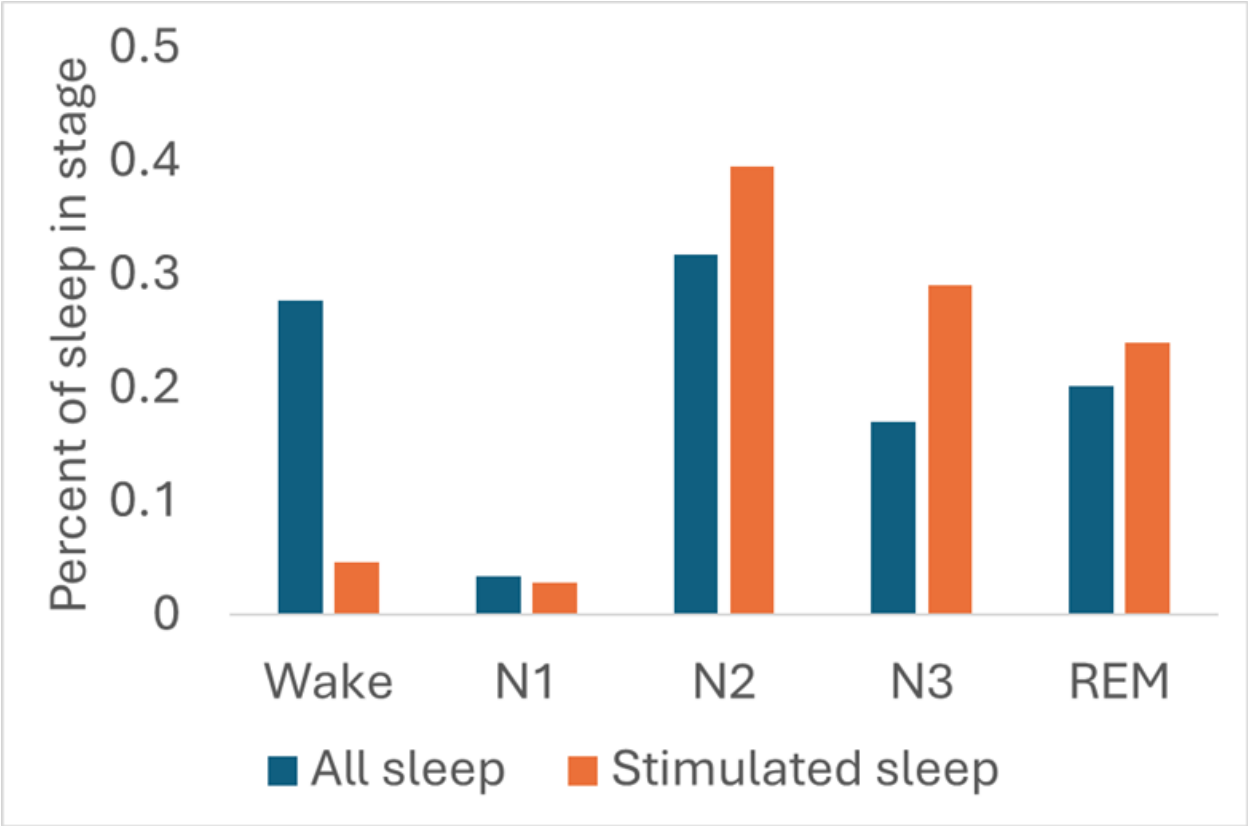


Figure S3: Stimulated sleep selected by the ML model+post processor is enriched for N2, N3, and REM sleep. The largest degree of enhancement was found for N3 sleep, and wake was

underrepresented in the stimulated sleep.