Harnessing Intelligent Audio-Gesture Interfaces For Wearables As A Sleep Aid

by

Daniel Alberto Jacobs Luengo

B.S in Economics, Texas A&M University (2013)

Submitted to the Integrated Design and Management Program

in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE IN ENGINEERING AND MANAGEMENT

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2024

Authored by: Daniel Alberto Jacobs Luengo
Integrated Design and Management Program
May 23, 2024

Certified by: Tony Hu
Program Director and Senior Lecturer, The Riccio Graduate Engineering Leadership Program, Thesis Supervisor

Accepted by: Joan Rubin
Executive Director, System Design and Management Program

Harnessing Intelligent Audio-Gesture Interfaces For Wearables As A Sleep Aid

by

Daniel Jacobs

Submitted to the Integrated Design and Management Program

on May 23, 2024, in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE IN ENGINEERING AND MANAGEMENT

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

ABSTRACT

Insomnia—difficulty in initiating and maintaining sleep—affects a significant portion of the global population. The mainstream adoption of wearable computing presents a unique opportunity to study and aid sleep at an individual level. Here we introduce Zzzonic, a smart sleep-aid application designed for smartwatches that leverages cognitive psychology and human-computer interaction (HCI) to facilitate sleep onset by engaging users in audio tasks as a form of intrusive thought control. A significant aspect of Zzzonic's functionality is its adaptive control system, which estimates sleep onset latency in realtime by monitoring indicators such as motion and user response. The system then progressively modifies the characteristics of the audio tasks to minimize sleep onset latency. This thesis evaluates Zzzonic through a series of user trials conducted throughout the development of the app, accessing the capacity to predict and control sleep onset. The results indicate accurately predicting sleep onset latency in realtime as a control signal is possible but there was no evidence indicating the system could minimize slope onset latency. The inclusion of more indicator signals and machine learning techniques is likely to significantly improve realtime sleep onset latency prediction. Future work on computer-modulated intrusive thought control would benefit from the evaluation of task design, intrusive thought indicators and identifying an adequate control framework.

Thesis supervisor: Tony Hu

Title: Program Director and Senior Lecturer, The Riccio Graduate Engineering Leadership Program

Acknowledgements

Samantha Chan and Nathan Whitmore, thank you for patiently lending me your expertise and equipment in support of early explorations of this project.

My thesis supervisor—Tony Hu—thank you for nurturing my curiosity, for your generous support and most of all for your incredible leadership as director of IDM—our maverick design family at MIT. I will always be deeply grateful for inviting me to participate in one of the richest adventures of my life.

My best friend Sunny—thank you for always being so close, even when you are not.

I especially thank my parents—Norelis and Richard Jacobs—who have both given me unbelievable and unconditional support. And especially my father—who helped me immeasurably on this project through many hours of discussion, analysis and debate—I hope invention continues to be an excuse to spend time together.

And finally my sister and best friend Natalie, who inspired this project—I hope this helps you find better sleep.

Table of Contents

Introduction

For a significant minority living with chronic insomnia[1]——to our best knowledge around 6-10% of the global population[1]——sleep is painfully elusive. People suffering from chronic insomnia feel it as an impairing, detrimental and largely unresolved ailment[2]——a drag on their quality of life.[3] This is consistent with an abundance of studies tying sleep deprivation to a decrease in wellbeing, particularly in the context of mental health, where insomnia is frequently observed alongside depression[4] and anxiety[5]. Some studies have even tied long-term sleep deprivation to a decreased lifespan[6].

Insomnia is a complex condition——it can have multiple causes] and it is addressed with multiple treatments[7]. Currently, our first-line treatment, cognitive behavioral therapy for insomnia (CBT-I) in combination with medications], is only effective in 70-80% [7]of cases and has drawbacks—such as accessibility and adherence. Given the degree by which persistent sleep deprivation can degrade a person's quality of life and the limits of our current treatment options, this thesis is, at one level, a shot at forging a new path for intervention.

Yet, one of the most intriguing aspects about insomnia—at least to me—is just how much of a mystery the condition can be to the people who experience it. For the most part, people with chronic insomnia only have a vague idea of how to manage it despite consulting their physicians[1], much less an understanding of the root of their own inability to sleep.[3] Even amongst physicians, there remains uncertainty[1] and misconception[1] about insomnia.

And this makes sense—given all the potential confounding factors, there is not a definitive consensus about what causes insomnia, what determines whether a treatment will be successful for an individual or much less what the root of an individual's sleeping difficulties is. Take sleep hygiene—a very broad set of healthy sleeping habits—presumably holistic because any one habit is unlikely to be a silver bullet. In at least one sense the current state of treatment is throwing spaghetti at the wall and seeing what sticks. The need for more customized treatments is certainly being recognized. So at a more general level, this thesis proposes a platform to enable targeted insomnia management strategies.

The core idea behind this thesis is loosely inspired by what is likely the earliest and most widely recognized western stereotype regarding the treatment of insomnia: counting sheep[2]——admittedly, most insomniacs are unlikely to find counting to be particularly helpful. *New Scientist* boldly wrote *"Sleep scientists discount sheep"*[3] after reporting on a popular but frequently misinterpreted 2002 study by A. Harvey[8]. Students with symptoms of insomnia were split into groups and asked to report their perceived sleep onset latency[4] after engaging in various types of mental distraction as they fell asleep. The conclusion was that certain

[1] In the context of this thesis, chronic insomnia is defined by the diagnostic criteria set by the American Academy of Sleep Medicine (AASM) in the ICSD-3-TR which requires the following conditions: 1) a person reports difficulty sleeping 2) and it results in some kind of daytime impairment 3) that occurs at least three times per week 4) over the course of three months 5) and it is not circumstantial 6) or related to another medical condition.

[2] Douglas Galbi wrote has delightful blog article describing the earliest known reference to counting sheep as a sleeping technique——a comical 12th century story "*The King and his Story-teller*" in a Latin text called *Disciplina Clericalis* by Petrus Alphonsi, whose telling suggests the origin is even older.

[3]Interestingly, counting sheep is never mentioned in the referenced study.

[4] **Sleep onset latency** (SOL) is the term for the amount of time it takes us to fall asleep.

mental distractions actually appeared to shorten the time someone believed it took them to fall asleep——at least temporarily——while others had no effect.

To be fair the results are not cut and dried but perhaps, like many fables——there may be some grain of truth to “counting sheep”. There is something deeply intuitive about seeking to unwind and relax before falling asleep——to occupy ourselves with something——even when we are already tired. So intuitive perhaps that I would postulate it is an adaptation——a sleep primer perhaps.

A recent study[9] identified higher concentration and diffusion of glutamate——a neurotransmitter——in the lateral prefrontal cortex——a region of the brain that largely manages cognitive control——in people that had engaged in high-demand cognitive work when compared to various controls. They proposed the resources required to restore balance constitute a metabolic cost that makes executing cognitive control more difficult. In other words——demanding more from our brain results would result in less executive power.

From the perspective of insomnia this is a fascinating conjecture[10]; the predominant model[11] of insomnia basically attributes it to a deficient control of anxious thoughts. Perhaps the very quality that someone with insomnia lacks is the quality we seek to induce by unwinding before we sleep[12]. It evokes moments when I have gone to bed directly after a long night of immersive work——only to be plagued by a sort of repetitive-obsessive dream about the work itself——certainly not restful sleep.

In any case, even if it is only a learned behavior, there is something to be said about the sheer ubiquity of our tendency to occupy ourselves with something before bed. A recent poll[13] conducted on behalf of the American Academy of Sleep Science (AASM) reported 29% of respondents read as part of their bedtime routines. Another AASM poll[14] reported that 80% of Americans had lost sleep viewing social media. Losing sleep is not an inability to sleep—but it does show that bedtime activities are not all made equal. My intuition says that somewhere between what engages us and what relaxes us is an opportunity to facilitate sleep.

This brings us into the current state of insomnia treatment——there is strong, highly corroborated data indicating CBT-I is effective[15]. But this should not trick us into feeling that the problem is solved. Far from it——a significant portion——20-30%[7] of people do not respond to treatment and for those who do there is still a significant chance of recurrence[15].

Part of this can be explained by adherence[15]——CBT-I asks a lot from a patient that is already fatigued in exchange for an improvement that may take some time to show. CBT-I is evolving to address this through digital CBT-I (dCBT-I) designed to be more engaging and accessible. But what if the treatment looked more like the very bedtime activities we seem so motivated to seek? Oddly enough——some of the most clinically doubtful sleep aids, such as white noise apps——are also enduringly popular.

Finally there are wearables——which have evolved into powerful computers in their own right and are now mainstream consumer device platforms with the capacity to reach millions of people. Given the extraordinary things deep learning is enabling and the richness of the data wearables are privy to——wearables seem like a grossly underutilized technology.

I believe there is an opportunity where all of these converge to deliver better sleep aid at scale and to exploit that scale to deliver better treatment.

Background

The Restless Mind: Pathogenesis

To understand the philosophy behind Zzzonic—in particular what the potential of a computational approach might be—it is useful to consider insomnia's complexity and what the current treatment options are. As mentioned previously, the prevailing pathophysiological theory of insomnia is one of hyperarousal—that is, an abnormal state of physical and or psychological alertness makes it difficult to fall asleep and stay asleep. Multiple models have been proposed to explain the hyperarousal but in general they fall into two unique but interconnected perspectives.

The most clinically prevalent, arguably, is the cognitive model[11], which primarily attributes insomnia to intrusive thoughts, worry and rumination that induce distress—making it difficult to relax and fall asleep. Eventually the mind becomes preoccupied with the lack of sleep itself, it disproportionately perceives environmental disturbances and eventually internalizes an exaggerated sense of sleep deficiency—reinforcing itself and further exacerbating real sleep deficits.

To construct a sense for this model consider that:

I. People with chronic insomnia most frequently attribute their sleeping difficulty to the emotions and thoughts they have in bed[16] [3]. Corroborating this, a series of studies by A. Harvey[17][18][19] recorded the cognitive state of people with insomnia right before sleep and compared them to people without sleeping difficulties——they found people with insomnia were more likely to be focused on worries, on wakefulness itself, unpleasant imagery and——as a whole—— to perceive this time to be more distressing and less controllable than those without insomnia. The effect of anxious presleep cognition on our ability to fall and stay asleep has been corroborated in many studies[12][20][21][22].

II. As we will review subsequently, cognitive behavioral therapy for insomnia (CBT-I)——which addresses our beliefs and behaviors about sleep——is currently the most clinically effective treatment for insomnia.[7][23]

III. There are poorly understood paradoxes between insomnia that can be objectively measured, insomnia that can not and differences in subjective perception of sleep onset and sleep quality between people with insomnia and people without. [24][25],[26][27]

IV. Finally, a consensus of studies have confirmed insomnia frequently coexists with other mental health disorders[5][28][29]. Together these points of view indicate at least some portion of insomnia is in the domain of the mind.

The physiological model, on the other hand, focuses on how hyperarousal springs from our bodies—this dysfunction then leads to problems initiating and maintaining quality sleep.[30] Some relevant considerations:

I. Studies have determined physiological differences between those with insomnia and those without[31].

II. Additionally there are people with objectively measured insomnia that have no identifiable psychiatric conditions[32]. Together, these findings build a picture of a condition that is also rooted in the body.

This thesis primarily focuses on the cognitive angle of insomnia and for that reason I will not elaborate further on the existing physiological models or pharmacological aids—but in general there is much still to understand about the biology of insomnia, and—at least for now—people with insomnia are generally hesitant to rely on the available sleep medications.[2][33][34] In some sense this is also true about CBT-I.[35]

In any case—the degree to which insomnia is a physical or psychological condition for most individuals is not obvious. One unifying theory is that a physiological predisposition[36], [37] can eventually evolve into or confound with cognitive insomnia. Sleep labs may seem like the answer except they often[5] just confirm what physician and patient already know—poor sleep quality—and sometimes they raise more questions than answers[3].

To add to the confusion there is an open question regarding confounding comorbidity[29], [38]. As mentioned before insomnia frequently coexists with anxiety[39] and depression[40] as well as other psychiatric disorders. Insomnia can be a symptom of these [41] but there is also evidence that it can actually precipitate anxiety and depression [42][43]. Classic chicken and the egg problem—as with before, realistically it is a little of both. A recent poll of physicians revealed a majority identify insomnia as a secondary disorder[44]—essentially a symptom—while clinical sleep experts disagreed—maintaining that treating insomnia as a primary condition resulted in better outcomes for both.

There are also many types of insomnia [45] yet the distinctions are hardly used in a clinical setting [44].

Finally, misconceptions about insomnia treatment are common. There is a well documented and reasonable apprehension to take sleep medication[33]—but there are studies that show that the combination of CBT-I and sleep medication can result in better remissions rates over an extended follow up period[46]. Sleep hygiene—which is often perceived to be ineffective[2][35] has been proven to enhance treatment effects—but only when prescribed an integrated component of CBT-1.[47] White noise machines on the other hand—are very popular, despite mixed evidence, at best, of their effectiveness[48].

So it is of no surprise, at least to me, that one of the most common sentiments amongst those living with chronic insomnia is perplexity around what insomnia is, what causes it and how to manage it[2][3], even amongst those who have sought help[44].

Controlling Insomnia: A control systems perspective

Given the complex dynamics of the insomnia system—the grander spirit of this thesis is the development of a system that can study sleep at scale and transform all the confounding variables and uncertainty of insomnia into targeted insights that not only give someone the relief of seeing the pattern in their insomnia but also reduce the effort of managing it.

A project by Yala et. al. encapsulates this spirit—utilizing a reinforcement learning framework to optimize personalized breast cancer screening policies that improve screening outcomes in a way that would be too complex for a radiologist to do alone [49]. Even in the domain of sleep,

[5] However they remain valuable for differential diagnoses.

Khaokaew et al. have recently explored how to combine large language models (LLMs) and sleep data in a system called *ZzzGPT* to conversationally probe sleep data and provide personalized insights tied to sleep efficiency predictions[50].

An ideal system could extend into direct control of the sleep environment. Haghayegh et al. developed a smart mattress with a novel temperature control to improve sleep[51]—which could help alleviate thermoregulation disorders[52] that disrupt sleep. Modeling the insomnia system to this degree would require a significant number of people to contribute detailed sleep data—and the best way to motivate this might be to facilitate sleep in return.

There is also research exploring control near the sleep system itself. The use of VR biofeedback at bedtime to facilitate sleep has been explored and has shown positive results downregulating physiological arousal. [53] A. Horowitz et al. has developed a system, *Dormio,* designed to exploit the hypnagogic state in service of creativity through computer controller tracking of biosignals and user interaction [54]. Closely related to this project is an investigation to predict sleep onset in real-time for closed-loop sleep control in a mice model [55] as well as an evaluation of reaction time as a useful indicator for an intrusive thought control system. [56]

Returning to the cognitive model of insomnia—if the core issue is so heftily a preoccupied mind as we lay in bed—then presumably anything that controls negative thoughts from inundating the mind—without significantly arousing—it might actually aid sleep. I imagine a system that is compelling, like a puzzle, adjacent to sleep, like listening to music in bed and highly perceptive of the sleep state

- Some studies have found playing *Tetris*, a game, reduced the frequency of intrusive memories, although there were conflicting outcomes in the reduction of the related distress[57]

- In the context of sleep the most widely cited experiments were conducted by A Harvey et al.—one study showed certain tasks seemed to improve subjective[8] measures of sleep onset—their long-term effectiveness was uncertain.

- Another study did identify a reduction in the frequency of intrusive thoughts and objective[1] sleep onset latency.

It is safe to say that all the minimum elements that would be required to fairly assess the potential of a system like this—administration by a third-party, natural sleep environment, accurate real-time measures of sleep onset and intrusive thoughts—have not yet come together.

Treatment: CBT-I

The first-line treatment for insomnia is currently cognitive behavioral therapy for insomnia[23][58](CBT-I)—sometimes in conjunction with sleep medication—in one analysis 45% of patients treated with CBT-I were able to increase their total sleep time (TST) and reduce their sleep onset by 50%.[7] Additionally CBT-I has fewer known side effects than sleep medication and longer lasting results than sleep medications alone [7], presumably a result of addressing a cause for insomnia more directly.
CBT-I is a multi-component treatment that functions by attempting to replace detrimental beliefs, conditioning and behaviors regarding sleep with education, discussion and management strategies, typically over the course of 4-10 weeks.[47]

I. Cognitive Therapy (CT)

The core objective of cognitive therapy is to identify and replace detrimental beliefs about a person's capacity to fall asleep that interfere with inducing and maintaining sleep. This is an important component of the existing treatment protocol that a theoretical, distraction-based intervention alone would not address—for that reason we refer to distraction-based interventions as an aid rather than a treatment. Still—much like the inclusion of medication with CBT-I—it could conceivably improve results in conjunction with CBT-I by facilitating adherence and positive conditioning with the sleep process.

That being said, the emergence of digital cognitive behavioral therapy for insomnia (dCBT-I) is a natural complement that in recent studies has proven to be effective [59][60] although not quite at the level of traditional CBT-I yet. This could be integrated into distraction-based intervention to produce stronger outcomes. Given the current state of large language models, it is reasonable to believe that artificial intelligence might one day allow dCBT-I to approximate the performance of face-to-face CT at a fraction of the cost.[61][60] For example a chat based interaction might be able to track changing beliefs over time—evaluating the effectivity of the treatment on the fly.

II. Sleep Restriction Therapy (SRT)

This component addresses conditioning that drives people with insomnia to seek to lay in bed longer—reinforcing detrimental beliefs about their sleep efficacy. By restricting the amount of time a person with insomnia lays in bed to the amount that induces them to fall and stay asleep quickly—patients can gradually reinstate confidence in their sleep efficacy.

This is probably the most problematic individual component of CBT-I for patients to adhere to since they are already sleep-deprived.[34] This is another element of CBT-I that distraction-based interventions would not address—but this is a natural complement for wearable devices—which can monitor vitals to optimize bedtimes, automatically manage bedtimes and use location data in the home to support adherence.

III. Stimulus Control Therapy (SCT)

SCT addresses conditioning that drives people with insomnia to engage in bedtime activities that extend their sleep onset—such as watching TV—because they can not sleep. This in turn prolongs their sleep onset, reinforces negative beliefs about their sleep efficacy and creates detrimental associations with the sleep environment. In SCT patients abstain from these activities.

This is one dimension of CBT-I that would be an inherent component of distraction-based interventions since engagement with the system would necessarily exclude these other activities—with the added effect of potentially being more inherently motivating given the interactive element.

IV. Sleep Hygiene (SH)

Sleep Hygiene is a broad set of healthy sleeping habits that include things such as exercising, regular bedtimes, conditioning the sleep environment exclusively with sleep, avoiding naps and the consumption of substances that are known to interfere with sleep—like caffeine. No evidence has been found that prescribing these alone improves insomnia [1] perhaps because it is difficult to adhere to [1] or because it is insufficiently targeted [1] (what works for one person might not work for another) but sleep hygiene has been shown to improve results as a component of CBT-I [1].

One element of sleep hygiene is to engage in relaxing and somewhat monotonous activities before bed—this would be an inherent characteristic of a distraction -based intervention. Additionally sleep hygiene is particularly well suited for an intelligent sleep management system that might track and support adherence as well as identify what is most and least effective on an individual basis.

V. Relaxation Techniques

The goal of relaxation techniques is to train a patient to induce physical or mental arousal at will via techniques such as progressive muscle relaxation and biofeedback. This is another element a distraction-based intervention inherently addresses—with the added benefit of being administered by an independent entity, perhaps making it easier to practice since it does not require training or discipline.

Considering the clinical effectiveness of CBT-I there may appear to be little need for an alternative but this is not the case. The most significant problem with CBT-I is that it does not address the needs of a significant group of people who do not respond—approximately 20-30%[15]—and for those who do, insomnia will remerge in a follow up period in about 40-50% of cases[15]. One might imagine that:

- Perhaps some cases of insomnia are simply unrelated to a person's mental relationship to sleep. In this case it would be valuable to be able to identify it—allowing a person to focus on other management strategies.

- It might also simply represent the people who have not successfully adhered to the treatment[15] in this case we should seek ways to make it easier for people to adhere to CBT-I.

Finally there are problems with accessibility—it can be expensive and feel inconvenient.[60] Adherence and accessibility are among the core elements dCBT-I strives to address. Still this is a lot of inertia to overcome for a population already plagued by fatigue and exhaustion.

In this way a theoretical distraction-based intervention would have a strong advantage since it is a passive solution requiring little effort on the users end while filling the natural instinct to occupy the mind when you cannot fall asleep. It would also have the added strength of being independent of a person's beliefs, meaning it can help prevent remissions when a person's state of mind falters—as sleep medication does in conjunction with dCBT-I. Finally it is a non-pharmacological approach for those that might seek it. Just like CBT-I is a holistic set of therapies that are stronger together than alone—it is likely that any positive effect, if any, from a distraction-based intervention would best perform as an interpreted element of CBT-I.

Having mentioned some of the potential benefits it is important to note there also a number of conceivable risks that will need to be evaluated:

I. Little is known about how the character of a distraction affects our thoughts—much less its tendency to affect our sleep onset. Based on our initial testing it seems easy to extend someone's sleep onset by arousing them too frequently. The effect, if any, in the other direction is unclear.

II. Even if it was effective—that benefit might not be perceptible by people who already have distorted beliefs about sleep—failing to address the fundamental issue.

III. It is also conceivable that the experience itself could become an indicator of sleep onset latency—becoming a feature for selective attention to latch onto, exacerbating insomnia. For example, a user might learn to associate an element of the interaction with sleep onset latency—focusing on it as a person with insomnia might detrimentally focus on the clock.

IV. There is also the possibility of run-away effects—a particular distraction could arouse someone—the system could respond by adjusting to their level of alertness which could then further arouse someone until they are fully awake.

V. Its effectiveness could also be short-lived. It is reasonable to think a person might grow fatigued with a simple repetitive task—losing the capacity to engage and distract the user.

I recognize an intervention of this kind is highly speculative—but I also believe forging a new option for insomnia requires exploring uncertain and unconventional paths.

Sound Asleep: The effect of sounds on sleep

Finally—a brief note on sounds and sleep. The use of white noise as a sleep aid has grown steadily popular since it was first introduced in the 1960s[62]. A poll conducted on behalf of the AASM in 2023 indicated that 18% of respondents reported using a white noise machine or app when they slept[63]. Corroborating this trend, as of this writing, four of the top ten apps for iOS in the Health & Fitness category are white noise apps.

This enduring demand for white noise sleep aids presumably means they are perceived to be effective and in any case endorse findings about people's search for non-pharmacological sleep aids. The prevailing explanation for white noise sleep aids work is that they mask environmental noise. Indeed, a study evaluating insomniac's subjective experience found a proportion attributed their lack of sleep to sound disturbances.[2], [48] This is not surprising under the cognitive model of insomnia, which attributes this to disproportionate concern of sleep disturbances via selective attention. Assuming it effectively masks noises then this would in fact ease some of the detrimental selective attention.

Yet, evidence about the clinical effectiveness of white noise as a sleep aid measured objectively is mixed[64]. This is also not particularly surprising since the majority of these studies have not focused on the subjective angle of sleep—the ones that noted subjective improvements in sleep typically reported an improvement.

Sounds may also induce physical relaxation. One randomized control trial observed that EEG showed decreased alpha power (the predominant brainwave while lying in bed with your eyes closed) as well as increased heart rate variability (HRV)—an indicator of decreased stress when listening to Tibetan singing bowls when compared to performing progressive muscle relaxation (PMR) and a control waiting list[65]—they also corroborates this noting subjective perceptions of relaxation.

Zzzonic builds on white noise in order to create a relatively neutral soundscape as well as a means of reaching a population that is already seeking app-based sleep aid but it does not necessarily intend to mask noise or stimulate relaxation. Instead we assumed a good interaction modality would allow a user to close their eyes. That being said, there appears to be a potentially fruitful convergence between white noise sleep aid apps, dCBT-I and computer-controlled sleep that could be exploited to deliver easier and effective sleep aid at scale.

The Stairway to Sleep: Breaking down and tracking the sleep onset period.

Sleep Stages

The sleep cycle—which typically lasts 90 to 110 minutes—is normally characterized by 5 stages: wake, rapid eye movement (REM) and non-REM (N1-N3) and repeats approximately 4 to 5 times throughout the night.[66]

1. Wake—laying in bed, still, with your eyes closed is typically characterized by the domination of alpha waves in the brain, which are associated with relaxation and meditation. It is also associated with full muscle tone and will most likely contain the most motion. At this stage the heart rate will stabilize around a person's resting heart rate. This marks the start of the sleep period Zzzonic is concerned with.

2. N1—is the lightest stage of sleep—it constitutes about 5% of total sleep time and lasts somewhere between 1 and 5 minutes. It is not considered particularly restful. Muscle tone is still present but it begins to decrease along with shallower breathing and a descending heart rate. During this stage a person is still capable of responding to stimuli but their alertness and reaction time might drop. [67] There might be some loss in environmental awareness—notably—it does not constitute what most of us would perceive as sleep. People with insomnia have shown a higher probability of transitioning back from N2 to N1[68] and overall more time in N1 and less time in N3[69].

3. N2—is the beginning of "deep" sleep—it lasts around 25 minutes and then gets progressive longer—constituting about 45% of sleep. Muscle tone is markedly reduced at this stage but not completely absent. A person might still be able to respond to stimuli but their response rates and reaction times will have probably dropped markedly[1]. People in this stage grow but largely disconnected from the environment. Heart rate and body temperature continue to drop steadily. This also constitutes what is considered restful, restorative sleep—making it an important factor in sleep quality. A person will most likely lose consciousness at some point during this stage. This is the end of the portion of sleep that Zzzonic is trying to track, currently.

4. N3—the deepest stage of sleep—constitutes about 25% of total sleep time and lasts about 20 to 40 minutes. It is composed by the predominance of low-frequency delta waves. A person is unlikely to wake up or respond to stimulus. Along with N2 this defines restful sleep.

5. REM—is associated with dreaming and not considered a restful stage of sleep although it is important. It constitutes about 25% of sleep. Our brain activity at this stage is similar to when we are awake. This first occurs after 90 minutes and lasts 10 minutes getting progressive longer.

Polysomnography (PSG)

PSG is the gold standard of sleep staging——producing a holistic and detailed characterization of the sleep cycle through a variety of signals from the body including EEG, eye movement, actigraphy and heart rate. It is also relatively intrusive, removing a person from their natural sleep environment and attaching numerous sensors, both of which can introduce bias. The cost and logistical complexity of PSG is a concern when thinking about highly accurate characterization of sleep at scale.

Actigraphy and Consumer-grade wearables

Consumer-grade wearables, on the other hand are relatively cheap and accessible——if not as precise——and have the potential to enable more natural studies that reach more people for longer periods of time. These wearables primarily characterize sleep states using motion and are sufficiently accurate in estimating metrics such as sleep onset latency and sleep efficiency in most use cases. But this is not always the case—actigraphy has a tendency to underestimate the sleep onset latency during the early stages of sleep for example[70] .

Historically, actigraphy devices reduced the raw acceleration signal into "activity counts" to characterize the activity level for a particular epoch, a result of hardware limitations at the time that made storing and processing the whole signal impractical. As a result, actigraphy-based sleep scoring algorithms have typically been limited to making binary classifications of an epoch (sleep or awake)—so called sleep/wake algorithms—having limited information to make the additional discriminations typically seen in polysomnographic sleep scoring. Another implication of "activity counts" is that most widely cited actigraphy-based sleep scoring algorithms were developed around this feature. Being device-specific, "activity counts"are typically proprietary, with little information out there on their determination, making actigraphy based sleep scoring algorithms rather difficult to use without one of these devices or software. This has been a challenge for this project.

Still, many important assessments about sleep health can be made using these two states alone. As recent as 2020, Apple's sleep-scoring algorithm only determined sleep or awake states[1]. As of May 2024, Apple's actigraphy-based sleep-scoring algorithms breaks-down sleep into "Core Sleep", "Deep Sleep" and "REM" although it is unclear to what degree, if any, this is enabled by the use of other signals such as heart rate. It is worth noting that in this case N1 seems to be clumped somewhere between "Awake" and "Core Sleep", perhaps suggesting the issues in discriminating against such a light-phase of sleep.

Behavioral Response Monitoring

Although characterizing sleep stages via interaction may seem counterintuitive, it actually has a surprisingly early precedent. The first recorded use of auditory stimuli to characterize sleep was in 1862, when Ernst Kohlschütter wrote *Messungen der Festigkeit des Schlafes* distinguishing sleep depth by measuring differences in responsiveness to auditory stimuli.

A study from 1997 sought to evaluate the potential of behavioral response monitoring (BRM) in identifying sleep onset latency and compare its performance to actigraphy by comparing both to PSG. A subject would respond to audio tones or blinking lines as they fell asleep. The study found that BRM was significantly more accurate than actigraphy—which is consistent with its tendency to underestimate SOL. Relevantly, the BMR using audio tones resulted in more responses and arousals prior to and during N1 than using light. The SOLs also tended to overestimate—which is consistent with some of our own results. For the tones stimulus nights sleep latencies shortened over successive SOLS trials although no explanation is provided for this except that people seemed to be sleepier.[71]

More recently a team investigating brain signals that might further characterize N1 used motor responses to auditory stimulus to investigate the relationship between behavioral responses and brain signals. The study found that the number of responses and the reaction time of the responses appeared to decrease in a gradual way across N1. Interestingly they also found that sound didn’t really influence the response rate, this would imply the tasks or sound were not producing an identifiable effect on the person—this has implications for the control system.[67]

Zzzonic

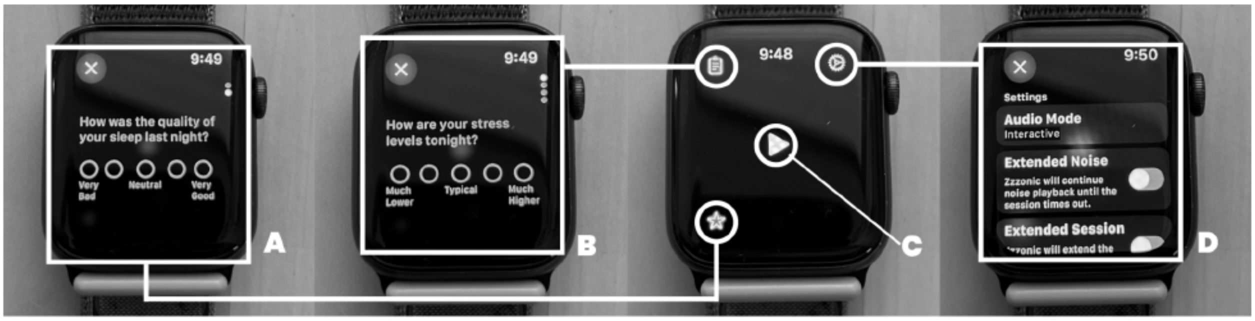


Figure 1 A) Postsleep subjective rating of sleep onset latency and sleep quality. B) Presleep survey for control. C) Start button. D) User settings.

Zzzonic Interface

Zzzonic is a smart, sleep-aid app designed for wearable devices and guided by the following principles:

i. Significant evidence points to anxious pre-sleep cognition as a culprit of insomnia.

ii. There is a mainstream market demand for wearables—these, already armed with a growing number of sensors, have evolved into powerful computers capable of running apps and machine learning inference independently. Advances in machine learning—in particular deep learning—are enabling new levels of prediction, analysis and control in the healthcare setting.

iii. The emerging use of dCBT-I and popularity of white noise apps presents a unique intersection to deploy an integrated digital platform for sleep-aid.

Interacting with Zzzonic goes as follows:

I. A user lays in bed with their smart watch and a nearby wireless audio speaker—once comfortably tucked in—they double tap the start button (fig.1C) to begin the experience. At this point the screen goes dark and the interface goes from being visual-touch to audio-gesture.

II. The system starts by presenting a user with auditory stimuli—two consecutive instrumental notes over a white noise soundscape. These two tones constitute a “task” of variable difficulty—the task is to select whether the second tone is higher or lower pitched than the first—and the user is asked to respond as quickly as possible.

III. The user can respond by gently flicking their wrist counterclockwise or clockwise, indicating a response of “higher" or “lower” respectively. As little as a 17° tilt is required for a system to recognize a response. The user is free to move in bed and interact with the system in whatever sleeping position they find most comfortable.

IV. After a user responds, the system indicates whether the response was correct or incorrect by playing a high-pitched or low-pitched Tibetan singing bowl sound.

From the perspective of the user this constitutes most of the interaction except that they may perceive the frequency of audio tasks to get faster or slower (what we refer to as cadence) and the tones harder or easier to distinguish (what we refer to as difficulty). They will also perceive the overall volume to increase or decrease.

The idea is to create a somewhat monotonous yet compelling focal point that a user can restrain their mind around. Once the user is engaged in the sequence of tasks, the system attempts to smoothly transition them into a state of unconsciousness. It does this by increasing the intensity (the combination of cadence, difficulty and volume) if it estimates the person is too mentally preoccupied—or by decreasing the intensity if it estimates the person is primed to fall deeper into the sleep onset period. Once the system estimates a user is near sleep onset it will shut off automatically—if the user ceases to respond after some time the system will also assume the user is unconscious and shut off on its own.

Zzzonic treats the human-computer interaction as a control system, seeking to minimize predicted sleep onset latency—itself a function of feedback signals from the body as well as the interaction—by adjusting the available control variables—in this case cadence, difficulty and volume. This gives it an added advantage

One way to frame the balance Zzzonic strives to strike is to consider that if a cognitive task is too demanding the mind will be aroused and the task will prevent a person from falling asleep. Alternatively if a task is insufficiently engaging the mind has too much room to wander and it may not restrain cognition enough to be effective. The system must adjust the experience to simultaneously match and drive a user's state of sleepiness at any given point in time—and it depends on estimating how sleepy a user is to do so.

As an interface, Zzzonic is intentionally designed to minimize arousal and closely fit into the natural sleep system. This is why, besides a minimal visual user interface for settings and basic control, the entire experience is a hands-free, closed-eyes interaction that can be comfortably engaged from any position in bed. Replacing the visual modality with audio output eliminates the need for light, which has been shown to disrupt sleep in the context of environmental light and minimizes stimulation of the visual cortex.

Aside from the primary interaction the user has the option of providing relevant information that may impact their sleep onset latency and sleep efficiency before a session (**fig.1A**) as well as a rating of the sleep onset latency and sleep efficiency the next morning (**fig.1B**). Additionally a user can choose a number of settings regarding the frequency distribution of the white noise and duration of white noise (**fig.1D**). For example background noise can persist throughout the night or be programmed to shut-off once the person has fallen asleep.

Zzzonic Predictor

Zzzonic makes periodic estimates of sleep onset latency to use as the necessary control feedback signal. To do this the system relies on actigraphy as well as behavioral response monitoring features derived from the interaction to predict a person's sleepiness.

Zzzonic primarily operates when a user is still conscious enough to respond to stimuli—which ranges from the fully awake state to somewhere within N2, when a person typically falls completely unconscious. This range is bisected by the first and lightest stage of sleep, N1, the typical marker of sleep onset. It encompasses N1 because a person can still respond to stimuli at this point. This is relevant because it implies that traditional sleep staging (awake, N1, N2, N3 and REM) does not provide adequate resolution to distinguish the ~20-30 minutes prior to unconsciousness (a nearly binary set of values—Awake, N1 and N2). For this reason, Zzzonic attempts to directly predict sleep onset latency rather than any discretization of the sleep onset period[1]—providing a continuous measure of the user's sleep state.

By nature of its interactivity, the Zzzonic interface can use the very responses used for intrusive thought control to gain additional information about a user's sleepiness—which is particularly useful in the sleep onset period since the accuracy of actigraphy based sleep staging is typically weakest during this period. These features can serve as proxies for muscle tone or hearing and cognitive acuity, which are known to deteriorate as a person approximates sleep onset[]. These response-dependent indicators can be combined with response-independent indicators readily available in many smart watches, such as actigraphy, breathing rate, heart rate and body temperature. Additionally, Zzzonic makes note of factors such as emotional state, stress, consumption, physical exhaustion and the sleep environment via the presleep survey (**fig.1B**)—which can all play significant roles in the sleep onset outcome on any given night.

The interactive nature of the app also enables an additional function: a lack of response from the user is a robust signal suggesting unconsciousness and a flag for the system to shutdown. This, along with a person's subjective perception of sleep onset (**fig.1B**) can be used in the learning mechanism of the model. Finally the current implementation is an intentionally simple linear model—meaning there is a lot of potentially upside in leveraging machine learning.

Zzzonic Controller

The Zzzonic controller module prescribes the intensity (cadence, difficulty and volume) of the audio tasks performing the intrusive thought control using a periodic estimate of a user's sleep onset latency as a feedback signal. This is currently a simple rational function, designed to make audio tasks exponentially sparse and quiet as a person grows sleepier, giving the mind increasing space to descend into deep sleep.

A more sophisticated controller might actually try to minimize the sleep onset latency function with a broader set of control variables—such as control of the sleep environment itself and targeted insomnia management protocols. Additionally, the difference between a person's objective measures of sleep onset latency and their subjective perception of sleep onset latency, collected via the postsleep rating (**fig.1B**) might be a useful feedback signal for the controller to evaluate insomnia and counteract potential selection attention.

Results

Experiment

Initial Experiment Design

When it comes to research on the role of distraction on sleep onset latency, the most relevant experiments required patients to self-administer distractions and did not include an objective measure of sleep.

Given this, the initial goal for this project was to collect data on the relationship between independently administered intrusive thought control and objective sleep onset latency in a controlled manner. The goal was to control for as many factors as possible while carrying out the experiment in as natural of an environment as possible.

For this reason, our experiment was designed with a within-subject framework (every participant would receive the experimental condition—interactive mode—as well as two control conditions, noise-only and silence) which allowed for control of the physiological and psychological differences between participants. Participants would either be normal sleepers or people self-reporting sleeping difficulties that met the requirements of transient insomnia as defined by the AASM, which would discriminate against the effects of the various conditions on the two groups of people—who might react differently.

Each subject would spend two nights under each condition and these would be assigned randomly in order to control for weekday effects and the first-night effect (FNE)—which is when sleep quality declines in an unfamiliar environment. The conditions would be administered remotely via the same version of the app in order for the participants to sleep at home and to reduce any effect the perception of being observed might have on their sleep—additionally it would ensure controlled administration of the condition.

We would collect actigraphy and interaction data via the device during the interactive portion of the sessions as well as throughout the night in addition to two short surveys. The first would be presented remotely at the beginning of each session via the user interface—this would be used to control for any factors that might impact the results, such as stress level, consumption of caffeine, etc. The second was presented the next morning via the user interface to collect and observe subjective measures of sleep quality and sleep onset latency from the previous night.

The same actigraphy based sleep/wake classifier would be used to get a consistent and objective measure of sleep quality and sleep onset latency based on actigraphy—which would make precision of the algorithm less important. At the end of the study, average sleep onset latency and average sleep efficiency across the groups would be compared for statistical significance.

Testing

II. After some initial testing it became clear that precise functioning of the sleep onset latency (SOL) predictor and controller would be necessary before assessing the potential effect of computer-modulated intrusive thought control on sleep onset latency since an obviously dysfunctional controller would easily bias the results. For this reason we shifted our focus on exploring and evaluating a basic SOL predictor and controller as a proof of concept.

This would require an ensemble of indicators. Our initial intention was to include a number of other signals that are typically available on wearables:

I. Heart rate is a component of polysomnographic (PSG) sleep scoring[1]. As of watchOS 10.4, Apple does not allow real-time access to heart rate measurements for commercial apps except for use during an active workout[1]. With standard HealthKit privileges, the effective sampling frequency for heart rate measurements ranged between 0.1 to 1HZ, a function of private operating system rules. The sparse and irregular sampling introduced the potential for signal aliasing while limiting its utility as a real-time feedback signal.

II. HRV is also a component of PSG scoring[1][1] and has the additional potential of serving as a stress indicator[1]—making it a good candidate for SOL prediction. For the time being HRV could only be sampled approximately once every 4 hours.

III. Respiratory is another component of PSG scoring[1] this, in addition to Apple's sleep own sleep stage scoring analysis, were only available for query once the sleep session had ended, excluding their use as a feedback signal.

Given the current limits and the added complexity of interfacing with the HealthKit, these variables were excluded in the scope of this project but have the potential to be included through more clever signal processing, an exception from Apple or via another device.

From all the potential response-independent variables (heart rate, HRV, respiratory rate, sleep analysis, motion) available on Apple Watch Series 6 and newer, we ultimately chose to focus on motion and interaction derived features at this stage.

This consisted of evaluating three indicators we arrived at through process of elimination during testing:

I. Total motion—the $\log_{10}$ of the sum of all the magnitudes of the acceleration vectors from triaxial acceleration measurements sampled at 50Hz within a 60 second epoch.

II. Response latency—the time interval between the perceptible end of an audio task and the beginning of a user's response—i.e. reaction time.

III. Gesture magnitude—the sum of all the magnitudes of the acceleration vectors from triaxial acceleration measurements sampled at 50Hz within the duration of the gesture.

The indicators would be evaluated by assessing their capacity to change gradually over time in a way that reasonably indicated a relationship to sleepiness—this was necessary since traditional sleep scoring would not provide sufficient resolution, particularly in the initial phases of sleep which is our concern.
We would also test the predictive and controlling functions:

I. SOL predictor—this is composed of two functions. The first is three linear functions (one for each indicator) that were empirically parametrized—these take a moving average of the indicator measurement and output individual predictions of the SOL. The second takes the three individual SOL predictions and outputs an equally weighted composite average. For reasons explained below this would be evaluated by determining how closely it had managed to predict the moment someone stopped responding.

II. SOL controller—a rational function that takes as input the SOL prediction and outputs an exponentially large parameter (cadence) as the SOL prediction becomes smaller, which defines the distribution frequency of the audio tasks. Since we would not be able to compare sleep onset latency unbiasedly across interactive and control nights, this would be evaluated, for now, by its capacity to induce some discernible change in “activity” in the user—i.e. sleepiness.

We had hoped to evaluate the SOL prediction using the device itself (Apple’s sleep analysis) but we eventually discovered a number of problems with this:

I. The most immediate being that the operating system appeared to unilaterally terminate Zzzonic while the device was in Sleep Focus mode, which is required for sleep analysis. In other words, we had difficulty running Zzzonic and Sleep Focus simultaneously. It is possible efficiency improvements might make Zzzonic compatible with Sleep Focus but we were not able to address this within the scope of the thesis.

II. The more general and problematic issue is that actigraphy based sleep-wake algorithms will tend to classify user responses as awake time—resulting in a sleep onset latency that probably lags the first appearance of N1. We confirmed this in early tests using a Dreem Headband which is more accurate at scoring sleep than actigraphy alone. An implementation of Oakley’s algorithm consistently classified the interactive portion of the sessions as awake yet using the Dreem Headband, we were able to observe people fully interacting with an early version of the system in the N1 and N2 sleep stages, which is consistent with a number of studies showing people can respond to verbal auditory stimuli while they sleep, including during N2[72]. Studies evaluating sleep/wake algorithms have identified it is particularly insensitive during the early moments of falling asleep[1]. Future work needing to evaluate behavioral response monitoring systems will benefit from relying on a second more holistic device, such the Dreem Headband for development and evaluation.

III. Additionally the lack of transparency in traditional sleep/wake algorithms make it difficult to apply the same algorithm to data from different devices. Luckily, the activity count algorithm for certain devices has been described well in one study[73] and implemented as a Python package which could be combined with the Oakley algorithm to produce a usable measure of sleep onset latency and sleep efficiency for this experiment. The Oakley algorithm was chosen since it has been shown to have higher sensitivity for people with sleep disorders [74]. Importantly, it’s been shown that using raw accelerometer data from a different device, including the Apple Watch, produces similar accuracy under the same “activity count” algorithms.[75] The implication being we can reasonably expect that inputting the raw acceleration signals from the Apple Watch will produce similar “activity counts'' and sleep-awake classifications.

IV. Initially we assumed the moment a person stopped responding (unconsciousness) could serve to mark sleep onset accurately, as long as the person confirmed they had fallen asleep rather than intentionally ignoring the system once we followed up. In practice we found this is not a particularly accurate indicator of sleep onset either since it also significantly lagged the first appearance of N1, as confirmed by the Dreem Headband. Nevertheless, this became the most adequate option so we adjusted our evaluation accordingly. Still—our grander objective remains to assess the potential of a technology like this as an intervention.

Results

It is worth noting these results are based on data that comes from multiple versions of the app while we were iteratively building, debugging, calibrating and evaluating the prediction and control system. Here we report the average performance of the sleep onset indicators, predictor and controller across this entire period.

In total we captured 60 sessions of which 49 sessions were deemed valid—lasting more than 20 minutes—across multiple versions of the app (v.0.1.1 - v0.1.9) with an average duration of 4 hours and 12 minutes. Of these, 23 were interactive nights, with the interactive portion of the session lasting an average of 22 minutes. 5 sessions occurred on Friday, Saturday and Sunday nights, and 44 sessions occurred in the rest of the week.

These sessions represent 11 people, 7 male and 4 female. 7 users identified as having sleeping difficulties and 4 users identified as having no sleeping difficulties. Although the sample size was small, we checked for a statistically significant difference in sleep efficiency using the control sessions of each group, showing a 20% decrease in sleep efficiency ($t = 0.04$) $P(<0.05)$ in the group with sleeping difficulties.

I. Indicators

We observed two passive variables (independent of a user response) throughout all 49 sessions. Total motion was negatively correlated with the first 30 minutes of a session for the control and insomnia groups, -0.029 (p-value <0.005) and -0.034 (p-value <0.005) respectively. Additionally there was a clear distinction between the insomnia and control groups, with the insomnia group having a baseline total motion of at least 7% higher than the control group. Motion interval, which is a function of time, was positively correlated with the first 30 minutes of a session for the control and insomnia groups.

The interactive variables (dependent on a user response) were observed throughout 23 interactive sessions. Those characterizing the motion of the gesture (magnitude of the gesture, rotation rate of the gesture, rotation of the gesture and duration of the gesture) are functions of one another and were negatively correlated with the progression of the audio tasks sequence. Gesture magnitude was negatively correlated for both the insomnia and control groups with -0.044 (p-value <0.005) and -0.024 (p-value <0.005) respectively.

Neither of the variables characterizing the cognition of the response (response latency and response accuracy) had a significant correlation with the progression of the audio tasks sequence. Response latency for both groups was nearly flat with -0.003 and -0.003 slopes respectively (p-value <0.005). Response latency was also affected by a persistent bug in some devices that was difficult to control for and might have affected the results.
Response accuracy was almost uniform across sessions of varying length for both groups. In a few cases we observed nearly 100% accuracy for the whole session.

II. SOL Predictor

For all interactive sessions, the system shut off by means of a person ignoring the system (involuntarily or voluntarily) rather than by reaching an estimated SOL <= 5 minutes, indicating a consistent overestimation of the SOL.
On average the model overestimated the SOL by approximately 12.5 minutes. The best performing predictions within the normal range for sleep onset (10 - 30 minutes) lagged the moment a user fell unconscious by approximately ~6 minutes, these represented people in both control and insomnia groups.

The same users in both the insomnia and control groups reported that the system had shut off at the adequate moment and that experience had been generally pleasant as well as that the controller had behaved erratically and extended their sleep unpleasantly with different versions of the app.

III. SOL Controller

No identifiable correlation between the control variables and the activity of the users emerged. Users did report feeling like their sleep onset was extended when the cadence was too frequent or the volume too high but made no mention of the difficulty.

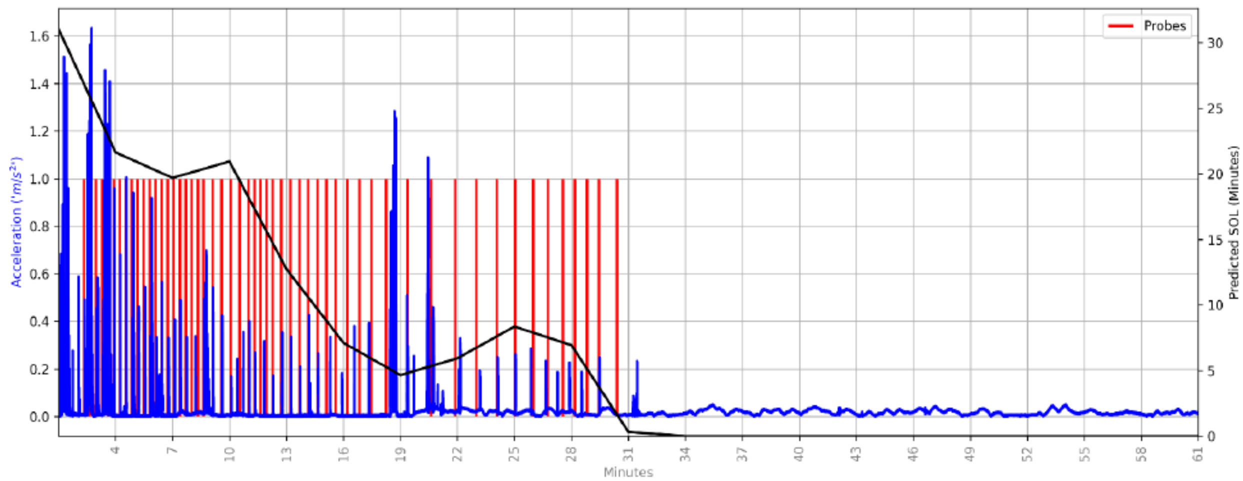


Figure 2 shows a session with an interactive duration of ~30 minutes. The blue line tracks accelerometer readings, the vertical red lines mark the distribution of audio tasks over time, the black line marks the trajectory of the predicted SOL over time.

Discussion

In total we were able to use data from 49 sessions—these were sessions that had not ended as a result of the app crashing or being prematurely terminated by the operating system. Of these 23 included an interactive portion.

The users in the insomnia group all reported having sleeping difficulties that matched the criteria of transient insomnia. The control consisted of 4 people who identified as normal sleepers. The average sleep efficiency for both groups was calculated across the 26 control sessions (silence) using an implementation of the Oakley algorithm resulting in a statistically significant lower sleep efficiency in the insomnia group compared to the control. The point of this was to have more certainty each group actually represented their population.

Ideally, each participant would have participated equally in each condition to ensure bias introduced from individual users was controlled for but this was not possible since persistent bugs with the interactive mode on certain Apple Watch devices frequently terminated sessions. Sessions were distributed as randomly as possible to control for weekday differences.

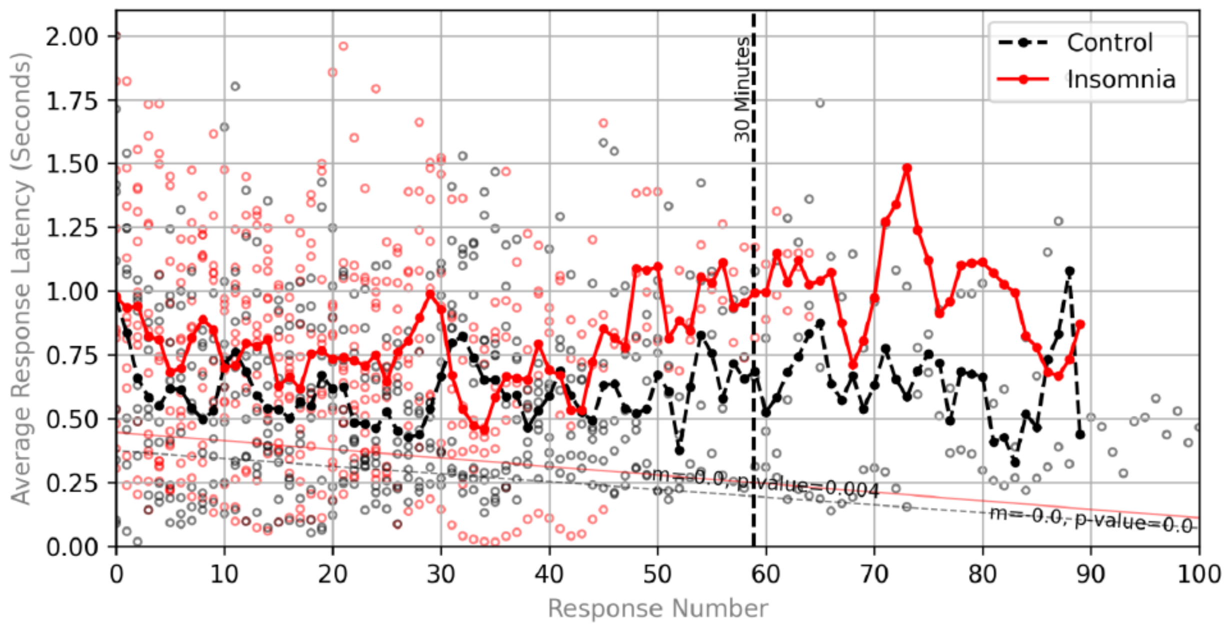


Figure 3 shows the average response latency in seconds (thick lines) at each point in the sequence of audio tasks across all interactive sessions for both the insomnia and control groups. The regression lines (thin lines) for the entire distributions are also plotted. In order to get a sense for the elapsed time, the vertical line (black dashed) marks response number 59 which approximately marked 30 minutes on average. The distribution of points (circles) for each population is also included to get a sense for the variance.

Response Latency

Response latency—a user's reaction time to the audio task—was hypothesized to increase as a user's drowsiness increased as investigated in similar studies.[1][67] We were unable to identify a significant correlation with the progression of the session—even after attempting to normalize it across sessions of varying lengths and calibration settings. This makes it the most ambiguous of the indicators we evaluated since it appeared—in multiple individual sessions—to progress upwards or be a weighty driver of the SOL prediction. There was also no obvious difference in response latencies between the insomnia and control groups—other investigations seemed to be able to infer insomnia from response latencies during the daytime, presumably from being more tired.[76]

A key factor in interpreting the results for response latency is that it was the most recalibrated indicator—that is, the most redefined across different versions of the app. It was also the most affected by bugs, disqualifying at least two sessions—all of which means the results may be noisy. Besides the uncertainty, response latency was critical in triggering the shut-off process automatically which served as the closest estimate of sleep onset. Regarding the noise—this has some extrinsic and intrinsic explanations.

Extrinsically—a key problem might be a poor noise-to-signal ratio—response latency typically ranged between 0.25 and 1.25 seconds—which could easily be buried by noise such as inattentiveness while the user is still tucking into bed, disproportionally slow responses as a user learns to use the system or occasional lagging in the system's gesture detection. Like the rest of the indicators, response latency was particularly noisy at the beginning. Additionally some sessions ended voluntarily—and these tended to be shorter, which would imply that additional noise was reintroduced from a person returning to wakefulness. This is one explanation for why the typical trajectory of response latency is flat in the first half of the session and then increases slightly in the second half of the session.

Response latency has a few characteristics that potentially make it intrinsically variable. Assuming it is a good measurement of attention or focus then it is reasonable to expect that it might vary as quickly and frequently as our thoughts. As an indicator this means we would be particularly interested in how it has tended to change across longer periods of time as opposed to how it has changed from moment to moment. This has a few implications:

i. If response latency is a high frequency signal then its change over time is a low frequency signal which potentially means the indicator is vulnerable to aliasing. Perhaps this is why it appeared to be a clearer signal to the predictor (which used a smoothed average as input) and in the latter half of a session (where a slower cadence of audio tasks represented a lower sampling rate of the signal) both of which effectively work as a low-pass filter. Response latency might benefit from anti-aliasing techniques.

ii. Although less important, its potential behavior as a low frequency signal also brings into question to what degree it might work as a feedback signal. If the required signal duration is too long it may not function for short sessions or may even have limited use in sessions of typical length.

Returning to its intrinsic variability—there is also the consideration that response latency is probably a function of a number of variables that introduce variability:

i. Although we did not identify any correlations between difficulty and response latency, it is reasonable to believe response latency would increase on average when the pitches are harder to distinguish, either as a result of a small difference in pitch or low volume. For our testing, difficulty was programmed to decrease as a user grew sleepier. In theory this might cancel out any increases in response latency coming from sleepiness, explaining its flat trajectory although our observations of response accuracy suggest this is less likely.

ii. Individual differences in a user's hearing and cognitive acuity or interaction style could result in a substantially different range of response latencies.

Finally, it remains possible that response latency is only a useful indicator at some yet unknown point along a user's progression towards unconsciousness—also explaining the tendency to increase in the second half of longer sessions.

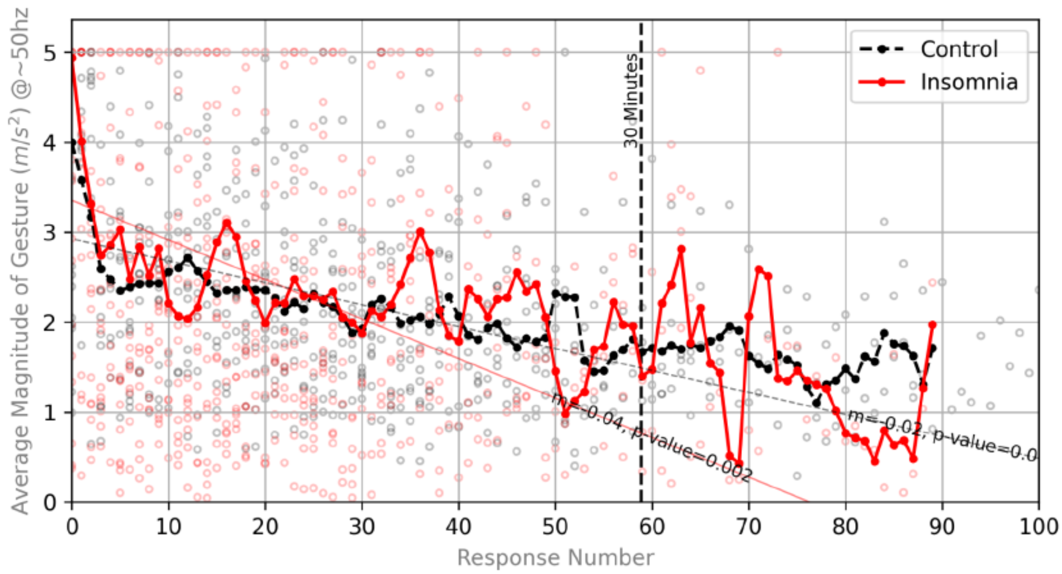


Figure 4 shows the average gesture magnitude in m/s^2 sampled at 50Hz at each point in the sequence of audio tasks across all interactive sessions for both the insomnia and control groups (thick lines). The regression lines for the whole distributions are also plotted [thin lines].

Gesture Magnitude, Rotation Rate, Rotation and Duration

Gesture magnitude—the strength of a user's flick of the wrist—and its correlates were the best performing continuous indicators that could be uniquely derived from the interaction. The hypothesis was that these served as proxies for muscle tone, which has been observed to decrease as a person grows sleepier[77]. We consistently observed that the strength of a user's response as measured by these indicators dropped as the session progressed or increased when a user became more alert—such as before a person voluntarily stopped the session.

Of these, gesture magnitude had the strongest correlation with the progress of the session and was assumed to be the most explanatory—for this reason it was selected as the indicator. The rest were highly correlated to gesture magnitude and assumed to be redundant. A more sophisticated model could conceivably make better use of including them. Additionally there may be cases where the redundancy might be useful. For example, if two users differ widely in gesture magnitude then gesture rotation might be easier to calibrate.

The gesture magnitude typically decreased sharply within the first 5 minutes of a session before stabilizing into a gradual descent. The initial drop probably represents the user calibrating their own strength at the start of each session and should be considered when setting the bounds of the indicator.

Our analysis shows the typical decrease in the strength of the response was nearly twice as fast in the insomnia group than in the control group—the insomnia group also seemed to vary considerably more in the strength of their responses. Given the sample size, this is probably just an artifact of variance in the insomnia group. One explanation for the variance is that a few users in the insomnia group participated in the testing remotely—and did not benefit from training on how to use the system. Thus they might have been unaware they could respond gently. There were also more people in the insomnia group, representing more potential gesture styles.

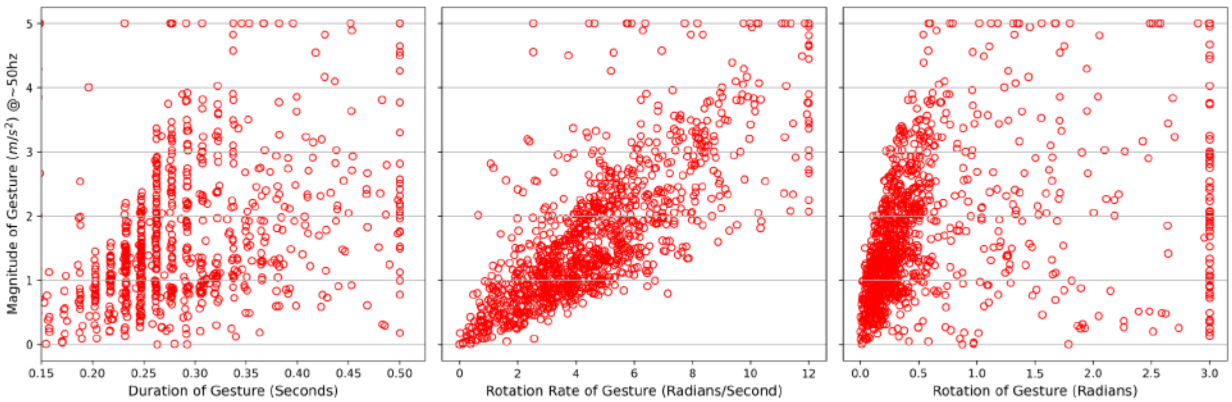


Figure 5 shows correlation plots for each alternative gesture indicator plotted against gesture magnitude for every data point in the dataset.

One matter these indicators bring up is the question of to what degree muscle tone represents physical relaxation versus sleepiness. Although connected[1] this could potentially indicate a person is relaxed but not sleepy.

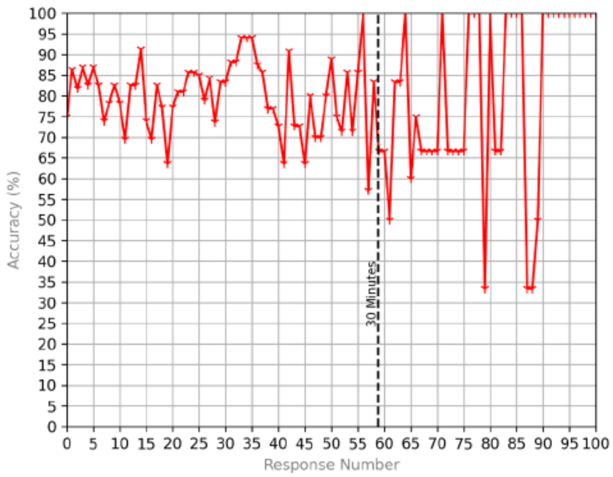
Response Accuracy

Response accuracy—whether a user answered the audio task correctly—was hypothesized to decrease as a user grew tired, a proxy for the decrease in cognitive and hearing acuity that has been observed when people grow sleepy in other settings[1][1]. The tendency for the response latency SOL to decrease does suggest some measurable degradation of cognitive acuity might exist. Despite this, our testing found the current audio task design does not allow the use of response accuracy as a feedback signal since there was no clear relationship with the progress of the session.

We observed nearly perfect accuracy in sessions lasting more than 30 minutes and no general tendency to decrease over time—despite programming the difficulty to increase if a user was not sleepy. This indicates, at the very least, that the audio tasks did not offer a way to measure the drops in cognitive or hearing acuity, if any, that might be useful to infer a user is sleepy. Beyond the possibility that no measurable relationship exists, there are a few potential explanations for the lack of variability:

I. It is possible the difficulty of the current tasks needs better calibration, currently the minimum difference in pitch between tones is 1/12th of an octave, which might be too easy to discriminate. Tests with an early prototype that allowed for smaller pitch differences showed somewhat more variability in response accuracy. A separate implication is that it can be difficult to determine if changes in response accuracy are due to sleepiness or task difficulty—easier tasks will naturally have higher accuracies than more difficult ones.

II. A binary outcome (correct or incorrect) makes response accuracy a less sensitive and more volatile feedback signal since there are no degrees of correctness. In theory a moving average, as implemented, would improve the sensitivity but it does not appear to be sufficient, especially as audio tasks become more sparse. Additionally, in any case there was a fifty-fifty chance of getting it correct.



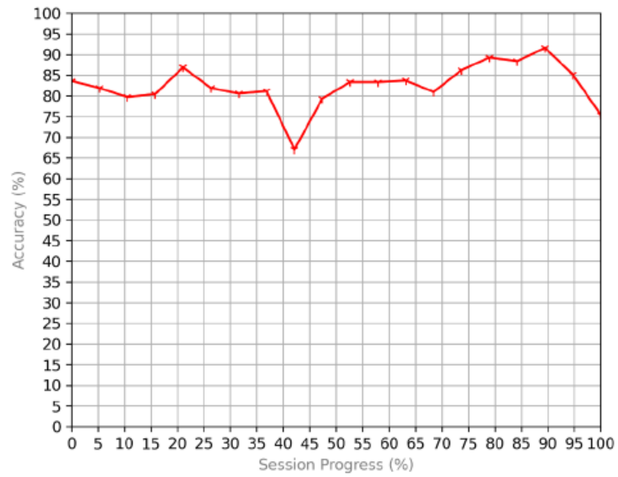


Figure 6 [Left] shows the accuracy as a percentage of correct guesses across all sessions and groups for each audio task in the sequence. **Figure 7** [Right] shows the accuracy as a percentage of correct guesses across normalized progress of the session.

Although the variance of the average accuracy appeared to increase with time—this effect almost entirely disappeared once the lengths of the sessions were normalized. This is not surprising since there were fewer long sessions. Users reported answering incorrectly when they forgot in what direction to flick their wrist or while they were still busy tucking into bed. This might represent a missed opportunity—it is worth considering how the task can be reframed to gather more information. For example, one study[1] asked participants to identify deviant sounds, which reduces the chances of random guesses.

Total Motion, Motion Interval

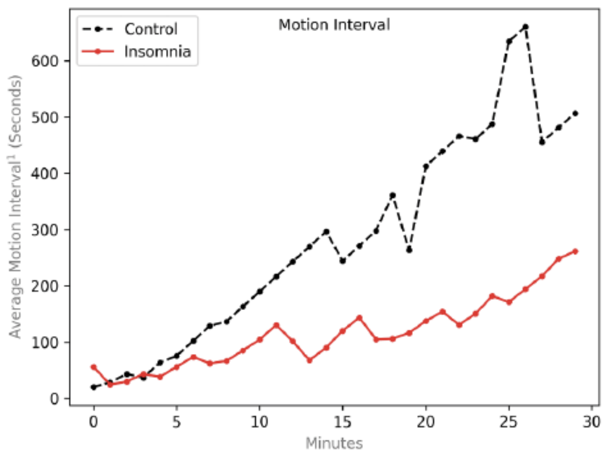
Total motion—a measure of a user's total movement during a period of time—is a form of standard actigraphy—and has already been thoroughly investigated[1][1]. Still, we were interested in understanding its suitability as a feedback signal. Our testing showed our implementation consistently decreased as the session progressed in a relatively gradual and continuous manner—presumably tracking a person's tendency to move less frequently and more gently (including breathing) as they grow tired.

The primary issue with motion generally as a feedback signal is its potential range which varies immensely between moments of relative stillness and strong movement—making it a volatile indicator. It also contains multiple signals at various scales—for example, the frequency of large movements might reveal comfort levels while small movements might reveal breathing patterns. For this reason we tracked total motion and motion interval—the time interval between large movements—separately. Additionally we used the $\log_{10}$ of total motion to attenuate large movements, since sessions were primarily composed of moments of relative stillness.

Our tests revealed motion interval was not suitable for a few reasons:

I. It had a tendency to erroneously interpret the passing of time as sleepiness and to change abruptly—making it an insensitive and volatile indicator.
II. It was also a sparse measurement, providing no insights between large events.

III. Including both total motion and motion interval is partially redundant, since total motion also considers the frequency of movements, particularly larger ones. This has the effect of doubling the influence of large movements.



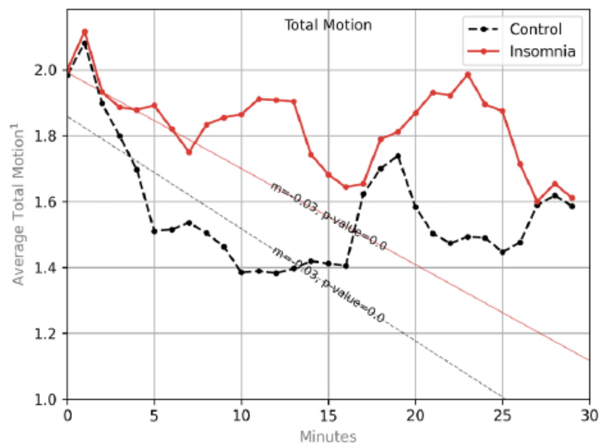
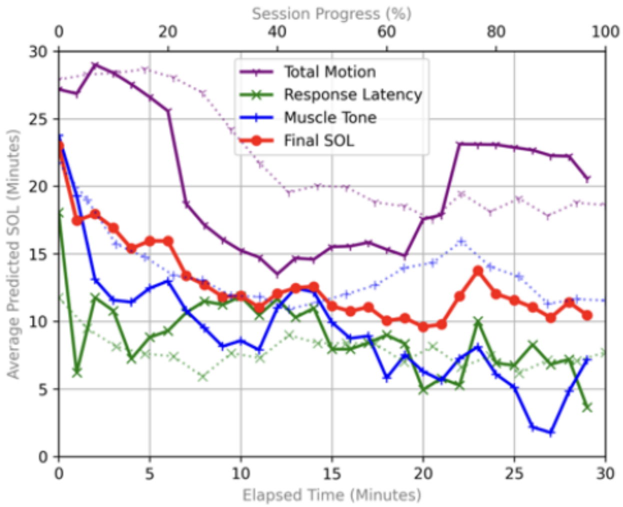


Figure 8 [Left] shows the average motion interval for all sessions for each 60-second epoch, separated by group.

Figure 9 [Right] shows the average total motion for all sessions for each 60-second epoch, separated by group as well as the regression lines for each group.

For these reasons we ultimately decided to only include total motion which already provides continuous information and can be fed into a number of filters to focus on different signals. Since our model largely ignores large movements, future work could investigate how to make better use of this information.

One issue we ran into with both total motion and motion interval was calibration. Unlike the other indicators there was significant variability in the bounds of the indicator between users, particularly between the insomnia and control groups. For example—the lower bound of total motion was significantly higher for the insomnia group, who tended to move more. Similarly, the lower bound for motion interval was significantly lower for the insomnia group, who tended to move more frequently.



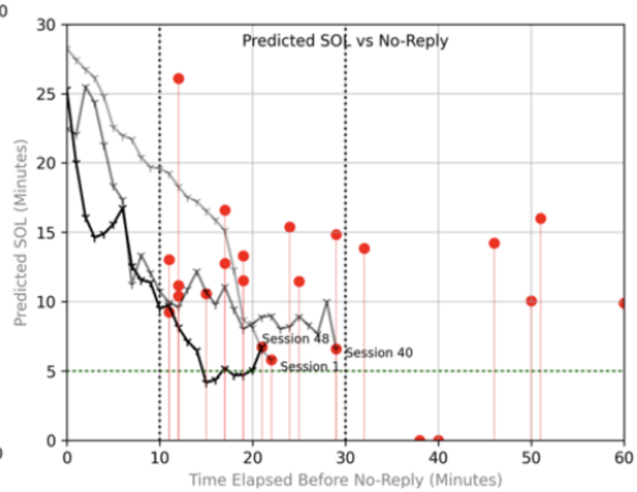


Figure 10 [Left] shows the trajectory of the average composite SOL across time as well as the SOL predictions for its component parts [solid lines]. It also shows the average trajectory of the composite SOL and its component parts after normalizing for the session length.

Figure 11 [Right] plots the predicted SOL against the real SOL (the moment someone stopped responding) [red dots] as well as the trajectory of the predicted SOL over time for the top 3 best performing sessions [black solid lines]. The dotted lines mark the target range predicted SOL <= 5 minutes AND 10 minutes < Real SOL < 30 minutes.

Sleep Onset Latency Predictor

The sleep onset latency (SOL) predictor consistently overestimated the SOL by approximately 12 minutes. **Fig.10** shows the average SOL prediction broken down by component after normalizing for the session length which plateaus at an estimated 12 minute SOL after about 12 minutes of elapsed time. **Fig.11** shows 14 of interactive sessions (60%) ended with a SOL prediction between 10 and 15 minutes.

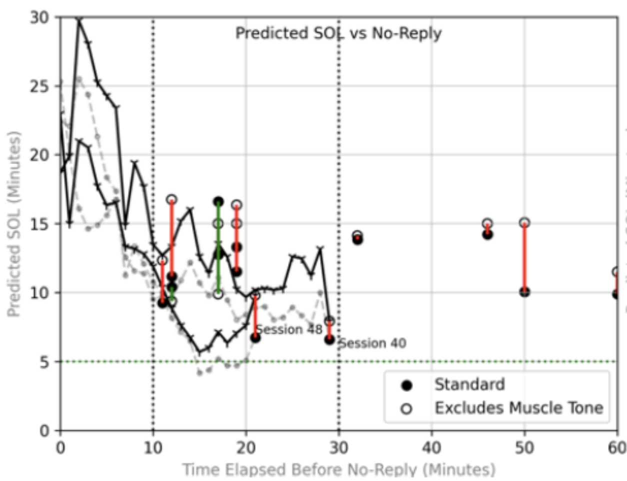
I. Indicator Calibration and Weighting

A persistent issue throughout our testing was the calibration of the indicators—as previously discussed there were significant differences in the upper and lower bounds of each indicator across groups, individuals and even sessions. This had the effect of saturating one or more of the indicators and preventing the SOL from ever descending since it was an equally weighted average of the three components.

Ideally indicators would be calibrated for each user. The most performant SOL predictions were manually calibrated to the user—even then variability remained between nights. We attempted to define these ranges automatically by observing some initial portion of each session but it was difficult to characterize a user's "sleepy" behavior with this information alone. It seems likely that a good calibration will depend on at least a few nights of data.

It is also reasonable to believe that indicators will vary in importance throughout the progress of a session. Dynamically adjusting the weights to emphasize their explanatory power over time might improve performance. It is worth nothing that the information we get from a user's response carries more "weight" when they are sleepy (i.e. it has a longer effect on the experience)—this is true for all response derived metrics and is perhaps a desirable feature assuming the SOL predictions are accurate.

II. Indicator Performance



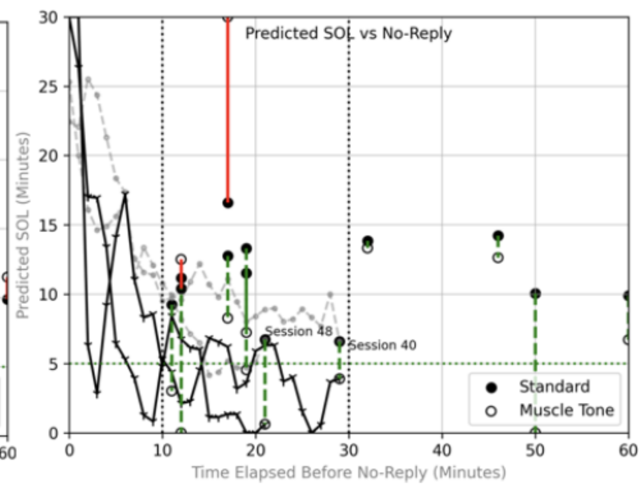


Figure 12 [Left] is a modified version of Figure 11 showing the simulated performance of the model when excluding the gesture magnitude (muscle tone) indicator. The green and red lines represent decreasing and increasing errors respectively. The black lines show the new trajectories of the composite SOL for the top 2 best performing sessions.

Figure 13 is a modified version of Figure 11 showing the simulated performance of the model using the gesture magnitude (muscle tone) indicator alone.

As mentioned earlier, the gesture magnitude indicator had a tendency to drop sharply at the beginning of a session followed by an oscillating descent between 0 and 15 minutes which has been observed in other studies tracking gestural responses[71]. The first drop probably represents the initial adjustment period while a user calibrates how much strength a response requires. This is consistent with the observation that gesture magnitude was the least likely to saturate the SOL prediction—suggesting the upper bound was well above the indication range.

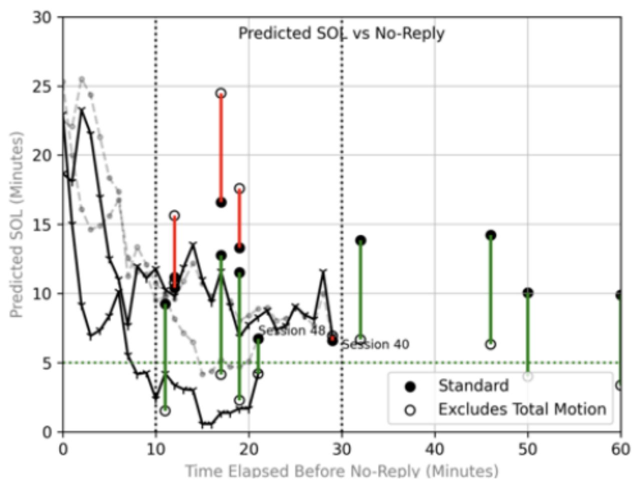
Fig.4 shows this most likely constitutes the initial drop between 4.5 m/s^2 and 2.5 m/s^2 still leaving an approximate 1 m/s^2 of range to use as indication. Calibrating to this smaller range does mean it may become more vulnerable to variability—future use of gesture magnitude should consider this adjustment effect.

The observations in figure x suggest the oscillating descent is more likely to represent real changes in muscle tone since it continues to descend rather than plateau. Gesture magnitude appears to be the “best” indicator when used alone (**fig.13**), improving the SOL prediction in almost every session of this set, some of which are brought into the target range. But this should not necessarily be attributed to its power as an indicator alone.

I. For example, the worst performing simulated session in (**fig.13**) is a result of a calibration setting that saturates the SOL prediction—one might imagine that this might happen more frequently if the upper bound is decreased to better focus on its sleep-related range.

II. Second, the simulated results are most likely underestimations of the real SOL—in a few sessions this results in an SOL prediction that reaches the target range within 10 minutes—which is probably too early. It remains possible that gesture magnitude is a front loaded indicator, becoming less important after 10 minutes. This would be another explanation for its tendency to flatten out (**fig.10**).

III. It would be valuable to investigate what the explanatory decay of each indicator is. Alternatively, excluding gesture magnitude produced relatively little damage to the simulated SOL predictions, implying it behaved relatively independently.



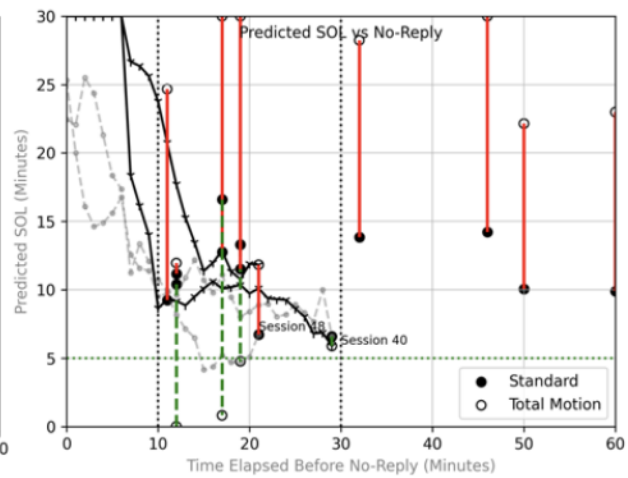


Figure 14 [Left] is a modified version of Figure 11 showing the simulated performance of the model when excluding the total motion indicator. The green and red lines represent decreasing and increasing errors respectively. The black lines show the new trajectories of the composite SOL for the top 2 best performing sessions.

Figure 15 is a modified version of Figure 11 showing the simulated performance of the model using the total motion indicator alone.

Total motion is the most likely culprit for the stagnation of the SOL predictions during testing. **Fig.14** shows the simulated results of using total motion alone, which significantly worsens the predicted SOL for a majority of sessions. The reason for this is upper bound saturation of the predicted SOL, which can be seen by the proximity of the resulting SOL predictions to the top of the range as well as the new trajectories of the best session SOL prediction—which only appear after ~7.5 minutes.

For these sessions, a significant portion of time was spent with one third of the SOL indicating full wakefulness—this probably prevented the composite SOL from descending below 12 minutes. Naturally, excluding total motion improved all the sessions that were affected by this.

Our focus on small-scale motion carries the implication that significant time must pass before total motion measurements enter the indicator range—thus it might benefit from dynamic weighting. Once the indicator was in range it had a tendency to drop sharply—this was probably a moving average effect before stabilizing into very gradual descents with little variability which are characteristics of a good feedback signal. This brings into question to what degree this represents breathing rate and if so, whether total motion is the best way to track it.

Finally, we observed that sufficiently large or frequent movements could bump the indicator beyond its upper bound, abruptly re-saturating the predicted SOL—implying the $\log_{10}$

transformation helped but didn't fully resolve the scale and volatility issues. This is consistent with the observation that the total motion component of the predicted SOL tended to rise in the second half of the session (**fig.10**).

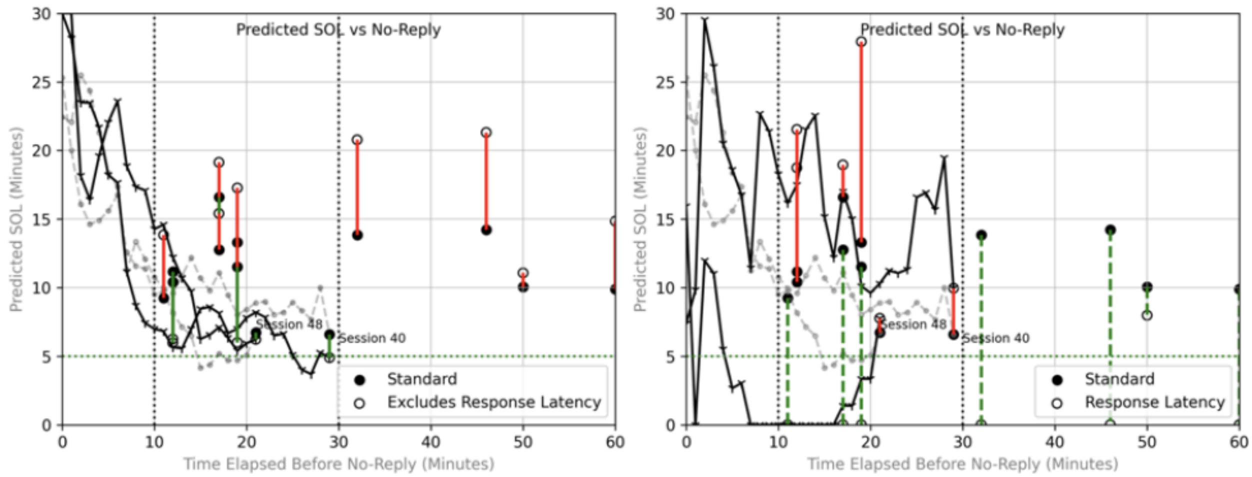


Figure 16 [Left] is a modified version of Figure 11 showing the simulated performance of the model when excluding the response latency indicator. The green and red lines represent decreasing and increasing errors respectively. The black lines show the new trajectories of the composite SOL for the top 2 best performing sessions. **Figure 17** is a modified version of Figure 11 showing the simulated performance of the model using the response latency indicator alone.

Given the uncertainty around response latency, we initially suspected it might have been behind the underperformance of the SOL predictor. As mentioned earlier, response latency was the most difficult indicator to calibrate. This contradicted its performance as an input to the SOL predictor which demonstrated a tendency to decrease throughout a session and data from other studies that have used it successfully[67] —this has a few explanations:

Response latency was the most likely indicator to "bottom-out" the SOL prediction—meaning we had trouble setting the lower bound. It spent considerable time indicating a user was asleep in a number of sessions—this is apparent from the number of sessions that "improve" to the minimum range in **fig.17**. This is a likely explanation for why its component of the SOL quickly plateaued around 7.5 minutes on average (**fig.10**). This is also consistent with other studies using reaction times, which had similar upper bounds but different lower bounds. [67]

There were sessions with more suitable calibrations that did work though—descending over the long run—but it remained highly variable as a feedback signal. If it can be properly calibrated, response latency would benefit from heavier smoothing as well as considering it a low-frequency signal.

Excluding it largely worsens the performance of the model because these happen to be the same points that were affected by total motion, attenuating the error.

Overall this demonstrates a strong need for flexible and accurate calibration of each of the indicators in the product context. Dynamic weighting is also likely to improve the performance—future work should investigate the explanatory decay of each indicator.

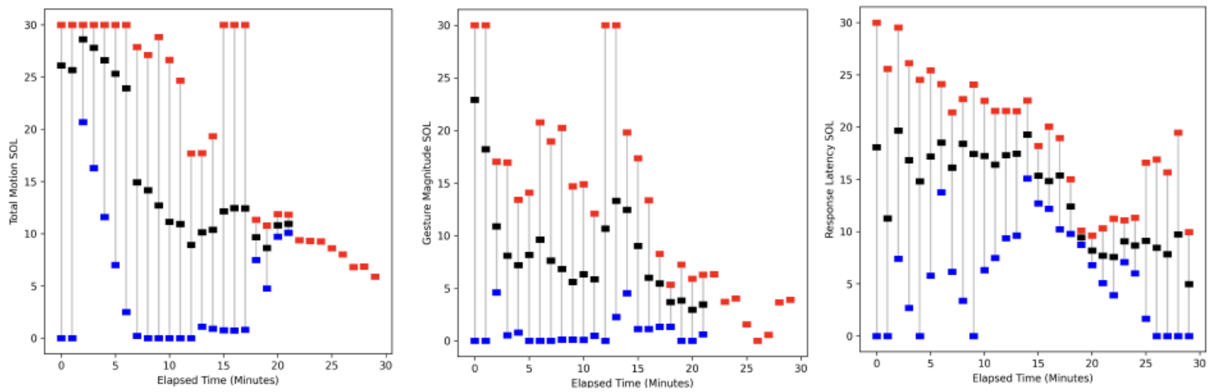


Figure 18, 19, and 20 (Left, Center, Right) shows the variability of each indicator over time after removing the most severely miscalibrated sessions. Plotted are the max (red), mean(black) and min(blue) values for each minute across all sessions.

This analysis also demonstrates the most reliable performance of an SOL predictor will probably come from an ensemble of indicators that balance their strengths and weaknesses as a whole. It is interesting to note that the best performing sessions, such as session 40 and 48 changed minimally when one of the indicators was excluded.

Given the performance of such a simple model, there is a lot of potential upside in integrating more signals and employing machine learning. These results suggest real-time sleep onset latency prediction on a wearable device is probably possible.

Moment of Unconsciousness as Sleep Onset Indicator

Finally, as mentioned in the experiment section, we observed that sleep onset, as measured by the Dreem Headband, typically preceded the moment someone fell completely unconscious. This has a number of curious implications—one is that this is probably another marker the predictor has overestimated the SOL.

Ideally the system would shut off as a result of identifying the real moment of sleep onset, necessarily this precedes the moment a person stops responding involuntarily. Yet in our tests nearly every interactive session ended by means of unconsciousness. And despite this, we found that users reported the system shut off at the correct moment and that the experience was pleasant when they perceived the cadence of auditory tasks was adequate, implying the SOL prediction error was smaller. In contrast, the same user would report the system detrimentally extended their sleep onset when they perceived the cadence of the auditory tasks was too fast, implying this error was larger. In other words the moment of unconsciousness could happen with a nearly accurate or grossly inaccurate SOL estimate, which reduces its utility as an objective measure of SOL.

Another implication is—what then, if not unconsciousness, defines the right moment to shut off? Defining the right moment will ultimately be essential for the effectiveness of any predictor. For example the first instance of N1 (i.e. typical sleep onset) might be too soon since the mind may still have a chance to ruminate.

At the other end we have already observed that unconsciousness (i.e. some point between N2 and N3, inclusive) can be too late. Ideally there exists a point of no return, a threshold before unconsciousness beyond which it is too late for intrusive thoughts to dominate the natural tide of sleep. Presumably it exists since people with insomnia ultimately do fall asleep, albeit with more difficulty.

This point would also, ideally, be able to be identified via indicators. The interactive element allows for some interesting possibilities. For example, during our testing of an early version of the system along with the Dreem Headband, we observed that a frequent, fixed cadence seemed to indefinitely maintain a user between wakefulness and N1, never beyond N2. Adjusting the cadence allowed a user to enter N2.

Aside from suggesting that the right cadence is critical to avoid disrupting sleep, it also suggests there is a cadence that marks a person's ability to fall asleep. Perhaps crossing a similar interaction based threshold can define the right moment to shut off.

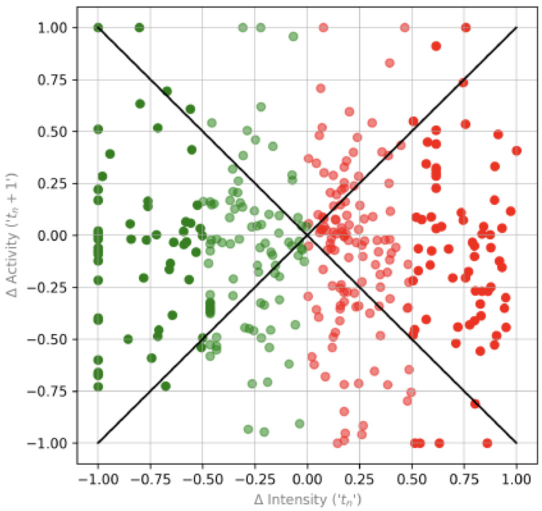


Figure 21 is a correlation plot showing a current change in intensity prescribed by the controller between time t_n and t_{n-1} plotted against the future change in activity between t_{n+1} and t_n.

Sleep Onset Latency Controller

Superficially, the implementation of the Oakley sleep/wake algorithm indicates that the interactive sessions extended the SOL for users in both the control and insomnia groups by 5 minutes and 7.5 minutes respectively.

That being said, this is a poor assessment of the real SOL since the algorithm—which is based on motion—tended to designate the entire interactive block of gesture responses as awake. As mentioned in before, we know this is not necessarily the case from our experiments with the Dreem Headband and corroborating studies. Additionally there were many factors we were not able to control for that might have extended the SOL during interactive sessions.

Looking at the data we were unable to identify patterns demonstrating the SOL controller was influencing the users' SOL in any way. Our hypothesis was that at a minimum, a discernible increase in alertness would emerge as the controller hiked the interactive intensity—this was the worst case scenario and a reasonable assumption given user reports.

Fig.21 is a correlation plot showing the current change in intensity prescribed by the controller between time t_n and t_{n-1} plotted against the future change in activity between t_{n+1} and t_n. An ideal controller would produce an X pattern, demonstrating a consistent capacity to induce an increase or decrease in the user's future activity by increasing or decreasing the interactive intensity.

Our expectation was to see an accumulation of correlated points in the upper right quadrant. Instead points were randomly spread—this was true for a number of relationships across time (e.g. t_n - t_{n-1} vs. t_{n+5} - t_n) and after attempting to control for the heterogeneous distribution of audio tasks over time as well as dysfunctions of the controller and predictor.

Although future intensity is a function of a user's current activity (the controller uses activity as a feedback signal) the assumption was a user's future activity should be relatively independent of current intensity. Still, the least weak correlations were between the mean changes of distant points in time (e.g. t_n and t_{n+5}) which probably represents the underlying tendency for both intensity and activity to descend over time.

Overall, the controller's performance was not surprising given the number of factors it potentially depends on, the most important being an accurate SOL prediction and a well calibrated control protocol. Additionally, the activity metric as currently defined brings with it a fair amount of noise—which further obfuscates the effects, if any, of the controller.

Future work should consider a less arbitrary control protocol as well additional proxies for intrusive thoughts. It would also be valuable to study the engagement characteristics of different tasks and evaluate them through more precise and informative equipment, such as EEG.

Users in both groups sometimes reported that the experience had been pleasant and that it had not disturbed them or that it had been erratic and that it extended their sleep onset latency . A reduced sleep onset was not mentioned.

I. The most pertinent control factor seemed to be cadence—the distribution of audio tasks over time—which users attributed the most interference too. For example—on a few occasions users reported that too many audio tasks in a row had extended their sleep onset—basically preventing them from growing unconscious. Conversely, too few audio tasks also seemed to induce a sort of arousal—a few users reported preemptively anticipating the next task if the cadence was too slow—noting a faster cadence would have been pleasant. Although subjective, the positive and negative association does imply this was the most performant control factor.

II. Volume was noticed by users but typically in a negative sense—either the volume was too high, too low or it lasted too long. This implies thoughtful control of volume is important since it will primarily disrupt rather than engage. Otherwise the soundscape was described as pleasant and relaxing.

III. The varying difficulty of the tasks was never mentioned, probably being the most inconsequential. Our testing shows response accuracy was often uniformly correct—implying difficulty—as currently implemented—was not a strong control factor since it was hardly perceived. Response accuracy has been used successfully in other studies though[67]. More complex auditory tasks might have a more noticeable effect on the user but it is unclear whether this would be detrimental.

A number of unusual system behaviors reported by users helped us identify bugs throughout the development process. These included pitch invariance (i.e. exhaustive repetition of the same audio task), extreme variations in the cadence (nearly constant or very sparse audio tasks), sudden changes in volume and sound glitches, sudden termination of the app, gesture recognition failure and indefinitely long sessions. Most of these were acutely noted by users—either partially or fully awakening them. Although this does not detract from the potential efficacy of the system, it does suggest there is little room for error—the system can easily become disruptive.

Design Considerations

Finally, our testing revealed some important insights from a design perspective. Multiple users noted that the battery of the Apple Watch was dead or significantly drained the following morning—and produced a sort of "range anxiety" before a session. The draining of the battery was worse when the extended audio playback feature was set. A natural design progression would be a wearable-smartphone hybrid, offloading the analysis and audio playback to the smartphone while the connected wearable serves as sensor. This might also resolve the issues with the OS based app terminations.

Additionally, in order to run Zzzonic in the background, the operating system required a persistent Bluetooth connection. This meant users without a bluetooth speaker or those who did not wish to disrupt a partner they were sharing the bed with were forced to use headphones. This was described as uncomfortable and probably affects a person's capacity to relax.

References

[1] M. M. Ohayon, “Epidemiology of insomnia: what we know and what we still need to learn,” *Sleep Med. Rev.*, vol. 6, no. 2, pp. 97–111, Apr. 2002, doi: 10.1053/smrv.2002.0186.

[2] K.-P. Yung, K.-F. Chung, F. Y.-Y. Ho, W.-F. Yeung, and T. H.-Y. Ng, “The Experience of Chronic Insomnia in Chinese Adults: A Study Using Focus Groups and Insomnia Experience Diaries,” *Behav. Sleep. Med.*, vol. 14, no. 4, pp. 406–428, 2016, doi: 10.1080/15402002.2015.1017097.

[3] L. Rezaie, H. Khazaie, and F. Yazdani, “Exploration of the experience of living with chronic insomnia: A qualitative study,” *Sleep Sci.*, vol. 9, no. 3, pp. 179–185, Jul. 2016, doi: 10.1016/j.slsci.2016.07.001.

[4] S. Suh, H. Kim, H.-C. Yang, E. R. Cho, S. K. Lee, and C. Shin, “Longitudinal Course of Depression Scores with and without Insomnia in Non-Depressed Individuals: A 6-Year Follow-Up Longitudinal Study in a Korean Cohort,” *Sleep*, vol. 36, no. 3, pp. 369–376, Mar. 2013, doi: 10.5665/sleep.2452.

[5] D. J. Taylor, K. L. Lichstein, H. H. Durrence, B. W. Reidel, and A. J. Bush, “Epidemiology of Insomnia, Depression, and Anxiety,” *Sleep*, vol. 28, no. 11, pp. 1457–1464, Nov. 2005, doi: 10.1093/sleep/28.11.1457.

[6] A. N. Vgontzas *et al.*, “Insomnia with Short Sleep Duration and Mortality: The Penn State Cohort,” *Sleep*, vol. 33, no. 9, pp. 1159–1164, Sep. 2010.

[7] M. D. Mitchell, P. Gehrman, M. Perlis, and C. A. Umscheid, “Comparative effectiveness of cognitive behavioral therapy for insomnia: a systematic review,” *BMC Fam. Pract.*, vol. 13, p. 40, May 2012, doi: 10.1186/1471-2296-13-40.

[8] A. G. Harvey and S. Payne, “The management of unwanted pre-sleep thoughts in insomnia: Distraction with imagery versus general distraction,” *Behav. Res. Ther.*, vol. 40, no. 3, pp. 267–277, 2002, doi: 10.1016/S0005-7967(01)00012-2.

[9] A. Wiehler, F. Branzoli, I. Adanyeguh, F. Mochel, and M. Pessiglione, “A neuro-metabolic account of why daylong cognitive work alters the control of economic decisions,” *Curr. Biol. CB*, vol. 32, no. 16, pp. 3564-3575.e5, Aug. 2022, doi: 10.1016/j.cub.2022.07.010.

[10] M. O. Harrington, J. E. Ashton, S. Sankarasubramanian, M. C. Anderson, and S. A. Cairney, “Losing Control: Sleep Deprivation Impairs the Suppression of Unwanted Thoughts,” *Clin. Psychol. Sci.*, vol. 9, no. 1, pp. 97–113, Jan. 2021, doi: 10.1177/2167702620951511.

[11] A. G. Harvey, “A cognitive model of insomnia,” *Behav. Res. Ther.*, vol. 40, no. 8, pp. 869–893, Aug. 2002, doi: 10.1016/s0005-7967(01)00061-4.

[12] D. A. Kalmbach, D. J. Buysse, P. Cheng, T. Roth, A. Yang, and C. L. Drake, “Nocturnal cognitive arousal is associated with objective sleep disturbance and indicators of physiologic hyperarousal in good sleepers and individuals with insomnia disorder.,” *Sleep Med.*, vol. 71, pp. 151–160, Jul. 2020, doi: 10.1016/j.sleep.2019.11.1184.

[13] “AASM Bedtime Routines Survey, Bedtime Routines,” American Academy of Sleep Medicine, 2023. [Online]. Available: https://aasm.org/wp-content/uploads/2024/01/sleep-prioritization-survey-2023-bedtime-routines.pdf

[14] “AASM Sleep Prioritization Survey, Social Media Impact on Sleep,” American Academy of Sleep Medicine, 2022. [Online]. Available: https://aasm.org/wp-content/uploads/2022/09/sleep-prioritization-survey-social-media.pdf

[15] A. Muench *et al.*, “We know CBT-I works, now what?,” *Fac. Rev.*, vol. 11, p. 4, Feb. 2022, doi: 10.12703/r/11-4.

[16] K. L. Lichstein and T. L. Rosenthal, “Insomniacs’ perceptions of cognitive versus somatic determinants of sleep disturbance,” *J. Abnorm. Psychol.*, vol. 89, no. 1, pp. 105–107, 1980, doi: 10.1037/0021-843X.89.1.105.

[17] J. Nelson and A. G. Harvey, “An exploration of pre-sleep cognitive activity in insomnia: imagery and verbal thought,” *Br. J. Clin. Psychol.*, vol. 42, no. Pt 3, pp. 271–288, Sep. 2003, doi: 10.1348/014466503321903384.
[18] A. G. Harvey, “Pre-sleep cognitive activity: A comparison of sleep-onset insomniacs and good sleepers,” *Br. J. Clin. Psychol.*, vol. 39, no. 3, pp. 275–286, 2000, doi: 10.1348/014466500163284.
[19] J. Nelson and A. G. Harvey, “Pre-sleep imagery under the microscope: a comparison of patients with insomnia and good sleepers,” *Behav. Res. Ther.*, vol. 41, no. 3, pp. 273–284, Mar. 2003, doi: 10.1016/s0005-7967(02)00010-4.
[20] A. Ballesio, V. Ghezzi, M. Vacca, C. Ottaviani, and C. Lombardo, “Effects of Presleep Cognitive Intrusions on Subjective Sleep and Next-Day Cognitive Performance in Insomnia,” *Behav. Ther.*, vol. 51, no. 5, pp. 688–699, Sep. 2020, doi: 10.1016/j.beth.2019.09.003.
[21] A. Narmandakh, A. J. Oldehinkel, M. Masselink, P. de Jonge, and A. M. Roest, “Affect, worry, and sleep: Between- and within-subject associations in a diary study,” *J. Affect. Disord. Rep.*, vol. 4, p. 100134, Apr. 2021, doi: 10.1016/j.jadr.2021.100134.
[22] J. Wuyts *et al.*, “The influence of pre-sleep cognitive arousal on sleep onset processes,” *Int. J. Psychophysiol.*, vol. 83, no. 1, pp. 8–15, Jan. 2012, doi: 10.1016/j.ijpsycho.2011.09.016.
[23] J. M. Trauer, M. Y. Qian, J. S. Doyle, S. M. W. Rajaratnam, and D. Cunnington, “Cognitive Behavioral Therapy for Chronic Insomnia,” *Ann. Intern. Med.*, vol. 163, no. 3, pp. 191–204, Aug. 2015, doi: 10.7326/M14-2841.
[24] A. Castelnovo *et al.*, “The paradox of paradoxical insomnia: A theoretical review towards a unifying evidence-based definition,” *Sleep Med. Rev.*, vol. 44, pp. 70–82, Apr. 2019, doi: 10.1016/j.smrv.2018.12.007.
[25] L. Rezaie, A. D. Fobian, W. V. McCall, and H. Khazaie, “Paradoxical insomnia and subjective–objective sleep discrepancy: A review,” *Sleep Med. Rev.*, vol. 40, pp. 196–202, Aug. 2018, doi: 10.1016/j.smrv.2018.01.002.
[26] J. D. Edinger *et al.*, “Insomnia and the eye of the beholder: Are there clinical markers of objective sleep disturbances among adults with and without insomnia complaints?,” *J. Consult. Clin. Psychol.*, vol. 68, no. 4, pp. 586–593, 2000, doi: 10.1037/0022-006X.68.4.586.
[27] M. T. Bianchi, K. L. Williams, S. Mckinney, and J. M. Ellenbogen, “The subjective–objective mismatch in sleep perception among those with insomnia and sleep apnea,” *J. Sleep Res.*, vol. 22, no. 5, pp. 557–568, 2013, doi: 10.1111/jsr.12046.
[28] J. Fernandez-Mendoza and A. N. Vgontzas, “Insomnia and its Impact on Physical and Mental Health,” *Curr. Psychiatry Rep.*, vol. 15, no. 12, p. 418, Nov. 2013, doi: 10.1007/s11920-013-0418-8.
[29] K. A. Khurshid, “Comorbid Insomnia and Psychiatric Disorders,” *Innov. Clin. Neurosci.*, vol. 15, no. 3–4, pp. 28–32, Apr. 2018.
[30] G. S. Richardson, “Physiological Models of Chronic Insomnia,” *J. Clin. Sleep Med.*, vol. 01, no. 04, Oct. 2005, doi: 10.5664/jcsm.26389.
[31] G. S. Richardson, “Human physiological models of insomnia,” *Sleep Med.*, vol. 8, pp. S9–S14, Dec. 2007, doi: 10.1016/S1389-9457(08)70003-0.
[32] A. N. Vgontzas, J. Fernandez-Mendoza, D. Liao, and E. O. Bixler, “Insomnia with objective short sleep duration: The most biologically severe phenotype of the disorder,” *Sleep Med. Rev.*, vol. 17, no. 4, pp. 241–254, Aug. 2013, doi: 10.1016/j.smrv.2012.09.005.
[33] J. M. Y. Cheung, D. J. Bartlett, C. L. Armour, J. G. Ellis, and B. Saini, “People with insomnia: experiences with sedative hypnotics and risk perception,” *Health Expect.*, vol. 19, no. 4, pp. 935–947, 2016, doi: 10.1111/hex.12388.
[34] N. Vincent and C. Lionberg, “Treatment Preference and Patient Satisfaction in Chronic Insomnia,” *Sleep*, vol. 24, no. 4, pp. 411–417, Jun. 2001, doi: 10.1093/sleep/24.4.411.
[35] J. M. Y. Cheung, D. J. Bartlett, C. L. Armour, and B. Saini, “Treating Insomnia: A Review of Patient Perceptions Toward Treatment,” *Behav. Sleep. Med.*, vol. 14, no. 3, pp. 235–266, May 2016, doi: 10.1080/15402002.2014.981818.
[36] C.-J. Harvey, P. Gehrman, and C. A. Espie, “Who is predisposed to insomnia: A review of

familial aggregation, stress-reactivity, personality and coping style," *Sleep Med. Rev.*, vol. 18, no. 3, pp. 237–247, Jun. 2014, doi: 10.1016/j.smrv.2013.11.004.
[37] J. G. Ellis *et al.*, "The natural history of insomnia: predisposing, precipitating, coping, and perpetuating factors over the early developmental course of insomnia," *Sleep*, vol. 44, no. 9, p. zsab095, Sep. 2021, doi: 10.1093/sleep/zsab095.
[38] R. M. Merrill and I. S. Gibbons, "Comorbidity of Sleep Disorders, Mental Illness, and Migraine or Headaches," *SN Compr. Clin. Med.*, vol. 5, no. 1, p. 283, Dec. 2023, doi: 10.1007/s42399-023-01622-w.
[39] L. Staner, "Sleep and anxiety disorders," *Dialogues Clin. Neurosci.*, vol. 5, no. 3, pp. 249–258, Sep. 2003.
[40] D. Riemann, L. B. Krone, K. Wulff, and C. Nissen, "Sleep, insomnia, and depression," *Neuropsychopharmacology*, vol. 45, no. 1, pp. 74–89, Jan. 2020, doi: 10.1038/s41386-019-0411-y.
[41] D. Nutt, S. Wilson, and L. Paterson, "Sleep disorders as core symptoms of depression," *Dialogues Clin. Neurosci.*, vol. 10, no. 3, pp. 329–336, Sep. 2008.
[42] D. Riemann and U. Voderholzer, "Primary insomnia: a risk factor to develop depression?," *J. Affect. Disord.*, vol. 76, no. 1, pp. 255–259, Sep. 2003, doi: 10.1016/S0165-0327(02)00072-1.
[43] N. Breslau, T. Roth, L. Rosenthal, and P. Andreski, "Sleep disturbance and psychiatric disorders: A longitudinal epidemiological study of young Adults," *Biol. Psychiatry*, vol. 39, no. 6, pp. 411–418, Mar. 1996, doi: 10.1016/0006-3223(95)00188-3.
[44] R. M. Benca, S. M. Bertisch, A. Ahuja, R. Mandelbaum, and A. D. Krystal, "Wake Up America: National Survey of Patients' and Physicians' Views and Attitudes on Insomnia Care," *J. Clin. Med.*, vol. 12, no. 7, p. 2498, Mar. 2023, doi: 10.3390/jcm12072498.
[45] M. L. Perlis and P. Gehrman, "Types of Insomnia," in *Encyclopedia of Sleep*, Elsevier, 2013, pp. 199–202. doi: 10.1016/B978-0-12-378610-4.00175-3.
[46] C. M. Morin *et al.*, "Cognitive-Behavior Therapy, Singly and Combined with Medication, for Persistent Insomnia: Acute and Maintenance Therapeutic Effects," *JAMA J. Am. Med. Assoc.*, vol. 301, no. 19, pp. 2005–2015, May 2009, doi: 10.1001/jama.2009.682.
[47] J. Williams, A. Roth, K. Vatthauer, and C. S. McCrae, "Cognitive Behavioral Treatment of Insomnia," *Chest*, vol. 143, no. 2, pp. 554–565, Feb. 2013, doi: 10.1378/chest.12-0731.
[48] S. M. Riedy, M. G. Smith, S. Rocha, and M. Basner, "Noise as a sleep aid: A systematic review," *Sleep Med. Rev.*, vol. 55, p. 101385, Feb. 2021, doi: 10.1016/j.smrv.2020.101385.
[49] A. Yala *et al.*, "Optimizing risk-based breast cancer screening policies with reinforcement learning," *Nat. Med.*, vol. 28, no. 1, pp. 136–143, Jan. 2022, doi: 10.1038/s41591-021-01599-w.
[50] Y. Khaokaew *et al.*, "ZzzGPT: An Interactive GPT Approach to Enhance Sleep Quality." arXiv, May 06, 2024. Accessed: May 22, 2024. [Online]. Available: http://arxiv.org/abs/2310.16242
[51] S. Haghayegh *et al.*, "Novel temperature-controlled sleep system to improve sleep: a proof-of-concept study," *J. Sleep Res.*, vol. 31, no. 6, p. e13662, 2022, doi: 10.1111/jsr.13662.
[52] L. C. Lack, M. Gradisar, E. J. W. Van Someren, H. R. Wright, and K. Lushington, "The relationship between insomnia and body temperatures," *Sleep Med. Rev.*, vol. 12, no. 4, pp. 307–317, Aug. 2008, doi: 10.1016/j.smrv.2008.02.003.
[53] M. de Zambotti *et al.*, "A virtual reality-based mind–body approach to downregulate psychophysiological arousal in adolescent insomnia," *Digit. Health*, vol. 8, p. 20552076221107887, Jan. 2022, doi: 10.1177/20552076221107887.
[54] A. Haar Horowitz, T. J. Cunningham, P. Maes, and R. Stickgold, "Dormio: A targeted dream incubation device," *Conscious. Cogn.*, vol. 83, p. 102938, Aug. 2020, doi: 10.1016/j.concog.2020.102938.
[55] D. Huffman, A. Ajwad, F. Yaghouby, B. F. O'Hara, and S. Sunderam, "A real-time sleep scoring framework for closed-loop sleep manipulation in mice," *J. Sleep Res.*, vol. 30, no. 4, p. e13262, 2021, doi: 10.1111/jsr.13262.
[56] C. Gao, W. Wei, G. Shi, H. Hu, R. Jiang, and Z. Zhou, "Extracting interaction effect of cognitive task and unwanted cognitive activity in pre-sleep period," presented at the International Academic Workshop on Social Science (IAW-SC-13), Atlantis Press, Oct. 2013,

pp. 637–642. doi: 10.2991/iaw-sc.2013.141.
[57] L. Iyadurai *et al.*, “Preventing intrusive memories after trauma via a brief intervention involving Tetris computer game play in the emergency department: a proof-of-concept randomized controlled trial,” *Mol. Psychiatry*, vol. 23, no. 3, pp. 674–682, Mar. 2018, doi: 10.1038/mp.2017.23.
[58] J. Walker, A. Muench, M. L. Perlis, and I. Vargas, “Cognitive Behavioral Therapy for Insomnia (CBT-I): A Primer,” *Klin. Spetsialnaia Psikhologiia Clin. Psychol. Spec. Educ.*, vol. 11, no. 2, p. 123, 2022, doi: 10.17759/cpse.2022110208.
[59] Ø. Vedaa *et al.*, “Effects of digital cognitive behavioural therapy for insomnia on insomnia severity: a large-scale randomised controlled trial,” *Lancet Digit. Health*, vol. 2, no. 8, pp. e397–e406, Aug. 2020, doi: 10.1016/S2589-7500(20)30135-7.
[60] B. Erten Uyumaz, L. Feijs, and J. Hu, “A Review of Digital Cognitive Behavioral Therapy for Insomnia (CBT-I Apps): Are They Designed for Engagement?,” *Int. J. Environ. Res. Public. Health*, vol. 18, no. 6, p. 2929, Mar. 2021, doi: 10.3390/ijerph18062929.
[61] P. Chandra, G. Joshi, and R. Bhagwat, “ChatGPT’s Evolution in Reshaping Cognitive Behavioral Therapy,” in *2023 IEEE Engineering Informatics*, Nov. 2023, pp. 1–9. doi: 10.1109/IEEECONF58110.2023.10520423.
[62] P. Green, “The Sound of Silence,” *The New York Times*, Dec. 27, 2018. Accessed: May 23, 2024. [Online]. Available: https://www.nytimes.com/2018/12/27/style/white-noise-machines.html
[63] “AASM Sleep Trends Survey - Sleep Trends,” American Academy of Sleep Medicine, 2023.
[64] E. Capezuti, K. Pain, E. Alamag, X. Chen, V. Philibert, and A. C. Krieger, “Systematic review: auditory stimulation and sleep,” *J. Clin. Sleep Med.*, vol. 18, no. 6, pp. 1697–1709, Jun. 2022, doi: 10.5664/jcsm.9860.
[65] C. Rio-Alamos, R. Montefusco-Siegmund, T. Cañete, J. Sotomayor, and A. Fernandez-Teruel, “Acute Relaxation Response Induced by Tibetan Singing Bowl Sounds: A Randomized Controlled Trial,” *Eur. J. Investig. Health Psychol. Educ.*, vol. 13, no. 2, Art. no. 2, Feb. 2023, doi: 10.3390/ejihpe13020024.
[66] A. K. Patel, V. Reddy, K. R. Shumway, and J. F. Araujo, “Physiology, Sleep Stages,” in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2024. Accessed: May 23, 2024. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK526132/
[67] M. Strauss, J. D. Sitt, L. Naccache, and F. Raimondo, “Predicting the loss of responsiveness when falling asleep in humans,” *NeuroImage*, vol. 251, p. 119003, May 2022, doi: 10.1016/j.neuroimage.2022.119003.
[68] Y. Wei *et al.*, “Sleep Stage Transition Dynamics Reveal Specific Stage 2 Vulnerability in Insomnia,” *Sleep*, vol. 40, no. 9, p. zsx117, Sep. 2017, doi: 10.1093/sleep/zsx117.
[69] N. S. Gooneratne *et al.*, “Case-Control Study of Subjective and Objective Differences in Sleep Patterns in Older Adults with Insomnia Symptoms,” *J. Sleep Res.*, vol. 20, no. 3, pp. 434–444, Sep. 2011, doi: 10.1111/j.1365-2869.2010.00889.x.
[70] D. Fekedulegn, M. E. Andrew, M. Shi, J. M. Violanti, S. Knox, and K. E. Innes, “Actigraphy-Based Assessment of Sleep Parameters,” *Ann. Work Expo. Health*, vol. 64, no. 4, pp. 350–367, May 2020, doi: 10.1093/annweh/wxaa007.
[71] M. L. Blood, R. L. Sack, D. C. Percy, and J. C. Pen, “A Comparison of Sleep Detection by Wrist Actigraphy, Behavioral Response, and Polysomnography,” *Sleep*, vol. 20, no. 6, pp. 388–395, Jun. 1997, doi: 10.1093/sleep/20.6.388.
[72] B. Türker *et al.*, “Behavioral and brain responses to verbal stimuli reveal transient periods of cognitive integration of the external world during sleep,” *Nat. Neurosci.*, vol. 26, no. 11, pp. 1981–1993, 2023, doi: 10.1038/s41593-023-01449-7.
[73] “Quantification of acceleration as activity counts in ActiGraph wearable - PMC.” Accessed: May 23, 2024. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9279376/
[74] M. R. Patterson *et al.*, “40 years of actigraphy in sleep medicine and current state of the art algorithms,” *NPJ Digit. Med.*, vol. 6, p. 51, Mar. 2023, doi: 10.1038/s41746-023-00802-1.
[75] R. G. Weaver *et al.*, “Evaluation of a device-agnostic approach to predict sleep from raw

accelerometry data collected by Apple Watch Series 7, Garmin Vivoactive 4, and ActiGraph GT9X Link in children with sleep disruptions," *Sleep Health J. Natl. Sleep Found.*, vol. 9, no. 4, pp. 417–429, Aug. 2023, doi: 10.1016/j.sleh.2023.04.005.
[76] Z. Su *et al.*, "Exploring the relationship between response time sequence in scale answering process and severity of insomnia: a machine learning approach".
[77] R. D. Ogilvie, "The process of falling asleep," *Sleep Med. Rev.*, vol. 5, no. 3, pp. 247–270, Jun. 2001, doi: 10.1053/smrv.2001.0145.