Sleepalyzer: Use of a Portable Psychomotor Vigilance Test to Measure Performance

By Jenny Chen

Executive Summary

Sleepalyzer is a smartphone implementation of the PVT-A and related algorithms to accurately and portably measure a person's reaction time and drowsiness. Here we present a breakdown of not just the application, but also potential market uses for the application. Ultimately, our goal was to create a smartphone application that was just as good, if not better, than the lab version.

Overview

Sleep deprivation is a serious issue that can impact not just an individual's mental and physical health, but also the safety of those around him or her. The number of sleep related accidents range from 10-30%[1] with a corresponding economic impact of $43 to 56 billion[2]. Lack of sleep in

[1] Effect of driving duration and partial sleep deprivation on subsequent alertness and performance of car drivers

[2] Neurocognitive Consequences of Sleep Deprivation

the United States alone, according to a study led by RAND Europe, results in a loss of up to $411 billion a year in lost productivity and mortality[3]. How much sleep a person gets has also been shown to have adverse effects on a person's mood, cognitive performance, and motor performance.[4] The dangers of sleep deprivation is well documented, yet 20% of individuals are considered to have insufficient sleep. This issue is of widespread concern for many industries, ranging from transportation and the military to doctors and astronauts. Quite simply, sleepiness can be dangerous -- it can lead to both physiological and physical stress as well as lapses in judgement

The Psychomotor Vigilance Test (PVT) was developed as a reaction task that was free of aptitude and learning effects, but also sensitive to sleep loss as well as any other factors that would impede a person's alertness. It is also impossible to cheat -- you cannot be faster than you actually are. The PVT, and its related algorithms, the PVT-A (adaptive, 10 minutes) and PVT-B (3 minutes), are sensitive to fatigue by requiring an individual to maintain constant vigilance throughout the task, and sleep deprived or less alert individuals will tend to make more false starts or lapses. Various versions of test have been developed and has even been used at places such as the International Space Station, but none have been released on a mobile platform and are limited to largely research or laboratory settings because of the precision requirements of the test (10-15 ms error range). Having a way to accurately measure and detect alertness provides an objective measure for how well a person might be able to perform a task or even simply provide a consumer more information on what affects his or her reaction time.

Market Opportunity

Unfortunately, the current laboratory only implementation limits the usage of the PVT as is. An accurate mobile implementation of the PVT has not yet been accomplished, much less a version released to the public. Yet, a mobile implementation would give organizations and researchers a portable and practical way of evaluating a person's fitness to perform tasks that require attention.

There is great incentive for companies to engage in fatigue reduction programs. An average Fortune 500 company could save nearly $40 million a year if half of its workforce engaged in a sleep-health program, and providing a way to track and measure alertness would be a vital part of any such program[5]. Workers often experience low levels of energy or poor sleep, with 38% of 29,000 surveyed saying they had experienced fatigue in the past two weeks[6]. This is a direct loss in productivity as well as a decrease in safety in working conditions. A smartphone version of the PVT provides a user friendly way of measuring and tracking alertness, facilitates more widespread adoption, and has direct benefits to companies.

---

[3] Why Sleep Matters: Quantifying the Economic Costs of Insufficient Sleep
[4] Effects of Sleep Deprivation on Performance
[5] Calculator shows hidden costs of fatigued workforce
[6] Study: Workplace Fatigue Common, Costly

Trends and Sentiments

Over the past few years, consumers have become more and more interested in purchasing or using products that maximize their well-being. There is a sense that consumers want to be more in control of their health, while on the other side, employers and insurers have an incentive to make sure individuals are healthy. Regardless of the reason, consumers are becoming much more health conscious. This has been reflected in the consumer goods industry at least since 2015, when health conscious foods grew more and more popular, especially for companies whose sales were on the smaller side.

This includes the area of sleep and technology, with apps appearing regularly in both the App and Google Play Store. As of two years ago, there were at least 51 unique apps around this purpose[7]. The sleep market is surprisingly diversified, in that the solutions range from furniture and lighting to mediation and sleep apparel. However, ultimately, the goal behind each of these products is to improve a person's sleep health or quality, which is not necessarily something that Sleepalyzer does directly as of now. This represents a market in which Sleepalyzer could potentially pivot to help fill, especially as consumers seek out new resources.

Size and Growth of Market

Reflecting the overall trend toward personalized health, the sleep health market has been growing as well and is worth between $30 to $40 billion. More importantly, it has grown about 8% every year and this rate does not seem to be decreasing any time soon[8]. In the last few years, sleep startups have raised more than $700 million in seed funding. In 2016-2017, the funding amounted to just under $300 million.

One of the most well-known sleep startups is Caspar, which makes mattresses, and this company raised a $55 million Series B, and later Target even attempted to buy the app for $1 billion before changing its mind[9]. It is important to note that the definition of a "sleep health" startup varies widely, from physical products to sleep tracking apps, and more. However, it seems that the market itself is on an upward trend.

Stakeholders

The main stakeholders or interested parties would include both individuals, employees, and employers. Researchers and academics in the field may also be interested in such a product.

Especially for people in professions that require long hours, such as doctors, truck drivers, pilots, and military personnel, alertness can be correlated with how well they can perform their job and can have a direct impact on the safety of both themselves and others.

---

[7] Overview of smartphone applications for sleep analysis
[8] Investing in the growing sleep-health economy
[9] Chasing dreams may be the next sleeper hit for venture capitalists

For context, this is just a chart estimating lost productivity due to sleep deprivation. Clearly, employers have an incentive to avoid this.

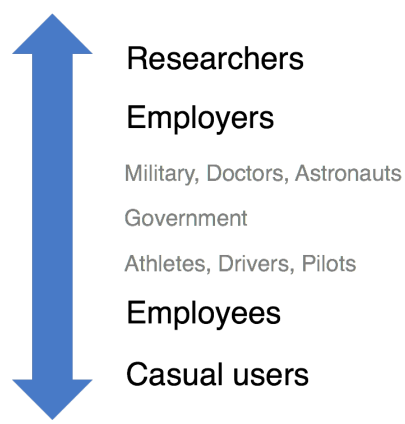
Total Working Time Lost Across Five OECD Countries

Country	U.S	UK	Germany	Japan	Canada
# Full-time workers (in thousands)	121,490	22,733	28,965	47,790	14,559
# Part time workers (in thousands)	27,340	8,296	11,245	14,000	3,387
Days lost (full-time): < 6 hours	528,377	87,372	60,545	185,289	22,089
Days lost (full-time): 6 to 7 hours	479,643	64,447	90,023	282,009	41,888
Days lost (part-time): < 6 hours	118,906	31,885	23,506	54,280	5,139
Days lost (part-time): 6 to 7 hours	107,938	23,519	34,951	82,614	9,745
Days lost: total	**1,234,864**	**207,224**	**209,024**	**604,191**	**78,861**
Hours lost: total	**9,878,910**	**1,657,792**	**1,672,192**	**4,833,532**	**630,886**

Excerpt from Table ES. 1 Rand

Other potential usages could include: medical screening (e.g., in waiting rooms), fit-for-duty testing (e.g., prior to a work shift, in truck drivers on highways), combined use with activity trackers (e.g., Fitbit), and enhancing sleep and performance in professional athletes. Researchers would also be able utilize the tool in their studies and in clinical trials. There are clear and real detriments toward being sleep deprived, and it has been shown that sleep deprived subjects performed at a level 1.37 standard deviations lower than the performance level of the non-sleep-deprived subjects[10]

Sleepalyzer would be useful and valuable to not just employees or employers who want to assess sleepiness, but also the general consumer who wants more data on what may affect their reaction times or simply a way to track performance over time. Quite simply, this provides consumers and users the data to make more informed decisions about what may affect their reaction times and alertness.



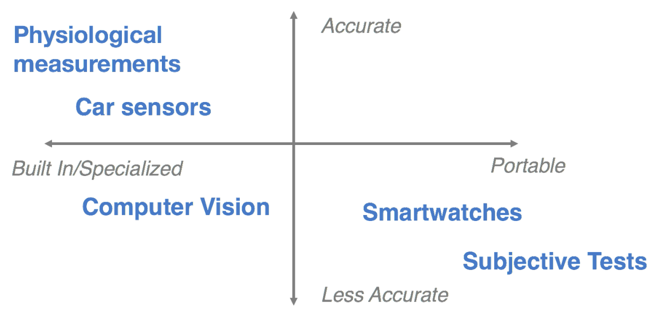
However, this does mean the potential range of users is quite large and it is not feasible to simply target all of these potential users all at once. We propose that starting with researchers who understand the value of the PVT algorithms is likely the easiest step to take, as in some ways the “marketing” aspect is already done. From there, endorsements from the scientific community would come regularly and provides a base from which to expand the market.

[10] Effects of Sleep Deprivation on Performance

Competitive Analysis

Detecting drowsiness is a simple but important test. There are various sleepiness tests and surveys that currently exist. Especially for driving, many car manufacturers have made attempts to include technologies that can detect a sleepy driver.

There are existent sensors that use the steering wheel and deviation from a lane in order to measure drowsiness. In addition, computer vision techniques have been used to gain clues from a driver's face, but this technique suffers from lighting issues. There are also physiological measure (e.g. heart rate, EEG) that may indicate drowsiness[11]. While physiological signals actually tend to be the most accurate, they are also obviously the most intrusive. Most of these techniques are viable, but either too specialized or too intrusive for a test that should be taken quickly and easily.



Indeed, the landscape for sleepiness detection is currently divided into what amounts to two sections. One contains highly specialized technology that can measure alertness with high accuracy and precision. The other contains highly portable techniques that are easily accessible, but are not reliable. For example, many subjective measures depend on rating scales based on self-evaluations by the individuals. However, while these are simple to administer and not time consuming, this introduces unintended bias and purposeful falsification[12].

Stats on existing sleep apps:
- 33 apps on App Store
- Average price for sleep apps is $1.12
- Average rating is 3.8 stars
- Sleep Cycle
  - subscription costs $29.99
  - 75.1K users

Current sleep apps and health apps in general are gaining in popularity. However, many of these sleep apps don't measure how sleepy you are -- they are generally centered around waking you up -- but are in the same market as Sleepalyzer for the average consumer. However, one thing to note is that these apps are not validated by research and generally not scientifically proven[13]. This is a major advantage for Sleepalyzer. The Psychomotor Vigilance Test (PVT) is peer-reviewed and has been extensively validated to be sensitive to deficits in attention from sleep loss and circadian misalignment.

Sleepalyzer also intends to address this market gap between high accuracy and high portability through an adaptive smartphone solution of the PVT-A. We envision a product that can be opened on your phone or tablet, the tests taken and results and metrics immediately given to the user.

[11] Detecting Driver Drowsiness Based on Sensors: A Review
[12] Measurements of sleepiness and fatigue
[13] Overview of smartphone applications for sleep analysis

This is not necessarily a novel idea. Attempts have been made in the past to create a working smartphone app for the PVT. However, the main challenge is in the latency in the touch screen of a smartphone, which while it has improved, is still a major technical challenge. There is an existing version of the PVT on Palm OS (Thorne 2005) and a version on a PC (Khitrov 2013). Obviously, a PC is not portable and the Palm OS is quite outdated and does not implement the version of the algorithm we want.

In addition, the PVT-A algorithm has never been implemented on any platform, which gives Sleepalyzer a competitive advantage in having a faster, scientifically proven version of the PVT available.

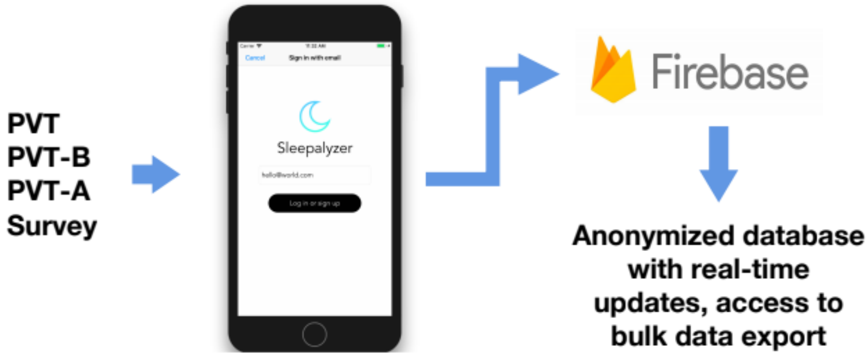
Product Analysis and Process

Sleepalyzer's main deliverable for this project is an accurate smartphone implementation of the PVT-A (along with variations of the PVT). Since the reaction time tests themselves already existed and were validated with 30+ years of research by Dr. Dinges and Dr. Basner[14], our approach to the problem consisted mainly of working with the doctors to gather the information we needed to implement the tests, and researching the best ways to implement the features we targeted.

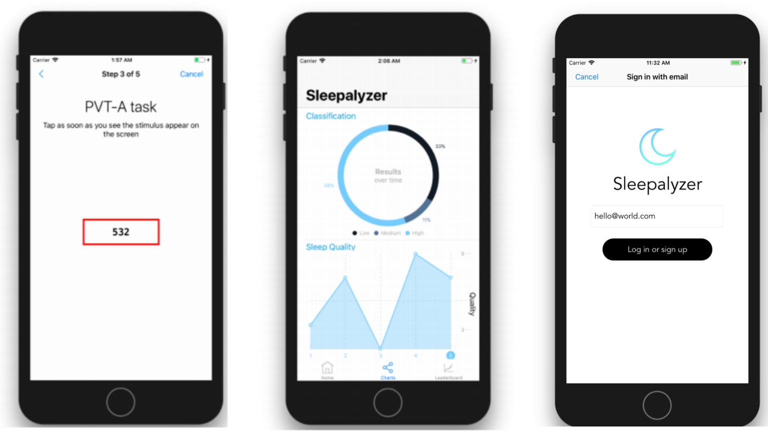
Our first step was to implement the PVT-A algorithm in Swift. Using the research papers provided by Dr. Dinges and Dr. Basner, we worked to implement the Bayesian algorithm and explored the experimental data that our medical advisors provided us. The next step was to actually write the code in the app, but this involved research on what the best framework to develop the app on would be. We ended up using ResearchKit, which is provided by Apple as a platform to develop medical apps on. However, we still needed to customize quite a bit of the tasks/views provided by ResearchKit to fit the PVT-A. Specifically, we needed to register taps at certain times, reveal stimulus at certain times, keep track of overall test time, and detect false starts and lapses.

All of this data would eventually feed into the PVT-A and other algorithms to ultimately determine a person's sleepiness level on every single iteration of the test. We experimented with reducing the upper bound of test time as well as purposefully testing with different scenarios (e.g. making reaction times purposefully slow, lots of false starts) to help qualitatively validate that the task was working as expected. In addition to being used in-app, the data is also stored to Firebase, which provides real-time updates and the ability to export and analyze data in bulk.

[14] See some relevant papers: here, here, here, and here



Visually, the PVT-A test in the app looks something like what is shown below. The color of the box's outline gives immediate feedback on how a user is performing based on what the algorithm has calculated. Each color is scaled based on performance. Once the algorithm has reached its threshold, we end the test and display the user's results. We also allow the user to end the task early through the cancel button. This then fast forwards them past the rest of the test. Aside from the suite of different PVT tasks, we also have two other sections of the app. One is a survey containing qualitative questions that the doctors mentioned they would need in their research, and the other is a tab for charts/graphs, which draws from the user's survey answers and PVT test results to show various visualizations of the data.



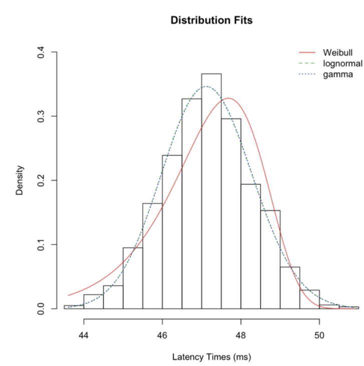
Technical Challenges

One of the key challenges we faced was combating the inherent latency of a smartphone touchscreen, which can range from 50 to 100 milliseconds; for comparison, the PVT-A and related algorithms can only tolerate a latency of about 10-15 milliseconds. Thus, while development of the major features of the app was ongoing, we were also conducting research in parallel on possible ways to mitigate this. After considering a variety of different options, including using other sensors (accelerometer, microphone, etc.) to better detect a “true” tap event, or implementing a calibration phase (this would have required additional hardware), we ended up using a slow-motion camera to measure latency over a large number of trials. Using these measurements, we were able to find a distribution of latency that we could use to correct the recorded reaction times in-app.

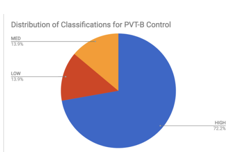
This obviously means we are limited to the device we tested on, which is the iPhone 6s and there could be potentially better ways to measure and account for latency. More research would likely need to occur here, and options range from using additional sensors, adding a small hardware sensor, measuring all device types (roughly), and/or adding a calibration phase.

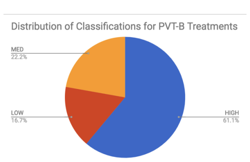
Evaluation

Our steps for evaluation are threefold. The first part of our evaluation was to assess the distribution of measured touch screen latencies on iPhone devices and find an expected latency value from that distribution. As mentioned above, we used a slow-motion camera to record over 100 touch screen latencies. Using R, we fitted various statistical distributions to this latency data, including the the Weibull, Gamma, and LogNormal distributions. We found that the data does not quite fit to a normal distribution, so taking the mean of the latencies was not the most optimal way of calculating expected latency. Ultimately, we found that the data fit nicely to the Gamma distribution, which features a tail that is skewed to the right).



Using the expected value associated with the Gamma distribution, we were able to find the expected latency value associated with touch screen responses on the iPhone device. This value is then subtracted from measured times for responses in our algorithm.

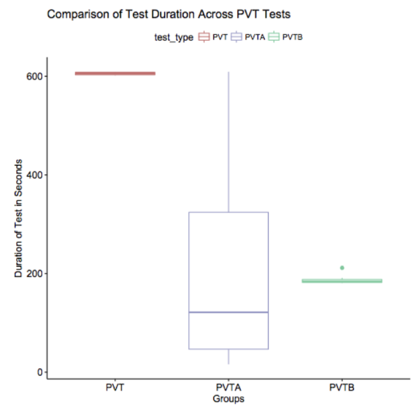




Next, we wanted to evaluate how accurate our implemented PVT-B (3-min test) and PVT (10-min test) are. To do so, we used the iPad app Joggle’s PVT-B and PVT test as a gold standard (there is no PVT-A on a portable platform). Using the tests on Joggle, we are able to obtain key metrics on each test performed and compare them to the same metrics from our implemented versions. The Joggle app is used widely by cognitive science researchers, but it is an expensive ($100) app and takes a lot more storage since it contains a wide suite of tools in addition to the PVT tests.

Our experimental framework was to seek out a sample of students and perform a 2-sample paired t-test. That is, for each student sampled, we make them take the Joggle PVT-B test and our implemented iPhone PVT-B test. We compare the metrics from the Joggle version and our version, and we see if these differences are significantly different or not. We did the same thing for the PVT test. We ended up sampling over 30 different students in our data collection process. We measured several metrics, like number of stimuli, number of lapses, median reaction time, and classification metric (used to determine high/medium/low status). Across the board for both the PVT-B and PVT tests, we found that the differences were insignificant (p-value above 0.9 in many cases, indicating strong insignificance) using significance level 0.05. What this reveals is that our iPhone implementation is able to replicate the performance of the currently adopted PVT tests in academia.

Lastly, we wanted to assess whether or not our implementation of the PVT-A algorithm is able to effectively reduce the time of the test. Using duration of test as our metric of interest, we performed 2-sample t-tests with the PVT-A experimental data and the PVT-B experimental data. We repeated this process with the PVT experimental data. We conducted a 1-sided test (testing to see if PVT-A significantly reduced the time of test) using significance level 0.05, and we found that there was significant evidence that the PVT-A was able to effectively reduce the time of test. This is important because there is no an existing iPad or iPhone implementation of the PVT-A algorithm, which gives Sleepalyzer a competitive advantage in being the first to have an implementation of the faster research-verified PVT-test.



Revenues

There are various ways we could generate revenue from the app, mostly following the standard ways of generating revenue from a mobile app, including, but not limited to advertising, freemium, use and analysis of data. Most of the consumer sleep applications currently on the market are free with some type purchasing add on. For example, SleepCycle, one of the most popular sleep health apps has a $30 subscription option and adds additional data analysis and features to the base version of the app. Sleepalyzer would likely have to follow a similar model in this market.

Sleepalyzer has potential to hit many different types of users, as outlined in the Stakeholders section. This provides additional opportunities to use various revenue models depending on the market targeted. For example, researchers, consumers, and employers may be interested in very different types of metrics and data analysis. Researchers likely care very much about the usability of the application whereas everyday consumers are almost just as focused on how the application itself looks. However, it is possible that each customer segment could be charged differently according to their needs, which in turn provides potentially additional revenue generation.

Costs

The next step would to be to look at costs for the application. In terms of costs, a lot of the upfront research validating the app has already been done so costs generally would center around further developing the app and marketing costs. To optimize for the best user experience, it would be a good idea to A/B test different features in order to provide a comprehensive app, focus development, and improve user metrics.

However, in our evaluations on Sleepalyzer as well as our own experience, many participants complained about the length of the PVT-B, which is a fixed three minute version. It is true the PVT-A can be shorter, but it also has the potential to be longer and up to ten minutes in length. Hence, it is likely that any commercial, consumer facing app of Sleepalyzer will need to include a version of the PVT that is at most one minute and ideally less than that. This would require more research by likely Dr. Dinges and Dr. Basner and does represent a time and resource sink and cost.

In addition, a consumer facing of the app would likely need additional features that bring it up to par and potentially better than existing sleep apps. This may include adding features like an alarm clock, a sleep diary, and other features that allow the consumer more flexibility and range of use. There would be a development cost associated with this.

It is important to note that these would not be costs for a research-only based version of the app, though the market size for such a research only app would also be smaller. Employers may also have different requirements. It is important to properly survey the different user requirements and select the customer segment that seems the most feasible initially before expanding.

Other concerns and costs would include privacy of the data and scale. As it is currently set up, the architecture may not allow for significantly large sets of users.

Conclusion

There are three main avenues we see Sleepalyzer expanding into. One to recognize is that the algorithm itself has value -- it is a scientifically reviewed and proven test that could be interesting to outside health vendors, other researchers, and consumers. Two other markets to approach are the everyday consumer and professionals/companies.

Consumers could see this app as a more complete sleep health app that has scientific backing and companies as a way to potentially incorporate this into a fatigue or health program. For example, this may be an application that an individual could bring to their doctor or their doctor could use to help gauge a person’s alertness or reaction time. Sleepalyzer reflects the growing trend toward health consciousness and provides a data driven approach toward solving a problem.

Sleepalyzer, though initially focused toward addressing the issue of fatigue, is a test that can be applied to any issue that affects alertness and reaction time. Another potential use case would be that of opioid addiction. This means Sleepalyzer has an impact not limited just to sleepiness.

Indeed, Sleepalyzer fits into the niche of being both accurate and portable. Using the app, a user is able to get an objective measure of their alertness level and receive feedback immediately. Ultimately, fatigue is a real issue and sleep matters. Sleepalyzer helps address that.