Cognitive Rhythms: Unobtrusive and Continuous Sensing of Alertness Using a Mobile Phone

Saeed Abdullah[1], **Elizabeth L. Murnane**[1], **Mark Matthews**[1], **Matthew Kay**[2], **Julie A. Kientz**[2], **Geri Gay**[1], **Tanzeem Choudhury**[1]
[1]Cornell University,[2]University of Washington
{sma249, elm236, mjm672}@cornell.edu,
{mjskay, jkientz}@uw.edu, {gkg1,tkc28}@cornell.edu

ABSTRACT
Throughout the day, our alertness levels change and our cognitive performance fluctuates. The creation of technology that can adapt to such variations requires reliable measurement with ecological validity. Our study is the first to collect alertness data in the wild using the clinically validated Psychomotor Vigilance Test. With 20 participants over 40 days, we find that alertness can oscillate approximately 30% depending on time and body clock type and that Daylight Savings Time, hours slept, and stimulant intake can influence alertness as well. Based on these findings, we develop novel methods for unobtrusively and continuously assessing alertness. In estimating response time, our model achieves a root-mean-square error of 80.64 milliseconds, which is significantly lower than the 500ms threshold used as a standard indicator of impaired cognitive ability. Finally, we discuss how such real-time detection of alertness is a key first step towards developing systems that are sensitive to our biological variations.

Author Keywords
Alertness; Performance; Sleep; Circadian Rhythms

ACM Classification Keywords
J.3 Life and Medical Sciences: Health

INTRODUCTION
Our cognitive abilities wax and wane over the course of the day [57]. In particular, *alertness* — a central component underlying a wide range of cognitive functions from learning to problem-solving to memory consolidation [39] — fluctuates in daily cycles called *circadian rhythms* [61]. Alertness is also influenced by a number of other factors, including sleep and one's internal body clock type (for example, whether one is an "early bird" or a "night owl") [13].

Given the key role alertness plays in cognitive performance, improving individuals' everyday alertness could have far-reaching positive impacts: it could improve productivity [3], help enhance learning outcomes [32], and combat road accidents and occupational errors caused by fatigue [17]. These areas are ripe for novel technological solutions. HCI and UbiComp researchers have recently shown interest in studying related aspects of cognitive performance such as attention and boredom (e.g., [41, 52]). However, to build effective solutions in this space, we must account for the fluctuating nature of alertness and the behavioral, environmental, social, and *biological* factors driving those fluctuations — something that current work rarely does [57]. Moreover, alertness patterns are different from one person to the next [13], necessitating *personalized* rather than generalized recommendations.

Novel UbiComp systems built on *continuous, individual* assessment of alertness therefore have the potential to significantly improve learning [4], occupational safety [17], work performance [3], and overall quality of life [3]. What could we build if we knew how alert someone was at all times of day? Unfortunately, existing methods for assessing alertness are cumbersome and ill-suited to continuous measurement: typically, alertness is assessed through stimulus-response tests, such as the widely-used Psychomotor Vigilance Task (PVT) [18]. This is a 2–10 minute task, during which the participant is presented with visual stimuli at random time intervals and asked to respond (e.g., press a button or touch a screen) as soon as they see the stimulus. Various statistical summaries of the resultant *response time* have been used as a measure of alertness. Studies utilizing tests like PVT are normally conducted in controlled lab environments under artificial conditions (e.g., to assess alertness after enforcing 85 hours of total sleep deprivation [64]). While the PVT has been implemented on smartphones [31], which makes it amenable to use in the wild (e.g., in experience sampling), any such use still requires at least 2 minutes of a person's undivided attention for each assessment — rendering it unsuitable for continuous measurement over extended periods of time.

We present an alternative: continuous, *unobtrusive* assessment of alertness using machine learning, backed by features of an individual's smartphone use and validated against PVT. This approach opens the door for more personalized, context-aware tools that can model, adapt to, and provide feedback on our alertness (and in turn our cognitive performance) in real time. In this work, we focus on collecting data in the wild in order to develop passive approaches for estimating alertness.

UbiComp '16, September 12–16, 2016, Heidelberg, Germany

ACM 978-1-4503-4461-6/16/09...$15.00
DOI:http://dx.doi.org/10.1145/2971648.2971712

We make the following specific contributions:

- Moving outside the controlled environment of a laboratory, we use experience sampling to assess alertness as people go about their everyday lives. We also collect diverse streams of self-reported and passively-sensed data, including body clock type, sleep information, and stimulant use. Using a validated smartphone-based PVT tool [31], we operationalize (measure) alertness in two ways: as *response time* and as *relative response time (RRT)* — the percent deviation of a response time from that person's overall mean response time [31, 62]. The widespread adoption and high usage levels of smartphones suggest that our methodology and models could potentially be applied at-scale.

- Collecting data over 40 days with 20 participants, we:

  – **Extend lab findings from the cognitive performance literature into the wild.** As the first study to focus on patterns of alertness using data captured in the wild, we seek to compare and extend findings from previous laboratory based work. Replicating prior findings is not only important for advancing scientific knowledge but it also validates our methodology. Among other comparisons, we demonstrate that alertness across the day is influenced by local time, internal body clock time, daylight savings, hours slept, and stimulant intake.

  – **Demonstrate passive alertness sensing on a smartphone.** Towards the vision of unobtrusive and continuous assessment of cognitive performance that can be deployed among large, diverse, and distributed populations, we describe novel methods for automatically estimating alertness. Specifically, we develop models grounded in chronobiology that can predict response time with an RMSE of 80.64ms (which is significantly lower than the 500ms threshold used as a standard indicator of impaired cognitive ability [7]). Using only passively sensed data, models for RRT also perform well with an average RMSE of 10.87% across all participants (i.e., we can detect as small as an 11% deviation in response time from an individual's baseline, which is significantly lower than the average daily peak-to-peak change in RRT of 29.4%).

- Lastly, we interpret key findings and discuss the broader implications of our work for the UbiComp and context-aware computing fields, specifically focusing on the opportunities our findings offer in the areas of scheduling, education, and accident-prevention.

RELATED WORK

Cognitive Performance

Our cognitive functions can be categorized into three main domains: i) attention, ii) executive functioning, and iii) memory [57]. These domains are usually assessed by measuring *performance* at specific tasks.

An essential factor in optimal performance is *alertness* given that it is a core subcomponent of the attentional system, which modulates sensory, motor, and cognitive processing [61]. Indeed, fatigue — a state of diminished alertness — is linked to motor vehicle accidents, industrial disasters, and other occupational errors [17] and has been equated to alcohol intoxication in terms of its negative impact on performance [36].

HCI and UbiComp research on cognitive performance has so far focused on understanding and modeling various aspects of attention. Research on patterns and contexts of attentional states in a workplace environment found that focused attention peaks during mid-afternoon and that online activities can reflect attentional states — associating, for instance, heavier use of productivity software with focused periods and internet surfing and window switching with boredom [41]. Leveraging such usage behaviors, other work has developed machine learning algorithms to automatically infer boredom [43, 52] from mobile phone use (e.g., the amount and types of apps used), along with contextual information (e.g., light levels and charging status) and demographics. However, while such studies have reported a time-of-day effect on attention, the fact that they do not take any chronobiological factors into consideration hinders both the range of analyses explored as well as the ability to offer biologically-informed explanations as to *why* particular trends are observed.

Circadian Rhythms of Performance

While cognitive performance is well studied, only recently has a more complete and nuanced understanding emerged from the field of chronobiology regarding how and why performance (and alertness specifically) fluctuates throughout the day. Specifically, these fluctuations are driven by the human body's "biological clock", which influences our sleep-wake patterns and numerous behavioral, mental, and physical processes that follow a roughly 24-hour cycle, known as our *circadian rhythms*. These rhythms are endogenous and rooted in our genetics, though external cues (predominantly sunlight) adjust them to the local environment [13].

Early work has connected cognitive performance with rhythms in body temperature (a well-known biomarker for circadian rhythms) — associating increased body temperature with better performance and vice versa [33]. More recently, a broad set of empirical work has confirmed these performance rhythms, including for alertness. Prominent among findings are a short-term, mid-day dip during which sleepiness increases and alertness drops [13, 46] as well as an evening alertness rebound [38].

While all (healthy) individuals experience these variations in cognitive performance and alertness levels throughout the day, the amplitudes and phases of these patterns display individual variability. A person's *chronotype* represents his or her unique circadian rhythms and lies on a spectrum from early to late types. Early types tend to be more alert earlier in the day while late types are more alert later [27, 61]. Beyond alertness, higher level cognitive processes including orientation and executive functioning also show differences depending on chronotype and time of day [42].

Age can also influence alertness patterns. While older adults are typically more alert in the morning, younger adults tend to be more susceptible to distraction in the morning and are more alert in the afternoon [25]. In addition, alertness levels can be impacted by external factors such as caffeine or alcohol consumption, food intake, and physical activity [57]. Finally, sleep can have marked impacts on alertness and cognitive performance, with inadequate sleep degrading alertness and extending sleep improving it [22].

Sensing Rhythms of Performance

Unfortunately, chronobiology research and performance modeling is typically constrained by challenges in approximating ecologically valid scenarios, since most studies are conducted in controlled, artificial lab setups. The emergent area of "circadian computing" aims to develop novel computational techniques for passively sensing biological rhythms using more broadly deployable, unobtrusive, and technology-mediated approaches. Its preliminary steps have focused on sleep sensing using smartphone screen on-off patterns [2] and social media data [47]. In this study, we go beyond sleep to consider daily cognitive performance. Specifically, we extend the findings from previous laboratory-based studies from chronobiology, cognitive psychology, and neuropsychology to introduce new methods for in-situ, at-scale computational modeling and prediction of alertness performance.

METHOD

Participants

In this study, we set out to detect daily variations in alertness for a student population. We focus on college students, who biologically tend to be significantly late types [55]. They are also known to experience significant dissonance between their internal biological rhythms and their socially-constrained scheduling [32, 55] — for instance, in the timing of their classes, meetings, and social engagements, among other activities. Chronic sleep deprivation and its impacts on cognitive performance and alertness are serious problems for this population [60] and can lead to drug and alcohol abuse, increased mental health problems, impaired academic performance, and learning deficits [11]. Regarding cognitive performance specifically, procrastination is just one manifestation of dips in alertness that is commonly observed in college students [15]. Students also tend to be heavy smartphone users [58] and are therefore well-suited for a passive sensing methodology based on smartphone usage behaviors.

Our inclusion criteria required participants to be an Android phone user and willing to participate for the full duration of the study. Through public mailing lists, recruitment portals, and snowball sampling, we recruited 20 individuals: 7 males and 13 females who all fall in the 18 – 29 year old age group.

Assessment Instruments

Momentary Assessments

To reiterate, we focus on alertness due to its well-established impact on almost every aspect of cognitive ability. To allow us to monitor alertness throughout the day along with relevant factors mentioned previously, our participants installed an Android application on their mobile phones that we developed to deliver a brief ecological momentary assessment (EMA) comprised of subjective assessments and an objective assessment.

To capture subjective measures of alertness and fatigue, our EMA included validated, brief questions from the Chalder Fatigue Scale [14] and Fatigue Visual Analogue Scales (VAS) [45]. Participants also checked off activities they had done in the past hour — specifically, consuming caffeine [64], exercising [12], using nicotine [23], napping [12], consuming alcohol [26], eating [54], and loafing (e.g., cyberloafing) [63] — to supply information about stimulants and anti-stimulants known to impact cognitive performance and sleep.

To objectively assess alertness, we used the Psychomotor Vigilance Task (PVT), a reaction test commonly used to measure alertness [40]. While the original PVT requires special hardware, we employed a validated Android smartphone implementation, PVT-Touch [31], which shows a visual stimulus at random intervals to the user, who responds by touching the screen. Response times are measured in milliseconds, and various statistical summaries of these times have been shown to be indicative of alertness [40]. One distinct advantage of using the PVT to assess cognitive performance over a long period of time is its immunity to practice or learning effects [37]. Further, brief versions of the PVT have proved sensitive to changes in alertness [7]. We administered a 3-minute version, a duration validated for alertness assessment [7, 8].

Processing PVT data first involves removing false starts — touch events before the stimulus is shown. In our dataset, false starts amounted to 2.85% of all touch events. Continuing to follow previous work [31, 62], we operationalize alertness as relative response time (RRT), computed as follows. First, since a PVT session includes multiple visual stimuli tests, we calculate the median response time ($MRT_{s,p}$) for each session s per person p. We also remove outlier sessions with $MRT_{s,p}$ falling outside (mean $\pm\, 2.5 \times$ SD) for each participant [62]. In our sample, 6.4% of all sessions were removed as outliers. Next, we take the mean $MRT_{s,p}$ across all of participant p's sessions to establish an individual baseline for participant p. Finally, we compute the RRT of a given session as its percentage deviation from p's individual baseline[1]. That is, given a PVT session s for a participant p with a median reaction time for that session of $MRT_{s,p}$, the corresponding RRT is calculated as:

$$RRT_{s,p} = \left(1 - \frac{MRT_{s,p}}{MMRT_p}\right) * 100,$$

where $MMRT_p = \frac{1}{N}\sum_{i=1}^{N} MRT_{i,p}$, is the mean MRT averaged across all N sessions from participant p. Positive RRT values thus indicate increased alertness, and negative RRT values indicate decreased alertness.

We placed the PVT at the end of the EMA battery, after the subjective assessments, because PVT-induced fatigue can cause a reverse biasing effect on subjective responses; plus

[1]Our code for computing and analyzing RRT is available at `https://github.com/saeed-abdullah/alertness-ubicomp-2016`

since the PVT's user burden is higher, placing it first might result in decreased compliance. Our current ordering allows participants to quit the EMA without completing the PVT but having at least completed the subjective questions. However, to protect against data quality issues arising from such non-compliance especially to the PVT portion, we instructed participants during an onboarding interview to only begin the EMA if they expected to have 5 minutes without significant distraction. We also provided instructions and demonstrated how to perform the PVT and tested it thoroughly on each person's phone; our instructions stressed focusing on accuracy and speed when responding to the stimuli.

We delivered the EMA four times per day at the start of 6-hour-long morning, afternoon, evening, and late night time windows defined by prior work [1] via passive phone notifications, which research has shown are successful in improving compliance [9]. Participants could complete the EMA anytime within a given window. (As mentioned, we instructed participants that an immediate response to phone notifications was not necessary but rather to wait to undertake EMAs until they had at least 5 distraction-free minutes). Once that time window ended, its notification would expire and a new one would be delivered in order to avoid redundant or temporally-mislabeled assessments. We chose this study design both to increase data coverage throughout the day as well as to improve data quality since the PVT task requires sustained attention without distraction.

Sleep and Chronotype Information

Participants completed a daily sleep journal in the form of a survey that included questions about bed time, minutes to fall asleep, number of wakeups during the night, wake time, and total sleep duration. Participants were reminded everyday at 10:30AM through a phone notification to complete this survey. Prior research validates the reliability of self-report journaling for in-situ nightly sleep measurement [50].

We used collected sleep information to assess individual chronotype. Since the comparison of chronotypes requires a single reference point, we use the well-established mid-sleep point on free days (MSF) — the halfway point between going to sleep and waking up — as the marker for individual chronotype [55, 56]. Because a majority of the population compensates for sleep debt accumulated on work days by sleeping longer on free days, this "oversleep" on free days is taken into consideration using a corrected measure of mid-sleep (MSF_{SC}): $MSF_{SC} = MSF - 0.5\,(SD_F - (5 * SD_W + 2 * SD_F)/7)$ [56], where SD_F and SD_W are sleep duration on free days and work days, respectively. $(5 * SD_W + 2 * SD_F)/7$ represents the averaged sleep duration across the week.

The mean and distribution of chronotypes in our dataset is consistent with prior related work on the identical age group [2]. Given that our young sample is predominantly late types as expected, we follow prior work [25] and consider anyone with $MSF_{SC} <$ 5:00AM as "early" and those with later mid-sleep points as "late". We also administered the Morningness-Eveningness Questionnaire (MEQ) [27] to confirm these classifications as early or late.

Given that chronotype influences our circadian rhythms in alertness performance, prior studies have introduced the concept of internal time [56]. We can think of internal time as a corrected measure of time that factors in an individual's chronotype. That is, while "external time" (ExT) represents the hours elapsed since midnight (12:00 AM), "internal time" (InT) is calculated using an individual's *biological* midnight (MSF_{SC}) [62]: $InT = ExT - MSF_{SC}$. We later present analyses based on internal time.

Phone Instrumentation

Our sensing framework incorporated phone probing that runs in the background to collect usage data. In this study, we focus on screen on/off events. We pre-processed phone probes by filtering out any data collected between periods when a participant began and completed EMAs to avoid overestimating a user's phone usage, when such interactions were actually due to conditions caused by participation in the study.

Compensation was based on the duration of participation ($5 for each week), sleep journal completion ($0.50 for each entry), and the number of completed EMA assessments ($0.20 for each entry). The Cornell University Institutional Review Board approved all procedures.

RESULTS

Over our 40 day study, our 20 participants provided an average of 2.52 (sd: 0.79) subjective alertness assessments, 2.46 (sd: 0.8) stimulant assessments, and 2.05 (sd: 0.87) objective alertness assessments per day. The average EMA compliance rate across all participants was 63.7%. Excluding the late night session (12:00AM - 5:59AM) when participants are likely to be asleep, the average compliance rate was 79.9%. Sleep diaries had a similar compliance rate of 72.8%.

While a number of studies have looked into cognitive performance and chronotype, ours is the first to use real world data collected over an extended period of time. This provides us an opportunity to compare with and extend previous laboratory based findings. Such replication is important for the scientific process and also validates the reliability and generalizability of our methodology based on data captured in the wild. Moreover, these steps help to introduce key chronobiological findings about cognitive performance to the UbiComp community, which can enrich the realm of ubiquitous systems aimed at enhancing our cognitive abilities.

We therefore first focus on comparing our results with previous laboratory-based studies. Then building on those findings, we later move on to presenting our novel methods for unobtrusive, continuous alertness assessment and prediction.

Replicating and Extending Extant Lab-Based Findings

Influence of Chronotype and Time-of-day on Alertness

We begin by looking into how alertness levels change across the local time of day. As a reminder, we operationalize alertness as relative response time (RRT) based on PVT data. Following prior work (which also facilitates comparison with it), we aggregate all participants and bin RRT values computed across sessions at 2-hour intervals [46, 62]. We find that RRT varies as a function of local time as shown in Figure 1.

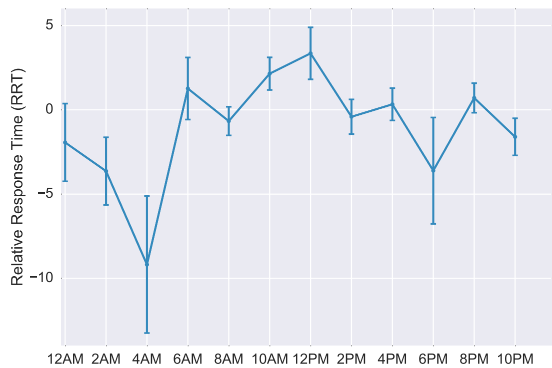

Figure 1. Relative response time (RRT) with standard error of mean (SEM) over the day. Positive and negative values of RRT indicate alertness higher and lower than individual baseline, respectively.

First, we see a local, short-term afternoon dip around 2:00PM, which is consistent with a well-known postprandial (i.e. after meal time) dip in core temperature and cognitive performance [13, 46]. We also see that RRT dips around 4:00AM and 6:00PM. Similar dips in alertness in relation to local time have been observed through controlled studies [61, 62], though the dips in our dataset happen later in the day, likely because our participants are later types ($MSF_{SC} = 05:56 \pm 0.94$ hrs) than those from past studies (e.g., $MSF_{SC} = 05:19 \pm 1.75$ hrs [62]). We further see that RRT improves during the late evening and night, as expected again since late types are the majority in our sample.

Next, we examine how RRT patterns differ between early and late types. Specifically, we calculate the difference in median RRT (using mean RRT does result in similar findings) between the groups over any given time period and normalize it by average daily change in RRT. As shown in Figure 2, this comparison between chronotypes reveals a striking difference between the groups' alertness levels depending on the local time of day. We find that early types are more alert in the morning (17% higher RRT compared to late types), and their alertness worsens in the later phase of the day (15% lower RRT compared to late types). Performing an ANOVA test on response time between early and late types across these time periods shows a significant difference with F(5, 964) = 5.36, $p < 0.001$. The fact that these results align with prior findings that early types are more alert in the morning while late types reach optimal performance later in the day [27, 61] helps validate the reliability of our alertness assessment.

While cognitive performance literature and UbiComp research have predominantly focused on alertness variations over local ("external") time, recent studies have shown that alertness can also be affected by our biological, "internal" time [62]. (Formula for internal time provided in the Method section). Figure 3 shows the variation in RRT for our sample according to internal time. Similar to prior research [62], we see RRT improves during early afternoon with a local peak around biological 2:00PM (i.e., 14 hours since biological midnight, MSF_{SC}). We notice the dip in alertness, however, occurs much later than in prior work, at around 12:00PM likely because our participants have later chronotypes.

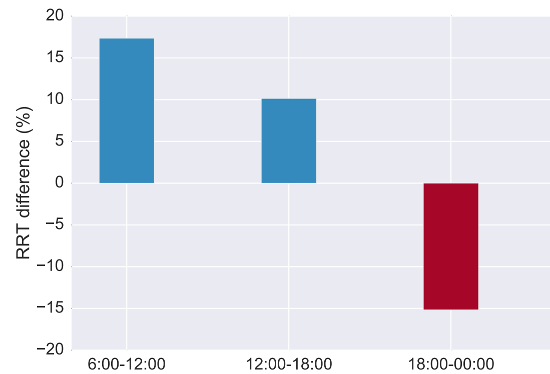

Figure 2. RRT of early chronotypes compared to late chronotypes across the day. Blue and red indicate higher RRT for early and late types, respectively. In the morning, early chronotypes display much higher alertness than late types, while the opposite is observed later in the day. Response time difference between early and late types across the day is also statistically significant: F(5, 964) = 5.36, $p < 0.001$.

Daylight Savings Time (DST) Impacts Alertness

More than 70 countries practice Daylight Savings Time (DST), which impacts around 1.6 billion people worldwide [30]. This twice-yearly 1-hour forward or backward clock shift is a *social clock* change, rather than an internal one. Similar to jet lag experienced after travel, DST can therefore significantly impact physiological and behavioral functioning as a result of the circadian disruption it produces [30]. While debates around DST have mostly focused on economic advantages [34], a number of studies have pointed out the resultant increased risk of driving accidents [59], deterioration in academic performance [21], and fragmented rest-activity cycles [35]. These issues can potentially be attributed to difficulties our internal circadian clocks experience in adjusting to the sudden social clock change. Because such disruptions can also impact cognitive abilities, we compare alertness before and after DST for early and late types in our sample.

While past research has shown that DST can result in higher workplace injury [6] and lower productivity due to cyberloafing [63], *our study is the first to look into the impact of DST on alertness using PVT data*, which allows us to investigate the *cause* rather than the *symptoms*. The Spring DST transition can affect circadian stability for days [30], an impact

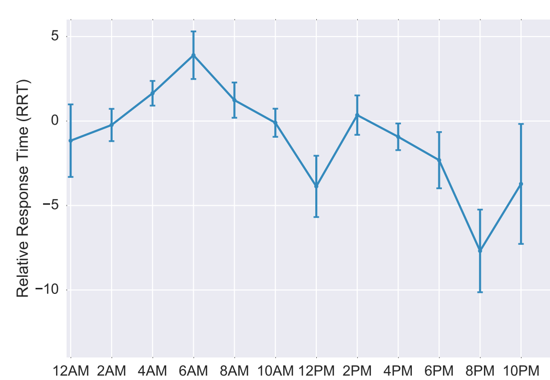

Figure 3. RRT with standard error of mean (SEM) with internal time on X-axis.

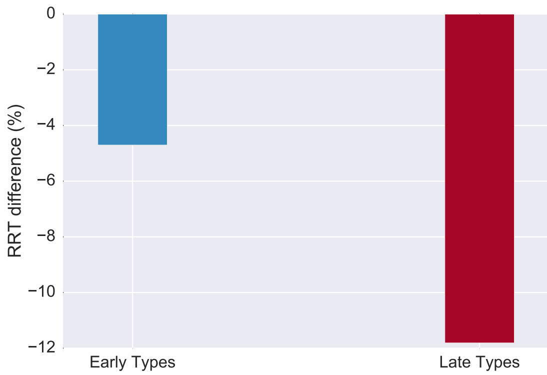

Figure 4. RRT change for early and late chronotypes before and after Spring DST. While the DST transition negatively affects both early and late types, late types suffer more. The difference in RRT before and after DST for both early ($t = -2.37, p = 0.02$) and late ($t = -2.52, p = 0.01$) types is statistically significant.

we believe our data can reflect since it spans 23 days before and 17 days after DST. To examine the effects of DST, we calculate the difference in median RRT (using mean RRT does result in similar findings) in the morning sessions before and after DST and normalize it by average daily change in RRT. As shown in Figure 4, RRT drops after DST for both early and late types (-4.70% and -11.8%, respectively). This difference in RRT before and after DST for both early ($t = -2.37, p = 0.02$) and late ($t = -2.52, p = 0.01$) types is statistically significant. Thus the overall alertness of both types degrades after Spring DST — though late types' alertness suffers significantly more than early types. Future study would do well to investigate whether alertness is similarly affected after the Fall DST transition.

Sleep Affects Alertness

Well known are the positive and negative impacts of adequate and deprived sleep, respectively, on numerous cognitive domains — though alertness is one of the earliest and most severely degraded by sleep deprivation [22]. We therefore next investigate the effect of our participants' sleep patterns on their alertness throughout the day. Given that the required duration of sleep varies across individuals [62], we compute sleep duration relative to individual sleep need. Sleep need is calculated as average sleep duration across work days (SD_W) and free days (SD_F) [62]: Sleep need $=$ $(SD_W * 5 + SD_F * 2)/7$

Previous research on shift-workers has found that reduced sleep duration negatively impacts alertness during morning shifts [62]. In our dataset, however, we do not find any such statistically significant difference in morning RRT between adequate (relative sleep need ≥ 1.0) and inadequate (relative sleep need < 1.0) sleep (t-score $= -1.45, p = 0.14$). We believe this contrasting result is explained by the fact that our sample consists of significantly later chronotypes than the studied shift-workers, which means our participants are more likely to exhibit low alertness levels in the morning regardless of their sleep duration the night before.

Going beyond prior studies, we also consider how sleep duration on nights following the daylight savings transition factors into alertness. We find that on nights after Spring DST, obtaining adequate sleep does improve RRT during the following day. Specifically, median RRT after nights with inadequate sleep falls by 17.49% (t-score $= -2.11, p = 0.03$). We find no similar relation that is statistically significant before the DST transition. This finding suggests that sleep deprivation has a more negative impact after Spring DST, when one is experiencing an already disrupted circadian system.

Stimulant Use Impacts Alertness

Stimulants can improve or reduce short-term alertness and cognitive performance. Caffeine has been shown to momentarily enhance alertness and counteract the negative effects of sleep loss [64], nicotine intake can improve alertness [23], and napping has been suggested as a preventive countermeasure against performance deterioration during long periods of wakefulness (at least for shift-workers, including those in the aviation industry [12]). On the other hand, alcohol is known to disrupt cognitive abilities including alertness [26], and a heavy meal can exacerbate daytime sleepiness and negatively impact cognition [54]. However, most studies about how stimulants impact cognition focus on shift-workers and are conducted in lab environments with artificial conditions, which means the results of such work might not be applicable to more general populations or everyday contexts.

We therefore aim to examine the impact of stimulants on alertness in a real-world setup. As a reminder, part of our EMA framework asked about activities done in the past hour that are known to improve alertness (i.e., positive stimulants) or diminish alertness (i.e., negative stimulants) [12, 23, 26, 54, 64], as shown in Table 1. Note that though loafing may be considered taking a "break", it is actually a negative stimulant since it represents activities that result in cognitive depletion rather than restoration (e.g. playing video games that require sustained attention).

Positive Stimulants	Negative Stimulants
Caffeine Consumption, Exercising, Napping, Nicotine Intake	Alcohol Consumption, Food Intake, Loafing

Table 1. Groupings of stimulants based on how they impact alertness.

We find that RRT increases by 5.08% across all participants after the use of positive stimulants, while we observe the opposite effect following the use of negative stimulants, which see a -1.37% drop in RRT. The difference in RRT immediately after taking positive vs. negative stimulants is statistically significant as well ($t = 2.21, p = 0.03$), indicating that stimulants can lead to short-term changes in alertness, which is consistent with the aforementioned lab based findings.

Comparing Self-Assessed to Objectively-Measured Alertness

In some settings, it is infeasible to perform objective alertness assessment using instruments like the PVT, making self-assessment the only practical option (e.g., during long haul driving, where there is a high risk of fatigue-related accidents due to deteriorated alertness). Researchers are therefore interested in determining the reliability of self-assessed alertness and how it relates to objective measurements.

As mentioned, we asked participants before each PVT test to rate their energy levels, concentration, and tiredness on 5-point scales. Participants' scores are highly-correlated with each other, supporting the internal validity of responses. To compare these subjective reports with our objective PVT measurements, we first group data based on high and low subjective scores using median as a threshold. Following prior work, we then compare the subjective high and low scores with RRT [19]. As shown in Table 2, we find that RRT differs significantly between periods with high and low self-assessed ratings, indicating that participants have a good awareness of their reduced cognitive capability, at least on a coarse level. This reliability of self-assessment is similar to findings reported by past studies that compared subjective and objective alertness for shift-workers [5, 19] — though it is worth noting that fatigue might impair a more fine-grained subjective assessment (i.e., "*how* tired are you" rather than "*are* you tired"), as found in prior work in the context of response time and road accidents [51].

Self-Assessed Variable	RRT difference between self-assessed alertness states
Energy	$t = 2.06, p = 0.04$
Concentration	$t = 2.76, p = 0.005$
Tiredness	$t = -2.1, p = 0.03$

Table 2. RRT differs significantly between self-assessed high and low alertness, indicating fatigued individuals are usually aware of reduced capability.

Novel Contributions: Passive Alertness Assessment

These results based on real world data confirm and augment previous laboratory based findings. Building on these results, we next focus on developing a novel means of continuously, automatically, and unobtrusively assessing alertness that can be deployed at scale.

Phone Usage Reflects Alertness

So far, we have explored how various internal and external factors (e.g., chronotype, internal time, external time, DST, and sleep) can influence an individual's alertness throughout the day and, importantly, how an awareness of biological rhythms can provide deeper interpretations of these observations. Given that our overarching goal is the development of methods for capturing and modeling these fluctuations in alertness, we next delve into whether technology-mediated behavioral traces can serve to *reflect* alertness patterns. While recent work has looked into the relationship between usage patterns and *self-reported* attention and boredom [41, 52], ours is the first study to consider objective alertness based on PVT — a clinically-validated tool.

We focus on usage patterns from smartphone interactions, which have been shown to reflect various aspects of our daily life related to both cognitive performance [41, 43] and circadian rhythms [2]. Moreover, individuals from this age group are the heaviest users of mobile technologies [53, 58]. We believe our sample's usage behaviors are representative of this population of interest since our participants' overall amount of phone use aligns well with that observed in prior work [20].

Based on related research, we hypothesize that more frequent, prolonged usage may signal a decreased ability to focus for an extended period of time [41, 47, 52] and so define two metrics to represent such aspects of engagement: i) *burstiness*, the total number of usage sessions in a given hour, where each session is marked by unlocking the phone and ii) total *duration* of phone usage in a given hour.

To begin, we systematically compare how usage patterns based on these metrics differ between periods with high (≥ 0) and low (< 0) RRT. The distribution of high and low groups is balanced with 56.98% and 43.02% of instances, respectively. We find that mean burstiness during high RRT periods is greater (9.07 ± 0.45) than periods with low RRT (7.9 ± 0.30), and the difference is also statistically significant (t-score $= 2.14, p = 0.03$). In other words, our participants initiated more sessions (i.e., unlocked their phones more) during periods when they were more alert. In contrast, the duration of phone usage per burst shows an opposite trend, with a mean duration of 116.37 ± 5.5 seconds/burst during high RRT periods compared to 123.47 ± 6.62 seconds/burst during low RRT. That is, when more alert, participants checked their phones frequently but for shorter lengths of time, while during low alertness, participants engaged in more sustained use. To further assess relations between RRT and phone use behaviors, we next compare short usage sessions (sessions less than 30 seconds, based on prior work [20]) between different alertness levels. From our dataset, we find that the mean number of short sessions during high RRT is 20% more than periods with low RRT (t-score $= 1.97$, $p < 0.05$).

Overall, these findings suggest that during periods of high vs. low alertness, phone use behaviors vary considerably and so can potentially be leveraged for passive alertness assessment.

Predicting Alertness

While response time tests like the PVT have been extensively used in lab studies, they are difficult to deploy in real world settings for extended periods due to the burdensome 2–10 minute time commitment necessary for meaningful results. A more unobtrusive, subtle, and contextually-embedded way to measure alertness is thus necessary to enable a new suite of UbiComp applications and systems that can accommodate variations in cognitive ability throughout the day and across circumstances. A main focus of this study is therefore exploring whether alertness can be predicted automatically using behavioral and contextual information — specifically, local time, internal time, sleep duration, relative sleep need, stimulant intake, subjective-assessment scores (energy, concentration, tiredness), and phone usage patterns (burstiness, duration, mean time between consecutive sessions, and short sessions under 30 seconds) as features. These features were selected based on our previously presented findings about phone usage patterns that are reflective of alertness variations along with behavioral and biological cues that extant literature suggests as relevant to cognitive performance.

Sleep and cognitive performance research has used response time in a number of different ways beyond RRT to assess cognitive ability. The ability to estimate response time is thus valuable to a large community and widely applicable, so we begin by attempting to predict response time from PVT data. We use Stochastic Gradient Descent (SGD) with the Huber loss function [28] for estimating response time. We use SGD because of its fast convergence speed and scalability for large scale learning, and we choose the Huber loss function since it is robust against outlier data points. SGD is known to be sensitive to feature scaling, so we standardize all features to have zero mean and unit variance. We also randomly shuffle the training data after each epoch, which prior work suggests can improve performance and convergence speed [10]. We randomly select 10% of the training data to choose model parameters (e.g., hyper-parameter α). The best performing model uses L1 norm as the regularization (penalty) term with hyper-parameter $\alpha = 10^{-7}$ and learning rate set to $\gamma_t = (\alpha \cdot t)^{-1}$ [10]. Using 10-fold cross validation, a generalized model for all participants results in a root mean square error (RMSE) of 83.81 milliseconds (ms). We also train personalized models using data from each participant; as expected, individual models further improve accuracy — average RMSE across all participants from 10-fold cross validation is 80.64ms.

The performance of our developed model using real-world data is very encouraging. To contextualize the accuracy of the models: in sleep loss and cognitive performance studies, a response time higher than 500ms is often considered as a "*lapse*" — a standard measure of impaired cognitive ability [7, 40]. The average RMSE of 80.64ms achieved by our developed models — a much higher granularity than the standard definition of lapse — indicates that these models could reliably be deployed instead of using PVT, especially to reduce burden for scenarios lasting extended periods of time.

Given these promising results in modeling response time, we next focus on modeling RRT. Since one of our key goals is estimating alertness unobtrusively, we now use a reduced set of features that can all be captured passively using smartphone sensors — specifically, local time, internal time, sleep duration, relative sleep need, phone usage burstiness, mean duration of phone usage sessions, average time between successive phone usage sessions, and frequency of short (less than 30 seconds) phone usage sessions. (Note that though we use self-reported sleep diaries in this study, it is possible to instead reliably perform passive sleep assessment using smartphone data [2, 44], which would enable the calculation of information such as chronotype, internal time, and relative sleep need without active user input). We again use SGD with Huber loss function, standardize features to zero mean and unit variance to avoid scaling issues. We randomly select 10% of the training data to choose model parameters (e.g., hyper-parameter α). The best performing model uses L1 norm as the regularization term with $\alpha = 10^{-8}$ and learning rate set to $\gamma_t = \gamma_0 \cdot t^{-\frac{1}{4}}$ where $\gamma_0 = 0.01$ is the initial learning rate. Using 10-fold cross validation, our generalized model achieves RMSE of 11.39% across all participants. Training individual models again produces further improvement and an average RMSE of 10.87% across all participants. That is, we can detect as small as an 11% deviation from individual baseline; given that daily peak-to-peak change in RRT averaged over all participants is 29.4%, this reaffirms the feasibility of using our models in place of PVT tests.

To evaluate the importance of each feature in modeling both response time and RRT, we perform feature ranking with recursive feature elimination (RFE) [24]. At each step of RFE, a model is trained on the entire dataset and the feature that contributes least to the model (as measured by the absolute weight assigned to each feature) is discarded. This procedure continues recursively until there is only one feature left. Table 3 shows our results from this process. For response time modeling, the highest ranked features (i.e., self-perceived energy, internal time, stimulant use) have also been identified by prior research as well-established modulators of alertness [19, 57, 62]. The ranking of the reduced feature set used for RRT modeling demonstrates the importance of taking individual chronotype into consideration as well as the informativeness of phone usage patterns, which is consistent with our earlier findings that phone checking and engagement behaviors exhibit distinct trends during periods of high and low alertness.

Overall, we think the performance of our models speaks to the feasibility of deploying our tools in real-world settings — and in turn, opening new UbiComp possibilities, from sensing to intervention in a range of domains and application areas.

Rank	Response Time	RRT
1	Energy rating	Internal time
2	Internal time	Avg. time between phone usage sessions
3	Stimulant intake	Short session frequency
4	Avg. time between phone usage sessions	Phone usage duration
5	Concentration rating	Relative sleep need

Table 3. Top ranking feature groups for modeling response time and RRT based on relative feature elimination (RFE).

DISCUSSION

Cognitive performance — particularly alertness — is known to vary significantly across the day as a result of multiple individual factors including chronotype, a sleep pressure that mounts as the day progresses, and social obligations. While previous studies have looked into temporal alertness trends, most were conducted in controlled laboratory setups or did not take biological factors into consideration. Our study is the first to use alertness data captured in ecologically valid settings over an extended period of time in order to develop passive sensing techniques grounded in chronobiology.

We first focused on replicating and extending the findings of previous lab-based studies both in order to push forward scientific knowledge as well as to validate our employed methodology and its generalizability. Furthermore, since the biological factors responsible for the fluctuations in our cognitive ability have mostly been unexplored by UbiComp researchers, replication also serves as a way to acquaint the community with chronobiological findings that are pertinent and valuable to the field. Our findings were mostly consistent with prior studies. We found that alertness drops in the early morning, peaks around noon and again in the evening, and we saw clear evidence of the well known mid-day dip. Comparing alertness across chronotypes, we confirmed that early types are more alert earlier in the day, while late types peak during evening. We further extended previous findings about the impact of the Spring Daylight Savings Time (DST) transition — a socially enforced change in time that can produce circadian disruption — and found that DST resulted in declined performance, with late types affected more severely.

Beyond extending previous work, we also set out to develop models capable of detecting and predicting temporal trends in alertness, which is a keystone of the attentional system and has considerable influence on a wide range of other cognitive functions [57]. Capable of assessing alertness in the wild, our models for both response time and RRT achieved high accuracy. Comparing to PVT tests, which can take as long as 2–10 minutes to complete, our passive-prediction approach is low burden and low cost, and it enables the real time measurement of performance in real world settings.

Sensing and Design Implications

Increasingly, technology is improving through personalization and context-awareness, but most systems have yet to recognize and support the timings of our internal body clocks. The alertness modeling techniques we have provided here could enhance context-aware computing systems as well as those from the emerging area of *circadian computing*: technologies that are aware of our internal timings and can play to our biological strengths [2].

Systems with this sort of empirical foundation may have a better chance of success since they can target the biological roots of a problem rather than just the symptoms. For instance, tools designed to limit cyberloafing behaviors might first help people identify and address other aspects of their personal behavior or environment that are disruptive to their circadian rhythms (e.g., caffeine use or trying to do cognitively demanding tasks at biologically ill-suited times). Our research lays the groundwork for the development of such circadian-aware systems that can help optimize performance, encourage restorative activities during alertness dips, or simply help us become more aware and accepting of our biological capabilities (and limitations) — ultimately leading to healthier, safer, and more sustainable working schedules.

Applications in Education and Scheduling

Given the connections among alertness, learning, and memory [4], the techniques we have introduced here for passive, in-situ alertness assessment offer substantial positive impacts for educational settings. Various studies have demonstrated that the learning schedules for college students [16] and high school students [32] (who both tend to have later chronotypes) run contrary to their attentional rhythms, resulting in increased fatigue and negative academic outcomes.

On an institutional level, our assessment methods provide a way to capture large-scale, ecologically valid data about students' cognitive performance rhythms, which could be used to advocate for more widespread reform in educational scheduling. At the class-level, this type of data could inform the creation of lecture schedules that account for collective group chronotype or support instructors in class planning (e.g., when choosing members for a group project). On an individual level, alertness assessment tools could help students identify personal productivity patterns in order to make more informed decisions when arranging class and study schedules that better align with their own cognitive rhythms.

More generally, the ability to predict alertness can enable novel forms of event scheduling (for both groups and individuals) that reflect the dynamic and oscillating nature of cognitive performance. For instance, a calendar system could intelligently suggest meeting slots based on the alertness profiles of participants. Also, given the rise of distributed workplaces, such systems could also facilitate team management by pairing collaborators who are better synchronized in terms of sensed alertness patterns. Similarly, a circadian-aware activity recommender system might suggest the best times of day for an individual to do either more or less cognitively-demanding tasks. Such recommendations would be based on one's chronotype, idiosyncratic sleep-related behaviors, current time-of-day, and overall sensed performance rhythms.

Applications in Accident Prevention

The ability to perform real-time, in-situ alertness detection could be particularly valuable for improving road safety. Driver fatigue is a major cause of road accidents: the National Transportation Safety Board (NTSB) estimates that 30% of all road accident fatalities in the US involve fatigue [49]. While a number of systems have attempted to predict driver fatigue using a variety of sensors (e.g., computer vision [29]), these systems tend to be costly and are considered more invasive. Our unobtrusive and passive methods for alertness sensing could complement increasingly ubiquitous driver assistance systems (e.g., that provide lane departure and forward collision warnings) in order to further increase driver and passenger safety — and crucially, potentially take effect before the driver even steps into the car.

While sensing based on smartphone usage *during* driving would be inapplicable, phone use leading up to entering the car could be used to infer alertness while driving, or a person's historic alertness patterns could be used to determine expected alertness during the current driving period. Further, the phone-based signals we focused on in this paper are only the first steps, and the assessment framework could be extended to unobtrusively mine additional technology-mediated behavioral cues from within the driving context (e.g., steering wheel interactions or radio tuning patterns).

Limitations and Future Work

First, it would be valuable to extend our study to samples of more diverse ages and occupations, particularly since our population of interest has a greater proportion of late chronotypes than is found in other groups. Broadening our focus to older populations would be particularly worthwhile given that a strong time-of-day effect has been observed in this population as well [57]. That said, we believe that a population such as ours — with known conflicts between obligations and biologically-preferred sleep-wake schedules — is an appropriate starting point. Relatedly, we apply a useful but coarse categorization of chronotypes into two groups; since most participants in our study are late types, such a categorization makes sense, but a larger study with more diverse chronotypes would enable more sophisticated analyses that treat chronotype as a continuous variable.

Along with broadening our attention to other populations, expanding our work to other aspects of cognitive performance would also be valuable. While we have focused on daily variations in alertness, rhythmic changes also occur in higher order cognitive functions like working memory and executive control [57]. A logical next step would be to deploy similar momentary assessment instruments in order to measure such additional aspects of cognitive performance.

Lastly, prior studies on smartphone use have considered the impact of context on usage. However, much work is left to be done to explore how a dynamic model of alertness together with an understanding of biological rhythms can be used to help *explain* smartphone habits and observed patterns of use. We have investigated how a straightforward set of phone usage metrics can reflect alertness, and related research has recently undertaken a more in-depth analysis and interpretation of smartphone application use trends in relation to different alertness stages [48]. Moving forward, given that people's behaviors related to work, entertainment, and social networking span multiple devices, a broader inspection of overall technology usage including computers, tablets, and TVs might provide additional insights or boost prediction performance.

CONCLUSION

Our cognitive performance varies considerably and predictably over the course of a day. Beyond behavioral, environmental, and social factors, our internal circadian rhythms play a particularly significant role in influencing our alertness on a moment-to-moment basis. However, previous studies that have focused on understanding patterns of alertness across the day either do not consider such key biological factors or are mostly done in controlled laboratory environments.

This paper presents the first chronobiology-informed study to collect real-world alertness data over an extended period of time. Expanding the findings of previous work, our data confirm that alertness fluctuates significantly over the course of the day and that these fluctuations differ between early types and late types. We also show that Daylight Savings Time (DST) has a negative impact on alertness, with late types suffering more; that alertness drops after a night of inadequate sleep following DST; and that stimulants can lead to short-term changes in alertness. Based on these results, we develop models that can accurately, continuously, and unobtrusively assess alertness.

Our contributions lay the foundation for a new class of circadian-aware systems — technology that can model, adapt to, and provide feedback about the dynamic cognitive variations experienced by individuals and groups, in order to support better collaboration, stabilize cognitive performance, and improve overall well-being.

Acknowledgements

This work was supported by a grant from the Robert Wood Johnson Foundation and the Health Data Exploration Project, the Intel Science & Technology Center for Pervasive Computing, Semiconductor Research Corporation, and the National Science Foundation (DGE-1144153 and IIS-1344613).

REFERENCES

1. Khaled Abdel-Kader, Manisha Jhamb, Lee Anne Mandich, Jonathan Yabes, Robert M Keene, Scott Beach, Daniel J Buysse, and Mark L Unruh. 2014. Ecological momentary assessment of fatigue, sleepiness, and exhaustion in ESKD. *BMC nephrology* 15 (2014), 1.
2. Saeed Abdullah, Mark Matthews, Elizabeth L Murnane, Geri Gay, and Tanzeem Choudhury. 2014. Towards circadian computing: early to bed and early to rise makes some of us unhealthy and sleep deprived. In *Proc. of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 673–684.
3. Bruce M Altevogt, Harvey R Colten, and others. 2006. *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem*. National Academies Press.
4. Gary Aston-Jones. 2005. Brain structures and receptors involved in alertness. *Sleep medicine* 6 (2005), S3–S7.
5. Joseph V Baranski, Ross A Pigeau, and Robert G Angus. 1994. On the ability to self-monitor cognitive performance during sleep deprivation: a calibration study. *Journal of Sleep Research* 3, 1 (1994), 36–44.
6. Christopher M Barnes and David T Wagner. 2009. Changing to daylight saving time cuts into sleep and increases workplace injuries. *Journal of applied psychology* 94, 5 (2009), 1305.
7. Mathias Basner, Daniel Mollicone, and David F Dinges. 2011. Validity and sensitivity of a brief psychomotor vigilance test (PVT-B) to total and partial sleep deprivation. *Acta astronautica* 69, 11 (2011), 949–959.
8. Mathias Basner and Joshua Rubinstein. 2011. Fitness for duty: A 3 minute version of the Psychomotor Vigilance Test predicts fatigue related declines in luggage screening performance. *Journal of Occupational and Environmental Medicine* 53, 10 (2011), 1146.
9. Frank Bentley, Konrad Tollmar, Peter Stephenson, Laura Levy, Brian Jones, Scott Robertson, Ed Price, Richard Catrambone, and Jeff Wilson. 2013. Health Mashups: Presenting statistical patterns between wellbeing data and context in natural language to promote behavior change. *ACM Transactions on Computer-Human Interaction (TOCHI)* 20, 5 (2013), 30.
10. Léon Bottou. 2012. Stochastic gradient descent tricks. *Neural Networks: Tricks of the Trade* (2012), 421–436.
11. Walter C Buboltz Jr, Franklin Brown, and Barlow Soper. 2001. Sleep habits and patterns of college students: a preliminary study. *Journal of American College Health* 50, 3 (2001), 131–135.
12. John A Caldwell, Melissa M Mallis, J Lynn Caldwell, Michel A Paul, James C Miller, and David F Neri. 2009. Fatigue countermeasures in aviation. *Aviation, Space, and Environmental Medicine* 80, 1 (2009), 29–59.

13. Julie Carrier and Timothy H Monk. 2000. Circadian rhythms of performance: new trends. *Chronobiology international* 17, 6 (2000), 719–732.

14. Trudie Chalder, G Berelowitz, Teresa Pawlikowska, Louise Watts, S Wessely, D Wright, and EP Wallace. 1993. Development of a fatigue scale. *Journal of psychosomatic research* 37, 2 (1993), 147–153.

15. Nancy L Digdon and Andrew J Howell. 2008. College students who have an eveningness preference report lower self-control and greater procrastination. *Chronobiology international* 25, 6 (2008), 1029–1046.

16. Angela K Dills and Rey Hernandez-Julian. 2008. Course scheduling and academic performance. *Economics of Education Review* 27, 6 (2008), 646–654.

17. David F Dinges. 1995. An overview of sleepiness and accidents. *Journal of sleep research* 4, s2 (1995), 4–14.

18. David F Dinges and John W Powell. 1985. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments, & Computers* 17, 6 (1985), 652–655.

19. Jill Dorrian, Nicole Lamond, and Drew Dawson. 2000. The ability to self-monitor performance when fatigued. *Journal of Sleep Research* 9, 2 (2000), 137–144.

20. Denzil Ferreira, Jorge Goncalves, Vassilis Kostakos, Louise Barkhuus, and Anind K Dey. 2014. Contextual experience sampling of mobile application micro-usage. In *Proc. of the 16th international conference on Human-computer interaction with mobile devices & services*. ACM, 91–100.

21. John F Gaski and Jeff Sagarin. 2011. Detrimental effects of daylight-saving time on SAT scores. *Journal of Neuroscience, Psychology, & Economics* 4, 1 (2011), 44.

22. Namni Goel, Mathias Basner, Hengyi Rao, and David F Dinges. 2013. Circadian rhythms, sleep deprivation, and human performance. *Progress in molecular biology and translational science* 119 (2013), 155.

23. William Griesar, Daniel Zajdel, and Barry Oken. 2002. Nicotine effects on alertness and spatial attention in non-smokers. *Nicotine & tobacco research* 4, 2 (2002), 185–194.

24. Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. 2002. Gene selection for cancer classification using support vector machines. *Machine learning* 46, 1-3 (2002), 389–422.

25. Lynn Hasher, David Goldstein, and Cynthia May. 2005. It's About Time: Circadian Rhythms, Memory, and Aging. (2005).

26. Ian Hindmarch, JS Kerr, and N Sherwood. 1991. The effects of alcohol and other drugs on psychomotor performance and cognitive function. *Alcohol and Alcoholism* 26, 1 (1991), 71–79.

27. Jim A Horne and Olov Ostberg. 1975. A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International journal of chronobiology* 4, 2 (1975), 97–110.

28. Peter J Huber and others. 1964. Robust estimation of a location parameter. *The Annals of Mathematical Statistics* 35, 1 (1964), 73–101.

29. Qiang Ji and Xiaojie Yang. 2002. Real-time eye, gaze, and face pose tracking for monitoring driver vigilance. *Real-Time Imaging* 8, 5 (2002), 357–377.

30. Thomas Kantermann, Myriam Juda, Martha Merrow, and Till Roenneberg. 2007. The human circadian clock's seasonal adjustment is disrupted by daylight saving time. *Current Biology* 17, 22 (2007), 1996–2000.

31. Matthew Kay, Kyle Rector, Sunny Consolvo, Ben Greenstein, Jacob Wobbrock, Nathaniel Watson, and Julie Kientz. 2013. PVT-touch: Adapting a reaction time test for touchscreen devices. In *Pervasive Computing Technologies for Healthcare*. IEEE, 248–251.

32. Paul Kelley, Steven W Lockley, Russell G Foster, and Jonathan Kelley. 2014. Synchronizing education to adolescent biology:'let teens sleep, start school later'. *Learning, Media and Technology* (2014), 1–17.

33. N Kleitman, S Titelbaum, and P Feiveson. 1938. The effect of body temperature on reaction time. *American Journal of Physiology* 121, 2 (1938), 495–501.

34. Matthew J Kotchen and Laura E Grant. 2011. Does daylight saving time save energy? Evidence from a natural experiment in Indiana. *Review of Economics and Statistics* 93, 4 (2011), 1172–1185.

35. Tuuli A Lahti, Sami Leppämäki, Jouko Lönnqvist, and Timo Partonen. 2008. Transitions into and out of daylight saving time compromise sleep and the rest-activity cycles. *BMC physiology* 8, 1 (2008), 3.

36. Nicole Lamond and Drew Dawson. 1999. Quantifying the performance impairment associated with fatigue. *Journal of sleep research* 8, 4 (1999), 255–262.

37. Nicole Lamond, Sarah M Jay, Jillian Dorrian, Sally A Ferguson, Gregory D Roach, and Drew Dawson. 2008. The sensitivity of a palm-based psychomotor vigilance task to severe sleep loss. *Behavior Research Methods* 40, 1 (2008), 347–352.

38. Julian Lim and David F Dinges. 2008. Sleep deprivation and vigilant attention. *Annals of the New York Academy of Sciences* 1129, 1 (2008), 305–322.

39. Julian Lim and David F Dinges. 2010. A meta-analysis of the impact of short-term sleep deprivation on cognitive variables. *Psychological bulletin* 136, 3 (2010), 375.

40. Sylvia Loh, Nicole Lamond, Jill Dorrian, Gregory Roach, and Drew Dawson. 2004. The validity of psychomotor vigilance tasks of less than 10-minute duration. *Behavior Research Methods* 36, 2 (2004), 339–346.

41. Gloria Mark, Shamsi T Iqbal, Mary Czerwinski, and Paul Johns. 2014. Bored mondays and focused afternoons: The rhythm of attention and online activity in the workplace. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*. 3025–3034.

42. Robert L Matchock and J Toby Mordkoff. 2009. Chronotype and time-of-day influences on the alerting, orienting, and executive components of attention. *Experimental brain research* 192, 2 (2009), 189–198.

43. Aleksandar Matic, Martin Pielot, and Nuria Oliver. 2015. Boredom-computer interaction: Boredom proneness and the use of smartphone. In *Proc. of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 837–841.

44. Jun-Ki Min, Afsaneh Doryab, Jason Wiese, Shahriyar Amini, John Zimmerman, and Jason I Hong. 2014. Toss'n'turn: smartphone as sleep and sleep quality detector. In *Proc. of the ACM conference on Human factors in computing systems*. ACM, 477–486.

45. Timothy H Monk. 1989. A visual analogue scale technique to measure global vigor and affect. *Psychiatry research* 27, 1 (1989), 89–99.

46. Timothy H Monk, Daniel J Buysse, Charles F Reynolds, and David J Kupfer. 1996. Circadian determinants of the postlunch dip in performance. *Chronobiology international* 13, 2 (1996), 123–133.

47. Elizabeth L Murnane, Saeed Abdullah, Mark Matthews, Tanzeem Choudhury, and Geri Gay. 2015. Social (media) jet lag: How usage of social technology can modulate and reflect circadian rhythms. In *Proc. of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 843–854.

48. Elizabeth L Murnane, Saeed Abdullah, Mark Matthews, Matthew Kay, Julie A Kientz, Tanzeem Choudhury, Geri Gay, and Dan Cosley. 2016. Mobile Manifestations of Alertness: Connecting Biological Rhythms with Patterns of Smartphone App Use. In *Proc. of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM.

49. National Transportation Safety Board. 1999. Evaluation of U.S. Department of Transportation Efforts in the 1990s to address operator fatigue. Safety report NTSB/SR-99/01.

50. Sanjay R Patel, Najib T Ayas, Mark R Malhotra, David P White, Eva S Schernhammer, Frank E Speizer, Meir J Stampfer, and Frank B Hu. 2004. A prospective study of sleep duration and mortality risk in women. *SLEEP* 27, 3 (2004), 440–444.

51. Pierre Philip, Patricia Sagaspe, Jacques Taillard, Nicholas Moore, Christian Guilleminault, Montserrat Sanchez-Ortuno, Torbjörn Akerstedt, and Bernard Bioulac. 2003. Fatigue, sleep restriction, and performance in automobile drivers: a controlled study in a natural environment. *SLEEP* 26, 3 (2003), 277–284.

52. Martin Pielot, Tilman Dingler, Jose San Pedro, and Nuria Oliver. 2015. When attention is not scarce - detecting boredom from mobile phone usage. In *Proc. of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 825–836.

53. Jacob Poushter. 2016. Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies. PEW.

54. LA Reyner, SJ Wells, V Mortlock, and JA Horne. 2012. 'Post-lunch' sleepiness during prolonged, monotonous driving — Effects of meal size. *Physiology & behavior* 105, 4 (2012), 1088–1091.

55. Till Roenneberg, Tim Kuehnle, Peter P Pramstaller, Jan Ricken, Miriam Havel, Angelika Guth, and Martha Merrow. 2004. A marker for the end of adolescence. *Current Biology* 14, 24 (2004), R1038–R1039.

56. Till Roenneberg, Anna Wirz-Justice, and Martha Merrow. 2003. Life between clocks: daily temporal patterns of human chronotypes. *Journal of biological rhythms* 18, 1 (2003), 80–90.

57. Christina Schmidt, Fabienne Collette, Christian Cajochen, and Philippe Peigneux. 2007. A time to think: circadian rhythms in human cognition. *Cognitive Neuropsychology* 24, 7 (2007), 755–789.

58. Aaron Smith. 2015. The smartphone difference. PEW.

59. Neeraj Sood and Arkadipta Ghosh. 2007. The short and long run effects of daylight saving time on fatal automobile crashes. *The BE Journal of Economic Analysis & Policy* 7, 1 (2007).

60. Daniel J Taylor, Kendra C Clay, Adam D Bramoweth, Kevin Sethi, and Brandy M Roane. 2011. Circadian phase preference in college students: relationships with psychological functioning and academics. *Chronobiology international* 28, 6 (2011), 541–547.

61. HPA Van Dongen and David F Dinges. 2005. Circadian rhythms in sleepiness, alertness, and performance. In *Principles and practice of sleep medicine*. 4, 435–443.

62. Céline Vetter, Myriam Juda, and Till Roenneberg. 2012. The influence of internal time, time awake, and sleep duration on cognitive performance in shiftworkers. *Chronobiology International* 29, 8 (2012), 1127–1138.

63. David T Wagner, Christopher M Barnes, Vivien KG Lim, and D Lance Ferris. 2012. Lost sleep and cyberloafing: Evidence from the laboratory and a daylight saving time quasi-experiment. *Journal of Applied Psychology* 97, 5 (2012), 1068.

64. Nancy Wesensten, William Killgore, and Thomas Balkin. 2005. Performance and alertness effects of caffeine, dextroamphetamine, and modafinil during sleep deprivation. *Journal of sleep research* 14, 3 (2005), 255–266.