Validating a Mobile Phone Application for the Everyday, Unobtrusive, Objective Measurement of Sleep

Shaun Lawson[1], Sue Jamison-Powell[2], Andrew Garbett[3], Conor Linehan[1], Erica Kucharczyk[4], Sanne Verbaan[5], Duncan Rowland[1], Kevin Morgan[4]

[1] University of Lincoln, Lincoln, United Kingdom

[2] University of Leicester, Leicestershire, Leicester, United Kingdom

[3] Culture Lab, School of Computing Science, Newcastle University, UK

[4] Loughborough University, Loughborough, Leicestershire, United Kingdom

[5] The Hague University of Applied Sciences, The Hague, Netherlands

{slawson | clinehan | drowland }@lincoln.ac.uk, sjp66@le.ac.uk, a.garbett@newcastle.ac.uk, {E.Kucharczyk | K.Morgan }@lboro.ac.uk, scverbaan@gmail.com

ABSTRACT
There is an identified need for objective, reliable, and scalable methods of measuring and recording sleep. Such methods must be designed for easy integration into people's lives in order to support both sleep therapy and everyday personal informatics. This paper describes the design and evaluation of a mobile phone application to record sleep, the design of which has substantive foundation in clinical sleep research. Two user studies were carried out which demonstrate that the application produces valid measurements of sleep quality and high levels of usability, whilst not seriously disturbing sleep or the sleep environment. These findings suggest that the app is suitable for both everyday sleep monitoring in a personal informatics context, and for integration into sleep interventions.

Author Keywords
Sleep, insomnia, mobile computing, personal informatics.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
It is well reported (e.g. [3,10, 14, 37]) that even moderate disturbance of sleep can have severe detrimental effects on aspects of daily life– including our cognitive and decision making abilities, our relationships with family and friends and on our ability to function effectively at our place of work. Serious sleep disorders, such as insomnia, and chronic sleep loss can further lead to significant debilitating health conditions such as cardiovascular problems [10] and diabetes [38]. Additionally, chronic sleep problems are often comorbid in that they are linked, in complex ways to, and are often indicative of, other health conditions [28]. However, despite these complexities of the pathogenesis of sleep problems, there are well-established interventions that can improve sleep problems for most people. For instance Cognitive Behavioral Therapy (CBT) for insomnia (CBT-I) has been repeatedly shown to be effective in clinical trials (e.g. see [29]). Despite the existence of clinically-proven behavioral therapies, as well as the promotion of regular and effective sleep by public health organisations, problems associated with sleep are prevalent in society. For instance, it is estimated that insomnia affects 30% of the world's population at some point during their lifetime [28]. Sleep disturbance and its consequences remain an urgent research challenge not for only for health professionals but as a multi-disciplinary priority.

Researchers within the CHI, personal/health informatics, and persuasive computing communities are beginning to identify opportunities where interactive systems may play a positive, and perhaps transformative, role in addressing society's sleep problems. In perhaps the most substantial appraisal of this topic to-date, Choe et al [11] identify design opportunities for using computing to support healthy sleep behaviors and suggest that the creation of unobtrusive solutions that allow users to monitor their sleep schedules is an area of immediate and compelling opportunity. Such solutions would potentially have widespread and substantial uptake across all manner of sleep-related research and intervention projects, as well as by people in the general community. Indeed, existing solutions to measuring sleep either lie firmly in the domain of clinical sleep monitoring (namely actigraphy and polysomnography) and are obtrusive, expensive and inaccessible by everyday users, or, in the case of widely available web and phone-based apps,

CHI 2013, April 27–May 2, 2013, Paris, France.

are not validated to accurately record sleep in an objective and repeatable manner [11].

In this paper we address these issues and describe the design, validation and evaluation of a mobile phone application, running on the Android platform, that accurately, reliably and objectively records sleep behaviour in a manner that is unobtrusive, usable and acceptable for everyday use. The application – hereafter referred to as the *Sleepful app,* has substantive foundation in previous clinical sleep research. It uses a stimulus-response paradigm to record periods of sleep and wakefulness which can be used to characterize sleep behaviour (wake/sleep periods) and generate objective indexes of sleep quality, such as sleep efficiency (SE), that are widely used by the sleep research community. Such measurement and characterization of a person's sleep can be subsequently used, for instance, as an assessment tool and as an outcome measure in clinical research, as well as for everyday personal informatics.

We begin by discussing the background to the design of the Sleepful app in which we review previous relevant research and the commercial applications that are available for measuring sleep. We touch upon the challenges of designing 'for the bedroom' and for users who, for much of the time, are asleep, as well as discussing the recording of an aspect of behaviour that is potentially sensitive, intimate and private. We go on to describe the subsequent design and deployment of the Android app. We then describe a user evaluation (Study 1) designed to validate the app's ability to accurately measure SE in comparison to actigraphy, as well as to understand the usability and obtrusiveness of the interactive experience. We then describe a second user evaluation (Study 2), which used polysomnography to further determine the app's ability to measure sleep/wake states and to ascertain whether or not the app measurably interrupted sleep. We finally discuss the immediate opportunities for exploitation of our findings by the CHI, health and personal informatics and sleep research communities.

BACKGROUND

Behavioural Sleep Therapies

The simplest possible approach to sleep therapy is the instruction of what is termed *sleep hygiene*, which is essentially a set of good practices regarding behaviour in the bedroom and in the hours leading up to sleep [36]. For example, people are recommended not to work or watch television in the bedroom, not to drink coffee after 6pm and to take plenty of exercise. Often poor sleep patterns can be much improved through such simple changes in behaviour. Bauer et al [6] recently described an interactive application that reminded people – using subtle cues and reminders on the wallpaper of their smartphone – about sleep hygiene.

For more severe sleep disorders, such as chronic insomnia, clinical and review evidence has shown that successful outcomes can be achieved with CBT: a psychological approach that addresses styles of thinking and behavioural habits that maintain or exacerbate psychological symptoms. Through systematic, structured programmes of cognitive re-structuring and behaviour change, patients are helped to develop their own resources to combat dysfunctional practices. Controlled trials have demonstrated that CBT is indeed effective in the treatment of insomnia (e.g. [30]). CBT interventions are also particularly well suited to self-help style programmes that require minimal therapist input.

Sleep Measurement

The goal of sleep therapy is to improve sleep behavior. Therefore some objective means of recording such behaviour is inherently necessary to measure the efficacy of any sleep intervention, whether this is improved sleep hygiene or a course of CBT. Whereas residential clinical sleep labs can provide an environment suitable for objective monitoring of patients by researchers there is a clear need to measure and record people's sleep behaviour in natural settings i.e. in their own home.

The most important measurements of a person's sleep are the times they go to bed and get up, and an accurate characterization of the periods spent awake or asleep between these times. In particular, measurements such as the time taken to initially fall asleep (sleep onset latency, SOL), time awake over night after sleep onset (WASO) and Total Sleep Time (TST) are important metrics used by sleep researchers and clinicians. Once a repeatable means of gathering such behavioral sleep data is possible then an accurate assessment of the person's sleep efficiency (SE) can be made. SE is an overall measurement of a person's sleep quality and is simply ratio of the time spent asleep (TST) to the amount of time spent in bed (SOL + WASO + TST). SE is normally 85 to 90% or higher for people who do not suffer sleep problems [17]. The problem of accurately measuring such behaviour, however, is a challenging one.

There have been many proposed means of gathering sleep behaviour data. The simplest method is the use of paper based diaries which are completed manually by the sleep patient. Like any self-reporting system however, sleep diaries are prone to manual recording errors. For instance, Espie et al [16] found that patients overestimated subjective sleep onset latency and sleep durations in paper based sleep diaries. In fact, it is self-evident that manual reporting of sleep behaviour is more prone to error than other self-reporting of health behaviour (e.g. food consumption) due to the reduced state of awareness people usually find themselves in and around their times of sleeping.

The most common way of automatically measuring sleep behaviour when conducting sleep-related research is through actigraphy. This involves a person wearing an accelerometer-based device on their wrist. A number of commercial wrist-mounted actigraphy devices (or 'actiwatches') are available on the marketplace and the sleep research community has developed a substantial body of work focusing the validation of such devices (e.g. see [4, 35]). Actigraphy – in a general sense - records movement, the most commonly encountered example of actigraphy being pedometers. In sleep, actigraphy is used to measure the "restlessness" of an individual in bed. Although individuals do make many types of movements during sleep, actigraphy typically records movement above a threshold, which determines that the user is awake.

Further measurement of the *state* a person is in during sleep can also give deeper insight into their sleep behaviour. Polysomnography (PSG) uses encephalographic (EEG), electromyographic (EMG), and electrooculagraphic (EOG) methods to record electrical impulses in the brain that represent areas of activity that represent sleep and wake states. It is also used to measure individual states of sleep, from the initial onset of sleep to slow wave sleep (deep sleep from which an individual is less likely to wake). PSG is considered to be the gold standard in the measurement of sleep, however it requires specialist equipment and expert analysis. Hence, whilst this is ideal for sleep professionals, it is unsuitable for use by the general public.

Sleep and personal informatics

Recently the concept of personal informatics has emerged as a research area within the CHI community [26]. This reflects an emerging desire for people to gather personally relevant information for the purposes of self-knowledge, self-reflection and self-monitoring. As the authors in [26] emphasize, self-knowledge has many benefits, such as fostering insight, increasing self-control, and promoting positive behaviors; all of such benefits are potentially applicable to sleep. The large number of 'sleep apps' that are available online and on mobile phone marketplaces, is evidence that sleep is one area where many people, not just those with identified sleep problems, are motivated to engage with personal informatics. Many of these apps (e.g. *Sleep Journal*, *YawnLog*) are little more than digital diaries, or journals, which facilitate self-report of sleep habits; others (e.g. *SleepCycle*) however utilize accelerometers in mobile phones to automatically monitor movement during sleep and therefore infer sleep patterns. More sophisticated systems use bespoke sensor devices to replicate actigraphy (e.g, *WakeMate*, *FitBit*, *SleepTracker*) or even simplified EEG measurements (e.g. *Zeo*).

Further recent reviews of these apps and services can be found in Choe et al [11] and in Kay et al [21]. It is clear that the goal of such systems is to allow personal recording of sleep in ways that are unobtrusive and fit with people's everyday lives. In an interview study, Choe et al. [11] confirmed that that people were most interested in sleep monitoring technology that was unobtrusive and that introduced a minimum of additional devices e.g. they wanted such technology to be implemented through smartphones. However, as also observed by Choe et al [11], very few, if any, of these existing commercial systems have undergone any rigorous validation or testing to determine their reliability or comparative performance against well-established methods such as actigraphy or PSG. The prevalence of such apps, however, does vividly illustrate the large scale interest in measuring sleep expressed by large numbers of people, and not simply those people undergoing sleep therapy.

THE DESIGN OF THE SLEEPFUL APP

Design Challenges in a Sleep Setting

Measuring sleep accurately

The purpose of the Sleepful app is to accurately measure sleep. In particular it needed to measure sleep/wake states and be able to generate data that could be used to derive common sleep quality measurements such as sleep efficiency (SE). In order to ascertain its accuracy in doing this it had to be designed so that it could be used in conjunction – or in parallel – with existing sleep measurement approaches including actigraphy and PSG. Additionally, in order for it to be recognized by the sleep research community, the app ideally had to be grounded in previous clinical sleep research that had evaluated active systems for measuring sleep in situ in the bedroom.

Maintaining sleep hygiene

Deployment of interactive technology such as phones, TVs, laptops, and MP3 players in the bedroom is completely juxtaposed to good sleep hygiene, as such devices represent unnecessary sources of arousal in a sleep setting [7]. Therefore the design and deployment of any interactive technology for recording sleep is immediately problematic. Specifically, if an application required the user to be cognitively and physically engaged with the technology, this would be unsuitable for use in a) the bedroom and b) those suffering from insomnia, where further disruption to their sleep may exacerbate their sleep disorder. Whereas typically a mobile application should be designed to actively engage the user, we must, in this context, design an application to be as calm and non-stimulating as possible. Hence, for instance, our interface design approach was to use low-key, cool, colour schemes and low contrast text with minimum configuration options and utilities.

Taking partners needs into account

In work related to that of Choe et al [11], Aliakseyeu et al [1] highlight interaction design opportunities around sleep. They draw attention to the use of typical bedroom technology such as TV systems and services in light of sociological issues surrounding sleep. For instance they highlight the very real problem of bedroom *partners* being disturbed by the engagement with interactive technology by users. Since most adults share their bed with a partner at least for some time in their life, this is also a potential general design issue for sleep measurement technology. Thus, the Sleepful App design ideally demanded a style of interaction that would not disturb partners. We note, however, that not only have previous researchers constructed social systems to actively share sleep behaviour (e.g. [23]) but also that recent research suggests that bedroom partners can play a significantly positive role in sleep behaviour and hygiene [34].

Maintaining Privacy

Choe et al [12] describe work in understanding attitudes towards the technological monitoring of round the clock activities in household settings as a personal informatics and Ubicomp problem. They report that if people mentioned a location in the home where they did not want their activity to be monitored, then the most frequently mentioned place was, indeed, the bedroom. However, they also reported that only 4.6% of categorized activity types roughly corresponded to sleep-related activities (sleeping, snoring, napping, staying up late). Whilst accelerometer based actigraphy has indeed been used to measure sexual activity in deliberate clinical trials [5] as well as in sexual health social marketing campaigns [13] it is self-evident that most people would not wish such activity to be tracked as a by-product of a sleep measurement system. Therefore, a further design consideration of our Sleepful app was that it should not measure anything other than sleep.

Basis for design: the stimulus response paradigm

Taking into account the above design considerations, we designed a mobile application that built upon previous research by Riley *et al* [33] who conducted a sleep measurement and intervention trial that utilized a bespoke hardware platform to deliver a course of sleep restriction therapy. In turn Riley et al [33] based their work on a large body of earlier sleep research conducted primarily by Lichstein and colleagues (e.g. see [22]) but utilized by many other researchers (e.g. [16]) in which a stimulus (a short duration, low volume audio tone) is emitted from a device every 10 minutes. The basic concept is that if a user is awake, they will respond to the tone via a recording device; whilst if they are asleep they will not. Users may also toggle the state of the system to being in or out of bed (e.g. if they decide to use the bathroom during the night). Hence, this is an *active* method of measuring sleep behaviour (including SOL, WASO and TST) using a *stimulus response* paradigm. According to Riley et al [33] recording sleep in this manner demonstrates greater than 90% agreement on measures of SE with more intrusive forms of measurement, such as PSG, while only incurring modest sleep disruption due the audio tones.

Lichstein and colleagues deployed their early active sleep recording method using microphones and tape recorders: i.e. users had to *speak* when they heard the tone. Riley et al [33] deployed their measurement system on a re-programmable eight-bit microprocessor handheld computer device previously designed for monitoring and scheduled prompting of smoking behavior. The device, which they termed an Insomnia Assessment and Treatment Program (IATP), though rather large (around 80mm x 50mm x 15 mm), was mounted on the wrists of trial participants using a Velcro strap when they got into bed.

As mentioned above, Riley et al [33] showed that the IATP was indeed able to accurately record participants sleep behavior. However, they also reported some drawbacks of the system: 20% of their participants suffered sleep disturbance resulting from the audio tones, as these were not reprogrammable in terms of frequency, or volume; additionally users reported only moderately positive ratings on ease of use and general satisfaction. This is unsurprising given that the device could be seen as large, cumbersome and unlike any modern computing device the users would be likely to have encountered. The IATP was also only able to record data to the local device and therefore was limited both in terms of the amount of data that it could store as well as in issues around accessing the data for analysis.

We chose to build upon the work of Riley et al [33], and Lichstein and colleagues prior to that, by redeploying their stimulus response approach on a modern smartphone platform. As a mobile phone is a familiar, ubiquitous device that is very often kept in the bedroom next to the owner we believe that it resolves the issues of using unfamiliar devices for recording as well as the effects of a new device and its physical placement in the bedroom.

Implementation

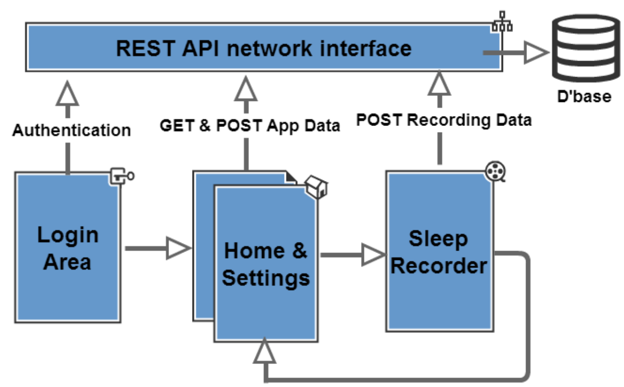
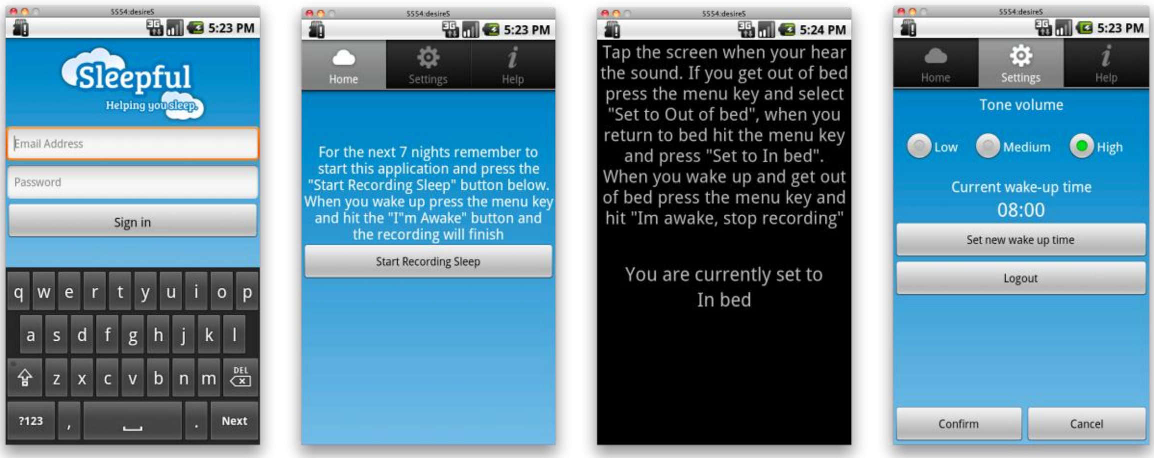


Figure 1. Functional design of Sleepful app

We implemented and evaluated the Sleepful app on Android devices. The app is server-driven: it pulls user profile data (including login information) from a HTTP server and pushes recorded sleep data for users back to the server using a custom built REST Application Programming Interface (API) via the smartphone's data connection. This solves limited data logging issues and allows easy access to sleep data by users themselves as well as, if necessary, clinicians or researchers. The user interface to the app was designed to feature cool, low contrast colouring using standard Android UI widgets that made use of both the hardware buttons of the device as well as the capacitive touch screen for input. The app consists of a very simple sequence of connected screens linked as shown in Figure 1.



(a) (b) (c) (d)

Figure 2. (a) Sign-in (b) Home (c) Sleep Recording and (d) Settings pages for the Sleepful app

The user initially signs into the application (see Figure 2(a)), and their details are authenticated through the API. Upon a validating the user's credentials the user is then taken to the home screen (shown in Figure 2(b)). The home screen relays instructions to the user, prompting them to record their sleep on a nightly basis. When the user enters their bed they begin recording their sleep by pressing the 'Start Recording Sleep' button located in the Home screen. Once the user hits the 'Start Recording Sleep' button on the phone then the app shows a black screen with simple instructional text (see Figure 2(c)) and begins to emit the low audio tones previously discussed.

We chose to emit a short 4-second low frequency (95Hz) low volume tone every 15 minutes. If a user is awake when they hear the tone then the expectation is that they respond to the tone by touching the screen of the mobile device – an interaction that is simple and possible to do in dark or poorly lit settings. If the participant should exit or enter their bed they are also required to toggle a soft switch within the application that sets the participant's status to 'Out/In of bed'. These tone responses and status toggles are recorded locally by the app and pushed at regular intervals to the HTTP server. The application allowed users to modify settings such as tone volume (see below) and wake up times within the application (see Figure 2(d)) which would also be pushed to the server for logging.

Pilot Testing

The Sleepful application was pilot tested with seven individuals over the course of seven nights. Several participants reported back that the tone was too loud and woke them up when they were on the cusp of sleep. However, one participant reported the opposite; that they couldn't hear the tone. This suggested that the tone setting should be dynamic, to support individual differences in hearing. The function to adjust the volume of the tone was added to the application. Pilot testers also reported that the screen was too bright for a dark bedroom, with two users covering up the handset at night. The text on the recording screen was originally white, which affected the overall brightness of the display. This text was adjusted to dark grey, which meant that the text was very feint in daylight, but was bright enough to be seen in a darkened room. Following the pilot study we progressed with two substantive user trials which are now described.

STUDY 1: ACTIGRAPHY COMPARISON AND USABILITY

The aims of Study 1 were a) to assess the accuracy of the Sleepful app's sleep efficiency (SE) measurement against the well-established and validated technique of wrist actigraphy and b) to investigate the usability of the application in light of the design challenges outlined above.

Participants and screening

Thirty-eight participants were recruited in Lincoln, UK. Two participants were excluded from taking part after exceeding the parameters on screening questionnaires (see below). Of the remaining, five did not complete: two of these gave reasons not connected to the app, whilst three reported that they found themselves to be "over prepared" for the tones and this caused anxiety leading to their withdrawal from the study (discussed further below). Of the thirty-one participants who completed the study, five did not supply at least four days data (either because they forgot to use the app, or wear the wrist actigraphy device) and therefore their data was excluded from the statistical analysis; however their comments are included in the qualitative element of the evaluation below. The remaining twenty-six participants, whose data were included in the statistical analysis consisted of five males (with a mean age of 38.49 years [SD = 11.28]) and 21 females (with a mean age of 32.75 years [SD = 13.11]). Participants were given information about the study before completing a consent form and two questionnaires which screened participants who may have had serious sleep disorders. Firstly, the Epworth Sleepiness Questionnaire [19] was used to eliminate people with very severe clinical sleep problems. Secondly, the Restless Legs Syndrome (RLS) Screening questionnaire [2] was also presented. The presence of RLS

[15] did not prevent participants from taking part in the trial, but would result in special treatment of their Actiwatch data.

Materials

Two standardized usability questionnaires, the NASA Task Load Index (TLX) [18], and the Computer System Usability Questionnaire (CSUQ) [25] were used to understand whether there were any significant problems with the design of the user interface of the application. Timings were also collected for each task, taken from video footage of the usability testing.

Participants completed a Daily Sleep Diary each morning. For each night of the trial the sleep diary collected details of: bed time, sleep onset latency (SOL), the number of times they woke up during the night, wake after sleep onset (WASO), final wake time, the time they got up, and how long they spent in bed. This data was used in the automatic interpretation of the Actiwatch data by its associated commercial software [32] which then generated SE scores. Data collected by the Sleepful app comprised of all periods of wake/sleep states from which we also calculated SE scores. Participants also completed subjective ratings of their sleep quality each morning. They were given opportunity in the diary to make comments regarding their use of the Sleepful application over the course of the trial. Personal details including age and whether they currently shared a bed with a partner. were also collected.

Participants were supplied with either an HTC Wildfire S or an HTC Desire S smartphone, both running identical versions of the Android OS. Each phone was supplied with a 3G data card to connect to the HTTP server. A commercial wrist actigraphy device, a CamNtech Actiwatch 4 actigraphy device [9], was also given to the participants. The device was programmed to record activity between 0.35 - 7.5 Hz at an epoch of 0.25 minutes.

Procedure

Initial Appointment

The initial appointment in which the participants were supplied with the Sleepful application on a mobile phone, took approximately 30 minutes. After informed consent had been obtained, participants were given a form to provide personal details. The participants were then asked to familiarise themselves with the mobile handset and were given written instructions. The participants were asked to complete 8 typical tasks on the Sleepful app, each of which was timed after which they completed a TLX questionnaire. The tasks included: opening the app, logging in, setting a wake time, starting/stopping sleep recording, responding to a sleep tone, toggling in/out of bed status, adjusting tone settings. Any participants who struggled to use the application were given extra help until they reported being comfortable. The participants were informed that they would be contacted to return the equipment after the trial period (7 nights) had completed.

Debrief

A follow-up appointment, in which the participant returned the equipment, took place one week after beginning the trial. This appointment lasted approximately 5 minutes Participants completed the CSUQ and were asked for their comments on using the application. Feedback in the form of their SE score was given to the participants a few days after the debrief appointment.

Results

Data Treatment

Sleep diaries were used to support the analysis of actigraphy data. SE scores were calculated for each day of the trial and a weekly average was calculated using the Actiwatch software. The daily and weekly SE scores were also calculated using the Sleepful mobile application. A minority of participants ($n = 10$) used both the app and the Actiwatch for the full seven nights. However, all participants used both for at least 4 days. Thus, data for 4 days from each participant was used in analysis. This was deemed to be acceptable as checks on the accuracy of the first day's data against the remainder showed no difference ($Fr[3] = 1.99$, $p>.05$), suggesting that participants were able to use the equipment accurately from the start of the trial. None of the participants in this analysis suffered with Restless Legs Syndrome.

Accuracy of SE Measurement

A Pearson's product moment correlation coefficient was used to compare, on a day by day basis, sleep efficiency (SE) data gathered via the application, with that recorded via actigraphy. The correlations demonstrate a fairly strong concordance. The mean SE was calculated for the four days' data for both the Actiwatch (Mean = 82.47%, SD = 9.4) and the Sleepful Application (Mean = 79.25%, SD = 14.02). A paired sample t-test showed that there was no significant difference between the SE scores ($t(25) = 1.617$, $p>.05$) measured by actigraphy and the Sleepful app.

Usability

The results from the NASA TLX measure, suggest that the app was judged as relatively easy to use. All mean scores for mental, physical and temporal demand, performance and frustration fell within the bottom 25^{t} percentile of possible scores. Comparisons of TLX scores showed no significant difference in perceived difficulty between participants who had previous smartphone experience and those that did not ($p>.001$, accounting for adjustments for multiple comparisons) The post-trial usability questionnaire (CSUQ), which examined the general experience of using the application during the week, also suggested that there were no significant problems with the interface design. The scores, like the TLX measure, fell within the bottom 25th percentile of scores and once more there was no significant difference between those with smartphone experience, and those without on the questionnaire as a whole, and each individual subscales ($p > .01$).

User reported feedback

When participants were debriefed, we asked them about their experiences of using the Sleepful application. Note that a side-by-side comparison of the *experience* of actiwatches and the app is problematic because of the very different nature of the systems.

Ease of Use

Participant 29, who found the Actiwatch too uncomfortable to wear in bed (and thus did not form part of the accuracy sample), said that the "first time using touch screen was surprisingly easy" and that she didn't have to refer to the instructions after the first night of use. Participant 31 found that she had to consult the instruction sheet supplied to remember how to toggle in and out of bed, but that the application was generally "easy to use, I don't have to worry about it". Participant 12 also "didn't have to think about how to use it". It is interesting to note that whilst participants freely spoke about the tone and brightness, on the whole they needed to be prompted on the ease of use of the app. This is very encouraging, as one of the design challenges was to create an artifact that did not overly engage the user. If users were immediately focused on how they used the app, even if they indicated that this was a positive experience, it would suggest that they were engaged with the use on some level. Given that the participants had to be prompted also suggests that there was little conscious engagement with the app.

Participant 12 reported vivid dreams whilst using the application that stopped once she finished the trial. This raised the question of whether the application, while functioning as an accurate recording device, was having some unintended and detrimental effects on the sleep of participants; effects that were difficult to detect through actigraphy or sleep diary. Such potential issues were investigated in Study 2 (below).

Brightness

Despite modifications made to screen brightness after pilot trials, 8 participants (n = 31) still reported problems with the level of brightness in dark bedrooms. Conversely, one participant said that they found the screen too dim to read the instructions for toggling in and out of bed in the middle of the night. If an individual is acclimated to a very dark room, any light may seem very bright. 4 of the participants who found the app too bright did acclimate to the light after a period of use, whilst the other four moved the phone so it was out of their eye line, or covered it up with a cloth or piece of paper. It is important to note that no participants withdrew from the study due to unacceptable levels of brightness. This suggests that whilst levels of brightness may cause slight disturbance, users will be able acclimate or make adjustments for use, but that some brightness is necessary in order for individuals to use the app effectively.

Audio tones

None of the participants reported that the tone was loud enough to wake them, or prevented them from falling asleep. Participants were advised to test the volume of the tone in their bedroom as acoustics were likely to be different to those in the lab. The participants did make use of the volume adjustment feature of the app, with some reporting that they changed the volume over the course of the nights that they used it. The majority of the participants used the app with the volume set to low, but 8 (25%) participants used the app with the louder volume settings, suggesting that the ability to adjust the volume for individual levels of hearing is an important one.

Sleep Partner

Participants who shared a bed with a partner were also asked whether their sleeping partner had any adverse experiences during the trial and, in particular, whether they noticed the tones. Eighteen of 31 participants (58%) had sleep partners. Only two of those participants reported that the app was found to be problematic by the partner. Although we would obviously desire there to be no detection at all, this was considered an acceptable level.

A state of preparedness

Three participants withdrew from the study due to experiencing a mental state of "preparedness" after using the app for the first night. They felt that they were waiting for the tone to sound which resulted in them being less relaxed than they would normally be, before falling asleep. Of those that completed the trial, Participant 5 said that she was "very aware of being awake, when she wasn't normally, Participant 22 said that she was "lying awake waiting for it to sound", and Participant 24 said that when she woke up she was "trying to stay awake for the next buzzer". Participant 22 also reported that the tone "wakes you up when you are about to drop off", this was echoed by Participants 14 and 29 who felt that the tone sounding brought her back "from the edge of sleep" (P14) and "started to drift off and woke to the tone" (P29), however both reported that this improved as they continued to use the app over the week. It is worth noting that none of these participants had sleep disorders and reported no problems in naturally going to sleep. It may be the case that individuals who suffer from serious sleep disturbance may react differently to the tone, and this will be addressed when the app is tested in a clinical population.

Contrastingly, three participants reported that the app produced in them a positive mental state. Participant 9 said that she "slept better than normal" after the first night of use. Participant 10 said that the act of setting the app was almost affirmative "I am going to sleep now" which meant that she settled down to sleep very quickly, rather than tossing and turning. Participant 33 reported that on the third night of use she was "now into the swing of things, I was used to the phone and the app" and that as she went to be she "shall almost miss the nightly presence of the bleeper" it was "somehow rather comfortable". This "preparedness", whether it be negative - lying waiting for the tone to sound, or positive – placing the user in a state where they are prepared to sleep, shows that app was effecting some users' thought processes. However, users that persevered with the app despite initial negative consequences found that this state diminished over time. Users that experienced a positive state of preparedness found that this state was sustained.

Discussion

The overall Sleep Efficiency scores for the week showed no significant difference between the Sleepful app and the Actiwatch data. It is upon figures of SE that an individual's sleep therapy is created; hence our choice of this metric as the sole measurement in this Study. Direct comparison of other measures of sleep such as WASO and TST here would be unreliable/unhelpful due to inherent differences in size of epoch in the two systems (the actigraph epoch is 0.25 seconds, the Sleepful app epoch is 15 minutes) though further research would be useful in determining any differences in these additional metrics. The usability measures together with the debriefing interview demonstrate that the app fulfilled the design criteria to offer low intensity user engagement of low complexity. The simplicity of use, together with the repeated use over time meant that using the app became instinctive for some participants. Therefore we are satisfied that this application is fit for purpose in this respect. The correlational analysis of the day by day measurement data was good, but did not reflect as strong as a relationship to suggest that the Sleepful Application measured sleep efficiency in the same way as the Actiwatch technology. However, actigraphy has been found to overestimate SE [31] and, given the epoch differences, it is reasonable to experience some discrepancy in the scores.

STUDY 2: POLYSOMNOGRAPHY COMPARISON & SLEEP DISTURBANCE ANALYSIS

The aim of Study 2 was to a) compare synchronous measures of sleep/wake detection using overnight PSG, wrist actigraphy, and the Sleepful app; and b) evaluate the impact of the app on inter sleep arousals as measured by PSG. PSG records electrical impulses in the brain that represent areas of activity that characterize sleep and wake states and is considered to be the gold standard objective measure of sleep. For a more detailed explanation see [19].

Participants

Participants were recruited through newspaper advertisements and community noticeboards in Loughborough. Inclusion criteria included being able to accommodate home PSG on two consecutive nights, whilst exclusion criteria included assessment via the Epworth Sleepiness Questionnaire [19] and the RLS questionnaire [2]. Those with insomnia symptoms were not required to cease taking hypnotic medication for the purpose of the study. Five participants (3 women and 2 men) aged between 30 and 51 years of age (mean 37.20, SD= 8.64) completed the overnight sleep recordings. Participants had a mean score of 8 (SD= 8.64; participant range 1-15) on the Pittsburgh Sleep Quality Index [8] a score of 5 or above indicates clinically significant poor sleep). Data are reported from the second night of PSG recording.

Method

We conducted polysomnography following the American Academy of Sleep Medicine Guidelines [19] and included encephalographic (EEG), electromyographic (EMG), and electrooculagraphic (EOG) monitoring. Sleep stages and arousals were scored in line with AASM criteria by an experienced researcher using Somonologica Studio (Somnologica Studio, Medcare Flaga Medical Devices). For reliability assessments, randomly selected 15 minute sections from each quartile of nightly recordings were scored by a second rater. Mean reliability was 84.5% (range 83.4%- 98%). All sleep recordings met the 82 % inter-rater concordance guideline as stated by Iber et.al [19]. Participants wore a Philips Respironics actigraphy device [32] and completed a standard sleep diary for the 7 night duration of the study. Days 1 to 5 allowed the participant time to adjust to having sleep monitored by actigraphy. On the 6th and 7th evenings, overnight PSG was conducted in the participant's home (PSG 1 & PSG 2). PSG 1 allowed participant's to get used to overnight sleep recording and this data was not used in analysis.

Results

Concordance of sleep measurement with Sleepful app

Percentage agreements were computed using point by point comparisons to test the sleep/wake detection ability of the app. Each of the tones was first marked on the PSG record at the time point when it was generated and the percentage agreement was calculated. An average of 29.4 tones was generated during a sleep period (range; 24-32); a total of 147 across the recorded sleep periods of all participants. Of these tones, 125 showed sleep/wake concordance between PSG and the application (an 85% agreement rate). Of the 22 tones which lacked concordance, 18 of the tones were responded to when the PSG record showed sleep, while 4 tones were ignored during periods of PSG wake. Tones responded to whilst the PSG showed sleep may reflect the PSG sleep scoring, which require that any 30 second epoch scored is scored as sleep if the majority of the trace shows sleep activity, and the preceding and following epoch shows the same activity. Thus, if a participant woke momentarily to respond to the tone and then returned directly to sleep within a 30 second window (i.e. the second half of one epoch and the first half of the following epoch), then that epoch could be scored as "sleep". These results are very similar to those of Lichstein et al [27] where analysis of disagreements in PSG vs a spoken response to a tone showed that 54% of non-concordance was due to the PSG scored as sleep when participants had responded to the stimulus.

Examination of Sleepful intrusiveness

The possible intrusiveness of the app on sleep was assessed in three ways adopting the methodology of Lichstein et al [27] by i) examination of the percentage of awakenings associated with a tone ii) examination of the % of tones associated with awakenings; and iii) comparison of tone and non-tone related awakenings. Note that it is generally difficult to infer a causal relationship between tone and awakening as people experience a number of spontaneous awakenings or transitions to wake stage throughout the night. When 23 of the tones were generated, the participant was already in a period of wakefulness (in excess of 30 seconds). The remaining 124 tones occurred during Stage 2 sleep. Of these tones, 7 (5.6%) were associated with an awakening within 6 seconds of tone generation. By comparison, 58 spontaneous (not associated with tone) awakenings from sleep (meeting the definition outlined earlier) occurred during all recorded sleep periods. Of the total 65 awakenings experienced during the sleep period, 10.8% occurred within 6 seconds of a tone occurring, and 89.2% occurred at other times, suggesting that participants awoke more regularly in the absence of a tone. Hence we can say with some confidence that the application does not appear to be problematic in terms of waking users. In future we intend to improve understanding of any disturbance through exploration of smartwatches and variation of the tonal parameters.

Discussion of polysomnography

Results suggest a high concordance in sleep-wake detection between PSG sleep recording and the app. The level of accuracy shown by the app is consistent with its use as a personal sleep monitor capable of detecting personal and relevant changes in sleep structure. In addition, it does not appear that the application caused significant disturbances in the sleep of users. While some instances were recorded of awakenings in response to the tone, these represented only 11% of the total awakenings recorded in the data.

CONCLUSIONS

We have described the development and evaluation of a mobile phone app for everyday, unobtrusive monitoring and recording of sleep. The contribution of the work is in the design and development of an app that meets all identified design requirements for measuring sleep and, most particularly, in its validation through a comparison of its effectiveness with established sleep measurement methods used in clinical sleep research. The app was also rigorously tested in terms of usability as well as for any potentially negative effects on the sleep of the user and on the sleep setting (including on sleep partners) using established, well-reported, methods. We conclude that the app could immediately be offered for use in a number of settings:-

Everyday personal informatics by casual users. There is an evident enthusiasm for people to measure their sleep for self-awareness and reflection and perhaps to better understand their own health without immediate recourse to professional health services. The Sleepful app allows users to measure sleep without the prohibitive cost of traditional sleep measurement devices. In our own future work we plan to incorporate the Sleepful app into an online social network platform which will allow people to view, reflect upon and share visualisations of their sleep. In recent related work Kay et al [21] describe *Lullaby* which allows people to reflect on the unconscious experience of sleep (and its disturbance) through sophisticated visualisations.

Integration with novel, interactive digital sleep interventions. In future work we plan to incorporate the Sleepful app into a computerised CBT-I intervention for people with proven sleep problems. The CBT-I will be delivered online in parallel with a sleep intervention based on sleep restriction therapy (as used in [33]) which can be delivered through simple modification of the app (i.e. through the functionality to set 'sleep windows'). The inbuilt server based data logging of the app lends itself ideally to monitoring of the efficacy and adherence to the sleep restriction therapy by users. Although we plan to pursue this work ourselves we also offer the app and its design for use by other researchers.

Utilisation as a cheap, easy to use sleep measurement tool in any sleep research. Our app has been validated against proven, well-established sleep measurement techniques (actigraphy and PSG). With careful experimental design it could feasibly be utilised as a sleep measurement tool in any sleep research. Again the inbuilt uploading of sleep data is ideal in this setting as researchers, or clinicians, would be able to monitor sleep behaviour remotely.

ACKNOWLEDGMENTS

This research was funded by EPSRC grant EP/I000615/1.

REFERENCES

1. Aliakseyeu, D., Du, J., Zwartkruis-Pelgrim, E., and Subramanian, S. Exploring Interaction Strategies in the Context of Sleep. *Proc Interact 2011*. Springer. (2011).
2. Allen RP, Picchietti D, Hening WA, Trenkwalder C, Walters AS, Montplaisi J. Restless legs syndrome: diagnostic criteria, special considerations, and epidemiology.. *Sleep Med, 4*, 2, (2003) 101-119
3. Altena, E., van der Werf, Y., Strijers, R.L.M., & van Someren, E.J.W., Sleep loss affects vigilance: effects of chronic insomnia and sleep therapy. *Journal of Sleep Research, 17*, 3, (2008) 335-343.
4. Ancoli-Israel, S., Cole, R., Alessi, C., Chambers, M., Moorcroft, W., Pollak, C.P. The role of actigraphy in the study of sleep and circadian rhythms. *SLEEP*, 26(3), (2003) 342–392.
5. Andersen, M.L., Poyares, D., Alves, R.S., Skomro, R., & Tufik, S. Sexomnia: Abnormal sexual behaviour during sleep. *Brain Res Rev* (2007) 56:271-82.
6. Bauer, J.S., Consolvo, S., Greenstein, B., Schooler, J., Wu, E., Watson, F., & Kientz, J. ShutEye: encouraging awareness of healthy sleep recommendations with a mobile, peripheral display. *Ext. Abstracts CHI2012* ACM Press (2012) 1401-1410.

7. Brick, C.A., Seely, D.L., & Palermo, T.M. Association Between Sleep Hygiene & Sleep Quality in Medical Students. *Behavioral Sleep Med, 8*, 2, (2010) 113-21.
8. Buysse, D., Reynolds III, C., Monk, T., Berman, S., & Kupfer, D. The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. *Journal of Psychiatric Research*, 28(2), (1989) 193.
9. Cambridge Neuro Tech. http://www.camntech.com/
10. Chandola, T., et al. The Effect of Short Sleep Duration on Coronary Heart Disease Risk is Greatest Among Those with Sleep Disturbance. *SLEEP*, 33 (6) (2010).
11. Choe, E.K., Consolvo, S., Watson, N.F., & Kientz, J.A. Opportunities for Computing Technologies to Support Healthy Sleep Behaviors. *Proc CHI 2011*. ACM Press (2011) 3053-3062. ACM.
12. Choe, E.K., Consolvo, S., Jung, J., Harrison, B., & Kientz, J.A. Living in a Glass House: A Survey of Private Moments in the Home. *Proc UbiComp 2011*.
13. Condom08 project. (2011) http://bit.ly/v3GBtW
14. Drake, C.L., Roehrs, T., & Roth, T. Insomnia causes, consequences, and therapeutics: an overview. *Depress Anxiety*;18 (2003) 163–76.
15. Earley, C.J. Restless Legs Syndrome. *N. Engl. J. Med*, 348, (2003) 2103–2109.
16. Espie, C., et al Use of the Sleep Assessment Device (Kelley and Lichstein, 1980) to validate insomniacs' self-report of sleep pattern. *Journal of Psychopathology and Behavioral Assessment*. 11 (1), (1989) 71-79.
17. Frankel, B.L., Buchbinder, R., Coursey, R.D., & Snyder, F. (1976). Recorded and reported sleep in primary chronic insomnia. *Arch. Gen. Psychiatry,* 33, 615–23
18. Hart, S.G., & Staveland, L.W. Development of the NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload. Advances in psychology, 52,* (1988) 139-183.
19. Iber, C., Ancoli-Israel, S., Chesson Jr, A., & Quan, S. *The AASM Manual for the scoring of sleep and associated events*. Westchester, IL: American Academy of Sleep Medicine (2007).
20. Johns, MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *SLEEP 14*, 6, (1991). 540–545.
21. Kay, M., Choe, E.K., Shepherd, J., Greenstein, B., Consolvo, S., & Kientz, J.A. (2012). Lullaby: A Capture & Access System for Understanding the Sleep Environment. *Proc UbiComp 2012*. ACM Press (2012).
22. Kelley, J E., & Lichstein, K L. A sleep assessment device. *Behavioral Assessment*, 2, (1980) 135–146.
23. Kim, S., Kientz, J.A., Patel, S.N., & Abowd, G.D. Are you sleeping?: Sharing portrayed sleeping status within a social network. Proc CSCW '08. (2008). 619–628.
24. Leger, D., Scheuermaier, K., Philip, P., Paillard, M., & Guilleminault, C. SF-36: Evaluation of quality of life in severe and mild insomniacs compared with good sleepers. *Psychosom Med* (2001) 63:49–55.
25. Lewis, J.R. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, *7*, 1, (1995) 57-78.
26. Li, I., Dey, A., Forlizzi, J., Höök, K. & Medynskiy, Y. Personal informatics and HCI: design, theory, and social implications. *Ext. Abstracts CHI2011* ACM Press (2011). 2417-2420.
27. Lichstein, K., Nickel, R., Hoelscher, T., & Kelly, J. Clinical validation of a sleep assessment device. *Behaviour research and Therapy*, 20, (1982) 292.
28. Mei, E., & Buysse, D.J. Insomnia: Prevalence, Impact, Pathogenesis, Differential Diagnosis, and Evaluation. *Sleep Med Clin*, 3, 2, (2009) 167-174.
29. Morin, C.M., Vallieres, A., Guay, B.; et al. Cognitive behavioral therapy, singly and combined with medication, for persistent insomnia: a randomized controlled trial. *JAMA*, 301, (2009) 2005-2015.
30. Morin, C.M., Bootzin, RR., Buysse, DJ., et al. Psychological and behavioral treatment of insomnia: Update of the recent evidence (1998-2004). *SLEEP, 29*, 11, (2006) 1398-1414.
31. Paquet, J., Kawinska, A., & Carrier, J. Wake detection capacity of actigraphy during sleep. *SLEEP, 30*, 10, (2007) 1362-1368.
32. Philips Actiware http://www.learnactiware.com/
33. Riley, W.T., Mihm, P., Behar, A., & Morin, C.M. A computer device to deliver behavioral interventions for insomnia. *Behav Sleep Med*. 8(1) (2010) 2-15.
34. Rogojanski, J., Carneya, C.E., & Monsonb, C.M. Interpersonal factors in insomnia: A model for integrating bed partners into cognitive behavioral therapy for insomnia. *Sleep Medicine Reviews* (in press).
35. Sadeh, A., Hauri, P.J., Kripke, D.F., & Lavie, P. The role of actigraphy in the evaluation of sleep disorders. *SLEEP, 18*, 4, (1995) 288-302.
36. Stepanski, EJ., & Wyatt, JK. Use of sleep hygiene in the treatment of insomnia, *Sleep Medicine Reviews*, 7(3) (2003) 215–225.
37. Varkevisser, M. & Kerkhof, G.A. Chronic insomnia and performance in a 24-h constant routine study. *Journal of Sleep Research, 14*, 1, (2005). 49-59.
38. Vgontzas, A.N., Liao, D., Pejovic, S., et al. Insomnia with objective short sleep duration is associated with type 2 diabetes: a population-based study. *Diabetes Care* 32 (2009) 1980–1985.