A qualitative study of sleep trackers usage: evidence of orthosomnia

Samuel Aupetit[1], Gäetan Dubroca[1], Sara Escaich[1], Philippe Cabon[2]

[1] ERGO-CENTRE, France [2] LATI/University of Paris Descartes, France

ABSTRACT

This article deals with digital sleep trackers that give information about sleep duration and quality in everyday life. Despite the number of these devices, scientific and independent studies of their usage are rare. This work aims at describing the acceptance (short term) and appropriation (long term) of these devices in order to identify their benefits and limitations. Data collection combines questionnaires on sleep and user experience, a sleep diary, and interviews on human computer interface. The main results concern the evidence of "orthosomnia", an anxiety phenomenon, that is affecting people who obsess over the results of their sleep trackers.

KEYWORDS

HCI; Technologies; Acceptance; User centred design; Sleep.

Introduction

Individual sleep trackers are more and more numerous in private life for monitoring the duration and the quality of sleep in light of well-being and sleep performance. These new devices correspond to changes in the field of health that move from a focus on the disease towards a personalized preventive health management (Swan, 2012). This domain called "quantified self" corresponds to considerable economic and societal stakes. In this context, sleep monitoring is a major concern in everyday life given the impact of sleep on health. Over the past century, sleep duration has been reduced by approximately 90 minutes (Heslop et al., 2002) due to artificial lighting and social activities. Recent studies point out that low sleep duration (less than 7h) is associated with increased cardiovascular risk (Wingard et al., 1982) and significantly decreased life expectancy (Heslop et al., 2002). In terms of safety, driving drowsiness is one of main cause of highway fatalities (ASFA, 2018). The selling argument of the trackers available on the market refers to a benefit in comfort, health and user performance. However, we do not have a body of data that evaluate scientifically and independently the real usage of these devices.

Thus, the aim of this paper is to make a qualitative study of the use and acceptance of sleep trackers, in order to gain knowledge on their real usage and to identify their benefits and limitations. How people really use these devices? What types of problems do they face? What are the positive and negatives effects on everyday life and sleep?

This study refers to French cognitive ergonomics (Leplat, 1990) that presents conceptual and methodological tools for data collection and analysis in real world environment. This approach is linked with works on the dynamics of use ranging from an *a priori* assessment of the imagined usage of a technology to an *a posteriori* assessment of the actual use (Terrade et al., 2009). As a result, it is possible to specify three moments: "acceptability" refers to people's representations of a future or possible technology (Bobillier-Chaumon & Dubois, 2009); "acceptance" refers to attitudes during the first uses of the device influenced by its utility and usability. These two moments

represent the short-term assessments of the technology. The last moment refers to "appropriation" i.e. attitudes during regular use of technology which is a long-term assessment.

The study includes 2 studies. The first one is a "picture" of the use of these devices after a week. We have collected the first moments of experience with the selected devices in order to describe acceptability and acceptance. The second study is a longitudinal assessment of the appropriation and the effects of the devices on everyday life.

Study 1. "Picture" of acceptability and acceptance

The first step of the project was to make a benchmark of all the existing sleep trackers. This benchmark led to identifying 65 products. We decided to study the usage of 6 of them: Beddit (Sleep Tracker), Oura Ring (Oura), Emfit (QS), Hugone (Seven Hugs), Sleep Cycle (Alarm Clock), Fit Bit (Charge 2). The selection criteria were (1) the form of the device (bracelet, Smartphone, ring, watch…), (2) the type of sensors (accelerometer, light meter…), (3) data collected, (4) the sales on the French market, and (5) selling price.

Our decision to study the acceptance in such a systematic and detailed manner meant we had to investigate a relatively small sample of sleep trackers and users. This study design, which focuses on the details of individual experience, can produce findings which would be difficult to obtain with more quantitative approaches. We considered 10 users for each sleep tracker according to Nielsen (1993), to collect reliable data i.e. not influenced by the intrinsic aspects of the user. This kind of population is enough to detect 90% of usability problems.

30 participants were recruited for this study (16 men, 14 women; mean age 48). All subjects received a questionnaire on their motivation to participate, their relationship to digital technology and sleep, and a questionnaire to investigate their sleep habits and problems. Each subject was studied over a period of 3 weeks. The first week was the control week without any devices. During the second week, the participant had to use a first system. They used the second device the third week. The order of presentation of the systems was counterbalanced. In total, 82 weeks of experimentation were considered. Our approach is based on the combination of data from different levels:

- A daily sleep diary. It allows the evaluation of sleep duration, sleep latency, perceived quality of sleep and a sleepiness score thanks to the Karolinska Sleepiness Scale (KSS). The KSS is a subjective scale assessing alertness from 1 (very alert) to 9 (very drowsy). This scale was validated against physiological recordings and cognitive performance measures. Values higher than 7 are associated with physiological signs of drowsiness (microsleeps) and a significant impairment of performance;
- Observation of the use, questionnaire on user experience (Attrakdiff, Lallemand et al., 2015) and usability (System Usability Scale, Brooke, 1986) in order to evaluate the acceptance of the devices;
- Interview data were obtained to complete all the previous cited data.

Study 2. Longitudinal study of appropriation

45 subjects who have used a sleep tracker for 1 month to 2 years have been recruited (27 men, 18 women; mean age 52). We conducted an interview with each participant to analyze their usage dynamics. The same questionnaires on user experience as study 1 were completed.

Results

The results section gives a factual presentation of the main results obtained for the two studies. After presenting the principal benefits (knowledge on sleeping), several limits perceived by the users are identified: an increase in sleeping delay, a problem of usability, the lack of users help, and the evidence of orthosomnia.

The main benefit: "know more about my sleep!"

The results show that participants appreciate having information on sleep quality (percentage of each sleep steps, sleep time vs bed time), the sleep quantity (duration of sleep) and the sleep conditions (air quality, temperature, humidity). During the interview, they declared to have a very limited knowledge of what constitutes their sleep e.g. a subjective feeling of sleep, an estimation of duration, a feeling of asleep, sensation of night awakening. So, the main expectation they have concerns the capacity for the trackers to describe their sleep by objective data.

"Having information about my sleep environment was very interesting"
"The application was useful because I see the curves, it gives me an idea of the state of my sleep that I did not know"
"I noticed that when I play sports my sleep is more restorative"
"My opinion has changed, at first I was very sceptical because for me it was a gadget, it's really useful it allows to know each other better"
"What happens at night we do not know; the device helps me to better understand my sleep"

Sleep diary results: a limited usefulness and an increase in sleeping delay

Table 1 presents the results obtained with the sleep diary for each device. It is a comparison between the experimental week (with system) and the control week (without system). A positive result means an increase of the item (sleep duration, sleep delay, perceived sleep quality, drowsiness) compared to the control week, and *vice versa* for the case of a negative result.

Table 1. Sleep diary results

	Fitbit	HugOne	Beddit	SleepCycle	Emfit	Oura Ring	**Average**
Sleep duration	+ 3.2%	+5.4%	-2.4%	+4.6%	+2.7%	+3.5%	**+2.8%**
Sleep latency	+9.3%	+9.5%	+5%	+3.7%	+33.3%	0	**+10.1%**
Perceived sleep quality	-3.2%	-4.4%	-4.4%	-3.3%	-12.3%	-6.7%	**-5.8%**
Sleepiness score of Karolinska	-2.6%	-2.8%	0	-2.56%	+6.4%	-18.1%	**-3.3%**

Results present an increase in sleep delay (+10.1%), sometimes even very significantly (+33.3% for Emfit). Data also show a reduction in sleep quality compared to a week without a system (a reduced score of 12% for Emfit). Finally, for most trackers, there is a decrease in the Karolinska Sleepiness Scale score indicating that participants felt more alert during the week with device (-18% for Oura Ring). The increase in sleep time reported by participants is relative (+ 2.8% on average). In summary, the improvement of sleep conditions is not significant after a week of experimenting. We can even identify an increase in sleep delay with the trackers. Interviews conducted in the longitudinal study tend to confirm this limited usefulness but not the importance of sleeping delay. This phenomenon can be associated with the first steps of usage and the learning phases.

Evidence of orthosomnia

Several interviews highlight a phenomenon that can be compared to anxiety and deals with finding a perfect sleep. Providing an objective measurement of sleep, that can be compared at an individual level but also at an interpersonal level, seems to create or potentiate this deviance for some users. The consequences can be the prolongation of bedtime to gain points or the removal of any activity that is not supposed to increase the score:

“I really want to be over 80 out of 100 ... I do not know what it represents but I trust the device, it must be that I sleep better if I get closer to 100."
"I eat less or eat more in the evening and I look if it has changed anything about my sleep to gain some points!"
“I am not insomniac, I understand that my days influence my nights. But I pay more attention to my sleep with this device and I'm not sure it's the best thing to do!”
“During the week, I took a lot less outings with my friends because I think my score will be reduced! So I watch TV at home and see the effects on sleep and it works… I win 4 points!”
“One night I slept somewhere other than in my bed and I thought I was going to sleep badly and lost some points. The device will record me a bad night and he will make me a critic and tell me that the sleep was not restorative, it I very stressful”.

These results are in line with a recent American study published in the Journal of Clinical Sleep Medicine (Baron et al., 2017) dealing with activity trackers (bracelets) including day and night measurement (Call Watch and Fit Bit). It reveals that the wearing of these systems would favour what researchers have called "orthosomnia": an excessive preoccupation with one's sleep, preventing one from sleeping properly. The authors of the study adopted the word "orthosomnia" to describe what would become for some a real obsession. This quest for "ideal sleep" can, it seems, be amplified by the use of these bracelets. Orthosomnia is defined as the quest for perfect sleep, without insomnia, without nightmares, with regular cycles and ensuring its function of recovery of the body. However, this quest is so ubiquitous that it can be counterproductive: by being obsessed with sleep, one becomes less sleepy. Our work shows the similar excessive control of sleep that leads to increased anxiety and altered sleep conditions. Some subjects seem to be obsessed with the device: it's the first thing they do when they get up in the morning, they are disturbed when the system does not work, they diligently scrutinize all the graphics of the device and this several times in the day. This “dependence” is the case specifically for the users most concerned about their sleep and therefore the most in need of information and help:

"It makes addict! I absolutely want to compare each night! "
"It becomes an essential reflex to start the day. This is an important issue for me, sleep and so this device are important! "
"When I have a drink I think about the app. When I play sports I think about the app! "
"The first thing I do in the morning, before going to the bathroom,
it's to see if I slept well or not! "
" I look all day curves: those of the night but also those of the day because I wear the ring also the day like that I am always followed!"

A problem of usability

The results presented in Table 2 highlight the weaknesses of all the devices in terms of usability. The pragmatic quality is close to the average score (HugOne has a negative score), while the System Usability Scale establishes the average usability of the systems tested at 68 points, which is a score that is just acceptable for a usable product (the HugOne score is below the 50 points mark).

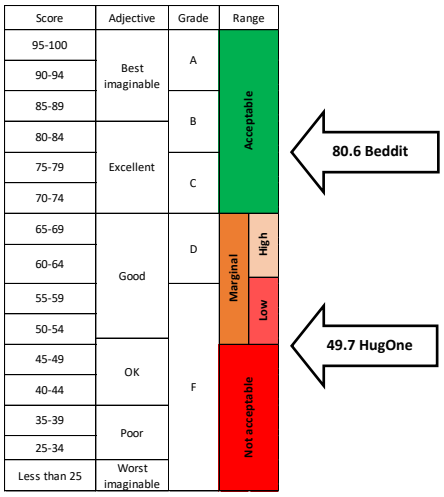
Table 2. Usability results obtained by questionnaires

	Pragmatic quality arising from AttrakDiff survey Scoring ranging from -3 at 3	Usability arising from SUS survey Scoring ranging from 0 at 100
Fitbit	0.49	62.5
HugOne	-0.32	49.7
Beddit	0.78	80.6
SleepCycle	0.18	67.5
Emfit	0.18	64.8
Oura Ring	0.48	65.7
Average	**0.3**	**65.1**

One of the main limitations consists in the complexity of the Human Computer Interface. Reading and understanding the displayed data tends to discourage users. The usability scores obtained by questionnaires and interviews show that a system can have a poor usefulness (gain knowledge and optimizing sleep) due to perceiving the usability as limited (accessibility and understanding of complex graphics). This is the case of the Emfit system, which has the highest theoretical utility, but which is not used by participants because its interface is too complex (limited accessibility, scientific vocabulary, too many graphics…).

"The application is different if I use Ipad mini, iPhone 6s and android Samsung. It's a little cheated, it's not the same product! I took the iPhone from my husband but before it was easy »
"I can see the data and the graphics, but I do not know what it is, I do not understand!"
"I did not see that you had to click to see the detail"
"If you do not have a minimum of computer skills, it's very difficult to use this device"
"The figures are very beautiful, but it is difficult to understand"
"It seems to be very scientific and I need a translation!"

Figure 1 shows the results for Beddit and HugOne obtained by the System Usability Scale questionnaire. The score obtained by Beddit is 80.6/100 and that obtained by HugOne is 49.7/100.



These are the two extreme scores collected. Beddit is rated as having "Excellent" usability while HugOne's interface is rated as "Not Acceptable". These results perfectly reflect the data collected with the other questionnaires and interviews.

A lack of effective help

The results show that the most useful systems were those allowing users to gain a better understanding of their sleep. The selling point of these devices suggests also, most of the time, to improve the sleeping conditions. The existing help takes the form of generic advice; "do not drink coffee before going to bed", "do not eat just before going to bed", "turn off the light". This advice does not match the expectations of users and thus cannot improve their sleep.

Figure 1. SUS results for Beddit and HugOne

"One night I did not sleep well at all, I woke up a lot of times. The system saw that I had a very bad sleep, the rating was not good at all, however it gave me as a solution to drink less alcohol ... thank you but I never drink a drop of alcohol. I did not know what to do in that case!"

The lack of help is even more annoying when the user is really concerned about their sleep. The quantity and quality of information provided to the users (very little deep sleep, a high heart rate, frequent awakening) are not associated with concrete advice, and that causes a state of anxiety. The participants declared they considered the displayed data as medical information and trusted it. So, the devices are considered as frustrating: they show and alert users regarding their sleep problems, but only provide ineffective "tips" that cannot solve these problems.

"I saw in the data a peak at over 100 heart rates one night, it really worried me! I would have liked the advice of a doctor to tell me if it was normal or not...! "

"I saw on the curve that I stopped breathing during the night: it means that I breathe no more, that I made an apnoea! I needed to have a professional opinion because if I made sleep apnoea it could be dangerous! "

Discussion

The study tends to show the acceptance of some sleep trackers by pointing out their main benefits and risks from the users' point of view. The devices allow the user to know more about their sleep to adapt their behaviour during night and day. The results show that the participants find a real usefulness in having information on the quality, the quantity and the environment of their sleep. This can help fight against sleep inertia and associated risks. However, several limitations have been identified e.g. poor accessibility, limited usability and lack of adapted and effective advice to improve sleep. The study highlights a specific limitation that can have important effects on everyday life of users: "orthosomnia" i.e. a new sleep disorder arisen from the pursuit of a perfect night's sleep and associated anxiety. This phenomenon is associated with the fact that number of patients who are seeking treatment for self-diagnosed sleep disturbances such as insufficient sleep duration and insomnia due to periods of light or restless sleep observed on their sleep tracker data. These trackers provide them an objective measure of their sleep, most of the time a score out of 100 points. This result is in line with a recent American paper (Baron et al., 2017).

The methodology used in this study combining (1) sleep trackers usage and effects on sleep analysis, (2) qualitative and quantitative data, (3) "picture" and longitudinal study of usage, succeeds in providing interesting knowledge on the acceptance of these devices. The future systems should take these results into account in order to match the users' needs (better accessibility, personalized displayed data, adapted advice) and to attempt to avoid orthosomnia (e.g. removing the typical indicator expressed in percentage) that is an important issue, regarding the number of human activity trackers available on the market.

Acknowledgements

This project was supported by the MAIF Foundation.

References

Baron, KG., Abbott, S., Jao, N., Manalo, N., & Mullen, R. (2017). Orthosomnia: are some patients taking the quantified self too far? Journal of Clinical Sleep Medicine, 13(2), 351–354.

Bobillier-Chaumon, M., & Dubois, M. (2009). L'adoption des technologies en situation professionnelle : quelles articulations possibles entre acceptabilité et acceptation ? Le Travail Humain, 72, 355-382. DOI:10.3917/th.724.0355

Heslop, P., Smitha, G.D., Metcalfe, C., Macleod, J., & Hart, J. (2002). Sleep duration and mortality: The effect of short or long sleep duration on cardiovascular and all-cause mortality in working men and women. Sleep Medicine, 3, 305–314.

Lallemand, C., Koenig, V., Gronier, G., & Mar5n, R. (2015). Création et validation d'une version française du questonnaire AttrakDiff pour l'évaluation de l'expérience utilisateur des systèmes interactifs. Revue Européenne de Psychologie Appliquée. doi:10.1016/j.erap.2015.08.002

Leplat, J. (2000). L'analyse psychologique de l'activité en ergonomie. Toulouse: Octarès.

Nielsen, J. (1993). Usability Engineering. Boston: Academic Press.

Swan, M. (2012). Sensor Mania! The Internet of things, wearable computing, objective metrics, and the quantified self 2.0. J Sens Actuator Netw, 1, 217-253.

Terrade, F., Pasquier, H., Reerinck-Boulanger, J., Guingouain, G., & Somat, A. (2009). L'acceptabilité sociale : la prise en compte des déterminants sociaux dans l'analyse de l'acceptabilité des systèmes technologiques. Le Travail Humain, 4, 383-395.

Wingard, D., Berkman, L., & Brand, R. (1982). A multivariate analysis of health related pratices: nine years mortality follow-up of the Alameda County study. American Journal of Epidemiology, 116, 76, 5-775.