Wi-Mind: Wireless Mental Effort Inference

Tilen Matkovič
University of Ljubljana
Ljubljana, Slovenia
matkovic.tilen@gmail.com

Veljko Pejović
University of Ljubljana
Ljubljana, Slovenia
veljko.pejovic@fri.uni-lj.si

UbiComp/ISWC'18 Adjunct,, October 8–12, 2018, Singapore, Singapore
ACM 978-1-4503-5966-5/18/10.
https://doi.org/10.1145/3267305.3274114

Abstract
From not disturbing a focused programmer, to entertaining a restless commuter waiting for a train, ubiquitous computing devices could greatly enhance their interaction with humans, should they only be aware of the user's cognitive load. However, current means of assessing cognitive load are, with a few exceptions, based on intrusive methods requiring physical contact of the measurement equipment and the user. In this paper we propose Wi-Mind, a system for remote cognitive load assessment through wireless sensing. Wi-Mind is based on a software-defined radio-based radar that measures sub-millimeter movements related to a person's breathing and heartbeats, which, in turn allow us to infer the person's cognitive load. We built and tested the system with 23 volunteers engaged in different tasks. Initial results show that while Wi-Mind manges to detect whether one is engaged in a cognitively demanding task, the inference of the exact cognitive load level remains challenging.

Author Keywords
wireless sensing; signal processing; cognitive load

ACM Classification Keywords
H.5.2. [Information Interfaces and Presentation (e.g. HCI)]: User Interfaces; H.1.2. [Models and Principles]: User/Machine Systems

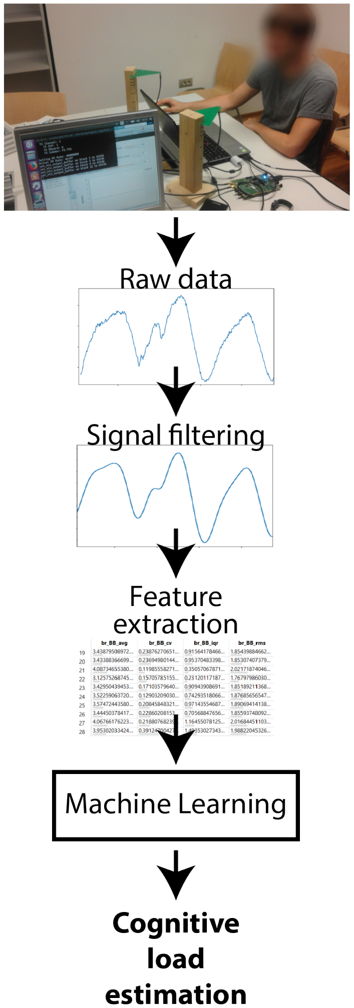

Figure 1: Wi-Mind scheme - idea for wireless cognitive load inference.

Introduction

As our reliance on ubiquitous computing devices grows, so does the need for seamless interaction with these devices. The postulates defined by Mark Weiser in 1991 call for “calm” technology that blends in with the environment, understands the user, and works towards fulfilling the user’s needs [26]. Unfortunately, almost thirty years later we are surrounded by devices that remain completely oblivious to our needs, and that contradict Weiser’s vision by getting in the way of our actual intents. Mobile communication devices are a prime example of such a conflicting technology, as an average smartphone user receives around 100 push notifications per day, most of which are disruptive [17].

Understanding a human user encompasses multiple aspects of human consciousness, from sensing one’s emotions, over inferring one’s goals, to perceiving one’s fatigue. Recent research, however, has shown the link between a user’s interruptibility and her immersion in a task at hand [18, 20], making the inference of mental effort a promising potential enabler of improved human-computer interaction. While to date research in understanding one’s mental effort has been tested mostly on intrusive methods, with notable exceptions of camera-based approaches [2, 16], here we explore the prospects of devising a *wireless* non-intrusive vital sign radar monitor to infer a user’s cognitive load. We design and implement a software-define radio-based wireless system prototype and through real-world experiments on a group of 23 volunteers evaluate its ability to sense physiological signals and through machine learning connect these to a user’s mental effort.

Background and Related Work

Inferring someones cognitive load is challenging and can be done in multiple ways, e.g. by subjective self-evaluation after completing some task or by observing the person’s performance on the task. One example for such measurement is NASA TLX (Task Load Index), where participants report their load after completing a task [13]. However, these highly subjective evaluations can be also correlated with more objective physiological signals, which are results of a human autonomic nervous system and cardiovascular system reaction. Some of these signals include heart rate blood pressure [8], heart rate variability [21], respiratory changes [10], brain activity [11], galvanic skin response (GSR) [7, 22], eye movement [11], pupil size, and facial expression [25]. These can be measured with special equipment, e.g. nasal thermistor, chest respiration strap, ECG (Electrocardiogram), sphygmomanometer (blood pressure monitor), smart watch, electroencephalography (EEG), etc. One thing in common for all these monitors is – they are intrusive, i.e. they require a body contact.

Recent advancements in technology enabled non-intrusive vital signs’ monitoring, such as camera-based approaches to measuring heart rate variability [16] and detecting pulse from head motions in a video [4]. In 2015, Adib et al. introduced Vital-Radio [3], a wireless sensing technology for monitoring breathing and heart rate without body contact that exploits the fact that wireless signals are affected by the motion in the environment. More specifically, chest movements due to human inhaling/exhaling and skin vibrations due to heartbeats can be captured by observing reflected radio waves’ phase variation. Similar wireless-based vital signs monitoring systems include TensorBeat [24], which employs channel state information (CSI) phase difference data to estimate breathing rates for multiple persons with commodity WiFi devices, an ultra-wideband (UWB) radar by Huang et al.[14], and impulse-radio (IR) UWB Doppler radar-based solutions [6, 15].

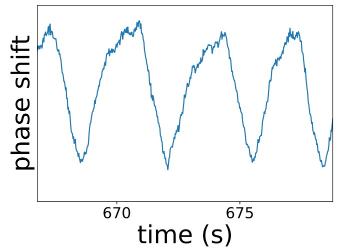

Figure 2: Wave phase shift of the reflected signal through time - persons inhales and exhales can be seen clearly.

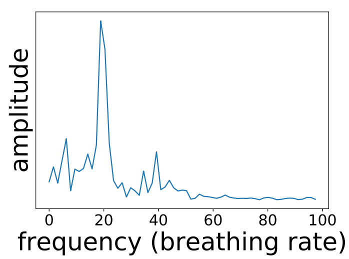

Figure 3: Frequency domain of the signal from Figure 2. The highest peak represent the highest probability of breathing rate.

In terms of applications, Zhao et al. used a technology similar to Vital-Radio, called EQ-Radio, for analysing radio frequency (RF) reflections off a person's body to recognize the emotional state [28]. To infer cognitive load unobtrusively, Abdelrahman et al. use thermal imaging cameras focused on a persons forehead and nose [2], while McDuff et al. use a five-band digital camera to detect cognitive stress [16]. While promising, the need for frontal camera placement might limit the applicability of the above approaches (e.g. for inferring a car driver's engagement). Urh and Pejovic use smartphone sensing to infer task engagement, however, their work remains at a coarser granularity as it, among other features, concentrates on location, time, and calendar events [23].

Wi-Mind System

In this paper we present Wi-Mind, a system for wireless cognitive load inference. The system is based on the premise that a person's vital signs, such as respiratory rate and heartbeat rate, correlate with that person's cognitive load. In Figure 1 we sketch the system that consists of a *wireless module* for collecting vital signs data and a *machine learning module* for inferring one's cognitive load based on the collected data. A user is stationary (seated) in an office setting and engaged in a mental task. One antenna of the wireless module is placed on the right, the other on the left side of the person (see Figure 1, top image), and are used to unobtrusively obtain data corresponding to the person's vital signals. The data is further filtered and processed, and forwarded to the machine learning module that then makes the final inference about the person's cognitive load.

Wireless monitoring

The idea for unobtrusive vital signs data collecting was taken from the already mentioned Vital-Radio system [3]. Recent advancements in CPU capabilities and signal processing algorithms have led to *software defined radio* (SDR) – a concept that enables highly customizable transmission/reception through a symbiosis of radio front-end hardware and signal processing on a general purpose computer. The core of Wi-Mind is an SDR implementation of the Frequency Modulated Carrier Wave (FMCW) radar based on a slightly modified *gr-radar* module [27] running on top of the GNUradio SDR framework [1]. This radar allows us to observe very fine movement of the user's body (predominantly chest), which may correspond to breathing and heartbeats. The hardware we use consists of an Ettus Research USRP B210 radio front-end that has two directional antennas – one for transmitting the signal to the object, the other for receiving the signal reflected off the object.

A phase shift of the electromagnetic wave sent from one antenna, reflected off the body, and captured at the other antenna, corresponds to the distance the wave has traveled. In Figure 2 we see larger phase shift variations that correspond to a person's inhale-exhale cycles, as well as smaller variations on top of these, corresponding to heartbeats. One way to obtain vital signs from the signal phase shift is to calculate the Fast Fourier Transform (FFT) of the signal and then single out the highest peak in the frequency domain. The position of the peak corresponds to the breathing rate – i.e. if the person's respiratory rate is 20 breaths per minute, then the FFT will have the highest peak at the value 20 (Figure 3). In order to cope with noise, we filter the signal with a band-pass filter. We suppress any peaks that are below 5 or above 100 breaths per minute (the average breathing rate of an adult human is around 12 to 20 breaths per minute). Slightly more difficult is the extraction of the heartbeat rate. Since heartbeats are seen as the higher frequency vibrations on top of breathing, we filter the signal with a bandpass filter from 60 to 150 beats per minute (average heartbeat rate of adult human goes

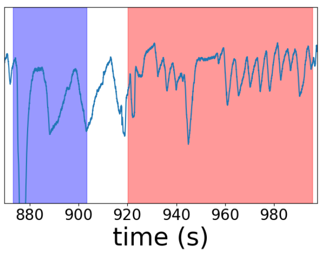

Figure 4: Acquired wireless signal while relaxing (blue) and solving some task (red). Unmarked (white) area is presents a time frame when user is clicking on "next slide" button and is not used in feature extraction.

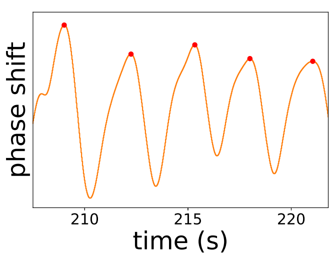

Figure 5: Filtered signal where each red dot/peak represents one breath. This data is later used to compute the inter-breath interval features.

from 60 to 100 beats per minute), and single out the highest peak within that interval.

Cognitive load inference
The machine learning (ML) module processes the breathing and heartbeat signals collected by the wireless module and infers the person's cognitive load. To train the ML model, however, we need to acquire vital signals from a person engaged in tasks of different complexity.

Wi-Mind is geared towards sedentary mental task load inference, thus, we collect the data in an office setting with an application Haapalainen et al. constructed to elicit different cognitive load burden [12]. The application runs on a PC and presents the user with six *task types*:

- **Finding hidden pattern (HP)** – find a given pattern in multiple images;
- **Finding A's (FA)** – choose all words that have a letter "A" in them;
- **Gestalt completion (GC)** – from a partial image find out what would the whole picture represents and write down the answer;
- **Number comparison (NC)** – in two parallel lists of numbers find those that are equal;
- **Scattered X's (SX)** – in a set of images find letters "X" and click on them;
- **Pursuit test (PT)** – connect values on the left side to the corresponding values on the right side following entangled lines connecting the two sides.

Each of these tasks is presented three times, with three different *difficulty levels* (e.g. task HP is performed at easy, medium and hard level). While we certainly expect that this objective label already correlates with a person's cognitive load, we also rely on the NASA-TLX questionnaire to infer a person's subjective feeling about the load. The questionnaire is presented to users after each of the tasks.

In total, there are 18 different tasks type/difficulty combinations, and from each we obtain the following:

- task complexity (objective label);
- task load index (subjective score calculated from NASA-TLX questionnaires);
- task completion time (calculated from the app logs);
- number of correct answers (calculated from the comparison with correct answers).

As users are working on the above tasks, we also collect the vital signs with the Wi-Mind wireless module (explained above). Before and after each task (task is considered as task of one difficulty and one task type) there is a short break of 30 seconds, where a user is instructed to relax (see Figure 4). Further, there is a short transition period when a user advances from the break to the task. The break and the transition ensure that physiological signs between adjacent tasks do not interfere. Inspired by [9], for each completed task we extract the following features from the wireless signals:

- **Respiratory signs**: average breathing rate, standard deviation of inter-breath intervals (see Figure 5 for inter-breath intervals), square root of the mean of the squares of differences between adjacent inter-breath intervals, interquartile rank of inter-breath intervals, average of inter-breath intervals, coefficient of variation of inter-breath intervals, ventilation I:E (inspiratory:expiratory) ratio calculated from intervals between each inhalation and exhalation, number of zero-crossings, stan-

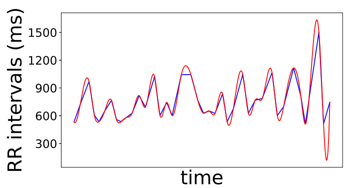

Figure 6: Heartbeat interval variability through time.

Actual \ Predicted	0	1
Relaxed (0)	287	127
Busy (1)	113	265
Accuracy (%)	70	
AUC	0.77	

Table 1: Relaxation/business detection confusion matrix. Random forest algorithm was used and tested with leave-one-participant-out validation.

dard deviation and mean value between time intervals between them. The following features were experimental, i.e. they were not mentioned in the article [9], but we still tried to find some connection between physiological signs and cognitive load: total spectral powers of the filtered signal in the following power bands: 6-12 beats, 12-18 beats, 18-24 beats, 24-30 beats; area around maximum value in frequency domain; mean value, standard deviation, median value, interquartile rank of raw/filtered signal;

- **Heartbeat signs**: average heartbeat rate, average RR interval (see Figure 6 for RR interval variability), standard deviation of RR intervals, square root of the mean of the squares of differences between adjacent RR intervals, percentages between adjacent RR intervals that are greater than *x* ms (x = 20, 50, 70), interquartile rank and coefficient of variation of RR intervals. The following features were experimental: total spectral power of the filtered signal in power bands up to 150 beats, up to 40 beats, 40-80 beats, 60-100 beats, 80-120 beats, 100-150 beats; area around maximum value in frequency domain; mean value, standard deviation, median value, interquartile rank of the raw/filtered signal.

Finally, a machine learning model is trained to predict one of the target metrics (e.g. task engagement) from the above features. In the next section we describe the preliminary results of model training and testing.

Data Collection, Machine Learning Model Construction and Evaluation

In a quiet air-conditioned room we recreated the setup from Figure 1 and ran the cognitive tasks application on a PC, while collecting wireless signals with Wi-Mind. With each participant we collected their demographics, explained the experimental protocol, and had them complete the tasks uninterrupted. The average time for completing the experiment was around 45 minutes. In total we had 23 volunteers, aging from 20 to 38, 17 male and 6 female.

To construct the ML model we use Orange, a popular data mining toolkit [5]. We extract the above respiratory and heartbeat features from the wireless signals and feed them to different classifiers (Random Forest (RF), Naive Bayes (NB) and Support Vector Machines (SVM)). The classification accuracies are evaluated through cross validation.

In the preliminary step we were curious to see whether the acquired data can at least be used to discern between a person being busy and resting. To evaluate such a basic classifier, we divide the data into *relaxing* (30 second intervals when a participant is instructed to relax) and *busy* (while solving task) time frames (see Figure 4 to get the idea for relax/busy intervals). To have equal properties and not having biased data in sense of different time intervals, we removed the intervals where users took less than 30 seconds to solve the task and included only the center 30 seconds when a user is solving a task. The ratio between relax and busy instances is 414:378 (52.3% : 47.7%). The confusion matrix results for classifying relaxed and busy time frames with a random forest-based classifier with 100 trees are shown in Table 1. We see that the classification, although far from perfect, to an extent manages to separate “relaxed" from “busy" states.

Next, we try to predict the cognitive load increase/decrease. If we look at the Figure 4 again, to mark the increase in the cognitive load, we merged the relax and busy intervals and constructed another set of features: the breathing rate difference and the heart rate variability difference between the beginning and end of the merged interval. If the user goes from the relaxed to busy state, the instance is labeled “in-

	RF	SVM	NB
Accuracy (%)	83	54	77

Table 2: Classification accuracies of “increase” or “decrease” of cognitive load. Leave-one-participant-out validation is used.

Task type	Accuracy (%) RF	SVM	NB
HP	39	43	38
FA	33	33	31
GC	38	38	39
NC	33	29	32
SX	48	50	26
PT	47	57	35

Table 3: Classification accuracies of task complexities by task type for different classifiers tested with leave-one-participant-out validation for each task separated.

creasing”, otherwise the cognitive load is “decreasing”. Instances ratio “increasing” to “decreasing” is 368:413. From the Table 2 we see that classification for the binary classification problem work well, at least with some classifiers (RF and NB). This is not surprising, as there are clearly differences in breathing rates when user is going to start solving some task versus finishing it.

The final goal of Wi-Mind is to infer the level of the user’s cognitive load. Here we assume that the complexity of the task at hand is reflected in a user’s cognitive load. While this is true, besides the task characteristics, a participant’s characteristics and the interaction between the two, also influence the expressed mental effort [19]. Thus, we do not expect our models to perfectly explain task complexity through wireless sensing.

We focus on data collected while a person was actively solving a task and try to learn the difficulty of the task. As mentioned, we have three types of task difficulties: *low*, *medium* and *high* and six types of tasks. Because we removed segments shorter than 30 seconds, the ratio between low:medium:high is the following: 27%:36.5%:36.5%. The results for the overall data classification, tested with leave-one-task-out validation, with data from all tasks bundled together, show no improvement over a baseline majority vote classifier, which has 36% accuracy. However, once we group data and build a separate classifier for each of the task types, we observe that the inference’s accuracy varies with the task type (see Table 3). The prediction is the best for the PT (pursuit test) with the average accuracy of 57% with SVM, 47% with RF and 35% with NB algorithm. The second one is the SX (Scattered X’s) with the average accuracy of 48% with RF, 50% with SVM, and 26% with NB algorithm. However, the rest of the classifiers mostly fail to outperform the baseline. In PT task, similar to GC, users have to type on keyboard which generates some noise in the wireless domain. The most likely reason for a slightly higher accuracy of classification with this task are wireless signal amplitude changes caused by extensive hand movement as the user is engaged in typing, and reflected in some of the calculated features.

Conclusions

In this paper we presented Wi-Mind, a wireless cognitive load inference system. We implemented Wi-Mind in SDR and experimental evaluated the system. The results show that Wi-Mind can, to some extent, identify whether a person is engaged in a task or not and when a user is just starting or finishing some cognitive load related task.

However, our analysis is still in early stages. Immediate improvements could include testing Wi-Mind with a higher number of volunteers or with users with different age/physical fitness, in order to make our dataset bigger and more representative, additionally filter the signal for irregular noise (e.g. limb motion), and others. Furthermore, we plan to explore different methodological paths. First, the objective task difficulty label almost certainly does not reflect the actual user engagement, nor perception – a well trained user might find all of the given tasks easy. Thus, we also plan to evaluate the ability to infer the subjective NASA-TLX index metric. Second, features we used are based on experienced from intrusive means of measuring vital signs. Wireless signal phase data we collected might hide additional features potentially related to cognitive engagement. Being feature oblivious, a convolutional neural network might represent a promising approach. Finally, our models are built on the combined data of all users. In future, we will examine models built for groups of similar users.

Acknowledgements
The authors would like to thank Eija Haapalainen, SeungJun Kim, Jodi F. Forlizzi, and Anind K. Dey from CMU for making their cognitive load measurement software available, Martin Frlin for his help with the software adaptation, and anonymous reviewers for valuable comments.

REFERENCES
1. 2018. GNU Radio. `https://www.gnuradio.org`. (2018). [Online; accessed 26-July-2018].
2. Yomna Abdelrahman, Eduardo Velloso, Tilman Dingler, Albrecht Schmidt, and Frank Vetere. 2017. Cognitive heat: exploring the usage of thermal imaging to unobtrusively estimate cognitive load. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 33.
3. Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C Miller. 2015. Smart homes that monitor breathing and heart rate. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. ACM.
4. Guha Balakrishnan, Fredo Durand, and John Guttag. 2013. Detecting pulse from head motions in video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
5. Janez Demšar, Tomaž Curk, Aleš Erjavec, Črt Gorup, Tomaž Hočevar, Mitar Milutinovič, Martin Možina, Matija Polajnar, Marko Toplak, Anže Starič, and others. 2013. Orange: data mining toolbox in Python. *The Journal of Machine Learning Research* 14, 1 (2013), 2349–2353.
6. Amy Droitcour, Victor Lubecke, Jenshan Lin, and Olga Boric-Lubecke. 2001. A microwave radio for Doppler radar sensing of vital signs. In *Microwave Symposium Digest, 2001 IEEE MTT-S International*, Vol. 1. IEEE, 175–178.
7. Mai ElKomy, Yomna Abdelrahman, Markus Funk, Tilman Dingler, Albrecht Schmidt, and Slim Abdennadher. 2017. ABBAS: an adaptive bio-sensors based assistive system. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM.
8. Tycho K Fredericks, Sang D Choi, Jason Hart, Steven E Butt, and Anil Mital. 2005. An investigation of myocardial aerobic capacity as a measure of both physical and cognitive workloads. *International Journal of Industrial Ergonomics* 35, 12 (2005), 1097–1107.
9. Martin Gjoreski, Hristijan Gjoreski, Mitja Luštrek, and Matjaž Gams. 2016. Continuous stress detection using a wrist device: in laboratory and real life. In *Mental Health workshop with ACM UbiComp'16*. Heidelberg, Germany.
10. Mariel Grassmann, Elke Vlemincx, Andreas von Leupoldt, Justin M Mittelstädt, and Omer Van den Bergh. 2016. Respiratory changes in response to cognitive load: A systematic review. *Neural plasticity* 2016 (2016).
11. Martijn Haak, Steven Bos, Sacha Panic, and LJM Rothkrantz. 2009. Detecting stress using eye blinks and brain activity from EEG signals. *Proceeding of the 1st driver car interaction and interface (DCII 2008)* (2009), 35–60.
12. Eija Haapalainen, SeungJun Kim, Jodi F. Forlizzi, and Anind K. Dey. 2010. Psycho-physiological Measures for Assessing Cognitive Load. In *ACM UbiComp'10*. Copenhagen, Denmark.

13. Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

14. Xinming Huang, Ling Sun, Tian Tian, Zeyan Huang, and Edward Clancy. 2015. Real-time non-contact infant respiratory monitoring using UWB radar. In *Communication Technology (ICCT), 2015 IEEE 16th International Conference on*. IEEE.

15. O Boric Lubecke, P-W Ong, and VM Lubecke. 2002. 10 GHz Doppler radar sensing of respiration and heart movement. In *Bioengineering Conference, 2002. Proceedings of the IEEE 28th Annual Northeast*. IEEE.

16. Daniel McDuff, Sarah Gontarek, and Rosalind Picard. 2014. Remote measurement of cognitive stress via heart rate variability. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*. IEEE.

17. Abhinav Mehrotra, Mirco Musolesi, Robert Hendley, and Veljko Pejovic. 2015. Designing Content-driven Intelligent Notification Mechanisms for Mobile Applications. In *UbiComp'15*. ACM, Osaka, Japan, 813–824.

18. Abhinav Mehrotra, Veljko Pejovic, Jo Vermeulen, Robert Hendley, and Mirco Musolesi. 2016. My Phone and Me: Understanding People's Receptivity to Mobile Notifications. In *ACM CHI'16*. San Jose, CA, USA.

19. Fred GWC Paas, Jeroen JG Van Merriënboer, and Jos J Adam. 1994. Measurement of cognitive load in instructional research. *Perceptual and motor skills* 79, 1 (1994), 419–430.

20. Veljko Pejovic, Abhinav Mehrotra, and Mirco Musolesi. 2015. Investigating The Role of Task Engagement in Mobile Interruptibility. In *Smarttention, Please! Intelligent Attention Management on Mobile Devices Workshop (with MobileHCI'15)*. Copenhagen, Denmark.

21. Mahmoud Shakouri, Laura H Ikuma, Fereydoun Aghazadeh, and Isabelina Nahmens. 2018. Analysis of the sensitivity of heart rate variability and subjective workload measures in a driving simulator: the case of highway work zones. *International Journal of Industrial Ergonomics* 66 (2018), 136–145.

22. Yu Shi, Natalie Ruiz, Ronnie Taib, Eric Choi, and Fang Chen. 2007. Galvanic skin response (GSR) as an index of cognitive load. In *CHI'07 extended abstracts on Human factors in computing systems*. ACM, 2651–2656.

23. Gašper Urh and Veljko Pejovic. 2016. TaskyApp: Inferring Task Engagement via Smartphone Sensing. In *ACM UbitTention workshop'16*. Heidelberg, Germany.

24. Xuyu Wang, Chao Yang, and Shiwen Mao. 2017. TensorBeat: Tensor decomposition for monitoring multiperson breathing beats with commodity WiFi. *ACM Transactions on Intelligent Systems and Technology (TIST)* 9, 1 (2017), 8.

25. Zixuan Wang, Jinyun Yan, and Hamid Aghajan. 2012. A framework of personal assistant for computer users by analyzing video stream. In *Proceedings of the 4th Workshop on Eye Gaze in Intelligent Human Machine Interaction*. ACM, 14.

26. Mark Weiser. 1991. The Computer for the 21st Century. *Scientific American* 265, 3 (1991), 94–104.

27. Stefan Wunsch. 2018. gr-radar - GNU Radio Radar Toolbox. `https://github.com/kit-cel/gr-radar`. (2018). [Online; accessed 20-July-2018].

28. Mingmin Zhao, Fadel Adib, and Dina Katabi. 2016. Emotion recognition using wireless signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. ACM.