This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Theme Article: Managing Attention

Toward Cognitive Load Inference for Attention Management in Ubiquitous Systems

Veljko Pejović
University of Ljubljana

Martin Gjoreski
Jožef Stefan Institute

Christoph Anderson
University of Kassel

Klaus David
University of Kassel

Mitja Luštrek
Jožef Stefan Institute

***Abstract*—From not disturbing a focused programmer to entertaining a restless commuter waiting for a train, personal ubiquitous computing devices could greatly enhance their interaction with humans, should these devices only be aware of their users' cognitive engagement. Despite impressive advances in the inference of human movement, physical activity, routines, and other behavioral aspects, inferring cognitive load remains challenging due to the subtle manifestations of users' mental engagements via vital signal reactions. These signals are traditionally captured with expensive, obtrusive, and purpose-built equipment, preventing seamless cognitive load inference for human–computer interaction adaptation. In this article, we present our achievements toward enabling large-scale unobtrusive cognitive load inference. Our approaches rely on mining sensor data collected by commodity wearable devices, and software-defined radio-based wireless radars. We also discuss further related research avenues, as well as ethical issues surrounding automatic cognitive load inference.**

Digital Object Identifier 10.1109/MPRV.2020.2968909

Over merely three days in October 2018, five children were killed in distinct car crashes at school bus stops in Indiana, Mississippi,

Authorized licensed use limited to: Institute Jozef Stefan. Downloaded on May 06,2020 at 21:23:57 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Pennsylvania, and Florida. The reasons for these crashes prompted the officials to declare a *significant problem* with distracted driving caused by the use of mobile computing and communication devices. Washington State enacted "DUI-E: driving under the influence of electronics," a bill that outlaws all use of handheld electronics behind the wheel, allowing officers to pull a driver over simply for picking up her phone. Yet, the law is unlikely to hit its intended target, as research shows that simply being aware of a recent smartphone notification, even without interacting with a mobile device, results in as much distraction and low performance in attention-demanding tasks as if a person was actively using the phone.[1]

Distractions claimed 3166 lives on American roads in 2017, yet the impact of inconsiderate interruptions initiated by pervasive computing devices is not confined to driving only. Mobile phones and instant messaging have penetrated deep into the office culture, where they cause reduced productivity and increased stress levels.[2] According to Basex's estimates, in 2010, an average knowledge worker lost 2.7 h each day on unnecessary interruptions. The economic impact of these interruptions translates to roughly $751 billion in the United States alone.[3]

The worrying evidence we provide is not a call for plunging our mobile devices into total silence. Indeed, mobile notifications represent the most appropriate means of initiating communication with a remote party and are essential for receiving timely and relevant, sometimes lifesaving, information. Instead, following Mark Weiser's vision,[4] we argue that the interaction between humans and pervasive computers should be made as seamless as possible. Despite tremendous advances in computing and sensing capabilities, our devices miss a key feature for realizing the above vision—the ability to detect when our attention is focused on a particular task.

What does it mean for our attention to be focused on something? Why do some tasks require more cognitive resources than others? Is there anything today's computers can *sense* that is somehow related to a person's cognitive engagement? Are our mathematical tools mature enough to reliably detect one's cognitive engagement purely from sensed data? Finally, can cognitive load inference be brought to masses? Can it be done unobtrusively and with cheap commodity devices? This article aims to answer the above questions and to provide guidelines for the evolution of considerate pervasive computing devices.

ATTENTION, COGNITIVE LOAD, AND PHYSIOLOGICAL RESPONSES

Ashcraft defines attention as a mental process of concentrating effort on stimuli or mental events.[5] It is essential to understand the underlying mental processes, as well as the associated cognitive resources to manage attention efficiently.

Architecture of Cognitive Processes

Anderson *et al.*'s adaptive control of thought-rational architecture (ACT-R) is an experimentally validated model that allows simulating human cognition.[6] The model is based on the idea that distinct *cognitive resources*, represented as modules, are responsible for perceiving and interacting with the surrounding environment. Each module is equipped with an individual buffer in which information is stored as chunks—single units of declarative knowledge. The procedural resource coordinates the behavior of all connected modules. To implement coherent and goal-directed behavior, it selects production rules by searching for patterns within the modules' buffers. Each module can only place a single chunk within its buffer at a time.[7] ACT-R only supports the execution of a single task at a given moment as modules that store immediate task-related information or that keep track of active goals can only represent one particular problem state or goal at a time.

The threaded cognition theory[8] extends the ACT-R architecture to model multitasking behavior by allowing multiple goals to be active and stored in the goal module's buffer. Each goal generates an associated thread—a component that comprises all processing across mental resources to accomplish a particular task. Just like in multithreaded programming, multiple threads may be active at the same time. Yet, resources are exclusive—only a single thread can be active

Authorized licensed use limited to: Institute Jozef Stefan. Downloaded on May 06,2020 at 21:23:57 UTC from IEEE Xplore. Restrictions apply.

in a particular mental resource at any moment in time. Threads claim resources as soon as possible and release them once done with the processing, thus allowing other concurrent threads to proceed. In the case of multiple threads competing for the procedural resource, the least recently processed thread will be selected. Following the concepts of the threaded cognition theory, multitasking is then accomplished by multiple concurrent threads, automatically requesting, processing, and releasing resources.

Disrupted Thoughts

Interruptions are caused by external or internal stimuli, incoming audio–visual notifications on a smartphone being an example of the former, impulsive smartphone checking being an example of the latter. Tasks signaled by these stimuli might cause conflicts with other tasks that are currently being processed. Typically, conflicts occur when two or more tasks require the same mental resource at the same time. In particular, Borst *et al.* investigate conflicts involving the *problem state* resource. The authors build upon the threaded cognition theory and examine why complex tasks seem to cause stronger disruptions than simple tasks.[7] Depending on its complexity, a task may or may not require the problem state. At the moment of interruption, if both the primary and the interrupting task are complex, the problem state of the primary task is stored in the declarative memory where it starts to decay. On the return from the interruption, the primary task's problem state has to be retrieved from the memory, yet the process is tied with delays and errors, which tend to be more detrimental the longer the problem state has lingered in the declarative memory.

Experimental results, evaluating resumption time and errors when tasks of differing complexities are interrupted, confirm this theory. They show that the negative effects of interruption are correlated with task complexity—*the disruption is minimized when the task's cognitive burden is light enough not to require the problem state*. This observation directs us to the preferred way to initiate interaction in pervasive computing, the one where *interruptions are scheduled for moments of light cognitive load*. A similar interaction approach has been proposed by Bailey *et al.* ,[9] where the authors manually constructed task models (for office-related tasks) and empirically showed that higher level task boundaries within the models correspond with states of decreased cognitive load. Instead of constructing task models to detect periods of low cognitive load in advance, in our work, we examine the ability to infer these low-load periods directly.

Table 1. Sensors used for monitoring physiological changes.

Device	Measurement
Electrocardiography (ECG) sensor	Heart activity
Galvanic Skin Response (GSR) sensor	Sweating rate
Blood volume pulse (BVP) sensor	Blood flow through vessels
Blood pressure sensor	Blood pressure
Electroencephalography (EEG) sensor	Brain activity
Electromyography (EMG) sensor	Muscle activity
Contact skin temperature (ST) sensor	Temperature of the skin
Breathing-rate sensor	Breathing rate
Infrared camera	Temperature of the skin
Camera-based eye tracker	Pupil diameter

Physiological Responses to Cognitive Load

When humans experience a psychophysiological load, e.g., in the form of a demanding task, the activation of the sympathetic nervous system increases. The increased activation substantially speeds up certain processes within the body ("fight-or-flight" response).[10] It raises the heart rate, sweating rate, breathing rate, and blood pressure; the pupils dilate; the saliva flow decreases; the heart-beats become equidistant; the blood flow is restricted from the extremities, and is redirected toward the vital organs. After the load, the sympathetic nervous system response slows down, the parasympathetic nervous system inverts the physiological changes, and initiates rest and repair processes. These physiological changes can be captured via sensors, and we list some commonly used sensors in Table 1.

Traditionally, these signals have been measured in highly controlled environments using specialized equipment. Recently, the desire for *in*

Authorized licensed use limited to: Institute Jozef Stefan. Downloaded on May 06,2020 at 21:23:57 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

situ knowledge of cognitive load has led to advances in physiological signal-based cognitive load inference using commodity devices.[11,12] In the following sections, we showcase our efforts in this direction: "Commodity Wearables for Cognitive Load Inference" section presents a smart wristband-based cognitive load inference pipeline, while the "Wireless Sensing for Cognitive Load Inference" section presents Wi-Mind, a wireless sensing system for cognitive load inference. Unique in their ways, these systems fall under a broader umbrella of recent research efforts in the field of cognitive load inference in ubiquitous computing (for a more general context, we refer the reader to the work done by Anderson *et al.*).[13]

COMMODITY WEARABLES FOR COGNITIVE LOAD INFERENCE

Approximately 70 million fitness wristbands, smartwatches, and smart garments are sold every year. With embedded arrays of sensors, these devices represent an attractive source of physiological data that may be used for cognitive load inference. However, with constant pressure to cut costs, sensors found in these devices are cheap and often unreliable. Furthermore, we must cope with the unrestricted nature of these sensors' usage—unlike with traditional lab-based equipment, users of wearable devices can walk around, make limb movements, or introduce other artifacts that render the analysis of sensor data challenging.

To assess the potential for inferring cognitive load with low-cost wearables, we fitted volunteer participants with Microsoft Band 2 fitness wristbands, immersed them in situations designed to elicit different cognitive load, and, via the wristband, collected physiological data during the experiment. We then constructed models that aim to predict users' cognitive engagement based on the collected data.

Stimulating Cognitive Response and Collecting Physiological Data

Our experiments are designed to replicate sedentary work in an office, as we believe that such situations benefit the most from intelligent attention management. The experiments were performed in a quiet, normal-temperature room where a single participant was working on a PC with the wristband strapped to her nondominant arm. The total of 20 participants attended these experiments. The wristband measured: intervals between successive heartbeats (RR intervals), galvanic skin response (GSR, sampled at 1 Hz), skin temperature (ST, sampled at 1 Hz), and accelerometer data (sampled at 8 Hz).

The study comprised of six cycles of cognitive-load tasks. For each cycle, three variations of a randomly selected cognitive-load task type were presented on the PC. The tasks, akin to puzzles, were designed and previously validated by Haapalainen *et al.*[14] and include the following: *1)* Gestalt Completion (GC) test, where the subject is asked to identify incomplete drawings; *2)* Hidden Pattern (HP) test, where the subject has to decide whether a model image is hidden in other comparison images; *3)* Finding A's (FA) test, where the subject has to find the letter "a" in presented words; *4)* Number Comparison (NC) test, where the subject has to decide whether or not two displayed numbers are the same; *5)* Pursuit test (PT), where the subject has to visually track irregularly curved overlapping lines from numbers on the left side of a rectangle to letters on the opposite side; and *6)* Scattered X's (SX) test, where the subject has to find the letter "X" on screens containing random letters. The variations differed in the *designed* difficulty (i.e., easy, medium, and difficult) and thus in the expected cognitive load they elicit. After solving each of the three variations, the participants filled the NASA-TLX questionnaire to assess the *subjective* cognitive load posed by the tasks.

Physiological Data Processing and Feature Engineering

For the analysis, we take 60-s segments before each NASA-TLX questionnaire. Tasks of shorter duration were disregarded, as the delay between the nervous system commands and the corresponding physiological responses may prevent reliable detection by wearable sensors. We then filter the segmented data and extract features from each segment.

Sweating is one of the most characteristic responses to increased cognitive load. We concentrate on data collected by the wristband's GSR sensor, filter it using a sliding mean filter,

Authorized licensed use limited to: Institute Jozef Stefan. Downloaded on May 06,2020 at 21:23:57 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

and then extract the fast-acting (GSR responses) component and the slow-acting (tonic) component from the filtered signals. The fast-acting component is used to calculate the number of responses in the signal, the responses per minute in the signal, and the sum of the responses. The slow-acting component is used to calculate the derivative of the tonic component, and the difference between the tonic component and the overall signal. In addition, from the filtered GSR signal, we calculate: mean, standard deviation, 1st and 3rd quartile, quartile deviation, derivative of the signal, sum of the signal, sum of positive derivative, proportion of positive derivative, total spectral power of the signal in five frequency bands between 0 and 0.5 Hz with a 0.1 Hz span. All of these features have been analyzed in related studies on stress, emotions, and cognitive load monitoring.[15]

The activation of the subject's sympathetic nervous system triggered by cognitive load also "stabilizes" heart beating, leading to more uniform RR intervals. On the other hand, the rest periods between the tasks reverse this process, and the RR intervals become less regular, since "A healthy heart is not a metronome."[16] Heart rate variability analysis (HRV) is commonly used to quantify the dynamics of the RR intervals. The details on 12 HRV features we extract can be found in our earlier work.[12]

Finally, from the ST signal and the magnitude of the acceleration signal, we extract six statistical features: mean, standard deviation, quartile deviation, derivative of the signal, coefficient of variation, and difference between the maximum and minimum value in a segment.

Inferring Cognitive Load

The feature extraction process represents each segment by a set of features ready to be fed into machine learning (ML) algorithms. In our preliminary analysis,[12] we discovered that different task types require different models that connect physiological responses to task difficulty. This observation poses additional questions about the types of cognitive load elicited by different task types, and we plan to inspect it further in our future work. For now, however, our first experiment focuses on task-specific models aiming to predict a subjective measure, the NASA-TLX, and an objective measure, the designed task difficulty (easy/medium/difficult) of the task at hand. The best performing designed task difficulty predictor, a Naive Bayesian model, achieved precision, recall, and accuracy of about 50%, which is an improvement compared to the majority classifier's 33% accuracy.[12] While these figures are not very high, it is encouraging that the model tends to confuse the neighboring labels, i.e., easy–medium and medium–difficult more than distant labels (i.e., easy–difficult).

Interaction in the pervasive computing domain could be improved even if the devices could not reliably detect our level of cognitive engagement, but could only understand whether we are cognitively engaged at all. Therefore, in our second experiment, we construct models to discern between moments when users were solving tasks and moments when users were explicitly told to rest. Furthermore, we compare personalized models with general models. The personalized models were evaluated using leave-one-task-out for each person specifically. The general models were evaluated using leave-one-subject-out. Thus, they are completely person-independent. We report the results achieved by the Random Forest (RF) algorithm, since it is the most stable ML algorithm among those tested (Random Forest—RF, Support Vector Machine—SVM, Gradient Boosting—GB, AdaBoost—AB, KNN, Gaussian Naive Bayes—NB, and Decision Tree—DT). The results are presented in Table 2.

Table 2. Precision and recall of task versus rest classification for personalized and general models.

	Model			
Task	Personalized RF		General RF	
	Prec.	Rec.	Prec.	Rec.
HP	.72	.71	.73	.73
FA	.77	.78	.80	.80
GC	.61	.58	.47	.47
NC	.68	.70	.73	.72
SX	.71	.71	.78	.77
PT	.57	.59	.67	.65
Average	.68	.68	.70	.69

Authorized licensed use limited to: Institute Jozef Stefan. Downloaded on May 06,2020 at 21:23:57 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Interestingly, the general RF performs slightly better than personalized RF. This is probably due to the size of the training data. The highest precision and recall of 80% is achieved by the general RF for the FA task.

The Microsoft Band 2 is equipped with an array of sensors, but which of these sensors are the most promising when it comes to cognitive load inference? We retrained the General RF model for each sensor separately and measured the F1-score. The highest F1-score of 65% was achieved by the model trained with RR-related features, followed by acceleration- (F1-score of 63%), GSR- and skin temperature-related features (both with an F1-score of 55%). Therefore, detecting heart activity appears to be the most promising way forward for cognitive load inference using wearable sensors.

The experimental results are also influenced by the size of the dataset, the noise in the sensor data, and the physiological characteristics of the participants. To ameliorate these influences, one might increase the number of participants, rely on transfer learning methods to reuse labeled data from similar studies, and further examine the role of participants' physiological and psychological traits on the results.

WIRELESS SENSING FOR COGNITIVE LOAD INFERENCE

The physiological reaction elicited by the autonomic nervous system when a person is under cognitive load also manifests through body movements. For instance, intensified breathing leads to faster chest movement, while heartbeats lead to minute movements of the breastbone. *Wi-Mind*, a system developed at the University of Ljubljana, uses a software-defined radio (SDR) radar to unobtrusively detect these movements and infer one's cognitive load through an ML pipeline.

Detecting Movement

Wi-Mind's wireless monitoring module is based on Vital-Radio, a solution that uses frequency modulated carrier wave (FMCW) radar to pinpoint transmitted signal reflections corresponding to motion coming from a person.[17] It then uses phase changes in the reflected signal

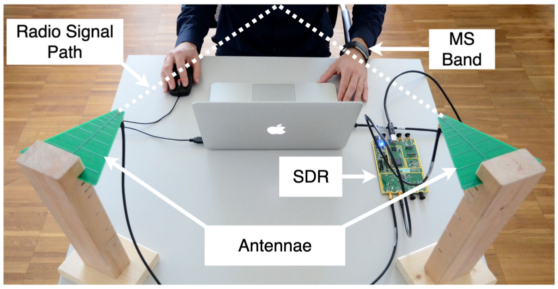

Figure 1. Wi-Mind in action.

to detect small movements of the person's body. Our signal processing implementation consists of a modified gr-radar extension for GNUradio open-source SDR framework running on a general-purpose computer. A commodity USRP B210 SDR front-end board with two antennas, one for transmission, and the other for the reception of the signal reflected off a person's body (see Figure 1) is connected to the PC. The raw signal corresponding to the phase of the reflected radio waves flowing from the USRP to the PC contains the information about the movements and is funneled into Wi-Mind's ML pipeline.

Extracting Breathing and Heartbeat Features

Wi-Mind extracts *breathing-related features* by first filtering the raw signal to focus on movements that may be caused by breathing. A normal breathing rate for an adult at rest is between 6 to 31 breaths per minute[18] and is modulated by task engagement, age, and other factors. To ensure that even the lowest breathing rates are not filtered out, yet to account for slow signal drifts due to posture changes, we use an empirically derived 0.083 Hz (corresponding to five breaths per minute) lower bound for the pass-band filter to single out breathing. With the upper bound of the filter, we aim to exclude heart beat-related interference. The normal resting heart rate frequency starts at 1 Hz[19] and is reflected in signal changes that are an order of magnitude smaller amplitude that breathing-induced changes. Consequently, we decided to use 1 Hz as the upper bound for the pass-band filter that singles out breathing.

From the filtered signal, we first calculate the Fast Fourier Transform (FFT), extract energy in different spectral bands, and then single out the highest peak in the frequency domain—this

Authorized licensed use limited to: Institute Jozef Stefan. Downloaded on May 06,2020 at 21:23:57 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

corresponds to the breathing rate. Furthermore, we calculate the difference between the average breathing rate in the first half and the average breathing rate in the second half of a time window. The change might indicate the start or the end of solving a mental task. We augment the above process with time-domain features, such as respiratory rate variability. For this, we use a peak detector on the filtered signal to pinpoint the times at which inhales happen, and then measure the interbreath interval. Finally, we calculate statistical measures from the raw signal, including mean, median, standard deviation, and root mean square value, as these features have been shown to be relevant for the inference of high-level features from vital signals.[20]

Heartbeat-related features are produced in a similar fashion. Yet, they are less reliable and remain more difficult to detect than breathing-related features due to much smaller chest movements. Wi-Mind first filters the raw signal, this time keeping frequencies between 0.83 and 2.5 Hz, corresponding to 50 and 150 heartbeats per minute. Unlike with the breathing rate, the highest peak of the filtered frequency domain signal need not correspond to the heart rate. Spillover from the much stronger breathing signal might lead to the second highest peak actually corresponding to the heart rate. Besides the average heart rate, we calculate the difference between the rate at the beginning and the end of the time window, and HRV features. We extract these from the time domain signal, using a peak detector to identify heartbeats, and then calculate the RR intervals, their variability, and other related statistics (further details on extracted features can be found in our previous work).[21]

Inferring Cognitive Load

The inference algorithm is the final part of Wi-Mind's ML pipeline. To build and test it, we conducted experiments with 23 participants exposed to the same set of tasks of different difficulties as described in the "Commodity Wearables for Cognitive Load Inference" section. As users were solving the tasks seated in front of a PC, Wi-Mind was transmitting and receiving the reflected wireless signals and calculating the above breathing and heartbeat-related features. We connected these features with information about a participant's task engagement. The participant was either engaged in a task of a particular objective difficulty (i.e., easy, medium, and difficult) or resting between the tasks.

The first experiment investigated whether wireless sensing can be used to detect a person's cognitive engagement. We used the same experiment of six cycles of cognitive load tasks interspersed with rest periods as in the study described in the "Commodity Wearables for Cognitive Load Inference" section. We divided the sensed data into periods when a user was working on a task and periods when a user was resting. We trained different binary classifiers (KNN, SVM, RF, and NB) to infer whether a user is busy or resting, using leave-one-subject-out approach. RF built upon the above-explained features provides the highest inference accuracy—70%, significantly above the 50% baseline. A closer inspection of the data reveals notable differences among individuals. For example, one person‘s resting breathing rate may fluctuate between 7 and 13 breaths per minute, while another person's resting rate may fluctuate between 8 and 22 breaths per minute. Consequently, normalizing breathing rate in a way that each user's resting rate is the same leads to slightly improved inference results, with RF achieving 75% accuracy.

The second experiment investigated whether we can detect the level of cognitive engagement for a person who was actively working on a mental task. A preliminary analysis of the subjective feeling of cognitive load, measured via the NASA-TLX questionnaire, indicated that designed difficulties of different task types do not elicit difficulty perceptions that are comparable across tasks. For instance, GC task‘s medium difficulty was often perceived to be more difficult than HP task's high difficulty instance.* Therefore, a common model for the designed difficulty prediction is bound to fail, and we build a separate model for each of the six task types.

The results of leave-one-subject-out cross-validation shown in Table 3 reveal varying predictability potential of different task types (between 36% and 44% accuracy, compared to 33% baseline),

*We note that similar was observed in our study with wearables (see the "Commodity Wearables for Cognitive Load Inference" section)

Authorized licensed use limited to: Institute Jozef Stefan. Downloaded on May 06,2020 at 21:23:57 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Table 3. Accuracy for binary and ternary task difficulty classification with Wi-Mind for different task types. The best performing model for the given task type is in the parentheses.

	Accuracy	
Task	Ternary (Easy/Med/Diff)	Binary (Easy/Diff)
HP	.36 (SVM)	.58 (RF)
FA	.36 (KNN)	.53 (KNN)
GC	.40 (NB)	.60 (RF)
NC	.44 (KNN)	.65 (RF)
SX	.38 (RF)	.59 (KNN)
PT	.37 (SVM)	.58 (RF)

with the highest success observed with the NC task. We also attempted to answer a toned down research question—can we discern among the extremes of cognitive engagement and detect whether a person was solving an easy or a difficult version of the same task. The results show that for certain tasks, such as NC, this is possible with slightly more than 65% accuracy (cf., 50% baseline).

> "The ability to focus attention on important things is a defining characteristic of intelligence."
> *Robert J. Shiller*

Wi-Mind's ML pipeline relies on carefully engineered features about breathing and heartbeats. Despite this, other features might provide additional information that could lead to higher accuracy of cognitive load inference. A neural network trained on raw wireless phase signal alleviates the need for crafting individual features and may implicitly take into account phenomena that are difficult to model via features, yet potentially relevant for cognitive load inference (e.g., sighing). We construct a long short term memory network and train it with raw signals corresponding to a person resting or working on a task. The network yields 75.2% accuracy of cognitive engagement prediction, which is on par with the best performing classical machine learning algorithm—RF.

ROAD AHEAD

Our attention is precious, so everyone competes for it: from ubiquitous applications on our phones and wearables, to email, instant messaging, TV, billboards and of course, fellow humans. However, few of us manage the attention well, possibly because evolution has not equipped us to deal with so many stimuli. So rather than to divert our attention, we should harness the power of ubiquitous computing to help us manage it. We envision future technology that will assess both the level of cognitive load and the type of mental resources currently engaged. The technology will then evaluate incoming messages and tasks, and decide when and whether to notify us about each so that we will be able to focus on what matters.

In this article, we discuss our recent efforts representing the initial steps toward efficient attention management. We recognize the link between physiological signals and a user's cognitive load, and build two systems, a wearable and a wireless one, that unobtrusively infer cognitive load. Compared to the state-of-the-art solutions,[14] our wearable-based system relies on cheap commodity devices, whereas our Wi-Mind wireless system moves away from any physical contact between the user and the equipment. In terms of the absolute performance, the systems we have developed are on par with those described in the literature,[22] yet less accurate than those that rely on specialized equipment and lab settings. Haapalainen *et al.* achieve around 80% accuracy in two-class classification,[14] whereas our wearable and wireless sensing achieve 60% accuracy for easy/difficult and 75% accuracy for rest/busy classification. The accuracy gap might stem from the modalities used as[14] relies on ECG and heat-flux-related features. Combining our approaches with thermal imaging[23] represents a potentially promising avenue for future research on unobtrusive cognitive load inference.

Despite the limited accuracy, the presented systems could already enhance a number of applications. For instance, being able to predict when someone is starting or finishing a task, e.g., doing homework, calculating a spreadsheet, etc., even with 75% accuracy could lead to improved notification management systems that, rather than interrupting indiscriminately, aim to find natural breaks as the most suitable moments for information delivery.[9]

Authorized licensed use limited to: Institute Jozef Stefan. Downloaded on May 06,2020 at 21:23:57 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Understanding Where Our Attention Is

The systems described in this article, although demonstrating the potential for unobtrusive inference with commodity devices, still have some shortcomings with respect to the accuracy of cognitive load detection. A comparison with lab studies conducted with specialized equipment points to a modest gap between the vital signs inference accuracy of these and our approaches. Introduction of new sensors, such as EEG, computer vision to capture facial expressions, and multimodal processing (e.g., using acceleration data to detect movements, and then account for these when extracting heartbeats from the PPG sensor)[24] could bridge this gap.

Inferring cognitive engagement remains challenging even with vital signs detected perfectly. Heartbeats, breathing and sweating do not only proxy one's task engagement, but also emotions, stress, physical activity, thermal regulation, and other aspects. Furthermore, cognitive load is a complex concept that reflects a task's inherent complexity (i.e., intrinsic load), the complexity of the task's representation (i.e., extraneous load), and the complexity of constructing the schema of the task (i.e., germane load). It may well be that accurate assessment of cognitive load requires taking into account the nature of the task at hand and adjusting the inference methodology accordingly. For instance, the highest inference accuracy achieved by Wi-Mind, a wireless system relying on breathing and heartbeat-related features, happens when users are solving the number comparison task. The wearable approach elaborated in the "Commodity Wearables for Cognitive Load Inference" section shows most promise when users are solving the finding 'A's task.

Measuring Required and Available Mental Resources

In computers, multitasking is managed by schedulers, which allocate computational resources based on the estimated task duration and resource availability. Should we wish that ubiquitous computing manages our real-world multitasking in a similar fashion, we need to empower it with the understanding of the mental resource requirements of real-world tasks. A great deal of (meta) data might be needed for predicting the complexity of an incoming task. For example, an e-mail from a supervisor might request a student to proofread a section of a research paper. The system needs to infer which, and to what extent, procedural, declarative, and perceptual resources are needed to read the e-mail. Before deciding whether to interrupt the student or not, the system should also understand the user's cognitive capacities, current cognitive engagement, and assess the potential for multitasking, among other things. ACT-R and the threaded cognition model (see the "Attention, Cognitive Load, and Physiological Responses section) present a solid foundation on which a holistic model providing a continuous picture of one's cognitive engagement could be constructed. Modeling cognitive load requires that a system maintain a computational representation of ongoing mental processes and the resources they occupy. A detailed model might even predict bottlenecks in mental resource allocation, thus ensuring that users are not overburdened with conflicting tasks.

Smooth and Ethical Attention Steering

We envision a future in which attention will be steered by subtle cues, rather than grabbed by buzzing notifications. Along which course should the attention be steered? Arguably, users should be in *the flow*—fully immersed in the task at hand, absorbed by the activity, not bored, but not frustrated either. The flow, however, defines the target level of cognitive engagement, not the content. Games and online social networking applications are very good at getting their users in the flow; yet, the overall effect on an individual's long-term well-being and productivity can be negative. Should effective and reliable attention management technology be available, it is easy to see that it would be particularly valuable at the workplace. But should we let our employers decide how to manage our attention? Perhaps to a degree, but ubiquitous technology makes it all too easy for work to intrude in our private lives. Attention is the most precious resource, and users should have the final choice about the purpose for which this resource is going to be used.[25] Consequently, we call for further debate on where and to what extent attention should be steered, as well as on how the steering can be done in the least intrusive, yet the most transparent and ethical way.

Authorized licensed use limited to: Institute Jozef Stefan. Downloaded on May 06,2020 at 21:23:57 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

ACKNOWLEDGMENT

The authors would like to thank T. Matkovič and M. Frlin for their help with the presented solutions' implementations and for conducting user studies. This work was supported by the European Union's Horizon 2020 Research and Innovation Programme under Grant agreement no. 815141 (DECENTER: Decentralised technologies for orchestrated Cloud-to-Edge intelligence).

References

1. C. Stothart, A. Mitchum, and C. Yehnert, "The attentional cost of receiving a cell phone notification," *J. Exp. Psychol., Human Perception Perform.*, vol. 41, no. 4, pp. 893–897, Jun. 2015.
2. G. Mark, D. Gudith, and U. Klocke, "The cost of interrupted work: More speed and stress," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, Apr. 2008, pp. 107–110.
3. J. B. Spira and J. B. Feintuch, "The cost of not paying attention: How interruptions impact knowledge worker productivity," Report Basex, 2005.
4. M. Weiser, "The computer for the 21st century," *Scientific Amer.*, vol. 265, no. 3, pp. 94–104, Sep. 1991.
5. M. H. Ashcraft, *Cognition*, 4th ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2006.
6. J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Qin, "An integrated theory of the mind," *Psychological Rev.*, vol. 111, no. 4, pp. 1036–1060, Oct. 2004.
7. J. P. Borst, N. A. Taatgen, and H. van Rijn, "The problem state: A cognitive bottleneck in multitasking," *J. Exp. Psychol., Learn., Memory, Cognition*, vol. 36, no. 2, pp. 363–382, Mar. 2010.
8. D. D. Salvucci and N. A. Taatgen, "Threaded cognition: An integrated theory of concurrent multitasking," *Psychological Rev.*, vol. 115, no. 1, pp. 101–130, Jan. 2008.
9. B. P. Bailey and S. T. Iqbal, "Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management," *ACM Trans. Comput.-Hum. Interact.*, vol. 14, no. 4, pp. 21:1–21:28, Jan. 2008.
10. W. B. Cannon, *The Wisdom of the Body*. New York, NY, USA: W. W. Norton and Company, Inc., 1932.
11. R. Rajan, T. Selker, and I. Lane, "Task load estimation and mediation using psycho-physiological measures," in *Proc. 21st Int. Conf. Intell. User Interfaces*, Mar. 2016, pp. 48–59.
12. M. Gjoreski, M. Luštrek, and V. Pejovic, "My watch says I'm busy: Inferring cognitive load with low-cost wearables," in *Proc. Ubittention Workshop*, Oct. 2018, pp. 1234–1240.
13. C. Anderson, I. Hübener, A.-K. Seipp, S. Ohly, K. David, and V. Pejovic, "A survey of attention management systems in ubiquitous computing environments," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 2, pp. 58:1–58:27, Jul. 2018.
14. E. Haapalainen, S. Kim, J. F. Forlizzi, and A. K. Dey, "Psycho-physiological measures for assessing cognitive load," in *Proc. 12th ACM Int. Conf. Ubiquitous Comput.*, Sep. 2010, pp. 301–310.
15. A. Alberdi, A. Aztiria, and A. Basarab, "Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review," *J. Biomed. Inform.*, vol. 59, pp. 49–75, Feb. 2016.
16. F. Shaffer, R. McCraty, and C. L. Zerr, "A healthy heart is not a metronome: An integrative review of the heart's anatomy and heart rate variability," *Frontiers Psychol.*, vol. 5, 2014, Art. no. 1040.
17. F. Adib, H. Mao, Z. Kabelac, D. Katabi, and R. C. Miller, "Smart homes that monitor breathing and heart rate," in *Proc. 33 rd Annu. ACM Conf. Human Factors Comput. Syst.*, Apr. 2015, pp. 837–846.
18. G. Benchetrit, "Breathing pattern in humans: Diversity and individuality," *Respiration Physiol.*, vol. 122, no. 2/3, pp. 123–129, Sep. 2000.
19. P. Palatini, "Need for a revision of the normal limits of resting heart rate," *Hypertension*, vol. 33, no. 2, pp. 622–625, Feb. 1999.
20. C. Maaoui, A. Pruski, and F. Abdat, "Emotion recognition for human-machine communication," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2008, pp. 1210–1215.
21. T. Matkovič and V. Pejovic, "Wi-Mind: Wireless mental effort inference," in *Proc. Ubittention Workshop*, 2018, pp. 1241–1249.
22. R. L. Charles and J. Nixon, "Measuring mental workload using physiological measures: A systematic review," *Appl. Ergonom.*, vol. 74, pp. 221–232, Jan. 2019.
23. Y. Abdelrahman, E. Velloso, T. Dingler, A. Schmidt, and F. Vetere, "Cognitive heat: Exploring the usage of thermal imaging to unobtrusively estimate cognitive load," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 3, Sep. 2017, Art. no. 33.

Authorized licensed use limited to: Institute Jozef Stefan. Downloaded on May 06,2020 at 21:23:57 UTC from IEEE Xplore. Restrictions apply.

24. M. B. Mashhadi, E. Asadi, M. Eskandari, S. Kiani, and F. Marvasti, “Heart rate tracking using wrist-type photoplethysmographic (PPG) signals during physical exercise with simultaneous accelerometry,” *IEEE Signal Process. Lett.*, vol. 23, no. 2, pp. 227–231, Dec. 2015.
25. H. A. Simon, *Designing Organizations for an Information-Rich World*. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 1971, pp. 37–72.

Veljko Pejović is currently an Assistant Professor with the Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia. His research interests include mobile computing, HCI, and resource-efficient computing. He received the Ph.D. degree in computer science from the University of California Santa Barbara, CA, USA. He is the corresponding author of this article. Contact him at veljko.pejovic@fri.uni-lj.si.

Martin Gjoreski is currently working toward the Ph.D. degree with the Department of Intelligent Systems, Jožef Stefan Institute, Ljubljana, Slovenia. His research focuses on monitoring human psychophysiological states using wearables. He received the M.Sc. degree in information and communication technologies from the Jožef Stefan Postgraduate School, Ljubljana, Slovenia. Contact him at martin.gjoreski@ijs.si.

Christoph Anderson is currently working toward the Ph.D. degree with the Department for Communication Technologies, University of Kassel, Kassel, Germany. His research focuses on the recognition of human attention and interruptibility, particularly through interdisciplinary approaches. He received the M.Sc. degree in computer science from the University of Kassel. Contact him at anderson@uni-kassel.de.

Klaus David is currently a Full University Professor and Head of the Department for Communication Technologies, University of Kassel, Kassel, Germany. His research interests include mobile networks, applications, and context awareness. He received the Ph.D. degree in physics from the University of Siegen, Siegen, Germany. He is a Member of the IEEE, ACM, and GI. Contact him at david@uni-kassel.de.

Mitja Luštrek is currently the Head of the Ambient Intelligence Group, Department of Intelligent Systems, Jožef Stefan Institute, Ljubljana, Slovenia. His research focuses on applying machine learning to sensor and other data in the domains of health, well-being, and ambient assisted living. He received the Ph.D. degree in computer and information science from the University of Ljubljana, Slovenia. Contact him at mitja.lustrek@ijs.si.

Authorized licensed use limited to: Institute Jozef Stefan. Downloaded on May 06,2020 at 21:23:57 UTC from IEEE Xplore. Restrictions apply.