Article

Using Psychophysiological Sensors to Assess Mental Workload in Web Browsing

Angel Jimenez-Molina[1,] *, Cristian Retamal[2] and Hernan Lira[3]

[1] University of Chile, Faculty of Physical and Mathematical Sciences, Department of Industrial Engineering; ajimenez@dii.uchile.cl
[2] University of Chile, Faculty of Physical and Mathematical Sciences, Department of Electrical Engineering; cristian.retamal@ug.uchile.cl
[3] University of Chile, Faculty of Physical and Mathematical Sciences, Department of Industrial Engineering; hlira@ing.uchile.cl
* Correspondence: ajimenez@dii.uchile.cl; Tel.: +56-2-2978-4056

Abstract: The mental workload induced by a Web page is essential for improving the user's browsing experience. However, continuously assessing the mental workload during a browsing task is challenging. In order to face this issue, this paper leverages the correlation between stimuli and physiological responses, which are measured with high-frequency, non-invasive psychophysiological sensors during very short span windows. An experiment was conducted to identify levels of mental workload through the analysis of pupil dilation measured by an eye-tracking sensor. In addition, a method was developed to classify real-time mental workload by appropriately combining different signals (electrodermal activity (EDA), electrocardiogram, photoplethysmography (PPG), electroencephalogram (EEG), temperature and eye gaze) obtained with non-invasive psychophysiological sensors. The results show that the Web browsing task involves on average four levels of mental workload. Also, by combining EEG with the PPG and EDA, the accuracy of the classification reaches 95.73 %.

Keywords: psychophysiological sensors; mental workload; Web browsing tasks; signal processing

1. Introduction

Although Web applications are often justified in terms of increasing the productivity of human tasks, they sometimes have the opposite effect, interrupting, reducing the performance of, or increasing the mental workload of the user [1–4]. A typical task in which this phenomenon may occur is Web browsing. In this task, the user fixes her/his gaze on and between Web elements, i.e., graphic or textual areas of a Web page, such as news, commercial advertisements, and menus [5–7]. In cognitive psychology, mental workload refers to the total amount of perceived mental effort used for learning or processing new information [8–11].

An important factor in measuring the effectiveness of a Web page is the user's browsing experience. It has been shown that the higher the level of user's browsing experience is, the lower the mental workload [3,4,12]. Every Web page has both an intrinsic and an extrinsic mental workload [3,13–15]. The former is related to the natural effort required to absorb new information, to the process of learning to navigate around the page, and to the process of becoming accustomed to the design of the page. The latter consists of the mental workload caused by the inclusion of unnecessary details or external interruptions, such as font styles that convey no meaning, commercial advertisement pop-ups, and irritating recommendations, which may have a negative effect on user's browsing experience.

Continuously assessing, at any moment, the mental workload involved in browsing tasks entails measuring it either when the user fixes her attention on a Web element or when her gaze switches from one element to another. This assessment of mental workload can enhance the user's

browsing experience in many ways: for instance, avoiding extrinsic mental workload by automatically identifying the most suitable moments to proactively deliver content to the user or preventing irritating intrusions from the environment; reducing intrinsic mental workload by keeping the Web page support interventions on stand-by and adapting graphic user interfaces in real time; and evaluating the likelihood of user's abandonment, frustration or techno stress, among other benefits. In addition, instantaneous classification of mental workload into intrinsic or extrinsic to the Web elements of a Web page would make it possible to detect short time windows of reduced cognitive burden to activate the delivery of different types of recommendations in a timely, unobtrusive manner, such as contextual news in newspaper portals or commercial advertisement pop-ups on various Web sites. In addition, it may be possible to enhance search tasks, for instance, for restaurants, flight tickets, or retail products, by providing relevant feedback to the search engine based on the user's cognitive status [6].

To realize the above requirements, it is essential to address the challenge of automatically assessing the mental workload in a continuous fashion while the user is engaged in browsing, that is, in real time, with high frequency and using very short time windows.

Many studies have focused on classifying mental workload in general by capturing and processing data using ever less invasive psychophysiological sensors [16–20]. This method is founded on the empirical demonstration of the correlation existing between psychological stimuli and physiological responses triggered by the nervous system. Moreover, mental workload has been shown to vary frequently within a short time span [21,22].

Although considerable research has been devoted to assessing mental workload on the scale of hours and minutes by using data extracted from psychophysiological sensors, less attention has been paid to time windows lasting seconds or less, such as when a user fixes her gaze on a Web element. Indeed, Bailey et al. [23] have recently proved that moments of reduced mental workload occur while the user's attention is transiting from one task to another. However, this was shown only for coarse-grained tasks, such as selecting a travel route among alternatives presented in a graphic interface or classifying a list of emails into various categories [23].

In this paper, the capabilities of psychophysiological sensors are leveraged to research the possibility of assessing mental workload in real time during a browsing task. This paper thus attempts to answer the following research questions:

- RQ1: Is it possible to identify levels with regard to a user's mental workload within very short time windows (order of milliseconds) based on psychophysiological signals recorded during a Web browsing task?
- RQ2: Is it possible to accurately classify in real time a user's mental workload, both when her gaze is fixed on a Web element and when her gaze is transiting from one Web element to another, by combining different non-invasive psychophysiological sensors?

In addition, based on the findings of Bailey et al. [23], this paper attempts to prove the following hypothesis:

- H1: Mental workload is significantly smaller when the user's attention is switching from one Web element to another than when she is focused on a Web element.

To answer these research questions and prove the stated hypothesis, an experiment was conducted in which 61 users performed a normal Web browsing task in front of a computer screen while their psychophysiological responses were measured by different sensors and recorded in a database. The gold standard with regard to answering RQ1 is pupil diameter because several previous studies have shown that, under controlled illumination conditions, this psychophysiological response is a valid and reliable indicator of mental workload [23–29]. Using clustering methods, this paper shows that, by processing the pupil dilation response, four levels of mental workload can be identified per user on average.

However, measuring pupil dilation with an eye tracker is not a realistic and practical method to classify mental workload, for example, in the open air, because it requires constant and controlled illumination conditions. Thus, in this paper, more practical and less invasive sensors are assessed to measure other psychophysiological responses, such as heart rate (HR), electrodermal activity (EDA), body temperature, and electrocardiogram (ECG). The electroencephalogram (EEG) sensor is also assessed because there have been important advances in the construction of portable EEGs and in algorithms to reduce motion-related artifacts [30] [31]. It is expected that before long, there will be EEG devices that only capture brain waves from the areas of the brain relevant to the assessment of mental workload, making them less invasive [32].

This paper shows that, using all the sensors and efficiently processing their signals using artificial neural networks, mental workload can be classified as proposed in RQ2, with 68.94 % accuracy, 66.62 % recall, and 76.92 % precision. However, using all the sensors and a multi-layer perceptron, it is possible to achieve 88.46 % accuracy, 88.84 % recall, and 88.85 % precision. Ultimately, the best performance is obtained by combining EDA, HR, and EEG, achieving 95.73 % accuracy, 94.25 % recall, and 95.6 % precision in the classification of mental workload. Furthermore, the hypothesis that mental workload is significantly smaller when the user's attention is switching from one Web element to another than when she is focused on a Web element is confirmed ($MSE = 1.7829$; $p - value = 0.00184 < 0.05$).

The contributions of this paper include (i) identifying different levels of mental workload required for Web browsing through the processing and analysis of pupil dilation measured by an eye-tracking sensor; (ii) developing a method for appropriately combining non-invasive psychophysiological sensors to classify real-time mental workload in small time windows with high accuracy (mean=99.1%, SD = 0.2772%) based on the behavior of the user's gaze in a Web browsing task; and (iii) leaving open the possibility of using gaze shifts from one Web element to another as the most appropriate time to provide the user with recommendations, for example.

This paper is organized as follows. Section 2 provides the background required to understand this research. Section 3 presents the related literature. The experiment conducted is described in Section 4, as well as the data processing and the machine learning methods applied to the data. The results are presented in Section 5 and are discussed in Section 6, while Section 7 concludes the paper.

2. Background

2.1 Assessment Methods

Cognitive resources are assets used by cognition to think, remember, make decisions, solve problems, or coordinate movements, such as perception, attention, short- and long-term memory, and motor control [33,34]. According to Navon et al. [35], these resources underlying human learning and information processing are limited [36].

Wickens [9], in his multiple resource theory, suggests that these resources can be used in parallel for multiple tasks, using several resources at once. However, when task demand is high, the resources allocated to that task are not available for another task if the same mental resources are required at the same stage of processing. Excessive use, moreover, can cause a state of overload known as cognitive resource depletion [37]. This overload means that the brain is unable to process new information, resulting in processing and/or execution errors [38].

Mental workload results from the different levels of resource demand, depending on the parallel tasks that the person is performing [8,9,21,22]. Excessive resource demand can cause distraction, increase errors, generate stress and frustration, and reduce the ability to undertake mental planning, problem solving, or decision-making [39,40]. One example is the distraction caused by unwelcome advertisements on a Web page while the user is browsing. In this case, the intermingling of the browsing task with the intrusion of commercial advertisements forces the user to divide attention and allocate cognitive resources to the new stimulus.

Traditionally, mental workload has been assessed in different situations using subjective methods [16] based on surveys, auto-perception scales, or think-aloud protocols [41–43]. These

methods are applied after the user has already finished the task, and the assessment of the mental workload depends of the user's final perception [44]. Therefore, these methods are constrained by the reporting bias introduced by relying on past memories and by the problem of ecological validity based on observing responses to hypothetical scenarios rather than behaviors in a real setting [45]. In addition, the static nature of these methods makes them unfit for real-time evaluation. The most widespread example of this method is the NASA Task Load Index, which measures the mental and physical performance, as well as the effort and frustration, of the user [46].

Performance-based methods have also been used, which measure indicators generated during task execution, such as the percentage of correct responses or execution time [3,16,17]. In this method, the user needs to be engaged in only one task. Its major restriction is the difficulty of assessing mental workload in near real time.

The attempts to find objective indicators to measure mental workload in real time are based on collecting contextual information, which can be captured mainly using psychophysiological sensors [47–49]. Indeed, there is ample empirical evidence in psychophysiology showing that some physiological responses are directly related to psychological factors such as stress, mental workload, and emotions [50–52]. That is, there is a correlation between the physiological responses triggered by the nervous system and psychological stimuli.

Psychophysiological responses are controlled by the autonomic nervous system (ANS), which regulates and coordinates bodily processes such as digestion, temperature, blood pressure, and many aspects of emotional behavior [53]. These actions occur independently of the conscious control of the individual. The ANS includes the sympathetic nervous system (SNS) and parasympathetic nervous system (PNS). The SNS controls actions required in emergency situations, such as stress and movement. It can cause heart rate acceleration, pupil dilation, and increased blood flow to the muscles, sweating, and muscle tension. The PNS controls the functions related to rest, repair, and relaxation of the body. The responses elicited by this system include a decrease in heart rate and blood pressure, stimulation of the digestive system, and pupillary contraction, among others [50,51].

2.2 Psychophysiological measurements

There are different types of methods to measure psychophysiological responses elicited complementarily by the SNS and PNS [54]. For instance, the device for tracking gaze is the eye tracker. It consists of a camera on the computer screen that works according to the "corneal-reflection / pupil-center" method [55]. It also allows the measurement of the variation of the pupil diameter. The pupillography measures changes in pupil size, which can be attributed to both parasympathetic inhibition, which explains the first dilation phase, and sympathetic activation, which explains the subsequent contraction phase [56,57]. Although pupil dilation can be triggered by a light reflex caused by changes in environment illumination or by a proximity or accommodation reflex to improve visual focus, it can also be caused by a psychosensory reflex associated with the cognitive or emotional engagement of the person while exposed to any sensory stimulus [58]. In contrast to changes in the two previous reflexes, changes in pupil size in this case are subtler, so a high-precision device or eye tracker is required for their detection [59].

The eye tracker is also used for tracking the eye to determine gaze position or movements within a scene, including two relevant measurements:

- *Fixations*: moments during which the gaze is relatively fixed or focused. They occur because sharp vision is only possible within a small area in the human eye called the fovea. It is useful to determine when eye fixation occurs because, in most cases, it coincides with attention.
- *Saccades*: rapid eye movements or jumps from one fixation point to another. Saccades follow a pattern (or trajectory) depending on several factors: what is currently being looked at, visual target tracking, experience, and emotions.

Another set of psychophysiological measurements is obtained by electroencephalography. This is based on recording the electrical activity of the brain measured on the scalp. The device used is the

EEG, which measures the voltage resulting from changes in ionic current flow within the neurons of the brain, produced by the brain's synaptic activity. There are five major brain waves: delta (1-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-25 Hz), and gamma (approximately 25 Hz). Fritz et al. [16] indicate that a decrease in the activity of the alpha band in conjunction with an increase in the activity of the theta band is associated with greater attentional demand and memory workload. Moreover, other studies have concluded that theta and delta bands are sensitive to stimuli involving difficult manipulation.

EDA is a psychophysiological response that can be assessed by measuring changes in the electrical properties of the skin. Skin conductivity varies with changes in skin moisture (sweat) and may reveal changes in the SNS. EDA is also known as galvanic skin response (GSR), and it is inexpensive to assess, easily captured, and robust. It is measured by attaching one or two electrodes usually to the fingers or toes. It is an indicator of psychological and physiological arousal. In addition, it serves to identify emotional states. EDA has two components: (1) a phasic component that changes rapidly and is related to external stimuli or a non-specific activity and (2) a tonic component or base signal that varies slowly and sets basic skin conductance. A classic behavior is that when arousal increases, there is an increase in sweat gland activity, decreasing electrical resistance, and thus increasing conductivity.

The cardiovascular system is particularly interesting for psychophysiology because it is highly sensitive to neurological processes and psychological factors such as stress. It is regulated by the ANS, which produces patterns of electrical activity that are fundamental for psychophysiological measurements [50]. Several studies associate changes in cardiac activity with psychological phenomena, such as mental work, perception, attention, problem solving, and signal detection [60].

An ECG is used to measure the electrical activity of the heart, using at least three electrodes attached to the chest. The electrodes collect the necessary data with regard to the electric waves that describe the cardiac cycle, based on which the HR or its variation (HRV) are obtained.

The human body constantly exchanges heat with the environment as part of the process of self-regulation to maintain homeostasis (internal balance of the body). Body temperature increases and decreases in relation to the energy exchanged. The regulation of blood flow to the skin and thermal radiation is considered a function of the ANS [61]. Studies conducted in this field, according to Genno et al. (1997) [62], suggest that skin temperature has potential as a psychophysiological measure of the individual.

3. Literature Review

This paper focuses on the measurement of mental workload while the user browses a Web site in front of her personal computer. However, the literature in this regard is scant. Thus, to start studying the measurement of mental workloads in various domains and to help understand the methodology associated with this type of research, this section focuses on two main points: the assessment of mental workload using psychophysiological sensors in general and the measurement of mental workload in Web environments.

3.1 Assessment of Mental Workload with Psychophysiological Sensors

A relevant study for this paper is that by Bailey et al. [23] who develop psychophysiological measures to assess the effect of interruptions on the performance of a person executing a task. They establish that interruption involves considerable negative effects, such as increased time to complete the task [63], a wider range of errors [64], additional efforts in decision-making [65] and mood changes such as increased frustration and anxiety [66–68]. For example, when an interruption occurs at a random time while performing a major task, the time to completion can increase by up to 30 %, up to twice as many errors can be committed, and user displeasure doubles, in contrast to when the interruption occurs at a pre-programmed time. Therefore, Bailey et al. empirically find that interruptions may have a lower cost if they occur at a time of low mental workload, hypothesizing that this may occur at the boundaries between subtasks when executing the general task [69]. As a test method, they assess mental workload by pupil dilation in three different tasks that include

respective subtasks. The first task consists of assessing two different routes between two cities on a monitor; the user must measure the distance and cost of the routes, tabulate the data, and, finally, discriminate and choose the shortest and most economical route. In the second task, the user must edit a document and correct spelling at three levels of complexity (editing a word, editing two words, and editing a complete sentence). The third task entails classifying nine emails involving explicit issues (low complexity) and ambiguous issues (high complexity) into four categories. Each of these scenarios is applied to 24 people (seven women) between 19 and 50 years of age. The main conclusions of the study are as follows: (i) mental workload varies during the execution of the three tasks, (ii) the mental workload decreases when performing subtasks compared to the general task, and (iii) different subtasks demand different levels of mental workload based on their complexity.

Other studies focus on training classifiers to process psychophysiological signal data in a time window in order to predict whether the load associated with a specific task is high or low [70]. For example, Haapalainen et al. [17] measure the mental workloads of basic tasks such as the resolution of problems on a monitor, visual perception, and cognitive speed by using an eye-tracking device, EEG, ECG, heat flow, and rate measurements. As a result, they find that ECG and heat flow together distinguish between tasks of high and low cognitive demand with 80 % precision.

Fritz et al. [16] seek to verify whether psychophysiological sensors are useful in measuring the difficulty of a computer code comprehension task with various levels of difficulty. The tasks are performed by software developers, who are monitored using an eye tracker and an electroencephalogram. Fritz et al. use the $Beta/(Alpha + Theta)$ ratio based on the evidence that beta increases with task execution, theta is deleted, and alpha is blocked. The models obtained classify task difficulty with 85 % accuracy.

Shi et al. [71] assess stress and arousal levels by measuring EDA for increasing levels of difficulty. The experiment consists of a transition interface in which the participants must respond to the requirements in three scenarios: (1) using gestures and speaking, (2) only speaking, and (3) only using gestures. The difficulty varies depending on level of visual complexity, number of entities, number of distractors, time limit, and number of actions to complete. The results indicate that there is a significant increase in the EDA signal as task difficulty increases.

Nourbakhsh et al. [72] confirm the effectiveness of EDA in discriminating between the difficulty of eight arithmetic tasks with four levels of difficulty. In addition, as an extension of the previous study, Nourbakhsh et al. measure mental workload using EDA changes and the number of blinks obtained from an eye-tracking device. The experiment is the same as in the previous study. This time, by combining both sensors, 75 % precision is achieved for the lowest level of difficulty.

Xu et al. [73] show that mental workload can be measured by pupil dilation if illumination changes. The experiment consists of arithmetic tasks that vary in difficulty depending on the number of digits.

In Ikehara et al. [18], an eye-tracking device, a pressure sensor for the mouse, an EDA sensor, and a pulse oximeter (for measuring HR and level of oxygen in the blood) are used. The experiment consists of selecting on a screen the fractions whose value is less than 1/3. There are two levels of difficulty in the experiment. The results indicate that EDA and pupil dilation have the greatest statistical significance in terms of detecting task difficulty.

Using an elastic neural network, Hogervost et al. [19] find that the best performance is obtained when EEG is combined with pupil dilation (91% accuracy) and when EEG is combined with peripheral physiology (89 %); with EEG alone, they obtain 86 % accuracy. In addition, using only the measurement of the electrode located in the Pz position (central parietal area of the head), they obtain 88 % accuracy.

3.2 Assessment of Mental Workload in Web Environments

Although the study of users' cognitive responses during Web browsing is an intriguing area, it remains little explored. Indeed, one of the few studies on the topic is that by Albers [3], who examines how mental workload theory applies to the design of Web sites using the tapping test method, which measures mental workload by focusing on performance. As in all the examples using

this approach, the tapping test adds an additional secondary task to the main one, measuring the performance of the participant to determine the level of mental workload induced. In this case, the main task is to browse two Web sites sequentially—with implicit mental workload controlled by design—and answer questions aloud in relation to the Web pages, while the secondary task is to rhythmically keep tapping per second. As mental workload increases, tapping begins to fall slowly and lose the rhythm, even losing it completely when there is cognitive overload. However, implementing a secondary task as required by this method prevents from generating a realistic scenario for the user and does not allow real-time measurement.

The most recent research regarding the observation of Web users' experience involves the measurement of their behavior as a reaction to different stimuli, such as notifications, and allows us to predict the user's response according to Navalpakkam & Churchill [74]. By comparing mouse pointer movement to eye tracking, they are able to determine a more user-friendly layout for a Web site, which improves the effectiveness of the notification. Finally, they conclude that gaze and mouse movement patterns contain important information in terms of assessing the user's status, determining if they are distracted from the assigned task or striving to fulfill it. The correlation between eye movements and mouse pointer movement predicts a Web user's different psycho-emotional states. They also conclude that the user is more likely to pay attention to notifications when they vary in position on the Web site rather than when they are fixed.

As summarized in Table 1, the measurement of mental workload using psychophysiological signals has been tested for a varied set of tasks. In addition, studies have investigated how mental workload is related to the design of a Web page. However, the abovementioned research provides no evidence regarding assessment of mental workload while browsing a Web site using multiple psychophysiological measures. There is also no reference to time overhead to determine how feasible it is to implement real-time measurement. Partial. Time Windows average length for classification of 23.7 s.

4. Materials and Methods

4.1 Participants

The initial experimental group includes 61 participants (19 women and 42 men), aged between 19 and 35 years (*mean age = 23.8* years, *SD = 3.2* years), all engineering students at the University of Chile, recruited through the institutional news Web application. None of them suffered from cardiovascular diseases or was taking medications that could have affected their normal behavior. All of them were familiar with browsing tasks. Each session had a duration of approximately 60 min. The final experimental group is composed of 53 people. Eight participants were discarded due to various problems during signal measurement and processing.

This research has the approval of the Research Ethics Committee at the Faculty of Physical and Mathematical Sciences at the University of Chile. In addition, all of the participants read an informed consent and agreed on signing it. The consent contained information about the procedure, purpose of the experiments, voluntary participation, right to decline to participate at any moment, how to access the research results and researchers' information.

4.2 Psychophysiological Sensors

Psychophysiological sensors have the advantage that measurements do not depend on the user's perception and are not under the control of the user.

Table 1. Related work analysis

Ref.	Small Time Windows	Real Time	Web Browsing Tasks	Multiple Psychophysiological Sensors
[15]	Partial. Time window average length for classification of 23.7 s.	Fulfills	Fails. Desktop-based tasks of visual perception and cognitive speed.	Fulfills. Eye tracker, EEG, ECG, heat flux and HR.
[14]	Partial. Sliding time windows of sizes from 5 seconds to 60 seconds, sliding 5 seconds between intervals.	Fulfills	Fails. Comprehension tasks of computer code.	Fulfills. Eye tracker, EEG
[34], [37]	Partial. Three silent reading tasks were performed. Each task consisted of four text slides and each slide was presented for 30 seconds.	Fulfills	Fails. Arithmetic tasks	Partial. EDA and blink.
[16]	Fails. Duration of tasks between 60 and 70 seconds.	Fulfills	Fails. Select the fraction whose value is less than 1/3.	Fulfills. Eye tracker, EDA, pulse-oximeter, mouse pressure sensor.
[17]	Fails. Classification in 2 minutes windows.	Fulfills	Fails. N-back task.	Fulfills. EEG; peripherals (EDA, respiration, ECG) and eye measures (eye tracker).
[18]	Fulfills. The average overall duration of the limits was 550 ms.	Fulfills	Partial. Measures the cost of interruptions in tasks such as: choosing a route, correcting spelling and classifying emails.	Partial. Only the pupillary dilation.
[3]	Not applicable.	Fails	Fulfills. Measure the cognitive load on a website.	Fails. Measurement of mental workload by performance: tapping test.
[36]	Fails. The participants interacted with each page for about 100-120 seconds.	Fulfills	Fulfills. Study the design of websites in a way that improves the effect of a notification.	Partial. Compare the tracking of the mouse with eye tracking.
[38]	Fulfills. Windows between 300 and 600 ms.	Fulfills	Fulfills. Predict the intention of clicking on a website.	Partial. Only the pupillary dilation.

In addition, they are becoming less intrusive and allow tasks to be performed in various scenarios, giving greater ecological validity to the experiments. They also allow real-time data capture [16,50].

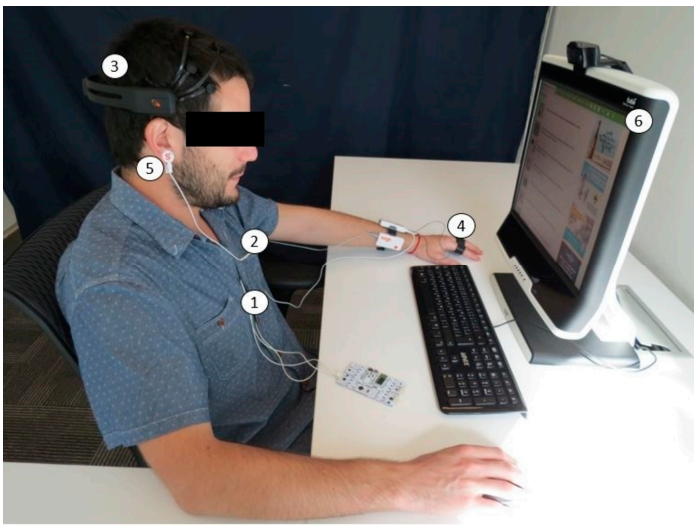

Figure 1. Participant with the sensors runs the experiment. The sensors are: (1) ECG, (2) axillary temperature, (3) EEG, (4) EDA, (5) PPG and (6) eye tracker.

For data acquisition, the following sensors were used: *GSR+*, optical pulse sensor, and Bridge Amplifier + unit, all from the Shimmer [75]; ECG BITalino [76] ; EEG Emotiv Epoc [77]; and Tobii T120 Eye Tracker [78]. Figure 1 shows an example of a volunteer outfitted with all the sensors.

To measure the EDA and HR signals, the Shimmer GSR+ unit sensor was used with a sampling frequency of 120 Hz. The position of the electrodes for measuring the EDA was the palm area of the proximal phalanx of the index and ring fingers of the left hand [79]. The optical sensor that functions as a photoplethysmograph (PPG) was attached to the lobe of the right ear [80]. The Shimmer Bridge Amplifier + unit sensor with a sampling frequency of 50 Hz was used to measure body temperature. The sensor was applied under the right armpit. This sensor was synchronized with the EDA and pulse sensors using a base provided by Shimmer together with Consensys software.

The BITalino BioMedical Development All-in-One Board with a sampling frequency of 1000 Hz was used to measure the ECG. The configuration of the three electrodes followed the lead II standard [81,82]. Before applying the electrodes, the skin was prepared by wiping it with alcohol to remove grease and impurities to reduce noise. In addition, an ECG gel was used. OpenSignals evolution software provided by the manufacturer was used [83].

To measure the EEG, the Emotiv EPOC EEG sensor with a sampling frequency of 128 Hz was used. The sensor was attached to the head, positioning the reference sensors first. To improve the conduction of the electrical signals of the brain, each electrode was previously hydrated. To capture the data and verify that the sensor was properly applied, the Emotiv Xavier Testbench software provided by the manufacturer was used.

The Tobii T120 Eye Tracker with a sampling frequency of 120 Hz was used to measure pupil dilation and for eye tracking. Tobii Studio software was used for calibration and to perform data collection [84]-

4.3 Experimental Procedure

A fictitious Web site was created whose basic configuration is shown in Figure 2. This layout of the Web elements was maintained through all the experiment. The elements within the Web site were seven news headings with their respective representative image, four rectangular advertisements, a typical navigation bar with a menu, the logo of the page in the upper left corner, and a bar at the bottom of the page.

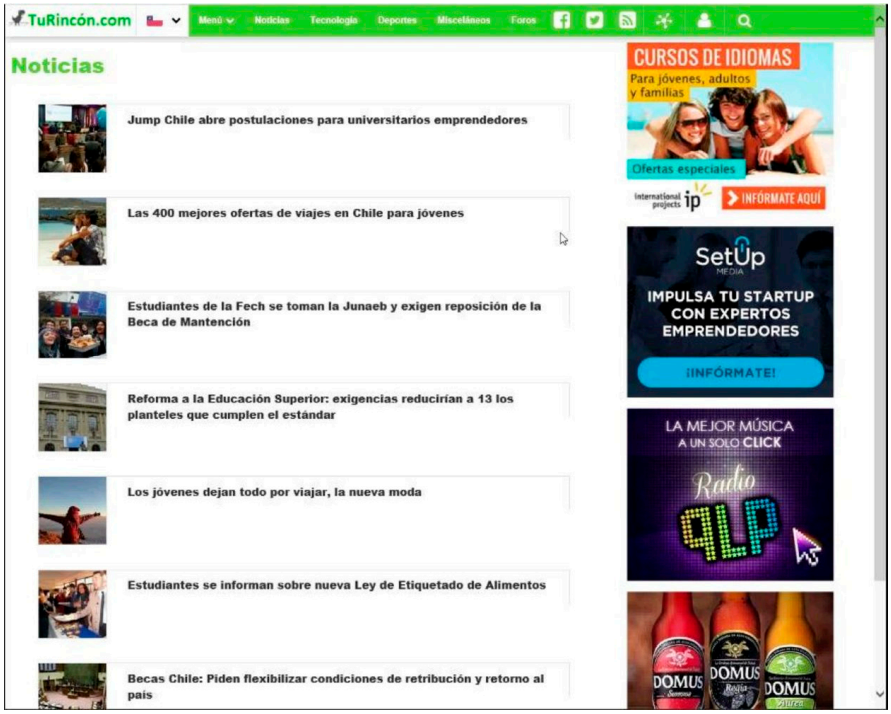

Figure 2. Example of a dummy Web page for the experiment.

Each participant was tested individually at the laboratory. A physically isolated experimental room was used to maintain the experimental configuration and the environment constant for all participants. In addition, the room did not receive any sunlight, to avoid the effects of infrared light on measurements and to maintain constant illumination conditions that do not affect pupil diameter measurements [85].

As soon as the participant arrived in the experimental room, the experiment was explained to her, and she was asked to read and sign the informed consent, as well as a questionnaire to get her basic anonymous information. The participant seated in front of the screen, and the sensors were connected in the following order: ECG, axillary temperature, EEG, EDA, and PPG; then the eye tracker was calibrated with the help of the participant (Figure 1).

Prior to the tests, each user underwent a relaxation period consisting of the visualization of three four-minute videos of landscapes with background instrumental music. Then, the participant was asked to take deep breaths for one minute with eyes closed and with soft background instrumental music. This procedure aimed to eliminate the Hawthorne effect – modification in the behavior of the subjects due to their awareness of being studied – and physiological effects similar to the "white coat" effect in measured signals [86]. Next, the participant was asked to maintain a fixed posture, sitting in front of the computer, without moving the head or the left hand, where the sensors were connected. The instructions were that the user could freely browse the Web site for as long as they wanted and indicate when they wanted to finish. Finally, all sensors were removed from the participant, while she was asked do not tell others the experimental procedure.

4.4 Data Analysis

4.4.1 Time Window Definition

Bailey (2008) [23] shows that mental workload decreases during transitions between subtasks. For this paper, the analysis of each Web element is considered a specific subtask and the passage between them as the transition period between subtasks. Thus, in this study, mental workload is assessed during two time windows:

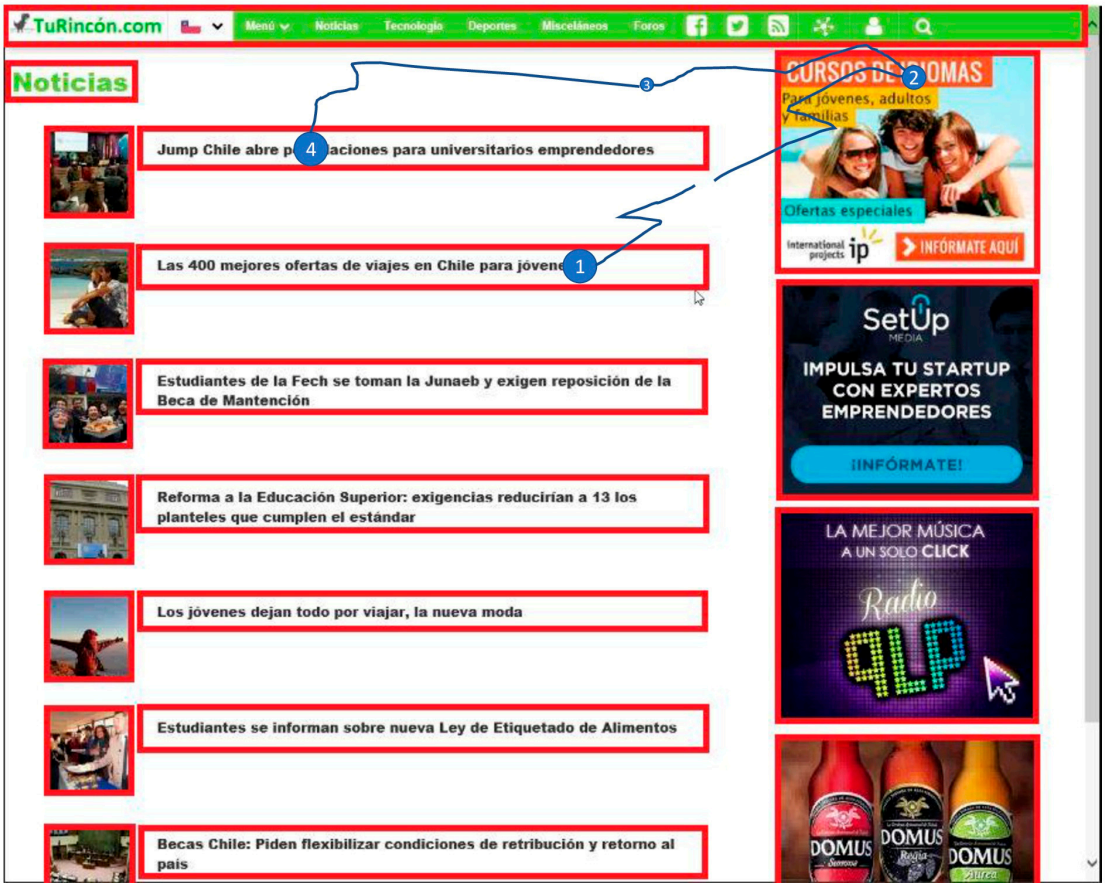

Figure 3. Example of active window and transition window.

- *Active window:* Time during which the user fixes her gaze on a specific area of interest (AoI), which may correspond to a news headline, an advertisement, or the menu bar of the Web site.
- *Transition window:* Time that elapses while the user is not fixing her gaze on any of the areas of interest. It can be a transition between two elements or towards the same element.

As illustrated in Figure 3, the red rectangles represent the studied AoIs; the blue circles represent fixations, which size varies in accordance with the fixation time and the blue lines represent the saccades. Thus, the time a fixation is into an AoI pertains to an active window. The time between two fixations, such as fixation one and fixation two, pertains to a transition window. Note that the transition window between fixation two and four add the fixation three, which does not fall into any AoI.

To discriminate between types of windows, the data file exported from the Tobii Studio program generates a column showing the AoI that the participant is inspecting for each sample. It discriminates between 3 values: when the user is not looking at the screen – inactive –, when the user is looking at a certain AoI – active window –, and when the user's gaze is directed outside the AoI – which is considered a transition window.

A long minimum time of 500 milliseconds is set to define a valid time window. This is based on the research of Loyola et al. [87], who assesses the identification of key Web elements in a Web site using eye tracking. This time span is selected to avoid possible contamination of the pupil signal by the analysis of a previous object. Time windows below the threshold are not considered for analysis and are therefore deleted. When the same Web element is analyzed before and after a deleted window, the two segments are joined, generating a window of greater length.

4.4.2 Data Preprocessing

The data exported from Tobii Studio contains the diameter of the left pupil, the diameter of the right pupil (both in millimeters), and the validation of the reliability of the capture of each pupil between 0 – high reliability – and four – the eye was not detected. On average for all participants and considering only valid windows, the reliability of the capture of the left pupil is 0.2469, and that of the right pupil is 0.22036; these are reliable values to validate the capture of pupil diameter data. As these values are an average for all the participants, the pupil data with the highest level of reliability are selected for each sample [16].

Next, signal distortion artifacts, such as saccades and blinks, are eliminated. A column in the extracted data shows if the sample is a fixation or a saccade, and this information is used to filter saccades. Furthermore, a linear interpolation between the values of the blinks detected is used. In addition, a Blackman window with a cut-off frequency of 2 Hz is applied as a low-pass filter.

EDA raw data provides the values of electric resistance of the skin in Kilohms [$k\Omega$]. To reduce noise and eliminate motion artifacts, two procedures are performed: first, a strict instruction is given to each participant not to move the hand or fingers where the electrodes are attached, and second, the signal is filtered with a low-pass cut-off frequency of 5 Hz. Furthermore, on the recommendation of the literature [88], capture resolution is reduced without risk of data loss. The EDA signal measured with a sampling frequency of 120 Hz is reduced to 10 samples per second. The phasic component is extracted by applying a median filter with a window width of -4/+4 and subtracting the average of the current sample [88]. This component allows the detection of peaks of the EDA signal. With slow transitions, the phasic component does not show major variations.

Regarding the electrocardiogram, the raw data yield values that must be transformed to millivolts [mV]. The processing of this signal consists of using a low-pass filter with a cut-off frequency of 100 Hz and applying the fast Fourier transform to obtain the characteristic shape.

The raw data of the PPG yield signal values in millivolts [mV]. From this signal, it is possible to obtain the HR. Previously, the PPG signal is processed using a low-pass filter with a cut-off frequency of 16 Hz with a Blackman window, obtaining a cleaner signal. Then, HR is obtained via the following steps: first, the peaks must be found; second, the time between them is substracted (Δt in [miliseconds / pulse]); third, they are converted from hundredths to seconds and from [seconds / pulse] to [pulses / second], which is then multiplied by 60 to convert to [beats/minute]. This is resume in the Equation (1):

$$HR = \frac{60}{\Delta t \cdot 100} \left[\frac{beats}{minute}\right] \quad (1).$$

The raw data yield body temperature values in degrees Celsius. The processing of this signal consists of using a low-pass filter with a cut-off frequency of 1 Hz, as concluded based on the data collection in Haapalainen et al. [17].

The EEG signal is subject to a wide variety of artifacts and noise [89,90]. Among the elements that cause artifacts are blinking, oculomotor activity, head movements, facial expressions that add noise due to the muscle electrical signal, and movement of the electrodes, among others. To eliminate the effect of head swinging, a high-pass filter with a cut-off frequency of 0.5 Hz is used. In addition, a low-pass filter with a cut-off frequency of 40 Hz is used to eliminate noise from the electrical grid (50 - 60 Hz). To eliminate outliers and decrease the effect of the blinking artifact a Hampel filter is used [91].

4.4.3 Feature Extraction

Feature extraction is performed based on time windows. Since signals have different scales, to be comparable objects, it is necessary to standardize them before extracting characteristics from them, as proposed by Guyon et al. [92]. To perform standardization, the classical $(x - \mu)/\sigma$ form is used, where x is the vector corresponding to the signal, and μ and σ are the mean and the standard deviation of the signal, respectively.

Table 2. Characteristics extracted by each signal.

Signals	Extracted Characteristics
Pupil	mean, standard deviation
EDA	Accumulated data, average as a function of time and spectral power
Phasic	average, absolute value of the maximum, number of peaks
ECG	mean, median, variance of ECGMAD (average absolute deviation)
PPG(HR)	mean, standard deviation, RMS of HR
T	mean, median
EEG	power and phase of the analytical signal obtained with the Transf. of Hilbert

A total of 44 characteristics pertaining to the different signals are extracted: two from pupil dilation, six from EDA, two from body temperature, three from ECG, three from PPG-HR, and two from each of the 14 EEG channels. Table 2 shows a summary of the characteristics, following which the obtained characteristics are presented in more detail.

Because it has been proven that pupillary response is an important indicator of the mental effort required to solve a task, it is selected as the gold standard by which to cluster windows and generate labels for cognitive levels. There are clustering cases in the literature regarding the development of Web tasks such as the study of Loyola et al. [87]. The selected characteristics are the mean and variance of the pupil diameter of the eye that displays greater reliability in its measurement.

Based on the findings of Nourbakhsh [72] and Shi et al. [71], the following characteristics are extracted from the processed EDA signal: accumulated normalized data, mean as a function of normalized time, and spectral power without normalized continuous component. Equation (2) shows the calculation of the normalized EDA signal. Each point in time t is added, where i corresponds to the participant, k and m is the total number of tasks; $m = 3$ in this case:

$$EDA_{normalized}(i,k,t) = \frac{EDA(i,k,t)}{\frac{1}{m}\sum_{j=1}^{m}\sum_{t=1}^{T_{ij}} EDA(i,j,t)} \quad (2).$$

Therefore, the data for each participant are normalized by dividing the task signal by the mean value of all the tasks for the subject. Then, the accumulated EDA characteristics are calculated as shown in Equation (3) and mean EDA is calculated according to Equation (4), where T is the total time for all the tasks:

$$EDA_{accumulated}(i,k) = \sum_{t} EDA_{normalized}(i,k,t) \quad (3),$$

$$EDA_{average}(i,k) = \frac{\sum_{t} EDA_{normalized}(i,k,t)}{T} \quad (4).$$

The following characteristics are extracted from the phasic component obtained: number of peaks, maximum modulus, and average of the phasic component of the window [16].

Based on the proposal by Haapalainen et al. [17], the following characteristics are selected for the ECG signal: median, mean, and variance of the ECG median absolute deviation (ECG_MAD), calculated using Equation (5):

$$ECG\_MAD = |ECG_i - median(ECG)| \quad (5).$$

The characteristics of the heart rate obtained from the PPG signal are selected based on the time domain characteristics used in Betella [93]. These are the mean, standard deviation, and root mean square of HR. Based on the proposal by Haapalainen et al. [17], the median and mean of the temperature are selected.

For the EEG signal, there are two main approaches: event-related potential (ERP) analysis and time-frequency signal analysis. The latter is selected because it is more closely related to the psychophysiological and structural processes of the brain [89]. It is used to study emotional-cognitive states in particular and is more advisable when studying a limited period or a relatively low amount of data, as is the case of the time-window study of this paper [94]. Among the different ways of analyzing the EEG signal in time-frequency are frequency bands with Fourier transform, Morlet wavelets, and Hilbert transform. All three show similar results according to Cohen [95]. Thus, the option of the Hilbert Transform ($\mathcal{H}\{eeg\}(t)$) is selected, which has the advantage of greater control over frequency filtering. The Equation (6) shows this transform:

$$\widehat{eeg} = \mathcal{H}\{eeg\}(t) = (h * eeg)(t) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{eeg(\tau)}{t-\tau}d\tau \quad (6),$$

where $h(t) = 1/\pi t$, eeg is the EEG signal and $\widehat{eeg}$ is the resulting analytic signal. Before applying this transform, a bandpass filter between 2 and 15 Hz is used to center the study in the theta (4 - 8 Hz) and alpha (8 - 12 Hz) frequency bands. These are related to states of mental activity and relaxation, respectively, where theta increases and alpha is suppressed when there is mental workload [94]. A complex signal called the "analytical signal" is then obtained, from which two characteristics are extracted. This is performed for each of the 14 channels of the EEG signal.

4.4.4 Clustering

Clustering is performed per participant to determine how many levels of mental workload the user presents based on the measurement of pupil diameter in order to label the database after ascertaining these levels. In Loyola et al. [87], the k-means method is used. Because an overestimation or underestimation of the number *K* of clusters affects the quality of the cluster, the optimal value of clusters is sought. The *K* value is tested from two onwards to obtain two curves. The index of Calinski & Harabasz (CH) and the internal measure of cohesion of the sum of the squares within the group (WSS) are selected to this end [96–98]. The stop rule is the value closest to the area where the curves interact. Figure 4 shows an example of this methodology for participant 59, where the intersection is generated at $K = 3$. Visually, the grouping can be validated considering Figure 5.

The Jaccard coefficient obtained using the bootstrap method is used as an external criterion for validating clusters, which assesses how stable the cluster is [96,97] . Values between *0.6* and *0.75* indicate that the group is measuring a pattern in the data, but there is no certainty as to which points should be grouped. Groups with stability values above approximately *0.85* can be considered highly stable (real clusters). There are participants who present well-defined clusters with Jaccard coefficients very close to one, and others with values far from acceptable.

On average, the coefficients are over *0.75*, so clustering is accepted as valid. For example, for the clusters in Figure 5, the Jaccard indices are *0.6288* (cluster 1), *0.9024* (cluster 2), and *0.8517* (cluster 3). Considering all the valid participants, the number of clusters varies between three and six levels of mental workload, and on average, there are approximately four levels of mental workload validated with acceptable cohesion indices (RQ1).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 December 2017 doi:10.20944/preprints201712.0021.v1

Peer-reviewed version available at Sensors 2018, 18, 458; doi:10.3390/s18020458

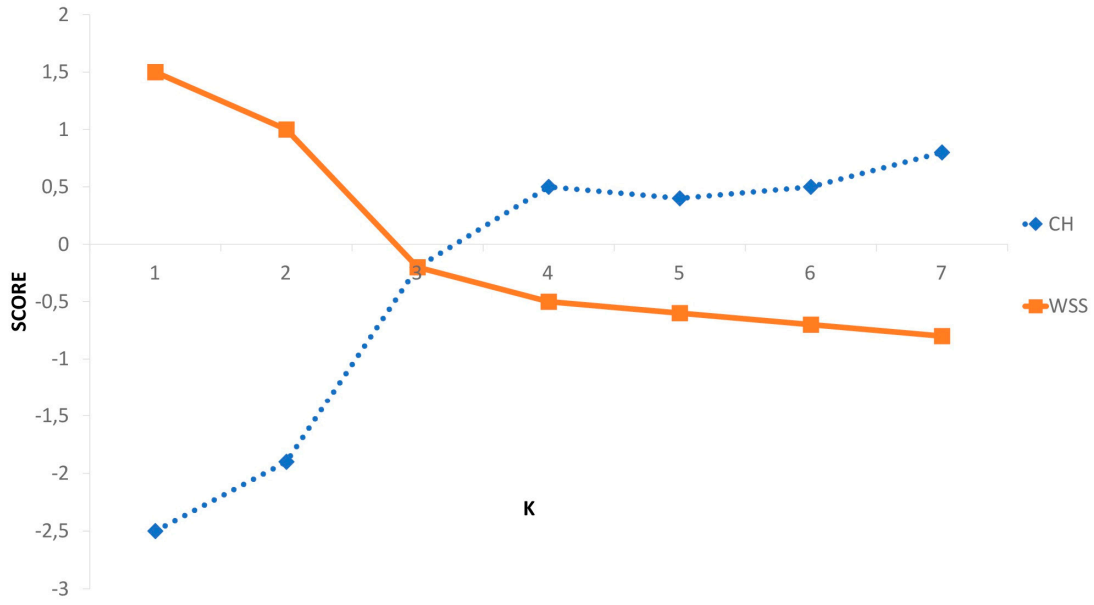

Figure 4. Optimal number of clusters according to the intersection method of CH and WSS curves for participant 59.

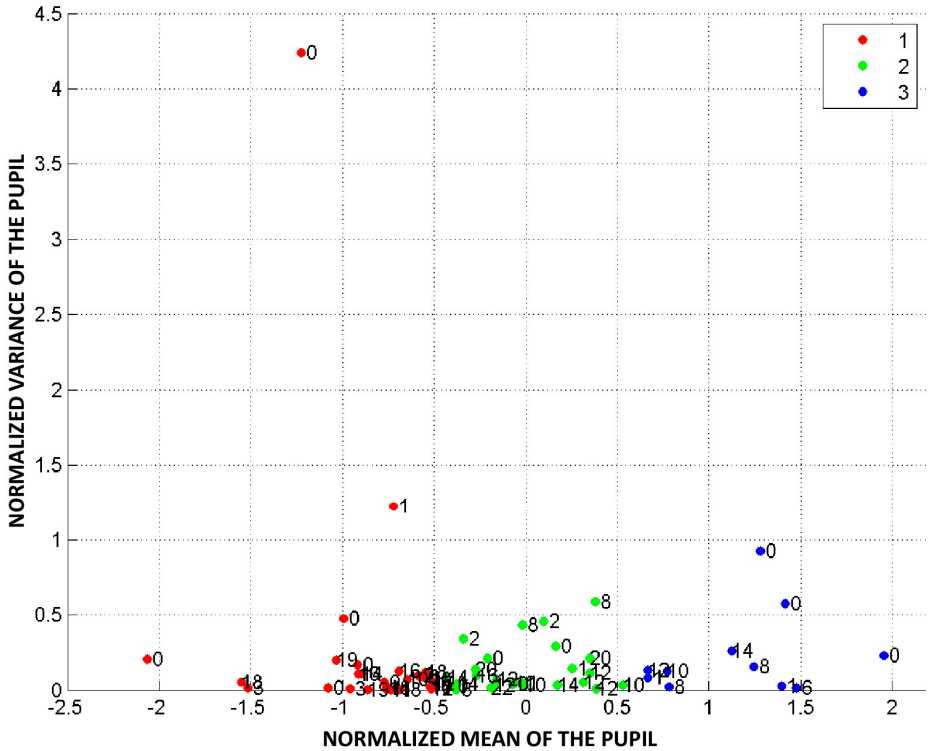

Figure 5. Optimal grouping of time windows according to their level of cognitive load for the participant 59.

4.4.5 Machine Learning Models and Feature Selection

To perform the classification, a training set is first generated with 70 % of the observations and then a test set with the remaining 30 %. To avoid biases, a 10-fold cross validation is performed in which the classes are distributed uniformly within each set. In addition, they are randomly selected while maintaining the proportions. Because clustering generates some classes containing very few

elements, sometimes just one or two, time windows that would not make sense to classify because they would only be in the training set and not in the test set, are deleted.

Two classification models are applied: artificial neural networks and recursive feature elimination (NN and RFE respectively) and two-layer multi-layer perceptron (MLP). Each classification result is obtained from the average resulting from executing each algorithm 100 times. Next, the implementation of each model is described.

An artificial neural network with a hidden layer is implemented, with all the artificial neurons completely interconnected and trained with the algorithm backpropagation. To calculate the number of neurons in the hidden layer (h), the heuristic method of the geometric pyramid rule is used with the expression $h = \sqrt{m \cdot n}$, which consists in calculating the square root of the product between the number of inputs (m) – number of characteristics – and the number of outputs (n) – number of classes. Therefore, the number of neurons in the hidden layer changes per participant. This classifier is combined with the RFE method for feature selection.

There is evidence in the literature regarding the use of the random forest and recursive feature elimination (RF-RFE) method for the selection of characteristics with good results when applied to the classification of mental fatigue with EEG signals [99]. This combines recursive elimination with random forest, that is, a set of decision trees that assesses features and generates a ranking following a score criterion. This method of feature selection is implemented with the Caret and Random Forest packages in R. The algorithm is executed using Matlab's Neural Network Toolbox with the toolbox's Neural Net Pattern Recognition nprtool. It is executed once the characteristics have been obtained – per participant – with the RFE method. Table 3 shows the characteristics selected for six participants as an example.

To test another way of improving the artificial neural network without using feature selection, a different neural network configuration is tested: MLP. For the implementation of the MLP neural network, the H2O package in R is used [100]. The programmed neural network has two hidden layers with 100 neurons each, with a rectified linear activation function, as used by Hinton [101]. The key, according to Hinton, to reducing overfitting is to include a *50 %* dropout for each layer, which prevents artificial neurons from co-adapting to training data.

Table 3. Selected features with the RFE method for different participants.

Participant	Selected features
1	maxPhasic, meanECGMAD, meanHR, powerEEG channel1, phaseEEG channel1, powerEEG channel3, phaseEEG channel12
2	medianTemp, avgGSR, powerEEG channel7, phaseEEG channel7, powerEEG channel1, powerEEG channel11
3	meanHR, meanPhasic, phaseEEG channel8, meanECGMAD, maxPhasic, rmsHR, accGSR, powerEEG channel5, phaseEEG channel4, phaseEEG channel6, phaseEEG channel12, avgGSR, stdHR, phaseEEG channel2, phaseEEG channel7, meanTemp, phaseEEG channel5, phaseEEG channel13, powerEEG channel13, phaseEEG channel1, medianECGMAD, powerEEG channel2, phaseEEG channel14, varECGMAD
4	meanHR, avgGSR, powerEEG channel1, phaseEEG channel9, powerEEG channel13, phaseEEG channel14, medianTemp, varECGMAD, phaseEEG channel2, rmsHR, mean-Phasic, phaseEEG channel8, medianECGMAD, phaseEEG channel1, phaseEEG channel12, numpksPhasic, phaseEEG channel11, phaseEEG channel13, maxPhasic, powerEEG channel8
5	avgGSR, meanTemp
6	medianTemp, powerEEG channel6, powerEEG channel5, meanPhasic

Thus, each neuron in the hidden layers is omitted at random from the network with a probability of *0.5*. In addition, another method added to avoid model overfitting is the L_1 and L_2 regularization method as a linear combination, as shown in Equation (7)**Error! Reference source not found.**. For this, the objective function for the artificial neural network is defined as $L(\boldsymbol{W}, \boldsymbol{B}|j)$, where $\boldsymbol{W}$ represents the weight matrix and $\boldsymbol{B}$ the column of bias vectors for each training example.

$$L'(\boldsymbol{W}, \boldsymbol{B}|j) = L(\boldsymbol{W}, \boldsymbol{B}|j) + \lambda_1 R_1(\boldsymbol{W}, \boldsymbol{B}|j) + \lambda_2 R_2(\boldsymbol{W}, \boldsymbol{B}|j) \quad (7),$$

where the values of λ_1 and λ_2 are parameters that weight the relative contribution of the penalty terms R_1 and R_2 (rule L_1 and L_2, respectively) in relation to the objective function $L(\boldsymbol{W}, \boldsymbol{B}|j)$. The values of $\lambda_1 = 10^{-5}$ and $\lambda_2 = 10^{-5}$ are determined as recommended in the H20 manual [100].

5. Results

5.1 Statistical Analysis

Based on Bailey (2008) [23], who showed a decrease in mental workload between subtask boundaries, the hypothesis that there is a decrease in mental workload in the transition time windows between the analysis time windows of one Web element and another is proposed. To verify the hypothesis, the mean pupil diameter within each window is selected as our gold standard. The objective is to determine if mean pupil diameter varies depending on whether it is in an active window or in a transition object. An analysis of variance with repeated measures (ANOVA-RM) is performed since the factors to be studied are within-subjects. For the analysis, the complete universe of windows of all the participants is considered.

As a result, a $p - value = 0.00184$ is obtained with a 95 % confidence interval, so the null hypothesis is rejected. In addition, as shown in Table 4, mean pupil diameter in the transition windows is smaller than in the active windows. Therefore, it is concluded for the data as a whole that the difference between mean pupil diameter in the active windows and the transition windows is statistically significant and that the diameter is smaller in the transition windows (H1).

5.2 Classification

Table 5 shows that the worst accuracy measure obtained with the NN-RFE is 45.24 % (participant 48) with five classes, and the best result is 95.24 % (participant 23) with two classes. The classification mean including all 53 participants yields the result of 68.94 % accuracy with a standard deviation of 11.54 %. The results of the classification according to the number of final classes are also analyzed. As shown in Table 6, there is a tendency for the classification percentage to decrease as the number of classes increases. In particular, an accuracy of 90.61 % is obtained for the classification of two classes, 73.34 % for three classes and acceptable values are obtained for four and five classes.

Regarding the MLP, in Table 7, the worst measure of accuracy obtained is 72.16 % (participant 17) with four classes, and the best result is 99.9 % (participant 9) with three classes. The classification mean including all 53 participants is 88.46 % with a standard deviation of 7.94 % for the accuracy measure (RQ2). The classification analysis is performed according to the number of final classes. As shown in Table 8, the trend observed for NN-RFE is maintained such that the higher the number of classes is, the lower the classification percentage, but with a break in the case of four and five classes.

Table 4. Standardized means of pupillary diameter for transition and active windows.

Factor	Mean	Standard Deviation
Transition	0.01924	0.97
Active	0.10737	0.84

Table 5. Results of classification using NN-RFE.

-	Accuracy (%)	Recall (%)	Precision (%)
Mean	68.94	66.62	76.92
Standard Deviation	11.54	13	11.62
Maximum	95.24	95.45	95.45
Minimum	45.24	39.99	46.29

Table 6. Average classification using NN-RFE by quantity of classes.

Number of classes	Accuracy (%)	Recall (%)	Precision (%)
2	90.61	89.99	92.15
3	73.34	71.69	83.56
4	63.29	59.57	70.75
5	57.94	54.92	64.35
6	53.85	53.61	66.31

Table 7. Results of classification using MLP.

-	Accuracy (%)	Recall (%)	Precision (%)
Mean	99.1	98.99	99.27
Standard Deviation	0.2722	0.3325	0.2174
Maximum	100	100	100
Minimum	98.26	97.96	98.64

Table 8. Average classification using MLP by quantity of classes.

Number of classes	Accuracy (%)	Recall (%)	Precision (%)
2	99.03	98.92	99.22
3	99.14	99.05	99.31
4	99.06	98.93	99.25
5	99.06	98.96	99.24
6	99.22	99.15	99.36

5.3 Evaluating Psychophysiological Sensors

To assess the performance of each sensor, the MLP neural network that obtains the best results with all the sensors is selected as a supervised learning model. Table 9 shows the results of assessing the performance of each sensor separately. The sensor with the best performance is EEG, with 88.78 % accuracy in the classification, slightly superior to the classification using all the sensors. The other sensors separately have a very low level of classification accuracy.

Table 9. Summary of sensor classification results for MLP with 100 neurons in each hidden layer and 400 epochs.

Sensors	Accuracy (%)	Recall (%)	Precision (%)
All	99.1	98.99	99.27
EDA	53.28	63.18	53.52
ST	42.37	76.41	35.7
ECG	46.22	66.81	45.55
PPG	46.69	76.58	43.88
EEG	92.27	94.74	94.94
EDA+PPG	63.36	68.33	66.5
EDA+EEG	96.23	97.44	97.73
PPG+EEG	95.89	96.78	96.56
EDA+PPG+EEG	97.29	98.01	98.36

Then, combinations of the three sensors with higher performance are tested: EEG, EDA, and PPG (HR). As shown in Table 9, the combinations with EEG provide the best results. The combination with the highest performance is EDA, PPG (HR), and EEG, with 95.73 %.

An important difference between the EEG sensor and the others is that it allows the extraction of a greater number of characteristics because the 14 electrodes contribute two characteristics each, for a total of 28 characteristics. This factor may explain the superior performance of this sensor compared to the rest. Therefore, it is concluded that it is possible to obtain good classification results for this experimental design with less than 5 sensors, even only with the EEG. The temperature sensor and the ECG can thus be discarded.

5.4 Evaluating Time Overhead

For mental workload classification in Web site browsing to lead to a real application, processing must be sufficiently fast, given that the time windows considered have a minimum length of 500 milliseconds. Table 10 shows the classification time for the algorithms used that yielded the best results.

The results show that, for the artificial neural networks and RFE model, the time window is very small at 0.0083 seconds on average, which ensures that this model can be implemented in real time with an acceptable classification mean of 68.94 % (see Table 5).

Table 10. Testing time for models.

	Mean [sec]	Standard Deviation [sec]
NN-RFE	0.0083	0.001
MLP	0.109	0.023
MLP EEG	0.1218	0.0041
MLP EDA+PPG	0.1056	0.0091
MLP EDA+EEG	0.1207	0.0098
MLP PPG+EEG	0.107	0.0037
MLP EDA+PPG+EEG	0.126	0.007

On the other hand, for the model based on MLP, the effect of its parameters given by the 100 neurons in the two hidden layers increases the processing time to an average of 0,1 seconds. So, this model can be implemented in real time as well with a classification average of 99.1 % accuracy for all sensors and a reasonable 63.36% for more portable sensors (PPG and EDA, see Table 9).

6. Discussion

The results of the statistical analysis determine that pupil diameter in the transition time windows is statistically and significantly lower than in the active windows of Web element analysis. Given the proven correlation between pupil dilation and mental workload, it is determined that there is a decrease in mental workload in the time windows between the analysis of one Web element and another (H1).

A possible application of the demonstration of this hypothesis (H1) is the generation of recommendation systems that support the user in during Web browsing according to her interest, that is, when she is not cognitively overloaded with new content. This is applicable, for example, to retail applications or advertisement.

Regarding the assessment of the psychophysiological sensors to estimate mental workload, with the exception of the EEG, the signals of the sensors used do not provide an appropriate level of classification by themselves for this type of task, although the combinations of signals with EEG stand out, obtaining very good results. One of the reasons may be the time constant of each signal; that is, signals such as skin temperature or conductivity take longer to react compared to electrical signals from the brain.

However, despite the fact that the combination of EDA and PPG (HR) does not provide better results than EEG alone, a reasonable level of accuracy (63.36 %) is obtained for its use in practice, even before portable EEG technology is available. The advantage of the EDA and PPG sensors is that they are non-invasive, portable, cheaper, and easily integrated into a board that transmits via Bluetooth or other wireless means to a gateway, such as a smartphone, from a wearable such as a smartwatch, a wristband, or a textile [88]. In addition, considering that the time overhead of the classification in each time window is very small (on average 0.1056 s, with a standard deviation of 0.0091 s), EDA and PPG are considered feasible alternatives for the first practical applications of the real-time assessment of mental workload in users browsing Web pages.

7. Conclusion

The study of human behavior and physiology when performing human-computer interactions activities is complex due to the multiple factors that affect each person in their performance and behavior with regard to this class of tasks. This research assesses the behavior of a user in the simple task of browsing freely through a fictitious Web page created specifically for this study, using psychophysiological sensors.

It is shown that for the complete data set, that is, considering the complete universe of windows of all the participants, pupil diameter – as a measure of mental workload – is significantly lower in the transition windows than in the active windows, with a significance of $p - value = 0.00184$ for a 95 % confidence interval. Therefore, patterns of low mental workload states are identified, and the hypothesis (H1) that it is indeed possible to measure mental workload in Web browsing activities and, moreover, that the mental workload of the user decreases in the transition from the analysis of one Web element to another while browsing freely is verified.

The unsupervised model of k-means analysis as a data mining technique is applied to the mean and variance of pupil dilation, based on which the Web browsing task involves on average four levels of mental workload. Thus, it is concluded that there are several mental workload states that can be determined (RQ1).

To classify levels of mental workload, the MLP neural network is used, which obtains a result of 88.46 % accuracy on average (RQ2). In addition, the electroencephalogram is the sensor that obtains the best results, classifying with 88.78 % accuracy. If the EEG is combined with the PPG and EDA, the accuracy of the classification rises to 95.73 %.

In terms of future lines of research, it is proposed to use the data to study Web users' mood behavior together with their cognitive behavior. In addition, it is proposed to focus the research on the EEG sensor, which showed superior performance, using other analytical approaches, such as wavelets and/or ERP, to determine the most relevant involved brain areas.

Supplementary Materials: The following dataset was submitted as supplementary material: "dataset_Sensors". It contains all the preprocessed, gathered sensor data for each participant.

Acknowledgments: This work was financed by the CONICYT FONDECYT program, project code: 11130252. The authors are thankful for the continuous support of "Instituto Sistemas Complejos de Ingeniería" (CONICYT: Proyecto Basal FBO16). In addition, the authors would like to thank Todd Pezzuti, Marcos Orchard, Javier Ruiz-del-Solar and all the experiment participants.

Author Contributions: Angel Jimenez-Molina initiated the research, conceived and designed the experiments; Cristian Retamal performed the experiments; Cristian Retamal and Hernan Lira analyzed the data; Angel Jimenez-Molina and Cristian Retamal wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Landauer, T. *The trouble with computers: Usefulness, usability, and productivity*; MIT Press, 1995; Vol. 11; ISBN 0262121867.
2. Gwizdka, J. Distribution of cognitive load in Web search. *J. Am. Soc. Inf. Sci. Technol.* **2010**, *61*, 2167–2187, doi:10.1002/asi.21385.
3. Albers, M. J. Tapping as a measure of cognitive load and website usability. *Proc. 29th ACM Int. Conf. Des. Commun. - SIGDOC '11* **2011**, 25, doi:10.1145/2038476.2038481.
4. Longo, L.; Rusconi, F.; Noce, L.; Barrett, S. The importance of Human Mental Workload in Web Design. *Proc. 8th Int. Conf. Web Inf. Syst. Technol.* **2012**, 403–409, doi:10.5220/0003960204030409.
5. Buscher, G.; Cutrell, E.; Morris, M. R.; Buscher, G., Cutrell, E. and Morris, M. R. 2009. What do you see when you're surfing?: Using eye tracking to predict salient regions of web pages. In *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*; In Proceedings CHI 2009, 21-30.: New York, New York, USA, 2009; p. 21.
6. Buscher, G.; Dumais, S. T.; Cutrell, E.; Buscher, G., Dumais, S. and Cutrell, E. 2010. The good, the bad, and the random: an eye-tracking study of ad quality in web search. In *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval - SIGIR '10*; In Proceedings SIGIR 2010.: New York, New York, USA, 2010; p. 42.
7. Dumais, S. T.; Buscher, G.; Cutrell, E. Individual differences in gaze patterns for web search. In *Proceeding of the third symposium on Information interaction in context - IIiX '10*; ACM Press: New York, New York, USA, 2010; p. 185.
8. Sweller, J. Cognitive Load During Problem Solving: Effects on Learning. *Cogn. Sci.* **1988**, *12*, 257–285, doi:10.1207/s15516709cog1202_4.
9. Wickens, C. D. Multiple resources and performance prediction. *Theor. Issues Ergon. Sci.* **2002**, *3*, 159–177, doi:10.1080/14639220210123806.
10. Parasuraman, R.; Sheridan, T. B.; Wickens, C. D. Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs. *J. Cogn. Eng. Decis. Mak.* **2008**, *2*, 140–160, doi:10.1518/155534308X284417.
11. Arbib, M. A.; Bonaiuto, J. J. *From neuron to cognition via computational neuroscience*; ISBN 9780262034968.
12. Jung J., Maier A., Gross A., Ruiz N., Chen F., Yin B. 2011. *Investigating the Effect of Cognitive Load on UX:A Driving Study.*; AutomotiveUI'11, November 29-December 2, 2011, Salzburg, Austria Adjunct

Proceedings.

13. Paas, F.; Tuovinen, J. E.; Tabbers, H.; Van Gerven, P. W. M. Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educ. Psychol.* **2003**, *38*, 63–71, doi:10.1207/S15326985EP3801_8.
14. Nguyen, Frank, Colvin-Clark, R. Learning Solutions, November 7, 2005. 2005,.
15. van Merriënboer, J. J. G.; Ayres, P. Research on cognitive load theory and its design implications for e-learning. *Educ. Technol. Res. Dev.* **2005**, *53*, 5–13, doi:10.1007/BF02504793.
16. Fritz, T.; Begel, A.; Müller, S. C.; Yigit-Elliott, S.; Züger, M. Using psycho-physiological measures to assess task difficulty in software development. In *Proceedings of the 36th International Conference on Software Engineering - ICSE 2014*; International Conference on Software Engineering (ICSE), Hyderabad, pp 402-413.: New York, New York, USA, 2014; pp. 402–413.
17. Haapalainen, E.; Kim, S.; Forlizzi, J. F.; Dey, A. K. Psycho-physiological measures for assessing cognitive load. In *Proceedings of the 12th ACM international conference on Ubiquitous computing - Ubicomp '10*; Proceedings of the 12th ACM International Conference on Ubiquitous Computing, pp. 301–310., 2010.
18. Ikehara, C. S.; Crosby, M. E. Assessing Cognitive Load with Physiological Sensors. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*; IEEE; p. 295a–295a.
19. Hogervorst, M. A.; Brouwer, A.-M.; van Erp, J. B. F. Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. *Front. Neurosci.* **2014**, *8*, 322, doi:10.3389/fnins.2014.00322.
20. Khusainov, R.; Azzi, D.; Achumba, I. E.; Bersch, S. D. Real-time human ambulation, activity, and physiological monitoring: Taxonomy of issues, techniques, applications, challenges and limitations. *Sensors (Switzerland)* 2013, *13*, 12852–12902.
21. Gazzaniga, M. S.; Ivry, R. B.; Mangun, G. R. Cognitive Neuroscience: The Biology of the Mind. *Q. Rev. Biol.* 2009, *84*, 196–197.
22. Davis, Stephen, Palladino, J.; Stephen F. Davis, Joseph J. Palladino, Christopherson, K. *Psychology*; Pearson, 2012; ISBN 0205848559.
23. Bailey, B. P.; Iqbal, S. T. Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Trans. Comput. Interact.* **2008**, *14*, 1–28, doi:10.1145/1314683.1314689.
24. Hess, E. H.; Polt, J. M. Pupil Size in Relation to Mental Activity during Simple Problem-Solving. *Science (80-.).* **1964**, *143*, 1190–1192, doi:10.1126/science.143.3611.1190.
25. Hahnemann, D.; Beatty, J.; Kahneman, D.; Beatty, J. Pupillary responses In a pitch-discrimination task. *Percept. Psychophys.* **1967**, *2*, 101–105, doi:10.3758/BF03210302.
26. Juris, M.; Velden, M. The pupillary response to mental overload. *Physiol. Psychol.* **1977**, *5*, 421–424, doi:10.3758/BF03337847.
27. Beatty, J. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychol. Bull.* **1982**, *91*, 276–92.
28. Hoeks, B.; Levelt, W. J. M. Pupillary dilation as a measure of attention: a quantitative system analysis. *Behav. Res. Methods, Instruments, Comput.* **1993**, *25*, 16–26, doi:10.3758/BF03204445.
29. Nakayama, M.; Takahashi, K.; Shimizu, Y. The act of task difficulty and eye-movement frequency for the "Oculo-motor indices." In *Proceedings of the symposium on Eye tracking research & applications - ETRA '02*; Proceedings of Eye Tracking Research and Applications..: New York, New York, USA, 2002; p. 37.

30. Li, G.; Lee, B. L.; Chung, W. Y. Smartwatch-Based Wearable EEG System for Driver Drowsiness Detection. *IEEE Sens. J.* **2015**, *15*, 7169–7180, doi:10.1109/JSEN.2015.2473679.
31. Brockmeier, A. J.; Príncipe, J. C. Learning recurrent waveforms within EEGs. *IEEE Trans. Biomed. Eng.* **2016**, *63*, 43–54, doi:10.1109/TBME.2015.2499241.
32. SmartCap Technologies Available online: https://www.smartcaptech.com/ (accessed on Dec 2, 2017).
33. Oulasvirta, A.; Tamminen, S.; Roto, V.; Kuorelahti, J. Interaction in 4-Second Bursts: The Fragmented Nature of Attentional Resources in Mobile HCI. *Proc. SIGCHI Conf. Hum. factors Comput. Syst. - CHI '05* **2005**, 919, doi:10.1145/1054972.1055101.
34. Fawcett, J. M.; Risko, E. F.; Kingstone, A. *The handbook of attention*; 2015; ISBN 9780262029698.
35. Navon, D.; Gopher, D. On the economy of the human-processing system. *Psychol. Rev.* **1979**, *86*, 214–255, doi:10.1037/0033-295X.86.3.214.
36. Matthews, G.; Warm, J. S.; Smith, A. P. Task Engagement and Attentional Resources. *Hum. Factors J. Hum. Factors Ergon. Soc.* **2017**, *59*, 44–61, doi:10.1177/0018720816673782.
37. Miettinen, M.; Oulasvirta, A. Predicting time-sharing in mobile interaction. *User Model. User-adapt. Interact.* **2007**, *17*, 475–510, doi:10.1007/s11257-007-9033-x.
38. Jimenez-Molina, A.; Ko, I.-Y. Cognitive resource-aware unobtrusive service provisioning in ambient intelligence environments. *J. Ambient Intell. Smart Environ.* **2015**, *7*, doi:10.3233/AIS-140299.
39. Norman, D. A.; Draper, S. W. *USER CENTERED SYSTEM DESIGN New Perspectives on Human-Computer Interaction Library of Congress Cataloging-in-Publication Data*; 1986; ISBN 0-89859-872-9.
40. Monk, C. A.; Boehm-Davis, D. A.; Trafton, J. G. The Attentional Costs of Interrupting Task Performance at Various Stages. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2002**, *46*, 1824–1828, doi:10.1177/154193120204602210.
41. Young, M. S.; Brookhuis, K. A.; Wickens, C. D.; Hancock, P. A. State of science: mental workload in ergonomics. *Ergonomics* **2015**, *58*, 1–17, doi:10.1080/00140139.2014.956151.
42. Matthews, G.; Reinerman-Jones, L. E.; Barber, D. J.; Abich, J. The Psychometrics of Mental Workload: Multiple Measures Are Sensitive but Divergent. *Hum. Factors J. Hum. Factors Ergon. Soc.* **2015**, *57*, 125–143, doi:10.1177/0018720814539505.
43. Greco, A.; Valenza, G.; Scilingo, E. P. Emotions and Mood States: Modeling, Elicitation, and Recognition. In *Advances in Electrodermal Activity Processing with Applications for Mental Health*; Springer International Publishing: Cham, 2016; pp. 45–54.
44. Guastello, S. J.; Shircel, A.; Malon, M.; Timm, P. Individual differences in the experience of cognitive workload. *Theor. Issues Ergon. Sci.* **2015**, *16*, 20–52, doi:10.1080/1463922X.2013.869371.
45. Dhami, M. K.; Hertwig, R.; Hoffrage, U. The Role of Representative Design in an Ecological Approach to Cognition. *Psychol. Bull.* **2004**, *130*, 959–988, doi:10.1037/0033-2909.130.6.959.
46. Hart, S. G.; Staveland, L. E. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Adv. Psychol.* **1988**, *52*, 139–183, doi:10.1016/S0166-4115(08)62386-9.
47. Halverson, T.; Estepp, J.; Christensen, J.; Monnin, J. Classifying Workload with Eye Movements in a Complex Task. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2012**, *56*, 168–172, doi:10.1177/1071181312561012.
48. Mehler, B.; Reimer, B.; Coughlin, J. F. Sensitivity of Physiological Measures for Detecting Systematic Variations in Cognitive Demand From a Working Memory Task. *Hum. Factors* **2012**, *54*, 396–412, doi:10.1177/0018720812442086.
49. Wang, S.; Gwizdka, J.; Chaovalitwongse, W. A. Using Wireless EEG Signals to Assess Memory

Workload in the n-Back Task. *IEEE Trans. Human-Machine Syst.* **2016**, *46*, 424–435, doi:10.1109/THMS.2015.2476818.

50. Cacioppo, Tassinary, Berntson. 2007. *The Handbook of Psychophysiology.*; Cacioppo, J. T., Tassinary, L. G., Berntson, G. G., Eds.; Cambridge University Press. Third Edition: Cambridge, 2017; ISBN 9781107415782.
51. Walsh, E. G.; Marshall, J. *Physiology of the nervous system*; Longmans London, 1957;
52. Mohino-Herranz, I.; Gil-Pita, R.; Ferreira, J.; Rosa-Zurera, M.; Seoane, F. Assessment of Mental, Emotional and Physical Stress through Analysis of Physiological Signals Using Smartphones. *Sensors* **2015**, *15*, 25607–25627, doi:10.3390/s151025607.
53. Nardelli, M.; Valenza, G.; Cristea, I. A.; Gentili, C.; Cotet, C.; David, D.; Lanata, A.; Scilingo, E. P. Characterizing psychological dimensions in non-pathological subjects through autonomic nervous system dynamics. *Front. Comput. Neurosci.* **2015**, *9*, 37, doi:10.3389/fncom.2015.00037.
54. Banaee, H.; Ahmed, M. U.; Loutfi, A. Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges. *Sensors.* **2013**, *13*, 17472–17500, doi:10.3390/s131217472.
55. Goldberg, J. H.; Wichansky, A. M. Eye Tracking in Usability Evaluation. In *The Mind's Eye*; In Hyona, J., Radach, R., & Deubel, H. (Eds.), The mind's eye: Cognitive and applied aspects of eye movement research (pp. 493-516). Amsterdam:Elsevier., 2003; pp. 493–516 ISBN 9780444510204.
56. Kuipers, J.-R.; Thierry, G. ERP-pupil size correlations reveal how bilingualism enhances cognitive flexibility. *Cortex* **2013**, *49*, 2853–2860.
57. Steinhauer, S. R.; Siegle, G. J.; Condray, R.; Pless, M. Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. *Int. J. Psychophysiol.* **2004**, *52*, 77–86.
58. Beatty, J.; Lucero-Wagoner, B. *The Pupilary System*; Handbook of psychophysiology, 2nd ed., (pp. 142-162). New York, NY, US: Cambridge University Press., 2000;
59. Zhou, J.; Sun, J.; Chen, F.; Wang, Y.; Taib, R. Measurable Decision Making with GSR and Pupillary Analysis for Intelligent User Interface. *ACM Trans. Comput. Interact. - Spec. Issue Physiol. Comput. Human-Computer Interact.* **2015**, *21*, 1–23, doi:10.1145/2687924.
60. Andreassi, J. L. *sychophysiology: Human Behaviour and Physiological response*; 2000; ISBN 080582832X.
61. Brioschi, M. L.; Siqueira, J. T. T. de; Teixeira, M. J. Avanços da imagem infravermelha na disfunção temporomandibular. *JBA j. bras. oclus. ATM dor orofac* **2006**, *6*, 34–41.
62. Genno, H.; Ishikawa, K.; Kanbara, O.; Kikumoto, M.; Fujiwara, Y.; Suzuki, R.; Osumi, M. Using facial skin temperature to objectively evaluate sensations. *Int. J. Ind. Ergon.* **1997**, *19*, 161–171, doi:10.1016/S0169-8141(96)00011-X.
63. Angeles-Aquino, F.; Cantu-Chapa, A. Subsurface Upper Jurassic stratigraphy in the Campeche Shelf, Gulf of Mexico. *West. Gulf Mex. Basin. Tectonics, Sediment. Basins, Pet. Syst.* **2001**, *75*, 343–352, doi:fzbgvk.
64. Latorella, K. A. Effects of Modality on Interrupted Flight Deck Performance: Implications for Data Link. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **1998**, *42*, 87–91, doi:10.1177/154193129804200120.
65. Speier, C.; Valacich, J. S.; Vessey, I. The Influence of Task Interruption on Individual Decision Making: An Information Overload Perspective. *Decis. Sci.* **1999**, *30*, 337–360, doi:10.1111/j.1540-5915.1999.tb01613.x.
66. Zijlstra, F. R. H.; Roe, R. A.; Leonora, A. B.; Krediet, I. Temporal factors in mental work: Effects of interrupted activities. *J. Occup. Organ. Psychol.* **1999**, *72*, 163–185, doi:10.1348/096317999166581.
67. Adamczyk, P. D.; Bailey, B. P. If not now when?: the effects of interruption at different moments

within task execution. *Proc. SIGCHI Conf. Hum. factors Comput. Syst.* **2004**, *6*, 271–278, doi:10.1145/985692.985727.

68. Bailey, B. P.; Konstan, J. A. On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. In *Computers in Human Behavior*; Pergamon, 2006; Vol. 22, pp. 685–708.
69. Miyata, Y.; Norman, D. Psychological issues in support of multiple activities. *User centered Syst. Des.* **1986**, 265–284.
70. Yoshida, Y.; Ohwada, H.; Mizoguchi, F.; Iwasaki, H. Classifying Cognitive Load and Driving Situation with Machine Learning. *Int. J. Mach. Learn. Comput.* **2014**, *4*, 210–215, doi:10.7763/IJMLC.2014.V4.414.
71. Shi et al. 2007. *Galvanic Skin Response (GSR) as an Index of Cognitive Load.*; CHI Extended Abstracts on Human Factors in Computing Systems.;
72. Nourbakhsh, N.; Wang, Y.; Chen, F.; Calvo, R. A. Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks. In *Proceedings of the 24th Australian Computer-Human Interaction Conference on - OzCHI '12*; Conference: Proceedings of the 24th Australian Computer-Human Interaction Conference.: New York, New York, USA, 2012; pp. 420–423.
73. Xu, J.; Wang, Y.; Chen, F.; Choi, E. Pupillary Response Based Cognitive Workload Measurement under Luminance Changes. In; Interact 2011, Part II. LNCS, vol. 6947:178-185. Springer, Heidelberg., 2011; pp. 178–185.
74. Navalpakkam, V.; Churchill, E. Mouse Tracking: Measuring and Predicting Users' Experience of Web-based Content. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*; In Proceedings of the ACM SIGCHI conference on human factors in computing systems (pp. 2963-2972). ACM Press.: New York, New York, USA, 2012; p. 2963.
75. Wearable Sensor Technology Available online: https://www.shimmersensing.com/ (accessed on Dec 2, 2017).
76. BITalino - Biomedical Equipment Available online: http://bitalino.com/en/ (accessed on Dec 2, 2017).
77. Emotiv Available online: https://www.emotiv.com/ (accessed on Dec 2, 2017).
78. Eye tracking technology for research - Tobii Pro Available online: https://www.tobiipro.com/ (accessed on Dec 2, 2017).
79. Villarejo, M. V.; Zapirain, B. G.; Zorrilla, A. M. A stress sensor based on galvanic skin response (GSR) controlled by ZigBee. *Sensors.* **2012**, *12*, 6075–6101, doi:10.3390/s120506075.
80. Ye, Y.; He, W.; Cheng, Y.; Huang, W.; Zhang, Z. A Robust Random Forest-Based Approach for Heart Rate Monitoring Using Photoplethysmography Signal Contaminated by Intense Motion Artifacts. *Sensors*. **2017**, *17*, 385, doi:10.3390/s17020385.
81. Nelson, C. V; Geselowitz, D. B. *The theoretical basis of electrocardiology*; Oxford University Press, 1976; Vol. 1; ISBN 019857374X.
82. Macfarlane, P. W.; Lawrie, T. D. V. *Comprehensive electrocardiology: theory and practice in health and disease*; Pergamon, 1989; Vol. 2;.
83. OpenSignals | Data Visualization Software | Bitalino Available online: http://bitalino.com/en/software/ (accessed on Dec 2, 2017).
84. Tobii Pro Studio eye tracking software, dedicated for UX Available online: https://www.tobiipro.com/product-listing/tobii-pro-studio/ (accessed on Dec 2, 2017).
85. iMotions Biometric Research Platform. 2016. *Eye Tracking Pocket Guide.*;
86. Parsons, H. M. What Happened at Hawthorne?: New evidence suggests the Hawthorne effect resulted

from operant reinforcement contingencies. *Science (80-.).* **1974**, *183*, 922–932, doi:10.1126/science.183.4128.922.

87. Loyola, P.; Martinez, G.; Muñoz, K.; Velásquez, J. D.; Maldonado, P.; Couve, A. Combining eye tracking and pupillary dilation analysis to identify Website Key Objects. *Neurocomputing* **2015**, *168*, 179–189, doi:10.1016/J.NEUCOM.2015.05.108.

88. iMotions Biometric Research Platform. 2016. *GSR Pocket Guide.*;

89. Li, X.; Chen, X.; Yan, Y.; Wei, W.; Wang, Z. J. Classification of EEG signals using a multiple kernel learning support vector machine. *Sensors.* **2014**, *14*, 12784–12802, doi:10.3390/s140712784.

90. Fauvel, S.; Ward, R. K. An energy efficient compressed sensing framework for the compression of electroencephalogram signals. *Sensors.* **2014**, *14*, 1474–1496, doi:10.3390/s140101474.

91. Quintero-Rincón, A. Preprocesamiento de EEG con Filtros Hampel. *ARGENCON 2012 - IEEE Lat. Am. Trans.*

92. Guyon, I.; Elisseeff, A. An Introduction to Feature Extraction. In *Feature Extraction*; Feature Extraction. Foundations and Applications 207:1-25: Berlin, Heidelberg, 2006; pp. 1–25.

93. Betella, A.; Zucca, R.; Cetnarski, R.; Greco, A.; LanatÃ, A.; Mazzei, D.; Tognetti, A.; Arsiwalla, X. D.; Omedas, P.; De Rossi, D.; Verschure, P. F. M. J. Inference of human affective states from psychophysiological measurements extracted under ecologically valid conditions. *Front. Neurosci.* **2014**, *8*, 286, doi:10.3389/fnins.2014.00286.

94. iMotions Biometric Research Platform. 2016. *EEG Pocket Guide.*;

95. Cohen, M. X. *Analyzing neutral time series data. Theory and Practice*; Massachusetts Institute of Technology., 2014;

96. Xu, R.; Wunsch, D. C. *Clustering*; IEEE Press, 2009, John Wiley & Sons, Inc., Hoboken, New Jersey, USA. (pp. 263-278), 2009; ISBN 0470276800.

97. Zumel, N.; Mount, J. (Computational scientist) *Practical data science with R*; Manning Publications Co. 1st Edition.; ISBN 9781617291562.

98. Calinski, T.; Harabasz, J. A dendrite method for cluster analysis. *Commun. Stat. - Theory Methods* **1974**, *3*, 1–27, doi:10.1080/03610927408827101.

99. Kai-Quan Shen, K.-Q.; Chong-Jin Ong, C.-J.; Xiao-Ping Li, X.-P.; Zheng Hui, Z.; Wilder-Smith, E. P. V. A Feature Selection Method for Multilevel Mental Fatigue EEG Classification. *IEEE Trans. Biomed. Eng.* **2007**, *54*, 1231–1237, doi:10.1109/TBME.2007.890733.

100. Candel, A.; Lanford, J.; LeDell, E.; Parmar, V.; Arora, A. *Deep Learning with H2O*; Published by H2O.ai, Inc. Third Edition., 2015;

101. Hinton, G. E.; Srivastava, N.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. R. Improving neural networks by preventing co-adaptation of feature detectors. **2012**.