Measuring Photoplethysmogram-Based Stress-Induced Vascular Response Index to Assess Cognitive Load and Stress

Yongqiang Lyu[1], Xiaomin Luo[2], Jun Zhou[1], Chun Yu[1], Congcong Miao[1], Tong Wang[1], Yuanchun Shi[1], Ken-ichi Kameyama[3]

[1]TNLIST, Research Institute of Information Technology, Tsinghua University, Beijing China
[2]BGI (Wuhan) Translational Research Center, Wuhan, China
[3]Interactive Media Lab., Corporate R&D Center, Toshiba Corp., Tokyo, Japan
luyq@tsinghua.edu.cn, luoxm@ luoxm@genomics.org.cn, kenichi.kameyama@toshiba.co.jp

ABSTRACT
Quantitative assessment for cognitive load and mental stress is very important in optimizing human-computer system designs to improve performance and efficiency. Traditional physiological measures, such as heart rate variation (HRV), blood pressure and electrodermal activity (EDA), are widely used but still have limitations in sensitivity, reliability and usability. In this study, we propose a novel photoplethysmogram-based stress induced vascular index (sVRI) to measure cognitive load and stress. We also provide the basic methodology and detailed algorithm framework. We employed a classic experiment with three levels of task difficulty and three stages of testing period to verify the new measure. Compared with the blood pressure, heart rate and HRV components recorded simultaneously, the sVRI reached the same level of significance on the effect of task difficulty/period as the most significant other measure. Our findings showed sVRI's potential as a sensitive, reliable and usable parameter.

Author Keywords
Cognitive Load; Mental Effort; Stress; Stress-Induced Vascular Response Index (sVRI); Photoplethysmogram.

ACM Classification Keywords
H.1.2 [User/Machine Systems Subjects]: Human information processing.

INTRODUCTION
Cognitive load refers to the load that performing a particular task imposes on the human cognitive system [1]. It reflects the demand for both the psychological and physiological resources of cognition, such as working memory, the processing units for visual/spatial and auditory/verbal information, and mental intent. Assessing the cognitive load that a particular system or task imposes on users is very important in human-computer interaction (HCI) design, as the information on cognitive load can be used to improve these systems [2]. Dynamic and real-time assessment of cognitive load is particularly useful for capturing the detailed changes subjects undergo during a task, which supports a precise analysis as well as improvement on design components [2-4] and the collection of continuous psychophysiological data for smart HCI applications [5-7].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

CHI 2015, April 18 - 23 2015, Seoul, Republic of Korea
Copyright 2015 ACM 978-1-4503-3145-6/15/04…$15.00
http://dx.doi.org/10.1145/2702123.2702399

In practice, the *mental effort*, which refers to the cognitive resources actually allocated to accommodate a task [1], is often used to reflect the actual cognitive load while participants are working on a task. Mental effort is linked to both cognitive capacity (cognitive resources available) and autonomic intent (arousal). Due to variations in arousal, the same person may perform differently on the same task at different times. Cognitive load or mental effort causes mental *stress*, which occurs when the task's cognitive load exceeds the subject's cognitive capacity, or may also occur when the subjects feel the cognitive load will exceed their cognitive capacity [8]. Mental stress often accompanies the cognitive process when subjects exert mental effort, e.g. executing tasks. This stress may give rise to emotions like anxiety, fatigue, and anger and has an influence on human health [5, 6]. Stress measurement is also very important for affective computing and health-related applications.

Researchers may employ subjective *work/task load* questionnaires [9, 10] to assess cognitive load, or stress state questionnaires [11] to assess the stress. Compared to these subjective measures, objective measures, including behavioral methods, such as behavior measurement and behavioral performance measurement [4, 12, 13], and physiological measures [14-23], such as heart rate variation, blood pressure and skin conductivity, can assess cognitive load and stress more sensitively and accurately. Nevertheless, these measures still suffer from some notable limitations in sensitivity, reliability and usability. Researchers often need to combine multiple measures together to improve the sensitivity and reliability of these measures. However, combining measures in this way is not usable in ubiquitous applications. In the era of ubiquitous applications and wearable devices [21-23], reliable and usable measurement of cognitive load and stress is critical.

In this study, we focus on studying a sensitive, reliable and relatively usable (real-time, low-cost, noninvasive and hopefully wearable) physiological measure to assess cognitive load and stress. We propose a PPG-based *stress-induced vascular responses index* (sVRI) to realize the goal by measuring the peripheral vasoconstriction tone (as Figure 1 shows).

Figure 1. A standard PPG waveform

The main contributions of our work are:

- As an addition to traditional measures, we propose the PPG-based stress-induced vascular response index (sVRI) as a novel physiological measure to assess cognitive load and mental stress, which is verified to be able to differentiate effortful conditions associated with tasks of varying difficulties.
- Differing from previous work, the sVRI employs a relative amplitude ratio between two contours in a PPG waveform, which is sensitive and relatively reliable. The sVRI can output in a nearly real-time fashion at a data rate of one per heartbeat (of the subject). We also found a low standard deviation of sVRI for participants compared to other measures.
- The sVRI shows promising usability in ubiquitous scenarios because of the easy generation of PPG in either a transmission mode or a reflection mode at peripheral parts of body (e.g. fingers, earlobes, forehead) by a low-cost optical-sensing apparatus (e.g. oximeter).

The rest of the paper is organized as follows. First, we analyze the related work on the measures of cognitive load and stress. Second, we provide the basic methodology and the detailed algorithm framework for sVRI. Third, we employ a classic cognitive experiment involving arithmetic calculation of three difficulty levels to examine the performance of sVRI as well as other known physiological measures including blood pressure, heart rate and heart rate variation (HRV). Next, we analyze the experimental results and discuss the limitations of sVRI and the study. Finally, we conclude the study.

RELATED WORK

Subjective measures

We classify the measures available to assess cognitive load and stress into two main categories: subjective and objective. Subjective *work/task load* questionnaires [9, 10] can assess cognitive load by evaluating the characteristics of the tasks that people process, such as task form, complexity, time demand and effort required. Stress state questionnaires [11] can assess mental stress by evaluating three primary categories: task engagement, distress and worry. Although subjective measures may distinguish the cognitive load and stress, they heavily rely on subjective recall and descriptions from participants after the task.

Objective measures: behavioral methods

Although objective methods may not distinguish cognitive load/effort from stress, they are more sensitive, reliable, and usable than subjective methods. The objective methods most commonly used are behavioral methods and physiological methods. Behavioral methods include behavior measurement and behavioral performance measurement. Behavior measurement focuses on user behaviors, such as keyboard and mouse behaviors [5, 6], touch screen behaviors [12], and physical factors of body [13]. Behavior measurement is promising for use in ubiquitous applications in spite of current drawbacks in sensitivity and accuracy. It is also meaningless to assess the cognitive load or stress in a single task only by measuring user performance due to the lack of a control for mental effort investment. However, dual-task-based performance measurement is available. Dual-task analysis is based on the assumption of limited cognitive resources that can be allocated flexibly to the primary task and the secondary task performed simultaneously [4]. The cognitive load induced by one of the tasks is thus assessed by measuring the performance change of that task or the other task. The main disadvantages of dual-task analysis lie in the need for dedicated design of homogeneous dual tasks and the sensitivity and reliability issues due to variable participants' attention [4].

Objective measures: physiological methods

Compared to behavioral measures, physiological measures can directly, sensitively, and reliably reflect subject's internal response to cognitive load and stress. The human body is controlled by nervous systems including the central nervous system (CNS), the autonomic nervous system (ANS), the somatic nervous system, etc. The ANS controls the functions of most organs and thus influences changes in related vital signs [3, 14], such as electrocardiograph (ECG), heart rate (HR), heart rate variation (HRV), pulse rate (PR), pulse rate variation (PRV) respiration, blood pressure (BP), pupil dilation and eye blink. The ANS is composed of the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS), which inhibit each other. The SNS is responsible for emotional arousal and the physiological adjustments supporting performance, and its activities increase during tasks that require active coping. The PNS inhibits SNS activities and increases its functions during tasks that require passive coping (e.g. response) [16]. Some vital signs are controlled by both the SNS and PNS (dually innervated), while some are controlled only by the sympathetic nervous system (sympathetic-only).

Cognitive load and stress relate to both the SNS and PNS. The dually innervated vital signs actually reflect the production of the both nervous systems. Different modes of the nervous regulations of the two systems may produce the same overt (observed) vital-sign results [15], which may not be informative as to the underlying mechanism [16]. Isolating measures with a specific mapping to the SNS and PNS and then combining them may help assess the cognition and mental process further [16]. SNS activation increases as the central processing complexity of a task increases, and sympathetic-only vital signs are regarded as direct and sensitive indices for autonomic effort, as well as resulting stress in comparison with the PNS-related methods [3, 15-17].

Based on the nervous or biological systems that physiological methods utilize to assess cognitive load or stress, they are split into ANS measures and non-ANS measures. ANS measures are further broken up into dually innervated measures and the sympathetic-only measures.

Non-ANS measures

Some physiological measures available for assessing cognitive load or stress are linked to non-ANS nervous systems. For example, Electroencephalography (EEG) is linked to the central nervous system (CNS) and electromyogram (EMG) is controlled by both the somatic nervous system and the autonomic nervous system [18]. In addition to EEG, neuroimaging techniques, such as positron-emission tomography (PET) and functional magnetic resonance imaging (fMRI) are also very important for measuring brain activation during task execution [4]. Researchers may also employ biomarker-identifying techniques to examine the psychophysiological activities, for example, examining the cortisol or adrenaline in saliva or blood samples [19]. The main drawbacks of non-ANS methods are the complexity of usage, high cost and possible insensitivity to the level of mental effort investment [16].

Dually innervated measures

A range of physiological measures are based on dually innervated vital signs [3, 15], such as the HR/HRV (ECG), respiration, blood pressure, PR/PRV (pulse) and eye activities. Because these measures still suffer from sensitivity and reliability issues, researchers often employ multiple measures to examine the cognitive/mental process of participants in order to improve the assessment sensitivity and reliability. For example, Brouwer et al. [3] used EEG, HRV, respiration, skin conductance, pupil size and eye blink to study the mental effects induced by task difficulties, and found different sensitivities and reliabilities among the measures. Tan et al. proposed a smart video-medicated collaboration system with the user's mental feedback indicated by electrodermal activity (EDA), blood pressure and respiration [7]. Hjortskov et al. [20] also compared the performance of HRV and blood pressure on the effect of mental stress.

These measures have some other drawbacks as well. HRV is inconvenient to measure (multiple ECG leads needed), weak in real time (need at least 2min integral), sensitive to noise (breathing and motion artifacts), and complicated to analyze (manual artifacts corrections are often required and there are multiple components to analyze) [6, 16, 20]. PRV [21, 22] is also dually inverted, which is originated from the heart but may only have comparable sensitivity with HRV for healthy users at rest [24]. Heart rate, respiration, blood pressure (BP) and eye activities are vulnerable to internal and external noise, require obtrusive measurements, and have limitations in their ability to track information in real-time (e.g., BP).

Sympathetic-only measures

Sweat glands are controlled only by SNS, which affect the skin conductivity that is often measured by the electrodermal activity (EDA) method (also known as the galvanic skin response). The effectiveness of EDA has been demonstrated in previous studies [3, 7, 14]. However, EDA may have notable individual differences and is susceptible to the influence from skin conditions as well as ambient temperature and humidity. Over the course of a long testing process, the sweat that leads to EDA changes may be greatly influenced by ambient factors, which can result in sizeable differences between the recorded and actual cognitive and mental response of the user.

Peripheral vasoconstriction is another sympathetic-only vital sign. Iani et al. [16] investigated whether peripheral arterial tone reflects changes in mental effort. Using a modified version of the Sternberg memory task, they found that peripheral arterial tone measured by a pressure-based plethysmograph amplitude was sensitive to task difficulty. Luo et al. [19] reported a strong relationship between participants' PPG amplitudes and cortisol levels. Cortisol is regarded as the biomarker of mental stress. This study verified the physiological linkage between PPG amplitudes and mental stress induced by negative stimuli. Signal-improving techniques for guarding PPG against motion artifacts [25, 26] are also well studied and have been applied in clinical-standard pulse oximeters. Compared with pressure-based plethysmograph measurement, PPG measurement may be more sensitive and stable. PPG amplitudes can also be obtained using only minimal contact, e.g. the surface-sensing method in reflection mode [23], or without contact, e.g. image-based methods via a webcam [21]. PPG amplitudes in the time domain can also be analyzed after converted to the frequency domain [27]. However, PPG amplitudes may have notable temporal changes and individual differences, which can cause severe reliability and usability issues.

STRESS-INDUCED VASCULAR RESPONSE INDEX

In this study, we propose a PPG-based *stress-induced vascular response index* (sVRI) to assess cognitive load (through mental effort) and stress. This index measures a relative ratio of two contours in a PPG waveform, which is sensitive and has limited individual differences. The algorithms for the sVRI also guarantee its reliability and real-time updating.

The featured waveform patterns of sVRI

The effect of mental effort and stress results in peripheral vasoconstriction that is reflected in PPG appearance. Compared with a normal PPG waveform when the participant is in rest (without mental effort), the extracted patterns denoted by the average amplitudes of the two contours as shown in Figure 2, A1 and A2, can show significant differences when a mental task is imposed.

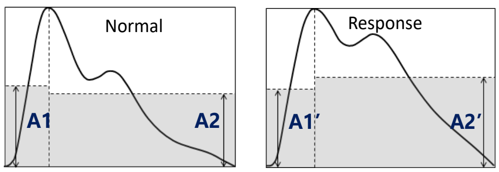


Figure 2. Quantitative response to mental effort, which is reflected in the PPG waveform patterns by the average amplitudes of A1 and A2. A1 and A2 vary (to A1' and A2' respectively) as the vascular response to mental effort.

There is an obvious A2-amplitude elevation in *response* mode in comparison with the *normal* mode. However, A2, A1 and the maximum amplitude of the waveform may have severe temporal changes and individual differences, which causes the notable reliability and usability issues of traditional PPG-based methods. In this study, we define an instant sVRI as:

$$sVRI = \frac{A2}{A1} \quad (1)$$

which can measure dynamic mental effort and stress sensitively and reliably. A1 and A2 are calculated through the area integrals of the two contours of the waveform, and sVRI increases as mental effort or stress increase.

Algorithm framework for reliable and real-time sVRIs

Based on the patterns extracted from the PPG, the sVRI can be implemented as Figure 3 shows. The process starts with the original PPG waveforms, followed by three procedures: the featured pattern extraction (FPE), dynamic sliding windowing (DSW) and statistical index moderating (SIM).

Using the basic signal filtering methods [25, 26], the featured pattern extraction (FPE) can be fed with the high-quality PPG waveforms. In FPE, the average amplitudes of the two contours, A1 and A2, as Figure 2 shows, are calculated by area integral and the instant value of sVRI is calculated using Equation 1.

In order to output reliable and real-time sVRIs, we process the PPG waveforms using the dynamic sliding windowing (DSW) procedure, which frames the PPG waveforms into partially overlapping data windows of varying sizes. One window frames a series of PPG waveforms and outputs one sVRI value.

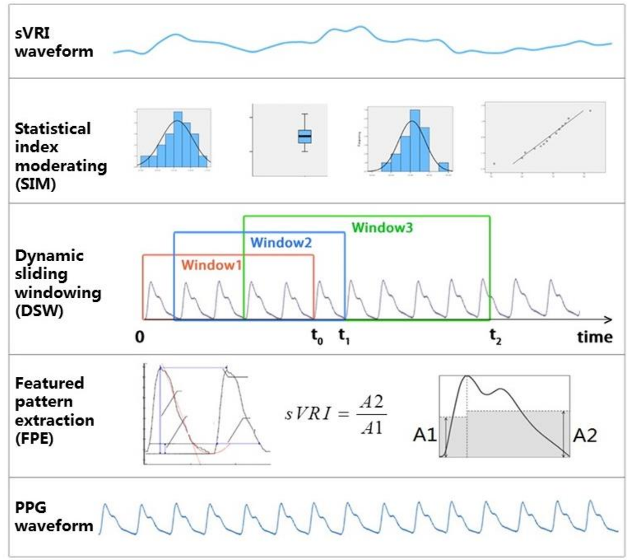


Figure 3. The algorithm framework. Bottom to top: the process begins with the original PPG waveforms; through featured pattern extraction, dynamic windowing and then statistical moderating, the sVRI waveform is achieved.

Although digital filtering in the original PPG (before our algorithms) can mitigate motion artifacts, some distorted waveforms may still exist. Therefore, first, basic digital filtering manipulates the data to select the qualified waveforms to proceed, which is an empirical parameter-based waveform selection scheme. Second, the statistical index moderating (SIM) is invoked to process the window to output a reliable sVRI value corresponding to the window. Third, a new window follows the window before it with certain duration backwards. For instance, as depicted in Figure 3, window 2 (in blue) slides backwards with a pulse beat (similarly with heart rate) after window 1 (in orange). Sometimes, the window might slide with a bigger step to lower the computation required, for example window 3 (in green) move two pulse beats after the window 2, which lowers the data rate for sVRI output. The window sizes may also differ due to the varying proportions of the filtered data. In order to pass enough valid data to the following statistical moderator (SIM), the window sizing process is adopted. For a period with more distorted waveforms, the window size increases and the sVRI data rate decreases.

In order to guarantee the reliability of sVRI, we employ statistical index moderating (SIM) to process and output the sVRI for a window, which assumes a normal distribution holds in the data window. The outliers are filtered by Grubbs' test, which filters out the data too far away from the expectation value (measured by multiples of the standard deviation). The expectation value after filtering is considered as the final output sVRI of the data window. It is noted that this statistical process is actually a pessimistic moderating, which may delay to indicate the sharp mental changes that may be caused by abrupt situations, e.g. a sudden frightening event.

Based on the algorithm flow above, the sVRI is generated in a reliable and real-time fashion with a limited starting delay while the initial data windows are pre-cached.

HYPOTHESES

In this study, we propose the sVRI to assess cognitive load (via mental effort) and stress. We designed an experiment with three task difficulties and three testing periods to evaluate the following specific hypotheses:

Hypotheses A (period effect): The sVRI should show a significant difference between the baseline (before task) and the task-processing period (in task).

Hypotheses B (difficulty effect): The sVRI should show significant differences with the task difficulty varying.

Hypotheses C (monotone effect): The sVRI should increase as mental effort increases, which is induced by greater task difficulty.

METHOD

This study employed a classic mental task of arithmetic calculation using certain digits of number to subtract 6/7/13/77 [28], e.g. using a two-digit number to subtract 7. This task had low requirements for learning [3]. While the participants were performing the tasks, their performance data and physiological data, including sVRIs, were recorded simultaneously. Questionnaires from the participants about their experience were also collected. Finally, the hypotheses were examined based on the analysis of those data.

Participants

There were 40 participants taking part in the experiment. In order to reduce the effects on performance caused by individual capabilities or backgrounds [3], we recruited university students with similar educational backgrounds and arithmetic capabilities. Participants were skilled at arithmetic according to the pretests on their arithmetic capabilities, and none of them had prior experience with the experimental content. Participants were aged between 19 and 25 years old (22 ± 1.7), with 22 males and 18 females. All participants were healthy and did not use any medications. All were right-handed and had normal or corrected-to-normal vision. The experiment was approved by the local ethics committee and performed in according with the ethical standards as laid down in the Declaration of Helsinki. Each participant was paid 40 US dollars.

Apparatus

The sVRI, ECG (for HRV) and blood pressure of the participants were measured throughout the experiment. The PPG for sVRI was measured using a standard clinical finger-clip PPG sensor at a sampling rate of 500HZ [19] and was processed by simultaneous sVRI algorithms. The ECG was measured using a SMOTE PSG device (Compumedics Ltd., Australia) at a sampling rate of 200Hz [24]. Blood pressure was measured by an OMRON HEM-7112 blood pressure monitor. The arithmetic problems were presented on a 19” LCD monitor, and participants used a 101-key standard keyboard to input the answers and then press the Enter key to submit.

Setting

The experiment took place in a sound-attenuated, temperature-controlled and electrically shielded room. Room temperature and humidity during the experiment were held constant. For each participant, the sVRI sensor was placed on the left index finger of the participant; the cuff of the blood pressure monitor was bound to the left upper arm and kept at the same height as the heart of participant. For the ECG, self-adhesive 1 1/2” electrodes with 7% chloride wet gel were attached to the participant’s chest in a standard configuration of leads. Once the experiment started, the participant was asked to try to avoid movement.

Stimuli

This study employed a classic mental task format, i.e. the arithmetic calculation that uses certain digits of number to subtract 6/7/13/77 [28]. The task lasted 6 min and the number of completed calculations as well as the number of correct answers were counted and presented to the participants instantly. We set three levels of task difficulties, i.e. *easy*, *medium* and *hard*, and provided a simple and clear GUI for performing the calculations. The main font was black and was presented on a light grey background. In the easy and medium levels, participants had 10 s maximum to complete each calculation, after which it would be counted as wrong and the next question would appear. Real-time results were presented in the interface. In the easy level, participants were asked to subtract 6 or 7 (at random) from random three-digit numbers; while in the medium level, participants were asked to subtract 13 or 77 (at random) from random four-digit numbers. As a special case in this study, the hard level was more challenging; it required participants to calculate the solution to random four-digit numbers subtracted from random five-digit numbers in 15 s. This case was designed to investigate the cognitive load (mental effort) under very high stressors.

Protocol

Each participant was tested individually on three tests (one for each task difficulty) within a two-week period, and was randomly assigned a task difficulty for each test without being told the level of difficulty. Each test lasted about 50 min as shown in Figure 4. At the beginning of the test, the experimenter provided the general information and instructions about the experiment and informed the participants that their physiological responses would be recorded at rest and during mental tasks. The participants were also asked to treat the experiment seriously and to try their best on the task.

The period before the task lasted 30 min, and was composed of a 20min rest period (including preparation time) and a 10min *pre-task* (baseline) period. After attaching the recording devices, the participant was left alone in the room for about 10 min, and was verbally instructed to rest quietly, move as little as possible, and wait for further instructions. In the following 10min pre-task period, an experimenter (only an experimenter in situ) recorded the participant's physiological data as the baseline before starting the arithmetic task. After that, the participant was instructed to complete the arithmetic calculations in 6 min. This was the *in-task* (working) period. When time was up, the participant was instructed to stop and rest for 10 min, which was the *post-task* (recovery) period. The physiological measures were also recorded during the in-task and post-task. In the final 5min period, the participant was asked to fill out questionnaires assessing the feelings about the tasks and the effort on the tasks.

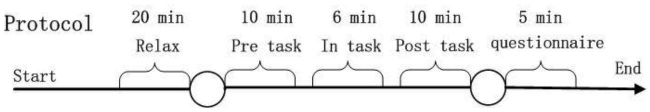


Figure 4. The experiment protocol.

Dependent variables

For each participant, during each of the three task difficulties and during each of the three testing periods, we recorded and determined the values of a range of dependent variables to measure task performance and physiology as described below.

Task performance measures were the amount of total completed calculations (total completions) and the proportion of the correct answers (accuracy). We considered the participants with more correct answers as better performers. In contrast, we considered the participants with many completions but few correct answers, or with few completions but high rate of correctness, as worse performers. In addition, participants also filled out questionnaires subjectively rating task difficulty and their efforts on the task; this data was used to support the performance data.

The raw signals of the PPG for sVRI were recorded via a clinical PPG sensor and processed by the algorithms in this study to form sVRI. We used the average sVRI value in each testing period to examine significance and for comparison with other physiological measures.

The blood pressure (BP) measure includes the systolic and diastolic blood pressures. The blood pressure values are susceptible to human factors, and therefore the measurement process was controlled on human factors in the tests according to the clinical standards. The experimenter read the systolic/diastolic BP values from the BP monitor, who had received prior training in our partner clinics. During the pre/in-task period, BP was recorded three times, and during the post-task period, BP was recorded five times.

As a measure of heart rate (HR), we determined the mean R-R (R-peaks) interval (RRI) from the ECG for each testing period via the Kubios HRV v2.1 (University of Eastern Finland). ECG was recorded throughout the pre, in and post periods.

Based on the R-R intervals, three measures of heart rate variability (HRV) over a 5min period were computed, including the root mean squared successive difference (RMSSD) [3, 14], the high-frequency HRV (HF), and the mid-frequency HRV (MF). HF was the power in the high frequency range of 0.15 – 0.4 Hz of the R-R intervals by Welch's method [3] and MF was the power in the range of 0.04 – 0.15 Hz. The absolute power of the HF and MF within the two bands was submitted to a natural logarithm (ln) transform to approach normal distribution. The RMSSD, HF and MF were all computed via the Kubios HRV software (v2.1, University of Eastern Finland).

Analysis

First, the performance data was examined to confirm task difficulty and participants' effort. A repeated-measures ANOVA was used to analyze the total completions and accuracy as within-subject factors (both corresponded to the three difficulty levels). The Box-plots plots were also used to help justify the outlying performance by assuming the target performance of the participants were compliant with the normal distribution.

Second, each dependent physiological variable was analyzed using a 3*3 repeated-measures ANOVA with the testing period (three levels: pre-task, in-task and post-task) and task difficulty (three levels: easy, medium and hard) as within-subject factors. An effect of a variable varying with the testing period should be reflected by a main effect of the testing period. The effect of a variable varying with the task difficulty was analyzed through the 3 difficulty * 1 in-task period repeated-measures ANOVA because the baselines should not be affected by the task difficulties. The average values of each period were used to analyze the physiological measures.

In the ANOVA tests, a Huynh-Feldt correction of the degrees of freedom was used when the sphericity assumption was violated. Effect sizes for effects as explored by the repeated measures ANOVAs were determined by computing partial eta squared (η^2). All statistical analysis were performed with the SPSS 13.0 (formerly SPSS Inc. and now IBM, USA) for windows, and were conducted at significance level = 0.05, two-tailed.

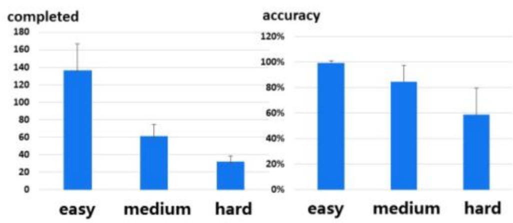
RESULTS

Performance Data

Raw Data

First, the performance data was analyzed to confirm 1) if the task difficulties were properly set, and 2) if the participants invested the expected amount of effort.

The performance data is shown in Figure 5. The mean of the total completions was maximal for the easy task (M = 136, SD = 30), intermediate for the medium task (M = 61, SD = 13) and minimal for the hard task (M = 32, SD = 6), $F(2, 78) = 511.49$, $p = 0.000$. Similarly, the mean accuracy was maximal for the easy task (M = 99.07%, SD = 2.15%), intermediate for the medium task (M = 84.78%, SD = 12.42%) and minimal for the hard task (M = 58.66%, SD = 21.07%), $F(2, 78) = 124.51$, $p = 0.000$.



(a) Total completions (b) Accuracy

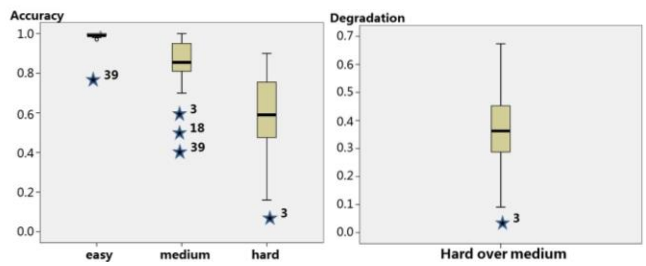
Figure 5. The performance data (means and SDs) on different task difficulties: (a) the amount of total calculations completed (b) the accuracy.

Filtering Effortless Participants

Recognizing that some participants might have been putting too little effort into the task, we utilized both subjective and objective methods to analyze which participants did not meet our expectation for effort:

1) Using questionnaires. We collected the questionnaires and analyzed participants' feedbacks on how much effort they had invested (a five-point Likert scale) on each task. We found five participants expressed that they might not have invested enough effort during the hard task. Others thought they invested enough effort during all tasks.

2) Using outlying performance analysis. In addition to analyzing the subjective questionnaires, we employed a statistical outlier-filtering method by assuming a normal distribution of the samples. We analyzed the schematic box-plots (Figure 6) with respect to the amount of total completed calculations, the accuracy and the performance degradation respectively. Looking at the number of calculations completed (not shown in Figure 6), we found that there were three outperformers, who we did not exclude from further analysis because our aim was to remove effortless participants. As shown in Figure 6(a), we determined three participants (No. 3, 18 and 39) had significantly worse performance than other participants on one or two tasks. As Figure 6(b) highlights, participant No. 3 was the only negative outlier for the performance degradation between the hard task and the medium task. No outliers found for the performance degradation between the medium task and easy task. We found that participant No.3 was also one of the five participants who, based on self-assessment using the questionnaire, had invested less effort into the tasks, while the other outliers were not. Therefore, we excluded the participant of No.3, and used the data from the remaining *39* participants for further analysis.



(a) Accuracy. (b) Performance degradation.

Figure 6. Identifying effortless performance. (a) Accuracy vs. task difficulties and (b) Performance degradation calculated as the total completions in hard task over that of medium task.

Physiological data

Effect of testing period

Table 1 shows the mean, and corresponding standard deviations, of all physiological measures during the testing periods and for different task difficulties. The measures are sVRI, systolic blood pressure (sysBP), diastolic blood pressure (diaBP), HR (RRI), RMSSD, MF and HF. A repeated-measures ANOVA of 3 periods (pre, in and post) * 3 difficulties (easy, medium and hard) was performed for each of those measures, which we used to first examine the physiological responses during different testing periods.

Measure	**Easy task**			**Medium task**			**Hard task**		
	Test period			Test period			Test period		
	Pre (baseline)	**In**	**Post**	**Pre (baseline)**	**In**	**Post**	**Pre (baseline)**	**In**	**Post**
sVRI	0.837±0.035	0.878±0.043	0.846±0.035	0.832±0.025	0.892±0.036	0.846±0.023	0.835±0.024	0.900±0.059	0.853±0.027
sysBP	103.4±10.1	116.75±12.4	105.0±10.7	104.0±10.2	118.3±12.56	110.9±30.7	103.3±10.0	118.56±16.1	104.2±10.1
diaBP	64.6±6.5	75.8±8.6	67.1±8.1	65.9±6.6	78.2±8.3	67.5±7.1	64.0±6.3	74.5±8.0	66.4±6.3
HR	830.4±112.2	752.3±111.7	829.0±108.4	839.3±119.1	737.1±107.8	825.1±109.8	849.0±122.8	748.6±164.2	847.9±121.0
RMSSD	36.5±14.2	29.9±14.1	32.8±12.8	39.4±13.8	30.7±10.5	35.4±13.2	38.3±16.1	32.4±14.0	35.3±15.4
MF (ln)	6.23±0.72	6.00±0.70	6.19±0.73	6.59±0.70	6.33±0.78	6.53±0.70	6.40±0.64	6.09±0.80	6.22±0.78
HF(ln)	6.08±0.84	5.56±1.22	5.96±0.89	6.40±0.89	5.82±1.12	6.11±1.04	6.27±1.34	5.83±1.11	6.06±1.07

Table 1. Results of physiological measures. The physiological data were recorded during the pre (baseline), in and post task periods for the easy, medium and hard task difficulty levels respectively.

Table 2 shows the results of the within-subject effects for the in-task period compared to the baseline and post-task values. We can see that each physiological measure showed a very significant effect of the testing period: the sVRI, sysBP and diaBP significantly increased, and the HRV measures (HR using RRI, RMSSD, MF and HF) significantly decreased, while participants performed the tasks in comparison with the baselines and post-tasks, ($p < 0.0001$). The results were consistent with previous studies [3, 16, 20].

Measure	Period			
	p	*F*	*df1, df2*	η^2
sVRI	0.000	121.8	1.3, 47.6	0.754
sysBP	0.000	31.02	1.4, 51.6	0.449
diaBP	0.000	180.1	1.4, 52.6	0.826
HR(RRI)	0.000	129.4	1.4, 53.6	0.773
HRV:RMSSD	0.000	23.84	1.72, 65.4	0.385
HRV:MF	0.000	9.85	2, 76	0.206
HRV:HF	0.000	19.56	1.3, 50.8	0.340

Table 2. Within-subject effects of the period by 3*3 repeated-measures ANOVA: p and F values for the main effects of period, η^2 for the effect sizes.

Effect of task difficulty

A variable's sensitivity to mental effort as induced by varying task difficulty should be shown by a main effect of task difficulty. Although this effect should primarily be determined by the in-task period, we still first examined the 3*3 ANOVA results (partially shown in Table 2). We found that the MF ($F(2, 76) = 6.387$, $p = 0.003$), sysBP ($F(1.3, 50.4) = 3.728$, $p = 0.049$) and diaBP ($F(2, 76) = 3.797$, $p = 0.027$) showed significant effects of task difficulty, which suggested that MF and diaBP may vary notably on the baselines according to their full freedom degrees (2 and 76). The sVRI ($F(2.1, 81.3) = 4.601$, $p = 0.011$) and HR ($F(3.3, 124.8) = 2.941$, $p = 0.031$) also showed significant interactive effect of testing period * difficulty with low freedom degrees, which may suggest that the main effects should be primarily reflected by the in-task values.

Next, we analyzed the in-task values of the physiological measures on their own, without the out-of-task values. Table 3 lists the results by using 3 difficulties * 1 period (in-task) repeated measures ANOVA.

Measure	Difficulty			
	p	*F*	*df1, df2*	η^2
sVRI	**0.019 ***	4.18	2, 76	0.099
sysBP	**0.002 ****	6.66	2, 76	0.149
diaBP	0.054	3.15	1.8, 68.2	0.077
HR (RRI)	**0.025***	4.41	1.5,58.4	0.104
HRV:RMSSD	0.295	1.24	2, 76	0.032
HRV:MF	**0.033***	3.56	2, 76	0.086
HRV:HF	0.139	2.03	2,76	0.051

Table 3. The effect of task difficulty on the in-task physiological values.

We can see that the sVRI, sysBP, HR and MF measures reached significance on the effect of the task difficulty, while the other physiological measures did not. These results were consistent with the findings of previous studies as well [3, 16, 20]. Additionally, in order to eliminate the influence of the baseline on the subsequent measurements, we also examined the change ratio (*cr*) for the in-task *values*, which is defined as:

$$cr = \frac{value - baseline}{baseline} \times 100\% \qquad (2)$$

From the results using change ratios, we found that sVRI (F=5.50, p=0.006), diaBP (F=35.60, p=0.000), HR (F=5.94, p=0.004) and MF (F=5.24, p=0.007) showed greater significance on the effect of task difficulty in comparison with the results obtained without using the change ratios (the results shown in Table 3). However, on the contrary, sysBP (F=3.28, p=0.043) and HF (F=0.91, p=0.409) showed less significance on the results using change ratios. These results may reflect the reliability issues of the blood pressure and HRV measurements, and from these results it can also be concluded that the sVRI was sensitive and reliable.

We also examined the trends of the mean values of the variables reaching significance on task difficulty. Figure 7 shows the trends of the means for the sVRI, sysBP (systolic BP), HR (RR interval) and MF with standard errors. The sVRI and sysBP values were taken from Table 1, and HR and MF used the change ratios to the baselines. The sVRI and sysBP increased and mid frequency decreased with the task difficulty increasing. However, the RR interval did not show the significant decrease (HR's increase) from the medium to hard task, which might suggest the reliability issue of the heart rate.

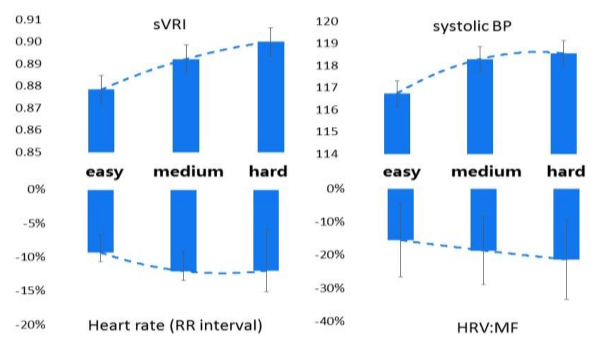


Figure 7. The mean (with standard errors) in-task values of the sVRI, systolic blood pressure, heart rate (RRI) and mid frequency as a function of task difficulty.

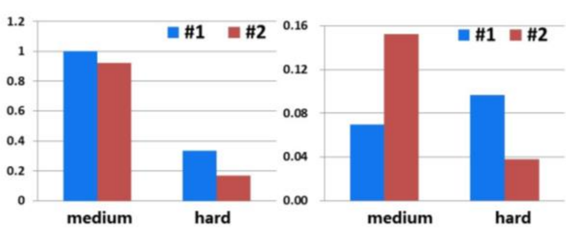
Other factors

We also examined the LF/HF (LF is the MF in this study) results for HRV. We found that the LF/HF only showed significance on the testing period (F=5.80, p=0.005), and did not show significance on the difficulty (F=1.34, p=0.269), which was similar with HF. In addition, we checked the 2min HRV using the last 2 minutes in each period for our calculation. We found a slight difference from the 5min HRV: 1) HR, MF and HF still showed significance on the period (p <= 0.001) but the RMSSD was no longer significant (p=0.541). 2) Only MF still maintained significance on the task difficulty (F=6.826, p=0.002), and HR did not (p=0.211).

We also tested for gender effects on the task performance and on the sVRI by t-tests. The gender effects were significant on the total completed calculations but not on the accuracy in the easy and medium tasks (p=0.004 and 0.015 respectively). The sVRI did not show significant gender effects for the three difficulties (p=0.584, 0.050, 0.062 respectively).

Performance and mental effort

We inspected the performance and mental effort using sVRI to understand them better. Figure 8 compares two participants on the performance and sVRI respectively. The performance was the ratio of correct answers (normalized to 1.0 by the best performance), and can be used to evaluate the outcome of the participants objectively. We can see from Figure 8(a) that participant #1 outperformed #2 about 9% in the medium task and about 100% in the hard task. We know the performance is related to individual capability, mental effort, and task complexity. With the task given, it becomes possible to evaluate differences in individual capabilities when mental effort is measurable using sVRI. Figure 8(b) shows the sVRIs in order to display and compare these participants' mental effort. We can see that participant #1 invested less mental effort (smaller sVRI) than #2 during the medium task, but achieved better performance, which implies that participant #1 might have better personal capability. Participant #1 invested increased effort on the hard task, while participant #2 seemed to give up during the hard task according to the large reduction (more than 70%) of the sVRI compared to the medium task. Based on this information, it follows logically that participant #1 still achieved much better performance (about 2x) than participant #2 on the hard task.



(a) Performance (b) sVRI (mental effort)

Figure 8. Comparing two participants (#1 and #2) on the medium and hard tasks: (a) performance and (b) mental effort displayed using sVRI.

Discussion

Hypotheses

Based on the experimental results, we have confirmed that all the three hypotheses set at the beginning of this study hold.

Comparing the physiological measures

BP and HRV need very careful control to improve their reliability, and their sensitivity may often vary between different trials, which occurred during this study as well as previous studies [3, 16, 20]. The sVRI showed promising sensitivity and reliability in this study. In addition, the sVRI had a small SD in proportion to the mean value, below 5%, while other measures were all above 10%. The sVRI can also output in real-time at the rate of heartbeat (similar with heart rate), while blood pressure needs an air-pumping process (nearly 1 min) and the HRV needs at least a 2min integral. The sVRI is measured nonintrusively via pulse, while other measures may need more body involvement, such as the BP with arm, HRV with ECG (chest), etc.

Ubiquitous usability

We employed a clinical oximeter-based PPG sensor using the transmission mode in this study for our assessment of the sensitivity, reliability and basic usability of the sVRI. PPG waveforms can also be obtained through slight-contact methods by using the reflection mode on the fingers, forehead or earlobes, such as the surface-sensing method [23], or through contactless methods, such as the image-based method using a webcam [21], which may suggest the promising usability of sVRI in ubiquitous applications.

Limitations

The sVRI does not currently distinguish cognitive load from stress despite their differences. The sVRI may be suitable only for measuring autonomic mental effort (cognitive load) and the accompanying stress, and it does not reflect the PNS activities that also correlate to cognitive process and stress since it is sympathetic-only. In addition, high heart rates (e.g. above 120 bpm) may also affect the reliability of sVRI due to more distorted PPG waveforms. Furthermore, the algorithms and experiment still have some limitations in the following two aspects. 1) There were still some odd values appearing in the current output of sVRI, which requires further improvement. 2) The easy task was actually already challenging enough for the university students. The experiment therefore focused too much on the high stressors.

CONCLUSION AND FUTURE WORK

In this paper, we propose the sVRI as a novel physiological measure for assessing cognitive load and stress which can be added to currently known measures. The experiment, based on three levels of task difficulty and three stages of testing periods, showed that the sVRI reached the same level of significance on the effects of task difficulty and period as the highest significance from the blood pressure and HRV components. Our findings also suggested the sVRI's potential as a sensitive, reliable, and usable measure. In future work, we will focus on the current limitations of this method and study its extension for ubiquitous aplications. Futher comparisons with EDA may also be involved.

ACKNOWLEDGMENTS

We thank all the volunteers for their participation in our experiment. We thank all the anonymous reviewers for their insightful comments and helpful suggestions for improving the initial draft of this manuscript. We also gratefully acknowledge the grant from NSFC (#61201357).

REFERENCES

1. Paas, F., Tuovinen, J.E., Tabbers, H., et al. Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educ. Psychol. 38*, 1 (2003), 63-71.
2. Oviatt, S. Human-centered design meets cognitive load theory: designing interfaces that help people think. In *Proc. 14th ACM Multimedia*, ACM (2006), 871-880.
3. Brouwer, A.-M., Hogervorst, M.A., Holewijn, M., et al. Evidence for effects of task difficulty but not learning on neurophysiological variables associated with effort. *Int. J. Psychophysiol. 93*, 2 (2014), 242-252.
4. Brunken, R., Plass, J.L., and Leutner, D. Direct Measurement of Cognitive Load in Multimedia Learning. *Educ. Psychol. 38*, 1 (2003), 53-61.
5. Hernandez, J., Paredes, P., Roseway, A., et al. Under pressure: sensing stress of computer users. In *Proc. CHI 2014*, ACM (2014), 51-60.
6. Sun, D., Paredes, P., and Canny, J. MouStress: detecting stress from mouse motion. In *Proc. CHI 2014*, ACM (2014), 61-70.
7. Tan, C.S.S., Schning, J., Luyten, K., et al. Investigating the effects of using biofeedback as visual stress indicator during video-mediated collaboration. In *Proc. CHI 2014*, ACM (2014), 71-80.
8. Lazarus, R.S. *Psychological stress and the coping process*. McGraw-Hill (1966), New York.
9. Hart, S.G. and Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*, North-Holland (1988), 139-183.
10. Verwey, W.B. and Veltman, H.A. Detecting short periods of elevated workload: A comparison of nine workload assessment techniques. *J. Exp. Psychol. Appl. 2*, 3 (1996), 270-285.
11. Matthews, G., Campbell, S.E., Falconer, S., et al. Fundamental dimensions of subjective state in performance settings: Task engagement, distress, and worry. *Emotion 2*, 4 (2002), 315-340.
12. Gao, Y., Bianchi-Berthouze, N., and Meng, H. What Does Touch Tell Us about Emotions in Touchscreen-Based Gameplay? *ACM Trans. Comput.-Hum. Interact. 19*, 4 (2012), 1-30.
13. Dennerlein, J., Becker, T., Johnson, P., et al. Frustrating computer users increases exposure to physical factors. In *Proceedings of the International Ergonomics Association, Seoul, Korea*(2003).
14. Kreibig, S.D. Autonomic nervous system activity in emotion: A review. *Biol. Psychol. 84*, 3 (2010), 394-421.
15. Berntson, G.G., Cacioppo, J.T., and Quigley, K.S. Autonomic determinism: The modes of autonomic control, the doctrine of autonomic space, and the laws of autonomic constraint. *Psychological. Review 98*, 4 (1991), 459-487.
16. Iani, C., Gopher, D., and Lavie, P. Effects of task difficulty and invested mental effort on peripheral vasoconstriction. *Psychophysiology 41*, 5 (2004), 789-798.
17. Mulder, G. The concept and measurement of mental effort. In *Energetics and human information processing*, Springer (1986), 175-198.
18. Brouwer, A.-M., Hogervorst, M.A., Van Erp, J.B., et al. Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of neural engineering 9*, 4 (2012), 045008.
19. Luo, L., Xiao, L., Miao, D., et al. The Relationship between Mental Stress Induced Changes in Cortisol Levels and Vascular Responses Quantified by Waveform Analysis: Investigating Stress-Dependent Indices of Vascular Changes. In *ICBEB 2012*, IEEE (2012), 929-933.
20. Hjortskov, N., Rissén, D., Blangsted, A.K., et al. The effect of mental stress on heart rate variability and blood pressure during computer work. *Eur. J. Appl. Physiol. 92*, 1-2 (2004), 84-89.
21. Bousefsaf, F., Maaoui, C., and Pruski, A. Remote detection of mental workload changes using cardiac parameters assessed with a low-cost webcam. *Comput. Biol. Med. 53*(2014), 154-163.
22. Yoo, K.-S. and Lee, W.-H. Mental stress assessment based on pulse photoplethysmography. In *Proc. ISCE 2011*, IEEE (2011), 323-326.
23. Chigira, H., Maeda, A., and Kobayashi, M. Area-based photo-plethysmographic sensing method for the surfaces of handheld devices. In *Proc. UIST 2011*, ACM (2011), 499-508.
24. Schäfer, A. and Vagedes, J. How accurate is pulse rate variability as an estimate of heart rate variability?: A review on studies comparing photoplethysmographic technology with an electrocardiogram. *Int. J. Cardiol. 166*, 1 (2013), 15-29.
25. Wijshoff, R.W.C.G.R., Veen, J., Van Der Lee, A.M., et al. PPG motion artifact handling using a self-mixing interferometric sensor. In *SPIE2011*, SPIE(2011), 78940F-78940F-78913.
26. Wijshoff, R.W.C.G.R., Mischi, M., Veen, J., et al. Reducing motion artifacts in photoplethysmograms by using relative sensor motion: phantom study. *Journal of Biomedical Optics 17*, 11 (2012), 117007-117007.
27. Chen, C.-C., Lin, S.-C., Young, M.-S., et al. Accumulated Mental Stress Study Using the Meridians of Traditional Chinese Medicine with Photoplethysmography. *The Journal of Alternative and Complementary Medicine 20*, 11 (2014), 860-867.
28. Ritter, F.E., Reifers, A.L., Klein, L.C., et al. Lessons from defining theories of stress for cognitive architectures. *Integrated models of cognitive systems*(2007), 254-262.