Non-Invasive Measurement of Cognitive Load and Stress Based on the Reflected Stress-Induced Vascular Response Index

YONGQIANG LYU and XIAO ZHANG, Tsinghua University, Beijing, China
XIAOMIN LUO, BGI (Wuhan) Translational Research Center, Wuhan, China
ZIYUE HU, Tsinghua University, Beijing, China
JINGYU ZHANG, Institute of Psychology, CAS, Beijing, China
YUANCHUN SHI, Tsinghua University, Beijing, China

17

Measuring cognitive load and stress is crucial for ubiquitous human–computer interaction applications to dynamically understand and respond to the mental status of users, such as in smart healthcare, smart driving, and robotics. Various quantitative methods have been employed for this purpose, such as physiological and behavioral methods. However, the sensitivity, reliability, and usability are not satisfactory in many of the current methods, so they are not ideal for ubiquitous applications. In this study, we employed a reflected photoplethysmogram-based stress-induced vascular response index, i.e., the reflected sVRI (sVRI-r), to non-invasively measure the cognitive load and stress. This method has high usability as well as good sensitivity and reliability compared with the previously proposed transmitted sVRI (sVRI-t). We developed the basic methodology and detailed algorithm framework to validate the sVRI-r measurements, and it was implemented by employing two light sources, i.e., infrared light and green light. Compared with the simultaneously recorded blood pressure, heart rate variation, and sVRI-t, our findings demonstrated the greater potential of the sVRI-r for use as a sensitive, reliable, and usable parameter, as well as suggesting its potential integration with ubiquitous touch interactions for dynamic cognition and stress-sensing scenarios.

CCS Concepts: • **Human-centered computing → HCI design and evaluation methods**;

Additional Key Words and Pharses: Cognitive load, mental effort, photoplethysmogram, stress, reflected stress-induced vascular response index

ACM Reference format:
Yongqiang Lyu, Xiao Zhang, Xiaomin Luo, Ziyue Hu, Jingyu Zhang, and Yuanchun Shi. 2018. Non-Invasive Measurement of Cognitive Load and Stress Based on the Reflected Stress-Induced Vascular Response Index. *ACM Trans. Appl. Percept.* 15, 3, Article 17 (July 2018), 17 pages.
http://dx.doi.org/10.1145/3185665

We gratefully acknowledge the supports from ubihealth (Grant No. FP7-PEOPLE-2012-IRSES), the Chinese National Key Research and Development Program (Grant No. 2017YFB0403404), Tsinghua University Initiative Scientific Research Program, the Research Fund from Beijing Innovation Center for Future Chip (Grant No. KYJJ2016005), and the China Scholarship Council.
Author's addresses: Y. Lyu, X. Zhang, Z. Hu, and Y. Shi, 1-407, FIT Building, Tsinghua University, Beijing 100084, China; email: luyq@tsinghua.edu.cn; X. Luo, BGI (Wuhan) Translational Research Center, Wuhan 430074, China; email: luo.xiaomin@139.com; J. Zhang, 16 Lincui Road, Chaoyang District, Beijing 100101, China; email: zhangjingyu@psych.ac.cn.

http://dx.doi.org/10.1145/3185665

1 INTRODUCTION

Cognitive load refers to the load imposed on the human cognitive system by performing a particular task [1]. Cognitive load can induce mental stress, which often accompanies cognitive processes when subjects exert mental effort, e.g., executing tasks. This stress may give lead to certain negative emotions, such as anxiety, fatigue, and anger, which have generally negative effects on human health [2]. Thus, cognitive load and stress measurements are also important for smart health-care applications [3].

Sensing the cognitive load and stress in users is very important for ubiquitous human–computer interaction (HCI) applications to dynamically and sufficiently respond to user activities, e.g., mobile computing [4], smart driving [5], and smart context awareness and assistance [6]. These measurements can help to identify detailed mental changes by subjects during a task, thereby supporting precise analysis, smart reaction, and improvements to the design of components [7–9].

However, a large gap exists between the requirements of usable ubiquitous applications and the capabilities of current measures. For example, it is difficult to use subjective questionnaires [10–12] in a continuous manner without interfering with the user experience, although they are useful. Objective measures such as behavioral methods (such as behavior measurement [13, 14] and dual-task measurement [8]) and physiological measures [15] (such as heart rate variation (HRV), blood pressure (BP), and electrodermal activity (EDA)) still fail to establish a good balance among sensitivity, reliability, and usability, especially for ubiquitous applications and wearable devices [16–18], although these methods have specific advantages in some areas.

In this study, we developed a sensitive, reliable, and non-invasive measure for assessing cognitive load and stress. In particular, we developed a reflection-mode photoplethysmogram (PPG)-based stress-induced vascular response index, i.e., reflected sVRI (sVRI-r), for measuring the peripheral vasoconstriction tone in a slight-touch mode, which greatly extends the usability of the transmitted sVRI (sVRI-t) method proposed previously [28]. We conducted two experiments to verify the sensitivity, reliability, and usability of this new method. In experiment 1, we employed the classic cognitive tasks based on arithmetic calculations to show the current gold-standard sensitivity and reliability of sVRI-r compared with sVRI-t. In experiment 2, we employed the same arithmetic calculations used in experiment 1 to verify the sensitivity and reliability of sVRI-r and to show its advantages in terms of usability.

The main contributions of this study are as follows.

- We developed sVRI-r as a usable physiological measure to assess cognitive load and mental stress based on a slight touch of the user's finger on the sensor, and we verified that it could distinguish different levels of effort associated with corresponding task difficulties.
- We implemented sVRI-r with infrared and green light sources to demonstrate their differences in performances.
- Compared with sVRI-t, sVRI-r is non-invasive and it is promising for measuring cognitive load and stress dynamically during pervasive touch interactions.

Compared with the previously proposed sVRI-t method [28], we demonstrated the superior sensitivity and reliability of sVRI-r, which efficiently exploits the potential of sVRI in ubiquitous scenarios, such as touch interactions and wearable sensors. In the reflection mode and transmission mode [28], we used the same experimental settings and analysis methods to ensure a fair comparison. The remainder of this article is organized as follows. First, we discuss related research. We then provide the main technical details regarding sVRI, including those for sVRI-t and sVRI-r. In the following two sections, we present the results of a classic cognitive experiment, which we employed to compare the sensitivity, reliability, and usability of sVRI-t (experiment 1) and sVRI-r (experiment 2), as well as other known physiological measures, including BP and HRV. Finally, we describe the experimental results, discuss the limitations of the study, and give our conclusions.

2 RELATED WORK

In contrast to subjective measures such as *work/task load* questionnaires [19, 20] and stress-state questionnaires [21], physiological measures record objective human vital signs and they can support dynamic and reliable analyses. Physiological measures are usually linked to the human nervous system. For example, some are associated with the autonomic nervous system (ANS), which controls the functions of most human organs [7, 22], such as electrocardiography (ECG), HRV, pulse rate variation (PRV), EDA, respiration, BP, pupil dilation, and eye blinks. Additional measures are associated with other nervous systems, such as electroencephalography (EEG), which is linked to the central nervous system, and electromyography, which is linked to both the somatic nervous system and ANS [23]. Some medical methods are also employed in cognition-related research, such as neuroimaging techniques like positron-emission tomography (PET) and functional magnetic resonance imaging (fMRI) [8], and biomarker techniques based on cortisol or adrenaline examination in saliva or blood samples [24].

Before considering the advantages and disadvantages of these physiological measures, the nervous systems that activate and control them are first introduced. The ANS comprises the sympathetic nervous system (SNS) and parasympathetic nervous system (PNS). The SNS is responsible for emotional arousal and physiological adjustments that support performance. SNS increases its activities during tasks that require active coping, whereas PNS inhibits SNS activities and increases its functions during tasks that require passive coping [25]. Some vital signs are controlled by both the SNS and PNS (dually innervated) [23, 25], such as ECG, HRV, PRV, EDA, respiration, BP, pupil dilation, and eye blinks, whereas others are controlled only by the SNS [23, 25], such as EDA and peripheral vasoconstriction. The dually innervated vital signs actually reflect the outputs of both nervous systems. Therefore, the different modes of nerve regulation by the two systems may produce the same overt (observed) vital-sign results [26], which may not be informative with respect to the underlying mechanism [25], and thus the sensitivity and reliability could be lacking when measuring cognitive load and stress. By contrast, SNS-only vital signs are regarded as direct and sensitive indices of autonomic effort, as well as resulting stress, because SNS activation increases as the central processing complexity of a task increases [7, 25, 26], thereby indicating their potential for yielding sensitive and reliable measurements. Despite their capacity for directly and accurately reflecting cognitive activities, EEG, PET, and fMRI, as well as clinical biomarker techniques are difficult to use in ubiquitous scenarios due to the need for invasive attachments, high costs, and the loss of usability during long-term monitoring.

The first category of SNS-only measures is associated with skin conductivity. Sweat glands (eccrine) are controlled only by the SNS, where it affects the skin conductivity, which is often measured by EDA. The effectiveness of EDA for measuring the cognitive load or mental stress has been demonstrated in previous studies [7, 27]. However, EDA may exhibit notable individual differences and it is susceptible to the influence of skin conditions, as well as the ambient temperature and humidity.

The second category of SNS-only measures is associated with peripheral vasoconstriction, which is often measured using plethysmography techniques. Iani et al. [25] verified that the peripheral arterial tone measured by a pressure-based plethysmograph can reflect the changes in mental effort. Luo et al. [15] reported a strong relationship between the PPG amplitudes in participants and the cortisol level, which is clinically regarded as a major biomarker of mental stress. Studies have also utilized video plethysmography to approximately assess the cognitive load and stress using cameras [16] with imaging techniques. Compared with pressure-based plethysmograph measurements, PPG measurements are more sensitive and reliable, and they have greater potential usability due to the flexibility for contact or contactless measurement. At present, two main PPG-based analysis methods are employed to measure the cognitive load or stress: time-domain characteristic analysis and frequency-domain characteristic analysis. Time-domain characteristic analysis includes methods such as amplitude analysis [25], waveform and pulse rhythm analysis [16, 18], and defined feature-based analysis [29]. However, these methods all depend on the PPG amplitudes, which may exhibit notable temporal changes and individual differences, thereby causing severe reliability issues.

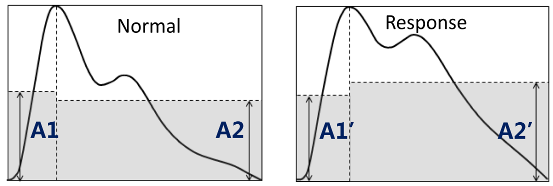


Fig. 1. Quantitative response to mental effort reflected by the PPG waveform patterns in terms of the average amplitudes of A1 and A2. A1 and A2 vary (to A1' and A2', respectively) as the vascular responses.

(a) Transmitted PPG sensor (b) Reflected PPG sensor

Fig. 2. PPG sensors in the transmission mode (a) and reflection mode (b). The transmission mode requires a light-emitting diode (LED) opposite the photodetector (PD), whereas the reflection mode requires that the LED and PD are both on the same side.

3 sVRI-r

In this study, we developed sVRI-r to non-invasively assess the cognitive load (through mental effort) and the resulting mental stress. sVRI-r shares the same index definition as sVRI [28] where it measures the relative ratio of two contours in the PPG waveform, which may be more sensitive and the individual differences are limited. As shown in Figure 1 [28], sVRI is motivated by the effects of mental effort and stress reflected in the PPG appearance. Compared with a normal PPG waveform when the participant is at rest (without mental effort), the average amplitudes of the two contours, as shown in A1 and A2 in Figure 1, can have significant differences when a mental task is imposed. An obvious A2-amplitude elevation exists in the response mode compared with the normal mode.

A2, A1, and the maximum amplitude of the waveform can exhibit severe temporal changes and individual differences, but the sVRI value defined in Equation (1) may be relatively stable with steadier morphological waveforms, which is the key feature that we verified in this study:

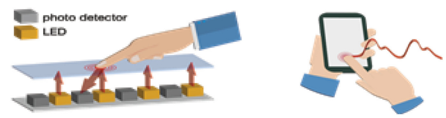
$$\mathrm{sVRI} = \frac{A2}{A1}. \tag{1}$$

A1 and A2 are calculated based on the area integrals of the two contours of the waveform, where the sVRI should increase with the mental effort or stress, as shown in Figure 1.

3.1 Measurement Modes

sVRI can be measured in two PPG sensing modes, i.e., with transmitted PPG sensors (Figure 2(a)) and reflected PPG sensors (Figure 2(b)), which are used to measure the sVRI-t and sVRI-r, respectively. sVRI-t employs a light-transmitting PPG sensor, where light needs to be transmitted through a body part (Figure 2(a)), e.g., by using a finger-clip sensor or placing a finger into the shallow groove of a sensor.

Due to the inconvenient requirements for the measurement apparatus, sVRI-t has disadvantages in ubiquitous HCI applications, such as mobile phones and wearable sensors. According to a previous study of reflected PPG sensors based on processing the signal from reflected light (Figure 2(b)), the usability of sVRI may be improved in HCI touch modalities via a similar sensing process, as depicted in Figure 3.



(a) sVRI-r for touch modality (b) A mobile scenario

Fig. 3. Potential applications of the sVRI-r in the touch modality (a) and an application scenario for mobile phones (b).

The transmitted PPG has a stable signal, and the current PPG sensing techniques and algorithm flow are sufficient to produce adequate sVRIs. However, compared with the transmitted PPG, the reflected PPG signal may become weaker and its waveform appearance may vary during continuous monitoring, thereby requiring a practical solution to maintain the signal quality and algorithm flow.

3.2 Signal Enhancement Techniques for sVRI-r

sVRI-r aims at obtaining more usable measurements for ubiquitous HCI applications, such as applications involving touch modalities. We use the same definition of the instant sVRI value employed in Equation (1), and we propose additional signal enhancement techniques and filtering techniques to improve the signal quality for sVRI.

The light-emitting diode (LED) in a PPG sensor transmits light to the skin, bone, and vascular blood. The photodetector (PD) collects the reflected light from these tissues after some of the light is scattered or absorbed. PPG waveforms are generated based on the changes in the pulsatile part (blood and vasoconstriction), which reflects less of the light to the PDs. Therefore, in the reflection mode, the PDs receive a much lower power density and the signal output is weaker compared with the transmission mode.

The PPG signal comprises two parts. The first is time-variant and it is regarded as the alternating current (AC) component. The second is basically static and it is regarded as the direct current (DC) component. The AC component is formed by the light reflecting from the pulsatile parts of the body, such as blood and vasoconstriction [30], whereas the DC component is formed by the light reflecting from other parts of the body, such as skin, bone, and other tissues. In addition, environmental light contributes to the final signal and degrades the signal-to-noise ratio. The PPG signal can be formulated as Equation (2):

$$S_f = S_{ac} + S_{dc} + S_{env}, \tag{2}$$

where the final signal S_f comprises the AC component S_{ac}, DC component S_{dc}, and environmental component S_{env}. S_{ac} is the signal that we aim to improve, and S_{dc} and S_{env} are those that we intend to inhibit. To enhance the signal, we employed two techniques in this study.

The first technique aims to enhance the light source from the LED. Assuming that the environmental light is stable, the signal-to-noise ratio for S_{ac} should improve if we increase the luminous intensity of the LED. The enhanced light source also increases the DC component, but this technique is still very important for improving the signal in an open environment, especially the signal-to-environmental-noise ratio. In this study, we employed multiple diodes as light sources to improve the signal. However, this method may cause power consumption challenges in practice. In this case, a high-frequency switch for the power supply might be a possible solution, which could also assist with improving the signal.

We also used green light (wavelength = 500 to 600nm), because it exhibits sensitive changes due to the superficial flow of blood in skin [31], which is beneficial for improving the S_{ac} signal compared with S_{dc}. We employed green-light LEDs with a wavelength of 570nm and compared the results obtained using infrared light LEDs to assess the potential suitability of green light.

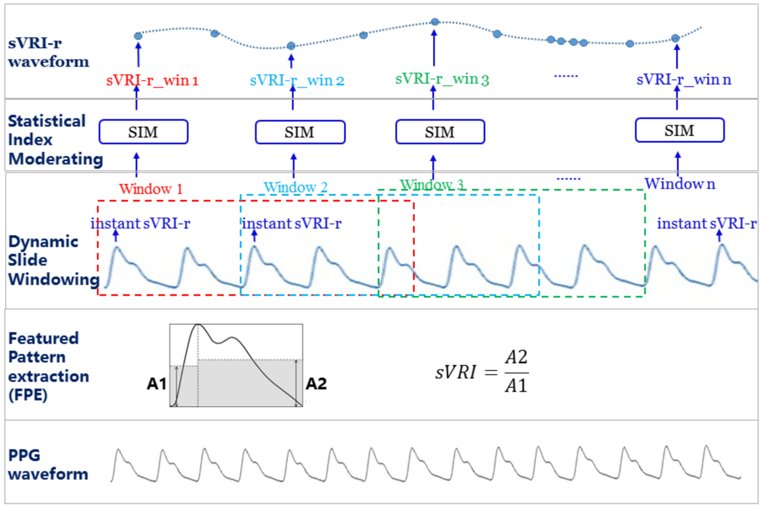


Fig. 4. Framework of the algorithm. Bottom to top: the process begins with the original PPG waveforms, before applying featured pattern extraction, dynamic windowing, and statistical index moderating to obtain the sVRI-r waveform.

3.3 Algorithm Framework for Reliable and Real-Time sVRI-r

The overall flow of the algorithm for calculating the sVRI-r is shown in Figure 4. The process starts with the capture of the original PPG waveforms, which is followed by three procedures comprising featured pattern extraction (FPE), dynamic sliding windowing (DSW), and statistical index moderating (SIM), which produce continuous sVRI output waveforms.

Using the traditional PPG signal filtering methods [32, 33], the FPE can be fed with the PPG waveforms. It should be noted that the PPG waveforms entered into this algorithm have basic requirements in terms of the signal quality, although how to improve the reliability of the sVRI-r outputs needs to be addressed. If most of the input PPG waveforms are noisy or distorted, no reliable sVRI-r can be expected. In highly ubiquitous scenarios such as wearable and natural interactions, enhanced signal quality guaranteeing techniques should be considered before applying the algorithm.

In FPE, the average amplitudes of the two contours, A1 and A2 (see Figure 1), are calculated based on the area integral, and the instant value of the sVRI-r is calculated using Equation (1). To output reliable and real-time sVRI-r waveforms, we employ a 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-r value. Digital filtering the original PPG (before the algorithm) can mitigate motion artifacts, but some distorted waveforms may still exist. Therefore, basic digital filtering manipulates the data to select the qualified waveforms for processing, which is an empirical parameter-based waveform selection scheme. SIM is then invoked to process the window to output a reliable sVRI-r value corresponding to the window. A new window follows the window with a certain backward duration. For instance, as depicted in Figure 4, window 2 (in blue) slides backwards with two pulse beats after window 1 (in red). Sometimes, the window might slide with a larger step to lower the computational requirements on ubiquitous devices such as watches. In addition, the window size may differ according to the varying proportions of the qualified original PPG waveforms to maintain the quality of the SIM process. The window sizing strategy is very important, particularly in ubiquitous scenarios with low quality input signals.

SIM processes the sVRI-r windows fed by DSW and calculates the window sVRI-r for each window, where it is assumed that a normal distribution holds in each window. The outliers are filtered by an outlying test, i.e., filtering those with very large distances from the expected value (the distance is measured by multiples of the standard deviation). The expected value of the instant sVRI-r after filtering is considered to be the final window sVRI-r, and those window values comprises the final output waveform. It should be noted that this statistical process is actually pessimistic moderating and there may be a delay before indicating rapid mental changes caused by abrupt situations, e.g., a sudden frightening event.

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

4 EXPERIMENT 1: TRANSMISSION MODE

In experiment 1, we aimed to test the performance of sVRI-t, which we mainly described in our previous study [28]. We present this experiment together with the reflection-mode experiment (experiment 2) to provide an overall view of the sVRI-related methods and to compare the two modes. We employed a classic cognitive task using arithmetic calculations to verify the sensitivity and reliability of sVRI-t compared with simultaneous recording of the BP, heart rate, and HRV components.

In this experiment, we treated mental effort as the actual cognitive load and it was ranked in arithmetic tasks with three testing periods and three levels of difficulty [28]. We tested the following specific hypotheses.

Hypothesis A (period effect): sVRI-t should increase as the task proceeds.

Hypothesis B (difficulty effect): sVRI-t should increase as the mental effort increases, which is induced by greater task difficulty.

Hypothesis C (reference validation): Some other parameters should also exhibit significant differences between test periods and task difficulty levels, such as the digital pulse amplitude (PA) and reflection index (RI) of PPG, and the HRV by ECG.

4.1 Methods

The arithmetic calculation task used in this study involved subtracting 6/7/13/77 from certain numeric digits [34], e.g., subtracting seven from a two-digit integer. This type of task has low learning effects [7] and it is considered appropriate for evaluating the pure cognitive load and accompanying stress. The performance data and the physiological data were recorded simultaneously for the participants while they performed the arithmetic tasks. Questionnaires were also collected from the participants about their experiences. Finally, the hypotheses were evaluated based on the analysis of the data.

Participants. Forty university students with similar educational backgrounds and arithmetic capabilities aged between 19 and 25 years (22 $\pm$ 1.7 years) who comprised 22 males and 18 females participated in the experiment to reduce the effects on performance caused by variations in individual capabilities and backgrounds [7]. All of the participants were healthy and were not using any medications. All were right-handed and had normal or corrected-to-normal vision. The experiment was approved by the institutional ethics committee and performed in accordance with the ethical standards established in the Declaration of Helsinki. Each participant was paid about 40 U.S. Dollars equivalent in domestic currency.

Apparatus. The sVRI-t (from PPG), ECG (for HRV), and BP were measured for the participants in the experiments. The sVRI-t was measured using a standard clinical finger-clip PPG sensor at a sampling rate of 500Hz and it was processed with the simultaneous sVRI-t algorithms. The ECG was measured by a SMOTE PSG device (Compumedics Ltd, Australia) at a sampling rate of 200Hz. BP was measured by an OMRON HEM-7112 BP monitor. The arithmetic problems were presented on a 19-inch LCD monitor and the participants used a 101-key standard keyboard to input their answers, before pressing the Enter key to submit.

Setting. The experiment was conducted in a sound-attenuated, temperature-controlled, and electrically-shielded room. The room temperature and humidity were kept constant during the experiment. For each participant, the PPG sensors were placed on fingers on the left hand and the cuff of the BP monitor was wrapped around the left upper arm and kept at the same height as the heart of the participant. Self-adhesive 1.5-inch electrodes

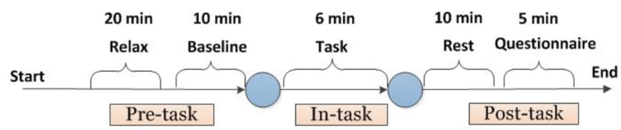


Fig. 5. Experimental protocol.

Fig. 6. Participant in the three test periods of a task.

were attached to each participant's chest with 7% chloride wet gel in a standard configuration of leads for ECG. After the experiment started, the participant was asked to try to avoid movements. The PPG signal during the air inflation/deflation process by the BP monitor was ignored to ensure signal reliability.

Stimuli. The stimulus used to induce cognitive load and stress was an arithmetic task, where certain numeric digits were subtracted from 6/7/13/77 [34]. The task lasted 6min, and we counted the number of completed calculations, as well as the number of correct answers, where the tests were presented to the participants instantly. We set three task difficulty levels, i.e., easy, medium, and hard, and we provided a simple and clear graphic user interface (GUI) to perform the calculations. Participants were allowed a maximum of 10s to complete each calculation for the easy or medium levels and 15s for the hard level. The answers given after this limit were counted as incorrect and the next question appeared. At the easy level, the subtrahends were 6 or 7 (at random) and the minuends were random three-digit numbers. At the medium level, the subtrahends were 13 or 77 (at random), and the minuends were random four-digit numbers. The hard level was most challenging, where it required the participants to subtract random four-digit numbers from random five-digit numbers.

Protocol. Each participant performed three tests (one for each task difficulty) individually within a two-week period. They were randomly assigned a task difficulty in each test without being told the level of difficulty in advance. Each test included three main task periods, i.e., pre-task, in-task, and post-task periods, which lasted for a total of approximately 50min, as shown in Figure 5. At the beginning of the test, the participants were provided general information and instructions regarding the experiment, and they were informed about the types of data recorded during the test. The participants were asked to treat the experiment seriously.

The pre-task period lasted 30min with a 20min rest period (including preparation time) and a 10min pre-task (baseline) period. In this period, the participants were left alone in the room for about 10min after the recording devices were attached, and they were verbally instructed to rest quietly and to move as little as possible. In the following 10min pre-task period, physiological data were recorded as the baseline. After the pre-task finished, the participants were then instructed to complete the arithmetic calculations in 6min, which was the in-task (working) period. When the time was complete, the participants were instructed to stop and rest for 10min, which was the post-task (recovery) period. In the last 5min, the participants were asked to fill out questionnaires to assess their experience regarding the tasks and the effort invested during the test. The experimental scenarios for the participants with attached testing apparatus are shown in Figure 6.

Dependent variables. For each of the three task difficulty levels and each of the three testing periods, we recorded and determined the values of a range of dependent variables to measure the task performance and physiology for each participant, as described in the following.

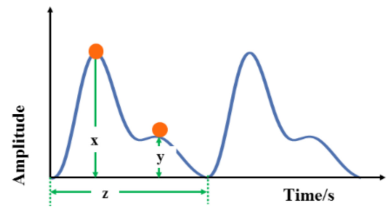


Fig. 7. Normal PPG. PA = x, RI = x/y, HR = 60/z.

The task performance was measured according to the total number of completed calculations (total completions) and the proportion of correct answers (accuracy). More total completions with better accuracy was considered better performance, whereas many completions but few correct answers or few completions but a high rate of correctness was considered bad performance. The participants were also asked to complete questionnaires to subjectively rate the task difficulty and their effort in the task, which were then used to help verify the performance measurements.

The BP is susceptible to human factors, and thus its measurement process was controlled based on human factors in the tests according to clinical standards. The experimenter had received prior training in our partner clinics, and they controlled the test process and read the systolic/diastolic BP values from the BP monitor. BP was recorded three times during each testing period and the average values for each period were used in the analysis.

We analyzed several typical HRV components in this study, i.e., heart rate (HR as the mean R–R interval from ECG, also denoted as RRI), root mean squared successive difference (RMSSD), high-frequency HRV (HF; the power in the high frequency range of 0.15–0.4Hz using Welch's method [7]), and low-frequency HRV (LF; the power in the range of 0.04–0.15Hz). We used 10min data to calculate the pre-task period HRV results and 6min data to calculate the in-task period HRV results.

sVRI-t was obtained using the PPG signals from a transmitted PPG sensor. We used the average sVRI-t value in each testing period to assess significance and for comparison with other physiological measures. The air-inflation and deflation process by the BP monitor severely affected vasoconstriction and blood flow, and thus the PPG signals obtained during these processes were ignored to guarantee the reliability of the data. We also compared three popular time-domain parameters of PPG, as shown in Figure 7, i.e., PA, RI, and HR.

Analysis. First, the performance data were analyzed to confirm the task difficulty and the effort made by the participants by repeated-measures ANOVAs with the total completions and accuracy as the within-subject factors. Box-plots were also employed to help justify the outlying performance, where we assumed that the target performance of the participants was compliant with a normal distribution.

Second, each dependent physiological variable mentioned above was analyzed using a two test-period * one task-difficulty repeated-measures ANOVA with the testing period (two levels: pre-task and in-task) as the within-subject factor. The effect of a variable that varied with the task difficulty was analyzed using three-task difficulty * one in-task period repeated-measures ANOVA, because the baselines were not expected to be affected by the task difficulty. The mean values for each period were used to analyze the physiological measures in this process.

In the ANOVA tests, Huynh–Feldt correction of the degrees of freedom was used when the sphericity assumption was violated. Sizes of effects explored by the repeated measures ANOVAs were determined by computing the partial eta squared ($\eta 2$). All of the statistical analyses were performed with SPSS 13.0 (formerly SPSS Inc., and now IBM, USA) for Windows and they were conducted at a significance level of $p = 0.05$ (two-tailed).

In addition, 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

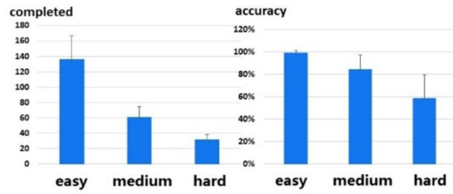


Fig. 8. Performance data (mean and SD) with different task difficulties.

$$cr = \frac{value - baseline}{baseline} \times 100\% \tag{3}$$

where *baseline* refers to the mean value in the pre-task period.

4.2 Results and Discussion

Performance Data. As shown in Figure 8 [28], the mean of the total completions was maximized for the easy task (mean = 136, SD = 30), intermediate for the medium task (mean = 61, SD = 13), and minimized for the hard task (mean = 32, SD = 6), $F(2, 78) = 511.49$, $p = 0.000$). Similarly, the mean accuracy was maximized for the easy task (mean = 99.07%, SD = 2.15%), intermediate for the medium task (mean = 84.78%, SD = 12.42%), and minimized for the hard task (mean = 58.66%, SD = 21.07%), $F(2, 78) = 124.51$, $p = 0.000$), which verified the correct task difficulty settings and normal effort of the participants.

Filtering Low-Effort Participants. Some participants may have made little effort in the task, which contradicted the experimental assumption regarding the expected mental effort investment, and thus they needed to be excluded to avoid interference with the results. We utilized subjective and objective methods to filter low-effort participants, as follows.

(1) Questionnaires. We collected the questionnaires and analyzed feedback from participants about how much effort they had invested (on a five-point Likert scale) on each task. We found that five participants indicated that they might not have invested sufficient effort during the hard task. Others thought that they invested sufficient effort during all of the tasks.
(2) Outlying performance analysis. To confirm the feedback in the questionnaires, we employed a statistical outlier-filtering method but also where we assumed a normal distribution for the samples. We analyzed the schematic box-plots with respect to the amount of total completed calculations, accuracy, and the degradation in performance. We provide more details of this technique for experiment 2.

Period effect. Table 1 shows the mean and corresponding standard deviations for all the physiological measures during the testing periods and for different task difficulties, where these measures comprise the sVRI-t, systolic BP (sysBP), diastolic BP (diaBP), HR (RRI), RMSSD, LF, HF, LF over HF (LF/HF), PA, RI, and HR (PPG). Repeated-measures ANOVA was performed for two periods (rest and task) * one difficulty (medium) for each of those measures, which we used to examine the physiological responses during different testing periods.

Table 2 shows the within-subject effects for the in-task period compared with the pre-task (baseline) period according to the 2*1 repeated-measures ANOVA. We can see that most of the physiological measures were affected significantly by the testing period, where sVRI-t, sysBP, diaBP, LF, LF/HF RI, and HR_PPG increased significantly, whereas HR_HRV, RMSSD HF, and PA decreased significantly while the participants performed their tasks compared with the rest conditions ($p \leq 0.001$). These results are consistent with those obtained in previous studies [7, 25, 35].

Effects of difficulty. A variable's sensitivity to mental effort induced by tasks with various difficulty levels should be represented as a main effect of the task difficulty. We analyzed the in-task values for the physiological measures without the out-of-task values. Table 3 shows the results according to three difficulties * one period

Table 1. Results for Physiological Measures

	Easy task			**Medium task**			**Hard task**		
	Test period			Test period			Test period		
Measure	**Pre(base)**	**In**	**Post**	**Pre(base)**	**In**	**Post**	**Pre(base)**	**In**	**Post**
sVRI-t	0.84 ± 0.04	0.88 ± 0.04	0.85 ± 0.04	0.83 ± 0.03	0.89 ± 0.04	0.85 ± 0.02	0.84 ± 0.02	0.90 ± 0.06	0.85 ± 0.03
sysBP	103 ± 10	117 ± 12	105 ± 11	104 ± 10	118 ± 13	111 ± 31	103 ± 10	118 ± 16	104 ± 10
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_HRV	830 ± 112	752 ± 111	829 ± 108	839 ± 119	737 ± 107	825 ± 110	849 ± 123	749 ± 164	848 ± 121
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
LF(n.u.)	45.7 ± 11.8	54.6 ± 13.2	47.6 ± 15.3	53.4 ± 9.05	61.8 ± 14.0	53.6 ± 15.5	50.1 ± 10.0	62.8 ± 15.1	47.9 ± 8.59
HF(n.u.)	53.9 ± 11.7	45.0 ± 13.2	52.2 ± 15.4	46.4 ± 8.99	37.9 ± 14.0	46.0 ± 15.4	49.7 ± 10.0	37.0 ± 15.1	51.3 ± 7.92
LF/HF	0.93 ± 0.44	1.38 ± 0.64	1.09 ± 0.70	1.22 ± 0.40	2.05 ± 1.32	1.38 ± 0.70	1.07 ± 0.38	2.23 ± 1.53	0.97 ± 0.31
PA	790 ± 321	581 ± 287	815 ± 287	960 ± 480	728 ± 421	1045 ± 453	779 ± 420	680 ± 699	913 ± 417
RI	0.60 ± 0.07	0.63 ± 0.09	0.61 ± 0.08	0.61 ± 0.07	0.62 ± 0.08	0.60 ± 0.07	0.68 ± 0.08	0.68 ± 0.09	0.68 ± 0.07
HR_PPG	77.1 ± 8.3	81.8 ± 9.4	77.9 ± 9.0	79.4 ± 10.1	82.7 ± 10.6	79.6 ± 10.6	78.2 ± 8.7	82.3 ± 10.1	76.3 ± 8.4

The physiological data were recorded during the pre- (baseline), in-, and post-task periods for the easy, medium, and hard task difficulty levels.

Table 2. Within-subject Effects of the Period According to 2*1 Repeated-measures ANOVA

	Period		
Measure	p	F	η^2
sVRI-t	**0.000***	139.313	0.786
sysBP	**0.001***	12.814	0.252
diaBP	**0.000***	195.897	0.838
HR_HRV	**0.000***	132.593	0.777
RMSSD	**0.000***	31.390	0.452
LF	**0.006***	8.451	0.182
HF	**0.003***	10.100	0.210
LF/HF	0.225	1.524	0.039
PA	**0.000***	54.763	0.646
RI	**0.026***	5.492	0.155
HR_PPG	**0.002***	11.337	0.274

p and F values are shown for the main effects of period and $\eta 2$ for the effect sizes.

(in-task) repeated-measures ANOVA, where the sVRI-t, sysBP, HR, LF, HF, and LF/HF measures were significant for the effect of the task difficulty, whereas the other physiological measures were not. These results are consistent with the findings of previous studies [7, 25, 35].

In addition, to eliminate the influence of the baseline on the subsequent measurements, we examined the change ratio (cr) for the in-task values. According to the results obtained 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 LF (F = 7.72, p = 0.012) were affected more significantly by the task difficulty compared with the results obtained without using the change ratios (results shown in Table 3). By contrast, sysBP (F = 3.28, p = 0.043), HF (F = 0.71, p = 0.498), and LF/HF (F = 1.81, p = 0.189) were not affected significantly according to the results obtained using change ratios. These

Table 3. Effect of Task Difficulty on the In-task Physiological Values

	Difficulty			
Measure	***p***	***F***	***df1, df2***	η^2
sVRI-t	**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:LF	**0.029***	4.72	1.4, 53.6	0.199
HRV:HF	**0.031***	4.61	1.8, 52.4	0.195
LF/HF	**0.032***	4.31	1.8, 56.7	0.185
PA	0.401	0.848	1.4, 42.8	0.027
RI	**0.000***	11.033	1.7, 52.1	0.269
HR_PPG	0.860	0.141	1.9, 57.4	0.005

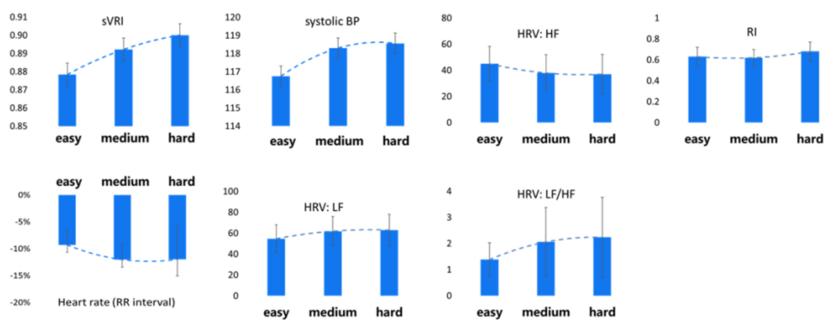


Fig. 9. Mean (with standard errors) in-task values for the sVRI-t, systolic BP, heart rate (RRI) HRVs, and RI as functions of the task difficulty.

results may reflect reliability issues with the BP and HRV measurements. These results also demonstrate that the sVRI was sensitive and reliable.

When we examined the trends in the mean values of the variables that were affected significantly by the task difficulty, as shown in Figure 9, we found that sVRI, sysBP, LF, and LF/HF increased, whereas HF decreased as the task difficulty increased. However, the RRI did not decrease significantly (HR increased) from the medium to hard task, which might suggest a reliability issue with the HR. RI different significantly with the task difficulty, but the effect was not monotonic, i.e., RI decreased from the easy to medium task, whereas it increased from the medium to hard task.

Verification of Hypotheses. Based on the experimental results, we confirmed Hypotheses A, B, and C. sVRI-t increased with the task period from rest to work, and with the task difficulty from easy to hard. Most of the other reference parameters obtained from PPG and ECG, such as PA, RI, and HRV, differed significantly between periods and difficulty levels, but only a few of them exhibited monotonic effects with the task difficulty, such as LF and HF.

5 EXPERIMENT 2: REFLECTION MODE

In experiment 2, we employed the same arithmetic calculation task as experiment 1 to induce the cognitive load and mental stress to verify the sensitivity and reliability of sVRI-r. sVRI-r was measured at three different

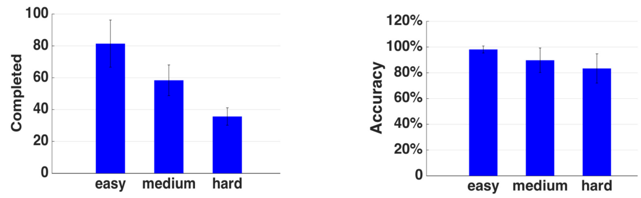


Fig. 10. Performance data (mean and SD) with different task difficulties.

locations on the left hand by three separate sensors. The first was on the ring finger for sVRI-t using a transmitted infrared PPG sensor, the second was on the middle finger for sVRI-r using a prototype reflected infrared sensor, and the third was on the thumb for sVRI-r using a prototype reflected green-light sensor (denoted as sVRI-rg).

In this experiment, we employed the sVRI-r to assess the cognitive load and stress. In the experimental verification, we treated mental effort as the actual cognitive load and ranked it in arithmetic tasks with three testing periods and three levels of difficulty. We tested the following specific hypotheses.

Hypothesis A (period effect): sVRI-r should increase as the task proceeds.

Hypothesis B (difficulty effect): sVRI-r should increase as the mental effort increases, which is induced by greater task difficulty.

Hypothesis C (reference validation): The simultaneous sVRI-t and sVRI-r values should agree with each other.

5.1 Methods

In total, 44 subjects participated in the experiment. The participants were aged between 18 and 32 years (22.6 ± 3.3 years), and they comprised 25 males and 19 females. The experimental procedure was identical to the previous experiment except for the apparatus employed. All of the PPGs were measured at a sampling rate of 500 Hz and they were processed using the simultaneous sVRI-r algorithms. The PPG signals for sVRI-t and sVRI-r were recorded via transmitted PPG sensors and reflected PPG sensors, respectively. We used the average value in each testing period to determine significance and for comparison with other physiological measures. The other variables were the same as those in experiment 1.

5.2 Results and Discussion

Performance Data. Similar to experiment 1, we analyzed the performance data to confirm that the task difficulty levels were appropriate and the effort of the participants. As shown in Figure 10, the mean of the total completions was maximized for the easy task (mean = 81, SD = 15), intermediate for the medium task (mean = 58, SD = 10), and minimized for the hard task (mean = 36, SD = 5), $F(2, 86) = 422.04$, $p < 0.001$). Similarly, the mean accuracy was maximized for the easy task (mean = 98.13%, SD = 2.71%), intermediate for the medium task (mean = 89.74%, SD = 9.51%), and minimized for the hard task (mean = 83.37%, SD = 11.38%), $F(2, 86) = 43.27$, $p < 0.001$). Compared with experiment 1, this cohort performed better in the hard tasks.

Filtering Low-Effort Participants. In this experiment, we used the same method as experiment 1 to confirm participants who obtained failures when investing adequate effort according to their questionnaires. After analyzing the box-plots, as shown in Figure 11, we found several outliers. However, only Nos. 15, 29, and 37 both reported failures in questionnaires and were marked as outliers in the box-plots. In addition, we found that some physiological data were missing for Nos. 32 and 39. Therefore, we excluded participant Nos. 15, 29, 32, 37, and 39, and used the data obtained from the remaining 39 participants in subsequent analyses.

Effect of Period. Table 4 shows the mean and corresponding standard deviations for all the physiological measures during the testing periods with different task difficulties. The measures comprised the sVRI-r using infrared light (sVRI-ri), sVRI-rg, and sVRI-t.

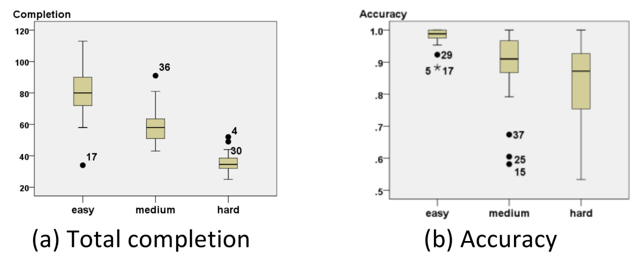


Fig. 11. Identification of low-effort performance: (a) total completed calculations vs. (b) accuracy.

Table 4. Results for Physiological Measures

	Easy task			**Medium task**			**Hard task**		
	Test period			Test period			Test period		
Measure	**Pre(base)**	**In**	**Post**	**Pre(base)**	**In**	**Post**	**Pre(base)**	**In**	**Post**
sVRI-ri	0.93 ± 0.05	0.94 ± 0.06	0.93 ± 0.05	0.92 ± 0.04	0.95 ± 0.05	0.93 ± 0.04	0.94 ± 0.05	0.97 ± 0.05	0.94 ± 0.05
sVRI-rg	1.00 ± 0.14	1.07 ± 0.10	1.05 ± 0.15	1.04 ± 0.17	1.12 ± 0.16	1.08 ± 0.18	1.05 ± 0.15	1.13 ± 0.13	1.10 ± 0.16
sVRI-t	0.91 ± 0.05	0.94 ± 0.05	0.93 ± 0.05	0.93 ± 0.04	0.95 ± 0.05	0.93 ± 0.04	0.93 ± 0.05	0.97 ± 0.06	0.93 ± 0.06

The physiological data were recorded during the pre- (Baseline), in-, and post-task periods for the easy, medium, and hard task difficulty levels.

Table 5. Within-subject Effects of the Period According to 2*1 Repeated-measures ANOVA

	Period		
Measure	p	F	η^2
sVRI-ri	**0.000***	22.334	0.370
sVRI-rg	**0.001***	13.539	0.263
sVRI-t	**0.000***	17.667	0.317

p and F values are shown for the main effects of period and $\eta 2$ for the effect sizes.

Table 6. Effects of Task Difficulty on the In-task Physiological Values

	Difficulty			
Measure	p	F	*df1, df2*	η^2
sVRI-ri	**0.018***	4.208	2, 76	0.100
sVRI-rg	0.053	3.047	2, 76	0.074
sVRI-t	**0.007***	5.330	2, 76	0.123

Repeated-measures ANOVA with two periods (baseline and in-task) * one difficulty was performed for each of the measures to check the within-subject effects (Table 5). We found that sVRI-ri, sVRI-rg, and sVRI-t all increased significantly when participants performed the tasks.

Effect of Difficulty. A variable's sensitivity to mental effort according to the task difficulty was represented as a main effect of the task difficulty. We analyzed the in-task values for the physiological measures without

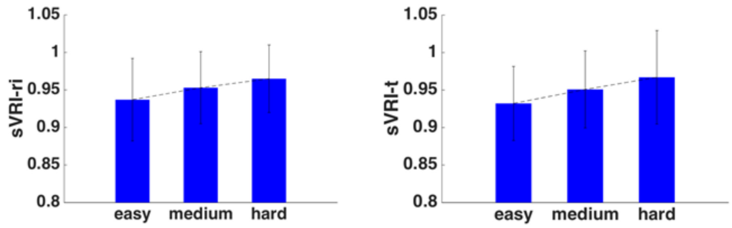
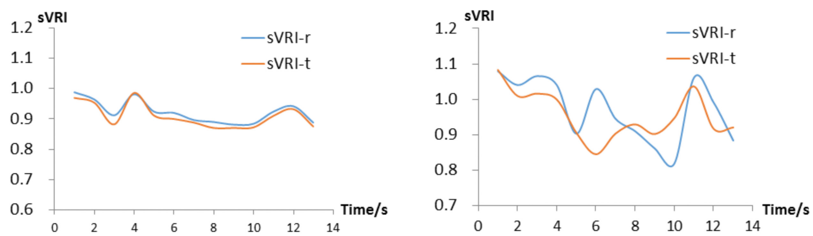


Fig. 12. Mean (SD) in-task values for the significant measures as a function of task difficulty.



(a) Fit with sVRI-t

(b) Inconsistency with sVRI-t

Fig. 13. Comparison of sVRI-ri and sVRI-t: fit with sVRI-t (a) and inconsistency with sVRI-t (b).

the out-of-task values. Table 6 shows the results obtained according to repeated-measures ANOVA with three difficulties * one period (in-task).

sVRI-ri and sVRI-t were affected significantly by the task difficulty, whereas sVRI-rg was not. We performed paired t-tests based on the three difficulty levels for sVRI-rg and found that only the comparison between the easy task and hard task was significant ($t(38) = -2.738$, $p = 0.009$), which may suggest that the green-light sensor implementation had limited sensitivity.

We also examined the trends in the mean values of the variables that were affected significantly by the task difficulty. Figure 12 shows the mean sVRI-ri and sVRI-t values with their standard errors, as also presented in Table 4. We can make the following conclusions based on these results. First, both sVRI-r and SVRI-t increased as the task difficulty increased from 0.93 (easy) to 0.98 (hard), where the mean change was about 5.4% and the standard errors were notable. The two measures had similar sensitivity and reliability levels. Second, the extent to which an individual's sVRI-ri or sVRI-t changed according to the task difficulty was similar to the difference in the sVRI values for two individuals with the same difficulty, where they were all around 5%. Thus, it is difficult to assess the level of the cognitive load and stress in individuals based simply on the absolute sVRI values. Our results only validated the performance of sVRI at analyzing the effects of period and difficulty with respect to the cognitive load and stress in the same subject.

Comparison of sVRI-t and sVRI-r. sVRI-r has advantages in terms of usability compared with sVRI-t due to its measurement method. According to the real-time outputs, we found that most of the simultaneous sVRI-t and sVRI-r waveforms agreed well with each other, as shown in Figure 13(a).

However, the sVRI-r methods were affected by the signal strength of the reflected light as well as environment noise, which substantially changed the PPG amplitudes and led to possible inconsistency with sVRI-t, as shown in Figure 13(b). In this study, we ignored the PPG waveforms with very low amplitudes when processing sVRI-r but the sensitivity of the sVRI-r method still degraded when distinguishing the difficulty levels ($p = 0.018$) compared with sVRI-t ($p = 0.007$).

Comparison of sVRI-ri and sVRI-rg. We found that the signal amplitude of the reflected green-light PPG was about 10% to 50% higher than that of the reflected infrared PPG, which may suggest the greater potential of green light for improving the signal quality. Furthermore, the sVRI-rg values were generally higher than the sVRI-ri values at the same time for a given participant. The standard deviations were also much higher for sVRI-rg (more than two times) than sVRI-ri, which was at a similar level to sVRI-t. In addition, sVRI-rg was not as sensitive as sVRI-ri at distinguishing the task difficulties. In summary, sVRI-rg may obtain stronger signals but its reliability still needs to be confirmed.

Hypotheses. Based on the experimental results, we confirmed Hypotheses A and B, but Hypothesis C was only partially confirmed. sVRI-r increased as the task proceeded and as the task difficulty increased. sVRI-r was sometimes inconsistent with sVRI-t due to degradation of the reflected signal quality.

6 LIMITATIONS AND CONCLUSIONS

Limitations. First, both the sVRI-t and sVRI-r cannot currently distinguish the cognitive load from stress despite their differences, and they may be suitable only for measuring autonomic mental effort (cognitive load) and the accompanying stress, and it does not reflect the PNS activities that are also correlated with cognitive processes and stress, because it is SNS-only. Second, the limited signal strength of the reflected light reduced its resolution of dynamic vasoconstriction. Future extensions to general touch modalities by adding a glass covering or surface material on the sensors would exacerbate these issues. In addition, high HR levels (e.g., above 120bpm) may degrade the signal quality for sVRI-r much more than sVRI-t due to the weak signals. Advanced techniques should be developed for preventing motion artifacts and for improving the sensitivity and reliability of sVRI-r while the target parts (e.g., fingers) move across the sensor surface.

Conclusions. In this study, we developed sVRI-r as a novel physiological measure for assessing cognitive load and stress, where it can be added to currently employed measures. Our experiment based on three levels of task difficulty and three testing periods showed that both the sVRI-r and sVRI-t methods were similarly affected significantly by the task difficulty and period, and their responses matched those of the most significant components from the BP and HRV. Our findings demonstrate the potential of sVRI-r as a sensitive, reliable, and usable parameter, especially for integration with ubiquitous touch interactions. In future research, we will focus on the current limitations of this method and improve its usability for various touch interactions.

ACKNOWLEDGMENTS

The authors thank all the volunteers for their participation in our experiment as well as the reviewers for their insightful comments.

REFERENCES

[1] F. Paas, J. E. Tuovinen, H. Tabbers et al. 2003. Cognitive Load measurement as a means to advance cognitive load theory. *Edu. Psychol.* 38, 1, 63–71.

[2] J. K. Kiecolt-Glaser, L. McGuire, T. F. Robles et al. 2002. Psychoneuroimmunology: Psychological influences on immune function and health. *J. Consult. Clin. Psychol.* 70, 3, 537–547.

[3] H. Sarker, M. Sharmin, A. A. Ali et al. 2014. Assessing the availability of users to engage in just-in-time intervention in the natural environment. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 909–920.

[4] S. Haim, R. Wang, S. E. Lord et al. 2015. The mobile photographic stress meter (MPSM): A new way to measure stress using images. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the ACM International Symposium on Wearable Computers*. 733–742.

[5] J. Healey and R. Picard. 2000. SmartCar: Detecting driver stress. In *Proceedings of the International Conference on Pattern Recognition*, 214, 218–221.

[6] H. Lu, D. Frauendorfer, M. Rabbi et al. 2012. StressSense: detecting stress in unconstrained acoustic environments using smartphones. In *Proceedings of the ACM Conference on Ubiquitous Computing*, 351–360.

[7] A.-M. Brouwer, M. A. Hogervorst, M. Holewijn et al. 2014. Evidence for effects of task difficulty but not learning on neurophysiological variables associated with effort. *Int. J. Psychophysiol.* 93. 2, 242–252.

[8] R. Brunken, J. L. Plass, and D. Leutner. 2003. Direct measurement of cognitive load in multimedia learning. *Edu. Psychol.* 38(1), 53–61.
[9] S. Oviatt. 2006. Human-centered design meets cognitive load theory: Designing interfaces that help people think. In *Proceedings of the ACM International Conference on Multimedia.* 871–880.
[10] S. G. Hart and L. E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Adv. Psychol.* 52, 6, 139–183.
[11] G. Matthews, S. E. Campbell, S. Falconer et al. 2002. Fundamental dimensions of subjective state in performance settings: Task engagement, distress, and worry. *Emotion* 2, 4, 315–340.
[12] W. B. Verwey and H. A. Veltman. 1996. Detecting short periods of elevated workload: A comparison of nine workload assessment techniques. *J. Exper. Psychol. Appl.* 2, 3, 270–285.
[13] J. Dennerlein, T. Becker, P. Johnson et al. 2003. Frustrating computers users increases exposure to physical factors. *Proceedings of the International Ergonomics Association Seoul Korea.*
[14] Y. Gao, N. Bianchi-Berthouze, and H. Meng. 2012. What does touch tell us about emotions in touchscreen-based gameplay? *ACM Trans. Comput.-Hum. Interact.* 19, 4, 1–30.
[15] L. Luo, L. Xiao, D. Miao et al. 2012. 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 *Proceedings of the International Conference on Biomedical Engineering and Biotechnology*, 929–933.
[16] F. Bousefsaf, C. Maaoui, and A. Pruski. 2014. Remote detection of mental workload changes using cardiac parameters assessed with a low-cost webcam. *Comput. Biol. Med.* 53(C), 154.
[17] H. Chigira, A. Maeda, and M. Kobayashi. 2011. Area-based photo-plethysmographic sensing method for the surfaces of handheld devices. *Proceedings of the ACM Symposium on User Interface Software and Technology.* 499–508.
[18] K. S. Yoo and W. H. Lee. 2011. Mental stress assessment based on pulse photoplethysmography. In *Proceedings of the IEEE International Symposium on Consumer Electronics.* 323–326.
[19] S. G. Hart and L. E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in Psychology*, A. H. Peter and M. Najmedin (Eds.). North-Holland, 139–183.
[20] W. B. Verwey and H. A. Veltman. 1996. Detecting short periods of elevated workload: A comparison of nine workload assessment techniques. *J. Exper. Psychol.: Appl.* 2, 3, 270—285.
[21] G. Matthews, S. E. Campbell, S. Falconer et al. 2002. Fundamental dimensions of subjective state in performance settings: Task engagement, distress, and worry. *Emotion* 2, 4, 315–340.
[22] S. D. Kreibig. 2010. Autonomic nervous system activity in emotion: A review. *Biol. Psychol.* 84, 3, 394–421.
[23] A.-M. Brouwer, M. A. Hogervorst, J. B. Van Erp et al. 2012. Estimating workload using EEG spectral power and ERPs in the n-back task. *J. Neural Eng.* 9, 4, 045008.
[24] L. Luo, L. Xiao, D. Miao et al. 2012. 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 *Proceedings of the International Conference on Biomedical Engineering and Biotechnology.* IEEE Computer Society, 929–933.
[25] C. Iani, D. Gopher, and P. Lavie. 2004. Effects of task difficulty and invested mental effort on peripheral vasoconstriction. *Psychophysiology* 41, 5, 789–798.
[26] G. G. Berntson, J. T. Cacioppo, and K. S. Quigley. 1991. Autonomic determinism: The modes of autonomic control, the doctrine of autonomic space, and the laws of autonomic constraint. *Psychological Review.* 98, 4, 459–487.
[27] J. Hernandez, P. Paredes, A. Roseway et al. 2014. Under pressure: Sensing stress of computer users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 51–60.
[28] Y. Lyu, X. Luo, J. Zhou et al. 2015. Measuring photoplethysmogram-based stress-induced vascular response index to assess cognitive load and stress. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* ACM, 857–866.
[29] C.-C. Chen, S.-C. Lin, M.-S. Young et al. 2014. Accumulated mental stress study using the meridians of traditional Chinese medicine with photoplethysmography. *J. Alt. Comp. Med.* 20, 11, 860–867.
[30] M. P. Fink. 2014. Analyzing the value of monitoring duodenal mucosal perfusion using photoplethysmography. *Crit Care.* 18, 5, 561.
[31] T. Tamura, Y. Maeda, M. Sekine et al. 2014. Wearable photoplethysmographic sensors—Past and present. *Electronics* 3, 2, 282.
[32] R. Wijshoff, J. Veen, A. M. Van der Lee et al. 2011. PPG motion artifact handling using a self-mixing interferometric sensor. In *Proceedings of the 11th Conference on Optical Fibers, Sensors, and Devices for Biomedical Diagnostics and Treatment,* SPIE (International Society for Optical Engineering). doi: 10.1117/12.874170
[33] R. W. Wijshoff, M. Mischi, J. Veen et al. 2012. Reducing motion artifacts in photoplethysmograms by using relative sensor motion: Phantom study. *J. Biomed. Optics* 17, 11, 117007–117007.
[34] F. E. Ritter, A. L. Reifers, L. C. Klein et al. 2007. Lessons from defining theories of stress for cognitive architectures. *Integr. Models Cogn. Syst.* 1, 254.
[35] N. Hjortskov, D. Rissén, A. K. Blangsted et al. 2004. The effect of mental stress on heart rate variability and blood pressure during computer work. *Eur. J. Appl. Physiol.* 92(1–2), 84–89.

Received February 2017; revised November 2017; accepted February 2018