Analysis on Mental Stress/Workload Using Heart Rate Variability and Galvanic Skin Response during Design Process

Xu Xu

A Thesis in the

Concordia Institute for Information Systems Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Applied Science

(Quality Systems Engineering)

Concordia University

Montreal, Quebec, Canada

April 2014

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: **Xu Xu**

Entitled: **Analysis on Mental Stress/Workload Using Heart Rate Variability and Galvanic Skin Response during Design**

and submitted in partial fulfillment of the requirements for the degree of

MASTER OF APPLIED SCIENCE (Quality Systems Engineering)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

______________________________ Chair
Dr. Roch Glitho

______________________________ Examiner
Dr. Mohammad Mannan

______________________________ Examiner (BCEE)
Dr. Lan Lin

______________________________ Supervisor
Dr. Yong Zeng

Approved by ______________________________
Chair of Department or Graduate Program Director

____
Dean of Faculty

Date ______________________________

Abstract

Analysis on Mental Stress/Workload Using Heart Rate Variability and Galvanic Skin Response during Design Process

Xu Xu

Mental stress/workload is now considered as an important element that affects design creativity. Many physiological measures such as Galvanic Skin Response (GSR), Electrocardiogram (ECG), Electromyogram (EMG), Blood Pressure (BP), Skin Temperature (ST), Blood Volume Pulse (BVP), respiration rate (RIP) and Electroencephalogram (EEG) are used to identify mental stress level (Karthikeyan, Murugappan, & Yaacob, 2011; Lee, Jo, & Lee, 2011; Michon, 1966; Roscoe, 1992; Ryu & Myung, 2005; Wilson, 2002).

This thesis aims to explore the relation between mental stress/workload and physiological measures and using those measures to identify the correlation between mental stress and design activities. One measure is Heart Rate Variability (HRV). The LF/HF ratio derived from HRV was used as an indicator of mental stress. The average value of LF/HF of different activities during a design experiment was computed and then clustered into seven levels which represent different levels of mental stress. The result showed that, during a cognitive design experiment, most of the activities in a design process were performed under low levels of mental stress and there was no correlation between types of design activities and levels of mental stress. Another measure is Galvanic Skin Response (GSR). Mean GSR of several design activities under different stress levels during a design experiment was calculated to investigate if it can be used as a reliable measure of mental workload. The result showed that mean GSR is significantly lower in baseline compared to working state and GSR in the pre-test resting state is significantly lower than GSR in the post-test resting state.

Acknowledgements

I would like to express my special gratitude to my advisor Dr. Yong Zeng, for his invaluable help, guidance, and encouragement during my Master study and research.

I would also like to thank my colleagues in the Design Lab. I am proud of being one of the members of Design lab. Thanks to Ms. Thanh An Nguyen for her assistance of designing experiments. Also, I would like to thank Mr. Suo Tan for his help in modifying HRV analysis program.

My appreciation is extended to all my friends in Canada who have shared the joyful and bitter moments during this portion of my life.

Last but not least, I would like to thank my family. They were always supporting me and encouraging me with their best wishes.

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1. Background and Motivation

Recently, increasing efforts were paid to understand design creativity. Numerous research was conducted to improve the creativity during a design process. For example: Goldschmidt & Tatsa (2005) proposed that good ideas are the base of creative design processes, Arciszewski et al (1995) presented that constructive induction can be used as the foundation of design creativity, Dorst & Cross (2001) analyzed empirical data on design processes, evaluated them on the overall quality and some aspects of creativity. They also applied their observation to a creative design model and suggested that the creative model needed refinements, Li et al. (2007) presented a qualitative model of creativity for product innovation which ascertain the direction and choose solution of improving creative product design, they have also found some key attribute that affect the creativity which included the thinking styles, knowledge, information, design methods and supporting tools. Nguyen & Zeng (2012) postulated that design reasoning follows a nonlinear dynamics which may become chaotic and design creativity is related to designer's mental stress through an inverse U shaped curve as illustrated in Figure 1.

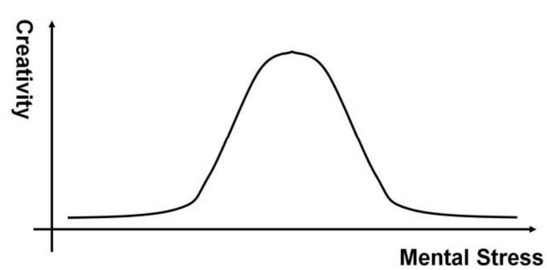

Figure 1. Relation between mental stress and creativity (Nguyen & Zeng, 2012).

Among those findings, the relation between designer's mental stress and creativity is a very new one and the relation has not been confirmed. In order to achieve the accurate result that mental stress is related to design creativity, the first and fundamental task is to find a suitable way to quantify mental stress during design activities. As a result, different physiological measures are identified to quantify mental stress. Those measures include: Galvanic Skin Response (GSR), Electrocardiogram (ECG), Electromyogram (EMG), Blood Pressure (BP), Skin Temperature (ST), Blood Volume Pulse (BVP), respiration rate (RIP) and Electroencephalogram (EEG) (Karthikeyan et al., 2011; Lee et al., 2011; Michon, 1966; Roscoe, 1992; Ryu & Myung, 2005; Wilson, 2002).

1.2. Objectives

The objectives of this thesis are to find relations between two physiological measures (Heart Rate Variability and Galvanic Skin Response) and mental stress/workload and to find relations between mental stress/workload and design activities.

1.3. Contributions

This thesis focuses on finding the relations between mental stress and physiological measures and correlation between mental stress/workload and design activities. The main contributions of this thesis include:

- After reviewing a series of papers related to mental stress and HRV, the reliability of LF/HF ratio calculated from HRV as an indicator of mental stress was confirmed.

- Relation between the numbers of design activities and mental stress levels was identified by analyzing HRV data gathered from a design experiment conducted in 2010.
- Relation between mean GSR and mental workload levels during design activities was identified which is slightly different from some other research.
- Relation between small workload change and corresponding mean GSR change was identified which confirmed the result of previous studies.

1.4. Research Methodology

This thesis has two major sections. One is to quantify mental stress during design activities with HRV and find the relations between design activities and mental stress. The other one is to conduct research on GSR to test if it is a reliable index of mental workload.

For the HRV work, previous work that conducted in our research group will be reviewed. The work was to choose an HRV parameter which best represents mental stress. Stroop color word test, introduced by Stroop (1935) was used as stressor. After confirming one reliable HRV parameter, the parameter of designers during a design experiment will be calculated by a Matlab based HRV analysis program-HRVAS (Ramshur, 2010). The design experiment was conducted by our research group which has been used in many research works. Then the value of the parameter will be clustered into seven parts which represent different stress levels. The numbers of design activities in different stress levels will then be calculated and the numbers will be analyzed by Friedman test and Wilcoxon test to evaluate in which stress level designers perform most of the activities.

For the GSR work, to achieve our goal, we conducted twenty-three design experiments. For each experiment, subjects were required to perform five design tasks and evaluate other designers' design result with the same design problem. During the experiment, subjects would generally experience several phases: pre-test resting, understanding (six tasks), solution generation (six tasks), evaluation (six tasks), and after-test resting. The workload for each design task will be evaluated by the subjects themselves when each design task ends. The GSR data will be monitored and recorded by CAPTIV-L7000-equipment which can also be used to record other physiological signals such as HRV, Respiration, etc. Then the average value of GSR for each phase is analyzed by a non-parametric statistical method-Wilcoxon Signed-Rank Test.

1.5. Thesis Organization

This thesis is organized as follow:

1. Section 2 will conduct a comprehensive review which focuses on identifying the relation between design creativity and mental stress, relation between HRV and mental stress, relation between GSR and mental stress, and limitation of previous research.

2. Section 3 will introduce two design experiments which were conducted in 2010 and 2013 respectively.

3. Section 4 will analyze data generated from those two experiments and present two analysis results respectively. In 2010, the data were numbers of design activities and HRV. In 2013, data include mean GSR, mental workload values evaluated by

the subjects.

4. Section 5 will discuss the result and generate conclusions based on a short discussion.

5. Section 6 will point out constrains of the thesis and future work.

2. Literature Review

2.1. Design creativity and mental stress

Creativity has been defined in many ways. For instance: Hennessey (2003) defines creativity as 'the process by which something judged (to be creative) is produced, Sternberg & Lubart (1999) defined creativity as that which 'produce(s) work that is both novel (i.e., original, unexpected) and appropriate (i.e., useful, adaptive concerning task constraints)', etc.

In recent years, more and more efforts were paid to understanding design creativity. In the paper presented by Li, et al. (2007), different factors which include intelligence of the designer, knowledge thinking style, personality, motivation, environmental context, etc. that may affect design creativity were analyzed (Sternberg & Lubart, 1995). An approach was then proposed based on the characteristics of product design process. The approach integrated the creative thinking methods of cognitive psychology, information technologies and computer application technologies which can be found in Figure 2.

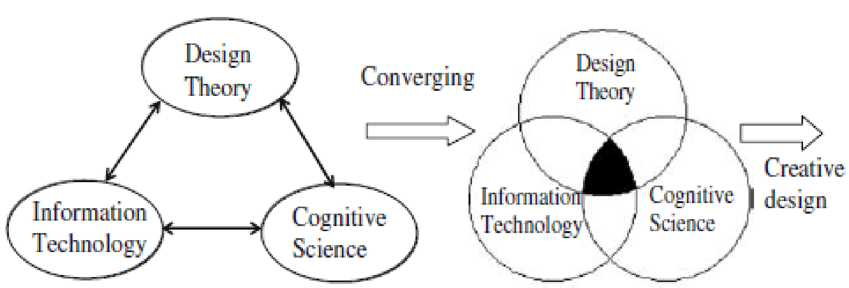

Figure 2. Creative design converging theory, information technology and cognitive science (Li et al., 2007).

Other models and concepts were also developed to understand creativity and to effectively implement creativity in product development process. For example, Polverini et al. (2011) proposed a design and managerial step-based framework which combined creativity, project management and technical development in industrial innovation. Zhang et al. (2012) developed a conceptual model of design for systematic innovation based on the rational problem-solving process and reflective practice. Sedenkov (2012) introduced a design theory and presented several requirements to design the theory. Luximon and Luximon (2010) proposed a method for creative product design using 3D CAD model as a source of inspiration. They use Juicy Salif, a lemon squeezer, as a source of inspiration to design jewelry rings. Filippi and Motyl (2010) investigated the relation and difference between the Theory of Inventive Problem Solving (TRIZ) and the Interaction Design (ID), and developed a more structured TRIZ-based approach to the ID process. Greene (2012) reasoned the failures of existing models of innovation and proposed corrections. Wang et al. (2010) introduced the creativity enhanced FBS model in which they implemented creativity in the function and the form layer and adopted the quotient space theory to represent the creative process mathematically. Graessler et al. (2012) discussed the low application rate of design methodologies in industrial practices and explained some general barriers lying in the nature of human problem solving, structural and organizational barriers. Since sketching is very important in creative design, Roth-Koch and Westkaemper (2010) reported a method to transform freehand sketches to digital three-dimensional (3-D) shapes. Nguyen and Zeng (2012) proposed a theoretical model of design creativity which illustrated the relation between mental stress and creativity.

Among those studies, it was found that very few research was paid to find the relation between mental stress and design creativity. However, in the research conducted by Nguyen and Zeng (2012) , a postulate was proposed which showed that design creativity is related to designer's mental stress through an inverse U shaped curve as illustrated in literature review section. Additionally, they reviewed previous research conducted by Tang and Zeng (Tang, 2009; Tang & Zeng, 2009) and hypothesized that the level of mental stress is positively related to workload and negatively related to mental capacity. They proposed that workload can be associated with the complexity of the problem. Mental capacity can be defined by a designer's knowledge, skills and affect. The relation between workload, mental capacity and mental stress can be found in Figure 3.

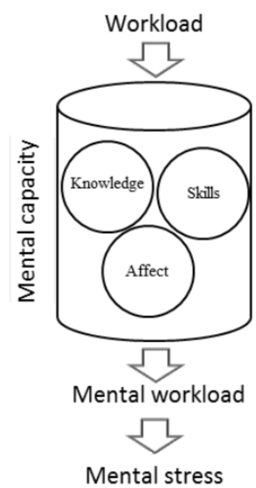

Figure 3. Relation between workload, mental capacity, and mental stress (Nguyen & Zeng, 2012).

2.2. Heart Rate Variability (HRV) and its relation to mental stress

HRV is known as the fluctuation in heart beat intervals, as shown in Figure 4.

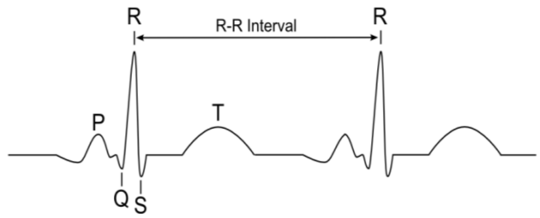

Figure 4. R-R Interval (Nguyen, Xu, & Zeng, 2013).

HRV has been used for monitoring mental health because of its strong dependency on the Autonomic Nervous System (ANS) (McCratyet al., 1995; Wijsman et al., 2011).

In the research that conducted by Boonnithi & Phongsuphap (2011), a number of HRV parameters which include mean heart rate, power spectrum in very low, low and high frequency ranges (VLF, LF, HF), etc. were computed and compared. HRV can be calculated by two methods: time domain and frequency domain. Time domain method could calculate a large number of parameters, but the paper only selected some of the most promising ones. For example, mean R wave to R wave interval (mRR), mean Heart Rate (mHR), Standard Deviation of RR interval (SDRR), Standard Deviation of Heart Rate (SDHR), etc. Frequency domain method estimated the Power Spectral Density (PSD) of the RR interval signals. A set of parameters (normalized very low frequency spectrum (nVLF), normalized low frequency spectrum (nLF), normalized high frequency spectrum (nHF), etc.) were generated to characterize HRV by three basic parameters: VLF, HF, and LF. Then, those parameters were evaluated by mental stress and normal state. It was found that mRR, mHR, nLF, nHF and SVI (Symphatovagal balance index, calculated by LF/HF) were more influenced by the effect of mental stress. The calculation methods of time domain and frequency domain can be found in Table 1.

Table 1. HRV calculation methods (time domain and frequency domain) (Boonnithi & Phongsuphap, 2011)

Measure	Unit	Formula (time domain)	Measure	Unit	Formula (frequency domain)
mRR	ms	$\sum_{i=1}^{N}(RR_i)$	*VLF*	ms^2	*Power spectrum from 0.003 to 0.04 Hz*
mHR	bpm	$\frac{\sum_{i=1}^{N}\left(\frac{60000}{RR}_i\right)}{N}$	*LF*	ms^2	*Power spectrum from 0.04 to 0.15 Hz*
SDRR	ms	$sqrt(\frac{\sum_{i=1}^{N}(RR_i - mRR)^2}{N-1})$	*HF*	ms^2	*Power spectrum from 0.15 to 0.4 Hz*
SDHR	bpm	$sqrt(\frac{\sum_{i=1}^{N}(60000/RR_i - mHR)^2}{N-1})$	*nVLF*	%	*VLF*100 /(VLF+LF+HF)*
CVRR		$SDRR * \frac{100}{mRR}$	*nLF*	%	*LF* 100 /(VLF+LF+HF)*
RMSSD	ms	$sqrt(mean((RR_{i+1} - RR_i)^2))$	*nHF*	%	*HF *100 /(VLF+LF+HF)*
pRR20	%	$\frac{Count(\lvert RR_{i+1} - RR_i \rvert)_{>20ms} * 100}{N-1}$	*dLFHF*	%	*| nLF – nHF |*
pRR50	%	$\frac{Count(\lvert RR_{i+1} - RR_i \rvert)_{>50ms} * 100}{N-1}$	*SMI*		*LF/(LF+HF)*
			VMI		*HF/(LF+HF)*
			SVI		*LF/HF*

In another paper "The effect of mental stress on heart rate variability and blood pressure during computer work" (Hjortskov et al., 2004), an experiment was conducted to find the correlation between mental stress and HRV/blood pressure. Twelve female students were involved in the experiment. They were asked to remember six random numbers showed on a computer screen and type them in the next four seconds. The experiment was separated into three sessions: Introductory Session (IS), Stress Session (SS), and Control Session (CS). In IS and SS, participants were treated unfriendly and monitored by a video camera. Before these two sessions, they needed to remember 12 common words (each word was viewed for 2 seconds). After those two sessions, they had to present the words to the video camera as many as possible. In CS, they were treated friendly and the camera was turned off. Each session consisted of four work period, each period was 3 minutes and they were separated by short breaks of 30 seconds. Each session was followed by an approximately 8 minutes rest period. During the experiment, HRV data and blood pressure data were recorded and analyzed. The result showed that blood pressure was higher when subjects were under stress compared to rest session and HRV is more sensitive and selective measure of mental stress (Hjortskov et al., 2004).

In 2011, an experiment was conducted by our research group and a thesis was then completed by Petkar (Petkar, 2011). The objective of the experiment was to choose an HRV parameter which can best represent mental stress.

In the experiment, Stroop color word test (Stroop, 1935), was used as a stressor. The Stroop test is a psychological experiment, which aims to demonstrate the impact of interference on subjects' reaction time. When the name of a color is displayed in a color not denoted by the name (e.g., the word "yellow" displayed in red font instead of yellow font), it takes longer and is easier to make mistakes for a subject to name the color of the word than when the name of a color is displayed in the same color denoted by the name (e.g., the word "yellow" displayed in yellow font) (Stroop, 1935).

In the experiment conducted by Petkar (2011), a stimulus word was displayed for 0.5 seconds. Then an answer list was displayed and subjects had to choose the right answer within 1.5 seconds. A blank screen was displayed for 1 second between two Stroop tasks. The Stroop test was divided into six difficulty levels were designed based on different combinations of color and color name in both stimulus word and the answer list (as shown in Figure 5). The difficulty levels were designed into a High interference (i.e. very high difficulty level follows very low difficulty level) and a Low interference (i.e. difficulty increase gradually). Twenty subjects (21-30 years old) volunteered in the experiment.

Figure 5. Six difficult levels (Petkar, 2011).

In the experiment, HRV data was analyzed using the program (HRV analysis) developed by Niskanen et al. (2004). With this program, HRV variables such as mean heart beat intervals (mRR), mean heart rate-number of heart beat per minutes (mHR), standard deviation of heart rate per minute (STD), low frequency of HRV in normalized unit (LF(n.u.)), high frequency of HRV in normalized unit (HF(n.u.)) and ratio of (LF/HF) can be calculated. After the experiment, those variables during High Stroop interference and Low Stroop interference were studied. LF(n.u.) is defined as LF(n.u.)= LF/(LF + HF) and HF(n.u.) is defined similarly.

In addition to the value of HRV variables, performance and response time of those subjects during the experiment were also calculated which are presented in Figure 6. The result showed that when the subjects were under Low interference, they would use less time and the performance would be better than High interference which represented that Low and High Stroop interference clearly have different load effects on subjects.

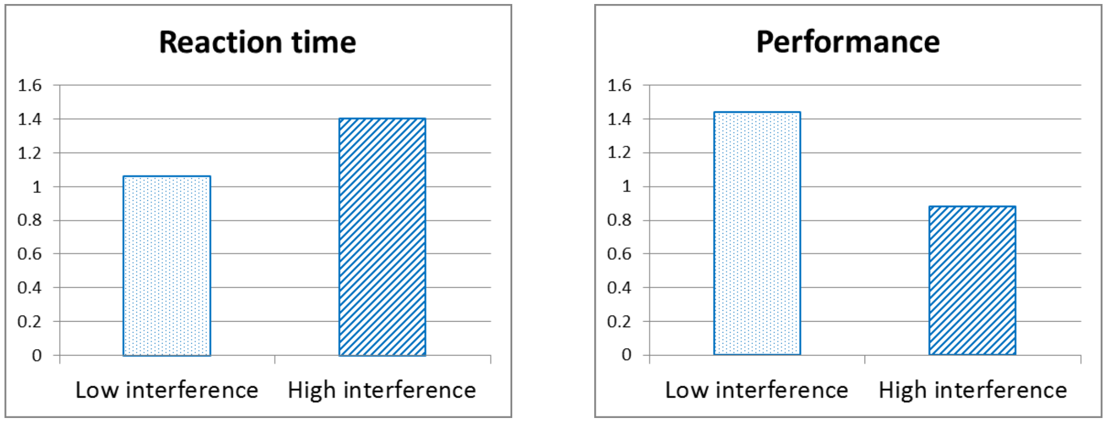

Figure 6. Reaction time and performance of Low and High interference (Petkar, 2011).

Two sided t-test was conducted between Stroop test and relaxing state at a 0.05 significant level. According to the result, it was found that LF, HF and LF/HF ratio have larger changes than mRR and mHR. Additionally, among those measures, LF/HF ratio experienced the largest changes which compared the values when subjects were relaxing and doing High Stroop interference test (Petkar, 2011).

Several Matlab based programs have been developed to analyze HRV data. One of them is HRV Analysis which was used in Petkar's research work (Petkar, 2011) we have mentioned in the previous content. It is a popular tool for assessing the activities of Autonomic Nervous System which cause mental stress (Niskanen et al., 2004). It was originally developed by Matlab. However, it can run without a Matlab installation. The sample interface of the software can be found in Figure 7, a sample output report of the software is shown in Figure 8.

Figure 7. Interface of HRV Analysis.

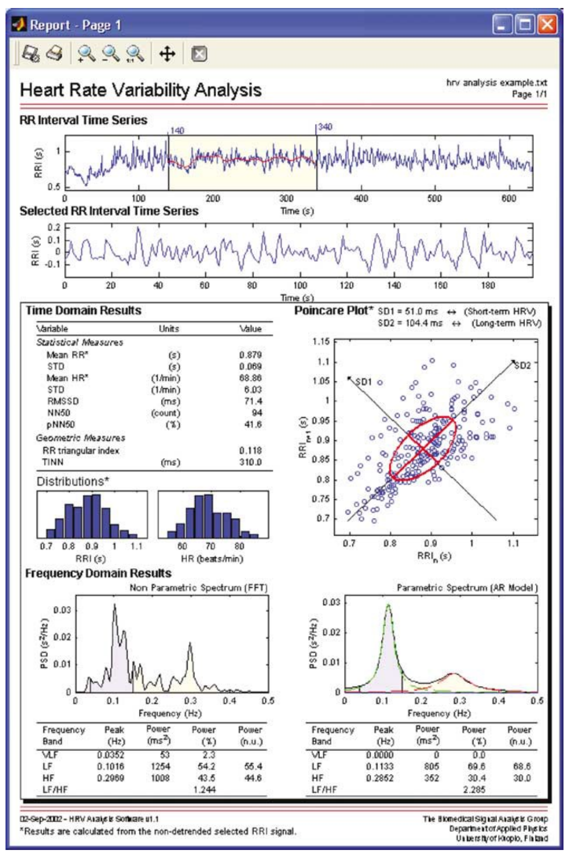

Figure 8. Sample output report.

HRVAS is another HRV analysis program which has been commonly used to calculate HRV parameters (Ramshur, 2010). This program implemented four major categories of HRV calculating techniques: statistical and time-domain analysis, frequency-domain analysis, nonlinear analysis, and time-frequency analysis. The interface of the software is illustrated in Figure 9. This software will also be used in HRV analysis in this thesis. This program will be the one we choose for data analysis since it can present HRV analysis result in every half second.

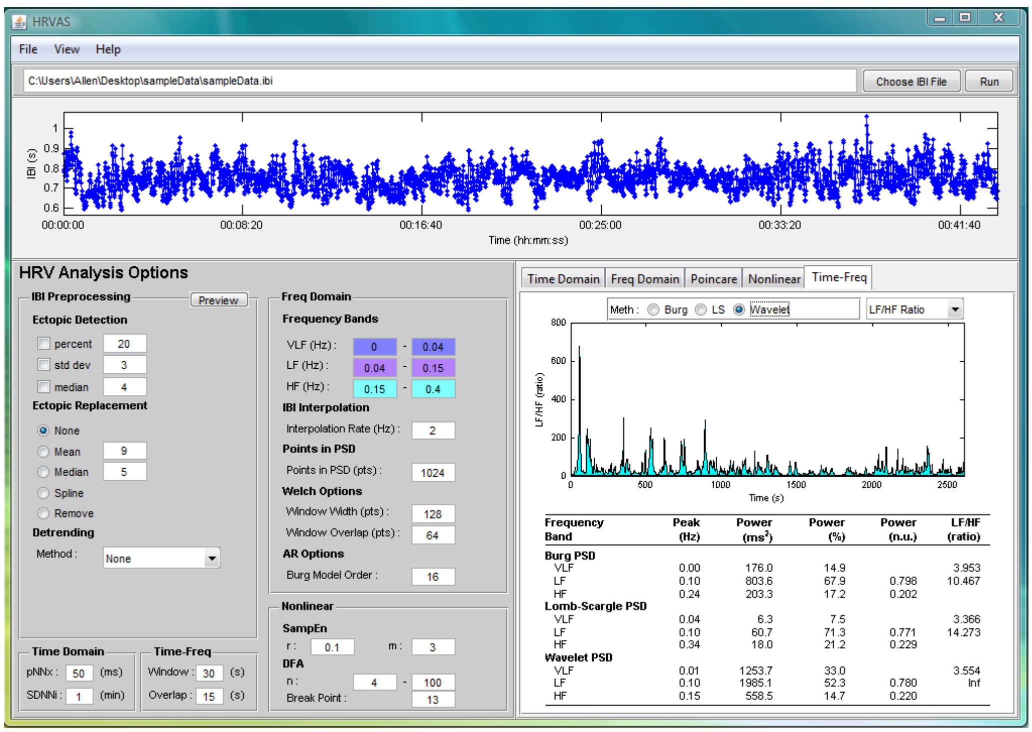

Figure 9. Interface of HRVAS.

2.3. Galvanic Skin Response (GSR) and its relation to mental stress/workload

GSR, also known as electrodermal activity (EDA), is a low-cost, easily-captured, and robust one for mental workload evaluation (Nourbakhsh, et al., 2012).

In the paper presented by Shi et al. (2007), Galvanic Skin Response (GSR) was used to evaluate people's stress and arousal levels when they were using unimodal and multimodal versions of the same interface. In the paper, the author's review showed that GSR variation and stress/SNS arousal were always linked (Selye, 1956). They also found that GSR increase or decrease respectively when the person was more or less stressed (Miller & Shmavonian, 1965). As a result, GSR was used as a control measure to assess the reliability and accuracy of a number of potential cognitive load indicators. The experiment was designed into 36 unique tasks in total, 12 in each set. Subjects had to use a traffic control management which tasks involved marking incidents on a map and the deployment of resources. Each 12 tasks were completed with a same interface, but with a different interaction condition: a) Multimodal interaction with hand gesture and speech input; b) Speech-only interaction; and c) Gesture-only interaction. The 12 tasks were separated into four levels, by providing the subjects with increasing difficulty levels. The four levels were illustrated in Table 2.

Table 2. Four difficulty levels (Shi et al., 2007).

Levels	Entities	Actions	Extras	Time
1 (L)	6	3	2	∞
2 (M)	10	8	2	∞
3 (M)	12	13	4	∞
4 (H)	12	13	4	90 sec.

The difficulty levels can be achieved in five ways:

1. **Visual Complexity:** The number of streets in beach map increased from 40 to 60.

2. **Entities:** The number of distinct entities in the task description was increased

3. **Actions to completion:** The minimum number of actions required for task completion.

4. **Extras:** The number of distracter entities (entities not needed for the task) increased.

5. **Time Limit:** The most difficult level was induced by implementing a time limit for completion.

Eleven subjects were involved in the experiment. After the experiment, mean GSR and accumulated GSR were calculated. An ANOVA test on accumulated GSR values for Low, Med and High levels was conducted, and showed that they are significant different. The result of the analysis showed that mean GSR of the subjects increases when his/her cognitive load increases. Additionally, it demonstrated that people experienced lower cognitive load levels when they use a multimodal interface (Shi et al., 2007).

In another research conducted by Nourbakhsh et al. (2012), two experiments were conducted to assess GSR data: one experiment is arithmetic task and another one is reading task. In arithmetic task, subjects had to perform tight arithmetic tasks with four difficulty levels. GSR data was collected and normalized. The result showed that there is an increasing trend between task difficulty and accumulative GSR. It also showed that the normalisation leads to more significant results between the four difficulty levels. In the reading task, subjects were required to perform three silent reading tasks. There were four slides of texts in each task and each slide was presented for 30 seconds. The subjects were asked to find words of certain lengths in each slide. The tasks were divided into three difficulty levels. The result also illustrated that normalisation produced significant result for average GSR and accumulative GSR results. Therefore, the authors concluded that normalisation effectively improves the significance of distinction between cognitive load levels for the mean and accumulative GSR (Nourbakhsh et al., 2012).

In a study of driver's cognitive load, a simulator and a field experiment were conducted to find the cognitive load on driving performance. In the experiment, ECG and GSR were used as measures of mental workload. The GSR analysis showed that mean GSR was higher in complex tasks than in baseline but GSR could not differentiate between task difficulties (Engström et al., 2005). In another study evaluating tasks at two levels of difficulty during computer work, GSR could significantly detect easy level from hard level but in contrast to previous results reported, GSR was higher at easy level than at hard level (Ikehara & Crosby, 2005). The authors interpreted this could be the result of stress associated with easy level because the participants were bored when doing easy tasks.

A very recent work on assessing cognitive load in stress and non-stress conditions with GSR reported that mean GSR could only show differences between task difficulties in non-stress condition, not in stress condition. In the meantime, GSR peak behaved consistently in both conditions. During the experiment, lack of control, task failure and social-evaluation were used as stressor (Conway et al., 2013). The study confirmed the importance of choosing GSR features when evaluating workload as well as the results presented in (Setz et al., 2010) that GSR peak reflects stress information.

3. Experiment Design

3.1. Design Experiment in 2010

In 2010, a design experiment was conducted by our research group. This experiment was designed by a PhD student (Thanh An Nguyen) from our research group. In the experiment, eleven graduate students with engineering background in the age range of 25 to 35 years old were selected as subjects. The experiment protocol is approved by Human Research Ethics and Compliance, Concordia University. Three devices-HRV monitor-Polar RS800G3, Electroencephalogram (EEG), and cameras-were used in this experiment to record the subject's physiological responses and design activities. In this article, we only analyzed HRV because result from EEG measurement were reported in another paper written by Nguyen and Zeng (2013).

In the experiment, each subject was required to perform one of the following design tasks chosen by the experimenter:

1. Design a vehicle that can transport an object between any two locations on earth within a few seconds.
2. Design a desk that helps a messy university student to keep things organized and tidy.
3. Design a house that can easily fly from one place to another place.

Subjects had to follow the following procedure:

1. Relax for 3 minutes with eyes open.

2. Relax for 3 minutes with eyes closed.

3. Solve the design problem chosen by the experimenter (no time limit).

4. After the design:

   a. Relax for 3 minutes with eyes closed.

   b. Interviewed by the experimenter.

5. Experiment ends.

During the experiment, subjects were allowed to use electronic tablet to express/sketch their design ideas and to make notes, go back and forth to check their previous sketches/notes at any time, use the internet. Once the experiment started, they are not supposed to ask any questions related to the design. However, they can ask for assistance regarding the use of the tools that used for design. Additionally, they would not be interrupted during the design task.

All the subjects used tablet to design. On average, each designer used 45 min to complete the design.

HRV was recorded by Polar RS810 (Figure 10. Polar RS800). The activities of all subjects and tablet monitor were recorded by cameras. Figure 11 shows a sample design activity and HRV data.

Figure 10. Polar RS800 (Lab, 2006-2013a).

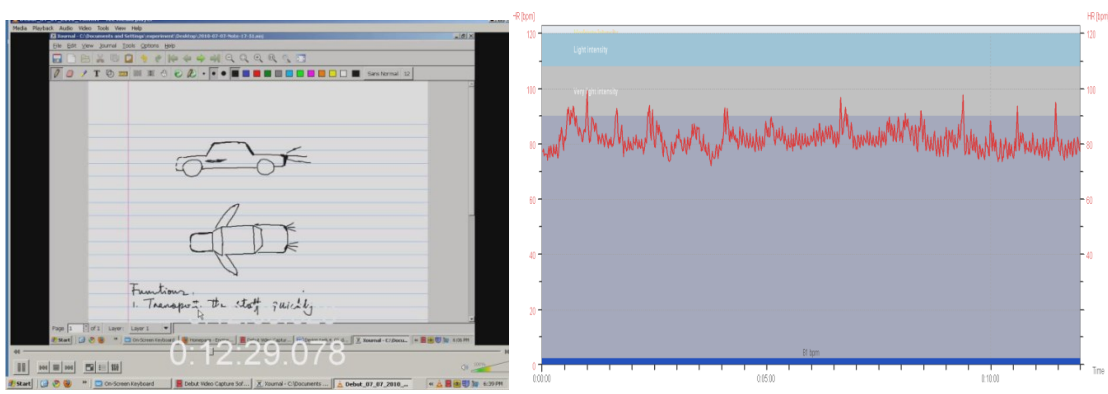

Figure 11. Sample design activity and HRV data (Nguyen et al., 2013).

3.2. Design experiment in 2013

In 2013, an upgraded design experiment was conducted which was developed by the same student (Thanh An Nguyen). It aimed to quantify designers' mental stress with physiological measures (EEG, HRV, GSR, Respiration, etc.).

Twenty-three (23) subjects (16 males and 7 females) with engineering background volunteered to participate in the experiment. At the beginning of the experiment, the subjects were required to rest for 3 minutes with eyes closed. Then they were asked to finish 6 design tasks. Each design task was separated into three parts: problem understanding part, solution expression part and evaluation part. In the problem understanding and solution expression part, the subjects were required to read a design problem (a sample question is shown in Figure 12 and then solve it. In the evaluation part, they needed to evaluate two proposed design solutions to the same design problem given in the problem understanding-solution expression part (an example is illustrated in Figure 13).

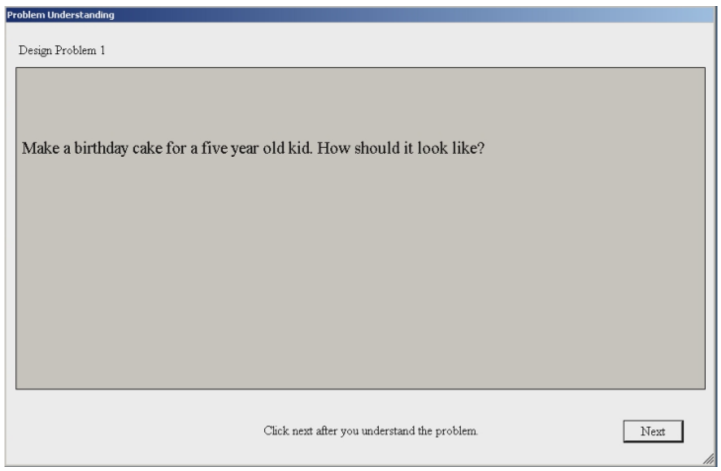

Figure 12. Sample design problem (Xu, Nguyen, & Zeng, 2014).

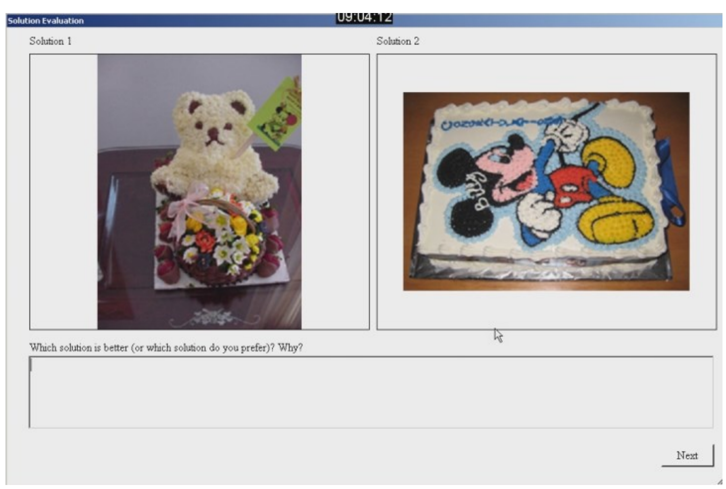

Figure 13. Evaluation part (Xu et al., 2014).

At the end of each part of each design task, the subjects were required to evaluate their mental workload by rating their mental demand, time demand, performance, effort, and stress level. This cognitive load evaluation method was developed by Hart and Staveland (1988). A screenshot of workload evaluation is shown in Figure 14.

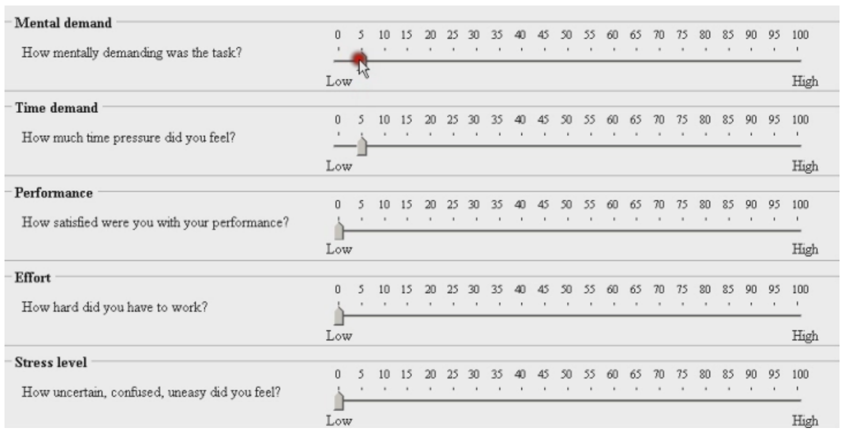

Figure 14. Workload evaluation.

After completing all six design tasks, the subjects rest for another 3 minutes with eyes closed.

The six design tasks were:

1. Make a birthday cake for a five year old kid. How should it look like?
2. Sometimes, we don't know which items should be recycled. Create a recycle bin that helps people recycle correctly.
3. Create a toothbrush that incorporates toothpaste. (Incorporate = include, combine)
4. In Montreal, people on wheelchair cannot use Metro safely because most of Metros have only stairs or escalators. Elevator is not an option because it is too

costly to build one. You are asked to create the most efficient solution to solve this problem.

5. Employees in an IT company are sitting too much. The company wants their employees to stay healthy and work efficiently at the same time. You are asked to create a workspace that can help employee to work and exercise at the same time.

6. Two problems with standard drinking fountain:

    - First one: filling up water bottle is not easy.
    - Second one: People who are too short cannot use the fountain and people who are too tall have to bend over too much.

    Create a new drinking fountain that solves the problems.

The complete procedure of the experiment was:

1. Rest for 3 minutes with eye closed.
2. Solve the problem 1 and evaluate workload using the workload evaluation table which can be found in Figure 14.
3. Evaluate two proposed design solutions (designed by others) to design problem 1 and evaluate workload with the same evaluation table.
4. Solve the problem 2 and evaluate workload using the workload evaluation table.
5. Evaluate two proposed design solutions (designed by others) to design problem 2 and evaluate workload with the same evaluation table.
6. Solve the problem 3 and evaluate workload using the workload evaluation table.
7. Evaluate two proposed design solutions (designed by others) to design problem 3 and evaluate workload with the same evaluation table.

8. Solve the problem 4 and evaluate workload using the workload evaluation table.
9. Evaluate two proposed design solutions (designed by others) to design problem 4 and evaluate workload with the same evaluation table.
10. Solve the problem 5 and evaluate workload using the workload evaluation table.
11. Evaluate two proposed design solutions (designed by others) to design problem 5 and evaluate workload with the same evaluation table.
12. Solve the problem 6 and evaluate workload using the workload evaluation table.
13. Evaluate two proposed design solutions (designed by others) to design problem 6 and evaluate workload with the same evaluation table.
14. Rest for another 3 minutes with eye closed.
15. Experiment ends.

There was no time limit for all of the subjects. On average, it took each subject one hour to finish the experiment. However, six subjects used approximately two hours to finish the experiment.

All the design tasks were performed on a tablet and were recorded. During the experiment, GSR data were recorded at a frequency of 32 Hz and were collected by CAPTIV-L7000 (Figure 15). Other data such as HRV, Respiration, EEG, etc. are also recorded. But in this thesis, only GSR data will be analyzed. In Figure 16, it illustrates an experiment in progress.

Figure 15. CAPTIV-L7000 (TEA, 2010).

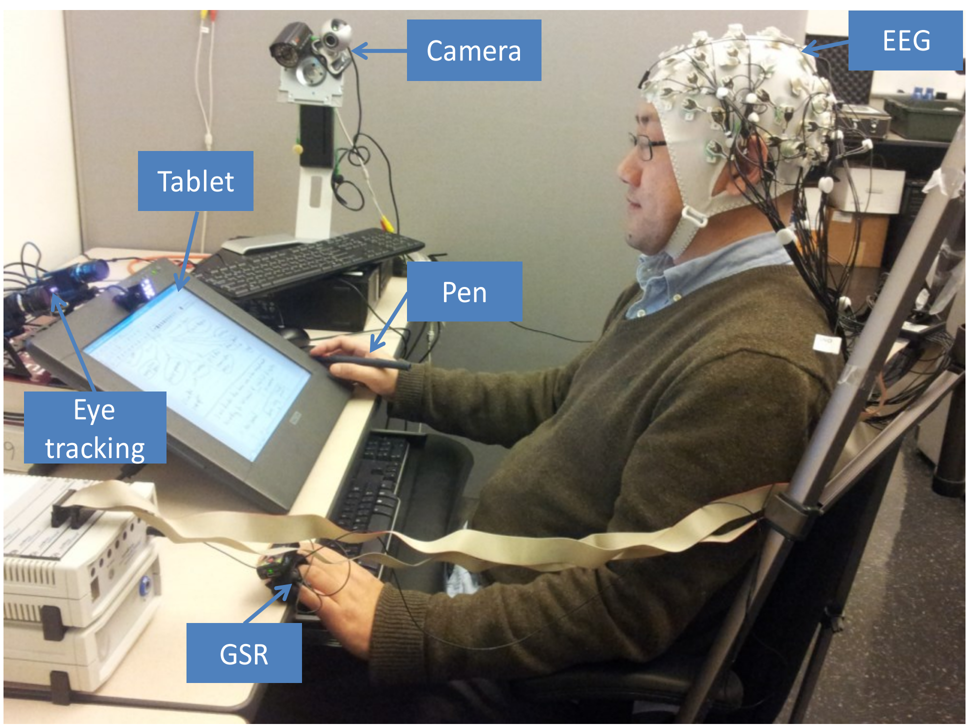

Figure 16. Experiment in progress (Lab, 2006-2013b).

4. Data Analysis and Result

4.1. 2010 experiment data

4.1.1. Data analysis for 2010 experiment data

Only seven subjects are selected for data analysis instead of eleven subjects due to problem with EEG signals (even though we do not analyze EEG signal in this thesis). The HRV data were analyzed by HRVAS which has been introduced in review section (Ramshur, 2010). The inputs of HRVAS were R-R intervals. A few modifications based on the source code of HRVAS were made with the help with another PhD student from our research group (Suo Tan). The purpose of the modifications was to export the LF/HF ratio data into an excel file for easier analysis.

Several steps were conducted to analyze the data:

1. Import R-R interval values into HRVAS, run the program.
2. Export the LF/HF ratio data calculated by wavelet analysis. The results demonstrate the LF/HF ratio every 0.5 seconds.
3. Segment the design process based on designer's activities which can be observed from the tablet screen. The start time and the end time will be recorded. The behaviors and description will also be recorded. Common segments include: write, pause, rest, etc. An example of several segments can be found in Table 3.

Table 3. Example of several segments (Nguyen et al., 2013).

ID	Start Time	End Time	Duration (s)	Description
1	00:00:00.00	00:11:08.646	668.65	He did some prepared work.
2	00:11:08.646	00:13:06.650	118.00	He looked the design question.
3	00:13:06.650	00:13:33.353	26.70	He operated the "program" and opened software.
4	00:13:33.353	00:13:40.165	6.81	He stopped for a while.
5	00:13:40.165	00:13:44.65	4.49	He opened the test page.
6	00:13:44.65	00:13:49.676	5.03	He stopped for a while.
7	00:13:49.676	00:13:50.300	.62	He opened a new page.
8	00:13:50.300	00:13:54.874	4.57	He stopped for a while.
9	00:13:54.874	00:14:02.709	7.83	He changed the colour of the page to yellow.
10	00:14:02.709	00:14:19.717	17.01	He stopped for a while.
11	00:14:19.717	00:14:22.456	2.74	He drew a vertical line.
12	00:14:22.456	00:14:35.871	13.41	He stopped for a while.
13	00:14:35.871	00:14:39.498	3.63	He drew another vertical line parallel the first one.
14	00:14:39.498	00:14:42.416	2.92	He stopped for a while.
15	00:14:42.416	00:14:50.655	8.24	He drew two horizontal lines to connect the vertical lines.

4. Calculate the average LF/HF ratio using Equation 1.

Equation 1. Calculation of average LF/HF ratio.

$$R_A = \frac{\sum_{i=1}^{n}(t_i \times r_i)}{\sum_{i=1}^{n} t_i}$$

In the equation, R_A is the LF/HF ratio for segment A, r_i is LF/HF ratio for duration t_i and n is the number of 0.5-second LF/HF components in segment A. For example, the average LF/HF ratio for segment X in Figure 17 can be calculated by:

$$R_X = \frac{0.5r_1 + 0.5r_2 + 0.5r_3 + 0.2r_4}{0.5 + 0.5 + 0.5 + 0.2}$$

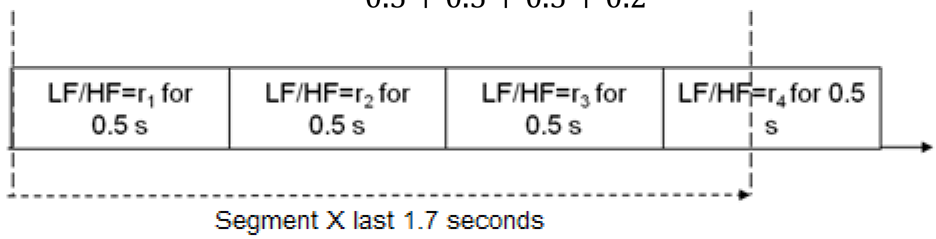

Figure 17. A sample calculation.

5. Cluster the all average LF/HF ratios into seven levels by K-means. K-means was applied twice the number of segments. The calculation was done in Matlab R2011b.

6. Collect the numbers of segments in different LF/HF ratio levels from those seven subjects and use the Friedman test to test if there are differences among different mental stress levels. The Feridedman test was conducted by Matlab built-in function.

7. Use one sided Wilcoxon signed rank test to check if the number of segments at low mental stress level is greater than the number of segments at high mental stress level.

Feridedman test is a non-parametric statistical test developed by the U.S. economist Milton Friedman which is similar to ANOVA. It can be used to detect differences in treatments across multiple test attempts (Friedman, 1937, 1940). This test can be conducted by a Matlab built-in function: $p = friedman(X,reps)$ (The MathWorks, 1994-2014).

One sided Wilcoxon signed rank test is a nonparametric test procedure for the analysis of matched-pair data, based on differences, or for a single sample, it is simple but powerful test for data analysis (Randles, 1988; Woolson, 1998). We use this test because our data are not normally distributed. The statistical analysis was conducted on a website-Social Science Statistics which provides a number of statistical test calculators includes Z-Test Calculator for 2 Population Proportions, Student T-Test Calculator for Comparing 2 Independent Means, Wilcoxon Signed-Rank Test Calculator, etc. The interface of Wilcoxon Signed-Rank Test Calculator can be found in Figure 18 (Stangroom, 2013).

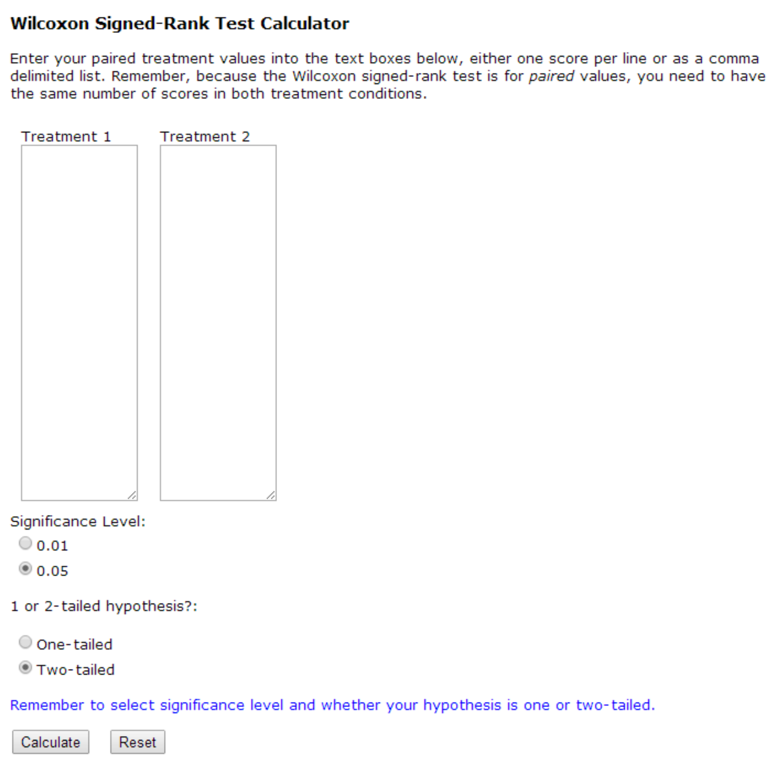

Figure 18. Interface of Wilcoxon Signed-Rank Test Calculator (Stangroom, 2013).

4.1.2. Result for 2010 experiment data

In 2010 design experiment, only seven instead of eleven subjects were selected for the data analysis due to the problem with EEG signals (even though we do not analyze EEG data in this thesis). The numbers of segments at each mental stress level for all seven subjects were calculated. The results are shown in Table 4 and then plotted in Figure 19. The trend of average numbers of segments in different stress levels can be found in Figure 20.

Table 4. Number of segments at different mental stress levels.

STRESS	SUBJECT 1	SUBJECT 2	SUBJECT 3	SUBJECT 4	SUBJECT 5	SUBJECT 6	SUBJECT 7	AVERAGE
LEVEL 1	72	91	95	295	12	51	27	91.85714
LEVEL 2	30	73	53	204	14	70	34	68.28571
LEVEL 3	19	41	37	114	11	47	21	41.42857
LEVEL 4	8	22	15	31	7	21	25	18.42857
LEVEL 5	5	3	6	24	3	15	13	9.857143
LEVEL 6	4	2	6	6	1	2	16	5.285714
LEVEL 7	1	1	2	2	1	1	4	1.714286

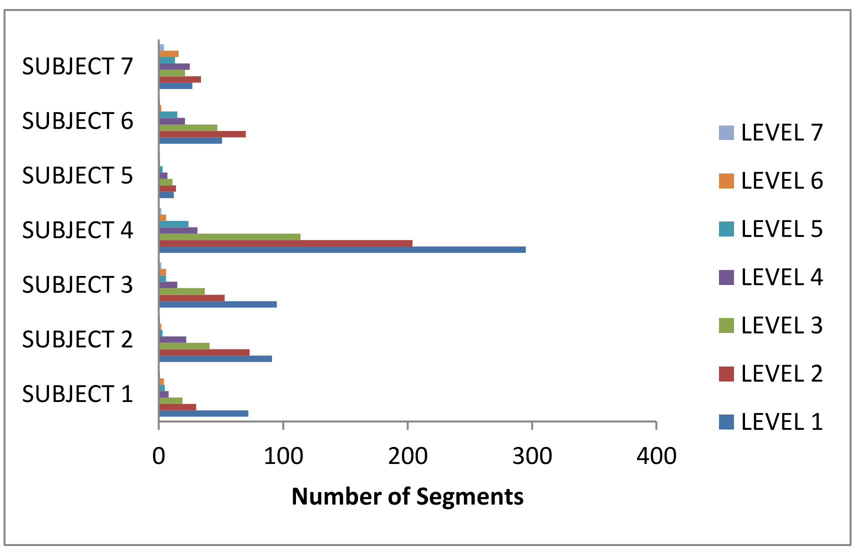

Figure 19. Number of segments at different mental stress levels.

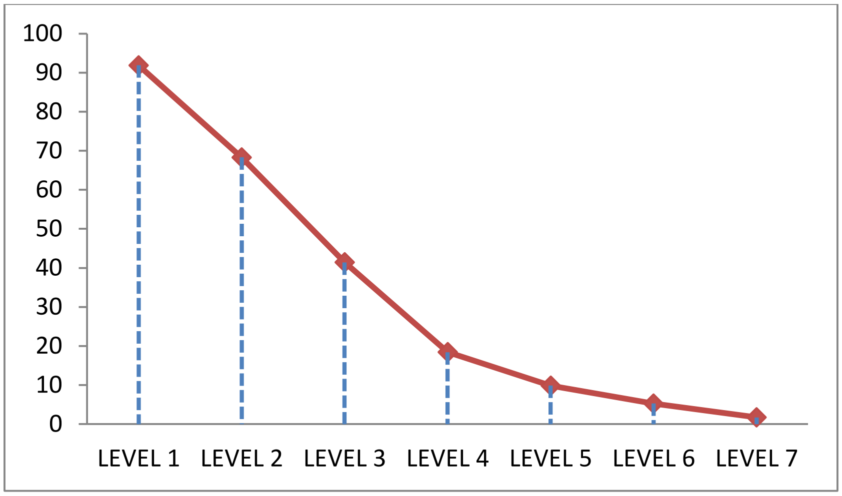

Figure 20. Trend of average numbers of segments in different stress levels.

According to Figure 19 and Figure 20, it can be observed that the number of segments generally decreases as subjects' mental stress level increases. As a result, we made a hypothesis that:

1. There are differences among different mental stress levels.
2. The number of segments at low mental stress level is greater than the number of segments at high mental stress level.

In this thesis, we used Friedman test to confirm the first hypothesis and Wilcoxon signed rank test to confirm the second hypothesis.

The result of Friedman test is shown in Figure 21.

Friedman's ANOVA Table

Source	SS	df	MS	Chi-sq	Prob>Chi-sq
Columns	187.357	6	31.2262	40.35	3.88094e-007
Error	7.643	36	0.2123		
Total	195	48			

Figure 21. Result of Fridman's ANOVA test.

For the second hypothesis, we add data form Table 4 of two different mental stress levels into the website and record the W value.

The result of W value can be found in Table 5 and Figure 22. Result of Wilcoxon signed rank test

Table 5. Result of Wilcoxon signed rank test.

Between Levels	W-Value	Critical W-Value	Between Level	W-Value	Critical W-Value
1,2 (n.s)	7	3	3,4	1.5	3
1,3	0	3	3,5	0	3
1,4	0	3	3,6	0	3
1,5	0	3	3,7	0	3
1,6	0	3	4,5	0	3
1,7	0	3	4,6	0	3
2,3	0	3	4,7	0	3
2,4	0	3	5,6 (n.s)	4	2
2,5	0	3	5,7	0	3
2,6	0	3	6,7	0	2
2,7	0	3			

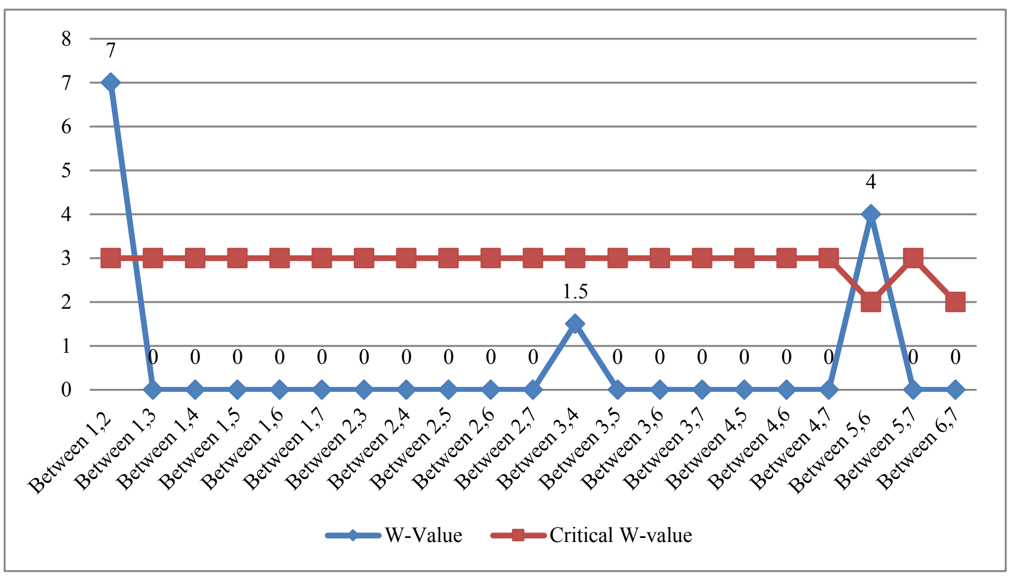

Figure 22. Result of Wilcoxon signed rank test.

In Table 5, most of the critical W value of samples is 3 except for stress level 5, 6 and 6, 7 because one sample is deleted in those two samples due to calculation in Wilcoxon test. The critical W-value for $n = 7$ and $\alpha = 0.05$ is 3, and for $n = 6$, the critical value is 2. As a result, most stress levels are different between each other except for level 1, 2 and 5, 6. An example of design activities at stress levels 6 and 7 is displayed in Table 6.

Table 6. Example of design activities at stress level 6 and 7.

	Simplified description	Level		Simplified description	Level
Subject 1	Pause	6	Subject5	Write	6
	Erase	6		Try to delete a word	7
	Modify sentences	6	Subject 6	Pause	6
	Pause	7		Scrub nose	6
	Fix a grammar mistake	6		Go back to the design solution, move slightly	7
Subject 2	Pause	6	Subject 7	Draw	6
	Erase	6		Move slightly	6
	Pause	7		Write	6
Subject 3	Draw	6		Move slightly	7
	Pause	6		Move slightly	7
	Pause	6		Write	7
	Fix a grammar mistake	6		Write	6
	Browse the internet	6		Move slightly	6
	Close an application	6		Move slightly	6
	Browse the internet	7		Write	6
	Choose an online article and read	7		Erase	6
Subject 4	Pause	6		Erase, scroll up	6
	Open a new page	6		Pause	6
	Try to erase the figure quickly	7		Scroll down	6
	Erase	6		Write	6
	Go to the "side view" page	6		Write	6
	Erase	6		Draw	6
	Go back to the question	7		Pause, move slightly	6
	Go back to the design page	6		Scroll down	6
				Draw	7

In this analysis, I helped Nguyen to compute segment for LF/HF, through which I have understood the principle of the Friedman test and One Sided Wilcoxon Signed Rank Test.

This lays a foundation for the experimental analysis conducted in 2013 (Nguyen et al., 2013).

4.2. 2013 experiment data

4.2.1. Data analysis for 2013 experiment data

Five subjects' data were excluded from data analysis from those 23 subjects. The reasons were: one subject was interrupted by emergency fire alarm, two subjects' data were not recorded correctly and the other two subjects only completed four instead of six design tasks.

In experiment design section, subjects needed to evaluate their workload after each part of each design task. In our research, we considered those workloads evaluated by subjects as their mental workload level. Also, most subjects felt uncomfortable at the beginning of the experiment, so the subjects' mental workload might be high due to being nervous rather than the difficulty of design problem. Therefore, we eliminated the GSR data of the first design task. Then for each individual, we ranked the rated workload from one to five according to subjects' workload evaluation as illustrated in Figure 14. Workload evaluation. Design problems are ranked accordingly. One sample data of a subject can be found in Table 7.

Table 7. Problems ranked according to rated workload level and corresponding GSR data of a subject.

Workload level	Problems	Workload	Mean GSR (problem understanding)
1	3	36	5.161659292
2	2	43	5.95566064
3	6	47	6.504022005
4	5	64	6.381673945
5	4	67	5.552206851

The GSR data were recorded by frequency of 32 Hz which means that we had 32 GSR values per second. The whole experiment was segmented in to about 50 parts according to design activities that subjects performed. Common activities include: start relaxing, end relaxing, read problem, start solution, end solution, workload evaluation, evaluation, etc. Then, GSR data were imported into a Matlab based program (the source code of the program can be found in Appendix) to collect GSR data corresponding to each activity developed with the help of Nguyen. Then we calculated five activities' corresponding average GSR values: resting (before task), problem understanding, solution expression, evaluation, and resting (after task). The results of solution expression were deleted because when we conducted experiments, we saw many subjects writing down their solution right after they read the problem. So, we assume when reading the problem, they may have some ideas (or generate a solution) in mind. The next stage is just to write it down. Therefore mental workload is mainly from the previous stage (problem understanding). However, this assumption has not been confirmed yet. Therefore, additional data analysis on mean GSR data of solution expression will be considered as a future work. One sample GSR data in one experiment is shown in Figure 23 which illustrates the GSR change of the subject from the beginning of the experiment until the end.

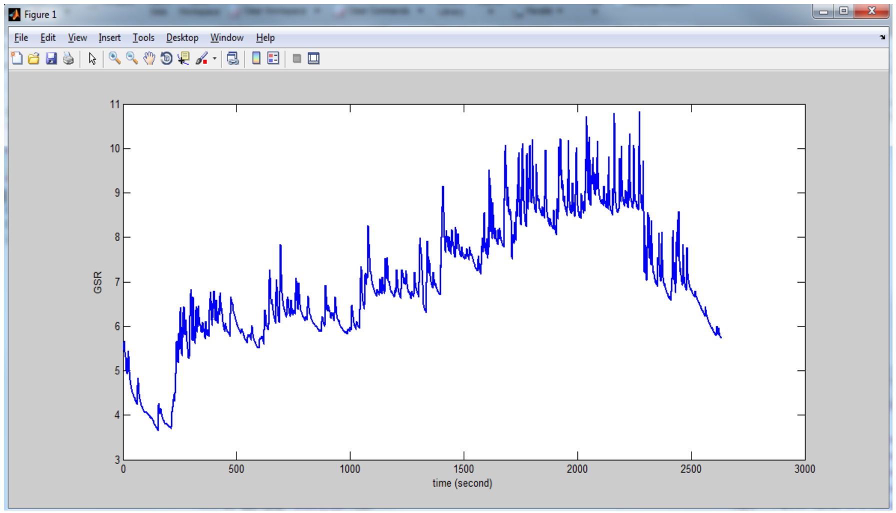

Figure 23. Sample GSR data.

Table 8. Mean GSR data in different levels of problem understanding and evaluation.

	Problem understanding					Evaluation				
Levels Subjects	1	2	3	4	5	1	2	3	4	5
1	0.259841	0.25	0.242848	0.259948	0.295249	0.24899	0.20328	0.28	0.24991	0.21917
2	1.648624	1.848919	2.241191	2.473393	1.567014	1.5827	0.80361	1.21985	1.20551	1.74147
3	0.539402	0.537476	0.505936	0.662063	0.579051	1.24215	1.32792	1.83266	1.41319	1.30954
4	2.050952	0.951671	1.762141	1.155493	1.062796	1.94022	1.75754	1.2715	1.31931	1.94383
5	2.022995	2.581881	1.654634	2.092898	1.731226	6.88629	6.1033	9.16762	7.99809	8.48912
6	2.871138	1.902992	2.780355	3.008828	3.209854	17.4592	18.0277	16.2367	15.2964	17.6724
7	3.779974	3.972972	3.113684	3.077174	3.628596	5.81872	4.99069	6.67018	6.1213	6.19438
8	1.995414	4.106184	2.52106	2.584797	2.101583	1.79016	2.05723	1.69226	2.53824	1.84092
9	1.1549	1.454947	1.472584	1.41194	1.22184	0.45885	0.45399	0.457	0.44871	0.49906
10	1.769064	1.865314	1.853888	1.940262	1.788349	2.68265	1.5203	1.51275	1.82099	2.53068
11	0.346601	0.359527	0.455382	0.468557	0.394506	3.06968	3.41093	2.44433	3.54173	2.97312
12	4.223428	4.638694	4.356514	4.26965	4.428406	2.01792	2.55189	2.595	2.18765	1.90285
13	0.805472	0.554811	0.673446	0.43486	0.813136	1.08684	1.5415	1.23504	1.3812	1.34727
14	0.50443	0.623551	0.590673	0.449012	0.879537	1.76163	1.81714	1.86954	1.78357	1.75465
15	1.39099	1.861202	1.784014	1.799733	2.039522	0.35	0.38	0.35818	0.41	0.5285
16	6.994717	6.624414	8.888702	7.023592	9.252866	2.9156	4.04769	3.04344	3.84339	3.51998
17	16.01455	19.33072	19.18364	18.78864	17.95556	0.32303	0.83552	0.69867	0.86146	0.46334
18	5.161659	5.955661	6.504022	6.381674	5.552207	0.28653	0.34204	0.42912	0.31696	0.33528

Mental workload levels (evaluated by subjects themselves) for problem understanding and evaluation are calculated and ranked. The result of mental workload levels can be found in the Appendix. Mean GSR in each mental workload level was also calculated. The result of mean GSR data in different levels of problem understanding and evaluation is illustrated in Table 8.

After those average GSR data were calculated, with Nguyen's instruction and our discussion, we decided to compare average GSR data:

- Between two design tasks at two difficulty levels.
- Between design state and resting state.
- Between pre-test and post-test resting.

It is noticed that we only compared data between resting state and its closest design state because we assume that this comparison is a good sample which can represent the overall trend between resting and design. In particular, we compared pre-test resting with design task No. 2, and we compared post-test resting with design task 6. The mean GSR of pre-test, post-test, and problem understanding of design problem 2 and 6 are illustrated in Table 9.

Mean GSR between resting before test (pre-test resting) and resting after test (post-test resting) was compared to see if the test had left any residual effect on the subject.

Since the average data we computed were not normally distributed, we used One Sided Wilcoxon Signed-Ranks Test to test the difference with the same website we used for analyzing 2010 experiment data. For example, if we want to compare whether GSR is different when designer is under resting (before design) and understanding problem 2.

We just copy the first two columns of Table 9 into the website (Stangroom, 2013) and collect the W-value to see if there exists significance between those two set of data. The calculation can be found in Figure 24. According to the result, it can be found that the W-value is 10. The critical value of W for $N = 18$ at $p \leq 0.05$ is 47. Therefore, the result is significant at $p \leq 0.05$. As a result, we may conclude that there exists difference in mean GSR when designer change from pre-test resting stage to understanding problem 2. Similar work will also conducted on comparing mean GSR data between two design tasks at two difficulty levels, design state and resting state, pre-test and post-test resting. And W-value will also be recorded when doing statistical analysis.

Therefore, the procedure for analyzing 2013 experiment data include:

1. Calculate workload of problem understanding and evaluation process for five design problems (the first design problem was deleted) and rank them according to their values to identify the workload level for each design problem of each subject.
2. Import GSR data into the Matlab based program to collect GSR data corresponding to each activity.
3. Collect the GSR data of pre-test resting, understanding five problems, evaluating five problems, post-test resting and calculate the average of those data.
4. Compare the difference of mean GSR two design tasks at two difficulty levels with one sided Wilcoxon Signed-Rank test, mean GSR value of problem understanding process and evaluating process will be compared respectively.

5. Compare the difference of mean GSR between pre-test resting with design task 2, post-test resting with design task 6 and pre-test resting with post-test with one sided Wilcoxon Signed-Rank test.

Table 9. Mean GSR of pre-test, post-test, and problem understanding of design problems 2 and 6.

Pre-test resting and design activities			Post-test resting and design activities		
resting (before design)	understanding problem 2	evaluating problem 2	resting (after design)	understanding problem 6	evaluating problem 6
1.223837488	2.050952381	1.58269697	1.086225522	1.062796385	1.741467269
1.469965066	1.902991597	1.520298165	1.623717927	2.780354767	1.512752404
0.543234113	2.581881259	1.327922374	1.484347467	1.731225937	1.413190184
0.310750536	0.29524911	0.28	0.23024549	0.242847731	0.203275862
0.951606931	1.648624277	1.840922022	1.853073188	2.241191196	2.53824263
1.437498403	3.628595875	2.444324918	2.812006412	3.972972425	3.541725959
4.175517301	4.106184388	2.551884913	2.030050853	1.995414414	1.902845835
0.810653336	1.22184	1.347273101	1.036577793	1.154899536	1.235040798
1.450705664	1.769063559	1.761634024	1.722367767	1.940262351	1.754649549
0.353303324	0.346601256	0.358177613	0.390789474	0.394505929	0.38
1.586248498	4.428405538	3.519979178	3.77349198	4.223427734	3.043435341
0.33999308	0.805472222	0.861458253	0.410546646	0.434860215	0.463340081
0.232172823	0.504429825	0.286532553	0.460086869	0.879537246	0.335274889
1.131893378	1.78401421	1.75753937	2.032664055	2.03952183	1.319305474
4.101611333	6.624413747	6.103294973	6.523044894	9.252866305	8.489114704
16.01246512	18.78863914	17.67235596	16.35283235	16.0145512	15.29634672
3.901399276	5.95566064	5.818718962	5.524531175	6.504022005	6.121301183
0.53275447	0.50593599	0.448711019	0.418801564	0.539401709	0.45884979

Success!

Explanation of results

We have calculated both a W-value and Z-value. If the size of *N* is at least 20 - see the Results Details box - then the distribution of the Wilcoxon W statistic tends to form a normal distribution. This means you can use the Z-value to evaluate your hypothesis. If, on the other hand, the size of *N* is low, and particularly if it's below 10, you should use the W-value to evaluate your hypothesis.

You should also note that if a subject's difference score is zero - that is, if a subject has the same score in both treatment conditions - then the test discards the individual from the analysis and reduces the sample size. If you have a lot of ties, this procedure will undermine the reliability of the test (and also suggests that the requirement that the data is continuous has not been met).

Treatment 1	Treatment 2	*Sign*	*Abs*	*R*	*Sign R*
1.223837488	2.050952381	-1	0.8271	12	-12
1.469965066	1.902991597	-1	0.433	8	-8
0.543234113	2.581881259	-1	2.0386	13	-13
0.310750536	0.29524911	1	0.0155	2	2
0.951606931	1.648624277	-1	0.697	11	-11
1.437498403	3.628595875	-1	2.1911	15	-15
4.175517301	4.106184388	1	0.0693	4	4
0.810653336	1.22184	-1	0.4112	7	-7
1.450705664	1.769063559	-1	0.3184	6	-6
0.353303324	0.346601256	1	0.0067	1	1
1.586248498	4.428405538	-1	2.8422	18	-18
0.33999308	0.805472222	-1	0.4655	9	-9
0.232172823	0.504429825	-1	0.2723	5	-5
1.131893378	1.78401421	-1	0.6521	10	-10
4.101611333	6.624413747	-1	2.5228	16	-16
16.01246512	18.78863914	-1	2.7762	17	-17
3.901399276	5.95566064	-1	2.0543	14	-14
0.53275447	0.50593599	1	0.0268	3	3

Significance Level:

- 0.01
- 0.05 (selected)

1 or 2-tailed hypothesis?:

- One-tailed (selected)
- Two-tailed

<u>Result Details</u>

W-value: 10
Mean Difference: 0.35
Sum of pos. ranks: 10
Sum of neg. ranks: 161

Z-value: -3.2881
Mean (*W*): 85.5
Standard Deviation (*W*): 22.96

Sample Size (*N*): 18

Result 1 - Z-value

The Z-value is -3.2881. The p-value is 0.0005. The result is significant at $p \leq 0.05$.

Result 2 - W-value

The W-value is 10. The critical value of *W* for *N* = 18 at $p \leq 0.05$ is 47. Therefore, the result is significant at $p \leq 0.05$.

Figure 24. Sample Wilcoxon Signed-Rank Test calculation.

4.2.2. Result for 2013 experiment data

The W values and critical W values for the GSR between different levels in problem understanding and evaluation can be found in Figure 25and Figure 26

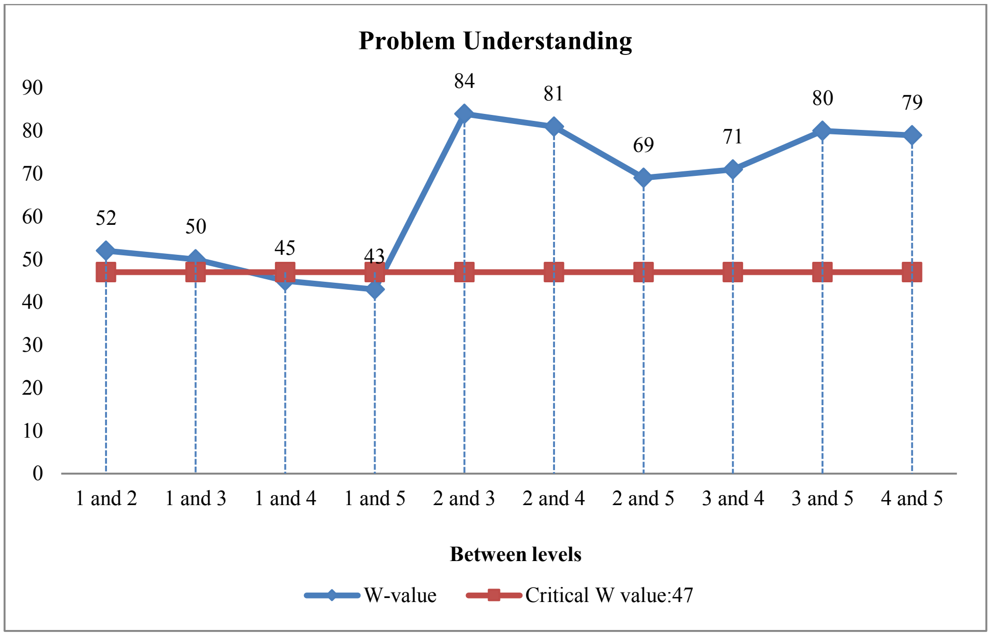

Figure 25. W-Value (problem understanding).

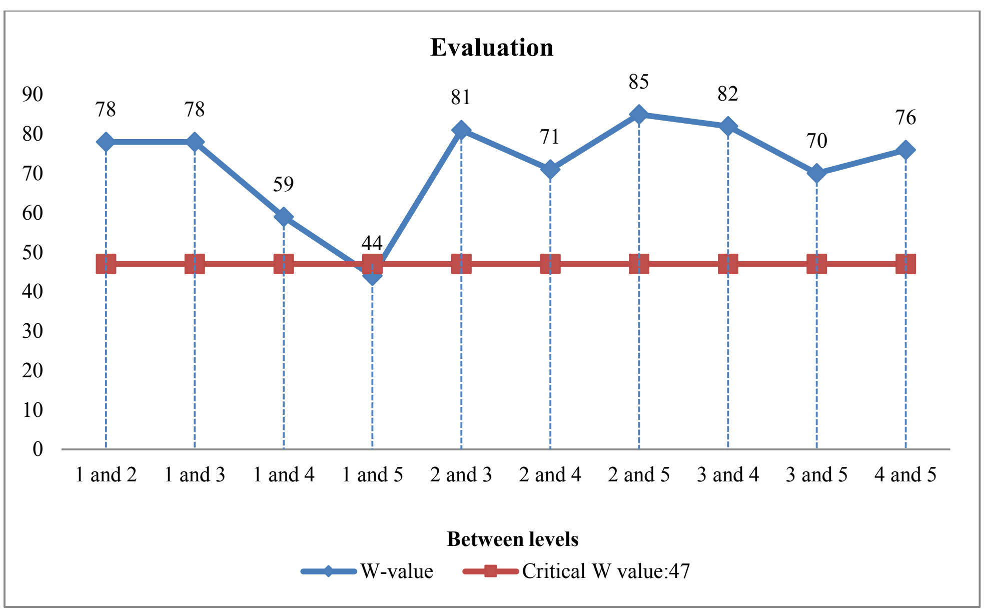

Figure 26. W-value (evaluation).

According to Figure 25, it can be observed that in problem understanding part, mean GSR of workload level 1 is only significantly lower than mean GSR of workload level 4 and workload level 5. As for Figure 26, it illustrated that only mean GSR of workload level 1 is significantly lower than mean GSR of workload level 5. No more significant differences can be found between other mental workload levels.

The comparison of mean GSR between resting state and design state is demonstrated in Figure 27. The GSR data illustrated in the figure are already normalized for better comparison. From this figure, it can be observed that mean GSR is generally lower when designers are resting compared to the value when they are understanding problems or evaluating others' designs. Additionally, mean GSR of pre-test resting is generally lower than mean GSR of post-test resting.

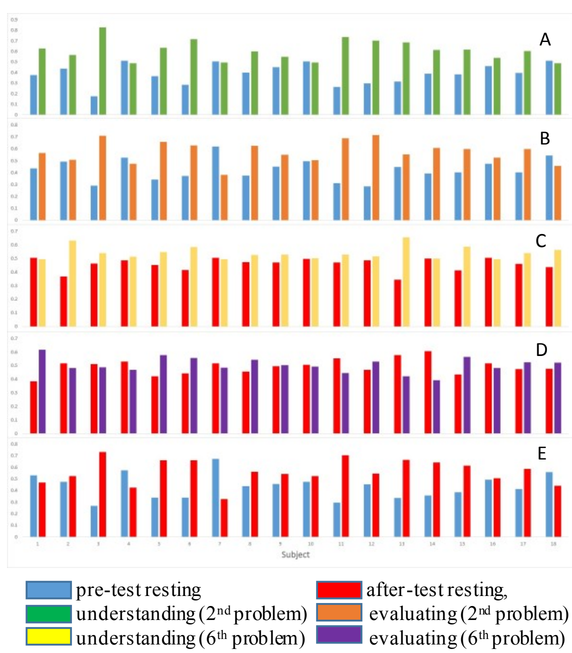

Figure 27. Comparison of average GSR data between resting state and different design states, resting state before experiment and resting state after experiment.

To confirm our observation, another Wilcoxon test was conducted. The result of the test can be found in Figure 28.

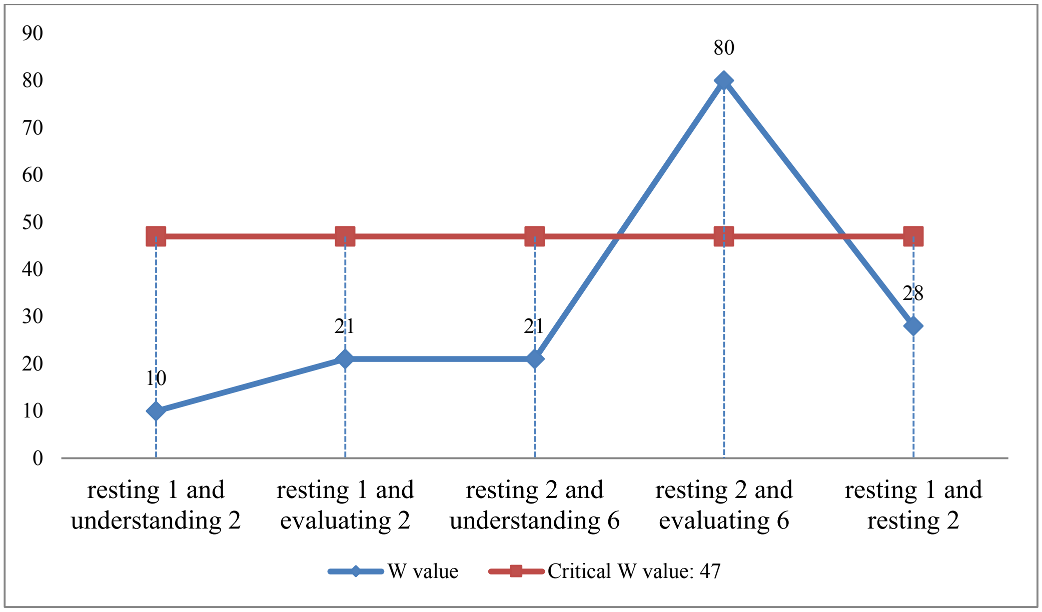

Figure 28. W-value (between resting and different design activities).

From Figure 28, it can be observed that only mean GSR between resting after design and evaluating the sixth problem showed non-significant. As a result, we may conclude that: the mean resting GSR before design task is significantly lower to mean resting GSR after design task, resting GSR is generally lower than designing GSR.

The experiment in 2013 was designed by Nguyen in our research group and I help her to conduct the experiment with our research group colleagues. When I was analyzing 2013 experiment data, Nguyen gave me a lot of suggestions on statistical analysis methods selection, data processing, programming, etc. which laid the foundation of success of the analysis (Xu et al., 2014).

5. Discussion and Conclusion

5.1. Discussion

According to the result of 2010 experiment data, in Friedman test (as shown in Figure 21), P-value (the probability >Chi-square) equals to $3.88*10^{-7}$ which is much lower than 0.05. Therefore, we may conclude that among those seven sets of data, there exist significant differences. As for the Wilcoxon test, the results showed that other except for level 1and 2, and level 5 and 6, number of design activities between other two different mental stress levels can be found significantly different. This can confirm our assumption that the number of design activities generally decreased with the increase of stress levels. This may imply that the higher mental stress is; the fewer design activities are performed. According to Table 6, we found that there seem few correlations between types of design activities and high mental stress levels.

As for 2013 experiment data, in this thesis, we are investigating if GSR is a reliable index of mental workload. The results showed that GSR is higher in design process comparing to GSR during resting. In addition, the average GSR of resting before design tasks is lower than the value of resting after design task. Finally, we found that GSR is only significantly different between very low and very high mental workload.

Compared to the result from (Berguer et al., 2001; Perala & Sterling, 2007; Shi et al., 2007), our results show slightly different. Their result illustrated the GSR increased as cognitive load/workload increased. However, our results demonstrated that only when difficulty levels of the design problems increased significantly, or when the subjects transit from resting phase to design phase, GSR increased. There might be three reasons:

- The difficulty levels of design problem in our experiment are not very different.
- When people are performing design activities, their GSR change might be different from when they are doing other activities.
- As reported in (Conway et al., 2013; Setz et al., 2010), in stress condition, mean GSR cannot differentiate among levels of workload. The subjects in the experiment experienced a certain level of stress (revealed in their rating after each task).

5.2. Conclusion

In this thesis, a comprehensive review on design creativity, mental stress, GSR, and HRV was conducted. After the review of design creativity and mental stress, an inverse U shape relation between mental stress and design creativity was observed. In addition to that, we have found that in design experiment, designer's mental stress generally increase as his/her mental workload increased. For the correlation between HRV and mental stress, a series of HRV parameters (mHR, mean RR, SDRR, SDHR, LF/HF ratio, etc.) were identified which may be affected by mental stress (Boonnithi & Phongsuphap, 2011; Hjortskov et al., 2004). It is also found that LF/HF ratio which is calculated by HRV is a very sensitive parameter which will increase with mental stress level increases (Petkar, 2011). Two HRV analysis software were identified to calculate HRV parameters and HRVAS was selected as our design experiment data analysis software (Niskanen et al., 2004; Ramshur, 2010). The review for the relation between GSR and mental stress/workload showed that mean GSR generally increases when the person is under stress (Shi et al., 2007), mean GSR was higher in complex tasks than in baseline, but GSR could not differentiate between task difficulties (Engström et al., 2005). However, some researchers found that GSR was higher at easy level compared to hard level. They interpreted that the reason might be the subjects were bored when they were doing easy tasks (Ikehara & Crosby, 2005).

As LF/HF ratio has already been confirmed to be a reliable indicator of mental stress, an analysis on identifying the relation between design activities and mental stress level was conducted with HRV data gathered from 2010 experiment data. The data was then segmented based on designer's activities. For each segment, LF/HF ratio was computed and then clustered into seven levels using k-means.

According to the results, it was found that:

1. Most of the activities in a design process were performed under low levels of mental stress.
2. The number of design activities decreases when the level of mental stress increases.
3. No correlation between types of design activities and mental stress level was found.

To find the reliability of mean GSR as an indicator of mental workload, another analysis on GSR data for 2013 experiment was conducted. Mean GSR for each activity (pre-test resting, understanding the design problem, evaluating other designers' solution for the same design problem, etc.) was calculated. Workload for understanding and evaluating for each subject was calculated and compared. Comparison of mean GSR between two different workload level activities, between resting and design activities was conducted and tested using one tail Wilcoxon Signed-Rank Test to see if there existed significant difference between mean GSR values during different activities.

According to the result, we conclude that:

1. Mean GSR can be a reliable measure of mental workload when designers have a great change in their mental workload.
2. Mean GSR is not very sensitive to small workload change when stress is involved. It is generally lower in resting phase compared to design phase. This confirms the results of previous studies.

6. Constrains and Future Work

This thesis has following constrains:

- In 2010 design experiment:
    - Subjects' data for analysis were limited.
    - The experiment procedure was too simple.
    - Only HRV data were analyzed.
- In 2013 design experiment:
    - The difficulty levels between design tasks were not significant.
    - Subjects' data for analysis were limited.
    - Only GSR data were analyzed.
    - GSR data of solution expression were not analyzed.

Additionally, only individual's mental stress was investigated, further research can be conducted on exploring the relation between stress of a company and its creativity.

As a result, the future work would include:

- Analyze other physiological signals which can be regarded as an index of mental stress/mental workload such as EEG (Electroencephalography), EOG (eye movements), and respiration rate, etc.
- Improve our design problems to make the difficulty between design tasks more significant.
- Conduct more experiments to improve the reliability of statistical analysis.
- Analyze HRV, respiration, EEG result of 2013 experiment data; compare the result to GSR result.

- Compare the difference of mean GSR data when subjects were expressing their solutions between two mental workload levels, and difference between problem understanding and solution expression on the same design task.
- Find ways to quantify companies' stress, and find the relation between stress and creativity in companies.

Publications

Nguyen, Thanh An, **Xu, Xu**, & Zeng, Yong. (2013). *Distribution of mental stresses during conceptual design activities.* Paper presented at the DS 75-7: Proceedings of the 19th International Conference on Engineering Design (ICED13), Design for Harmonies, Vol. 7: Human Behaviour in Design, Seoul, Korea, 19-22.08. 2013.

Xu, Xu, Nguyen, Thanh An, & Zeng, Yong. (2014). *Galvanic Skin Response as Index of Mental Workload in Design Activities*. Paper presented at the TMCE, Budapest, Hungary.

Xu, Xu, & Zeng, Yong. (2014). *Mitigation of the Risking Impacts on Whitemud Creek from 30th Avenue Storm Trunk Overflows Using Environment Based Design*, Project Report.

References

Arciszewski, Tomasz, Michalski, Ryszard S, & Wnek, Janusz. (1995). *Constructive induction: the key to design creativity.* Paper presented at the Preprints of the Third International Round-Table Conference on Computational Models of Creative Design, Heron Island, Australia.

Berguer, R, Smith, WD, & Chung, YH. (2001). Performing laparoscopic surgery is significantly more stressful for the surgeon than open surgery. *Surgical endoscopy, 15*(10), 1204-1207.

Boonnithi, Sansanee, & Phongsuphap, Sukanya. (2011). *Comparison of Heart Rate Variability measures for mental stress detection.* Paper presented at the Computing in Cardiology, 2011.

Conway, Dan, Dick, Ian, Li, Zhidong, Wang, Yang, & Chen, Fang. (2013). The effect of stress on cognitive load measurement. In P. Kotzé, G. Marsden, G. Lindgaard, J. Wesson & M. Winckler (Eds.), *Human-Computer Interaction – INTERACT 2013* (Vol. 8120, pp. 659-666): Springer Berlin Heidelberg.

Dorst, Kees, & Cross, Nigel. (2001). Creativity in the design process: co-evolution of problem–solution. *Design studies, 22*(5), 425-437.

Engström, Johan, Johansson, Emma, & Östlund, Joakim. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 8*(2), 97-120.

Filippi, Stefano, & Motyl, Barbara (2010). *Synergies between systematic innovation and interaction design for product development.* Paper presented at the TMCE.

Friedman, Milton. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association, 32*(200), 675-701.

Friedman, Milton. (1940). A comparison of alternative tests of significance for the problem of m rankings. 86-92. doi: 10.1214/aoms/1177731944

Goldschmidt, Gabriela, & Tatsa, Dan. (2005). How good are good ideas? Correlates of design creativity. *Design Studies, 26*(6), 593-611. doi: http://dx.doi.org/10.1016/j.destud.2005.02.004

Graessler, Iris , Haas, Volker , & Suchowerskyj, Wadym (2012). *Innovation based on applying design methodology.* Paper presented at the TMCE.

Greene, Richard Tabor (2012). *The culture work of innovating: Getting real about innovating in business 27 ways.* Paper presented at the TMCE.

Hart, Sandra G, & Staveland, Lowell E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload, 1*(3), 139-183.

Hennessey, Beth A. (2003). The social psychology of creativity. *Scandinavian Journal of Educational Research, 47*(3), 253-271.

Hjortskov, N., Rissen, D., Blangsted, A. K., Fallentin, N., Lundberg, U., & Sogaard, K. (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. doi: 10.1007/s00421-004-1055-z

Ikehara, C. S., & Crosby, M. E. (2005, 03-06 Jan. 2005). *Assessing cognitive load with physiological sensors.* Paper presented at the System Sciences, 2005. HICSS '05. Proceedings of the 38th Annual Hawaii International Conference on.

Karthikeyan, P., Murugappan, M., & Yaacob, S. (2011, 4-6 March 2011). *A review on stress inducement stimuli for assessing human stress using physiological signals.* Paper presented at the Signal Processing and its Applications (CSPA), 2011 IEEE 7th International Colloquium on.

Facilities of Design Lab. from http://users.encs.concordia.ca/~design/facility.html

Research Area of Design Lab. from http://users.encs.concordia.ca/~design/research.html

Lee, DaeSung, Jo, NamYong, & Lee, KunChang. (2011). A physiological approach to creativity under stress and non-stress conditions. In T.-h. Kim, H. Adeli, J. Ma, W.-c. Fang, B.-H. Kang, B. Park, F. Sandnes & K. Lee (Eds.), *U- and E-Service, Science and Technology* (Vol. 264, pp. 197-206): Springer Berlin Heidelberg.

Li, Yan, Wang, Jian, Li, Xianglong, & Zhao, Wu. (2007). Design creativity in product innovation. *The international journal of advanced manufacturing technology, 33*(3-4), 213-222.

Luximon, Ameersing, & Luximon, Yan (2010). *Creativity, inspiration and design exploration through modification of the juicy salif.* Paper presented at the TMCE.

McCraty, Rollin, Atkinson, Mike, Tiller, William A, Rein, Glen, & Watkins, Alan D. (1995). The effects of emotions on short-term power spectrum analysis of Heart Rate Variability. *The American journal of cardiology, 76*(14), 1089-1093.

Michon, JA. (1966). Tapping regularity as a measure of perceptual motor load. *Ergonomics, 9*(5), 401-412.

Miller, Lyle H, & Shmavonian, Barry M. (1965). Replicability of two GSR indices as a function of stress and cognitive activity. *Journal of personality and social psychology, 2*(5), 753.

Nguyen, Thanh An, Xu, Xu, & Zeng, Yong. (2013). *Distribution of mental stresses during conceptual design activities.* Paper presented at the DS 75-7: Proceedings of the 19th International Conference on Engineering Design (ICED13), Design for Harmonies, Vol. 7: Human Behaviour in Design, Seoul, Korea, 19-22.08. 2013.

Nguyen, Thanh An, & Zeng, Yong. (2012). A theoretical model of design creativity: nonlinear design dynamics and mental stress-creativity relation. *Journal of Integrated Design and Process Science, 16*(3), 65-88.

Nguyen, Thanh An, & Zeng, Yong. (2013). A physiological study of relationship between designer's mental effort and mental stress during conceptual design. *Computer-Aided Design*(0). doi: http://dx.doi.org/10.1016/j.cad.2013.10.002

Niskanen, Juha-Pekka, Tarvainen, Mika P, Ranta-Aho, Perttu O, & Karjalainen, Pasi A. (2004). Software for advanced HRV analysis. *Computer methods and programs in biomedicine, 76*(1), 73-81.

Nourbakhsh, Nargess, Wang, Yang, Chen, Fang, & Calvo, Rafael A. (2012). *Using Galvanic Skin Response for cognitive load measurement in arithmetic and reading tasks.* Paper presented at the Proceedings of the 24th Australian Computer-Human Interaction Conference.

Perala, Chuck H, & Sterling, Bruce S. (2007). Galvanic Skin Response as a measure of soldier stress: DTIC Document.

Petkar, Harshad. (2011). *Effects of working memory demand on performance and mental stress during the Stroop task.* Concordia University.

Polverini, Davide, Graziosi, Serena, & Mandorli, Ferruccio. (2011). A step-based framework to combine creativity, project management and technical development in industrial innovation. *International Journal of Product Development, 14*(1), 96-117.

Ramshur, John T. (2010). *Design, evaluation, and application of Heart Rate Variability Analysis Software (HRVAS).* (Master), The University of Memphis.

Randles, Ronald H. (1988). Wilcoxon signed rank test. *Encyclopedia of statistical sciences*.

Roscoe, A. H. (1992). Assessing pilot workload. Why measure heart rate, HRV and respiration? *Biological Psychology, 34*(2–3), 259-287. doi: http://dx.doi.org/10.1016/0301-0511(92)90018-P

Roth-Koch, Sabine , & Westkaemper, Engelbert (2010). *3-D shapes out of 2-D freehand sketches: Creative spatial interpretation of sketches.* Paper presented at the TMCE.

Ryu, K., & Myung, R. (2005). Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. *International Journal of Industrial Ergonomics, 35*(11), 991-1009. doi: DOI 10.1016/j.ergon.2005.04.005

Sedenkov, Vladimir (2012). *Design theory design: A new attempt.* Paper presented at the TMCE.

Selye, Hans. (1956). The stress of life.

Setz, Cornelia, Arnrich, Bert, Schumm, Johannes, La Marca, Roberto, Troster, G., & Ehlert, Ulrike. (2010). Discriminating stress from cognitive load using a wearable EDA device. *Information Technology in Biomedicine, IEEE Transactions on, 14*(2), 410-417.

Shi, Yu, Ruiz, Natalie, Taib, Ronnie, Choi, Eric, & Chen, Fang. (2007). *Galvanic Skin Response (GSR) as an index of cognitive load.* Paper presented at the CHI'07 extended abstracts on Human factors in computing systems.

Stangroom, Jeremy. (2013). Social Science Statistics. from http://www.socscistatistics.com/Default.aspx

Sternberg, Robert J, & Lubart, Todd I. (1995). *Defying the crowd: Cultivating creativity in a culture of conformity*: Free Press.

Sternberg, Robert J, & Lubart, Todd I. (1999). The concept of creativity: Prospects and paradigms. *Handbook of creativity, 1*, 3-15.

Stroop, J Ridley. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology, 18*(6), 643.

Tang, Yao. (2009). *Evaluation of mental stress during cognitive activities through HRV data and kinesics study.* (Master's Thesis), Concordia University, Montreal, Canada.

Tang, Yao, & Zeng, Yong. (2009). *Quantifying designer's mental stress in the conceptual design process using Kinesics study*. Paper presented at the ICED 09.

TEA. (2010). CAPTIV-L7000. from http://teaergo.com/docs/CAPTIV/L7000/EN/CAPTIV%20L7000_EN_v5.pdf

The MathWorks, Inc. . (1994-2014). Friedman's test - MATLAB friedman - MathWorks. from http://www.mathworks.com/help/stats/friedman.html

Wang, Yuxin, Liu, Zhijuan, Zhang, Xuguang, & Duffy, Alex. (2010). *Implementation and representation of creativity enhanced FBS model based on quotient space theory.* Paper presented at the TMCE.

Wijsman, Jacqueline, Grundlehner, Bernard, Liu, Hao, Hermens, Hermie, & Penders, Julien. (2011). *Towards mental stress detection using wearable physiological sensors.* Paper presented at the Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE.

Wilson, Glenn F. (2002). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *The International Journal of Aviation Psychology, 12*(1), 3-18. doi: 10.1207/S15327108IJAP1201_2

Woolson, RF. (1998). Wilcoxon Signed-Rank Test. *Wiley Encyclopedia of Clinical Trials*.

Xu, Xu, Nguyen, Thanh An, & Zeng, Yong. (2014). *Galvanic Skin Response as index of mental workload in design activities*. Paper presented at the TMCE, Budapest, Hungary.

Zhang, Qiang, Deniaud, Ioana , Caillaud, Emmanuel , & Baron, Claude (2012). *Analysis and model of systematic innovation for design.* Paper presented at the TMCE.

Appendix

Matlab code for calculating GSR data

```
col1 = m(:,1);

col2 = m(:,2);

col3 = m(:,3);

n=find (col3>=0);

col3 = col3(2:length(n));

start_index = 1;

for i=1:length(col3)

result1 = abs(col2 - col3(i));

min_val = min(result1);

end_index = find(result1 == min_val);

A = col1(start_index:end_index);

filename = ['result' int2str(i) '.csv'];

dlmwrite(filename, A);

start_index = end_index+1;

end

for i=1:60
```

```
filename = ['result' int2str(i) '.csv'];

col = csvread (filename);

row = col';

dlmwrite('test.csv',row,'-append');

end
```

Result of mental workload levels

	Problem understanding					Evaluation				
Levels Subjects	1	2	3	4	5	1	2	3	4	5
1	16.6	18.4	37.4	39.2	46	17	17.8	19.8	19.8	20.4
2	25	36	37	50	60	19	19	32	34	41
3	25	26	41.2	57.4	63.8	47.2	54.6	55.6	58.8	68.4
4	37	38	41	41	44	30.8	32	32	33	50.2
5	61	65	67.4	68.2	75.8	20.2	20.4	21.8	22	22.2
6	29.6	30.8	39.8	40	55	27.6	32	33.4	40	47.2
7	25	29.4	32	33	35	34	46	47	47	55
8	46	47	51	54	62	28	28	32	40	55
9	39.4	52.4	53.4	61.8	68.2	21.8	22.2	23.2	27	38.4
10	24	26	36.8	44	45	22	29.2	40.4	47.8	49
11	19.6	20.8	26.4	42.6	44.4	23	24	26	27.6	28
12	68.4	70	76.8	79.6	80	28	33	34	39	39
13	8.4	9	10	16.6	18.2	36.6	38.6	46	53.6	68.2
14	22.2	24.8	25	36	48.6	24	33	36.2	38.8	50
15	32	33	43	46	55	22.6	24	30	32.8	37.8
16	20.6	20.8	22	22.6	23.6	35.4	45.6	65	75.6	75.8
17	38.2	39.8	41.4	57.6	58.4	2.6	4.4	7.6	8	9.2
18	36	43	47	64	67	23.8	25.2	25.2	44	56.8