Proceedings of OZCHI 2005, Canberra, Australia. November 23 - 25, 2005.

DO PHYSIOLOGICAL DATA RELATE TO TRADITIONAL USABILITY INDEXES?

Tao Lin
University of Yamanashi
lintao@hci.media.yamanashi.ac.jp

Wanhua Hu
University of Yamanashi
hu@hci.media.yamanashi.ac.jp

Masaki Omata
University of Yamanashi
omata@hci.media.yamanashi.ac.jp

Atsumi Imamiya
University of Yamanashi
imamiya@hci.media.yamanashi.ac.jp

ABSTRACT

Task performance data and subjective assessment data are widely used as usability measures in the human-computer interaction (HCI) field. Recently, physiology has also been explored as a metric for evaluating usability. However, it is not clear how physiological measures relate to traditional usability evaluation measures. In this paper, we investigated the relationships among three kinds of data: task performance, subjective assessment and physiological measures. We found evidence that physiological data correlate with task performance data in a video game: with a decrease of the task performance level, the normalized galvanic skin response (GSR) increases. In addition, physiological data are mirrored in subjective reports assessing stress level. The research provides an initial step toward establishing a new usability method using physiology as a complementary measure for traditional HCI evaluation.

KEYWORDS: *usability evaluation, physiological measures, task performance, subjective assessment.*

1. INTRODUCTION

Over the past three decades, traditional usability studies in the human-computer interaction (HCI) field have been conducted from the perspective of interface guidelines and ergonomics. Usability evaluation methods are mainly based on two kinds of data: task performance data, such as task completion time and performance error rates, and subjective assessment data from a questionnaire. However, subjective measures may not be reliable when used in isolation, as they are subject to cognitive mediation. For example, Wilson and Descamps reported that users rate video quality lower when performing a difficult task than when performing an easier one (Wilson & Descamps, 1996). This evidence suggests that contextual variables may influence users' assessment. Furthermore, as the scale of an experiment becomes larger, accurate subjective results require a larger number of subjects and more time for analysis, with greater associated costs. In addition, we also found that, although users are able to complete many tasks at the same performance level, they still may have different opinions about usability of the interactive system because of varying physiological and psychological reactions. Some may feel almost no discomfort, while others report that they experienced considerable stress. Users' psychophysiological investments, such as the level of mental effort or stress/anxiety incurred, also seem to be an important aspect in the evaluation of human computer interaction.

As HCI researchers, we recognize that task performance and subjective assessment are essential elements in the usability evaluation framework, but they are not enough. An objective method for measuring psychophysiological investments should be integrated into the traditional usability evaluation framework. The evidence from physiology has suggested that physiological measures (e.g., skin conductance, heart rate, pupil size, respiration, blood volume pulse) open a window into users' state, reflecting reactions of

Additional copies are available at the ACM Digital Library (http://portal.acm.org/dl.cfm)
or ordered from the CHISIG secretary (secretary@chisig.org)
OZCHI 2005 Proceedings ISBN: 1-59593-222-4

autonomic nervous system (Andreassi, 2000). Moreover, with recent improvements in technology, they can conveniently provide a continuous, high-resolution data source. Thus, we believe that physiological measures are natural and reliable data sources to objectively evaluate the psychophysiological investments.

Physiological measures in HCI field are being explored in many studies. They show potential, for example, in assessing multimedia quality and measuring presence in a virtual environment. The ultimate goal of our research project is to create a new usability evaluation method based on traditional usability evaluation indexes (task performance and subjective assessment) and users' psychophysiological investments. In this experiment, we created a video game environment to elicit task performance, subjective assessment and physiological data, and analyzed their links and correlations as an initial step to establish the usability evaluation method. In addition, we examine physiological responses to frustrating events for clues to help explain the differences between physiological and task performance data.

2. PHYSIOLOGICAL MEASURES

Psychophysiology explores the relationship between the mind and body, and the influence they have upon each other. Physiological measures reflect involuntary reactions of the autonomic nervous system (ANS). Physiologists have used these measures as objective identifiers of human emotions such as anger, grief and sadness (Ekman, Levenson, and Friesen, 1983), while researchers in human factors have used them to determine mental effort and stress (Vicente, Thornton, and Moray, 1987). Based on previous research, we chose the physiological properties of galvanic skin response (GSR), blood volume pulse (BVP) and heart rate (HR), because they can be measured non-invasively and are good indicators of arousal. The GSR signal is an indicator of skin conductance. In 1956, Seyle linked GSR to stress and ANS arousal (Seyle, 1956). Recent research also shows that skin conductance varies linearly with the overall level of arousal and increases with anxiety and stress (Picard, 1997; Healey, 2000), and can also reflect both emotional response and cognitive activity (Boucsein, 1992). The BVP signal is an indicator of blood flow. BVP increases with negatively valenced emotions such as fear and anxiety, and decreases with relaxation (Picard, 1997; Healey, 2000). Heart rate is also considered to be a good indicator of overall activity levels, with a high heart rate associated with an anxious state and a low rate with a relaxed state (Frijda, 1986). Heart rate was automatically calculated from BVP with the Biograph software in our experiment.

3. RELATED STUDIES ON USING PHYSIOLOGY AS A METRIC OF HCI EVALUATION

Several studies reported by Wilson and Sasse show a novel method for assessing multimedia quality in the context of networked applications: physiological responses to degradations in media quality (audio and video) are taken as an objective measure of user cost (Wilson, 2001). They found significant increases in GSR and HR, and significant decreases in BVP for video shown at 5 frames per second versus 25 frames per second (Wilson & Sasse, 2000a), even though most subjects didn't report noticing a difference in media quality. Another main finding of this research is that subjective and physiological results do not always correlate with each other (Wilson & Sasse, 2000b). These discrepancies between physiological and subjective assessment support the argument for a 3-D approach to evaluating multimedia quality and other HCI evaluation areas.

Ward et al. used several physiological measures (HR, BVP, and GSR) to assess users' responses to well-designed and poorly designed web pages. No significant differences were found between users viewing the two types of web pages, in part due to large individual differences. However, distinct trends were seen between the groups when the data were normalized and plotted. Participants using the poor interface showed higher levels of arousal (Ward & Marsden, 2003; Ward, Marsden, Cahill and Johnson, 2002). Their study also provides an example of how physiological data can be fit into usability evaluation.

Michael et al. used physiological measures (GSR, HR, skin temperature) to evaluate presence in stressful virtual environments. The experiments found that the change in HR can satisfy the requirements for a reliable, objective measure of presence, and that change in GSR does to a lesser extent; change in skin temperature does not (Meehan, Insko, Whitton and Brooks, 2002).

In the domain of entertainment technology, an experiment was conducted to test the efficacy of physiological measures as evaluators of collaborative entertainment technology (Mandryk & Inkpen, 2004). Their results suggest that there are different physiological responses when a user is playing against a computer than when playing against a friend. These results are mirrored in the subjective reports provided by the participants. In addition, physiological measures were also applied in affective game design (Sykes, J. and Brown, S. 2003)

Scheirer et al. applied a pattern-recognition strategy known as Hidden MarKov Models to GSR and BVP data to detect states of frustration deliberately induced by a slow computer game interface (Scheirer, Fernandez, Klein and Picard, 2002).

4. EXPERIMENT DESIGN

In this experiment, participants' physiological stress was measured as a form of psychophysiological investment using GSR, BVP and HR. We chose a video game as the experimental task. Participants were required to play a popular 3-D video game called *Super Mario 64*, which was manufactured by NINTENDO®. The GSR data were collected at 64 Hz; BVP and HR data at 128 Hz. Questionnaire data, including subjective assessment of stress, participants' statistics and task performance data, were collected into a data file and analysed using SPSS 11.0 software.

4.1. Participants

Fourteen male and four female university students aged 19 to 31 participated in the experiment. Before the experiment, all participants were required to fill out a background questionnaire about their experience with the game, average game times and personal information such as sex, age and handedness. Nine of the 18 participants were experienced with the game, while the other nine were somewhat experienced or completely inexperienced.

4.2. Task

In our experiment, participants were required to play three parts of *Super Mario 64* as quickly and correctly as possible. Task One was to run all regulated paths up to a mountain and defeat King Bo-Bomb, which took about 140 seconds for an average skilled player. Task Two was to pound a wooden post into the ground while avoiding an attack from a monster. It generally took 55 seconds to play that part of the game. In Task Three, participants were required to go through a snow slide without falling from it. This usually took about 50 seconds. Screen shots of the three tasks are shown in Figure 1.

Figure 1: Screen shots of the experimental tasks (Task One, Task Two and Task Three)

Participants played each task continuously for 10 minutes, regardless of results. That is, participants could repeat a task many times in ten minutes. In order to ensure consistent experimental conditions, game settings were not changed during the course of each experiment.

4.3. Experimental Apparatus and Protocol

The experimental tasks were performed in an HCI laboratory. *Super Mario 64* was played on a NINTENDO^{64} and viewed on a 25-inch television screen. Physiological signal data were collected with the ProComp Infiniti System and BioGraph Software from *Thought Technologies*^{TM}. To measure GSR, two sensors were placed on the left fingers. BVP and HR were simultaneously measured using a sensor on the right fingers. The BVP sensor is sensitive to movement, which is the most likely cause for noisy data.

Therefore, participants were required not to move the finger attached to the BVP sensor while playing. Additionally, game output was recorded in order to synchronize it with the screen showing physiological data. An experimental scene is shown in Figure 2.

Figure 2: An experimental scene

The experiment was divided into four phases: a welcome phase, a practice phase, a game phase and a debriefing phase. During the welcome phase, participants signed a consent form with a detailed description of the experiment, its duration and its research purpose. Each participant also filled out a background questionnaire.

During the practice phase, instructions were read to each individual, describing the game rules, as well as a brief tutorial on how to complete the game tasks. Participants were then allowed to practice for approximately three minutes for each game task.

At the outset of the game phase, a 10-minute resting baseline (GSR, BVP, and HR) was gathered. Then, participants played three tasks, and each task lasted 10 minutes. After completing a task, each participant had about 15 minutes to rest and completed a questionnaire to assess levels of task difficulty and stress caused by the game. During the course of playing the game, their task performance data and physiological data were collected. Participants were neither encouraged nor discouraged to talk.

At the end of the game, participants discussed their impressions of the experiment.

5. RESULTS AND DISCUSSION

In this section, subjective data assessing stress and physiological data are first described and analyzed. Then the correlations between task performance and physiological data are investigated. Finally, physiological responses to frustration events are examined.

5.1. Subjective Data and Physiological Data

Participants' stress was assessed subjectively on a unidimensional scale, which derived form the RSME (Rating Scale Mental Effort) (Zijlstra, 1993). Ratings of perceived stress are indicated by a cross on a continuous line. The line runs from 0 to 150 mm, and every 10 mm is indicated. Along the line, at several anchor points, statements related to perceived stress are given, for example, almost no stress or extreme stress. The scale is scored by the measurement of the distance from the origin to the mark in mm.

Means for the stress scores were analysed using an ANOVA analysis to determine whether there is difference in stress perceived for the three tasks. The result shows significant differences ($F=8.25$, $df=53$, $P=0.001$) (see Figure 3). Task Three caused the highest stress level, followed by Task Two and Task One. Although the difference in stress-score means has statistical significance, the pattern is not always consistent. All the participants stated consistently on the questionnaire that they perceived the least stress during Task One. However, four participants reported that Task Two caused the highest stress level. The questionnaire for evaluating task difficulty could provide partial reasons for the discrepancies. All participants reported that Task One was easy, and five participants thought that Task Two was the most difficult, including the four participants who perceived the highest stress level during Task Two. In

addition, at the end of the tasks, when the four participants were asked why Task Two caused more stress than Task Three, they also explained that Task Two was very difficult to play. The results suggested that the discrepancy with the pattern could be attributed to task difficulty.

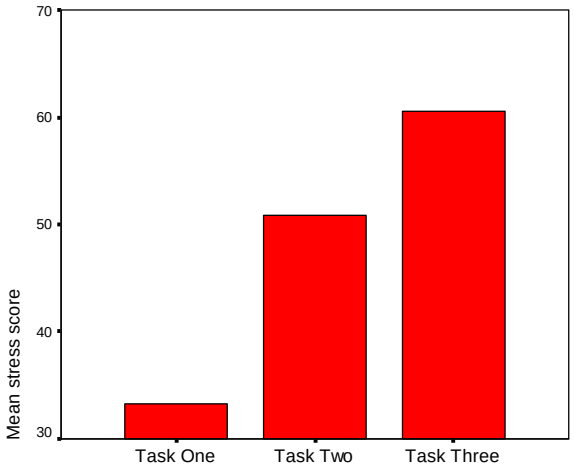

Figure 3: Mean stress score of three tasks.

We analysed GSR, BVP and HR data across three tasks. There were very large individual differences in physiological response, thus individual baselines have to be taken into account. To minimize the influence of individuals, we normalized GSR data using the formula (signal-baseline) / baseline. We found that there were significant differences in mean normalized GSR among the three tasks by using an ANVOA analysis ($F=11.6$, $df=53$, $P<0.001$) (see Figure 4).

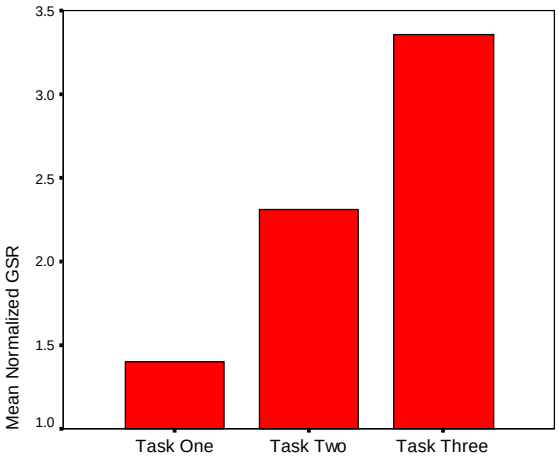

Figure 4: Mean normalized GSR of three tasks

The results are mirrored in the subjective report for assessing stress levels. As previous research suggested, GSR data can indicate the different levels of stress caused by the tasks. Moreover, GSR data from all participants were consistent with the pattern: Task Three caused the greatest normalized GSR, followed by Task Two and Task One. The four participants who perceived the highest stress levels during Task Two in subjective report were not affected by the task difficulty.

We also examined HR and BVP data but didn't find significant differences among the three tasks.

5.2. Task Performance Data and Physiological Data

Participants' success times for a task was defined as their task performance index. The more successes participants got, the higher task performance level was. The sum of success times for three tasks was analysed as overall task performance. Among participants, participant 7, who had 22 successes, reached

the highest task performance level. In order to examine the relationship between task performance data and physiological data, we classified participants into three groups with different task performance levels according their success times: low (0-7 successes), middle (8-15 successes) and high (16-22 successes).The mean normalized GSR across the three levels was shown in figure 5. An ANVOA analysis showed that there were significant differences among the three groups with different task performance levels (F=3.72, df=17, p= 0.04). The group with low task performance level experienced the greatest normalized GSR, followed by the group with middle task performance level and the group with high task performance level.

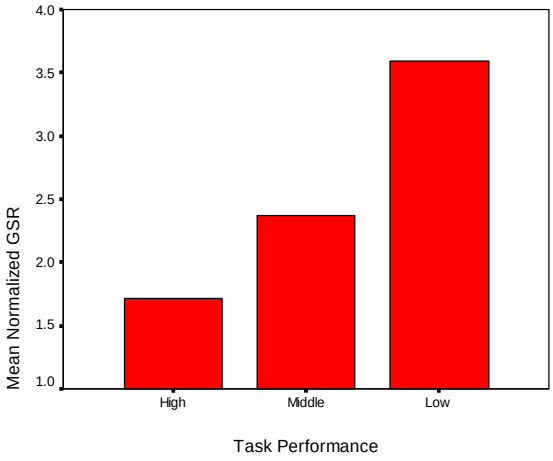

Figure 5: Mean normalized GSR across three task performance levels

We also analysed normalized GSR and success times for each task separately (see Figure 6). For Task One,

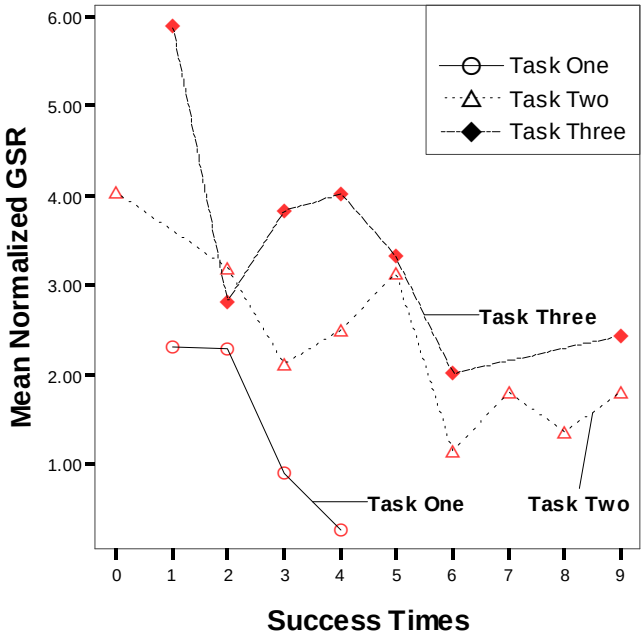

Figure 6: Mean normalized GSR across success times for three tasks

there is a consistently decreasing trend in normalized GSR with an increase in success times. Normalized GSR shows statistically significant differences across different success times (1, 2, 3 and 4 successes) (F=10.12, df=17, p=0.001). In addition, success times and normalized GSR also show a negative, statistically significant linear correlation (R=-0.77, P<0.01). For Task Two and Task Three, mean

normalized GSR doesn't decrease consistently with an increase in success times, but they still show an overall decreasing trend with an increase in success times. Correlation analysis for them shows that Normalized GSR of Task Two negatively and significantly correlates with its success times. ($R=-0.60$, $p=0.005$), and Task Three does not show significant correlations between normalized GSR and success times.

There seem to be reasons for the inconsistencies in the trend and the lack of statistical significance in Task Two and Task Three. The first is that the number of participants was not large enough distinguish to subtle task performance difference when tasks became more difficult. Moreover, there are larger differences in physiological responses between individual participants during difficult tasks than during easy task, which would also have contributed to the lack of statistical significance.

5.3. Physiological Response to Frustration Events

One of the advantages of physiological measures is that they provide a high-resolution, continuous and contextual data source. GSR is a highly responsive body signal, and, when collected at 64 Hz, it provides fast-response time-series data. In our study, we inspected GSR response to frustration events by examining small periods of time surrounding frustration events. In Lawson's theory of frustration (Lawson, 1965), frustration is described as "the occurrence of an obstacle that prevented the satisfaction of a need." Accordingly, we examined frustration events during Task One and Task Two by reviewing the recorded tapes, which mainly included the following kinds of events: being attacked by enemies, accidentally falling from a mountain road or a bridge, and being hit by an unexpected bomb.

Frustration events for Task One and Task Two were analysed. There were 355 frustration events in all. GSR data were windowed 5 sec prior to the frustration events and 10 sec after. For 256 frustration events, participants experienced more than 5% increases in GSR when frustrated. An example of a frustration event and its corresponding GSR response is shown in Figure 7.

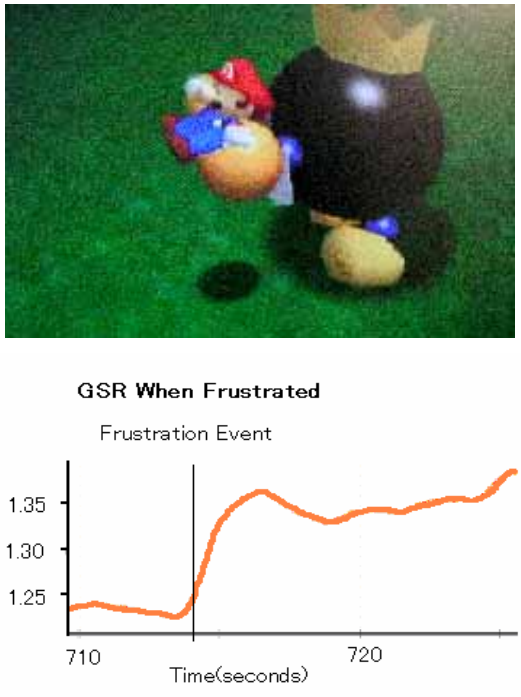

Figure 7: A participant was frustrated when attacked and experienced a large increase in GSR.

We also investigated the relationship of frustration events and task performance. Figure 8 shows the distribution of frustration events across success times for Task One and Task Two. The results suggest a trend: With an increase in success times, the mean number of frustration events decreases. Participants with poor performance experienced more frustration.

We also analysed participants' GSR response to their failures in Task Three, because once participants fell from the snow slide, they failed and had to start again. Ninety percent of participants' failures produced more than a 5% increase in GSR.

The link between success times and frustration events provides a reason for explaining the correlations of task performance and physiological data, although it is not comprehensive. When a frustration event happened, participants had to invest extra time and effort to deal with it. Consequently, task performance decreased and GSR responses changed.

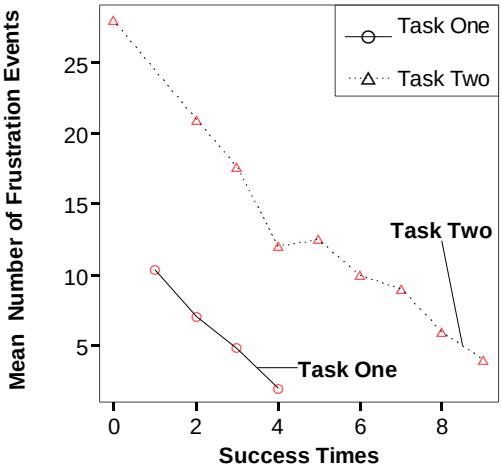

Figure 8: Distribution of frustration events across success times

6. CONCLUSIONS

While the results of this study are preliminary and warrant further investigation, there are three findings that indicate the necessity and promise for using physiology to evaluate usability. The first is that physiological measures are consistent with subjective measures and show significant sensitivity to changes in stress levels. Another finding is that, while we have been not able to disentangle the cause-effect relationship between task performance and physiological data, we did find a correlation between the two kinds of data. The finding suggests that different levels of task performance may be indicated by physiological data. In addition, we found that most frustration events produced remarkable physiological responses, and the participants with poor performance tended to experienced more frustration events. The poor task performance and greater physiological response could be attributed to these frustration events.

These results show the potential value of physiological data as a data source for usability evaluation. Physiological data not only provide a way to objectively measure psychophysiological investments and make it possible to evaluate usability at a more detailed level, but also provide some clues to explaining differences in task performance. The study takes an initial step toward establishing a new usability evaluation method using physiology as a complementary measure or as an independent measure for HCI evaluation.

7. FUTURE WORK

The ultimate purpose of our research is to establish a new usability evaluation method based on three kinds of data: user subjectivity, task performance data and physiological data, which not only evaluates the system's effectiveness and efficiency, but also takes physiological reactions into account. This paper presents an initial examination of the correlation between physiological data and the traditional index of usability evaluation. More steps are required to create this new usability evaluation method.

The experimental results showed that GSR data correlate to task performance data, but we don't conclude that there is a cause-effect relationship between the two kinds of data. We need more rigorous experimental conditions and analytical methods to understand the correlation. There are other physiological measures that may be useful for usability evaluation in HCI. For example, HR variability has been studied extensively. It has been used as an indicator of the extent of task engagement in information processing requiring significant mental effort (Sirevaag, Kramer, Wickens, Reisweber, Strayer and Grenell, 1993; Tattersall & Hockey, 1995; Wilson, 1993), and has been used to detect rapid transient shifts in mental workload (Kramer, 1991).

Eye tracking is another well studied measure. A major use of eye-tracking technologies in HCI is to ascertain what is being processed based on what is being looked at. The use of eye tracking enables usability evaluations to be conducted at a more detailed level, pinpointing specific areas within a display that may be causing usability problems and indicating how such issues change over time. In our usability evaluation framework, we will synchronize physiological responses with eye movements and establish a connection between elements of the interface and a user's current state, through which we are able to understand where usability problems are and how users react to them. In addition, eye tracking has potential to help explain task performance and to detect user fatigue and strain. For example, the number of fixations overall is thought to be negatively correlated with search efficiency, and longer fixations are an indication of difficulty in extracting information from a display (Goldberg & Kotval, 1999). It also has been found that pupil diameter decreases with fatigue (Hess, 1972; Lowenstein & Loewenfeld, 1964), and pupil diameter changes are related to positive effects (pupil dilation) and negative effects (pupil constriction) (Partal, Maria and Surakka, 2000). In our experiment, different levels of task performance produced different physiological responses. Eye tracking could provide clues about explaining the differences in task performance and physiological responses.

To create a new methodology, more subtle experimental manipulations should also be explored. For instance, experiments should be extended to different domains and a variety of tasks. Moreover, ensuring a sufficient number of participants is needed to increase the statistical power of the study.

8. ACKNOWLEDGEMENTS

We thank the members of the HCI group at the University of Yamanashi for their support of the research. This study was supported in part by the Grants-in-Aid for Scientific Research of the Japan Society for the Promotion of Science and by the RIEC of Tohoku University awarded to A. Imamiya.

9. REFERENCES

Wilson, F. & Descamps, P. T. (1996). Should We Accept Anything Less than TV Quality: Visual Communication, *International Broadcasting Convention*, 12th – 16 September 1996, Amsterdam.

Andreassi, J.L. (2000). *Psychophysiology: Human Behavior and Physiological Response* (4th Edition.). Mahwah, NJ: Lawrence Erlbaum Associates

Ekman, P., Levenson, R.W., and Friesen, W.V. (1983). Autonomic Nervous System Activity Distinguishes among Emotions. *Science*, 221(4616): 1208-1210.

Vicente, K.J., Thornton, D.C., and Moray, N. (1987) Spectral Analysis of Sinus Arrhythmia: A Measure of Mental Effort. *Human Factors*, 29(2): 171-182.

Seyle, H. (1956). *The stress of life*. New York: McGraw-Hill.

Picard, R.W. (1997). *Affective Computing*. Cambridge, MA: MIT Press.

Healey, J.A. (2000). *Wearable and Automotive System for Affect Recognition from Physiology*. PhD thesis (pp.22-45), Massachusetts Institute of Technology.

Boucsein, W. (1992). Electrodermal Activity. New York: Plenum Press.

Frijda, N. H. (1986). *The emotions.* Cambridge : Cambridge University Press.

Wilson, G.M.. (2001). Psychophysiological Indicators of the Impact of Media Quality on Users. *Proceedings of CHI 2001 Doctoral Consortium* (pp95-96). Seattle, Wash., USA: ACM Press.

Wilson, G.M. and Sasse, M.A. (2000). Do Users Always Know What's Good For Them? Utilizing Physiological Responses to Assess Media Quality. *Proceedings of HCI 2000: People and Computers XIV - Usability or Else!* (pp.327-339), Sunderland, UK: Springer.

Wilson, G.M. and Sasse, M.A. (2000). Investigating the Impact of Audio Degradations on Users: Subjective vs. Objective Assessment Methods. *Proceedings of OZCHI 2000: Interfacing Reality in the New Millennium* (pp135-142), Sydney, Australia.

Ward, R.D. and Marsden, P.H. (2003). Physiological responses to different WEB page designs. *International Journal of Human-Computer Studies, 2003, 59(1/2)*: 199-212.

Ward, R.D., Marsden, P.H., Cahill, B., and Johnson, C. (2002). Physiological Responses to Well-Designed and Poorly Designed Interfaces. *Proceedings of CHI 2002 Workshop on Physiological Computing.* Minneapolis, MN, USA.

Meehan, M., Insko, B., Whitton, M. and Brooks, F. (2002). Physiological measures of presence in stressful virtual environments. *Proceedings of the 29th annual conference on Computer graphics and interactive techniques* 2002, San Antonio, Texas July 23 – 26.

Mandryk, R.L. and Inkpen, K. (2004). Physiological Indicators for the Evaluation of Co-located Collaborative Play. *Proceedings of Computer Supported Cooperative Work* (CSCW 2004). Chicago, IL, USA.

Sykes, J. and Brown, S. (2003). Affective Gaming: Measuring Emotion Through the Gamepad. *In Proceeding of CHI 2003(732-733).* New York: ACM Press.

Scheirer, J., Fernandez, R., Klein, J., and Picard, R. (2002). Frustrating the User on Purpose: A Step Toward Building an Affective Computer. *Interacting with Computers, Vol. 14, No. 2*, pp. 93-118.

Zijlstra, F.R.H. (1993). *Efficiency in work behavior. A design approach for modern tools.* PhD thesis, Delft University of Technology. Delft, The Netherlands: Delft University Press.

Lawson, R. (1965). *Frustration: The Development of Scientific Concept*. New York: Macmillan.

Sirevaag, E. J., Kramer, A. E, Wickens, C. D., Reisweber, I., Strayer, D. L., and Grenell, J. F.(1993). Assessment of pilot performance and mental workload in rotary wing aircraft. *Ergonomics*, 1993, 36, 1121-1140.

Tattersall, A., and Hockey, G. (1995). Level of operator Control and Changes in Heart Rate Variability during Simulated Flight Maintenance. *Human Factors 1995, 37 (4)*, 682-698.

Wilson, G. F. (1993). Air-to-Ground training missions: A psychophysiological workload analysis. *Ergonomics, 36*, 1071-1087.

Kramer, A. F. (1991). Physiological metrics of mental workload: A review of recent progress. In D. L. Damos (Eds.). *Multiple-Task performance* (pp. 279-328.). London: Taylor and Francis.

Goldberg, J.H., Kotval, X.P. (1999). Computer interface evaluation using eye movements: methods and constructs. *International Journal Of Industrial Ergonomics* 24, 631–645.

Hess, E.H. (1972). Pupillometrics. In Greenfield, N.S., and Sternbach R.A. (Eds.). *Handbook of psychophysiology* (pp. 491-5310). New York: Holt, Rinehart and Winston.

Lowenstein, O. and Loewenfeld, I.E. (1964). The sleep-waking cycle and pupillary activity. *Annals of the New York Academy of Sciences*, 1964, 117, 142-156.

Partal, T., Maria, J. and Surakka, V. (2000). Pupillary responses to emotionally provocative stimuli. *Proceedings of ETRA 2000*, Florida, USA.