A Composite Measure for the Evaluation of Mental Workload

Lee-ming Wang[1], Vincent G. Duffy[1,2], and Yingzi Du[3]

[1] School of Industrial engineering, Purdue University, West Lafayette, IN 47906, USA
{lwang,duffy}@purdue.edu
[2] Regenstrief Center for Healthcare Engineering, and School of Agricultural and Biological Engineering, Purdue University, West Lafayette, IN 47906, USA
[3] Department of Electrical and Computer Engineering, Indiana university-Purdue University Indianapolis, Indianapolis, IN 46202, USA
yidu@iupui.edu

Abstract. Physiological measures have found reliable sensitivity to the variation of mental efforts to tasks of different difficulty levels. The sensitivity needs to be enhanced for further application. This paper proposed a composite measure consisting of three physiological measures, facial skin temperature, eye blinks and pupil dilation. The facial skin temperature will be measured by an infrared camera. One dimensional iris image will be used for the measurement of eye activities. All measurement will be done in real-time and unobtrusively. A preliminary experiment will be conducted for each measure to demonstrate their sensitivity. The combination then will be accomplished by factor analysis and regression analysis. Last, the analysis will focus on the improvement in sensitivity from the combination of individual measures.

Keywords: mental workload, facial skin temperature, pupil dilation, thermography, 1D iris identification.

1 Introduction

Multitasking is common in human daily activity. For example, driving consists of at least two tasks, tracking – maintain vehicles on the designated route, searching direction, and memorizing – keep in mind the direction to destination. Hence, it is reasonable to consider the 'multimodal information' embedding in the human behaviors with a composite evaluation index.

Past research has supported this idea. Previous research [1-4] showed facial skin temperature reveals sensitivity to the variation of mental workload. However, the researchers also noticed that the sensitivity may not be strong enough for practical application in complex tasks [4, 5]. The combination of various measures was suggested to enhance the explanatory power. Some related research on this concern has been done. Ryu et al. [6] found that in a dual-task experiment, none of the selected measures, excluding the subjective evaluation, was able to detect the variation of the mental workload for both tasks. Comparatively, when these measures

V.G. Duffy (Ed.): Digital Human Modeling, HCII 2007, LNCS 4561, pp. 460–466, 2007.

were combined, the significant difference of mental efforts for both tasks was successfully identified. The authors demonstrated, because the demands for each mental processing resource may be determined by different physiological measures, the combination of physiological measures is a good way to carry out the maximum explanatory power. Lin et al. [7] had a similar conclusion. They applied two composite measures, different combination of eye movement and hand motion, to evaluate the mental efforts of players in a video game. The combination was found be able to detect all the variation of mental efforts with regard to the games of different difficulty levels, while single measures failed to accurately detect that variation.

In this research, a composite measure, consisting of facial skin temperature, pupil dilation, and eye blink interval, is proposed in order to evaluate mental workload as well as user performance. These three physiological measures have been well studied as evaluation indices of mental workload. We believed these measures can reflect the mental demands of whole or partial different processing resources and therefore it is reasonable to combine them to achieve greater explanatory power for tasks in different domains.

Furthermore, in this research, real-time and non-intrusive implementation for each measure is proposed. Non-intrusive measures are preferred because they can introduce the minimum of disruption to the users and can bring most convenience in practical utilization. Real-time is also an important feature. The continuous surveillance of user status in tasks provides all the data over time, not only at the beginning and the end of the tasks [7]. The future work can be extended, with the higher reliability of composite measures, to analyze the effects of 'learning' and 'individual difference' on the evaluation of mental workload. In addition, adaptive automation provides a promising application [8]. If the measure helps a system with self-adjustment functions according the real-time performance of users, we believe that it can make contribution to many applications such as real-time aid system for operators, design of human computer interface, and quality control for complex systems such as e-service.

2 Measures

2.1 Facial Skin Temperature

Variation of facial skin temperature has received some attention as a physiological measure of mental status. This is because they showed the significant relationship to the change of mental status, and on the other hand, the non-contact measurement is practical with thermal infrared camera since human face emits mid- and far- infrared [2].

When sensations occur, the autonomic nervous system acts to mediate the response to the stimulus, thus resulting in the redistribution of blood flow and then the change of local skin temperature [2]. Genno et al [3] showed the temperature change in nose area when subjects experienced stress and fatigue. In a tracking task, the nose temperature dropped significantly at the beginning of the task and at the moment when an unexpected emergency signal alarmed during the task, while other parts of the face did not have observable patterns for change in temperature. In another task inducing fatigue, the nose temperature also showed significant drop. They quantified this relationship and applied the model to evaluate image preference. The results were in accord with the subjective evaluations for individuals. The temperature drop in nose area with increased mental workload/stress was also observed in other studies. Or et al [4] found that drop in nose temperature of drivers in relation to different driving conditions and environments such as virtual driving test/real car driving, city route/highway route, and with/without mental loading task. Veltman et al [5] detected the temperature drop in nose area of subjects in Continuous memory-Task on two difficulty levels. The research, though they have demonstrated the relationship between nose skin temperature and variation of mental workload, suggested further work in the investigation of measure sensitivity with more difficulty levels and the combination of other physiological measures to enhance the sensitivity and accuracy.

2.2 Pupil Dilation

The correlation between pupil dilation and increased mental efforts has been well known for decades. Beatty [9] investigated task-evoked papillary response through a series of studies. Pupil dilation was determined to be as a reliable indicator of mental efforts or processing load of tasks such as memory, language processing, reasoning, and perception. The results from Murata et al [10] also supported the fluctuation of pupil area to be a meaningful measure of mental efforts with the control of respiration. Iqbal et al [11] showed that papillary response, in interactive tasks, was an effective index of mental workload of computer users when the tasks can be decomposed subtasks in hierarchal structure. In addition to its usability as a objective measure, another attractive feature of pupil dilation is that its measurement can be non-intrusive, which causes minimal interference and enables continuous data acquisition. Therefore, it is expected pupil dilation can contribute explanatory power to the composite measure.

2.3 Eye Blink Interval

Eye movement is another promising indicator of mental efforts due to its sensitivity and the possibility of its non-intrusive and continuous measurement [12]. Blink interval is known for its positive correlation with increased mental workload, while blink duration tends to decrease against more intense workload. These physiological responses allow eyes to have more time to collect visual information for tasks. Van Orden et al. [12] found the blink frequency and duration declined as a function of the density of targets in their visuospatial task, where subjects had to identify targets on a computer screen and opt for an appropriate prosecution on a target that was reaching a specific location based on its identification (friend/enemy). Veltman et al. [13] investigated various physiological measures, including eye blinks, in a flight simulator. In their experiments, blink frequency and duration decreased from rest status when subjects perform both two different flight tasks. Moreover, their results showed that blink frequency increased when subjects met more difficult flight condition with arithmetic tasks. It was probably because subvocal activity of the memory process within arithmetic tasks stimulated the muscles of the eyelid, thus forcing the blinks. Similar results were found by Ryu et al. [6]. Eye blink interval revealed weak sensitivity to the mental workload in tracking tasks, but did not in the arithmetic task in which subjects has to memorize the digits to be manipulated. Previous research suggests the feasibility of eye blinks as a physiological measure; therefore in the proposed research, we adopt it as a component of the composite measure.

3 Experiment

Facial skin temperature, pupil dilation, and eye blink intervals are selected in the proposed research. Implementation of the measurement is real-time and non-intrusive in order to get the intermediate state of users and to avoid unnecessary processing load in addition to task-evoked one. In the following sections, section 3-1 will introduce the measurement in the experiment, section 3-2 describes the design of the experiment, and section 3-3 has the detail about the task.

3.1 Measurement

A thermal camera will be used to capture the facial skin temperature from subjects. The measurement locations on the face will be designated as Region of Interest (ROI). Using thermal camera, the background will be dark and the face will be brighter pixels due to the body temperature. The higher the temperature is, the brighter the pixel will be in the image. The selection of ROIs is determined by their capacity to reflect the variance of temperature due to mental workload. The temperature of a ROI is defined as the mean temperature of pixels (or partial hottest pixels) in it. The computation of mean temperature will be done for every frame in the thermal clip. Other factors that might affect subjects' facial temperature such ambient temperature, but it can be reduced by carefully controlling the environment.

A near infrared (NIR) camera will be used to capture the pupil changes and eye blink. NIR cameras are popular used in iris recognition [14, 15]. In the visible light wavelengths, it is difficult to get the detailed iris patterns from dark color eyes. But under the NIR light, small details from the iris (even dark iris) are visible. Du et al [14, 16] devised a methodology to recognize iris pattern using 1-D iris siganture. In their innovative approach, given an iris image, their method will automatically segment the iris out, which includes detection of pupil, limbic, eyelids, and eyelashes. The segmented iris is then transformed into polar coordinates. In the polar coordinates, all radial information will become linear; for example, the pupil area of our interest becomes rectangle, which can ease the task to calculate pupil size. The size of the same iris taken at different times may be variable in the image as a result of changes in the camera-to-face distance. Due to stimulation by light or for other reasons (such as hippus, the natural continuous movement of the pupil) the pupil may be constricted or dilated. These factors will change the iris resolution, and the actual size of the pupil in the image. However, it is shown that the limbic radius would be stable during pupil dilation or constrain. To solve these problems, we should use the ratio between the pupil radius to the limbic radius to measure the dilation of the pupil.

This methodology is feasible even if only partial iris image is available [17]. The measurement of pupil size and eye blink in this research will be based on Du's 1-D iris image approach. Pupil size will be detected for each frame in the clip. In addition, eye blink interval and duration can be estimated using difference images from consecutive video frames. Using proper setting up the NIR illuminater, the glare or the camera noise from the image acquisition can be reduced. In Ref. [14], it is showed that this method is very robust in pupil detection.

3.2 Experimental Design

The first stage is to conduct the preliminary experiment to demonstrate the selected physiological measures as indicators to mental workload induced by the designed tasks. In the present research, the selected physiological measures include facial skin temperature, pupil dilation, and eye blink intervals. Each of the physiological measures is anticipated to be sensitive to the subject's mental workload in tasks. And, on the other hand, it is also expected that these three measures will reveal sensitivity in a somewhat different way to the variation of workload for different difficulty levels. According to the three criteria, proposed by Kahneman [9], for physiological indicators for processing load, the measures in the presented research will be statistically tested for 'within-task', 'between-task', and 'between-individual' difference. An effective measure should be able to reflect the difference from Kahneman's criteria. The 'Within-task' criteria means the sensitivity a measure has for a task (i.e. difference of measurement from rest to performing task), while the 'between-task' criteria further confirms the intensity of the sensitivity (i.e. pairwise difference of measurement among difficulty levels). In the present research, we will focus on these two criteria and leave the third one as future work. Thus, the three measures, as well as a subjective measure, will be applied to tasks of different difficulty levels. ANOVA will be applied in all the statistical tests for the data analysis. And NASA-TLX will be used to obtain subject report immediately after execution of each task.

The second stage is to combine the three measures together. Ryu et al. [6] suggested two methods, factor analysis and regression analysis to achieve the combination. Although in their methods both led to similar results, the two methods will again be adopted in this research. By doing this we can have comparison for each method, and on the other hand, we can also have a chance to see if any redundancy exists in the selection of measures. The statistical test method will be the same as in the previous stage.

3.3 Task

The design of the task is aimed to provide a very-controlled task so that subjects' workload can be correctly evaluated at different levels of difficulty. A dual task is considered in this experiment since the composite measure is expected to be feasible in the complex environment. Subjects will need to perform a tracking task and mental arithmetic simultaneously as in [6].

In the tracking subtask, subjects have to control a rectangular frame by moving the cursor to track a ball on the screen, which travels on a trajectory unknown to subjects as in [3]. If the ball moves out of the frame, a buzzer will ring to alarm the subjects.

The difficulty of the tracking task is determined by the frame size, ball speed and moving path. For example, in a very difficult task a subject will need to use a small frame to enclose a fast ball which also makes many sudden turns. This subtask is designed with five difficulty levels.

When subjects keep the cursor following the ball, meanwhile they also need to perform mental addition on the numbers displayed on the screen and report the results orally. Subjects need to sum up two numbers and each number contains up to three digits, based on how difficult the mental arithmetic subtask is. And the mental addition subtask contains two difficulty levels: EASY and HARD.

Within-subject factorial design will be implemented in this experiment. Therefore, subjects will need to complete the task in ten conditions. Tasks in different conditions will be executed separately with the insertion of rest periods between each execution. Each execution and rest period will last five minutes. Further, it is expected to recruit 100 subjects for the experiment.

In the task, subjects have to identify the motion of the ball and immediately follow it. It is expected that the eye responses such as eye blinks and pupil dilation can reveal the mental activity of users corresponding to visual stimulus. In the meantime, subjects will have mental addition on the numbers displayed aside the ball-tracking subtask on screen. Mental arithmetic in the task may incur the intensive workload in working memory. The tendency toward subvocal rehearsal activity in the memory tasks can suppress the trigger of some eye related activities. Subvocal rehearsal activity may interfere with eye activities because eyelid, eyeball, and tongue share same area of primary motor cortex [13]. It is reasonable to expect that, even in the case that the measure using eye activities show poor sensitivity to subject's mental workload in the task, the other measure, the detection of facial temperature with thermography, can supplement the explanatory power. Hence, the proposed composite measure has the potential to pertain consistent sensitivity to mental workload during the task.

Previous research [4] indicated that facial skin temperature is sensitive, though not consistently, to mental workload in mental math tasks. Considering this perspective, it is also interesting to observe whether the addition of the measure of eye activities can enhance the overall explanatory power.

4 Conclusion

In this paper a composite physiological measure is proposed by combining measurement of facial skin temperature and eye related activities. Thermal images will be captured from an infrared camera to detect skin temperature, and the measurement of eye activities will be fulfilled by the detection of pupil boundary in one dimensional iris images. All measures will be applied to the subjects who will need to track a moving ball and perform mental arithmetic. Ten task conditions will be provided in order to test the sensitivity of the proposed measure. The analysis will focus on the improvement of sensitivity from the combination of individual measures.

References

1. Puri, C., Olson, L., Pavlidis, I., Levine, J., Starren, J.: StressCam: non-contact measurement of users' emotional states through thermal imaging. In: Proceedings of CHI 2005, pp. 1725–1728 (2005)
2. Pavlidis, I., Levine, j., Baukol, P.: Thermal imaging for anxiety detection. In IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications 2000, pp. 104–109 (2000)
3. Genno, H., Ishikawa, K., Kanbara, O., Kikumoto, M., Fujiwara, Y., Suzuki, R., Osumi, M.: Using facial skin temperature to objectively evaluate sensations. International Journal of Industrial Ergonomics 19, 161–171 (1997)
4. Or, C.K.L., Duffy, V.G.: Development of a facial skin temperature-based methodology for non-intrusive mental workload measurement, Occupational Ergonomics (in press 2007)
5. Veltman, H.J.A., Vos, W.K.: Facial temperature as a measure of operator state. In: Proceedings of the 1st International Conference on Augmented Cognition, 2005, pp. 22–27 (2005)
6. Ryu, K., Myung, R.: 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, 991–1009 (2005)
7. Lin, T., Imamiya, A.: Evaluating usability based on multimodal information: An empirical study. In: Proceedings of ICMI 2006, pp. 364–371 (2006)
8. Schmorrow, D., Stanney, K.M., Wilson, G., Young, P.: Augmented cognition in human-system interaction. In: Salvendy, G. (ed.) Handbook of human factors and ergonomics, John Wiley, New York (2006)
9. Beatty, J.: Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources. Psychological Bulletin 2, 276–292 (1991)
10. Murata, A., Iwase, H.: Evaluation of mental workload by fluctuation analysis of pupil area. In: Proceedings of the 20th Annual International Conference of the IEEE 1998, vol. 6, pp. 3094–3097 (1998)
11. Iqbal, S.T., Zheng, X.S., Bailey, B.P.: Task-evoked pupillary response to mental Workload in human-computer interaction. In: Proceedings of CHI 2004, pp. 1477–1480 (2004)
12. Van Orden, K.F., Limbert, W., Makeig, S., Jung, T.: Eye activity correlates of workload during visuospatial memory task. Human factors 43(1), 111–121 (2001)
13. Veltman, J.A, Gaillard, A.W.: Physiological workload reactions to increasing levels of task difficulty. Ergonomics 41(5), 656–669 (1998)
14. Du, Y., Ives, R., Etter, D., Welch, T.: Use of one-dimensional iris signatures to rank iris pattern similarities. Optical Engineering 45(3), 37201–37211 (2006)
15. Daugman, J.: How Iris Recognition Works. IEEE Transaction on Circuits and Systems for Video Technology 14(1), 21–30 (2004)
16. Du, Y., Ives, R., Etter, D., Welch, T.: A New Approach to Iris Pattern Recognition. In: Proceedings of SPIE 2004, vol. 5612, pp. 104–116 (2004)
17. Du, Y., Ives, R., Bonney, B., Etter, D.: Analysis of partial iris recognition. In: Proceedings of the SPIE 2005, vol. 5779, pp. 31–40.20 (2005)