Galvanic Skin Response (GSR) as an Index of Cognitive Load

Yu Shi
National ICT Australia
Bay 15, Technology Park
Eveleigh, NSW 1430,
Australia
yu.shi@nicta.com.au

Natalie Ruiz
National ICT Australia
Bay 15, Technology Park
Eveleigh, NSW 1430
Australia
natalie.ruiz@nicta.com.au

Ronnie Taib
National ICT Australia
Bay 15, Technology Park
Eveleigh, NSW 1430
Australia
ronnie.taib@nicta.com.au

Eric H. C. Choi
National ICT Australia
Bay 15, Technology Park
Eveleigh, NSW 1430
Australia
eric.choi@nicta.com.au

Fang Chen
National ICT Australia
Bay 15, Technology Park
Eveleigh, NSW 1430
Australia
fang.chen@nicta.com.au

CHI 2007, April 28 – May 3, 2007, San Jose, California, USA
ACM 978-1-59593-642-4/07/0004.

Abstract

Multimodal user interfaces (MMUI) allow users to control computers using speech and gesture, and have the potential to minimise users' experienced cognitive load, especially when performing complex tasks. In this paper, we describe our attempt to use a physiological measure, namely Galvanic Skin Response (GSR), to objectively evaluate users' stress and arousal levels while using unimodal and multimodal versions of the same interface. Preliminary results show that users' GSR readings significantly increase when task cognitive load level increases. Moreover, users' GSR readings are found to be lower when using a multimodal interface, instead of a unimodal interface. Cross-examination of GSR data with multimodal data annotation showed promising results in explaining the peaks in the GSR data, which are found to correlate with sub-task user events. This interesting result verifies that GSR can be used to serve as an objective indicator of user cognitive load level in real time, with a very fine granularity.

Keywords

Cognitive load, biosensors, multimodal interaction

ACM Classification Keywords

H.1.2 [User-Machine Systems]: Human information processing; H.5.2 [Information Interfaces and

Presentation]: User Interfaces–Input Devices and Strategies, Interaction styles.

Introduction

Cognitive load of a user represents an important factor for adaptive human computer interfaces, especially under high intensity work conditions and complex tasks. Cognitive load is variously described as the level of perceived effort associated with learning, thinking and reasoning [7]. Our research focuses on the detection of cognitive load fluctuations, based on specific patterns of interaction of users of multimodal interfaces. Multimodal interfaces are known to reduce the level of experienced cognitive load over unimodal interfaces [6]. Unobtrusive, real-time detection of cognitive load can be used to adapt the pace, volume and format of the information conveyed to the user, depending on their individual cognitive load experience. However, it is crucial to establish a reliable indicator of experienced cognitive load in order to assess which multimodal patterns are likely to occur at high or low levels of load, as there are large individual variations from person to person. Measures used by previous research rely either on subjective rankings of load, requiring user introspection, or physiological sensors [7]. While the former has been largely used and is believed to be consistent over time, it requires users to interrupt their current task and complete questionnaires or respond to interview questions, and cannot realistically be used as a real-time control indicator.

Physiological sensors on the other hand, seem to be ideal in providing real-time control indications of experienced load, despite their level of intrusiveness. The choice of sensors, their reliability and accuracy in reflecting cognitive load are still open issues discussed in the art. This paper begins with a short review of such technologies, and then presents the method and results of a user experiment that we conducted to assess the validity of Galvanic Skin Response (GSR) as an indicator of experienced cognitive load.

Physiological Measures of Stress and Arousal

Physiologists have long used these kinds of measures to evaluate stress, affective and arousal states. Recently, HCI researchers have used them to evaluate the usability of HCI systems and software [6]. When appropriately collected and correlated with other signals such as speech, facial expression and eye/body movement, these measures can help reveal user's cognitive state [2]. Such readings are useful because they are continuous and allow the signal to be measured at a high rate and at very fine granularity [7]. In contrast with subject reporting on the overall impression of the task, physiological sensors can uncover a number of fluctuations that may occur as the subject completes the task, indicating fluctuating levels of experienced cognitive load.

Galvanic Skin Response, (GSR), is a measure of conductivity of human skin, and can provide an indication of changes in human sympathetic nervous system (SNS). Research and empirical data have long linked GSR variation to stress and SNS arousal [8]. As a person becomes more or less stressed, the GSR increases or decreases respectively. Early work that investigated possibility of using GSR as a function of stress and cognitive activity can be found in [5]. More recent research has also linked GSR readings to cognitive activity [1] and established credible

Level	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.

Table 1. Cognitive Load Levels (L)ow, (M)id, (H)igh.

Figure 1. Subject-side set-up.

correlation between stress and cognitive functions [4]. All these works provides us with a theoretical basis for the use of GSR to measure cognitive load and its variations.

There has been a notable increased interest in recent years in using GSR as an objective measure of usability in HCI. Good reviews of such work have been presented in [2,3]. However, to the best of our knowledge, few have used GSR to evaluate cognitive load, especially in the evaluation of multimodal interfaces. The work presented in this paper intends to contribute to this area.

Experiment Design

A user experiment was designed and conducted in order to assess the reliability and accuracy of a number of potential cognitive load indicators. In this context, GSR was used as a control measure, to validate the effect of the cognitive load levels administered on the subjects. The application domain was traffic control management, where tasks involved marking incidents on a map and the deployment of resources. There were 36 unique tasks in total, 12 in each set. Each set of 12 tasks was completed with the same interface, but a different interaction condition: a) Multimodal interaction with hand gesture and speech input; b) Speech-only interaction; and c) Gesture-only interaction. In order to elicit natural multimodal interaction, the subjects were trained in all conditions during a preliminary session.

Four distinct levels of cognitive load were induced in each condition (3 tasks per level), by providing the subjects with tasks of increasing difficulty (see Table 1). The difficulty varied in five ways:

- Visual Complexity: The number of streets in each map increased from ~40 to ~60.
- Entities: The number of distinct entities in the task description was increased
- Actions to completion: The minimum number of actions required for task completion.
- Extras: The number of distracter entities (entities not needed for the task) increased
- Time Limit: The most difficult level was induced by implementing a time limit for completion.

In total, 11 random, renumerated subjects completed the experiment, though only the data for 5 have been analysed thus far. During the experiment, the subjects stood 2m in front of a wall-sized display (2x1.2m), as shown in Figure 1. A video camera capturing the upper body movements was placed on the right hand side. A second camera capturing the hand motion for automatic recognition was also on the right, closer to the subject. Subjective feedback was collected after the end of each load level. The subjects were required to rank the load experienced during the set of tasks, against other sets in the same condition. While not real-time, this control measure is assumed to provide a reliable indication of the level of difficulty perceived by the subject. GSR and blood volume pulse (BVP) were also monitored and recorded. Multimodal interaction patterns were later annotated.

The selected modalities for the experiment were speech, hand motion and a set of specific hand gestures, hence very prone to errors by automatic recognisers. We opted for a Wizard of Oz (WOz) implementation. Natural interaction patterns were expected to appear since the WOz technique allows intuitive input interaction while removing related limitations, such as input recognition error rates and misinterpreted semantic fusion of multimodal signals. Our wizard only performed manual recognition of speech and hand shapes, while hand motion detection was left to an in-house, video-based recognition module.

GSR Analysis

We used a GSR measurement apparatus from ProComp Infiniti System and Thought Technologies to collect the subjects' GSR during the experiment. The GSR was collected at a sample rate of 100Hz and measured in Siemens. In this study, we expected that:

E1: Users' GSR will be lower when using multimodal interface, compared to using unimodal interfaces,

E2: Users' GSR will show consistent increase when users perform tasks of increasing CL, across all interface conditions.

E3: There is a correlation between GSR reading variation and fine-grained interaction events, occurring during each task.

To validate E1 and E2, we calculated mean GSR, *mGSR*, across tasks, CL levels and all 5 subjects. As the time used to accomplish tasks of different cognitive load levels is different, we also used accumulated GSR, *accGSR*, to assess the total amount of stress or load experienced by a user during a task.

Figure 2 shows the mean GSR versus the interactive condition, that is, using multimodal interface or speech only interface, or gesture only interface. The *mGSR(modality)* is calculated as an average of all GSR values across all tasks, all cognitive load levels and all subjects. Though not statistically significant, it shows that the mean GSR for the use of multimodal interface (3.14 μSiemens) is lower than that for the use of speech-only interface (3.34 μSiemens), which in turn is lower than that for gesture-only interface (3.67 μSiemens), across the board.

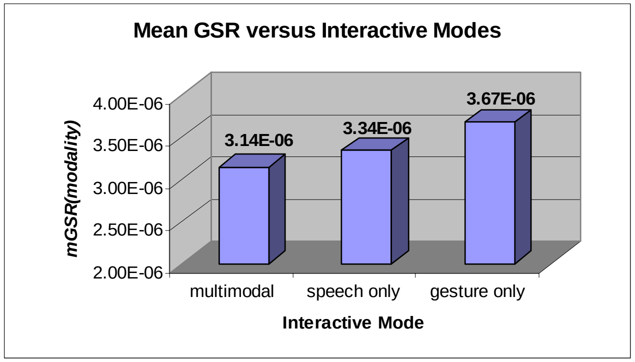

Figure 2. Mean GSR vs. Interactive Modes

Once analysis is complete, we expect the differences to attain significance. The fact that the mean GSR for gesture-only interface is higher than that for speech interface seems counter-intuitive. We reason that in the experimental set-up, the speech interface was fully wizard controlled, producing completely recognition-error free interactions. The gesture interface was not, real-time tracking was done with gesture tracking system with imperfect performance, which may cause some stress in users.

Figure 3 shows the mean GSR for the three interactive modes at three broad cognitive load levels (Low, Med and High, refer to Table 1). The *mGSR(modality, CL level)* is calculated as an average of all GSR values across all tasks of each broad load level (L, M, H) and across all subjects. It shows that the mean GSR increases for speech and multimodal modes when cognitive load level increases. However, for the gesture only interface, the mean GSR is highest when the cognitive load level is intended to be at its lowest. We offer the previously stated explanation: it can be challenging and stressful when user first uses automated hand gesture tracking, which in the case of the Low load level is always completed first. After a few tasks, users become more confident with gesture input, and the mean GSR drops and begins to increase again between Med and High levels.

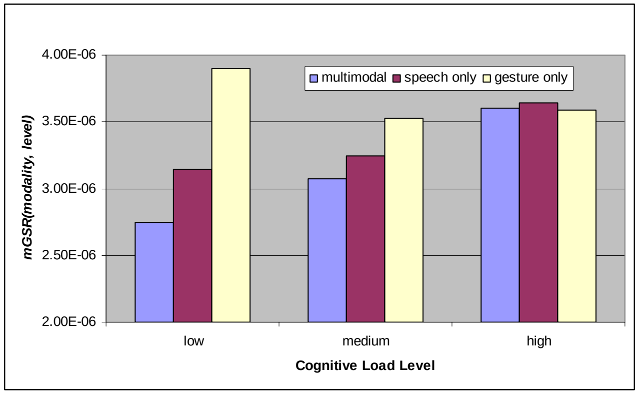

Figure 3. Mean GSR vs. CL and Interactive modes

Figure 4 shows the accumulated GSR for the three interactive conditions at three broad cognitive load levels. The *accGSR(modality, CL level)* is calculated as the sum of all GSR values across all tasks of each broad load level (L, M, H), and across all subjects:

$$accGSR(m,l) = \frac{1}{N_{subject}} \sum_{k=1}^{N_{subject}} \frac{1}{N_{task(level)}} \sum_{j=1}^{N_{task(level)}} \sum_{i=1}^{N_{time(j)}} GSR(m,l,k,j,i)$$

where the inner-most summation is over all GSR samples of a task, the middle averaging operation is over all tasks of a CL level, and the out-most averaging operation is over all subjects. We reason that the *accGSR* measure better reflects the total amount of cognitive load experienced by users. It is quite low for the Low level tasks when compared to the Med and High levels, because the time taken completing the tasks at this level is quite short. The *accGSR* value for multimodal mode at High is lower than that at Med because of the 90 seconds the time limit for High tasks, while the time for Med tasks is unrestricted and often exceeding 90 seconds, thus producing more cumulative cognitive load. An ANOVA test on *accGSR* values for Low, Med and High cognitive load levels, across all interactive modes, shows they are significantly different ($F=7.13$, $p=0.002$). Further analysis shows significant differences between Low and Med levels (two-tailed t-test, $p=0.0001$, <0.05) and between Low and High levels (two-tailed t-test, $p=1.29\times10^{-5}$, <0.05). This shows partial supports for E2, and evidence that GSR is capable of indicating cognitive load levels.

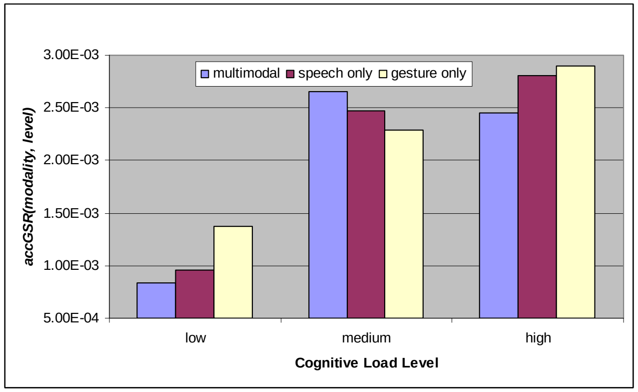

Figure 4. accGSR vs. CL and Interactive Modes

We then examined some typical GSR data sets for tasks. For example, Figure 5 shows a GSR data set for

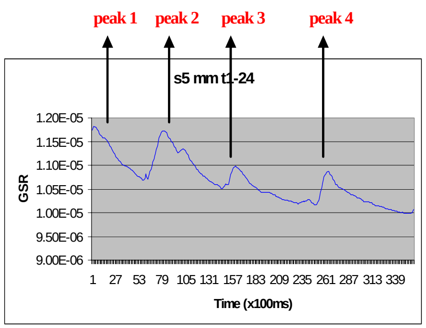

Figure 5. Typical GSR for a task

one multimodal task. We observe a combination of several large peaks (there are four in Figure 5) and several small peaks. GSR peaks usually indicate user frustration or stressful events. Despite the peaks, the GSR values decrease over time. After cross-examining this GSR data set with video recording of the subject performing the task and with multimodal annotation results, we can make some observations. Firstly, the large peaks in GSR readings seem to correspond to major interaction events during the task completion, for example, the first peak corresponds to time when the subject is reading task description. This is arguably the most stressful/cognitively loaded time. The other peaks in GSR correspond to the subject issuing various modal input constructions to complete sub-tasks. Secondly, after completing each sub-task towards the task goal, the subject becomes more confident and the GSR value gradually decreases, and reaches the lowest point (about 1E-05 Siemens) at the task completion time.

Conclusions and Future Work

We are investigating the relevance of GSR as an objective indicator of user's cognitive load and propose a number of GSR features that can provide further insights into the experienced level of cognitive load. Preliminary and partial analysis of GSR data from user experiments has shown that mean GSR across users increases as cognitive load increases. In addition, it suggests users experienced lower cognitive load levels when using a multimodal interface instead of a unimodal interface (such as speech-only interface or gesture-only interface).

For future work, we firstly would like to complete the GSR analysis for all 11 subjects who participated in the experiment, and perform further significance tests on the mean and accumulated GSR data. We also would like to explore in a more rigorous way the correlation between user's GSR variation and interactive behaviour, especially when using multimodal interfaces. We are developing a tool that will allow us to analyse, in real-time, changes in GSR and other modal measures.

References

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

[2] Lin, T., Hu, W., Omata, M. and Imamiya, A. (2005). Do Physiological Data relate To Traditional Usability Indexes? Proceedings of ozCHI 2005. Nov. 23-25, 2005. Canberra, Australia.

[3] 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.

[4] McEwen, B.S, Sapolsky, R.M. Stress and cognitive function. Journal of Current Opinion in Neurobiology, 1995.

[5] Miller, L.H, Shmavonian B.M. Replicability of two GSR indices as a function of stress and cognitive activity. Journal of Personality and Social Psychology. 1965 Nov;2(5):753-6.

[6] Oviatt, S., Coulston, R., and Lunsford, R. When Do We Interact Multimodally? Cognitive Load and Multimodal Communication Patterns. *In Proc. Int. Conf. on Multimodal Interfaces* (2004).

[7] Paas, F., et al. Cognitive load measurement as a means to advance cognitive load theory. Educational Psychologist (2003), 38, 63-71

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