Psychophysiological Assessment Tools for Evaluation of Learning Technologies

Richard H. Hall, Nicholas S. Lockwood, and Hong Sheng

Department of Business and Information Technology,
Missouri University of Science and Technology, Rolla, Missouri
{rhall,lockwoodn,hsheng}@mst.edu

Abstract. Research on the psychophysiological assessment of the impact of information technologies on humans is reviewed, with a particular focus on learning technologies and research carried out in in the Laboratory for Information Technology Evaluation (LITE) at Missouri S&T. Measures of arousal and valence are discussed first, including galvanic skin response (GSR), pupil dilation, and heart rate. This is followed by a discussion of the measurement of eye movement using eye tracking technologies. Lastly, a summary of the LITE lab research is provided. It is concluded that the measures are promising, based on these initial LITE lab results, though further work is needed to more accurately determine the appropriate constructs and contexts for optimizing the use of these tools.

Keywords: Learning Technologies, Psychophysiological Measurement.

1 Introduction

1.1 Psychophysiological Assessment of Learning Strategies

In recent years researchers have expanded their assessment tools for evaluating the impact of technologies on humans to include an array of psychophysiological measures, primarily measures of sympathetic nervous system activity such as heart rate, skin conductance, and pupil dilation. There are a number of reasons for this increased interest in these types of indices. First, subjective measures tend to be post-hoc, relying on learners' memories, or they interfere with experimental tasks, requiring a learner to interrupt an activity to provide self-report. Second, psychophysiological measures are particularly sensitive to the fleeting and non-conscious nature of emotional experience [1]. Third, advances in technology, have made data collection using psychophysiological measurement tools more efficient and user friendly [2]. There is a growing body of evidence that these tools can be more sensitive to some emotional experience than self report [3].

S. Yamamoto (Ed.): HIMI/HCII 2013, Part III, LNCS 8018, pp. 33–42, 2013.

1.2 Laboratory for Information Technology Evaluation

Missouri University of Science and Technology's Laboratory for Information Technology Evaluation (LITE) was founded in 1999. Since that time, LITE lab researchers have carried out a number of large and small-scale assessments of the impact of technologies on humans, with a particular focus on learning technologies. These projects have been sponsored by a number of external funding agencies such as the National Science Foundation; the U.S. Department of Defense, Department of Education, and Department of Energy; and the National Institute for Occupational Safety and Health. Results have been disseminated widely in the literature [4-7].

The LITE lab evaluations have been guided by a principal assumption that user responses are most accurately represented through the triangulation of data, based on multiple measurement and experimental approaches [4]. These measures have always included alternatives to traditional learning measures, such as indices of emotion. In recent years this has included a number of relatively new technologies, which allow for the detailed psychophysiological assessment of emotion, and the micro-behavioral measurement of eye tracking, as a measure of users' attentional focus.

2 Measures of Arousal and Valence

Russell [8] proposed that, fundamentally, emotion consists of two largely orthogonal dimensions, arousal and valance. Arousal represents emotional intensity and valence represents the pleasantness/unpleasantness of an event. Arousal is the dimension most commonly and most easily represented by psychophysiological measures, which measure the activity of the sympathetic branch of the autonomic nervous system. Increased sympathetic arousal results in the characteristic "fight or flight" response, in which many activities occur simultaneously to prepare an organism to flee or fight; including increased blood flow and heart rate to provide blood to skeletal muscles; increased skin temperature and perspiration to provide cooling; and increased pupil dilation to increase visual acuity [9]. Therefore, measures of heart rate, skin conductance, and pupil dilation can all serve as measures of sympathetic arousal. All of these measures demonstrate a human is aroused, though it is difficult to determine, without self-report and/or contextual information, whether or not this arousal represents a positive (e.g., engagement/excitement) or negative (e.g., fear) experience. Valence, on the other hand, is more difficult to measure via these psychophysiological measures, so researchers often use more traditional self-report measures to represent valence as a supplement to psychophysiological measures. However, there is some evidence that some of these psychophysiological measures can provide insight into emotional valence. This will be discussed in more detail below.

2.1 Galvanic Skin Response (GSR)

Overview. Measures of Galvanic Skin Response detect electrical conductance on the skin, caused by moisture, which is indicative of increased sympathetic nervous system activity. Increases in skin conductance are, thus, thought to represent increased emotional arousal [3, 10, 11]. There is some evidence that this arousal is linear, and that it can even provide some information on emotional valence [12].

LITE Lab Research. Measures of Galvanic Skin Response were first used in LITE lab research for the evaluation of a Virtual Reality System, used to train first responders for the aftermath of attacks utilizing weapons of mass destruction [13-15]. The rationale for using this measure of arousal was based on the importance of emotion for this type of "affectively intense learning" [14]. Psychologists have long pointed to the importance of the emotional congruence between learning activities and target performance [16, 17]. Therefore, one important criterion for a tool for teaching learners to perform effectively in emotionally charged environments, is that the learning itself evokes a significant emotional response.

In a pilot study of a desktop version of this virtual reality training system, participants' GSR was measured as they carried out the training. Consistent with expectations, GSR responses increased significantly during more stressful parts of the learning scenario, such as bomb explosions [13]. In a follow up evaluation, those who trained in the virtual reality system, which included emotional events such as bomb explosions, were compared to students who trained in a neutral environment, on learning outcomes. Although those trained in the emotional environment performed best in learning outcomes, GSR recorded during training did not significantly predict training group (emotional training versus neutral) [14].

GSR was also utilized in a more recent study, which consisted of an examination of users responses to web sites [6]. In this study, three types of web sites were evaluated, which differed systematically in degree of vividness and interactivity – low, medium, and high. In this research, users' GSR response again failed to significantly differ across experimental groups. However, the order of the means was interesting, in that those in the low and high vividness/interactivity group had the highest GSR readings, while the medium group had the lowest. It's possible that those in one of those groups were more frustrated, while those in the other were more engaged and excited, demonstrating the challenges of measuring valence via these types of measures, as discussed above.

2.2 Pupil Dilation

Background. Pupil size, as discussed, is a measure of sympathetic arousal, and has been found to predict emotional arousal in a number of studies [18, 19]. Interestingly, there is some evidence that pupil size can provide some insight into valence as well. Dilation size and duration appears to be greater [19] and varies more [20] for unpleasant stimuli.

One advantage of pupil dilation as a measure of emotion is that it appears to be particularly responsive to the granular nature of the emotional response. For example, in one study pupil dilation changed systematically with each successive digit presented in a short term memory task [21]. Due to the non-invasive nature of modern measurement tools, a second advantage of pupil dilation measures is their lack of interference during task performance.

One potential disadvantage of pupil dilation, as a measure of emotion, is that pupil size changes in response to other ambient stimuli as well, such as light intensity, color, and luminosity [22]. To combat this problem, researchers often control for variables such as dimensions of light, by keeping these variables consistent across experimental conditions [23, 24].

One other aspect of measures of pupil dilation is that they've been found to be particularly effective as measures of cognitive load [21, 25]. A number of studies in different contexts have found that pupil dilation increase with task demand or difficulty [21, 26, 27].

LITE Lab Research. The LITE lab carried out the evaluation of a Learning System called the "Rapid Development System", which is an educational technology developed to train Mechanical Engineers about control systems [28-30]. This project was a multi-year project and consisted of iterative evaluation during interface development, utilizing small sample usability testing, followed by evaluation of the system within the context of Mechanical Engineering classes.

As one part of this evaluation, cognitive workload was measured via pupil dilation. Interestingly, pupil dilation proved to be significantly related to learning style. Specifically, verbal learners exhibited a greater cognitive load while utilizing the learning tool than visually oriented learners. This makes intuitive sense, in that the system was graphically oriented, and made use of a number of animations, so it is quite possible that the more visual learners were more challenged and, subsequently, experienced a greater load. This is indirectly supported by the fact that verbal learners also found the interface more difficult to use, and scored lower on measures of perceived learning.

In other research carried out in the lab, researchers examined the impact of wait time and feedback on users responses to a web application [31]. In one group users were provided with feedback indicating a page was loading as they waited; whereas those in another group were not provided with this feedback. Consistent with expectations pupil dilation measures were significantly higher for those in non-feedback condition, representing frustration.

In another study, described above, LITE lab researchers carried out research on students' responses to web sites that differed in vividness/interactivity using a number of outcome measures [6]. In this study, of the three different psychophysiological measures of sympathetic activity utilized, pupil dilation was the only measure on which the groups significantly differed. Specifically, the group of students who experienced the high interactive/vivid web site demonstrated significantly greater measures of pupil dilation than those in the medium or low groups. Presumably, this was the result of the more engaging and arousing nature of the highly vivid and interactive site.

2.3 Heart Rate

Background. The final measure of sympathetic arousal that has been utilized in LITE lab research is heart rate. As with the other measures, heart rate increases and decreases with emotional arousal, however, there is substantial evidence that heart rate can also be used as a measure of emotional valence [3, 32]. First, while heart rate increases with all emotional arousal, the rate of deceleration is greater for responses to negative stimuli [33, 34]. The most promising aspect of heart rate as a measure of valence is the use of heart rate variability, as opposed simple increase/decrease in heart rate. Heart rate variability (HRV) is the oscillation of intervals between consecutive heartbeats. Multiple studies have found that higher HRV is associated with more pleasant emotional experience [35], and lower HRV is associated with unpleasant emotional experience [36, 37].

LITE Lab Research. Heart rate has been utilized in one study thus far conducted in the LITE lab, and this was in the form of heart rate variability (HRV) used to measure users' responses to web sites that differed in interactivity/vividness [6]. Although mean HRV increased systematically as expected, with the lowest level for those in the least vivid/interactive group and the highest level in the most vivid/interactive group, the mean differences among the groups were not significant.

3 Eye Movement

3.1 Background

Though pupil dilation is often measured via the same instruments, the measurement of eye movement is fundamentally different in comparison to the measures discussed above. First, it is not a measure of sympathetic nervous system arousal. Second, it does not represent what is traditionally viewed as a physiological response; rather it is a sort of micro-behavioral measure. However, it also differs from traditional behavioral measures, in that it provides information on non-conscious attention, can be collected without task interference, and is responsive to small gradations in time.

The measurement of eye movement is traditionally used to measure attention [38]. In fact, some researchers have suggested that eye movement is required for any complete theory of attention and visual processing [39, 40]. The measurement of eye movement has been used effectively in the field of psychology to provide insight into problem solving, reasoning, mental imagery and search strategies [38, 41, 42]. Eye tracking is thought to be particularly effective for providing information on moment-to-moment processing.

In order to better understand eye movement measurement, it's important to understand the way in which the eye works mechanically to scan the visual field. The eye views the world with a series of fixations and saccades. Although our visual perceptual experience is continuous, the actual physical activity of the eye is discrete, consisting of a series of momentary fixations connected by very quick movements referred to as saccades. When the eye fixates, it focuses on an object such that the object falls on the fovea which maximizes the perceptual fidelity of the object [43]. The duration of a typical fixation is 250 – 300 ms [44]. Saccades, on the other hand, are rapid eye movements with high acceleration and deceleration rates that last

30 – 120 ms [45]. Most modern eye trackers provide data based on fixations in one form or another as it is fixations, which presumably represent attention.

Typically eye-tracking data is represented visually in two ways. First, a “gaze plot” displays a series of fixations, in the form of circles, connected by lines that represent the saccades. The size of the dots represent the fixation duration, and numbers on the dots represent the order of the fixation. The gaze plot is typically used to represent the data from one individual, but data from a number of users can be combined to create a comprehensive gaze plot. On the other hand a “heat map” uses different colors, much as a weather map, where the colors differ based on the number and duration of fixations, such that an area of high fixation duration/number is represented as red, while green represents levels of low fixation duration/number. Heat maps typically represent the combined data from a number of users [46].

One of the most popular ways of utilizing eye tracking data quantitatively is through the use of “areas of interest” (AOI), which are researcher-defined areas within the visual stimulus. Typically, the total fixation time and/or the number of fixations within that area in proportion to other areas of the stimulus represent attention to the given area [28].

3.2 LITE Lab Research

Eye tracking was used extensively in the study described above, in which a Rapid Development System was evaluated [47]. The method was used in combination with more traditional usability techniques to identify problematic aspects of the interface via a series of iterative evaluations that occurred during development. For example, eye tracking demonstrated that users were not even noticing an important pop-up message, nor did they scroll down to the end of an important help menu. Eye tracking data also identified aspects of the interface that lead to user frustration, based on rapid and extensive saccade movement [48].

In this same study, the quantitative analysis of areas of interest found that students classified a visual learners attended more to the interface animations than the more verbal learners. This finding is consistent with previous research that found visual learners attend more to the visual aspects of a learning technology interface [49].

Another LITE lab study that utilized eye tracking consisted of the examination of the effect of time pressure on task completion strategy [50]. In this experiment, participants examined financial information with or without time constraints. An area of interest analysis supported the research hypothesis that learners under time pressure would tend more to use a non-compensatory solution strategy, which was represented by their reliance on summary tables and other visual information, as opposed to those without a time constraint who attended more to written text.

Lastly, in the study describe above in which users were provided with feedback (or not) while waiting on a web page to load [31], eye tracking data indicated that those without feedback were significantly less focused during the page load time than those in the non-feedback condition. This finding is consistent with the finding mentioned above that the mean pupil dilation measure for the non-feedback group was significantly higher, representing their frustration and boredom.

4 Conclusions

Table 1 is a summary of the LITE lab research reviewed.

Table 1. Summary of LITE lab Research Results

Measure	Study	Construct	Results
GSR	Virtual Reality Pilot [13]	Emotional Intensity	*Significant interface element differences
	Virtual Reality & Affectively Intense Learning [14]	Emotional Intensity	No signification group differences
	Web Site Vividness-Interactivity [6]	Emotional Intensity	No significant group differences
Pupil Dilation	RPS evaluation [28]	Cognitive Work Load	*Significantly related to learning style
	Wait Time and Feedback [31]	Cognitive Work Load	*Significant group differences
	Web Site Vividness-Interactivity [6]	Interest/Engagement	*Significant group differences
Heart Rate	Web Site Vividness-Interactivity [6]	Interest/Engagement	No significant group differences
Eye Movement	Evaluation of RPS [28]	Attention	*Significant learning style differences
	Time Constraints and Problem Solution [50]	Attention	*Significant group differences.
	Wait Time and Feedback [31]	Attention	*Significant group differences

Taking all of these studies together, the techniques applied seem relatively promising; in that seven of the ten studies found some significant effects utilizing these tools, consistent with expectations. Pupil dilation appears to be particularly promising as a measure of arousal, in that significant effects were found in all three studies that utilized this measure. However, it's important to note that this is a relatively small number of studies, and the measures did not always predict and differentiate as anticipated. Future work is needed to determine, more specifically, which measures are most effective for representing given constructs; and in what contexts are they most effectively utilized.

References

1. Eckman, I.: An Argument for Basic Emotions. Cognition and Emotions 6(3), 169–200 (1992)
2. Hudlicka, E.: To Feel or Not to Feel: The Role of Affect in Human Computer Interaction. International Journal of Human Computer Studies 59(1-2), 1–32 (2003)

3. Mandryk, R.K., Inkpen, K., Calvert, T.: Using Psychophysiological Techniques to Measure User Experience with Entertainment Technologies. Behaviour and Information Technology 52(2), 141–158 (2006)
4. Hall, R.H., Philpot, T., Hubing, N.: Comprehensive Assessment of a Software Development Project for Engineering Education. Journal of Learning, Technology, and Assessment 15(5), 4–42 (2006)
5. Luna, R., et al.: A GIS Learning Tool for Civil Engineers. International Journal of Engineering Education 26, 52–58 (2010)
6. Sheng, H., Joginapelly, T.: Effects of Web Atmospheric Cues on Users' Emotional Responses in E-Commerce. Transactions on Human-Computer Interaction 4(1), 1–24 (2012)
7. Truemper, J.M., et al.: Usability in Multiple Monitor Displays. ACM SIGMIS Database 39, 74–89 (2008)
8. Russell, J.: A Circumplex Model of Affect. Journal of Personality and Social Psychology 39(6), 1161–1178 (1980)
9. Brodal, P.: The Central Nervous System: Structure and Function. Oxford University Press, New York (2004)
10. Chanel, G.C., et al.: Boredom, Engagement and Anxiety as Indicators for Adaptation to Difficulty in Games. In: Proceedings of the International Conference on Entertainment and Media in the Ubiquitous Era. ACM Press, Tampere (2008)
11. Ward, R., Marsden, P.: Physiological Responses to Different Web Page Designs. International Journal of Human-Computer Studies 59(1-2), 199–212 (2003)
12. Lang, P.J.: The Emotion Probe: Studies of Motivation and Attention. The American Psychologist 50(5), 372–385 (1995)
13. Hall, R.H., et al.: Virtual Terrorist Attack on the Computer Science Building: A Research Methodology. Presence Connect 4(4), 12–15 (2004)
14. Wilfred, L.M., et al.: Training in Affectively Intense Learning Environments. In: Proceedings of the E-Learn Conference. AACE Press (2004)
15. Hilgers, M.G., et al.: Virtual Environments for Training First Responders. In: Proceedings of the Interservice/Industry Training, Simulation, and Education Conference, I/ITSEC (2004)
16. Ellis, H.C., Moore, B.A.: Mood and Memory. In: Dalgleish, T., Powers, M. (eds.) Handbook of Cognition and Emotion, pp. 193–210. Wiley, Hoboken (1999)
17. Eich, E., Macaulay, D.: Are Real Moods Required to Reveal Mood-Congruent and Mood-Dependent Memory? Psychological Science 11(3), 244–248 (2000)
18. Partala, T., Surakka, V.: Pupil Size Variation and an Indication of Affective Processing. International Journal of Human-Computer Studies 59(1-2), 185–198 (2003)
19. Partala, T.M., Jokiniemi, M., Surakka, V.: Pupillary Responses to Emotionally Provocative Stimuli. In: roceedings of the Symposium on Eye Tracking Research and Applications. ACM Press, Palm Beach Gardens (2000)
20. Janisse, M.P.: Pupil Size, Affect and Exposure Frequency. Social Behavior and Personality 2(2), 125–146 (1974)
21. Kahneman, D., Beatty, J.: Pupil Diameter and Load on Memory. Science 154(3756), 1583–1585 (1966)
22. Beatty, J., Lucero-Wagoner, B.: The Pupillary System. In: Handbook of Psychophysiology, pp. 142–162. Cambridge University Press, New York (2000)
23. Bradley, M.M., et al.: The Pupil as a Measure of Emotional Arousal and Autonomic Activation. Psychophysiology 45(4), 602–607 (2008)
24. Franzen, P., et al.: Sleep Deprivation Alters Pupillary Reactivity to Emotional Stimuli in Healthy Young Adults. Biological Psychology 80(3), 300–305 (2009)

25. Hyona, J., Tommola, J., Alaja, A.: Pupil Dilation as a Measure of Processing Load in Simultaneous Interpretation and Other Language Tasks. The Quarterly Journal of Experimental Psychology 48(3), 598–612 (1995)
26. Payne, D.T., Parry, M.E., Harasymiw, S.J.: Percentage of Pupillary Dilation as a Measure of Item Difficulty. Perception and Psychophysics 4, 139–143 (1968)
27. Wright, P., Kahneman, D.: Evidence for Alternative Strategies of Sentence Retention. Quarterly Journal of Experimental Psychology 23, 197–213 (1971)
28. Chintalapati, A., et al.: Evaluation of a Rapid Development System Using Eye Tracker. In: Proceedings of the American Association of Engineering Education (2010)
29. Tang, L., et al.: Development and Initial Analysis of a Mini CNC Rapid Development System. In: Proceedings of the American Association for Engineering Education (2010)
30. Tang, L., Landers, R.G.: Remote Use of a Linear Axis Rapid Development System. In: Proceedings of the American Association of Engineering Education, pp. 2010–2157 (2010)
31. Sheng, H., Lockwood, N.S.: The Effect of Feedback on Web Site Delay: A Perceptual and Physiological Study. In: Proceedings of the Workshop on HCI Research in MIS, Shanghai, China (2011)
32. Anttonen, J., Surakka, V.: Emotions and Heart Rate While Sitting on a Chair. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM Press, Portland (2005)
33. Bradley, M.M., Lang, P.J.: Affective Reactions to Acoustic Stimuli. Psychophysiology 37(2), 204–215 (2000)
34. Brosschot, J.F., Thayer, J.F.: Heart Rate Response is Longer After Negative Emotions than Positive Emotions. International Journal of Psychophysiology 50(3), 181–187 (2003)
35. Nolan, R.: Heart Rate Variability (HRV). European Heart Journal 17, 354–381 (1996)
36. Jonsson, P.: Respiratory Sinus Arrhythmia as a Function of State Anxiety in Healthy Individuals. International Journal of Psychophysiology 63(1), 48–54 (2007)
37. Brosschot, J.F., Van Dijk, E., Thayer, J.F.: Daily Worry is Related to Low Heart Rate Variability During Waking and the Subsequent Nocturnal Sleep Period. International Journal of Psychophysiology 63(1), 39–47 (2007)
38. Just, M.A., Carpenter, P.A.: Eye Fixations and Cognitive Processes. Cognitive Psychology 8, 441–480 (1976)
39. Findlay, J.M., Gilchrist, I.D.: Visual Attention: The Active Vision Perspective. In: Vision and Attention. Springer, New York (2001)
40. Jacob, R.J.K.: Eye Tracking in Advanced Interface Design, pp. 258–288. Oxford University Press, New York (1995)
41. Yoon, D., Narayanan, N.H.: Mental Imagery in Problem Solving: An Eye Tracking Study. In: Proceedings of the ACM Symposium on Eye Tracking Research & Applications. ACM Press (2004)
42. Zelinsky, G.J., Sheinberg, D.: Why Some Search Tasks Take Longer Than Others: Using Eye Movements to Redefine Reaction Times. In: Findlay, J., Kentridge, R., Walker, R. (eds.) Eye Movement Research: Mechanisms, Processes and Applications. Elsevier, Amsterdam (1995)
43. Pashler, H.E.: The Psychology of Attention. MIT Press, Cambridge (1998)
44. Salvucci, D.D., Goldberg, J.H.: Identifying Fixations and Saccades in Eye-Tracking Protocols. In: Proceedings of the Symposium on Eye Tracking Research and Applications. ACM Press (2000)
45. Palmer, S.: Vision Science: Photons to Phenomenology. MIT Press, Cambridge (1999)

46. Duchowski, A.T.: Eye Tracking Methodology: Theory and Practice. Springer-Verlag, Inc., London (2003)
47. Chintalapati, A., et al.: Evaluation of a Rapid Development System Using Eye Tracker. In: Proceedings of the American Association of Engineering Education, AC 2010-2210 (2010)
48. Goldberg, J.H., Kotval, X.P.: Computer Interface Evaluation Using Eye Movements: Methods and Constructs. International Journal of Industrial Ergonomics 24, 631–645 (1999)
49. Tsianos, N., et al.: The Learning Styles as a Basic Parameter for the Design of Adaptive E-Learning Environments. In: Proceedings of the International Conference on Open and Distance Learning Applications of Pedagogy and Technology (2005)
50. Sheng, H., Pochinapeddi, S.: User Pressure: A Psychophysiological Analysis of the Effect of Temporal Constraints of Information Processing and Decision Making. In: Proceedings of the International Conference on Human-Computer Interaction. Springer, Las Vegas (2013)