The Evaluation of a Physiological Data Visualization Toolkit for UX Practitioners: Challenges and Opportunities

Vanessa Georges
HEC Montréal
Montréal, Canada
vanessa.georges@hec.ca

François Courtemanche
HEC Montréal
Montréal, Canada
francois.courtemanche@hec.ca

Sylvain Sénécal
HEC Montréal
Montréal, Canada
sylvain.senecal@hec.ca

Pierre-Majorique Léger
HEC Montréal
Montréal, Canada
pml@hec.ca

Lennart Nacke
University of Waterloo
Waterloo, Canada
lennart.nacke@acm.org

Marc Fredette
HEC Montréal
Montréal, Canada
marc.fredette@hec.ca

ABSTRACT

The objective of this paper is to outline the challenges we have encountered during the development and evaluation of our Physiological Data Visualization (UX Heatmap) toolkit and discuss the opportunities that emerged from our experience. The main goal of the Physiological Data Visualization toolkit is to allow simpler and richer interpretation of physiological signals for UI evaluation, in order to reduce the barriers associated with the use of physiological measures in the fields of user experience design and research. Following a user test with 11 UX experts from the industry, we were able to better understand how and in which contexts they would use the proposed UX Heatmap toolkit in their practice.

Author Keywords

Interface design; heatmaps; physiological computing; affective computing, toolkits.

ACM Classification Keywords

H.2.1Design; Experimentation; HCI; Human Factors; Measurement; User interfaces. I.3.6 Computer graphics: Methodology and Techniques.

INTRODUCTION

Measuring the emotional state of users during the interaction is essential to the design of richer user experience. Users' emotional and cognitive states can be inferred using physiological signals such as electrodermal activity, heart rate, eye tracking, and facial expressions [1-2]. These measures can provide important temporal information to UX experts as to what the user is experiencing throughout the interaction without retrospective or social desirability bias [3]. However, these measures are still difficult to contextualize and interpret as they are not specifically associated with user behavior or interaction states. Physiological signals also require a certain degree of interpretation, as outputs need to be processed in order to transition from raw data to useful actionable insights. The objective of the proposed UX Heatmap toolkit is to address these issues and help UX experts incorporate physiological data in their user tests.

TOOLKIT OVERVIEW

Traditional gaze heatmaps are used in eyetracking as intuitive representations of aggregated gaze data [4]. Their main use is to help researchers and UX experts answer the question: "Where in the interface do people tend to look?" [5]. In the proposed UX Heatmap toolkit, the users' gaze now serves as a mean of mapping physiological signals onto the user interface. The resulting heatmaps represent the physiological signals' distribution over the interface, and can help answer the following question: "Where in the interface do people tend to emotionally or cognitively react more strongly?"

Toolkits aim at facilitating and fostering the adoption of emerging and complex technologies to new and non-expert populations [6]. Our toolkit aims to tackle the following challenges regarding the use of physiological measures in the prototyping and evaluation of user interfaces by UX experts: (1) data synchronization of multiple signals from various apparatus (e.g., visual attention using eyetracking and arousal using electrodermal activity), (2) valid physiological inferences (e.g., taking into account different signal latencies), and (3) the interpretation and contextualization of data using multiple types of visualizations. We thus designed a visualization toolkit that contextualizes physiological and behavioral signals to facilitate their use, see figure 1 [7].

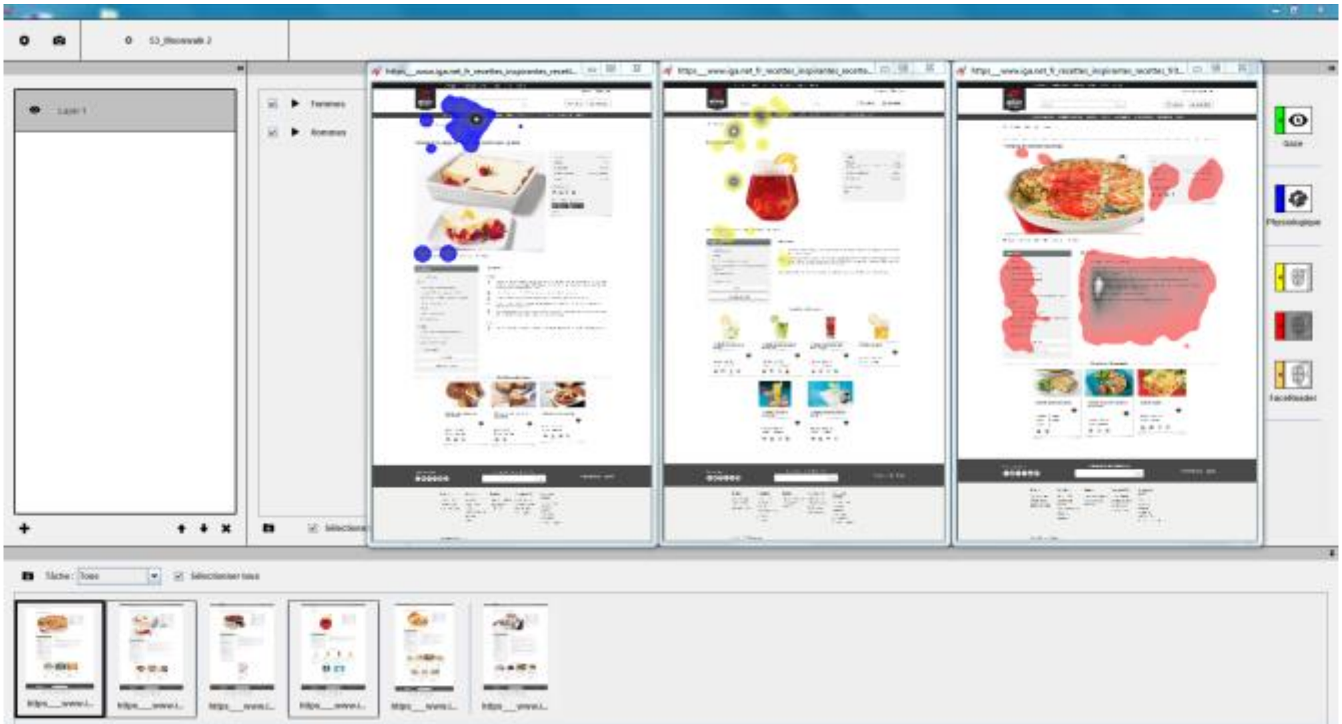

Figure 1. Screenshot of the UX Heatmap Toolkit Software Interface.

The Motives behind the Toolkit

Our recent work with industry has lead us to question a major discrepancy between industry and academic practices: while physiological measures are increasingly used in research, the adoption of these methods as UX evaluation tools remain uncommon in industry. From our experience interacting with the industry, there is a growing demand for more quantitative user research to provide data-driven recommendations based on physiological data such as eye tracking and emotional reactions. We therefore wanted to understand what can be done to facilitate their adoption in the industry.

Furthermore, physiological measures, in combination with traditional methods, can help practitioners to better evaluate the emotional dimension of user experience, which focuses on emotional responses triggered by the system interaction [8-10], as they each provide complementary information on how users feel about a system, game, or web interface [11]. While traditional evaluation methods can offer episodic data, i.e., before or after the interaction, physiological measures can provide moment-to-moment information [12]. For example, the addition of physiological measures can help practitioners identify the cognitive and emotional reactions users experienced using an interface, while a post-task interview can help delve further, after having identified these emotions. Therefore, this toolkit was built for UX experts, whose needs differ from those of academic researchers. While the former's needs are to analyze and adequately communicate findings to improve user experience, the latter's interest resides in the validation and understanding of phenomena, based on hypotheses [13].

PREVIOUS WORKS

Although there exists a series of HCI evaluations toolkits, the majority do not include physiological measures. Most usability toolkits simply guide the user toward the right tool to use for his or her context. However, certain toolkits do integrate physiological measures, tackling one aspect of the problem at hand. Most researchers have concentrated their efforts on finding ways to measure physiological signals and interaction states synchronously. For example, Kivikangas et al. [14] have developed a triangulation system to interpret physiological data from video game events. Dufresne et al. [15] have proposed an integrated approach to eyetracking-based task recognition as well as physiological measures in the context of user experience research. Other researchers have also developed tools that allow users' to manually assign subjective emotional ratings on visual interfaces [16] or to visualize emotional reactions in terms of GUI widgets [17]. While these research streams have produced interesting results, they are not easily transferable to new contexts of use, as they are based on internal information from the interactive system (e.g., video game logs, application events, or areas of interest). So, we looked to package and streamline this process and provide an integrated solution for UX experts.

THE EVALUATION OF THE TOOLKIT

To evaluate the proposed UX Heatmap toolkit, we conducted a study with eleven UX practitioners and consultants. Each interview lasted about one hour and a half, following a variation on the think aloud protocol, cooperative evaluation [18]. As such, participants were asked to talk through what they were doing; the interviewer taking on a more active role by asking questions along the way (e.g., 'Why?' 'What do you think would happen?').

UX experts were also asked to complete a user testing evaluation report using the toolkit. The PowerPoint report included a study summary, a research scenario and qualitative data. Participants were first briefed on the task at hand (i.e., complete a UX report for an online grocer using the UX Heatmap toolkit), before going through the partially completed report with the interviewer, to put them into context and get a sense of what was required of them. Participants had to complete two PowerPoint slides. The physiological signal data set used for the evaluation was collected in a previous study [19]. Practitioners were asked to: (1) generate and select data visualizations to include in their report using the toolkit, (2) interpret the results, and (3) provide recommendations to the client. The remainder of the time was used to discuss of the advantages and disadvantages physiological measures as a UX evaluation method, as well as the toolkit itself. This evaluation task was added to help UX experts integrate the information on physiological measures quickly and effectively, and to give them a concrete opportunity to use the toolkit in order to envision themselves using it in their own practice.

WHAT WE LEARNED

The goal behind the creation of the Physiological Data Visualization toolkit was to bring physiological measures to UX practitioners. Below are some of the findings we uncovered during the evaluation process of this toolkit.

Having a clear understanding of: **(1) the constraints and limitations of potential user's,** as this may limit the extent to which these users will consider the toolkits as a viable addition to their practice. The way in which researchers process and use information is quite different than the way practitioners do. For example, the automatization of certain functionalities (i.e. participant and layer creation), which we saw as superfluous, or nice-to-haves, were seen as crucial by practitioners in order to accelerate the interpretation of the visualizations generated with our toolkit. For researchers, this represent a minimal gain; as for practitioners faced with short development cycles, this is seen as saving, both time and money, two important attributes when choosing new methodologies.

(2) potential users' context of use, as this may limit the extent to which novice practitioners and researchers will consider these technologies as a viable addition to their projects. When asked about their intent to reuse the tool, ten out of eleven practitioners interviewed stated that they would use the toolkit in their practice. When inquired further, six of them declared that their use of the tool would depend on the projects, using it only in the interventions where emotions are an important component or if clients specifically requested them to use physiological signals. This could translate into a steep and ever present learning curve, as practitioners must re-learn how to use the toolkit and materials associated with the data collection of physiological signals at each use. As a result, it could be challenging for practitioners to master the toolkit if only used occasionally. Therefore, a barrier to entry exist for professionals unable to justify the financial investment due to sparse usage of such tools (i.e., eyetrackers, sensors).

Additional Challenges related to physiological measure usage by UX experts: **reliance on external sources and tools can become an issue.** For non-expert users, learning how to correctly take ownership of a new technology using a toolkit is already an endeavor; the addition to peripheral techniques and methods creates a much higher learning curve than they may have initially anticipated. Although the Physiological Data Visualization toolkit makes physiological measures more accessible to UX practitioners by addressing the interpretation of signals, there remains a need to better educate professional on some of the more technical aspects of physiological measurements. Data collection, experimental setup, and data extraction still have to be overseen by UX experts, representing an important time and financial constraints.

CONCLUSION

The objective of this paper was to highlight the challenges we have encountered during the development and evaluation of our Physiological Data Visualization toolkit and discuss the opportunities that emerged from our experience. Designing, building and sharing toolkits are a great way to bring new technologies to the larger HCI community. Understanding the end user is primordial to the development of a toolkit that will truly reach its goal and audience, which we uncover through the evaluation of out toolkit, which makes discussions on the methods and metrics that can be used to evaluate toolkits all the more interesting. By pursuing our working with potential users during the design and development phases of the toolkit, our objective is to continue to reduce the barriers associated with the use of physiological measures in the fields of user experience design and research.

ACKNOWLEDGMENTS

Authors want to thank Brendan Scully for manuscript revision and the UX practitioners who participated to this study This work was supported by the CRSNG *(Conseil de recherches en sciences naturelles et en génie du Canada).*

REFERENCES

1. Zhihong, Z., Pantic, M., Roisman, G.I. and Huang, T.S. 2009. A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 1: 39-58.
2. Calvo, R.A. and D'Mello, S. 2010. Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. *Affective Computing, IEEE Transactions on* 1, 1: 18-37.

3. King, M.F. and Bruner, G.C. 2000. Social desirability bias: A neglected aspect of validity testing. *Psychology and Marketing* 17, 2: 79-103.
4. Nielsen, J. and Pernice, K. *Eyetracking Web Usability*. New Riders, Berkeley, California, 2012.
5. Wooding, D.S. Fixation maps: quantifying eye-movement traces *Proceedings of the 2002 symposium on Eye tracking research & applications*, ACM, New Orleans, Louisiana, 2002, 31-36.
6. Cartwright, William, et al. "Geospatial information visualization user interface issues." *Cartography and Geographic Information Science* 28.1 (2001): 45-60.
7. Georges, Vanessa, et al. "UX heatmaps: mapping user experience on visual interfaces." *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2016.
8. Février, F., N. Gauducheau, E. Jamet, G. Rouxel, and P. Salembier, *La prise en compte des affects dans le domaine des interactions humain-machine*. Le travail humain, 2011. **74**: p. 183-201.
9. Mahlke, S. and M. Minge, *Consideration of Multiple Components of Emotions in Human-Technology Interaction*, in *Affect and Emotion in Human-Computer Interaction*, C. Peter and R. Beale, Editors. 2008, Springer Berlin Heidelberg. p. 51-62.
10. Hassenzahl, M., *The Interplay of Beauty, Goodness, and Usability in Interactive Products*. Human–Computer Interaction, 2004. **19**(4): p. 319-349.
11. Roto, Virpi, Marianna Obrist, and Kaisa Väänänen-Vainio-Mattila. "User experience evaluation methods in academic and industrial contexts." *Proceedings of the Workshop UXEM*. Vol. 9. 2009.
12. Zhihong, Z., Pantic, M., Roisman, G.I. and Huang, T.S. 2009. A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 1: 39-58.
13. Roto, V., Vermeeren, A.P.O.S., Väänänen-Vainio-Mattila, K., Law, E. and Obrist, M., Course notes: User Experience Evaluation Methods - Which Method to Choose? in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (Paris, France, 2013), ACM.
14. Kivikangas, M., Nacke, L. and Ravaja, N. 2011. Developing a triangulation system for digital game events, observational video, and psychophysiological data to study emotional responses to a virtual character. *Entertainment Computing* 2, 1: 11-16.
15. Dufresne, A., Courtemanche, F., PromTep, S. and Sénécal, S., Physiological Measures, Eye Tracking and Task Analysis to Track User Reactions in User Generated Content. in *Measuring Behavior*, (Eindhoven, The Netherlands, 2010), 218-222.
16. Huisman, G., Hout, M.v., Dijk, E.v., Geest, T.v.d. and Heylen, D. LEMtool: measuring emotions in visual interfaces *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, Paris, France, 2013, 351-360.
17. Cernea, D., Weber, C., Ebert, A. and Kerren, A., Emotion scents: a method of representing user emotions on gui widgets. in *IS&T/SPIE Electronic Imaging*, (2013), International Society for Optics and Photonics, 86540F-86540F-86514.
18. Dix, Alan. *Human-computer interaction*. Springer US, 2009.
19. Desrocher, C., et al. "The influence of product type, mathematical complexity, and visual attention on the attitude toward the website: The case of online grocery shopping." *Fourteenth Pre-ICIS SIG-HCI Workshop, Fort Worth, TX*. 2015.